Hydrology in the Humid Tropic Environment (Proceedings of a symposium held at Kingston, Jamaica, November 1996). IAHS Publ. no. 253, 1998.

A program for evaluating rainfall erosivity indices for tropical conditions

RICHARD COOKE

Department of Agricultural Engineering, University of Illinois at Urbana Champaign, 1304 W Pennsylvania Avenue, Urbana, Illinois 61801, USA

TERRENCE JOHNSON

Roy F. Weston/REAC, 2890 Woodbridge Avenue, Building 209 Annex, Edison, New Jersey 08837-3679, USA

JOSEPH LINDSAY

CARDI, University Campus, Box 113, Mona, Kingston 7, Jamaica

Abstract Erosivity indices are used to quantify the erosive potential of rainfall. Some of these indices have been specifically developed for erosion research under tropical conditions. This program can be used to evaluate indices based on the product of rainfall energy and characteristic intensity ($E{\cdot}I_t$), the product of rainfall amount and characteristic intensity ($A{\cdot}I_t$), or on threshold kinetic energy values ($KE > v$). Indices can be evaluated for individual rainstorms or for seasonal rainfall. In addition to the evaluation of erosivity indices, the program can also be used for digitizing existing rainfall charts to create a database of breakpoint rainfall, and for generating data sequences in which the correlation structure between rainfall amount and duration and the erosivity indices are preserved. Any of 18 commonly used probability distribution functions can be used to represent each rainfall variable. The program can be run on DOS-based computers with at least 256 K ROM. Use of the program's digitizing options requires a digitizing pad or similar device that can be accessed from an RS232 serial port.

Programa para evaluar índices de erosión de lluvias en zonas tropicales

Resumen Los índices de erosión se usan para cuantificar el potencial erosivo de las lluvias. Algunos de estos índices han sido desarrollados específicamente para zonas tropicales. Este programa se puede usar para evaluar índices de erosión que se basan o bien en el producto de la intensidad característica de la lluvia por su energía ($E{\cdot}I_t$) o por su cantidad ($A{\cdot}I_t$), o en valores límites de energía cinética ($KE > v$). Los índices se pueden evaluar para tormentas individuales o para todas las lluvias de una estación. El programa también se puede usar para digitalizar gráficos de lluvias, información que se puede almacenar en una base de datos, para generar secuencias de datos donde se preserva la correlación entre las propiedades de una lluvia y su índice de erosión y para representar cada una de las variables de la lluvia con una de las 18 distribuciones probabilísticas más comunes. El programa funciona en computadoras con sistema operativo DOS con al menos 256 K de ROM. Para usar las opciones de digitalización se necesita un digitalizador o similar. El digitalizador puede conectarse al puerto en serie RS232.

INTRODUCTION

Accelerated erosion and sedimentation pose serious threats to water resources, sustainable agriculture, and to the general quality and stability of the environment. Much accelerated erosion in tropical regions may be attributed to highly erosive rainstorms that are randomly distributed in space and time. Several indices have been developed to quantify the erosive potential of these rainstorms. However, there are many factors that limit the applicability of these indices, particularly the drudgery involved in calculating erosivity indices, and the unavailability of rainfall intensity measurements. This paper describes a microcomputer-based program developed to minimize the effects of the two factors mentioned above. This program was developed for IBM type computers, with a digitizing tablet or similar device required for some program options.

Erosivity indices

Rainfall erosivity is a measure of the erosive potential of rainfall. The most common erosivity index is the $E{\cdot}I_{30}$ index developed by Wischmeier & Smith (1978) for use in the Universal Soil Loss Equation (USLE). This index is a product of the kinetic energy (E) of the rainfall and the peak 30-min rainfall intensity. It can be evaluated for a single rainfall event, or values can be accumulated to yield seasonal or annual erosivity indices. In light of the high intensities and short durations of many tropical rainstorms, the maximum 30-min intensity is often replaced by the maximum 15- or 7.5-min intensities, giving rise to the $E{\cdot}I_{15}$ and the $E{\cdot}I_{7.5}$ indices, respectively. Other researchers have developed other indices that are based on the product of rainfall amount (A) and characteristic peak intensities. The most commonly used of these indices are the $A{\cdot}I_{30}$, $A{\cdot}I_{15}$, and the $A{\cdot}I_{7.5}$ indices.

Each of the erosivity indices mentioned above requires the evaluation of a characteristic peak rainfall intensity, and rainfall amount. Rainfall intensity can only be obtained from rainfall stations with recording raingauges. Only a small percentage of stations in tropical rainfall networks have such gauges. Of those that do, most of the historic rainfall records are stored on strip charts. These charts, one of which is shown in Fig. 1, are sometimes difficult to interpret, and working with them is always a tedious process.

COMPUTER PROGRAM

The program was written for use with DOS-based and Windows-based computers. The program consist of routines for digitizing rainfall charts, evaluating rainfall intensity parameters, calculating single event or seasonal erosivity indices, evaluating the correlations between indices, and generating synthetic rainfall or erosivity data.

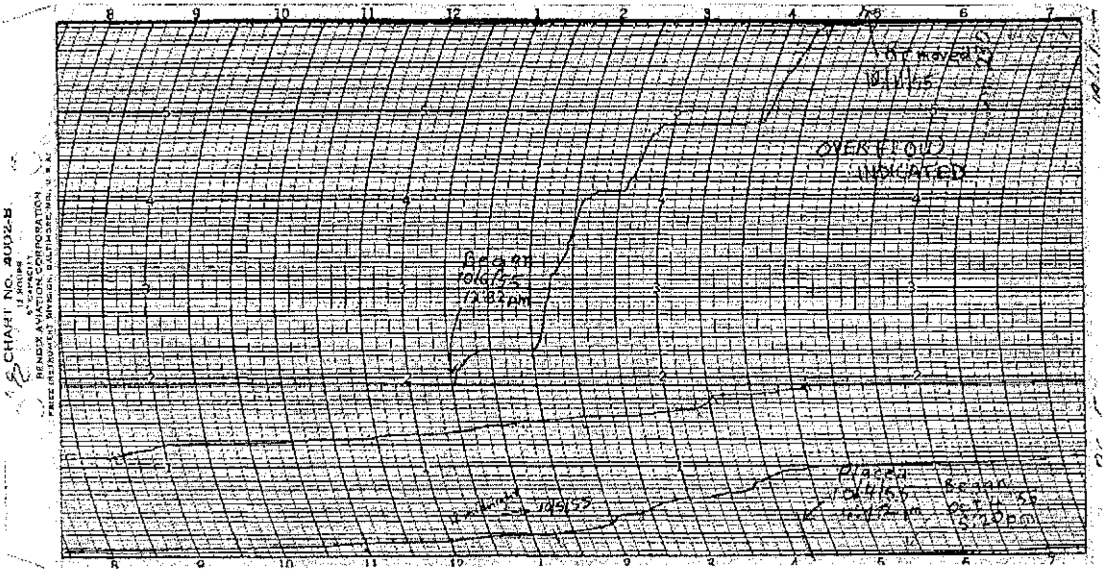

Fig. 1 Sample chart from a recording (weighing) raingauge.

Digitizing routine

The digitizing routine is set up for a CalComp DrawingBoard III, or a SUMMAGRAPHICS bit pad with an RS232 interface and with default communication protocol as shown in Table 1. However, it can be adjusted at runtime to match the actual configuration of the digitizer being used, as long as the connection is through an RS232 interface. Modified versions of the program for use with other digitizers can be obtained from the authors. The routine can be used with siphoning or weighing type gauges. Records from tipping bucket gauges are normally stored in digital forms, so it was unnecessary to include an option for these types of gauges. The digitized data is stored in ASCII format and may be exported to other applications. The data recorded for each rain event includes the year and the Julian day on which the event occurred, and cumulative precipitation along with the corresponding time.

The digitizing routine has been used extensively to process rainfall charts from recording stations in Trinidad and Tobago. On average, it reduces the time required to process a single rainfall event from three hours to two minutes. It has also been modified for digitizing runoff charts.

Table 1 Default digitizer communication protocol.

Location COM1
Parity: even
Stop bits: 2
Baud rate: 2400
Data bits: 7

Intensity parameters

Intensity parameters are calculated from the break point data produced by the digitizing routine. Alternatively, data from other sources can also be used if they conform to the format produced by the digitizing routine. Spline functions are fitted to the breakpoint data, facilitating the evaluation of peak intensities for periods ranging from 5 to 30 min. The data are also used to evaluate rainfall energy, itself a function of rainfall intensity and duration (Wischmeier & Smith, 1978).

Generating synthetic rainfall and erosivity data

The system used for rainfall/erosivity generation is divided into three stages, each programmed as a stand-alone module. Two of the modules are general-purpose modules that can be used in other simulation exercises, while the third relates specifically to rainfall generation. The stages, in order of progression, are:

(a) the generation of multivariate Gaussian (normal) data with the same correlation structure as the rainfall/erosivity properties;

(b) the transformation of each vector of the multivariate Gaussian data to the actual distribution chosen for the corresponding property; and

(c) the incorporation of the resulting data into a generated sequence of rainfall.

Generation of multivariate Gaussian data Consider the $n \times n$ correlation matrix Ω whose (i, j)th element is the correlation between *Property i* and *Property j*. Using Cholesky decomposition, Ω may be decomposed into the product:

$$\Omega = \mathbf{LU} = \mathbf{LL}^T \tag{1}$$

where $\mathbf{L}$ is a lower triangular $n \times n$ matrix and $\mathbf{U}$ is the transpose of $\mathbf{L}$. If $\mathbf{Z}$ is generated so that each row is formed through the relation:

$$\mathbf{Z}_R = \mathbf{L}\varepsilon^T \tag{2}$$

where ε is a vector of uncorrelated standard Gaussian variates, that is:

$$\varepsilon \sim \text{Gaussian } [0, 1] \tag{3}$$

then each of the vectors in $\mathbf{Z}$ conforms to a standard Gaussian distribution and the correlation matrix for the vectors is approximately equal to Ω (Cressie, 1991). This method can, therefore, be used to generate correlated multivariate data with a specific correlation structure. This is the procedure used in the first stage of the rainfall generator. The length of the vectors in $\mathbf{Z}$ is dependent on the number of realizations required.

Problems that can often arise include the specification of a correlation matrix that is not positive-definite, or one for which the matrix decomposition procedure is numerically unstable. Nonpositive-definite matrices can result from the use of subjective correlation information. Consider, for example, the generation of rainfall amount and intensity data for a number of adjoining raingauges in a network, one of which is a nonrecording gauge. Any correlation involving the intensity of rainfall at this gauge has to be subjective, since there are no intensity data for this location.

Quimby (1986), found that if the correlation matrix is written as:

$$\Omega = \mathbf{M} + \mathbf{I} + \mathbf{M}^T \tag{4}$$

where **M** has zeros on and above the diagonals, and **I** is the identity matrix, then

$$\Omega \approx \Omega^* = (\mathbf{M} + \mathbf{I})(\mathbf{M} + \mathbf{I})^T = \lambda\lambda^T \tag{5}$$

with the error in the approximation being $\mathbf{MM}^T$. **L** can be approximated by λ and there are no numerical errors associated with the determination of λ. This approximation can be determined even if Ω is not positive definite, in which case **L** would be indeterminate. One of the main inconsistencies in Quimby's approximation is that as long as the *i*th variable in the correlation matrix (*i* greater than one) is correlated with any of the other variables, the *i*th diagonal element in Ω^* is identically greater than one, that is, the correlation between the *i*th variable and itself exceeds one. Such a correlation value may be a mathematical convenience but it is a physical impossibility. Cooke (1993) modified the Quimby approximation to remove these impossible correlations. In this modification, **L** is approximated by λ^*, where:

$$\lambda^*_{i1} = \lambda_{i1} \tag{6}$$

$$\lambda^*_{ij} = \lambda_{ij}\left(\frac{1-\lambda^2_{i1}}{\sum_{k=2}^{i}\lambda^2_{ik}}\right)^{\frac{1}{2}} \quad 1 < j \leq i \tag{7}$$

In $\Omega^{**} = [\lambda^*][\lambda^*]^T$ there are no impossible correlations and all the diagonal elements are one.

A matrix which requires the least amount of global change to make Ω positive definite is given by:

$$\Omega_M = \Omega^{**} + \omega(\Omega - \Omega^{**}) \tag{8}$$

where Ω_M is a positive definite matrix that is closer in value to Ω than any other, and ω is a scalar quantity. For any specified correlation matrix, ω can be adjusted until the matrix just becomes positive definite. If the matrix is already positive definite then ω would be one. Since all the diagonal elements in Ω and Ω^{**} are one, the diagonal elements in Ω_M are always equal to 1.0 for all values of ω. The algorithm used in the rainfall generator returns the value of ω only in cases where it is not equal to one.

Transformation of multivariate Gaussian data The second stage of the rainfall generator involves the transformation of the correlated multivariate Gaussian data generated in the first stage. The matrix **Z** is transformed into the matrix **S** which has approximately the same correlation structure, but in which the vectors do not necessarily conform to Gaussian distributions. This transformation is effected by the use of the Taylor-Bender transform (Taylor & Bender, 1988, 1989), which utilizes

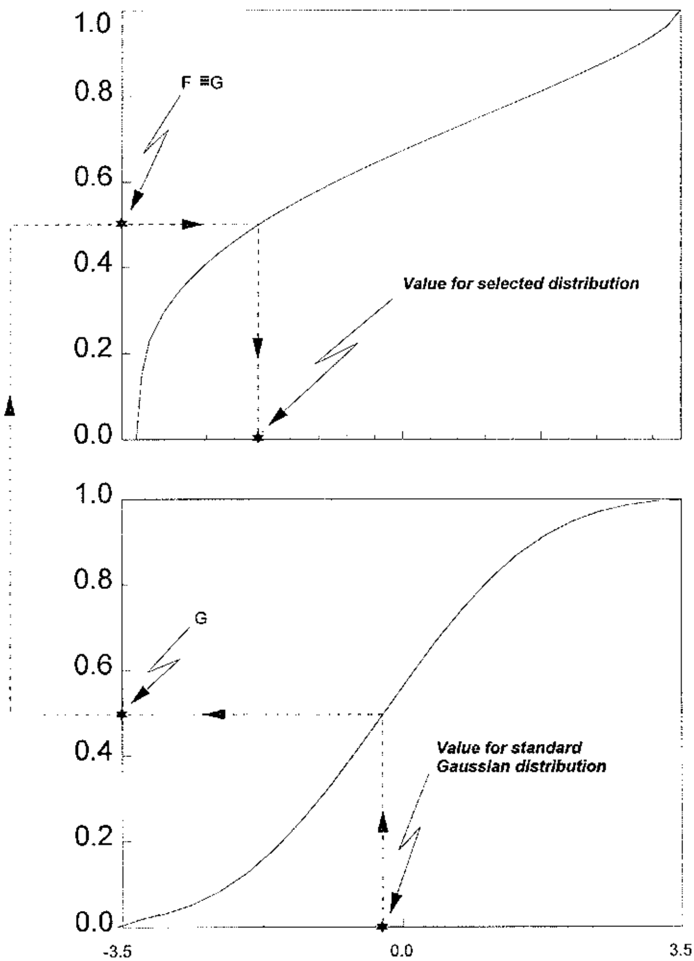

Fig. 2 Schematic diagram of the Taylor-Bender transform. In this application, the variate is being transformed from Standard Gaussian space.

the fact that the cumulative density function, F(x), is ~U[0, 1], regardless of the nature of the corresponding density function, f(x). If a random variable, x, that is distributed according to g(x) is to be transformed into a variable, y, distributed according to f(y), then:

$$y = \mathrm{INV}_f[\mathrm{G}(x)] \tag{9}$$

where INV[] represents the inversion of a cumulative distribution function, and the subscript, f, on the inverse transform function indicates that the inversion is done so that the resulting data is distributed according to f(x). Data from the standard Gaussian distribution may be transformed into any other distribution as shown schematically in Fig. 2.

The routine includes code to transform standard Gaussian data into data conforming to any of 18 homogeneous distributions or 162 mixtures of any two of these distributions. The homogeneous distributions that can be used are shown in

Table 2 Homogeneous distributions used in generator.

Gaussian (normal)
Log Gaussian (log normal)
Three parameter log Gaussian (log normal)
Exponential
Shifted exponential
Beta
Gamma
Three parameter gamma (Pearson type III)
Log Pearson type III (log gamma)
Inverted gamma (Pearson type V)
Gumbel (extreme value type I) for minima
Gumbel (extreme value type I) for maxima
Frechet (extreme value type II) for minima
Frechet (extreme value type II) for maxima
Three parameter Frechet for maxima
Weibull (extreme value type III) for maxima
Weibull (extreme value type III) for minima
Three parameter Weibull for minima

Table 2. These distributions where selected because they have been previously used in hydrological applications. Cooke *et al.* (1993) developed a microcomputer based routine for fitting these 18 distribution functions. The inclusion of mixture (heterogeneous) distributions is prompted by the nature of rainfall. In some locations, rainfall can result from convective or frontal processes. Ostensibly, a given distribution for a rainfall property, could be decomposed into two overlapping distributions. The representation of rainfall properties by heterogeneous distributions might lead to more accurate simulation.

Sequencing of rainfall properties Rainfall is sequenced by using a two-stage Markov chain process to determine sequences of events and nonevents. This involves the specification of four probabilities:

$$P(E/N) = \alpha \tag{10}$$

$$P(N/N) = 1 - \alpha \tag{11}$$

$$P(E/E) = \beta \tag{12}$$

$$P(N/N) = 1 - \beta \tag{13}$$

where P(*E*/*N*) is the probability of an event following a nonevent, P(*N*/*N*) is the probability of a nonevent following a nonevent, P(*E*/*E*) is the probability of an event following an event, and P(*N*/*N*) is the probability of a nonevent following a nonevent. Additionally, a flag is required to specify the definition of the conditioning event for each variable. For example, in the generation of rainfall amounts and durations for several stations in a network, the occurrence of rainfall at a particular location might either be conditioned by the occurrence of rainfall at that location on the previous day, or on the occurrence of rainfall at another location on the same day. The occurrence of a non zero value for rainfall duration at a location would be conditioned by the occurrence of rainfall at that location, with α and β being 0.0 and 1.0, respectively.

For each day of the simulation, the four probabilities are used to form a vector, **v**, whose elements are either one or zero depending on whether or not an event occurred for the corresponding property. The row of the matrix **S** corresponding to that day is modified by multiplying each element by the corresponding element in **v**:

$$\mathbf{S}_{R,i} = \mathbf{S}_{R,i} \times \mathbf{v}_i \tag{14}$$

The elements of **v** are regenerated for each day of the simulation.

Output data By the end of the third stage of the generation process, each vector in **S** now represents a sequence of a rainfall property or erosivity index over the duration of the simulation. The distribution of each property is exactly as specified in the input, while the correlations between the properties approximate the input conditions.

REFERENCES

Cooke, R. A. (1993) Development of phenomenologically-based distribution fitting procedures and spatial processes for mixed population soil properties. Unpublished PhD Dissertation, Virginia Tech., Blacksburg, Virginia, USA.

Cooke, R. A., Mostaghimi, S. & Woeste, F. (1993) VTFIT: a microcomputer-based routine for fitting probability distribution functions to data. *Appl. Engng Agric.* **9**, 401–408.

Cressie, N. (1991) *Spatial data analysis*. Wiley and Sons, New York, USA.

Quimby, W. F. (1986) Selected topics in spatial statistical analysis: nonstationary vector kriging, large scale conditional simulation of three dimensional Gaussian random fields, and hypothesis testing in a correlated random field. Unpublished PhD Dissertation, Department of Statistics, University of Wyoming, Laramie, Wyoming, USA.

Taylor, S. E. & Bender, D. A. (1988) Simulating correlated lumber properties using a modified multivariate normal approach. *Trans. Am. Soc. Agric. Engrs* **31**, 182–186.

Taylor, S. E. & Bender, D. A. (1989) A method for simulating multiple correlated lumber properties. *For. Products J.* **39**, 71–74.

Wischmeier, W. H. & Smith, D. D. (1978) *Predicting Rainfall Erosion Losses—A Guide to Conservation Planning.* Agriculture Handbook 537, USDA.