Environ Earth Sci
DOI 10.1007/s12665-011-0976-1

SPECIAL ISSUE

Rainfall-induced landslide susceptibility zonation of Puerto Rico

Chiara Lepore · Sameer A. Kamal · Peter Shanahan · Rafael L. Bras

Received: 30 July 2010 / Accepted: 10 February 2011

Abstract Landslides are a major geologic hazard with estimated tens of deaths and $1–2 billion in economic losses per year in the US alone. The island of Puerto Rico experiences one or two large events per year, often triggered in steeply sloped areas by prolonged and heavy rainfall. Identifying areas susceptible to landslides thus has great potential value for Puerto Rico and would allow better management of its territory. Landslide susceptibility zonation (LSZ) procedures identify areas prone to failure based on the characteristics of past events. LSZs are here developed based on two widely applied methodologies: bivariate frequency ratio (FR method) and logistic regression (LR method). With these methodologies, the correlations among eight possible landslide-inducing factors over the island have been investigated in detail. Both methodologies indicate aspect, slope, elevation, geological discontinuities, and geology as highly significant landslide-inducing factors, together with land-cover for the FR method and distance from road for the LR method. The LR method is grounded in rigorous statistical testing and model building but did not improve results over the simpler FR method. Accordingly, the FR method has been selected to generate a landslide susceptibility map for Puerto Rico. The landslide susceptibility predictions were tested against previous landslide analyses and other landslide inventories. This independent evaluation demonstrated that the two methods are consistent with landslide susceptibility zonation from those earlier studies and showed this analysis to have resulted in a robust and verifiable landslide susceptibility zonation map for the whole island of Puerto Rico.

Keywords Landslides · Landslide susceptibility · Rainfall-induced landsliding · Frequency ratio · Logistic regression · GIS · Susceptibility maps · Puerto Rico

C. Lepore (✉) · S. A. Kamal · P. Shanahan · R. L. Bras
Parsons Laboratory for Environmental Science and Engineering,
Department of Civil and Environmental Engineering,
Massachusetts Institute of Technology, Cambridge, MA, USA
e-mail: chlepore@mit.edu

Present Address:
S. A. Kamal
World Bank, Washington, DC, USA

Present Address:
R. L. Bras
Georgia Institute of Technology, Atlanta, GA, USA

Introduction

All major types of landslides occur within the territory of Puerto Rico. They include shallow soil slips, debris flow, debris slides, debris avalanches, and slumps (Larsen and Simon 1993). The factors contributing to the high landslide activity in Puerto Rico are the steep slopes and the relatively moist condition of soils due to abundant rainfall, especially in the northern part of the island. The southern part of Puerto Rico normally experiences less rainfall throughout the year, and landslides take place mostly during exceptionally heavy rainfalls (Campbell et al. 1985; Jibson 1986). Among the worst landslide disasters in the U.S. is the Mameyes landslide, a series of failures that took place in the area of Ponce and Coamo near the southern coast of the island during the passage of tropical storm Isabel in October 1985. At least 129 people were killed and more than 100 homes were destroyed. In addition to loss of human life, the direct and indirect costs from landslides, such as repair, replacement and maintenance of infrastructure, and loss of industrial, agricultural and forest productivity, are high. Moreover, besides the effects

Published online: 25 February 2011

landslides can have on the built environment and human life, landslides can have great effect on natural environments. This topic, among others, is the subject of the Luquillo Long-Term Ecological Research Program (Luquillo LTER, http://luq.lternet.edu). The Luquillo Experimental Forest (LEF), located in the northern part of Puerto Rico (Fig. 1), has been a center of tropical forestry research for nearly a century and recently joined the Critical Zone Observatory (CZO) network. Among the several research topics examined at LEF, disturbance and recovery of vegetation after landslides is one of the most important, focusing on the effects of landsliding on the ecology of the forest. Landslide recovery areas, for example, develop centers of biodiversity with pioneer species not found elsewhere in the forest (Shiels et al. 2008; Walker and Shiels 2008; Walker et al. 1996).

In landslide analysis, there are three key steps: susceptibility, hazard, and risk assessment (Einstein 1988), each step drawing on the results of the previous step. The components of this analysis are as follows:

- *Landslide susceptibility* is broadly defined as the probability of a landslide occurrence over a region. The probability is based on an empirical or modeled relationship between historical landslide inventory information and surface observables (Varnes and IAEG 1984). This implies that slope failures in the future will occur with the same probability under the same conditions that led to past and present instability. That said, the geomorphological conditions that lead to landslides in a particular area may vary in time (i.e., slopes will get gentler, vegetation can vary), and the resulting maps should be interpreted as intrinsically connected to specific characteristics of a territory rather than to a specific location (Guzzetti et al. 2005).
- *Landslide hazard* represents the probability of occurrence of a potentially damaging landslide within a specified period of time and within a given area (Varnes and IAEG 1984). The definition considers the location, the timing, and the magnitude of the landslide event. It is difficult to measure the magnitude of landslides, which exhibit wide diversity, varying from superficial to deep and from very slow to very fast. Moreover, it is often difficult to measure certain characteristics of landslides such as velocity, volume, and destructiveness. In most of the literature, the area of a landslide is taken as a good approximation of the magnitude of the event, especially in landslide hazard mapping.
- *Landslide risk* expresses the economic and social dimensions of slope failure. It is generally considered to be equal to the likelihood of death or injury, or to the expected monetary loss due to the occurrence of a landslide (Varnes and IAEG 1984). Risk is usually defined as the product of landslide hazard and vulnerability. Vulnerability represents the level of potential damage of a given element (e.g., building) subjected to a landslide of a given intensity.

Guzzetti et al. (1999, 2005), Van Westen et al. (1997, 2006), Van Westen (2000), Dai and Lee (2002), Santacana et al.(2003), Lee and Pradhan (2006; 2007), Lee et al. (2007), Zhu and Huang (2006), Arora and Gupta (2004), and Cheng and Wang (2007) have given a complete overview of the different methodologies commonly used in LSZ. Some methods give as output a quantitative evaluation of susceptibility in terms of a spatial probability of occurrence. Other quantitative methods define the susceptibility of a territory to landsliding using a numeric scale, not bounded between 0 and 1, with increasing values for increasing susceptibility. Qualitative methods are also used, mostly for very large areas.

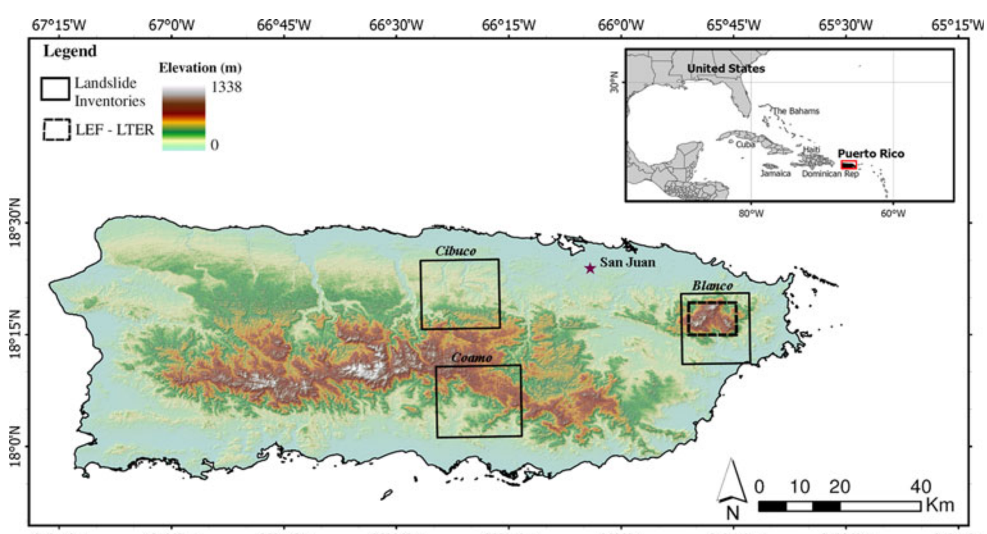

Fig. 1 Map of Puerto Rico showing the location of the three landslides inventories and the location of LEF-LTER

The common basis of the quantitative methods is to identify and obtain the correlations between landslide-inducing factors and the area where the landslides have been recorded and from those correlations define landslide-prone areas. Landslide-inducing factors include geomorphological, hydraulic, hydrological, and anthropogenic characteristics. Once the factors are identified they can be evaluated at the landslide site all together with a multiple regression technique, such as weighted linear regression (Hong et al. 2007), logistic regression (Cheng and Wang 2007; Dai and Lee 2002; Guzzetti et al. 1999), or discriminant analysis (Santacana et al. 2003), or individually in a bivariate statistical method (Lee and Pradhan 2006). More recent efforts have developed artificial neural network (ANN) approaches (Arora and Gupta 2004; Lee et al. 2007). All these methods have been frequently used in recent years because of the development of increasingly powerful GIS applications, which make the data analysis, manipulation, and management much simpler (Donati and Turrini 2002; Jimenez-Peralvarez et al. 2009).

LSZ or hazard mapping of Puerto Rico goes back to Monroe (1979). He proposed a LSZ for the island and assigned four categories of landslide susceptibility based on geologic maps and personal observations: (1) highest susceptibility for areas of past or current landslide activity, (2) high susceptibility for areas with slopes greater than 27° (50%) or that contain "slide-prone" rock and soil, (3) moderate susceptibility for all other sloping areas, and (4) low susceptibility for flat-lying areas. In the second category, the "slide-prone" rock includes the outcrop belts of the San Sebastian and Cibao formations in northern Puerto Rico (Monroe 1979) and the areas underlain by intrusive and volcanic rocks in the mountains of northeastern Puerto Rico (Briggs and Akers 1965). The rest of the island is broken by numerous faults and Monroe (1979) included all steep slopes near the faults as areas of high susceptibility. His work defined a large part of the island to have moderate susceptibility and only few areas, mostly already active, to have high susceptibility. Larsen and Torres-Sanchez (1998) completed a more detailed analysis of landslide susceptibility. They introduced a detailed landslide inventory for three large areas of the island (see Fig. 1). They considered a sort of conditional probability of failure for the three areas by developing simplified matrices that relate landslide occurrence to a set of geomorphological characteristics namely slope angle, slope elevation, slope aspect, and land use. In their study, however, they were limited by the quality and resolution of data available at that time. They found that hillslopes with gradients of 12° or more, elevation in excess of 300 m, or slope aspect facing the trade winds corresponded to an appreciable increase in landslide frequency. Larsen and Parks (1998) and Larsen et al. (2004), respectively, prepared detailed LSZ for two areas confined to the municipality of Comerío (74 km^2) in the center of the island and to the city of Ponce (400 km^2). Both classify the territory into four susceptibility classes based on Larsen and Torres-Sanchez's methodology at very detailed scales, 1:30,000 for Ponce and 1:20,000 for Comerío. Prior studies on LSZ in Puerto Rico have focused primarily on rainfall-triggered landslides, with the exception of Santiago and Larsen (2001), whose study of the San Juan metropolitan area concluded that it was generally not susceptible to earthquake-induced landslides.

The study presented here builds upon these previous efforts and develops a LSZ for the whole island of Puerto Rico. Consistent with previous studies, it focuses on rainfall-triggered landslides. Two statistical methodologies, frequently employed in the literature, are applied to the available data, the bivariate frequency ratio and multivariate logistic regression. The organization of the paper is as follows. The methodologies are introduced and discussed in the second section. The available data and their preliminary analysis are described in the third section. The results are presented and discussed in the fourth and fifth sections. The sixth section closes with remarks and conclusions.

Methods

Bivariate methodologies constitute a class of different methods used in LSZ applications (Van Westen 2000). In this study, the authors follow the so-called frequency ratio (FR) approach used, among others, by Lee and Pradhan (2006) and Lee et al. (2007). The methodology is quite simple both conceptually and in practice. For a given territory, the total area that encompasses the data set is A, and the portion where landslides occurred, A_{L}. Each landslide-inducing factor considered in the application (slope, curvature, etc.) is evaluated individually and reclassified into a series of bins (e.g., slope [0°–5°, 5°–10°,…]). A frequency ratio value is calculated for each bin i of each characteristic f as:

$$F_{fi} = \frac{A_{\mathrm{L}|fi}/A_{\mathrm{L}}}{A_{fi}/A}, \qquad (1)$$

where $A_{\mathrm{L}|fi}$ and A_{fi} are the portions of landslides and total area (A_{L} and A) with characteristic f having a factor value within a bin range i. The interpretation of the frequency ratio is as follows:

- $F_{fi} < 1$: there is proportionally less landslide area with characteristic f in bin i than there is total area with characteristic f in bin i;

- $F_{fi} \approx 1$: the percent of landslide area with characteristic f in bin i is in roughly the same proportion as the total area with characteristic f in bin i;
- $F_{fi} > 1$: there is a higher percentage of landslide area with characteristic f in bin i than there is total area with characteristic f in bin i, indicating a propensity of that value of characteristic f to lead to failure.

The final quantity of interest to LSZ is the Susceptibility Index (SI), calculated as:

$$\mathrm{SI} = \sum_{f=f_1}^{f_k} \sum_{i=1}^{n_f} F_{fi} \tag{2}$$

and is the sum over all landslide-inducing factors, $f = f_1,\ldots,f_k$, and bins, $i = 1,\ldots,n_f$. Higher values of SI indicate a higher propensity of the area to failure. One advantage of this methodology is its simplicity both in calculating susceptibility and in interpreting the results. The grid-based technique used to compare the surface observables (f factors) and landslides allows this methodology to be applied to large areas at any spatial scale or data resolution available.

Multivariate logistic regression analysis (LR method), the second methodology considered, is a technique commonly used to relate several independent explanatory variables (here identified as inducing factors f) to a dependent dichotomous variable (here landslide occurrences) using a nonlinear model. Slope failure can be represented through a binary variable $Y = [0, 1]$, indicating whether the failure has occurred ($Y = 1$) or not ($Y = 0$), and through some inducing factors, $X = [x_1,x_2,\ldots,x_n]^T$. Using the m observations, $(X_1, Y_1),\ldots,(X_m, Y_m)$, it is possible to express the probability of failure as a function of X. The probability of failure, P_L, becomes, hence, the expected value of Y given X.

$$P_\mathrm{L}(X) = P[Y = 1\,|X] = E[Y|X]. \tag{3}$$

A linear regression model is not applicable because of the dichotomous nature of Y, however, an available alternative is the so-called logit models (Hosmer and Lemeshow 2000; Liao et al. 1988). In the logit model, the natural logarithm of the ratio of the probability of failure to its complement (no failure) is related to the explanatory variables by the linear model:

$$z(x) = \ln\left(\frac{P_\mathrm{L}}{1 - P_\mathrm{L}}\right) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_n x_n, \tag{4}$$

where β_0 is the intercept of the model, β_i $(i = 1, 2,\ldots,n)$ are the coefficients of the explanatory variables x_i, $(i = 1, 2,\ldots,n)$. The logit $z(x)$ is a linear function of the observables, but the probabilities themselves are not. The probability of failure can be represented as:

$$P_L(X) = \frac{\exp\left(\beta_0 + \sum \beta_i x_j\right)}{1 + \exp\left(\beta_0 + \sum \beta_i x_j\right)}. \tag{5}$$

The coefficient β_i represents the effect of the ith landslide-inducing factor on the occurrence of the event: if positive, it indicates a positive correlation and vice versa. In a logit model, the explanatory variables can be either continuous or categorical. The flexibility of this model is particularly appealing for this application; most of the inducing factors considered are continuous (slope, elevation, etc.), but some are typically categorical (i.e., type of geology). In the case of categorical information, but also if the continuous values are binned in classes, it is common to make use of so-called dummy variable coding. Dummy variables classify each explanatory variable into a series of k bins and generate a $(k - 1)$-column matrix of 1s and 0s; the value of the bin corresponding to the actual variable value is 1, while the value for all other bins is 0, the first bin is left out as the reference category for the variable (Chau and Chan 2005).

The application of the LR method is more complex than the FR method, and the choice of the best numerical procedure for the estimation of the coefficients is not trivial. The parameters of the LR model are estimated using the maximum-likelihood method. Maximum likelihood computation for logit models may not converge due to numerical instabilities caused by their dichotomous nature (Allison 2008). Instabilities are usually avoided by introducing approximations in the numerical procedure, which can bias the final results, for example, by overestimating low values of probabilities and underestimating higher ones. The results shown here were obtained using *R*, a computer language for statistical computing (R Development Core Team 2005); *R* allows for a fast resolution of the maximum likelihood procedure, easy implementation of statistical tests, and most importantly in this application, easy handling of large data sets.

Discussion of methods

The choice of the inducing factors to include in either the FR or LR procedure is often dictated by the actual availability of information in the area of interest. Lithology and slope angle are commonly considered to be the most important factors in slope stability. Besides these, many other factors can be considered: curvature, distance to drainage, geology, land use, distance to fault, soil texture, distance to road, vegetation cover, geometry of the watershed (area, mean elevation, length, mean watershed angle), or geotechnical characteristics. Bivariate models including the FR method are based on the assumption of conditional independence between the various factors with respect to the probability of occurrence of landslides (Van Westen et al. 1997). Such an assumption is not always respected in applications to landslide susceptibility mapping since some of the variables are almost certainly correlated.

Once each factor has been parameterized and binned appropriately, its relative significance has to be determined. For example: what is the importance of slope angle relative to the distance to roads from the perspective of landslides? The FR method does not include rigorous statistical testing for determining the significance of the factors; typically, the best combination of inducing factors to include in the final SI calculation is based on the values obtained by the Area Under the Curve (AUC) of the Receiving Operating Characteristic (ROC) curves (Fawcett 2006). In the case of LSZ, the ROC plots the portion of landslides with a given susceptibility index versus the portion of the study area with the same susceptibility index. This ROC plot depicts the tradeoff between successful hit rates (true positives) and false alarm rates (false positives). The success of the susceptibility map is determined by taking the area under the ROC curve (AUC). An AUC value of 1 indicates a perfect fit of the model and a value of 0.5 represents a fit indistinguishable from random occurrence.

Factor selection for the LR method can be similar to the FR method (i.e., based on the use of the ROC) or based on more rigorous methods, such as common statistical model-building procedures. Forward, backward, or stepwise procedures are techniques that allow evaluation of the statistical significance of an added observable to the model. With forward selection, one starts with the simpler model and adds the next best-fitting variable until no further improvement is possible. In the case of backward selection, one starts with the most complex model and reduces the model successively. Stepwise selection is a combination of these procedures.

In the case of logistic regression, the significance of an observable is tested based on Akaike's Information Criterion (AIC), the Waldo tests, or the likelihood ratio test (2LL). The Waldo test compares the maximum likelihood estimate of the coefficients with their standard error; the ratio of these two quantities is then tested under the normality assumption. The 2LL test compares the log-likelihood of two fitted models, one nested within the other, to evaluate the significance of the additional coefficients. The AIC test computes both the sum of the log-likelihood and the number of parameters used in the fitted model. In case two different models have very similar likelihood values, AIC favors the one with the fewest parameters. The AIC and the 2LL tests are considered superior to the Waldo test because they do not make use of any normality assumption (Hosmer and Lemeshow 2000). For the application reported here, a stepwise procedure based on the AIC index, and two forward procedures based on the 2LL and on the ROC indexes have been applied. The LR method can model both continuous and categorical variables; in case a continuous variable is available it is important to control the assumption of linearity in the logit for the covariate (Eq. 4). In case this is not satisfied, it is possible to try simple transformations of the variable (e.g. log, power) or to proceed by classifying it in categories, similarly to the FR method.

In the LR method, one issue is the definition of the sample Y of occurrences and non-occurrences of landslides. For landslides ($Y = 1$), all available occurrences are used; the LR method can, in fact, be applied in "training mode" on subsets of landslide data to use the remaining landslide data to validate the results. In this study, the whole data set was used in order to extend the results to the whole island, and therefore to include as much information as possible in the analysis. For the sample of non-landslides ($Y = 0$), there are different options. One possibility is to use all of the available $Y = 0$ occurrences, that is the whole complementary area of $Y = 1$, with the risk of biasing the final results because of the vastly unequal proportions of 1 and 0 values (Guzzetti et al. 1999; Ohlmacher and Davis 2003). Another option is to use a sample of spatially randomly selected $Y = 0$ occurrences so as to have a non-occurrence sample equal in size to the occurrence sample (equal areas of landslides and non-landslides). Unlike the FR method, the LR method is robust when the data are auto-correlated, as often occurs when data are derived from a digital elevation model (DEM) (Mathew et al. 2008).

Application of LSZ to Puerto Rico

Puerto Rico is located in the northeastern Caribbean, 1,280 miles off the coast of Florida. The island is roughly rectangular and oriented along an east–west axis (Fig. 1). It measures approximately 160 km east to west and 55 km north to south, for a total area of a little less than 8500 km^2. (Its political territory includes several smaller islands that are not considered in this study.) Climate over the island varies significantly due to its varied topography, ranging from humid-tropical in the central mountain range and north coast to seasonal dry in the southern coastal plain (Larsen and Simon 1993). The prevailing trade winds are from the east and northeast, and much of the rainfall during the May through December wet season is associated with these winds. Annual rainfall ranges between 760 and more than 5,000 mm, with variations mostly due to the changes in land elevation over the central mountain range. The mean annual temperature varies with elevation, ranging from 23 to 27°C along the coastal plains to 19–23°C at the higher peaks. Topographically, the island is characterized by flat coastal areas and two main mountain ranges, the Cordillera Central mountain range in central Puerto Rico (maximum peak of 1,338 m) and the Sierra de Luquillo mountain range in the northeast (maximum peak of 1,074 m). The mountainous areas have common peak heights of 700–1000 m and are typified by moderate and steep slopes, with a pronounced perennial drainage network and several ephemeral streams (Larsen and Simon 1993; Monroe 1980; Pike 2006; Van der Molen 2002). The distinctive landforms of Puerto Rico are the result of the great variety of rock types present on the island, each with different erosional characteristics (Monroe 1980).

In order to apply LSZ methods to Puerto Rico, a database was constructed in the form of several maps, each representing a possible inducing factor for rainfall-induced landslides on the island of Puerto Rico. The spatial database was built using ArcGIS/ArcInfo software (ESRI, Redlands, California) and includes eight inducing factors: slope aspect, slope, land-surface elevation, curvature, distance from geological discontinuities (faults), distance from roads, geology type, and land-cover.

Land-surface elevation plays a key role in the LSZ of Puerto Rico, because of a strong orographic effect on the distribution of rainfall over the island (Daly et al. 2003). Altitude is available from two publicly available Digital Elevation Model (DEM) data sets. The Shuttle Radar Topography Mission (SRTM) provides DEM data at 1- and 3-arc-second resolution for all areas from 60°N to 60°S. The U.S. Geological Survey (USGS) National Elevation Dataset (NED) developed a seamless raster format at the same resolution by merging the best quality elevation data available across the United States. SRTM proved to have larger errors in highly vegetated areas (Simard et al. 2006; Weydahl et al. 2007), therefore the NED data was selected. By means of GIS software, the DEM was processed to define the other surface characteristics of the territory: slope, curvature, and slope aspect. The slope is calculated by isolating the maximum rate of change in the eight cells surrounding each elevation pixel. The curvature of a hillslope is defined as the rate of change of the slope. The aspect of a hillslope indicates the geographic orientation or "facing" direction of the slope. The hillslope aspect of a topographic feature affects its interaction with average precipitation structures, directionality of storms, and rain shadows or orographic enhancement of rainfall. Elevation can be modeled as a continuous variable or binned in classes, depending on the method used. For the FR method, elevation was divided into 50-m classes with a single class for values higher than 500 m above mean sea level (asl). Slope aspect is classified into eight compass orientations (N, NE, E, SE, S, SW, W, NW). Slope, certainly one of the most important factors in slope stability, is here analyzed in 5° bins with one bin for slopes higher than 40°. Curvature is classified as concave, flat, or convex morphology.

Other data included in the LSZ analysis are land use from the National Land-cover Data (NLCD, http://seamless.usgs.gov/faq/nlcd_faq.php); geology and geological discontinuities (Baweic 2001); and distance from roads (Census 2000 TIGER/Line Data, http://www.esri.com/data/download/census2000_tigerline/index.html). The NLCD data set defines several classes of land use. The areas considered in this analysis include only a few of these classes: dryland cropland pasture, grassland, savannah, evergreen needle forest, and herbaceous wetland. The geology of the island, as already mentioned, is complex; the original hundreds of geologic units listed in Baweic (2001) were grouped in five more general classes. Geological discontinuities, of great importance for landslide processes in Puerto Rico (Demoulin and Chung 2007; Monroe 1979), are represented by the distance from faults, divided into four classes. (Data on other geological discontinuities, including bedding planes, fold, joints, were not included as they were not readily available.) Finally, distance from roads, demonstrated as being an important factor in previous analysis (Larsen 1995), is included and binned into five classes. The resulting eight inducing factors, and their corresponding classes, are listed in the first and the second columns of Table 1.

The last and most critical source of information for the LSZ is the representation of historical landslide locations and sizes. Landslide (LS) inventory data are the most important inputs to an empirical or statistical LSZ analysis. Three separate landslide inventories were made available by Larsen (Larsen and Torres-Sanchez, 1998). Larsen and Torres-Sanchez (1998) present a large data set of landslides and use these data as point information only, applying a methodology very similar to the frequency ratio, but limited to the count of LS per unit area. The LS data set covers three areas of the island, Blanco, Coamo, and Cibuco (hereon referred to as "basins" although they are not hydrologic basins), shown in Fig. 1. Larsen and Torres-Sanchez (1998) developed landslide maps from 1:20,000-scale aerial photographs (with a 10-m × 10-m lower limit of observation) and GIS measurements. Landslides were identified by a range of geomorphic indicators, which included a sharp break or disruption in vegetation type, bare soil or soil with little vegetation re-growth, steep head, and side scarps and downslope debris deposits.

Inventories were available from Larsen and Torres-Sanchez (1998) as both mapped landslide areas (as a GIS polygon file) and tabulated landslide characteristics. The mapped data set, while providing the location of failures, is less complete than the tabulated data set in that mapped features are often only the linear alignment of the landslide scar without specification of width or area. In contrast, the tabulated data set includes landslide width and area, but not necessarily consistently with the mapped areas. To take advantage of all the data, the two data sets were matched so as to marry the geographic information from the mapped

Table 1 Landslide inducing factors considered in this work; FR method and LR method results

Factors	Classes		FR method				F	LR method
			Blanco AUC (entry)	Cibuco AUC (entry)	Coamo AUC (entry)	All 3 AUC (entry)		β
Slope aspect			0.68 (II)	0.745 (IV)	0.654 (IV)	0.675 (III)		−13.33 (±0.1) Intercept
	Flat	−1					–	–
	N	22					1.06	9.34 (±0.06)
	NE	67					1.25	9.36 (±0.04)
	E	112					1.19	9.40 (±0.03)
	SE	157					0.97	9.23 (±0.04)
	S	202					0.92	9.11 (±0.03)
	SW	247					0.93	9.10 (±0.04)
	W	292					0.79	8.95 (±0.04)
	NW	337					0.89	9.18 (±0.03)
Slope		5	0.701 (IV)		0.657 (V)	0.679 (V)	0.41	–
		10					0.70	0.10 (±0.015)
		15					1.00	0.29 (±0.02)
		20					1.24	0.41 (±0.02)
		25					1.35	0.41 (±0.025)
		30					1.41	0.35 (±0.03)
		35					1.29	0.25 (±0.05)
		40					1.51	0.32 (±0.07)
		>40					1.51	0.33 (±0.08)
Elevation		50	0.683 (I)	0.721 (I)	0.639 (I)	0.644 (I)	0.16	–
		100					0.40	0.69 (±0.03)
		150					0.51	0.82 (±0.02)
		200					0.63	0.94 (±0.01)
		250					0.78	1.02 (±0.02)
		300					1.04	1.30 (±0.04)
		350					1.50	1.46 (±0.03)
		400					1.52	1.57 (±0.03)
		450					1.66	1.65 (±0.04)
		>450					1.70	1.69 (±0.03)
Curvature	I	Convex	**0.71 (VII)**		**0.66 (VIII)**		1.10	
	II	Flat					0.85	
	III	Concave					1.17	
Dist. faults	I	100	0.703 (V)		0.658 (VI)	0.678 (IV)	1.24	–
	II	300					1.28	0.008 (±0.01)
	III	600					1.18	−0.09 (±0.03)
	IV	>600					0.95	−0.265 (±0.03)
Dist. roads	I	30		0.744 (III)	0.651 (III)		1.05	–
	II	90					1.07	0.0074 (±0.02)
	III	120					1.10	0.03 (±0.02)
	IV	180					1.17	0.07 (±0.01)
	V	>180					0.74	0.074 (±0.03)

Table 1 Landslide inducing factors considered in this work; FR method and LR method results

Factors	Classes		FR method				*F*	LR method
			Blanco AUC (entry)	Cibuco AUC (entry)	Coamo AUC (entry)	All 3 AUC (entry)		*β*
Geology	I	Nonvolclanocastic	0.704 (VI)	0.742 (II)	0.645 (II)	0.667 (II)	0.31	–
	II	Intrusive Terranes					1.72	0.90 (±0.07)
	III	Volcanoclastic					1.10	0.39 (±0.07)
	IV	Submarine Basalt					1.58	0.90 (±0.07)
	V	Alternation					0.65	0.86 (±0.10)
Land-cover	I	Dryland cropland pasture	0.696 (III)	**0.75 (V)**	0.659 (VII)	**0.70 (VI)**	1.04	
	II	Grassland					1.18	
	III	Savannah					0.61	
	IV	Evergreen needle forest					0.99	
	V	Herbaceous wetland					0.81	

Bold values indicate the final maximum AUC reached by the procedure

data set with the dimensional information from the tabulated data set and obtain a reconstructed data set. A reasonable assumption is that the generated landslide inventories have a resolution of ~ 3 m, equal to the minimum width of the landslide. However, the reconstructed data set also was upscaled to 30-m resolution to match the typical resolution of other data that define landslide-inducing factors. The upscaled landslide inventory was used in this study.

Figure 2 shows an example of the landslide inventory in the Blanco basin, with the landslides as originally mapped, the reconstructed landslide areas, and the upscaled areas. To assess the quality of the reconstructed and upscaled data employed in this study, these data are compared with previous landslide inventory analyses. Malamud et al. (2004) examined three well-documented landslide inventories and found that the landslides areas are well described by a three-parameter inverse Gamma distribution, and, for large values of landslide area, by an approximated inverse gamma distribution. In both cases, the power-law decay was found to approximately equal 2.4. Figure 3 shows the distribution of the tabulated areas, the reconstructed data, and the upscaled product from this study. The tail of the upscaled data set aligns well with Malamud et al.'s (2004) results (black continuous line). When the complete and approximated distribution (grey continuous and grey dashed lines) are fitted, respectively, to the reconstructed and to the upscaled data, a decay of about 2.55 is found, a close match to Malamud et al.'s decay. The complete inverse gamma distribution (grey continuous line) differs

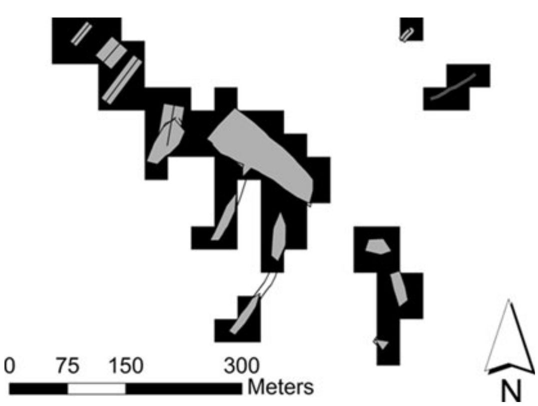

Fig. 2 Particular of the landslide inventory (Blanco area): the polylines given in the GIS files (*thin black lines*); the reconstructed areas (*grey areas*); and the upscaled landslide areas at 30 m (*black areas*)

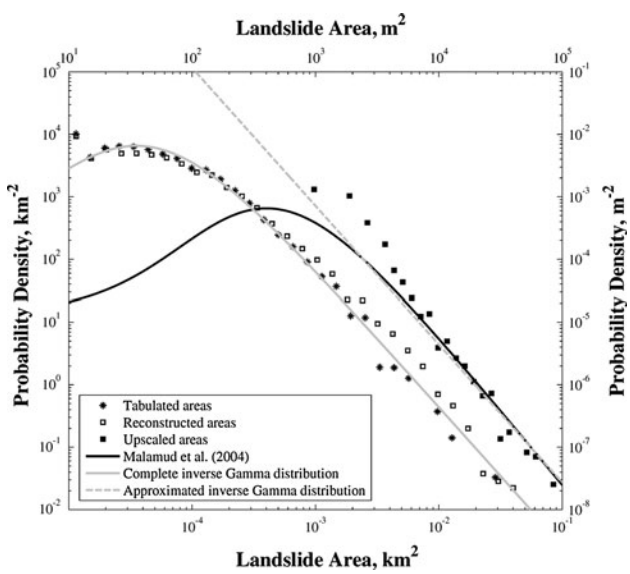

Fig. 3 Probability density of the landslide areas for the tabulated, reconstructed and upscaled data. The inverse gamma distribution obtained by Malamud et al. (2004) (*black line*), and the complete and approximated inverse gamma distribution fitted to the data from this work (*continuous and dashed grey lines*)

from Malamud et al.'s in the location of the rollover area. The left tail of data distribution in this study may suggest a more detailed description, in terms of resolution, of low LS area values in this inventory compared to the one used in Malamud et al.'s study. Figure 3 validates the landslide inventory by showing how its statistical properties agree well with previous broad studies of landslide probability density.

Results

Frequency ratio method

For the application of the FR method, the empirical distributions of $A_{L|fi}$ and A_{fi} are calculated for all three basins, Blanco, Cibuco, and Coamo, included in Larsen and Torres-Sanchez's (1998) data. Figure 4 portrays the landslide area (LS area, $A_{L|fi}/A_L$) and no-landslide area (NOLS area, A_{fi}/A) distributions for all eight factors. Figure 4 plots $A_{L|fi}/A_L$ and A_{fi}/A for the three separate basins (Blanco, Cibuco, and Coamo in light grey, black, and dark grey, respectively) and for all three together (black circles). Determination of the frequency ratio value by Eq. 1 corresponds to taking the ratio of the second row over the first row of Fig. 4. The distribution of the areas with respect to aspect is different in the three basins. The Coamo basin, located in the southern part of the island, has very few pixels oriented towards N to E. Similarly, the Cibuco basin, located in the northern part of the island, has little area oriented towards S. In contrast to the aspect, the distributions for the slope and elevation of NOLS areas are very similar among the three basins, whereas LS areas differ. Distance from roads, distance from faults, and curvature all have similar distributions within the three basins, whereas geology and land use can be very different depending on which of basins are considered.

The circles in Fig. 4 represent the LS and NOLS area distributions for the three basins together. Often they simply smooth the distributions obtained locally, while in some cases they can be very different from the other three distributions. Figure 4 makes clear that the F_f values obtained for the three basins separately can be different from those obtained considering the three basins together. Therefore, the goodness of fit of the model based only on local information (e.g., only on the Blanco basin) can be higher than that obtained using all three basins. On the other hand, using all the information available allows one to gain more complete knowledge about the landslides process on the island.

Which landslide-inducing factors to include in the final SI calculation were decided using a forward procedure based on the AUC of the ROC. SI was computed for different combinations of inducing factors; a specific factor was included in the final SI only if it increased the total AUC value. The selection algorithm stopped when no further increase of AUC was possible. Because of the clear differences shown by the data in Fig. 4, the analysis was repeated both for the three basins separately and for the case of a general model (when all the three basins were considered as one). The AUC values and the entry order for each of the factors are listed in Table 1 in columns 3–6. The F_f values are reported only for the case when all three basins are considered together (all 3); for this case, the final SI for the FR method can be obtained as:

$$\mathrm{SI} = F_{\mathrm{Elev}} + F_{\mathrm{Geo}} + F_{\mathrm{Asp}} + F_{\mathrm{Df}} + F_{\mathrm{Slp}} + F_{\mathrm{land\ cover}}. \tag{6}$$

Logistic regression

The areas considered in this study are quite large: at a 30-m resolution, they include for the combined three-basin data set a total of more than 1,000,000 NOLS and 11,000 LS pixels. It would be computationally impractical to consider all inducing factors in all pixels and, as stated in the

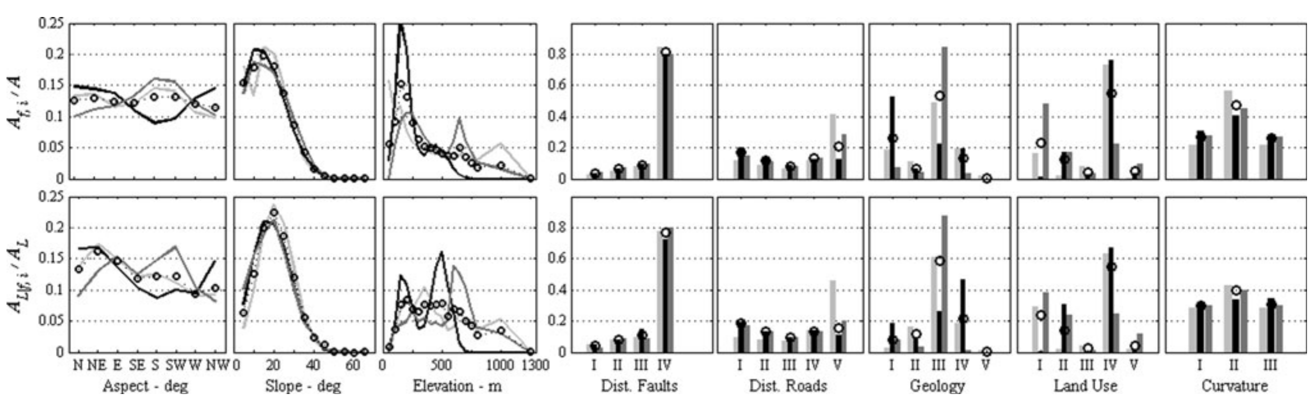

Fig. 4 NOLS ($A_{f,i}/A$) and LS ($A_{L|f,i}/A_L$) area distributions for all eight inducing factors: *light grey lines* and *bars* for Blanco basin, *black* for Cibuco basin, and *dark grey* for Coamo basin. The *circles* are for the case when all the three basins are considered together

discussion of the methods, including all pixels might bias the final results by overweighting NOLS pixels. On the other hand, extracting a smaller subset of pixels from the whole territory may lose some of the features of the process and bias the final results. The approach presented here falls in between the two options: 10 samples are generated, composed of all 11,000 LS pixels and of randomly selected NOLS pixels. The number of NOLS pixels ($Y = 0$ occurrences) is equal to 10 times the number of LS pixels; it is a large sample, yet not the whole area.

Once the samples are generated, the second step of the application is to define if any type of classification is needed. Here two variations are considered: the first with five factors (elevation, aspect, slope, distance from faults, and distance from roads) as continuous variables and the remaining three binned; the second with all eight landslide-inducing factors binned as in the FR application. As mentioned above, when any continuous observable is included in the logistic model, the assumption of linearity of Eq. 4 should be checked. In our data, the relationship is not linear, we therefore choose to use all eight landslide-inducing factors binned in classes.

The 10 samples were then fitted with the logistic regression model and the parameter values estimated by maximum likelihood. The factors to include were determined by three different procedures: a stepwise procedure based on the AIC index (stepAIC), a forward method based on the 2LL index (for2LL), and a forward procedure based on the AUC similar to that used with the FR method (forAUC). The use of three different procedures leads to slightly different determinations of included factors. The stepAIC method was the least parsimonious and for most of the cases included in the model all eight factors. The for2LL and forAUC methods converged on the same factors. The forAUC procedure led to results almost identical to the for2LL, confirming its ability to determine the correct factors when other statistically based procedures are not available. Table 1, column 8, reports the average values of the β coefficients obtained from the 10 samples and the number in brackets gives the variability of the estimates; the first bin of each inducing factor is missing its coefficient because of the dummy variable coding. The variability of the estimates is mostly very low and shows that the results are robust and representative of the entire area of the three basins. Introducing these coefficients in Eq. 4, we obtain:

$$z(x) = \beta_0 + \sum_{i=1}^{9} \beta_{\mathrm{Elev},i} X_{\mathrm{Elev},i} + \sum_{j=1}^{4} \beta_{\mathrm{Geo},j} X_{\mathrm{Geo},j} + \sum_{k=1}^{8} \beta_{\mathrm{Asp},k} X_{\mathrm{Asp},k} + \sum_{m=1}^{3} \beta_{\mathrm{Df},m} X_{\mathrm{Df},m} + \sum_{n=1}^{8} \beta_{\mathrm{Slp},n} X_{\mathrm{Slp},n} + \sum_{p=1}^{4} \beta_{\mathrm{Dr},p} X_{\mathrm{Dr},p}. \quad (7)$$

with the coefficients $\beta_{\mathrm{Elev},i},\ldots,\beta_{\mathrm{Dr},p}$ varying depending on the bin value. The final probability of failure P_L is obtained substituting Eq. 7 in Eq. 5.

Discussion

The final susceptibility maps obtained by the FR and LR methods both include six of the eight inducing factors. The factors for aspect, slope, elevation, distance from faults, and geology are chosen by both models, FR adds land-cover whereas LR selects distance from road.

Figure 5 shows the final ROC curves, in panels a and b for the FR models and panel c for the LR model. Panel a shows the ROCs for the three local models based on each basin separately; panel b instead evaluates the goodness of fit of the general model, obtained by aggregating the three basins. In panel b, the ROCs and the corresponding AUC values for the three separate basins are recalculated to show

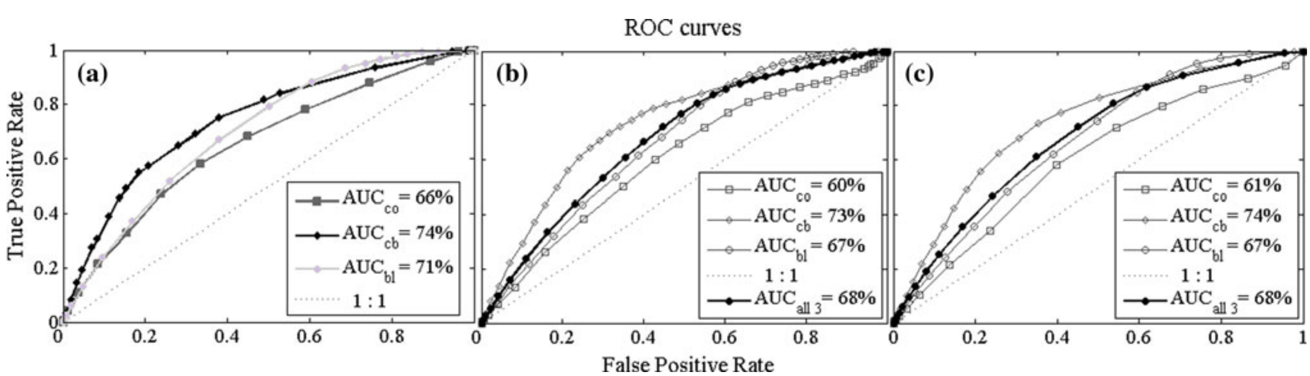

Fig. 5 ROC curves for the considered models: **a** ROC curves for local FR models for the three basins; **b** ROC curves for the general FR model applied to all three basins together and separately; **c** ROC curves for the LR model for all three basins together and separately. The *legend boxes* report the AUC values corresponding to each ROC curve; *subscripts* refer to basins as follows: *bl* Blanco, *ci* Cibuco, *co* Coamo

how the use of a more general model affects the goodness of fit in those three basins. Cibuco is the least affected by the general model, going from an AUC value from 74 to 73%, the Blanco AUC value is reduced from 71 to 67%, and Coamo goes from 66 to 60%. These results illustrate the good results obtained by this application, and an appreciable improvement in the goodness of fit when the local models are employed.

The LR method was not applied to the three basins separately, therefore panel c reports the ROC curves for the general model based on the logistic regression. The results in terms of AUC values are virtually identical to those obtained by the general model with the FR method. The application of the LR method is considerably more complex than the FR method. The results of the two applications, although different, are indistinguishable in terms of ability to fit the LS data. The FR method is therefore the preferable methodology for areas as large as the island of Puerto Rico.

The FR, and similarly the LR, results effectively capture the landsliding process of the three areas analyzed. In order to also investigate the feasibility of extending these findings to the whole island, the consistency of these results was tested with previous results, and the validity of the map for the whole island was assessed by comparison with other landslide inventories.

Having used the data set presented by Larsen and Torres-Sanchez (1998), good agreement with their results can be expected. Larsen and Torres-Sanchez found in slopes greater than 12° a threshold above which the frequency of landslides increases. Similarly, in Table 1, column 7, the F values for slope are greater than 1 for slope equal to or steeper than 15°. The findings in Table 1 also match Larsen and Torres-Sanchez relative to elevation, with a similar threshold behavior for elevations greater than 300 m asl, and for aspect with north and northeast facing slopes. Larsen and Torres-Sanchez did not include information related to geology, such as geologic formation or distance from geological discontinuities. Both factors, based on Monroe (1979) and on these analyses, are important landslide-inducing factors and improve the accuracy of the LSZ.

LSZ maps for local areas within Puerto Rico have been developed by Larsen and Parks (1998) and Larsen et al. (2004). Larsen and Parks (1998) developed a LSZ for the county of Comerio by considering only elevation, slope, aspect, and land use, whereas the LSZ for Ponce developed by Larsen et al. (2004) considered only slope and geology. In both cases, important inducing factors, such as geologic formations in the case of Comerio and elevation for Ponce, are left out.

In Fig. 6, panels a and b compare Larsen and Parks' (1998) map with the results of this study. Larson and Parks extended the results obtained by Larsen and Torres-Sanchez (1998) to Comerio County. They added information related to historical landslides used by Monroe (1979) (see area 1 in panel a and cross-hatched area in panel b). In panel b, the SI obtained by the FR method has been divided into four susceptibility classes, from low to very high. Although there is a certain consistency between the two maps, the map in panel b is dominated by high and very high susceptibility, whereas that in panel a is mostly moderate.

Panel c maps Larsen et al.'s (2004) results for the county of Ponce together with the location of some of the landslides registered after the heavy rainfall associated with the passage of Tropical Storm Isabel (Jibson 1989). (Landslides were mapped within an area of 362 km^2, delimited by the two horizontal dashed lines in panels c and d, and not for the whole area impacted by the rainfall event. The event was in fact more significant over the mountains north of the analyzed area and where panel d predicts high susceptibility to landslides.) According to Larsen et al. (2004), the data set used for the development of the LSZ of this small portion of the island is quite large, but was derived from a single event with highly localized and extremely intense rain cells. This makes the results not representative of an average rainfall condition but of a specific event and Larsen et al. note that the landslide distribution is influenced by the spatial variation in rainfall intensity during the storm. When panel c is compared to panel d there are some similarities in the extent of the susceptibility map, influenced by the very low elevation and slopes present in the coastal area. On the other hand, the results in panel d identify a band of moderate susceptibility, mostly corresponding to lower elevations and gentler slopes, whereas Larsen et al. map almost the entire county as highly susceptible. Also, their results are influenced by very local geologic formations (panel e, areas 2 to 4), which were not considered in the broader and more aggregated geologic classification used here. Finally, in panel d, while many of the mapped landslides fall in the high and very high susceptibility areas, some fall in the moderately susceptible area. This is consistent with the conditions of a tropical storm: the level of moderate susceptibility represents a lower probability of failure than a high or very high level but the rarer failures are associated with extraordinary events such as tropical Storm Isabel.

The only LSZ available for the whole territory is the one developed by Monroe (1979); his map gives a great amount of information about how geology plays an important role in the landsliding process on the island. However, Monroe (1979) defined almost the whole territory as moderately susceptible to landsliding, and the only areas mapped as highly susceptible are mostly active landslides at the time of the study. The study reported here, instead, aims at

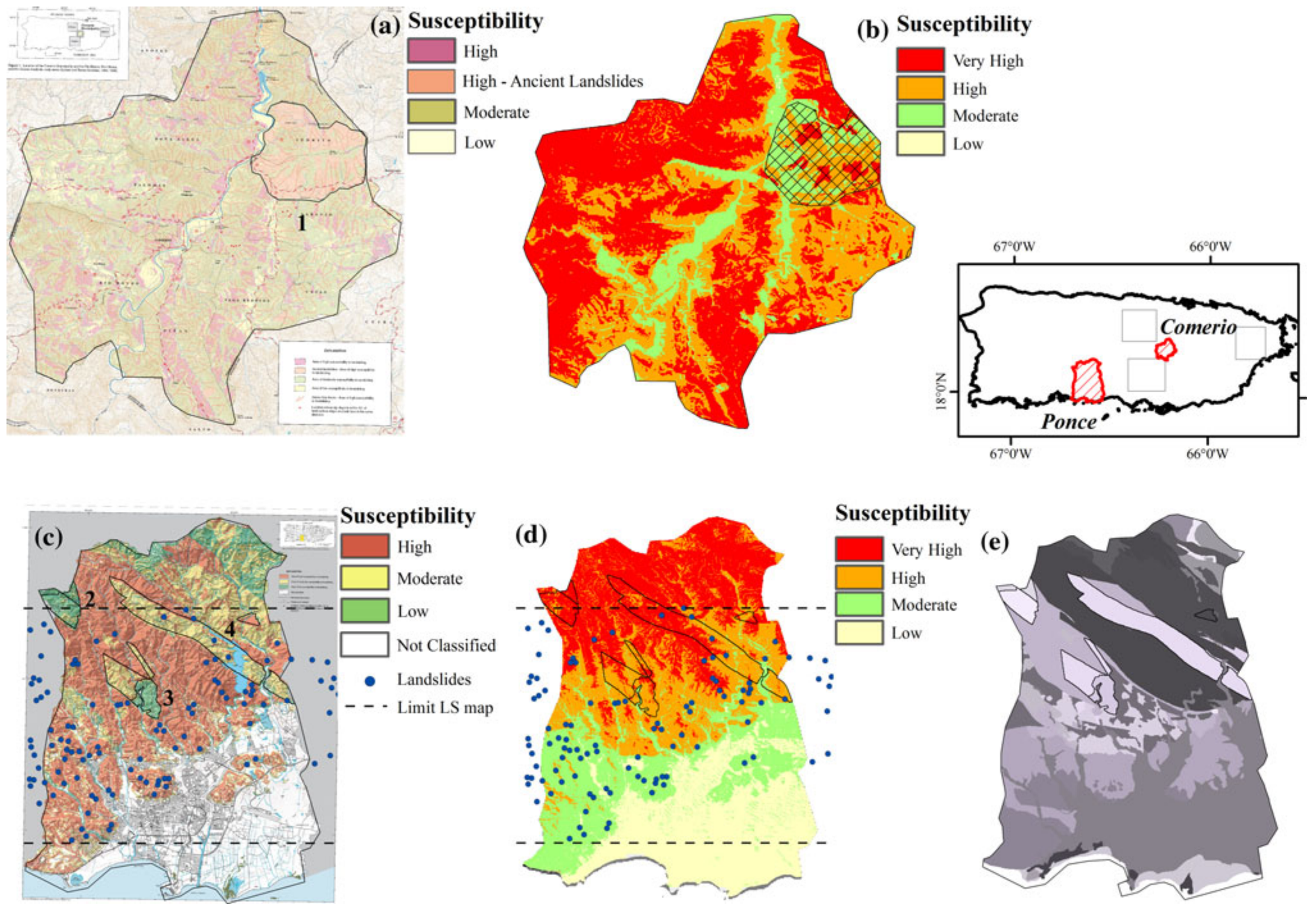

Fig. 6 Comparison of the FR method's results with previous literature. The first two panels show the Comerio area. Panel **a** reports the findings of Larsen and Parks (1998), whose analysis considered elevation, slope, aspect, and land use, and panel **b** shows results from this work. The last three panels show the Ponce region of Puerto Rico. Panel **c** reproduces the findings of Larsen et al. (2004), whose analysis considered only geology and slope, and panel **d** shows results from this work. Panels **c** and **d** show the landslides mapped after the tropical storm Isabel and the limit of this landslide inventory (*dashed lines*). Panel **e** depicts the geologic map for the area. Areas meriting special discussion are numbered 1 through 4 and addressed further in the text

defining the susceptibility of the territory of Puerto Rico based on more complete and detailed information on landslide-inducing features.

Pando et al.'s (2005) landslide inventory allows for an opportunity to evaluate qualitatively the susceptibility map obtained here for the whole island. Pando et al.'s inventory is the result of an extensive compilation of storms that triggered landslides between 1959 and 2003. Figure 7 shows the susceptibility maps obtained by the application of the FR method. The top row of Fig. 7 shows maps estimated using the local models for the three basins separately, while the bottom row shows results when the general model was applied to the whole island. The final SI values were divided into four categories, low (sand color), moderate (green), high (orange), and very high susceptibility (red). Note that different breaks in the SI will lead to different looking maps; the three breaks considered correspond to AUCs of 20, 50, and 80%. In the top row, the landslides that served to define LS-areas for our analysis are mapped—they appear as small black flecks on the maps. In the bottom row, the irregular cross-hatched areas depict the areas defined by Monroe (1979) as highly susceptible to landslides, and the hatched squares and circles show landslide locations given by Pando et al. (2005); the squares map landslides older than 1993 and the circles update the previous inventory with events recorded between 1993 and 2005. By comparing Pando et al.'s distribution of actual landslides with the final map, the validity of extending the results from this study to the whole island can be assessed.

Pando et al.'s landslide inventory confirms the high susceptibility of the northern outcrop belts identified by Monroe (1979). These belts are largely missing from the data set used in this application (only a small part of the belts lie within the Cibuco data set) and thus the factors that contribute to their vulnerability were inadequately covered in the analysis. In the absence of a more comprehensive LS data set that better covers the belts, a possible and reasonable way to consider Monroe's recommendation is to

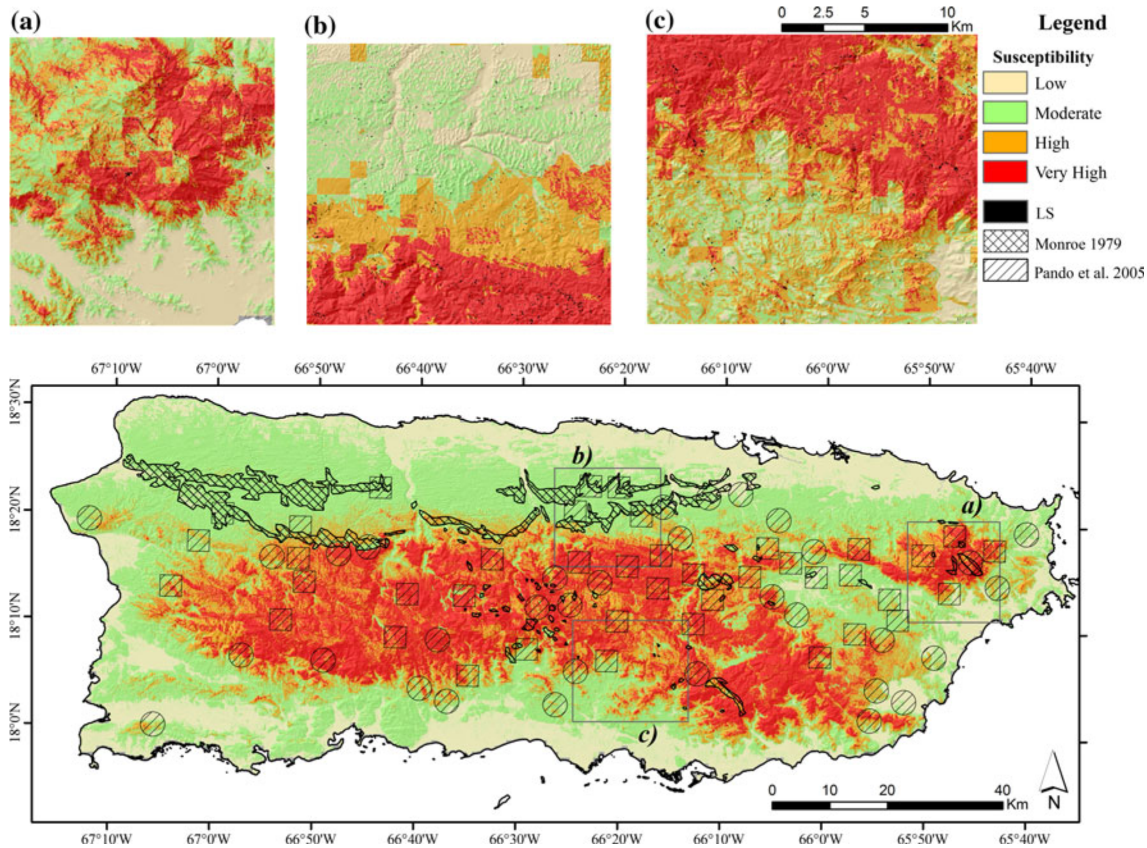

Fig. 7 Susceptibility map for the three local models (*top row*) for Blanco (**a**), Cibuco (**b**), and Coamo (**c**) basins and susceptibility map for the whole island of Puerto Rico (*bottom row*). The *hatched squares* and *circles* are adapted from Pando et al. (2005); the *irregular cross-hatched areas* represent high susceptibility areas as identified by Monroe (1979)

simply designate those belts as highly susceptible regardless of the SI values obtained. Larsen and Parks (1998) adopted a similar strategy when they defined the "ancient landslides" area for other geologic formations they knew to be highly vulnerable. On the other hand, Pando et al.'s data show how the LSZ developed by Monroe (1979) strongly underestimates the susceptibility of most of the island—in fact almost all Pando et al.'s landslide locations fall in Monroe's moderately susceptible class. In contrast, Pando et al.'s landslide locations agree well with the results presented here: the great majority of Pando et al.'s events fall within those areas indentified as having high or very high susceptibility to landsliding. Only a few of their landslides are located in areas mapped as moderate susceptibility and only one in an area of low susceptibility. The mapping also coincides with some isolated events on the coastal plains at the west of the island.

Conclusions

Two widely applied methodologies for LSZ—one based on a bivariate frequency ratio (FR method) and the other on logistic regression (LR method)—have been applied to Puerto Rico. Both methodologies correctly capture factors known to induce rainfall-triggered landslides in Puerto Rico: they both indicate aspect, slope, elevation, geological discontinuities, and geology as highly significant landslide-inducing factors, together with land-cover for the FR method, and distance from roads for the LR method. The LR method is a complete and robust methodology, grounded on rigorous statistical testing and model building, and relatively complex to apply. The FR method has the advantage of a simpler and more linear application. Despite its greater rigor, the LR method did not improve results over the much simpler FR method. Accordingly, the FR method was selected to generate the landslide susceptibility map for the island of Puerto Rico based on prior landslide inventories. The resulting landslide susceptibility predictions were compared with previous landslide analyses and a landslide inventory not considered in the analysis. This independent evaluation demonstrated that these methods were consistent with the LSZ from those earlier studies and that the results obtained by analyzing the three basins of Blanco, Cibuco, and Coamo successfully captured the landsliding process over the whole island and greatly improved the most recent island-wide zonation map. The analysis results in a robust and verifiable landslide zonation map for the island of Puerto Rico, which can be used in large scale planning and to assess, for example, the suitability of an area for urban development. Construction in areas of high to very high landslide susceptibility should predicated on an extensive site evaluation by engineering geologists and perhaps be limited to minor modifications of preexisting structures. As previously suggested, LSZs should also help in educating residents, municipal planners, and civil defense personnel about the risk of landslides in a territory, allowing them to work together to protect the populace from the hazards of landslides.

Acknowledgments The work has been funded by the National Aeronautics and Space Administration under Project NASA NNX07AD29G, Predicting Landslides Using Measurements of Precipitation from Space. We are thankful to Matthew C. Larsen, Associate Director for Water, U.S. Geological Survey, for providing landslide data. We appreciate the assistance and data provided by Prof. Frederick N. Scatena of the Department of Earth and Environmental Science, University of Pennsylvania, and the International Institute of Tropical Forestry of the U.S. Department of Agriculture Forest Service, Río Piedras, Puerto Rico. We gratefully acknowledge reviewer comments.

References

Allison PD (2008) Convergence failures in logistic regression. In: Proceedings of the SAS Global Forum 2008 Conference. SAS Institute Inc., Cary, NC. http://www2.sas.com/proceedings/forum2008/360-2008.pdf

Arora MK, Gupta RP (2004) An artificial neural network approach for landslide hazard zonation in the Bhagirathi (Ganga) Valley, Himalayas. Int J Remote Sens 25(3):559–572

Baweic WJ (2001) Geology, geophysics, mineral occurrences, and mineral resource assessment for the Commonwealth of Puerto Rico. US Geological Survey Open-File Report 98-38, one CD-ROM

Briggs RP, Akers JP (1965) Hydrogeologic map of Puerto Rico and adjacent islands. US Geological Survey Hydrologic Investigations Atlas HA-197

Campbell RH, Herd DG, Alonso RM (1985) Preliminary response activities and recommendations of the USGS Landslide Hazard Research Team to the Puerto Rico landslide disaster of October 7, 1985. US Geological Survey Open-File Report 85-719:13

Chau KT, Chan JE (2005) Regional bias of landslide data in generating susceptibility maps using logistic regression: case of Hong Kong Island. Landslides 2(4):280–290

Cheng Z, Wang J (2007) Landslide hazard mapping using logistic regression model in Mackenzie Valley, Canada. Nat Hazards 42:75–89

Dai FC, Lee CF (2002) Landslide characteristics and slope instability modeling using GIS, Lantau Island, Hong Kong. Geomorphology 42:213–228

Daly C, Helmer EH, Quinones M (2003) Mapping the climate of Puerto Rico, Vieques and Culebra. Int J Climatol 23:1359–1381

Demoulin A, Chung C-JF (2007) Mapping landslide susceptibility from small datasets: a case study in the Pays de Herve (E Belgium). Geomorphology 89(3–4):391–404

Donati L, Turrini MC (2002) An objective method to rank the importance of the factors predisposing to landslides with the GIS methodology: application to an area of the Apennines (Valnerina; Perugia, Italy). Eng Geol 63(304):277–289

Einstein HH (1988) Special lecture: landslide risk assessment procedure. In: Proceedings of 5th international symposium on landslides, Lausanne, Switzerland, pp 1075–1090

Fawcett T (2006) An introduction to ROC analysis. Pattern Recognit Lett 27(8):861–874

Guzzetti F, Carrara A, Cardinali M, Reinchenbach P (1999) Landslide hazard evaluation: a review of current techniques and their application in a multi-scale study, Central Italy. Geomorphology 31:181–216

Guzzetti F, Reinchenbach P, Cardinali M, Galli M, Ardizzone F (2005) Probabilistic landslide hazard assessment at the basin scale. Geomorphology 72:272–299

Hong Y, Adler R, Huffman G (2007) Use of satellite remote sensing data in the mapping of global landslide susceptibility. Nat Hazards 43:245–256

Hosmer DW, Lemeshow S (2000) Applied logistic regression, 2nd edn. John Wiley, New York

Jibson RW (1986) Evaluation of landslide hazards resulting from the 5–8 October 1985 storm in Puerto Rico. US Geological Survey Open-File Report 86-26

Jibson RW (ed) (1989) Debris flows in southern Puerto Rico. In: Schultz AP, Jibson RW (eds), Landslide processes of the Easter United States and Puerto Rico. Geological society of America Special Paper 236

Jimenez-Peralvarez JD, Irigaray C, El Hamdouni R, Chacon J (2009) Building models for automatic landslide-susceptibility analysis, mapping and validation in ArcGIS. Nat Hazards 50(3):571–590

Larsen MC (1995) How wide is a road? The association of roads and mass wasting in a forested montane environment, Puerto Rico. Am Geophys Union EOS Suppl 76(17):309

Larsen MC, Parks JE (1998) Map showing landslide susceptibility in the Comerìo Municipality, Puerto Rico. U.S. Geological Survey Open File Report 98-566

Larsen MC, Simon A (1993) A rainfall intensity-duration threshold for landslides in a humid-tropical environment, Puerto Rico. Geogr Ann Ser A 75(1/2):13–23

Larsen MC, Torres-Sanchez AJ (1998) The frequency and distribution of recent landslides in three montane tropical regions of Puerto Rico. Geomorphology 24(4):309–331

Larsen MC, Santiago M, Jibson RW, Questell E (2004) Map showing landslide susceptibility in the municipality of Ponce, Puerto Rico. US Geological Survey Scientific Investigations Map Report Number: I-2818 (http://pubs.usgs.gov/sim/2005/2818/)

Lee S, Pradhan B (2006) Probabilistic landslide hazards and risk mapping on Penang Island, Malaysia. J Earth Syst Sci 115(6):661–672

Lee S, Pradhan B (2007) Landslide hazard mapping at Selangor, Malaysia using frequency ratio and logistic regression models. Landslides 4:33–41

Lee S, Ryu J-H, Kim I-S (2007) Landslide susceptibility analysis and its verification using likelihood ratio, logistic regression, and artificial neural network models: case study of Youngin, Korea. Landslides 4:327–338

Liao SSC, Veneziano D, Whitman RV (1988) Regression models for evaluating liquefaction probability. J Geotech Eng 114(4):389–411

Malamud BD, Turcotte DL, Guzzetti F, Reinchenbach P (2004) Landslide Inventories and their statistical properties. Earth Surf Proc Land 29:678–711

Mathew J, Jha VK, Rawat GS (2008) Landslide susceptibility zonation mapping and its validation in part of Garhwal Lesser Himalaya, India, using binary logistic regression analysis and receiver operating characteristic curve method. Landslides 6(1):17–26

Monroe WH (1979) Map showing landslides and areas of susceptibility to landsliding in Puerto Rico. U.S. Geological Survey, Miscellaneous Investigations Series, MAP I-1148

Monroe WH (1980) Some tropical landforms of Puerto Rico. US Geological Survey Professional Paper 1159, 39 pp

Ohlmacher GC, Davis JC (2003) Using multiple logistic regression and GIS technology to predict landslide hazard in northeast Kansas, USA. Eng Geol 69(3–4):331–334

Pando MA, Ruiz ME, Larsen MC (2005) Rainfall-induced landslides in Puerto Rico: an overview. American Society of Civil Engineers, Geo-Institute & Geosynthetics 2005 Conference, Austin, TX

Pike AS (2006) Application of digital terrain analysis to estimate hydrological variables in the Luquillo Mountains of Puerto Rico. In: Proceedings of the fifth FRIEND world conference, Havana, Cuba. IAHS, St. Louis, pp 81–85

R Development Core Team (2005) R: a language and environment for statistical computing

Santacana N, Baeza B, Corominas J, De Paz A, Marturia J (2003) A GIS-based multivariate statistical analysis for shallow landslide susceptibility mapping in La Pobla de Lillet Area (Eastern Pyrenees, Spain). Nat Hazards 30:281–295

Santiago M, Larsen MC (2001) Earthquake-induced landslide susceptibility in the San Juan Metropolitan Area, Puerto Rico. US Geological Survey Open-file Report 01-39, 1 CD

Shiels AB, West CA, Weiss L, Klawinski PD, Walker LR (2008) Soil factors predict initial plant colonization on Puerto Rican landslides. Plant Ecol 195(2):165–178

Simard M, Zhang K, Ross MS, Rivera-Monroy VH, Castaneda-Moya E, Twilley R (2006) Using shuttle radar topography mission elevation data to map mangrove forest Height in the Caribbean. IEEE 0-7803-9510-7/06:1713-1716

Van der Molen MK (2002) Meteorological impacts of land use change in the maritime tropics

Van Westen CJ (2000) The modelling of landslide hazards using GIS. Surv Geophys 21:241–255

Van Westen CJ, Rengers N, Terlien MTJ, Soeters R (1997) Prediction of the occurrence of slope instability phenomena through GIS-based hazard zonation. Geol Rundsch 86(2):404–414

Van Westen CJ, Van Asch TWJ, Soeters R (2006) Landslide hazard and risk zonation—why is it still so difficult? Bull Eng Geol Environ 65:167–184

Varnes DJ, IAEG (1984) Landslide hazard zonation: a review of principles and practice. The UNESCO Press, Paris

Walker LR, Shiels AB (2008) Post-disturbance erosion impacts carbon fluxes and plant succession on recent tropical landslides. Plant Soil 313(1/2):205–216

Walker LR, Zarin DJ, Fetcher N, Myster RW, Johnson AH (1996) Ecosystem development and plant succession on landslides in the Caribbean. Biotropica 28:566–576

Weydahl DJ, Sagstuen J, Dick OB, Ronning H (2007) SRTM DEM accuracy assessment over vegetated areas in Norway. Int J Remote Sens 28(16):3513–3527

Zhu L, Huang J-F (2006) GIS-based logistic regression method for landslide susceptibility mapping in regional scale. J Zhejiang Univ Sci A 7(12):2007–2017