

## Regularized joint inverse estimation of extreme rainfall amounts in ungauged coastal basins of El Salvador

Michael J. Friedel

Received: 12 April 2007 / Accepted: 28 August 2007 / Published online: 1 November 2007  
Springer Science+Business Media B.V. 2007

**Abstract** A regularized joint inverse procedure is presented and used to estimate the magnitude of extreme rainfall events in ungauged coastal river basins of El Salvador: Paz, Jiboa, Grande de San Miguel, and Goascoran. Since streamflow measurements reflect temporal and spatial rainfall information, peak-flow discharge is hypothesized to represent a similarity measure suitable for regionalization. To test this hypothesis, peak-flow discharge values determined from streamflow recurrence information (10-year, 25-year, and 100-year) collected outside the study basins are used to develop regional (country-wide) regression equations. Peak-flow discharge derived from these equations together with preferred spatial parameter relations as soft prior information are used to constrain the simultaneous calibration of 20 tributary basin models. The nonlinear range of uncertainty in estimated parameter values (1 curve number and 3 recurrent rainfall amounts for each model) is determined using an inverse calibration-constrained Monte Carlo approach. Cumulative probability distributions for rainfall amounts indicate differences among basins for a given return period and an increase in magnitude and range among basins with increasing return interval. Comparison of the estimated median rainfall amounts for all return periods were reasonable but larger (3.2–26%) than rainfall estimates computed using the frequency-duration (traditional) approach and individual rain gauge data. The observed 25-year recurrence rainfall amount at La Hachadura in the Paz River basin during Hurricane Mitch (1998) is similar in value to, but outside and slightly less than, the estimated rainfall confidence limits. The similarity in joint inverse and traditionally computed rainfall events, however, suggests that the rainfall observation may likely be due to under-catch and not model bias.

**Keywords** El Salvador Monte Carlo Nonuniqueness Prediction Prior information Extreme rainfall Regionalization Return periods Recurrence intervals Regularization Uncertainty Ungauged

---

M. J. Friedel (✉)  
Crustal Imaging and Characterization Team, United States Geological Survey,  
Denver Federal Center, MS 964, Box 25046, Lakewood, CO 80225, USA  
e-mail: mfriedel@usgs.gov

## Notations

$A$	Drainage area in $\text{km}^2$
CN	Runoff curve number for a tributary basin
$d_r$	Vector of preferred regional dissimilarity between tributary basin parameter values
$d_a$	Vector of actual regional dissimilarity between tributary basin parameter values
$h$	Vector ( $n$ -dimensional) of observation values (recurrent peak-flow discharge values determined externally for areas regression equations)
$h_0$	Current model outputs
$j$	Regularization observations
$j_0$	Initial set of regularization constraints (soft prior information)
$I$	Identity matrix
$I_a$	Initial abstraction
$L_t$	Lag time
$L$	Hydraulic length of tributary basin
$L$	Operator indicating regularization observations acting on parameter set
$M$	Operator that maps parameter space to the space of observations
$M(p)$	Vector of simulated recurrent tributary basin peak-flow discharge and rainfall values determined using a rainfall-runoff model
$P$	Vector of parameter (rainfall-runoff curve numbers) and boundary condition (recurrent rainfall amounts) values being estimated
$p_0$	Current parameter values
$p - p_0$	Parameter upgrade vector
$p_j$	Rainfall-runoff parameter value in tributary basin $j$
$p_k$	Rainfall-runoff parameter value in tributary basin $k$
$p_i$	Prior information
$P$	Rainfall in mm
$q_{um}$	Unit peak discharge ( $\text{m}^3/\text{km}^2/\text{mm}$ )
$Q$	Diagonal weight matrix used in the measurement objective function
$R$	Depth of runoff in mm
$S$	Maximum potential difference between rainfall and runoff (tributary basin retention) in mm
$S$	Relative weight matrix used in the regularization objective function
$w_{ri}$	Regularization weights
$X$	Jacobian matrix of the operator
$X^T$	Transpose of matrix $X$
$Y$	Tributary basin slope
$Z$	Jacobian matrix of the operator
$\beta^2$	Regularization weight factor
$\lambda$	A diagonal matrix (Marquardt lambda) added to enhance inverse stability
$\Phi_m$	The model-to-measurement misfit described by a least-squares measurement objective function
$\Phi_r$	The model-to-regularization misfit described by a least-squares regularization objective function

## 1 Introduction

In tropical Latin America and other places of the world, monsoon storms and hurricanes are associated with rainfall events that result in natural disasters (extreme rainfall events), such as flooding, debris flows, and landslides (El Diario de 1999 British Broadcasting Corporation 2000, Crone et al. 2001). Of these natural disasters, extreme rainfall events result in flooding that accounts for about 15% of the total annual death toll worldwide (Hossain and Katiya 2006). Whereas the magnitude of rainfall can typically be determined from rain gauges, during hurricanes rain gauges are subject to inaccuracies due to undercatch, damage, and inaccessibility during flood situations (Curtis 2007). In developing countries, rainfall-induced flooding is more catastrophic (in terms of effects on human life and infrastructure) because river basins are sparsely or completely ungauged with no rainfall information available for design and/or warning based recurrent flood prediction (Hossain and Katiya 2006). Even where adequate surface gauge stations exist, the on-going and systematic decline of rainfall networks worldwide is recognized as a major concern for advancing flood-prediction related to extreme rainfall events, especially in basins that remain ungauged or are sparsely instrumented (Stoksa and Shikhlomanov et al. 2002). For these reasons, there is an immediate need for the development of new hydrological tools to determine the magnitude of extreme rainfall events and their effect on flooding in ungauged basins (Sivapalan 2003).

Hydrologists have developed a variety of numerical models to study and quantify the effects of rainfall on basin flooding (Bicknell et al. 1997, U.S. Army Corps of Engineers 1998, Rossmar 2005). The success of these numerical models to provide reliable answers, however, depends largely on the quality of model input information (such as, spatial and temporal rainfall measurements), conceptual model, and model parameterization (Friedel 2006a). Whereas rainfall information for model input in gauged basins is traditionally derived using storm events or intensity-duration-frequency curves (Cohn), this information is not commonly available in ungauged basins. In some cases, remote satellite sensing was used to derive rainfall relations for ungauged basins, however, scaling and calibration to local field measurements could be problematic (Laks 2004). Recently, multivariate statistical approaches (principal components and cluster analyses) were used to identify homogeneous regions from which rainfall hyetographs could be derived for all of Taiwan (Lin et al. 2005). Whereas multivariate statistical approaches show promise for regionalizing rainfall, the development of regional hyetographs using this approach requires information from many (on the order of 100 or more) spatially distributed rainfall gauges.

Assuming that rainfall information is available and conceptual basin model reasonable, model parameterization proceeds by supplying parameter values that control the rate exchange of simulated fluxes within a basin. Since most basin model parameters are unknown and/or abstract conceptual representations of physical phenomena, model parameter values must be either assigned from other sources of information (such as neighboring basins) or indirectly determined through some form of model estimation process (Friedel 2006a). Assigning model parameter values from neighboring basins to the ungauged basins being studied requires the application of a regionalization-based similarity (or dissimilarity) measure (Bloschl and Sivapalan 2005). One regionalization-based similarity measure is the parameter-value covariance; the equivalent measure of dissimilarity is the parameter-value variogram. Regionalization of basin models directly using the parameter-value covariance (or variogram) typically results in poor predictive performance

because of the inability to account for the underlying uncertainty associated with the so-called known model parameter values (Merz and Bloesch 2004).

Recently, the usefulness of regionalization based on variations in a regularized joint inverse approach were demonstrated for application to model calibration and prediction of streamflow in gauged basins (Doherty and Johnson 2003, Friedel 2006a, and Conined (Elliot et al. 2005) and unconined (Friedel, in press) peak-flow discharge and flood-flow depths in ungauged basins. In these applications, the regularized joint calibration of multiple basin and/or tributary basin models was conducted in which their estimated parameter values were permitted to deviate from a user-supplied parameterization scheme (soft prior information), as long as sufficient information content existed in the measurement constraints (that is, the constraint-information content facilitated improvements in the parameter values). In the study by Friedel (in press), normalization of the conceptualization and parameterization process, while maximizing the information content used to constrain the calibration process, facilitated the estimation of optimal (but not unique) parameter values from which the underlying prediction uncertainty of peak-flow discharge was quantified and validated.

In the regularized joint inverse approach presented by Friedel (in press), at least part of the peak-flow prediction uncertainty is associated with the ungauged tributary basin rainfall amounts supplied to the models. One alternative to supplying the ungauged tributary basin models with uncertain rainfall amounts is to estimate rainfall amounts as part of the model calibration process. In this article, the regularized joint inverse approach presented by Friedel (in press) is extended based on the hypothesis that peak-flow discharge measurements represent a similarity measure suitable for regionalization of extreme rainfall amounts of varying return periods (recurrent rainfall). To test the hypothesis, multiple tributary basin rainfall-runoff models are simultaneously calibrated using annual-maximum peak-flow values for the 10-year, 25-year, and 100-year recurrence intervals from tributary basins outside the study basins, and soft prior information in the form of spatial basin relations, to constrain the process. Since the estimated rainfall amounts are not dependent variables, a calibration-constrained Monte Carlo approach is used to quantify their limits of nonlinear uncertainty. Validation of the regularized joint inverse approach is done quantitatively by comparison of rainfall estimates based on traditional rainfall-frequency relations derived for recurrent intervals, and qualitatively by comparing the cumulative rainfall observed during Hurricane Mitch at a tributary gauge to the corresponding nonlinear rainfall prediction limits.

## 2 Regionalization of tributary basins parameter values and recurrent rainfall amounts

Estimation of the ungauged tributary basin rainfall-runoff parameters follows the regularized joint inverse modeling approach presented by Elliot et al. (2005), and Friedel (in press). Extension of this approach to incorporate estimation of basin rainfall is facilitated through modification of the Gauss-Marquardt-Levenberg parameter estimation algorithm to accommodate the incorporation of regional rainfall relations as soft prior information (regularization) and synthetic peak-flow recurrence information. The various mathematical elements used in the estimation of rainfall amounts for different return periods are presented in the following sections.

### 2.1 Gauss–Marquardt–Levenberg parameter estimation

In applying the Gauss–Marquardt–Levenberg parameter estimation algorithm (Cooley and Naff 1990), the estimation process seeks to minimize the difference between simulated and measured values described by some measurement objective function. In this study, the measurement objective function is described using a least-squares approach given by

$$\Phi_m = \frac{1}{2} \mathbf{M}(\mathbf{p}) - \mathbf{o}^T \mathbf{W} \mathbf{M}(\mathbf{p}) - \mathbf{o}, \tag{1}$$

where  $M$  is an operator that describes the action of a model under parameter estimation conditions (that is, the operator that maps an  $m$ -dimensional parameter space to the space of the  $n$  observations that are available as measurement-constraint information during the estimation process),  $\mathbf{p}$  is the  $m$ -dimensional vector that represents model parameter values;  $\mathbf{M}(\mathbf{p})$  is the vector of simulated values determined using a model; the  $n$ -dimensional vector that represents measurement observation values;  $\mathbf{W}$  is the matrix transpose; and  $\mathbf{W}$  is a weight matrix in which each diagonal element is proportional to the inverse of the squared error associated with the corresponding measurement. Also, the observations are oftentimes comprised of one or more types of measurement information (Friedel 2006) that are preprocessed to ensure: (1) homoscedascity (Bates and Granger 2006), (2) statistical independence of measurement noise (Kuczynski 1983), and (3) that information content associated with the measurement-constraints exert sufficient influence in the estimation process (Doherty and Johnston 2003; Friedel 2006a).

When minimization of the measurement objective function results in an optimal parameter set, the upgrade vector  $\mathbf{p} - \mathbf{p}_0$ , can be calculated directly by

$$\mathbf{p} - \mathbf{p}_0 = \frac{1}{2} [\mathbf{X}^T \mathbf{W} \mathbf{X} + \lambda \mathbf{I}]^{-1} \mathbf{X}^T \mathbf{W} \mathbf{e}_0, \tag{2}$$

where  $X$  is the Jacobian matrix (each row is comprised of sensitivities for a particular model output for which there is a corresponding field measurement) with respect to all elements of  $\mathbf{p}$ . These sensitivities are calculated at current parameter values, represented by  $\mathbf{p}_0$  for which corresponding model outputs are  $\mathbf{e}_0$ . For nonlinear models, the initial solution is typically not optimal and therefore requires the sensitivities  $X$  to be recalculated on the basis of the newly estimated parameter set. This process is then repeated until there is convergence to a user-specified minimum objective function value. To prevent instability of the parameter upgrade vector, the  $\mathbf{W} \mathbf{X}$  matrix of Eq. 2 is supplemented by the addition of a diagonal term comprised of the Marquardt parameter, multiplied by the identity matrix,  $\mathbf{I}$ .

### 2.2 Regularization procedure

In formulating the traditional estimation problem, convergence of Eqs achieved through observance of the principle of parsimony (Hill 1998). One problem with adhering to this principle is that there is no a priori means of knowing how many parameter values can be estimated. In this case, estimation of too few parameter values may result in the inability to predict anything of consequence, whereas estimation of too many parameter values relative to the number and/or limited information content (temporal and spatial variation of system stresses reflected in the measurements) in the constraint-measurements may result in an unstable solution where  $\mathbf{W} \mathbf{X}$  becomes singular, or near-singular (Doherty and Johnston 2003; Friedel 2005). Perhaps more importantly, the traditional

estimation approaches are not amenable to the solution of complex nonlinear inverse problems (such as those involving simultaneous calibration of multiple models and/or relations among parameters). One advantage of using the Gauss-Marquardt-Levenberg approach is the ability to incorporate regularization for solution of nonparsimonious models.

Regularization is a generalized mathematical framework that facilitates overcoming instability of the inverse problem through addition of observations (prior information). The explicit use of additional observations in which they do not change is during the inversion is direct prior information, whereas the implicit introduction of additional observations that may change values or relations as part of the estimation process is soft prior information. If the regularization observations, are represented by an operator, acting on the parameter set, the regularization relations can be represented by

$$L\hat{\phi} = r. \tag{3a}$$

The linearized form of this regularization operator can then be written as

$$Zp = r, \tag{3b}$$

where  $Z$  is the Jacobian matrix of the regularization operator if the regularization observations are given sufficient weight in comparison with the observation weights provided in  $W$ , the inverse problem is considered well-posed (Doherty).

This inverse problem is now iteratively solved for the parameter values using the modified upgrade equation given by

$$p = p_0 + [X^T W X + \beta^2 Z^T S Z + \lambda I]^{-1} [X^T W \hat{d} + r_0 + \beta^2 Z^T S Z \hat{r}], \tag{4}$$

where  $r_0$  represents the right side of Eq. (a) (when current parameter values are substituted for  $p$  in this equation);  $S$  is a relative weight matrix assigned to the regularization observations, and  $\beta^2$  is a regularization weight factor. The matrix supplied by the user and in many cases will be a set of weights applied individually to the regularization observations, such that those with greater weight are more rigidly enforced; in this case,  $S$  is a diagonal matrix.

The upgrade equation, Eq. (4), constitutes a minimization problem (deGroot-Hedlin and Constable 1990) in which a least-squares regularization objective function can be defined by

$$\Phi_r = \frac{1}{2} \hat{\phi}^T R \hat{\phi} + r^T S \hat{\phi} - r, \tag{5}$$

In solving the constrained inverse estimation problem, the regularization function is minimized subject to the constraint that  $\hat{\phi}_m$  rises no higher than a user-specified value, referred to as the target measurement objective function value. In this way, the user controls the level of model-to-measurement misfit required in the regularized inversion process. If the target measurement objective function cannot be achieved, the regularized inversion process simply minimizes  $\hat{\phi}_m$ . Where minimization of  $\hat{\phi}_m$  would otherwise be an unstable process due to parameter nonuniqueness, the stability of this process is maintained by seeking that set of parameters lying within the global minimum region that also minimizes  $\Phi_r$ . In either case, the regularization weight factor  $\beta^2$  operates as a Lagrange multiplier associated with the constrained minimization problem. In the algorithm used herein, the regularization weight factor is recalculated every iteration of the nonlinear parameter estimation process using a bisection algorithm (Doherty).

2.2.1 Observed and simulated recurrent peak-flow discharge in tributary basins

Traditionally, information used to constrain the nonlinear parameter estimation process is derived from streamflow observations recorded at gauge stations in the basin of interest (Doherty and Johnston 2003; Friedel 2006a). The lack of measured streamflow observations in the ungauged tributary basins therefore requires an alternative source of information to constrain the parameter estimation process. To derive information suitable for constraining the parameter estimation process, the annual-maximum peak-flow discharge in tributary basin streams is assumed to be runoff dominated due to the comparatively short time of concentration. Since the strongest correlation for runoff among basin attributes is the basin area (Seibert 1999), the observations can therefore reflect annual-maximum peak-flow values for the 10-year, 25-year, and 100-year recurrence intervals generated using regional (country-wide) regression equations. Each regional regression equation is fit to respective recurrent tributary peak-flow discharge values previously computed based on fitting the Pearson Type-III distribution to the logarithms of historical annual flood peaks (Weaver and Gamble 1993) recorded mostly at tributary gauge stations outside the study basins.

In contrast to the observed peak-flow discharge values used to constrain the joint nonlinear parameter estimation process, the simulated peak-flow discharge values for each tributary basin are computed simultaneously using separate rainfall-runoff models (U.S. Army Corps of Engineers 1999). The rainfall-runoff model components are based on mathematical relations intended to represent simplified meteorology, hydrology, and hydraulic processes in the tributary basins. For example, the runoff for a single tributary basin is computed using the empirical relation (Soil Conservation Service 1955) given by

$$R = \frac{P - I_a}{10} \left( \frac{P - I_a}{S} \right)^2 \tag{6}$$

where  $R$  is the depth of runoff in mm,  $P$  is the rainfall in mm,  $I_a = S * 0.2$ , and  $S$  is the maximum potential difference between rainfall and runoff (tributary basin retention) in mm.

Various tributary basin characteristics and model parameter values are required as rainfall-runoff model input to compute the meteorological and hydrologic processes. For example, rainfall return period amounts, duration, and cumulative distribution functions collectively define the input hyetograph that is used in all runoff calculations. Since these model input represent average values that are assumed to be distributed uniformly over each tributary basin, the assumption of uniform character over a tributary basin is anticipated to become less accurate with greater variance for larger tributary basins. Examples of tributary basin characteristics derived using a digital elevation model and geographic information system include area, length, and slope. In addition to tributary basin characteristics, tributary basin model parameter values (runoff curve numbers) that control the magnitude of simulated peak-flow discharge also must be provided.

The model curve numbers are characterized as lumped parameter values that account for runoff potential based on land use and soil type in the tributary basins. Other hydrologic response parameters that must be computed for use in the rainfall-runoff model are an implicit function of curve number that includes the lag time and initial abstraction given by

$$L_t = L^{0.8} \left( \frac{C_n - 1}{1900} \right)^{0.5} \tag{7}$$

$$S \leq \left\{ \frac{1000}{CN} \right\} \leq 10, \quad (8)$$

where  $L_t$  is the lag time, in hours;  $L_h$  is the hydraulic length of a tributary basin, in meters;  $S$  is the tributary basin slope, dimensionless;  $I_0$  is the initial abstraction, in mm; and CN is the dimensionless tributary basin runoff curve number.

Following the nonlinear solution of Eq. (7), the simulated peak discharge,  $q_p$ , is then computed for comparison to the measured values using

$$q_p = q_{um} A_m R, \quad (9)$$

where  $q_{um}$  is the unit peak discharge ( $\text{m}^3/\text{km}^2/\text{mm}$ ), and  $A$  is the drainage area in  $\text{km}^2$ .

### 2.3 Predictive uncertainty

In this study, the uncertainty associated with model parameter values and extreme rainfall amounts is computed using a calibration-constrained Monte Carlo approach (Friedel 2006). As the name implies, the calibration-constrained Monte Carlo approach is an inverse-based approach (as opposed to the conventional forward-based Monte Carlo approach) analogous to the generalized likelihood uncertainty estimation approach described by Beven and Binley (1992). In contrast to the forward-based Monte Carlo approach, the calibration-constrained Monte Carlo approach quantifies the uncertainty associated with parameter nonuniqueness that primarily arises as a consequence of insufficient information content available to constrain the nonlinear solution space during the model parameter estimation process. In this case, the lack of information content in measurement observations used to constrain the solution space results in many alternate parameters sets (nonuniqueness) that calibrate the rainfall-runoff model equally well; that is, the ending global objective function minimum is the same for the model calibrated using a number of alternate sets of initial parameter values. To ensure that the maximum range of uncertainty is explored, many alternate starting calibration parameter sets (derived by random sampling of uniform parameter distributions bounded by physically-based constraints) are used (Fig. 1). Following determination of a statistically significant number of estimated parameter sets, the nonlinear range uncertainty is determined and probability computed.

### 3 Case study

In 1998, Hurricane Mitch passed through the Republic of El Salvador resulting in flooding and landslides that killed 374 and displaced 55,864 persons (Helin 1999). Twelve of the country's 14 municipalities suffered significant damage (\$50M), but the hardest-hit areas were in the low-lying regions, particularly those associated with four basins that originate in the central mountains and extend to the coastal plain areas where they discharge to the Pacific Ocean (from west to east; Fig. 2): Paz (2,647  $\text{km}^2$ ), Jiboa (1,717  $\text{km}^2$ ), Grande de San Miguel (2,250  $\text{km}^2$ ) and Goascoran (1,750  $\text{km}^2$ ). In these four coastal basins, about 67,582 ha (163,000 acres) were flooded (including thousands of homes, businesses, seaport, and crops) and 15 major bridges were damaged or destroyed.

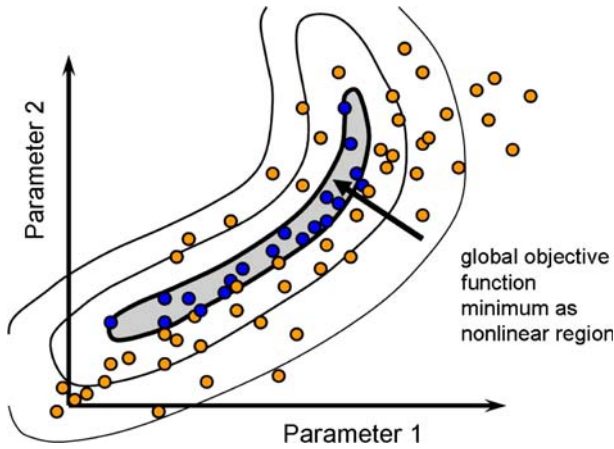


Fig. 1 Simplified illustration of predictive analysis using calibration-constrained Monte Carlo inverse approach

Both the Paz and Goascoran river basins share trans-country boundaries; for example, the Paz River serves as a border between El Salvador and Guatemala (925 km<sup>2</sup> in El Salvador and 1,722 km<sup>2</sup> in Guatemala), and the Goascoran River serves as a border between El Salvador and Honduras (1,315 km<sup>2</sup> in El Salvador and 435 km<sup>2</sup> in Honduras). Land use and vegetation in these coastal basins includes natural (deciduous brush, grassland, and pasture) and anthropogenic (cultivated areas of coffee, cotton, and sugar cane)

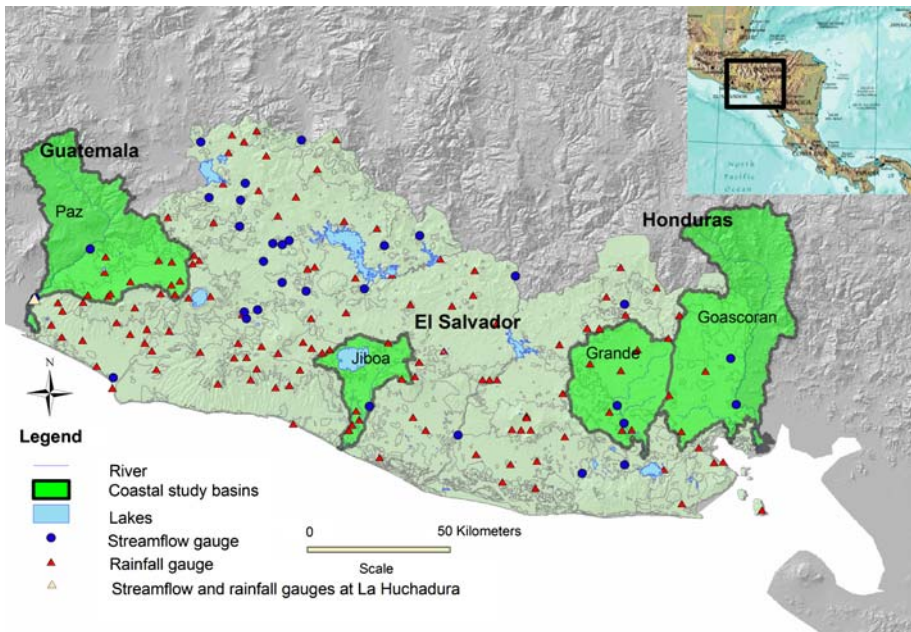


Fig. 2 Location map of coastal river basins studied in El Salvador: Paz, Jiboa, Grande de San Miguel (upper), and Goascoran

elements but in various proportions. For example, both the Jiboa and Grande de San Miguel are predominantly cultivated basins with minor amounts of deciduous brush in the Jiboa and urban influence in the Grande de San Miguel. The Paz and Goascoran basins have an even mix of both types of land use and vegetation.

Like other developing countries, the Republic of El Salvador has limited temporal and spatial content in meteorological and streamflow data. For example, there are currently fewer than 100 rainfall and 25 streamflow gauging stations (Fig. 1) that cover the entire country (area: 21,041 km<sup>2</sup>; elevation: sea level to 2,843 m), many with less than 5 years of record. At the time of Hurricane Mitch, there was only one streamflow and one rain gauge station co-located at La Hachadura in the Paz river basin, one streamflow gauge station at Villarias in the Grande de San Miguel river basin, and one surface meteorological station in each of the other coastal basins. The limited hydrological data in El Salvador is further exacerbated because most tributary basins (those that drain into a mainstem river) within the coastal river basins are numerous (in the hundreds) and remain completely ungauged. The limited hydrological data coupled with a need by the government to locate and build a new seaport prompted this study with an emphasis on estimating extreme rainfall amounts for 10-year, 25-year, and 100-year return periods in these four coastal river basins.

### 3.1 Rainfall-runoff curve numbers and recurrent rainfall amounts

Depending on the scale of study, the coastal basins could be divided into multiple tributary basins each having an associated curve number parameter. Whereas some rainfall-runoff studies assign a minimum tributary basin area threshold, doing so would generate hundreds of tributary basins. The actual number of tributary basins used in this study was determined based on the following criteria: (1) ensure a variation in tributary sizes (about 2D2,214 km<sup>2</sup>), (2) ensure a reasonable number of tributary basins and therefore curve numbers for estimation (20), (3) locate pour points coincident with the known gauge stations (Paz and Grande de San Miguel) and locations where simulated hydrographs are of interest in flood prediction (Friedel, in press), and (4) relative homogeneity based on physiographic characteristics. Whereas the first two criteria were important for testing the ability of the nonparsimonious parameter estimation process to converge, the third point was considered important for model validation. Application of these criteria resulted in 2, 2, 4, and 12 tributary basins associated with the respective coastal Paz, Goascoran, Jiboa, and Grande de San Miguel basins. For example, the 12 tributary basins identified in the Grande de San Miguel are shown in Fig. 2. Collectively, a total of 20 tributary basin runoff curve numbers and therefore models require joint estimation. With the primary objective of estimating three recurrent rainfall amounts for each of the four coastal basins, the total number of unknown parameter values requiring joint estimation is 32.

### 3.2 Observed peak-flow discharge in tributary basins

The lack of streamflow data in the ungauged tributary basins requires an alternative source of measurement information to constrain the joint inverse model parameter estimation process. To meet this need involves a two-step process: determine peak-flow discharge values from outside study basins and fit regression equations from which synthetic tributary basins observations can be generated. In the first step, annual peak-flow recurrence (10-year, 25-year, and 100-year) discharge values were determined by application of the

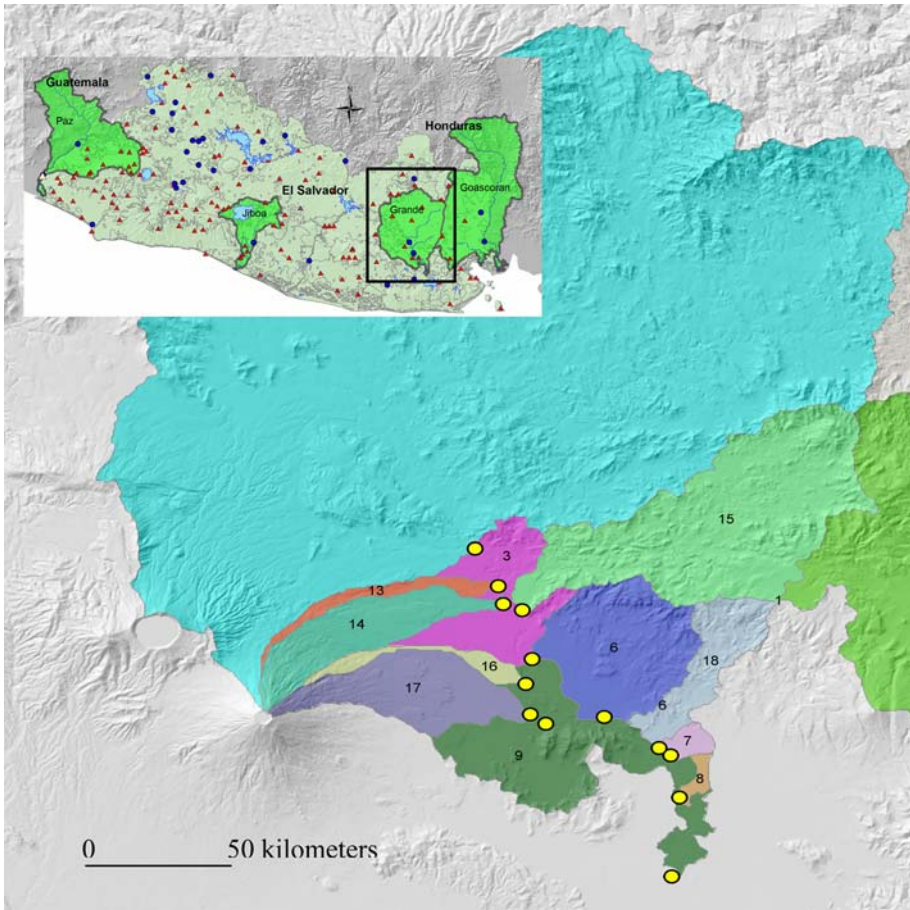


Fig. 3 Grande de San Miguel (upper) coastal river basin, El Salvador. Tributary basins used in the regularized joint inversion are numbered; associated outlets on mainstem are indicated by circles

method of moments and fitting a Pearson Type III distribution to the logarithms of annual flood peaks (Weaver and Gambler 1993) recorded at several gage stations: Villarias, Moscoso, La Hachadura, Montecristi, Bandera Santa Beatriz, Chilama La Libertad, San Pedro La Atalaya, Ceniza Conacaste Herrado, Torola Osicala (Table 1). The streamflow data from which the annual peak discharge values were determined represented an 18-year period of record. With the exception of the La Hachadura, and Villarias gauge stations, these tributary gauge stations were located outside the coastal study basins. All of the peak-flow discharge findings were consistent with recent area equations derived by Erazo (2004).

In the second step, separate regional (country-wide) regression equations were determined based on fitting peak-flow discharge values in log-log space to each set of recurrence data (Fig. 4). These three regression equations provide a means to generate synthetic peak-flow discharge values for the 20 ungauged tributary basins. In this way, the observation values of peak-flow discharge that are used to constrain the model parameter estimation process are considered unbiased (weighted mean of residuals on the

Table 1 Peak-flow discharge recurrence values determined by fitting a log-Pearson type-III distribution to historical streamflow information from selected gauge stations across El Salvador

Coastal basin	Tributary river basin gauge station	Drainage area (km <sup>2</sup> )	Peak discharge recurrence			Period of record
			10-year (m <sup>3</sup> s <sup>-1</sup> )	25-year (m <sup>3</sup> s <sup>-1</sup> )	100-year (m <sup>3</sup> s <sup>-1</sup> )	
Grande de San Miguel	Villarias	910	1,100	1,210	1,330	1955D1972
Grande de San Miguel	Moscoso	1,074	1,380	1,930	3,030	1955D1972
Paz	La Hachadura	1,991	3,560	5,050	7,560	1955D1972
Jiboa	Montecristo	429	444	579	799	1955D1972
Bandera	Santa Beatriz	422	602	935	1,740	1955D1972
Chilama	La Libertad	76.5	220	331	544	1955D1972
San Pedro	La Atalaya	102	333	423	570	1955D1972
Ceniza	Conacaste Herrado	168	202	293	436	1955D1972
Torola	Osicala	908	3,040	4,120	6,200	1955D1972
	Min	76.5	202	293	436	
	Max	1,991	3,560	5,050	7,560	

order of 10<sup>-2</sup>) and on-the-average correct (standard error of residuals of about 1), but have associated variability (uncertainty) indicated by scatter about the fitted line. Given that there are 20 tributary basins and three recurrent discharge periods, a total of 60 peak-discharge values were available as constraint information during the joint parameter estimation process.

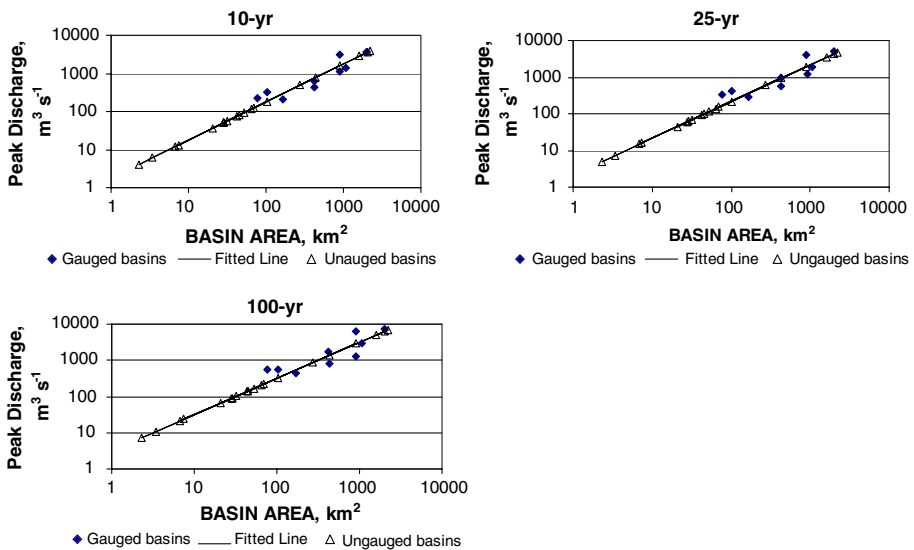


Fig. 4 Regional (country-wide) regression equations fit to peak discharge determined from historical streamflow information recorded at tributary basins outside study area

Whereas the observed information (synthetic measurements) used to constrain the parameter estimation process was derived from regional recurrent regression equations, the simulated peak-discharge values were generated (at each nonlinear parameter estimation iteration) from a set of rainfall-runoff HEC1 (U.S. Army Corps of Engineers 1998) models. The information needed as input to each rainfall-runoff model included tributary basin attributes (mean slope, flow length, and area) and hydrology (rainfall). The characteristic tributary basin attributes were determined using a geographical information system. Since the rainfall was assumed to be uniformly distributed within each tributary basin, a single cumulative rainfall amount, cumulative density function, and duration was assigned to each tributary basin model. The rainfall duration for each coastal basin differed as suggested by Erazo (2004) at El Salvador's Servicio Nacional de Estudios Territoriales (<http://www.snet.gob.sv>). Specifically, the respective rainfall duration for the Jiboa, Paz, Goascoaran, and Grande de San Miguel basins were 6-h, 1-day, 1-day, and 6-day. Whereas the tributary basin rainfall characteristics were assumed to be the same within a particular basin, the amount was independent with respect to the basins (Friedel, in press).

### 3.3 Regularization observations as soft prior information

In this study, the regularization observations (prior information) are defined by regional equations of dissimilarity; that is, the prior information is characterized by the difference between weighted logarithms of model parameters including tributary basin curve numbers (Friedel, in press) and coastal basin recurrent rainfall amounts. A general form of the prior information used to describe the regional dissimilarity is given by

$$p_{r_i} = \frac{1}{2} w_{r_i} \log q_j + w_{r_i} \log p_k \quad (9)$$

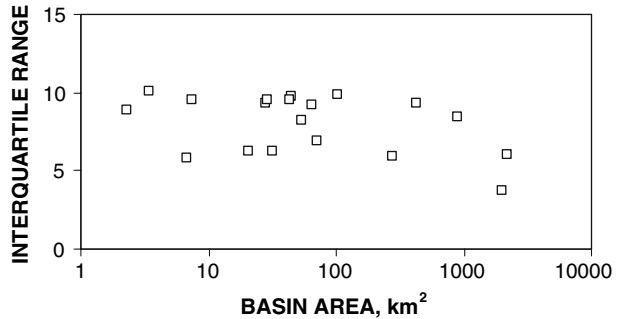
where the respective prior information and regularization weights  $w_{r_i}$  are set initially to 0 and 1 (homogeneous statistical structure analogous to a pure nugget parameter  $j$ ; and  $p_k$  is parameter  $k$ ).

Within the least-squares framework, this set of regularization equations describes the sum of squared differences between tributary basin curve numbers, and basin rainfall amounts, separated by some distance; therefore, the regularization equations are analogous to the spatial measures of dissimilarity described by a variogram: one variogram for tributary basin curve numbers, and one variogram for basin rainfall. However, there is no need to determine a priori the actual spatial structure. Rather, the regional relationship among basin curve numbers and basin rainfall amounts (spatial structure) is allowed to develop from the user-defined structural condition as part of the estimation process. That is, that spatial structure develops during the updating of rainfall-runoff curve numbers and recurrent rainfall amounts according to the available information content. The level of parameter parsimony achieved during the estimation of spatial structure is determined by the algorithm and information content in the synthetic peak-flow discharge values and parameter relations among tributary and coastal basins (soft prior information).

## 4 Results and discussion

A set of four parameter values (one runoff curve number and three recurrent rainfall events) was simultaneously estimated for each of the 20 rainfall-runoff models.

Fig. 5 Interquartile range for estimated tributary basin curve numbers



Application of the calibration-constrained Monte Carlo prediction approach resulted in tributary basin models with the same final objective function values. The consistency in final objective function values for these models indicated that the regularized joint inverse algorithm likely terminated at a global rather than local minimum in nonlinear parameter space. The overall quality of the model parameter sets was considered reasonably good as evidenced by consistency in high model correlation coefficient (0.99), and unbiased with a low mean of weighted residuals (on the order of  $10^{-4}$ ). The lack of any trend in the interquartile range of runoff curve numbers further corroborates the unbiased estimation of model parameter values with respect to basin size (Fig. 5). Since the near perfect model correlation values between observed and simulated profiles for each Monte Carlo trial are (by themselves) a poor indicator of model predictability (Friedman, 2005, 2006a), a statistical summary of estimated runoff curve numbers and rainfall amounts is presented in Tables 2 and 3.

All of the estimated tributary basin curve numbers reflect reasonable values when compared against other land use and soil groups (Soil Conservation Service, 1985). In assessing the estimated curve numbers, the values appear indicative of several land use and soil groups; for example, agricultural and soil group A (64); forest and soil groups ADC (30D70), and grass/pasture and soil groups ADB (39D61) (Soil Conservation Service, 1985). Whereas the higher values in estimated curve numbers are associated with the Grande and Paz basins, the lower values in estimated curve numbers are associated with the Jiboa basin. The lower values of estimated curve numbers in the Jiboa tributary basin may possibly be due to the influence of storage associated with Lake Illopango (Fig. 6). Given the complexity of land use and soil types in each basin, the estimated curve number more accurately reflects a composite value of land use and soil conditions, such as surface storage, agricultural development (coffee and sugar cane) and deforestation.

For a given tributary basin, the estimated curve number and rainfall amounts also can be described as random variables with statistical characteristics attributed to uncertainty in the model conceptualization, parameterization, and information content in the measurements used to constrain the solution (Table 2). The approximate range in nonlinear confidence intervals (difference between 5th and 95th percentiles; Fig. 6) associated with parameter values vary being attributed, in part, to the inability to capture this natural and anthropogenic variability, as well as variability in the synthetic tributary peak-flow discharge applied as constraints. The similarity in average and median curve numbers implies that the uncertainty associated with tributary basins may be approximated as Gaussian random variables. Likewise, the similarity in average and median (not shown) recurrent rainfall values suggests that uncertainty associated with rainfall in coastal basins may approximate as Gaussian random variables.

Table 2 Summary statistics for estimated tributary basin curve numbers

Tributary basin number	River basin	Drainage area (km <sup>2</sup> )	Estimated curve numbers, dimensionless				
			Minimum	Maximum	Average	Median	Standard deviation
1	Goascoran	276	47.7	62.0	55.9	56.6	4.7
2	Goascoran	2,213	56.7	71.9	65.1	65.9	4.8
3	Grande	28.2	45.6	63.7	56.3	58.7	6.9
4	Grande	900	52.1	68.6	62.1	64.6	6.3
6	Grande	45.1	39.1	57.9	49.9	52.4	7.1
7	Grande	3.4	33.4	55.5	44.8	46.9	7.8
8	Grande	2.3	52.6	70.1	62.9	65.2	6.6
9	Grande	53.2	60.3	76.5	70.0	72.4	6.2
13	Grande	7.4	46.6	64.6	57.2	59.7	6.8
14	Grande	28.7	41.4	59.9	52.2	54.6	7.0
15	Grande	102	48.7	66.5	59.2	61.6	6.7
16	Grande	6.8	42.9	61.3	53.7	56.1	7.0
17	Grande	32.1	43.1	61.5	53.8	56.3	7.0
18	Grande	20.7	37.3	56.3	48.3	50.7	7.1
10	Jiboa	431	35.2	51.4	45.3	46.8	5.2
11	Jiboa	42.9	34.6	51.4	45.1	46.6	5.4
12	Jiboa	64.1	34.6	51.4	45.1	46.6	5.4
19	Jiboa	70.3	50.0	67.6	61.3	63.0	5.7
22	Paz	2,032	64.6	72.4	69.4	70.0	2.6

The increase in range and standard deviation of estimated rainfall amounts with increasing recurrence interval reflects more uncertainty associated with larger and increasingly rare rainfall events (Table 2). Also, the cumulative probability distributions plotted for rainfall amounts indicate differences among coastal basins for a given return period and an increase in magnitude and range among coastal basins with increasing return period (Fig. 7). Validation of the regularized joint inverse approach to estimate extreme rainfall events was done in two ways: quantitatively by comparison of traditional rainfall estimates based on rainfall-frequency relations derived for recurrent intervals, and qualitatively by comparing the cumulative hurricane-induced rainfall observed at a tributary gage to the corresponding nonlinear rainfall prediction limits.

In the first validation approach, comparison of the median estimated rainfall amounts with 10-year, 25-year, and 100-year return periods for each of the basins with those rainfall amounts computed using traditional frequency-duration based approach and historical rain gauge data appear visually to be good (Fig. 8). However, the regularized joint inversion tends to estimate rainfall values that are larger (between 2% and 26%) than the traditional point rainfall frequency approach (Table 2). Also, it is noteworthy that the relative difference between these approaches appears to increase with increasing return period, but in all cases the regularized joint inverse approach may be considered to provide a more conservative estimate.

Of the four coastal basins being studied, the Paz was the only basin that had an operational rain gauge during Hurricane Mitch. It is interesting and important to note that

Table 3 Approach comparison for coastal basin rainfall amounts estimated to generate the T-year peak flow

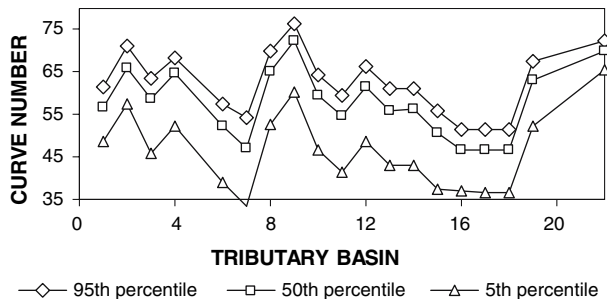
Peak-flow return period	Coastal basin	Estimated rainfall (mm)				
		Regularized joint inverse minimum <sup>a</sup>	Regularized joint inverse maximum <sup>a</sup>	Regularized joint inverse median <sup>a</sup>	Rainfall frequency analysis <sup>b</sup>	Percent difference <sup>c</sup> (%)
100-year	Grande	242.3	540.4	317.9	233.4	26.6
100-year	Jiboa	241.1	411.1	278.3	256.3	7.9
100-year	Paz	330	400.5	351.0	334.8	4.6
100-year	Goascoran	293.4	417.0	334.0	293.9	12.0
25-year	Grande	207.8	428.9	250.3	214.9	14.1
25-year	Jiboa	199	343.7	230.5	217.9	5.5
25-year	Paz	246.6	302.1	263.4	250.2	5.0
25-year	Goascoran	223.2	329.6	261.8	223.5	14.6
10-year	Grande	171.4	357.6	207.5	199.6	3.8
10-year	Jiboa	171.4	299.2	199.2	192.5	3.3
10-year	Paz	195	240.7	208.5	197.8	5.1
10-year	Goascoran	180.8	275.1	217.2	181.1	16.6
Observed rainfall at La Hachadura (mm)						
25-year	Paz			184.1		1-day
25-year	Paz			220.2		2-day

<sup>a</sup> New approach

<sup>b</sup> Traditional approach

<sup>c</sup> % difference= [(Regularized joint inverse median value Rainfall frequency analysis value)/Regularized joint inverse median value] \* 100

Fig. 6 Estimated tributary basin curve numbers (model parameters)



the rainfall amount of 184 (1-day) mm or 220.2 (2-day) mm observed at La Huchadura in the Paz tributary basin is similar in value to, but outside and slightly less than, the respective estimated 25-year recurrence nonlinear rainfall confidence limits of 246.6 mm and 302.1 mm. Explanations for the disparity between observed and estimated range of predicted rainfall amounts could be bias in the observations, bias in the model, or uncertainty in the observations used to develop the regression model. Whereas one might argue for prediction bias due to an overly simplified model structure (indicated by eigenvectors dominated each by an individual parameter), the similarity in joint inverse estimated and traditional rainfall frequency based approaches supports the notion of

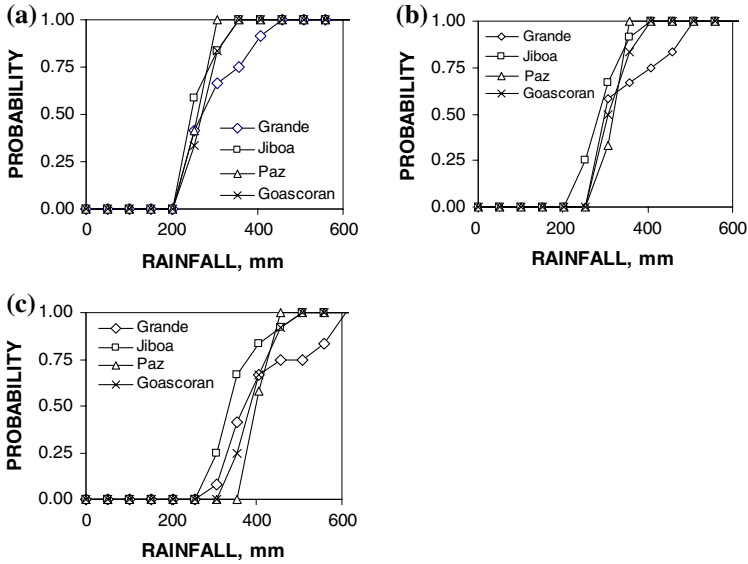


Fig. 7 Cumulative distribution functions for estimated basin rainfall (a) (10-year recurrence interval) (b) (25-year recurrence interval) (c) (100-year recurrence interval)

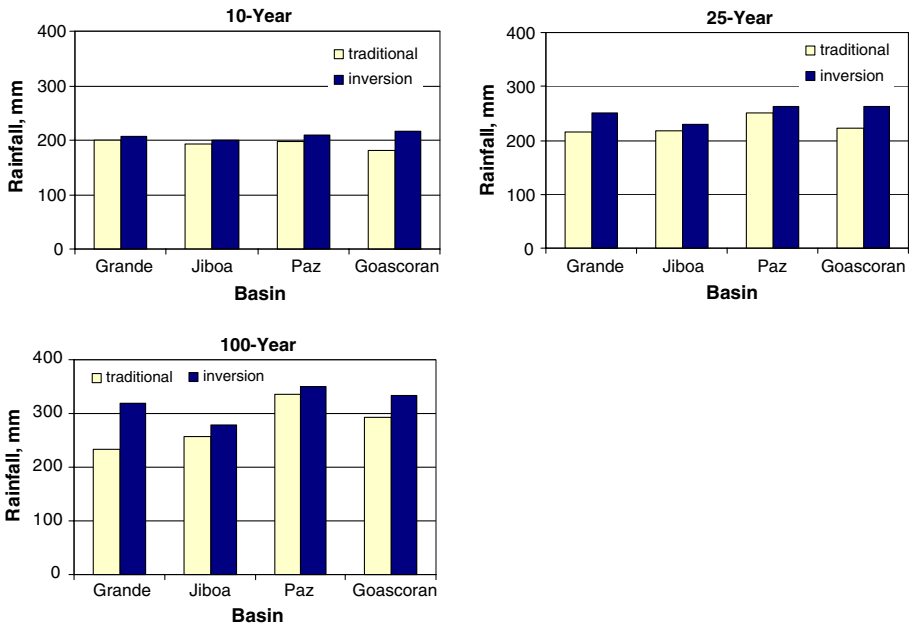


Fig. 8 Comparison of rainfall amounts estimated for 10-year, 25-year, and 100-year return periods using rainfall frequency-duration (traditional) and regularized joint inverse (inversion) approaches

rainfall observation bias due to under-catch and/or spatial variability as is oftentimes noted in the literature.

## 5 Summary and conclusions

A regularized joint inverse approach is presented and used to simultaneously estimate parameter values (tributary basin rainfall-runoff curve numbers and coastal basin extreme rainfall amounts) in four principal ungauged coastal river basins of El Salvador: Paz, Jiboa, Grande, and Goascoran. Application of a calibration-constrained Monte Carlo approach was used to establish the limits of nonlinear uncertainty in estimated rainfall-runoff curve numbers and rainfall amounts. The important results are as follows:

- ⊖ The level of parameter parsimony required to achieve a stable inverse solution was determined by the algorithm and information content in the synthetic peak-flow discharge values (generated using regional equations) and parameter relations among tributary and coastal basins (soft prior information).
- ⊖ Application of the calibration-constrained Monte Carlo technique revealed parameter nonuniqueness, because multiple parameter sets could be estimated using different initial starting values but each resulting in the same final objective function.
- ⊖ The estimated tributary basin curve numbers characterize composite values as they were similar to several land-use classes present in the coastal basins: agricultural and soil group A; forest and soil groups A–C, and grass/pasture and soil groups A–B.
- ⊖ The cumulative probability distributions for extreme rainfall amounts indicate differences among basins for a given return period, and an increase in magnitude and range among basins with increasing return interval.
- ⊖ The joint inverse estimated median rainfall amounts for all return periods were similar to those estimated using the traditional rainfall frequency-duration technique, but the range in uncertainty of estimated rainfall amounts increased with increasing return period.
- ⊖ The rainfall amount observed at the La Huchadura gauge station in the Paz tributary basin during Hurricane Mitch is similar in value to, but outside and slightly less than, the estimated 25-year recurrence rainfall confidence limits.
- ⊖ The similarity in regularized joint inverse estimates and traditional rainfall-frequency based estimates suggests that the rainfall observation at La Huchadura may be lower in magnitude due to under-catch and/or spatial variability, but not model bias.

Whereas the regularized joint inverse approach shows promise for simultaneous basin-parameter optimization and rainfall-event estimation in ungauged basins, there is a vast amount of uncertainty in the case study data of El Salvador (poor gage density, short periods of record, and uncertain peak-discharge and extreme rainfall data). A similar analysis needs to be conducted in a region where more extensive data (longer streamflow records and better rainfall frequency analysis) are available to provide additional insight into the conclusions/observations offered here.

**Acknowledgments** This work was carried out under the Earthquake Reconstruction Program administered by U.S. Agency for International Development, as part of technical assistance following the catastrophic earthquakes of January and February 2001 in El Salvador.

## References

- Bates B, Campbell E (2001) A Markov chain Monte Carlo scheme for parameter estimation and inference in conceptual rainfall-runoff modeling. *Water Resour Res* 37(4):937D947
- Beven KJ, Binley AM (1992) The future of distributed models: model calibration and uncertainty prediction. *Hydrol Process* 6:279D298
- Bicknell BR, Imhoff JC, Kittle JL, Donigan AS, Johanson RC (1997) Hydrologic simulation program-fortran, user's manual for version 11. U.S. Environmental Protection Agency EPA/600/R-97/-90, 755 pp
- Bloschl G, Sivapalan M (1995) Scale issues in hydrological modeling D a review. *Hydrol Process* 9(3D4):251D290
- British Broadcasting Corporation (2000) Flood emergency in El Salvador. <http://www.news.bbc.co.uk/1/hi/world/americas/914509.stm>. Accessed 7 September 2007
- Chow VT, Maidment DR, Mays LW (1988) Applied hydrology. McGraw-Hill, New York
- Cooley RL, Naff RL (1990) Regression modeling of ground-water flow. US Geological Survey Technical Water Resources Investigation: Book 3, Chapter B4, 232 pp
- Crone AJ, Baum RL, Lidke DJ, Sather DND, Bradley LA, Tarr AC (2001) Landslides induced by hurricane Mitch in El Salvador D an inventory and descriptions of selected features. US Geological Survey Open File Report, 01-144
- Curtis S, Crawford TW, Lecce SA (2007) A comparison of TRMM to other basin-scale estimates of rainfall during the 1999 Hurricane Floyd flood. *Nat Hazards* 43:187D198
- deGroot-Hedlin C, Constable S (1990) Occam inversion to generate smooth, two-dimensional models from magnetotelluric data. *Geophysics* 55(12):1624D1631
- Doherty J (2004) PEST: model-independent parameter estimation, version 5 of user manual. Watermark Numerical computing, Brisbane, Australia, 213 pp
- Doherty J, Johnston JM (2003) Methodologies for calibration and predictive analysis of a watershed model. *J Am Water Resour As* 39(2):251D265
- El Diario de Hoy (1999) Flooding in El Salvador (various articles October). <http://www.osb.org/pab/ES%20Flooding%20F99.htm>. Accessed 7 September 2007
- Elliot J, Smith ME, Friedel MJ, Litke D (2005) Post-Pre hydrologic hazards study for the 2002 Hayman, Coal Seam, and Missionary Ridge wildPres, Colorado. U.S. Geological Survey, Science Investigations Report 2004D5300, 125 pp
- Erazo AM (2004) Regionalization de caudales maximos y medios en El Salvador, Servicio Nacional de Estudios Territoriales. <http://www.snet.gob.sv/page.php?id=197&p=131&search=maximo>. Accessed 19 March 2006
- Friedel MJ (2005) Coupled inverse modeling of vadose zone water, heat, and solute transport: calibration constraints, parameter nonuniqueness, and predictive uncertainty. *J Hydrol* 312(1D4):148D175
- Friedel MJ (2006a) Predictive streamflow uncertainty in relation to calibration-constraint information, model complexity, and model bias. *Int J River Basin Manage* 4(1):1D15
- Friedel MJ (2006b) Reliability in estimating urban groundwater recharge through the vadose zone: managing sustainable development in arid and semiarid regions. In: Tellam JH, Rivett MO, IsraPlov RG (eds) Urban groundwater management and sustainability. NATO Science Series, IV. Earth and Environmental Sciences, vol. 74. Springer, Dordrecht, The Netherlands, pp 169D182
- Friedel MJ Probable flood predictions in ungauged coastal basins of El Salvador. *J Hydrol Eng* (in press)
- Helin J, Haigh H, Marks F (1999) Rainfall characteristics of Hurricane Mitch. *Nature* 399:316D316
- Hill M (1998) Methods and guidelines for effective model calibration. U.S. Geological Survey Water-Resources Investigations Report 98-4005, 90 pp
- Hossain F, Katiyar N (2006) Improving flood forecasting in international river basins. *EOS Trans (AGU)* 87(5):49D50
- Kuczera G (1983) Improved parameter inference in catchment models. 1. Evaluating parameter uncertainty. *Water Resour Res* 19(5):1151D1172
- Lakshmi V (2004) The role of satellite remote sensing in the prediction of ungauged basins. *J Hydrol Proc* 18:1029D1034
- Lin G-F, Chen L-H, Kao S-C (2005) Development of regional design hyetographs. *J Hydrol Proc* 19:937D946
- Merz R, Bloschl G (2004) Regionalization of catchment model parameters. *J Hydrol* 287:95D123
- Rossman L (2005) SWMM D storm water management model user's manual version 5. U.S. Environmental Protection Agency, EPA/600/R-05/040, 247 pp
- Shiklomanov AI, Lammers RB, Wosmarty CJ (2002) Widespread decline in hydrological monitoring threatens pan-arctic research. *EOS Trans* 83(2):16D17

- Siebert J (1999) Regionalization of parameters for a conceptual rainfall-runoff model. *Agric Forest Meteorol* 98/99:279–293
- Sivapalan M (2003) Prediction in ungauged basins: a grand challenge for theoretical hydrology. *Hydrol Process* 17(15):3163–3170
- Soil Conservation Service (1985) National engineering handbook, Sect. 4. U.S. Department of Agriculture, Washington, DC
- Stokstad E (1999) Scarcity of rain, stream gages threatens forecasts. *Science* 285:1199
- U.S. Army Corps of Engineers (1998) HEC-1 flood hydrograph package user's manual. U.S. Army Corps of Engineers Hydrologic Engineering Center, Davis, CA, version 4(1), 434 pp
- Weaver JD, Gamble CR (1993) Flood frequency of streams in rural basins of Tennessee. U.S. Geological Survey Water-Resources Investigations Report 92-4165