Predicting the landslides triggered by the 2009 96E/Ida tropical storms in the Ilopango caldera area (El Salvador, C.A.): optimizing MARS-based model building and validation strategies

E. Rotigliano[a], C. Martinello[a], M.A. Hernandéz[a,b], V. Agnesi[a], C. Conoscenti[a]

[a]Dipartimento di Scienze della Terra e del Mare – Università degli Studi di Palermo
[b]Facultad de Agronomía, Escuela de Posgrado – Universidad de El Salvador

Corresponding author: Christian Conoscenti – e.mail christian.conoscenti@unipa.it

Abstract
The main topic of this research was to evaluate the effect in the performance of stochastic landslide susceptibility models, produced by differences between the triggering events responsible for the calibration and validation datasets. In the Caldera Ilopango area (El Salvador), MARS (Multivariate Adaptive Regression Splines)-based susceptibility modeling was applied using a set of physical-environmental predictors and two remotely recognized landslide inventories: one dated at 2003 (1503 landslides), which was the result of a normal rainfall season, and one which was produced by the combined effect of the Ida hurricane and the 96E tropical depression in 2009 (2237 landslides). Both the two event inventories included shallow debris- flow or slide landslides, which involved the weathered mantle of the pyroclastic rocks that largely outcrop in the study area. To this aim, different model building and validation strategies were applied (self-validation, forward and backward chrono-validations), and their performances evaluated both through cut-off dependent and independent metrics. All of the tested models produced largely acceptable *AUC* (Area Under Curve) values, albeit a loss in the predictive performance from self-validation to chrono-validation was observed. Besides, in terms of positive/negative predictions, some critical differences arose: using the 2009 extreme landslide inventory for calibration resulted in higher sensitivity but lower specificity; conversely, using the 2003

normal trigger landslide calibration inventory led to higher specificity but lower sensitivity, with relevant increasing of Type-II errors. These results suggest the need of investigating the extent of such effects, taking multi-trigger intensities inventories as a standard procedure for susceptibility assessment in areas where extreme events potentially occur.

Key-words: landslide susceptibility, MARS, temporal validation, Ida hurricane, Caldera Ilopango (El Salvador).

1. Introduction

Landslides are among the most important causes of natural hazard in El Salvador (Rose et al. 2004), being triggered either by earthquakes or tropical storms. In particular, storms which frequently hit the country are responsible for the multiple activation of a large number of shallow and fast moving flow-like landslides, which cause life losses and severe damages disaster scenarios (e.g., CEPAL 2010; MARN 2011) either directly impacting along the slopes against inhabited areas or feeding debris floods phenomena along the streams. Predicting storm triggered multiple occurring landslides is of great importance in these areas where very steep slopes mantled by weathered pyroclastic rocks are exposed to such recurrent storm inputs, as in large part of Central America. In fact, landslide susceptibility models and their derived maps are among those tools that allow for cost/effectively mitigating the natural risk associated to storm events.

The territory of El Salvador is largely characterized by a young deeply incised tephra dominated landscape, so that a history of recurrent debris flow disaster events is already known. However, in El Salvador few studies have dealt with this topic. A logistic regression model for earthquake-induced landslides in the whole country has been assessed, using an inventory of 2001 seismically-induced landslides for calibration, but not being supported by any validation procedure (García-Rodríguez et al. 2008). García-Rodríguez and Malpica (2010) applied then to the same landslide inventory artificial neural networks method and validated the obtained model by means of Receiver Operating Characteristic (ROC) curves, albeit without *n*-folds precision and reliability analysis. For the extreme north-western sector of the country, principal component analysis was also applied (Kopačková and Šebesta 2007), obtaining a landslide susceptibility map which was calibrated with post-Mitch hurricane (1998) and post-2001 earthquake inventories for a study area of around 3.500 km^2, but again without a complete validation procedures.

On a national basis, MARN (Ministerio de Medio Ambiente y Recursos Naturales) produced in 2004 a landslide susceptibility map for the whole country at a 1:50.000 scale (MARN 2004) by applying a heuristic approach (Mora and Vahrson 1991, 1994) and using five controlling factors. Recently, the El Salvador territory was included in a regional landslide susceptibility scenario, obtained by applying a fuzzy based heuristic approach, with a low spatial resolution of 30 arcseconds (Kirschbaum et al. 2016).

In this paper, a stochastic approach to landslide susceptibility modeling for storm-triggered multiple debris flows events in El Salvador is proposed, exploiting the Multivariate Adaptive Regression Spline (MARS) technique (Friedman 1991; Conoscenti et al. 2015) and a complete validation scheme (Guzzetti et al. 2006; Frattini et al. 2010; Conoscenti et al. 2015, 2016; Lombardo et al. 2015, 2016; Cama et al. 2016). In particular, we focused on the assessment of a landslide susceptibility model for storm triggered landslides in the north-western inner slopes of the Ilopango Caldera in El Salvador (Figs. 1), where at the end of the first decade of November 2009 a multiple landslides event occurred, due to the combined effect of the Ida Hurricane and the 96E tropical depression (CEPAL 2010, 2011; MARN 2010, 2011).

Susceptibility models were prepared by applying MARS to regress a binary landslide-derived outcome (stable/unstable status) on a set of geo-environmental explanatory variables, which were derived from two available thematic (geology and soil use) maps and a 10m-cell digital elevation model (DEM). The model building and validation scheme exploited two original pre- and post-event landslide inventories (2003 and 2009, respectively), which were recognized by means of the Google Earth™ (GE) image databank and 3D-view integrated system (e.g., Costanzo et al. 2012a).

The main topic of this research was to evaluate the effect in the temporal prediction performance of a susceptibility models, produced by differences in the magnitude of the triggering events which caused the landslides forming the calibration and validation datasets.

The present research exploited free open data and software resources: Google Earth™, SAGA GIS (Conrad et al. 2015) and R software (R Core Team 2015).

2. Materials

2.1 Study area

El Salvador (Fig. 1) stretches SE-NW along the Central American Volcanic Front for about 250km in the Pacific side of Central America, near 150km inboard of the Middle American Trench, where the Cocos plate is subducted beneath the Caribbean plate (Agostini et al. 2006; Lexa et al. 2011). In particular, the Ilopango Caldera is located in the central graben system of El Salvador (Fig. 2a), between the coastal and the inner volcanic cordilleras, being one of the most dangerous calderas of Central America, with an area of around 200km^2, less than 20 kilometers east of the city of San Salvador (Rose et al. 2004). At least four exceptional eruptions in the last 100kys, the last of which near 2500 years ago, produced tephra layers and ignimbrites deposits which covered wide sectors of the central part of the country (Stoiber and Carr 1973). The caldera presently hosts a typical volcanic lake (Lake Ilopango), whose inner delimiting steep slopes are characterized by the outcropping of the weathered tephra layers. Such a geomorphologic setting is responsible for a large number of landslides which activate in the occasion of the tropical storm events associated with the recurrent either Atlantic and Pacific cyclones.

The study area (Fig. 2b) corresponds to an oblate catchment (about 5km long and 8km large) given by the convergence of several short highly steep streams into an alluvial plain named “Arenal de Cujuapa”, which progradates into the Ilopango lake with a marked delta-like head (“Punta El Pinar”). Actually, two main channels can be recognized in the alluvial plain, the southernmost of which corresponds to the ending branch of the Rio El Borbollón. The whole catchment drains the inner slopes of the northeastern sector of the Ilopango Caldera, which are characterized by the outcropping of Holocenic acid pyroclastic sequences, locally named “Tierra Blanca” (TB), belonging to the San Salvador formation (Quaternary). The latter covers the underlying pyroclastic deposits of the Cuscatlan formation, which are unburied by erosion along the valley bottom of the streams. Finally, in the upper sectors of the catchment, near the town of Cojutepeque, pyroclastites of the Bálsamo formation outcrop.

2.2 The Ida/96E event and related landslides

The Hurricane Ida developed on the 4^{th} of November as a tropical depression in the south-western sector of the Caribbean Sea, increasing its strength up to tropical storm grade on the 7^{th} of November, when it crossed the shoreline of Nicaragua, and to a second level hurricane at the midday of the 8^{th} (Avila and Cangialosi 2010). The hurricane then moved northward crossing the Caribbean Sea and the Mexico Gulf, weakening back to tropical storm and to depression on the 9^{th} and completely dissipating on the 12^{th}. During these same days, the low-pressure system 96E moved from the eastern Pacific Ocean causing intense rainfall between November 7^{th} and 8^{th} (CEPAL 2000, 2011). In these two days, Ida and 96E simultaneously struck an area of around 400 km^2 centered between Ilopango Lake and San Vicente Volcano, producing more than 300mm/24hrs at the Ilopango and San Vicente villages (Fig. 3).

In this area, large damages were recorded caused by floods and landslides with around two-hundreds deaths and a quarter of a billion dollars of economic losses (MARN 2010), the larger part of which in the north-western flank of San Vicente Volcano, where huge debris flow phenomena severely struck the villages of Verapaz and Guadalupe. At the same time, in the Ilopango Caldera area, hundreds of landslides triggered from steep slopes causing damages to cropland, rural houses and roads, as well as strongly affecting and modifying the connected fluvial system.

In order to prepare the required two (ante- and post-event) landslide inventories, a remote recognition was carried out through a systematic GE-based analysis, which was performed on two different epochs: one dated at 9/10/2003 (DigitalGlobe Catalog ID: 1010010002459C02) and one dated at 11/21/2009 (DigitalGlobe Catalog ID: 101001000AA5D801), the latter being taken just two weeks after the Ida/96E combined event. Unfortunately, the 2003 GE images were affected by a partial cloud coverage, so that the study area had to be subdivided into a 2003 cloud-free (CF) and a cloudy blind (CB) sector.

By comparing 2003 to 2009 rainfall data, it is clearly evident (Fig. 4) that 2003 can be considered as a "normal" rainfall year, during which the maximum 24h, 48h and 72h rainfall resulted far below the Ida/96E records. In the following, as a consequence, the 2003 and the 2009 landslide inventories were relatively considered as a "normal" and an "extreme" one, respectively.

It is worth to mention that, in the time span of some years, a large part of the 2009 landslide areas resulted as almost completely covered by vegetation and hardly recognizable on the field. At the time of our field survey (May 2015), the study area resulted as generally affected by dormant and active landslides, which were mainly classifiable as debris slides or debris flows. The warm-humid climate is in fact responsible for the fast growing of the vegetation, so that, with the exception of few cases of very recent landslides, the large part of the study area showed only smoothed forms of the previous slope failures (Fig. 5). Each landslide area was mapped as a polygon and represented by means of a landslide identification point (LIP; Costanzo et al. 2014), which was positioned on the highest point along the crown line. In light of the type of slope movement, LIPs were assumed as potentially suitable for detecting the site conditions responsible for the previous failures that, as such, can be used as diagnostic landforms (Rotigliano et al. 2011; Lombardo et al. 2014; Cama et al. 2015) for calibrating the predictive models. It is worth to note that, as a consequence, using a LIP inventory for calibrating the susceptibility models obviously led to estimating the probability for a pixel to be an initiating area, to be then integrated with propagation and/or runout stages modeling.

The two landslide inventories (Fig. 6) included 1503 and 2237 landslides, for 2003 and 2009 respectively. It is worth to note that 253 2009-cases corresponded to reactivation of 2003 landslides.

3. Methods

Landslide susceptibility modeling through stochastic approaches requires the definition of a set of independent variables or covariates, which are expected to play the role of predictors, and of a dependent variable, representing the outcome to be predicted. Differently from deterministic approaches, through the adoption of statistical methods proxy variables can be included as predictors, which potentially play an indirect role into the physics of slope failures; however, in basin scale studies, these proxy variables are the only available at reasonable costs. The predictors are selected among those geo-environmental variables that are supposed to have controlled the slope failure mechanisms responsible for the observed past landslide scenarios (Costanzo et al. 2012b); the latter directly expresses the spatial distribution of the outcome, in terms of stable/unstable status of each mapping unit and constitutes the calibration dataset. At the same time, statistical methods allow for verifying if and to what extent a single predictor does control the estimated susceptibility, as well as potential multi-collinearity between predictors.

Once a set of predictors is defined, a value for all of the variables is assigned to each of the mapping units (Rotigliano et al. 2012) in which the study area is partitioned. The application of statistical methods allows then for optimizing and testing for significance the quantitative relationships which link the probability of the observed outcome status (stable/unstable) and the site multivariate geomorphological conditions of each mapped pixel. The predictive capability of the calibrated susceptibility model is then submitted to quantitative validation tests, which must be based on the evaluation of the accuracy, precision and general reliability of the derived predictive images (i.e., the susceptibility maps) in matching the spatial distribution of one or more unknown validation landslide inventories.

3.1 The MARS method

Recently, the adoption of Multivariate Adaptive Regression Spline (MARS; Friedman 1991) has proved to strengthen the predictive skill of generalized linear modeling techniques. MARS is a non-parametric regression technique that aims at fitting un-linear relationships between predictors and outcome, by fragmenting their range into an optimized number of linear branches. Each branch defines into the covariate axis a basis function (BF) that is structured as hinge function delimited by knots. More complex BFs can be defined as the product of one or more hinge functions associated to different covariates. A particular case is the BF that corresponds to the model intercept, set to a constant value of 1.

The application of the MARS algorithm is based on a two stages procedure. In a first stage (forward pass) a model is generated by stepwise adding (starting from a constant only model) pairs of terms corresponding to the mirrored hinge functions generated by a knot. At each step, the added pair of terms that results in the regression giving the maximum reduction of the residual sum-of-squares error (RSS) is added. In light of the simple structure and fast computing, the searching of the best pair is run systematically (in a "brute force" fashion). This stage can be run up either a minimum RSS gain is obtained or the whole set of possible BFs are added. In the second stage (backward pass) MARS stepwise prunes the best fitting but typically overfitted model, by dropping out of the model at each step the single term whose removal results in the lowest Generalized Cross-Validation parameter (GCV; Craven and Wahba 1979). The criterion expressed by the GCV parameter is in fact the best compromise between fitting (low RSS) and model complexity, the latter depending on the number of terms. At each pruning step, a best model subset is then obtained.

MARS regression function is so given by:

$$f(x) = \alpha + \sum_{i=1}^{N} \beta_i \, h_i(x),$$

where α is the model intercept and the β_i the coefficients of the h_i basis functions obtained by knots-splitting the range of the x covariates.

In this research, MARS modeling was performed using the "earth" package (Milborrow et al. 2011) of R software. In order to reduce the complexity of the models, the maximum degree of interaction was set equal to 1, thus avoiding terms given by combinations of two or more BFs. The software semi-automatically determined the maximum number of terms entering the MARS models. The "evimp" function of "earth" was employed to estimate the variable importance, as a function of the number of entered model subsets. Only subsets equal to or smaller than the final model are considered to evaluate predictor importance (Milborrow 2015).

In light of its flexibility and fast/easy to apply software/hardware structure the MARS algorithm has been recently adopted in stochastic modeling of geomorphological phenomena, including soil erosion and landslides (e.g., Conoscenti et al. 2016, 2017). In this paper a first application to debris flow phenomena prediction through a time-partition based validation scheme is presented.

3.2 Predictors

The following covariates were assumed at the initial stage as potential predictors for slope failures in the study area: outcropping lithology (LIT), land use (USE), landform classification (LCL), elevation (ELE), steepness (STP), aspect (ASP), plan (PLN) and profile (PRF) curvatures, topographic wetness index (TWI) and terrain ruggedness index (TRI). The 10m pixel structure of the source DEM was adopted for partitioning the study area into mapping units (Cama et al. 2016). Table 1 details the DEM-derived covariates in the area and their related references, while the classes for each of the categorical predictors are listed in Table 2.

The selection of the predictors was based on largely adopted geomorphological criteria (Costanzo et al. 2012b) and was here also supported by a multi-collinearity analysis based on classic *VIF* (Variance Inflation Factor) estimation which exploited the "usdm" package (Naimi 2015). A *VIF* value of 10 was set to exclude collinear variables from the models (Heckmann et al. 2014; Jebur et al. 2014; Bui et al. 2016). All the variables resulted as not collinear and were so included into the final models.

3.3 Model building and validation strategy

According to the adopted research design, two validation schemes were applied (Tab. 3): chrono-validation, based on the 2003/2009 time partition, and self-validation, based on the spatial random partition of each of the two inventories (Chung and Fabbri 2003; Guzzetti et al. 2006; Cama et al. 2015, 2017; Lombardo et al. 2015). In particular, forward chrono-validation scheme was applied, by calibrating with 2003 and validating on 2009, whilst the opposite scheme was applied for backward chrono-validation. Moreover, due to the presence of the cloudy area in the 2003 GE coverage, chrono-validations schemes were adopted for predicting either the whole 2009 landslide inventory (2009_{ALL}) or the CF (2009_{CF}) subset. For the same reason, the backward chrono-validation procedure was performed only in the CF sector, by calibrating with 2009 landslides and validating in predicting the 2003s. By applying either time and random partition schemes, starting from the three available calibration dataset (2003CF, 2009CF and 2009ALL), the six models of Table 3 were obtained.

Comparing model A to model D, or model B to model E, allows investigating the role of the calibration inventory in the prediction skill of the derived susceptibility models. In fact, in both the two cases the same landslide scenario (2003 and 2009, respectively) was predicted by calibrating the susceptibility models either on a randomly partitioned of the same coeval landslides or on the whole dataset of the landslides recognized in the other epoch. The different model performances were then more clearly highlighted by directly comparing the model B to the model D. At the same time, in order to have reference levels for evaluating the performance of the temporal (chrono-validated) predictions, the 2003_{CF}, 2009_{ALL} and 2009_{CF} datasets were also submitted to random splitting based self-validation.

To estimate the potential role of the blind area in hampering the research strategy, B to C and E to F models were also compared.

Each dataset was balanced by adding to the positives (i.e. pixels hosting a LIP) an equal number of randomly selected negatives, corresponding to LIP-free pixels (Conoscenti et al. 2016). For temporal partition based validations, one hundred replicates were obtained by randomly multi-extracting a different subset of negatives both in the calibration and validation datasets. Self-validations were based on 10-folds with 10 repetitions cross-validation schemes, obtaining one hundred estimates of model parameters and performance metrics (Tab. 3).

The performances of the models were evaluated by adopting both cut-off dependent and independent metrics. In particular, the prediction skill of the model was evaluated by computing the *AUC* (Area Under Curve) in the ROC (Receiver Operating Characteristics) sensitivity Vs. fall out (1-specificity) plots, as well as from confusion matrixes by distinguishing the true/false positive/negative cases (i.e., TP, TN, FP and FN, respectively), obtained from Youden index optimized cut-off (Youden 1950).

For each of the validation procedures, the one hundred replicates allowed to obtain mean and variance of all the metrics enabling to estimate the model performances in terms of precision and reliability.

4. Results

In order to explore the structure of the models in terms of selected variables, the nsubsets criterion was adopted (Conoscenti et al. 2016), by counting the number of model subsets including each selected variable throughout the pruning pass, which is assumed as expressing the variable importance. Table 4 summarizes the results for the three calibrated models. With a threshold of variable importance of 1 or more, only 27 variables were extracted at least for one model, out of the 44 included at the first step of the modelling procedures, with larger set of variables included in the 2009CF and 2009ALL models.

Based on the comparison between the results of the three models, five main groups of variables can be defined: I, variables selected for all the three models; II, variables selected only for the 2009CF and 2009ALL models; III, variable selected only for the models calibrated in the CF area; IV, variables selected for the 2003CF and 2009ALL models; V, variables selected only for one single model. TRI and ELE are the most important variables, with very similar and high mean values. The Ia subgroup is completed by quite important and homogeneous variables. The high importance of North-eastern facing observed for the 2003CF model resulted as very lowered for both the 2009. The Ic subgroup includes variables which are very important for the two 2009 models, whilst a lowering of one order of magnitude is observed for 2003CF. The II group includes a large set of variables which are important for the two 2009 models (ASP_South and ASP_SouthEast, in particular), but not extracted throughout the pruning pass in the 2003CF calibration. SLO is selected as a quite important variable only for models calibrated in the CF sector, whilst group IV variables were extracted with varying importance, only for 2003CF and 2009ALL models. Finally, group V variables were extracted only for one of the calibrated models.

As regards the predictive performances, Figure 7 shows the averaged ROC curves that were obtained for the six models through their replicates, while, to ease the comparison of the global accuracy, a boxplot displaying each of the corresponding mean *AUC*s was prepared (Fig. 8).

The whisker symbols along the ROC curves (Fig. 7) attest for highly stable results through the replicates, with higher dispersion gradually shifting from true to false positives, in the direction of the lower scores. For the calibrated subsets, the frequency distribution of the scores shows a different shape in the intermediate range (0.7-0.3), with a more picked bi-modal trend for the 2003 model (Figs. 7a-c), resulting in a flat zone, where a wide range of scores is equally represented in terms of mapped pixels.

As regards the *AUC*s (Fig. 8), the 2003CF, 2009CF and 2009ALL self-validated models obtained similar excellent performances, with *AUC* values above the 0.8 threshold (Hosmer and Lemeshow 2000). At the same time, in the CF sector, the forward and the backward chrono-validations produced almost the same results in terms of *AUC*s, with largely acceptable values of 0.76 and 0.78, respectively. For the forward chrono-validations, only a slight performance decreasing was observed from the CF sector to the whole catchment (*AUC*=0.74); the same small difference was observed for the 2009 self-calibrated model, from 2009CF (*AUC*=0.83) to 2009ALL (*AUC*=0.81).

If cutoff-dependent performance metrics are taken into consideration (Tab. 5), it is evident that the loss in prediction skill from 2003 and 2009 self-validation (model A and model E) to forward and backward chrono-validation (model B and model D), respectively, depends on a sensitivity decreasing, which is more marked for the 2003 model, with no coupled loss of specificity. Furthermore, by directly comparing the backward (model E) to the forward (model B) chrono-validated models in the CF sector, in spite of the similar *AUC* performance (0.71 and 0.70, respectively), a marked higher sensitivity and lower specificity of the former arises. In both cases, the specificity does not change from self- to chrono-validation. It is worth to note that the two opposite behaviors of specificity and sensitivity compensated each other, so that the two models result in a similar accuracy.

Figure 9a,b shows the two susceptibility maps prepared by calibrating the models in the CF sector exploiting the 2003 and 2009 landslide inventory, respectively. The maps were obtained by averaging, for each pixel, the one hundred estimates of probability values. A map of the residuals is also shown (Fig. 9c), where the difference in the estimated score of the two models ($score_{2003}$-$score_{2009}$) was depicted. In spite of the similar general spatial pattern of the two prediction images, the 2009 model produced higher scores on average, whilst positive and negative residuals stretch along the north-westward and south-eastward slopes of the main SW-NE running pyroclastic ranges, respectively.

However, in terms of positive and negative predictions, if applying Youden index cut-offs, few pixels resulted as differently classified in the two maps (Fig. 10): less than 5% of the pixels with scores diverging for more than one susceptibility class; a larger percentage (13%) of pixels classified with a one class shift and crossing the cut-off score value.

5. Discussion

The analysis of the variable importance of the three calibrated models highlights that more variables are involved in the definition of the susceptibility for the extreme event datasets. At the same time, some variables play a role in the predictive models, no matter the intensity of the trigger, with two topographic factors showing the maximum importance: elevation (ELE) and Topographic Ruggedness Index (TRI). On the other hand, some variables (Topographic Wetness Index, Pasture and Crop Cultivation soil use) resulted as much more important (one order of magnitude) under extreme scenario, with the case of South and South-East facing, which are among the most important variables for the two 2009 models, but never extracted for 2003CF. Conversely, North-eastern facing has an importance index of more than 10 only for normal event condition. The SLO variable was selected only for the models calibrated in the CF sector, probably due to the geomorphologic conditions.

Figure 11 puts the main results of the validation tests inside the framework of the investigation strategy adopted in this research. The results attested that the 2003 landslide inventory allowed to calibrate a predictive model, whose *AUC* performance was estimated as very high and reliable, after a self-validation procedure was applied (model A); that was the only test we could have performed in 2003, before the 2009 the 2009 Ida/96E event. However, if trying to predict the sites where then debris flow and debris slide phenomena triggered (model B), a small *AUC* decreasing (from above to below the 0.8 threshold), but coupled with a relevant number of false negative occurrences (low sensitivity), arose: relying on a map prepared on 2003 would have resulted in 32% of missing positives (against the 22% expected on the basis of the 2003 self-validation test). An analogous *AUC* decreasing resulted for the backward chrono-validation (model D) with respect to the 2009CF self-validated model (model E), but caused by a moderate false negative prediction (miss rate) increasing, with only 21% of missing positives (against the 17% obtained from self-validation). It is worth to highlight that the model E showed the same accuracy of the self-validated model B in predicting the 2003 positives, suggesting the model calibrated with an extreme event landslide scenario of a different epoch (2009) as being able to reach the same performance in recognizing the sites of activation for a normal season landslide scenario of a self-validated one. Conversely, the model calibrated with this lower trigger landslide scenario resulted in a markedly lower sensitivity than the one calibrated under the extreme event (sensitivity = 0.68, against 0.83). In particular, the 2009-calibrated model resulted capable to detect as nearly as the 80% of the 2003 landslides, but expecting a higher number of positives, actually corresponding to 2003 stable sites (type-I errors), with low specificity and high number of false positives. The same model calibrated in 2003 recognized the negative locations in the 2009 landslide scenario with a higher performance than the 2009 self-validated itself (specificity = 0.72, against 0.63).

The results of this research seem to confirm non-linear stochastic relationships between predictors and outcome under different driving conditions, as the crossing with a more severe landslide scenario does not only correspond to a false-to-true conversion of the predicted positives (actually, a small decreasing of PPV is recorded for the 2003 forward chrono-validation), but also to positive occurrences for a number of predicted negatives. However, a similar but slighter effect is observed when models are calibrated with the extreme landslide scenario, which means the larger scenario does not fully include the smaller one.

In terms of geomorphological model, a more intense triggering of the slopes is responsible for the activation of large part of those site conditions which typically activate under normal triggering but together with other regions of the multivariate parameter hyperspace, having stable status under normal triggering, as attested by the 2009 models, which are controlled by more variables. This means that, if we focus on the applicative relevance of the prediction, exploiting landslide scenarios caused by more intense triggering events allows to fit large part of the normal-trigger caused landslides as well as the same extreme-trigger ones. At the same time, a source of errors in terms of successful positive predictions is introduced by extreme events, so that a moderate lowering of the sensitivity is to be expected. This could be due to the activation on 2009 of a secondary triggering mechanisms, caused by landslide coupling, which add a non-stochastic component to the spatial relationships between predictors and outcome, being rather controlled by morphodynamic slope connectivity. In fact, in a relevant number of cases, landslides in that extreme event scenario were triggered by the impact or the erosion (either laterally or at the foot of the slopes) of the moving mass detached from the primary slope failures. In Figure 12 a field example is given, highlighting a number of coupled landslides, in the 2009 landslide scenario. The same setting can be observed in Figure 5.

As regards the susceptibility maps, under an applicative perspective the 2009-calibrated models confirmed to be much more accurate in predicting positives, avoiding false negative predictions. Among the pixels predicted as negatives at 2003, but as positive at 2009, 227 out of 580 (39%) resulted unstable in the 2009 landslide scenario (Fig. 10); conversely, very few (0.5%) of the negative predicted pixels at 2009, but as positive at 2003, actually resulted unstable in 2003. Again, if considering the potential severity of a false negative prediction, the 2009 model confirmed to produce the more realistic and prudential prediction image in terms of potential damages.

Differences in temporal validations between models trained under normal or extreme event triggered landslide scenarios have been investigated in other papers (Lombardo et al. 2014; Cama et al. 2015 and references therein). However, in this research, deepening the analysis to cut-off dependent performance metrics highlighted that, together with the confirmation of a *AUC* decreasing from self- to chrono-validation, which could suggest using either one or the other model be the same, a clear difference arises in terms of type of predictive errors.

6. **Conclusions**

Predicting storm triggered landslides always poses the problem of the morphodynamic coherence between calibration and validation datasets. In fact, the prediction skill of a model can be hampered by a large difference between the trigger intensity of the event responsible for the calibration and one for the validation landslide dataset.

In the present research, a test was carried out in the Caldera Ilopango, which is a representative area of Central America, where recurrent extreme events occur striking landslide prone pyroclastic slopes.

Two different landslide inventories were exploited: one produced by normal rainfall, the other being the result of a very intense triggering storm (the Ida/96E 2009 event). The results confirmed the relevant role played by the triggering conditions both in the importance of the variable included in the susceptibility models, and in their predictive performance. As regards the predictors, it is worth to note that some variables were selected for both the two triggering scenarios, whilst some other only for the extreme event one, demonstrating that the slope failures occur under different mechanisms depending on the rainfall intensity. At the same time, in terms of predictive performances, the specificity of the predictive models resulted as not conditioned by the type of validation (chrono- or self-validation), nevertheless being higher for the model calibrated under normal event. Conversely, the sensitivity changes from self- to chrono-validation, with the models calibrated with a landslide inventory associated to normal trigger less capable to predict the sites of landslide activation under intense triggering and resulting in very critical type-II errors (high miss rate). On the contrary, models calibrated with extreme landslide scenarios resulted very efficient in self-predicting the positives as well as less critically limited in predicting the normal event triggered landslides.

It is worth to note that focusing only in an *AUC* estimation for assessing the quality of a susceptibility model could be misleading in terms of the applicative exploitation of the susceptibility maps, whose quality is critically dependent on the correctness of binary positive/negative discriminations.

This research demonstrated that validating on an extreme-event landslide inventory a susceptibility map calibrated with a normal landslide dataset does not result into a simple conversion from false to true positives (i.e., the turning of negatives but susceptible cases into positive), but that new susceptible conditions arises under intense triggering, which cannot be predicted if a normal event inventory is used for calibration. Conversely, extreme landslide inventories allow to calibrate susceptibility maps which are very effective in predicting the landslides produced by normal events but with limits in discriminating stable conditions.

Summarizing what above discussed, models calibrated with normal landslide scenario result in higher specificity (less Type-I error) but lower sensitivity (more Type-II error). To explain these differences, two main hypotheses are here suggested: the non-linear behavior in the trigger intensity dimension of regressed relationships which link predictors and outcome; the role of a non-stochastic (morhodynamic), related to the multiple coupled triggering between different landslides under extreme events. This point is obviously of great importance in terms of applicative consequences. In fact, it means that landslide susceptibility stochastic modeling requires multi-temporal calibration inventories, so to detect and estimate the effects of differences in the intensity of the trigger, optimizing positive and negative predictions. Strategies for integrating low and high trigger landslide inventories are to be issued and constitute the logical conclusive perspective of this research.

Acknowledgments

The present research was supported by a project funded by the Ministry of the Foreign Affairs of the Italian Government and carried out by the University of Palermo (resp. Prof. G. Giunta). Miguel Angél Hernandéz worked in this research as a PhD student of the Department of Earth and Marine Sciences of the University of Palermo (tutor E. Rotigliano, co-tutor C. Conoscenti). All the authors equally contributed to the research. The manuscript was linguistically reviewed by Maria Simona Romana.

References

Agostini S, Corti G, Doglioni C, Carminati E, Innocenti F, Tonarini S, Manetti P, Di Vincenzo G, Montanari D (2006) Tectonic and magmatic evolution of the active volcanic front in El Salvador: insight into the Berlín and Ahuachapán geothermal areas. Geothermics 35, 368–408. https://doi.org/10.1016/j.geothermics.2006.05.003

Avila LA, Cangialosi J (2010) Tropical Cyclone Report Hurricane Ida. National Hurricane Center. https://www.nhc.noaa.gov/data/tcr/AL112009_Ida.pdf

Beven KJ, Kirkby MJ (1979) A physically based variable contributing area model of basin hydrology. Hydrol Sci. Bull. 24, 43–69. https://doi.org/10.1080/02626667909491834

Bui DT, Tuan TA, Klempe H, Pradhan B, Revhaug I (2016) Spatial prediction models for shallow landslide hazards: a comparative assessment of the efficacy of support vector machines, artificial neural networks, kernel logistic regression, and logistic model tree. Landslides 13, 361-378. https://doi.org/10.1007/s10346-015-0557-6

Burrough PA, McDonell RA (1998) Principles of geographical information systems. Oxford University Press, New York

Cama M, Lombardo L, Conoscenti C, Agnesi V, Rotigliano E (2015) Predicting storm-triggered debris flow events: Application to the 2009 Ionian Peloritan disaster (Sicily, Italy). Natural Hazards and Earth System Sciences 15 (8), 1785-1806. https://doi.org/10.5194/nhess-15-1785-2015

Cama M, Conoscenti C, Lombardo L, Rotigliano E (2016) Exploring relationships between grid cell size and accuracy for debris-flow susceptibility models: a test in the Giampilieri catchment (Sicily, Italy). Environmental Earth Sciences 75 (238), 1-21. https://doi.org/10.1007/s12665-015-5047-6

Cama M, Lombardo L, Conoscenti C, Rotigliano E (2017) Improving transferability strategies for debris flow susceptibility assessment. Application to the Saponara and Itala catchments (Messina, Italy). Geomorphology 288, 52-65. https://doi.org/10.1016/j.geomorph.2017.03.025

CEPAL (Comisión Económica para América Latina y el Caribe, 2010) El Salvador: Impacto Socioeconómico, Ambiental y de Riesgo por la Baja Presión Asociada a la Tormenta Tropical Ida en Noviembre de 2009. Ciudad de México, pp. 21. http://repositorio.cepal.org/handle/11362/1382

CEPAL (Comisión Económica para América Latina y el Caribe, 2011) El Salvador: Evaluación de daños y pérdidas en El Salvador ocasionados por la depresión tropical 12E. http://www.transparencia.gob.sv/institutions/mag/documents/119851/download

CATHALAC (Centro del Agua del Trópico Húmedo para América Latina y el Caribe, 2009) Mapa de rutas de sistemas meteorológicos IDA y Baja 96E noviembre 2009. http://www.servir.net/images/desastres/2009-11-02_TT_Ida/ida_baja96_20091111.jpg

Chung CJ, Fabbri AG (2003) Validation of spatial prediction models for landslide hazard mapping. Natural Hazards 30 (3), 451-472. https://doi.org/10.1023/B:NHAZ.0000007172.62651.2b

Conoscenti C, Ciaccio M, Caraballo-Arias NA, Gómez-Gutiérrez Á, Rotigliano E, Agnesi V (2015) Assessment of susceptibility to earth-flow landslide using logistic regression and multivariate adaptive regression splines: a case of the Belice River basin (western Sicily, Italy). Geomorphology 242, 49–64. https://doi.org/10.1016/j.geomorph.2014.09.020

Conoscenti C, Rotigliano E, Cama M, Caraballo-Arias NA, Lombardo L, Agnesi V (2016) Exploring the effect of absence selection on landslide susceptibility models: A case study in Sicily, Italy. Geomorphology 261, 222-235. https://doi.org/10.1016/j.geomorph.2016.03.006

Conoscenti C Agnesi V, Cama M, Caraballo-Arias NA, Rotigliano E (2017) Assessment of Gully Erosion Susceptibility Using Multivariate Adaptive Regression Splines and Accounting for Terrain Connectivity. Land Degradation & Development. doi: 10.1002/ldr.2772

Conrad O, Bechtel B, Bock M, Dietrich H, Fischer E, Gerlitz L, Wehberg J, Wichmann V, Böhner J (2015) System for Automated Geoscientific Analyses (SAGA) v. 2.1.4, Geosci. Model Dev. 8, 1991-2007. https://doi.org/10.5194/gmd-8-1991-2015

Costanzo D, Cappadonia C, Conoscenti C, Rotigliano E (2012a) Exporting a Google Earth™ aided earth-flow susceptibility model: A test in central Sicily. Natural Hazards 61 (1), 103-114. https://doi.org/10.1007/s11069-011-9870-0

Costanzo D, Rotigliano E, Irigaray C, Jiménez-Perálvarez JD, Chacón J (2012b) Factors selection in landslide susceptibility modelling on large scale following the gis matrix method: Application to the river Beiro basin (Spain). Natural Hazards and Earth System Science 2(2), 327-340. https://doi.org/10.5194/nhess-12-327-2012

Costanzo D, Chacón J, Conoscenti C, Irigaray C, Rotigliano E (2014) Forward logistic regression for earth-flow landslide susceptibility assessment in the Platani river basin (southern Sicily, Italy). Landslides 11 (4), 639-653. https://doi.org/10.1007/s10346-013-0415-3

Craven P, Wahba G (1979) Smoothing noisy data with spline functions. Numer. Math. 31, 377–403. https://doi.org/10.1007/BF01404567

Frattini P, Crosta G, Carrara A (2010) Techniques for evaluating the performance of landslide susceptibility models. Engineering Geology 111 (1–4), 62-72. https://doi.org/10.1016/j.enggeo.2009.12.004

Friedman JH (1991) Multivariate adaptive regression splines. Ann. Stat. 19, 1–141. http://www.jstor.org/stable/2241837

García-Rodríguez MJ, Malpica JA, Benito B, Díaz M (2008) Susceptibility assessment of earthquake-triggered landslides in El Salvador using logistic regression. Geomorphology 95(3), 172-191. https://doi.org/10.1016/j.geomorph.2007.06.001

García-Rodríguez MJ, Malpica JA (2010) Assessment of earthquake-triggered landslide susceptibility in El Salvador based on an Artificial Neural Network model. Natural Hazards and Earth System Science 10(6), 1307-1315. https://doi.org/10.5194/nhess-10-1307-2010

Guzzetti F, Reichenbach P, Ardizzone F, Cardinali M, Galli M (2006) Estimating the quality of landslide susceptibility models. Geomorphology 81, 166–184. https://doi.org/10.1016/j.geomorph.2006.04.007

Heckmann T, Gegg K, Gegg A, Becht M (2014) Sample size matters: investigating the effect of sample size on a logistic regression susceptibility model for debris flows. Nat. Hazards Earth Syst. Sci. 14, 259–278. http://dx.doi.org/10.5194/nhess-14-259-2014

Hosmer DW, Lemeshow S (2000) Applied logistic regression, Wiley Series in Probability and Statistics. Wiley. http://dx.doi.org/10.1198/tech.2002.s650

Jebur MN, Pradhan B, Tehrany MS (2014) Optimization of landslide conditioning factors using very high-resolution airborne laser scanning (LiDAR) data at catchment scale. Remote Sens. Environ. 152, 150–165. http://dx.doi.org/10.1016/j.rse.2014.05.013

Kirschbaum D, Stanley T, Yatheendradas S (2016) Modeling landslide susceptibility over large regions with fuzzy overlay. Landslides 13, 485-496. doi: 10.1117/12.737835

Kopačková V, Šebesta J (2007) An approach for GIS-based statistical landslide susceptibility zonation: with a case study in the northern part of El Salvador", Proc. SPIE 6749, Remote Sensing for Environmental Monitoring, GIS Applications, and Geology VII, 67492R (29 October 2007). doi: 10.1117/12.737835

Lexa J, Sebesta J, Chávez J A, Hernández W, Pecskay Z (2011) Geology and volcanic evolution in the southern part of the San Salvador Metropolitan Area. Journal of Geosciences 56(1), 106-140. doi: http://dx.doi.org/10.3190/jgeosci.088

Lombardo L, Cama M, Maerker M, Rotigliano E (2014) A test of transferability for landslides susceptibility models under extreme climatic events: Application to the Messina 2009 disaster. Natural Hazards, 74 (3), 1951-1989. doi: https://doi.org/10.1007/s11069-014-1285-2

Lombardo L, Cama M, Conoscenti C, Märker M, Rotigliano E (2015) Binary logistic regression versus stochastic gradient boosted decision trees in assessing landslide susceptibility for multiple-occurring landslide events: application to the 2009 storm event in Messina (Sicily, southern Italy). Natural Hazards 79 (3), 1621-1648. https://doi.org/10.1007/s11069-015-1915-3

Lombardo L, Bachofer F, Cama M, Märker M, Rotigliano E (2016) Exploiting Maximum Entropy method and ASTER data for assessing debris flow and debris slide susceptibility for the Giampilieri catchment (north-eastern Sicily, Italy). Earth Surface Processes and Landforms, 41, 1776-1789. doi: 10.1002/esp.3998

MARN (Ministerio de Medio Ambiente y Recursos Naturales de Salvador, 2004) Memoria técnica para el mapa de susceptibilidad de deslizamientos de tierra en El Salvador.

MARN (Ministerio de Medio Ambiente y Recursos Naturales de Salvador, 2009) Análisis de las intensidades máximas de lluvia. Evento atmosférico de los días 7 y 8 de noviembre de 2009. Informe, pp. 23.

MARN (Ministerio de Medio Ambiente y Recursos Naturales de Salvador, 2010) Síntesis de los informes de evaluación técnica de las lluvias del 7 y 8 de noviembre 2009 en El Salvador: Análisis del impacto físico natural y vulnerabilidad socio ambiental. http://www.oas.org/summit/sisca/Download.aspx?type=C&lang=es&id=728

MARN (Ministerio de Medio Ambiente y Recursos Naturales de Salvador, 2011) Depresión Tropical 12E/Sistema Depresionario sobre El Salvador y otros eventos extremos del pacífico. https://reliefweb.int/report/el-salvador/depresi%C3%B3n-tropical-12e-sistema-depresionario-sobre-el-salvador-y-otros-eventos

Milborrow S (2015) Notes on the earth package [WWW Document]. URL http://www.milbo.org/doc/earth-notes.pdf

Milborrow S, Hastie T, Tibshirani R (2011) Earth: Multivariate Adaptive Regression Spline Models. R Software Package.

Mora S, Vahrson WG (1991) Determinación "A priori" de la amenaza de deslizamientos en grandes áreas utilizando in-dicadores morfodinámicos. In CODAZZI A (1992) Memoria del primer simposio sobre sensores remotos y sistemas de información geográfica para el estudio de riesgos natu-rales. Bogotá, Colombia, organizado por el Instituto Geográfico Ago-stín Codazzi, UNESCO, ITC Holanda, 259 - 273.

Mora S, Vahrson WG (1994) Macrozonation methodology by landslide hazard determination. Bull Soc. Eng. Geol. 31, 49-58.

Naimi B (2015) Uncertainty Analysis for Species Distribution Models. R Software Package. doi: 10.1111/j.1365-2699.2011.02523.x

R Core Team (2015) R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/

Riley S J, DeGloria S D, Elliot R (1999) A terrain ruggedness index that quantifies topographic heterogeneity. Intermountain Journal of Sciences, vol. 5, No. 1-4: 23-27.

Rose WI, Bommer JJ, López DL, Carr MJ, and Major JJ (eds, 2004) Natural hazards in El Salvador: Boulder, Colorado, Geological Society of America Special Paper 375, p.

Rotigliano E, Agnesi V, Cappadonia C, Conoscenti C (2011) The role of the diagnostic areas in the assessment of landslide susceptibility models: a test in the Sicilian chain. Natural Hazards 58(3), 981-999. https://doi.org/10.1007/s11069-010-9708-1

Rotigliano E, Cappadonia C, Conoscenti C, Costanzo D, Agnesi V (2012) Slope units-based flow susceptibility model: Using validation tests to select controlling factors. Natural Hazards, 61 (1), 143-153. https://doi.org/10.1007/s11069-011-9846-0

Stoiber RE, Carr MJ (1973) Quaternary volcanic and tectonic segmentation of Central America: Bulletin of Volcanology 37, 204-325. doi: https://doi.org/10.1007/BF02597631

Wilson JP, Gallant GC (2000) Digital terrain analysis. In: Wilson JP, Gallant JC (eds) Terrain analysis: principles and applications. Wiley, New York, pp 1–27.

Youden WJ (1950) Index for rating diagnostic tests. Cancer 3 (1), 32–35. doi: 10.1002/1097-0142(1950)3:1<32::AID-CNCR2820030106>3.0.CO;2-3

Zevenbergen LW, Thorne CR (1987) Quantitative analysis of land surface topography. Earth Surf. Process. Landf. 12:47–56. http://dx.doi.org/10.1002/esp.3290120107

Captions

Fig. 1 – Setting of the study area.

Fig. 2 – a) Location of the Caldera Ilopango. b) Drainage network and outcropping lithologies in the study area.

Fig. 3 – a) Tracks of the Ida and TD96E (mod. from Avila and Cangialosi 2010; CATHALAC 2009). b) 7-8 November 2009 cumulated rainfall in El Salvador (MARN 2009).

Fig. 4 – a) Average, 2009 and 2003 monthly rain at the meteorological station Ilopango. b) Comparison between the Ida/96E rainfall records and 2003 maximum ten critical cases for 24h, 48h and 72h durations.

Fig. 5 –Comparison between 2003 (a), 2009 (soon after the Ida/96E event; b) and 2015 (c) slope conditions on a representative sector of the study area (LIP: landslide identification point).

Fig. 6 – 2003 (a) and 2009 (b) landslide inventory maps.

Fig. 7 – ROC-plots for the six models and validation schemes (see Tab. 3).

Fig. 8 – AUC boxplots for the six models and validation schemes (see Tab. 3).

Fig. 9 – 2003 (a) and (b) 2009 landslide susceptibility maps. Map (c) and (d) frequency distribution of the residuals.

Fig. 10 – Differences in positive (P)/negative (N) predictions between the two models.

Fig. 11 – Graphical summary scheme of the adopted validation strategies and main performance metrics.

Fig. 12 – Field example of coupled multiple landslides in 2009.

Tab. 1 – List of the DEM-derived predictors.

Tab. 2 – List of the categorical predictors.

Tab. 3 – Characteristics of the validation schemes adopted for the six susceptibility models.

Tab. 4 – Summary of the variable importance index for the three calibrated models (NS = not selected).

Tab. 5 – Summary of the validation metrics for the six susceptibility models.

Figure3 Click here to access/download;Figure;fig3.tif

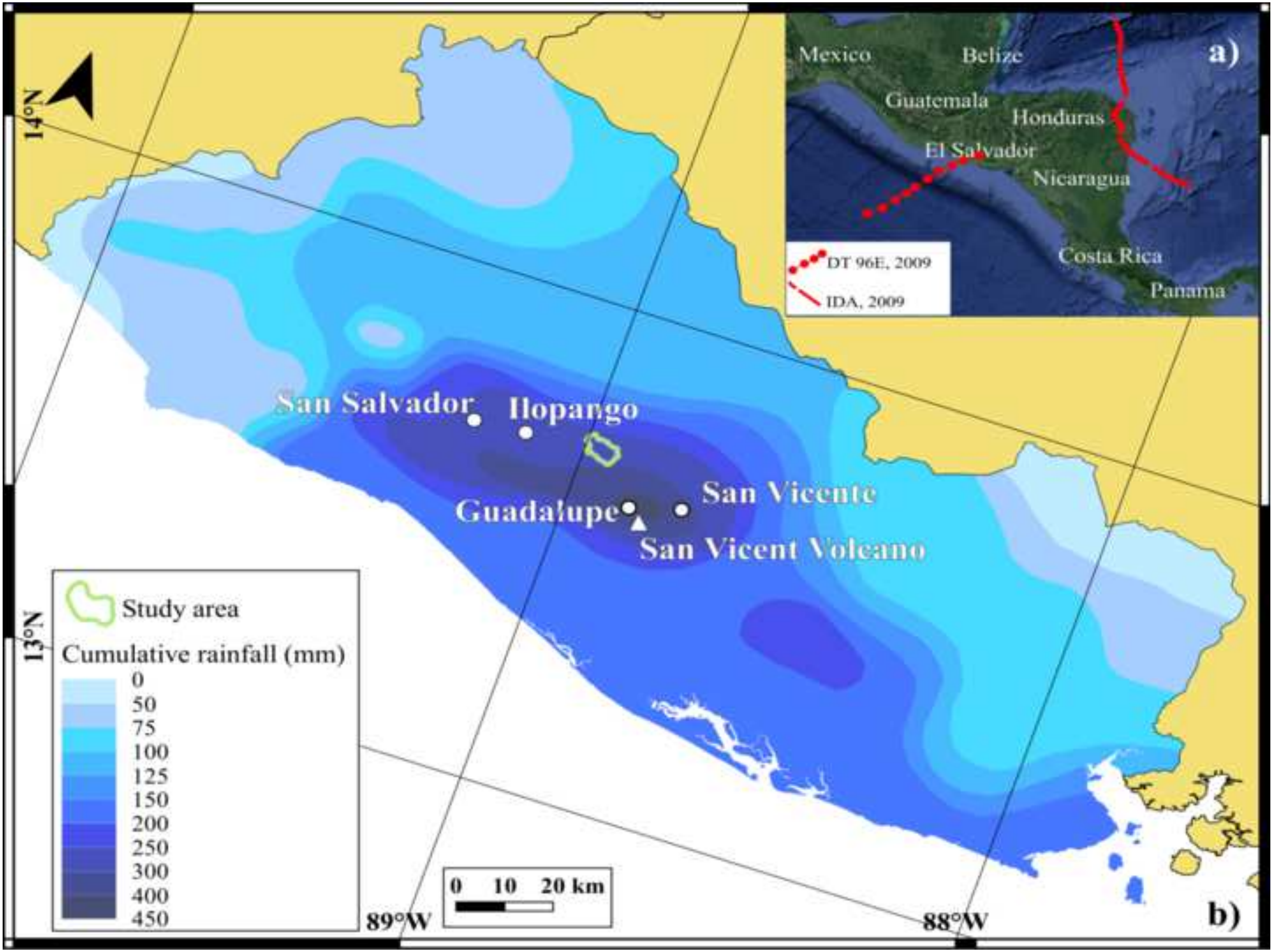


Figure2

Click here to access/download;Figure;fig2.tif

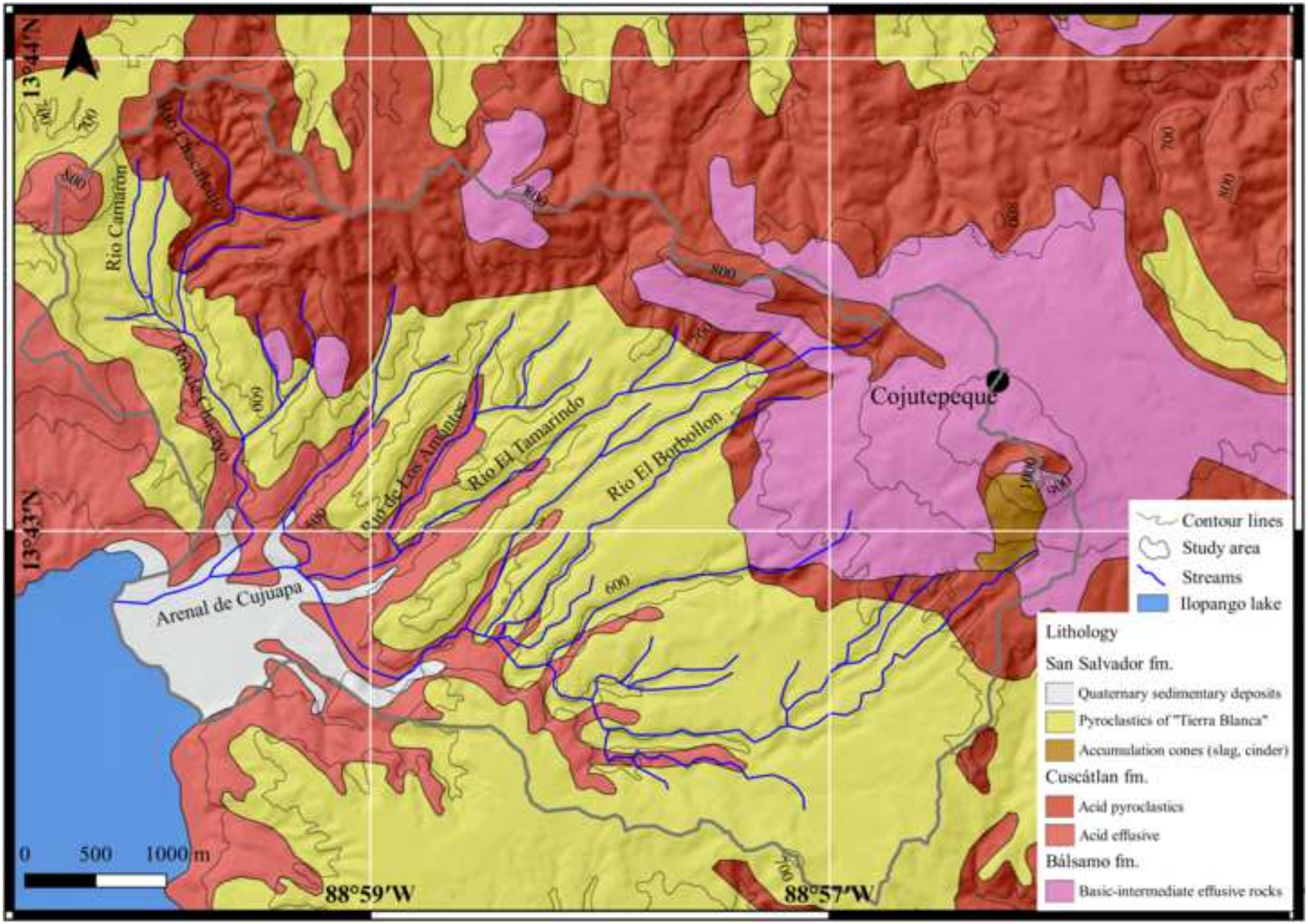


Figure5

Click here to access/download;Figure;fig5.tif

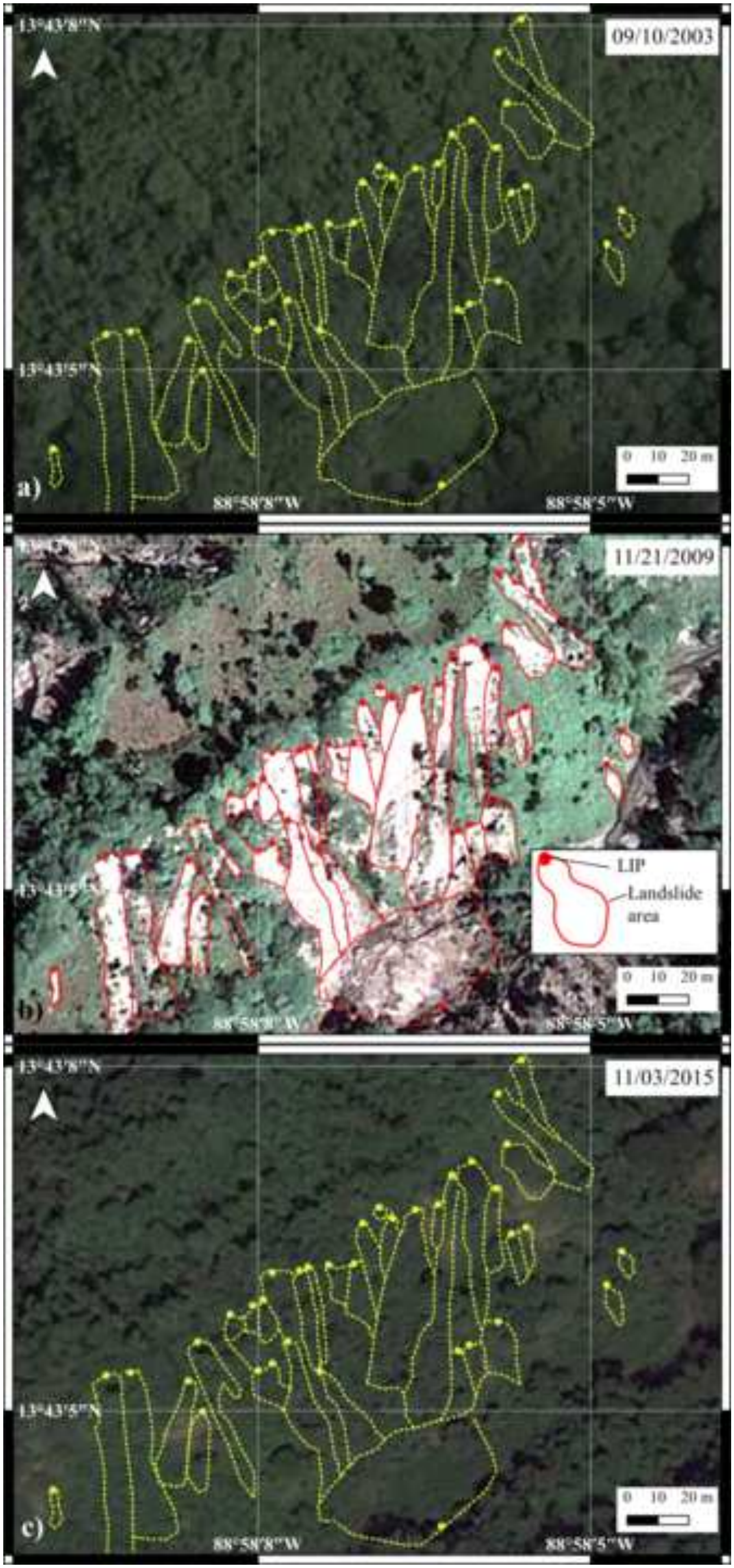


Figure4

Click here to access/download;Figure;fig4.tif

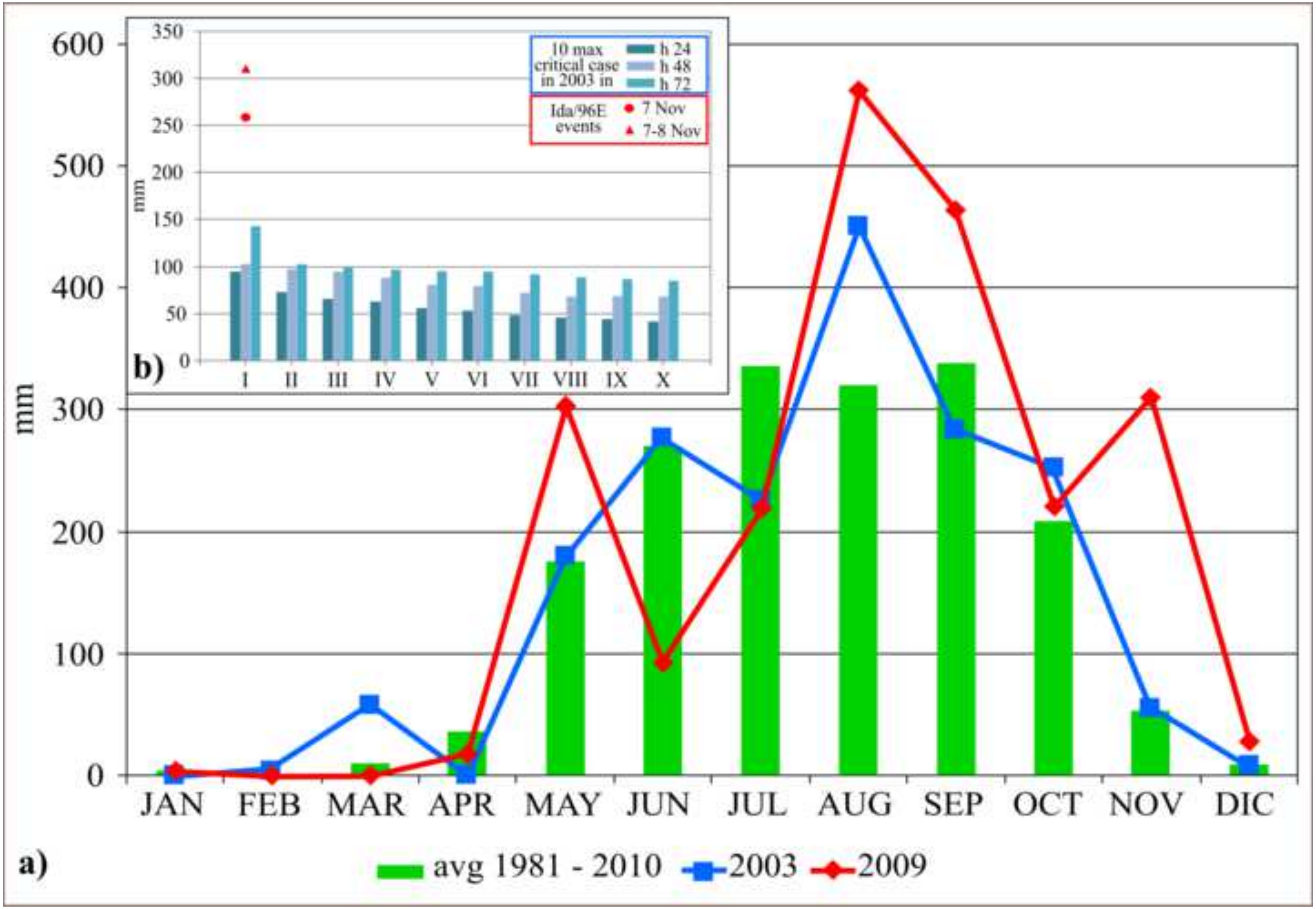


Figure8

Click here to access/download;Figure;fig8.tif

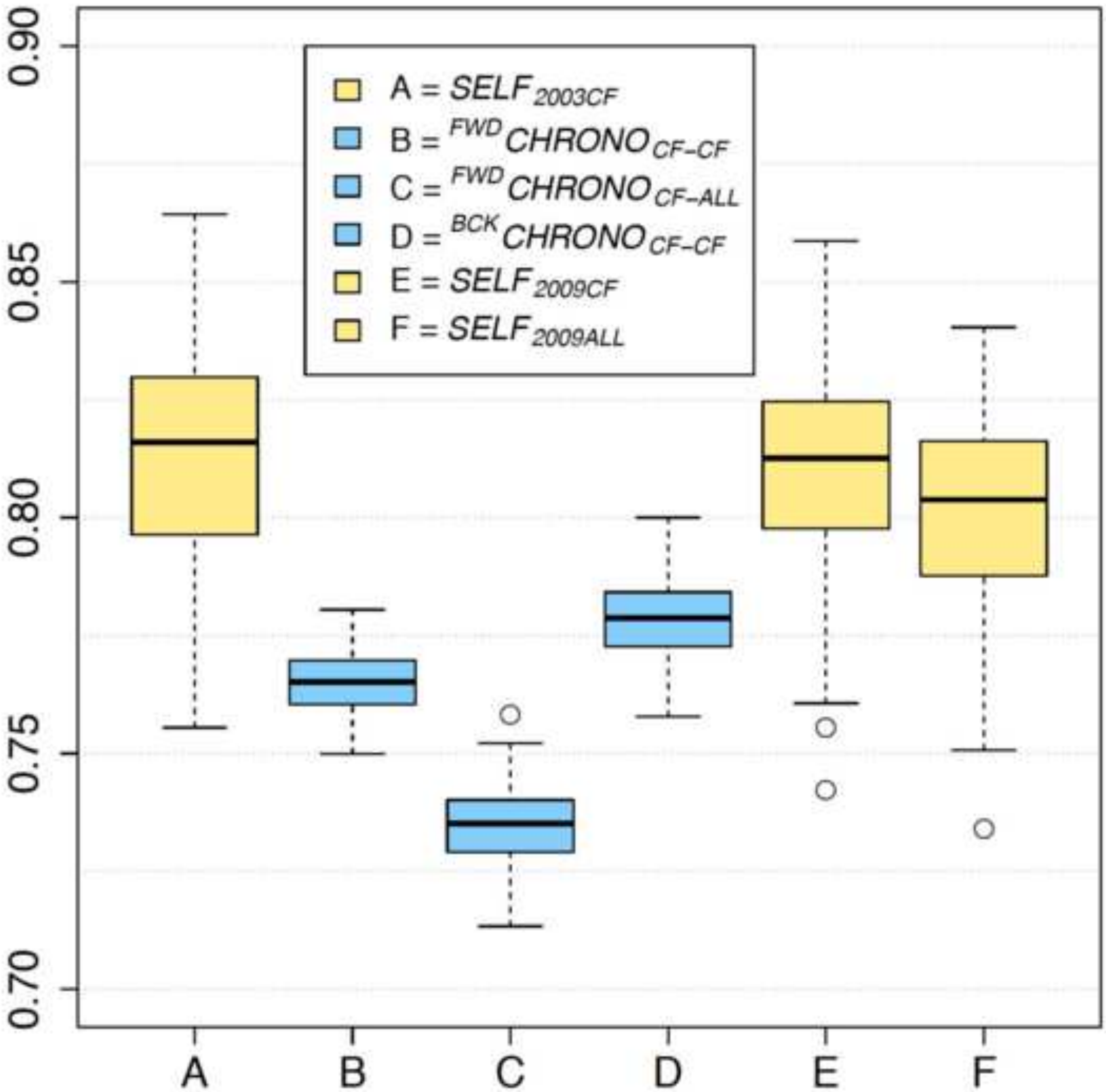


Figure10

Click here to access/download;Figure;fig10.tif

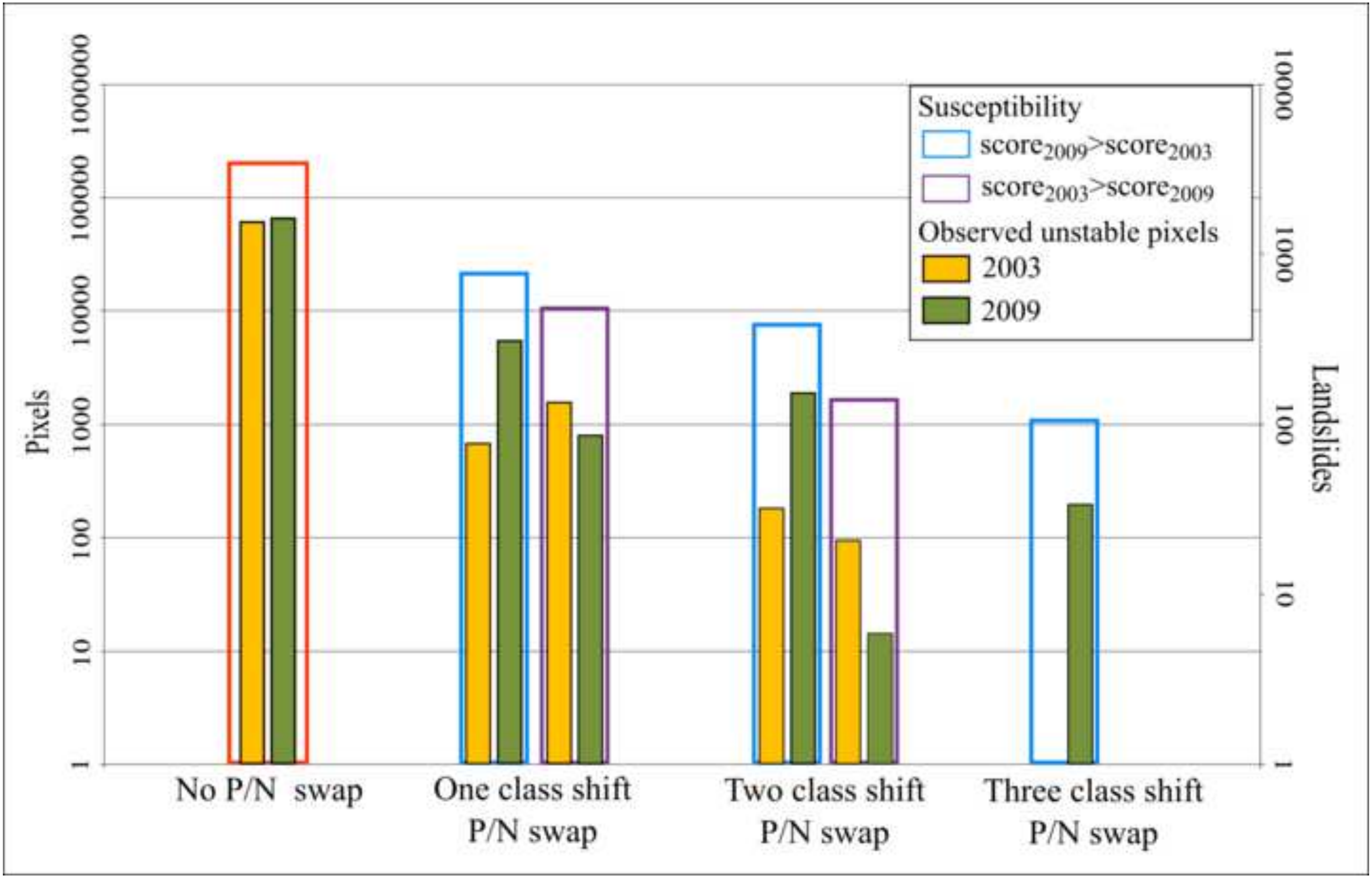


Figure11

Click here to access/download;Figure;fig11.tif

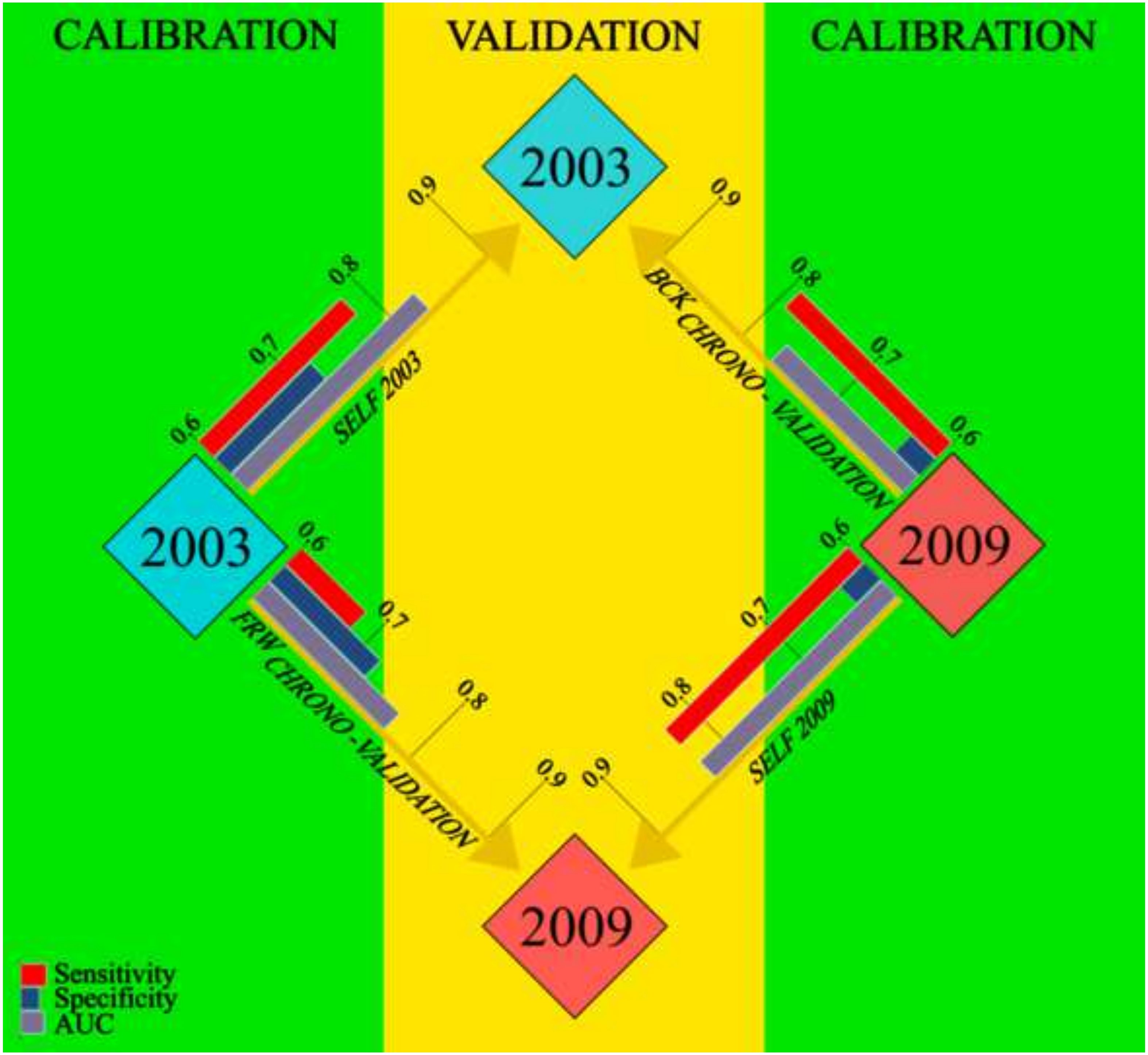


Figure9

Click here to access/download;Figure;fig9.tif

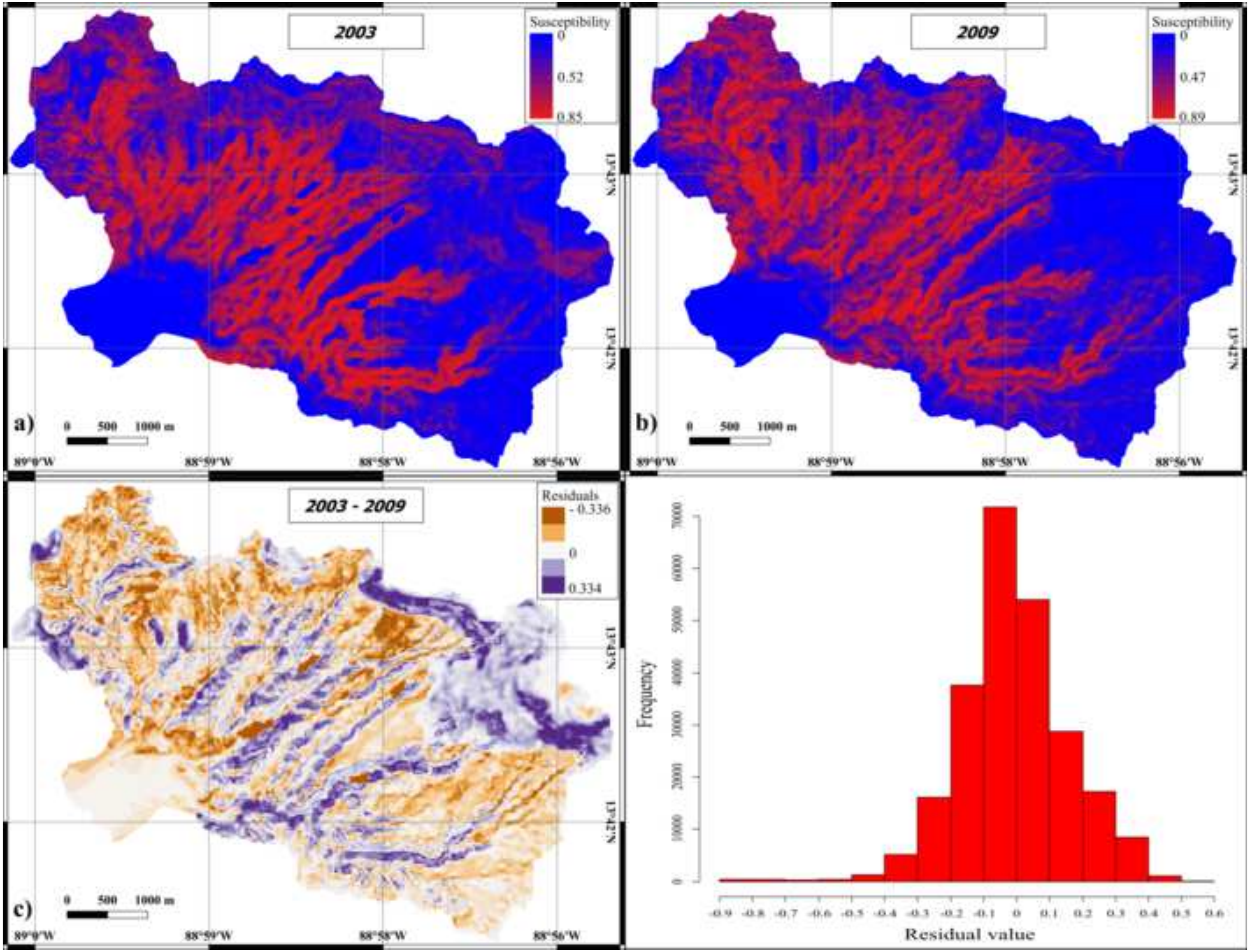


Figure12

Figure1


90°W
85°W
15°N
10°N
0
100
200 km
MEXICO
BELIZE
NORTH AMERICA PLATE
CARIBBEAN PLATE
GUATEMALA
HONDURAS
EL SALVADOR
NICARAGUA
MIDDLE AMERICAN TRENCH
COCOS PLATE
COSTA RICA
PANAMA PLATE

Figure6

Click here to access/download;Figure;fig6.tif

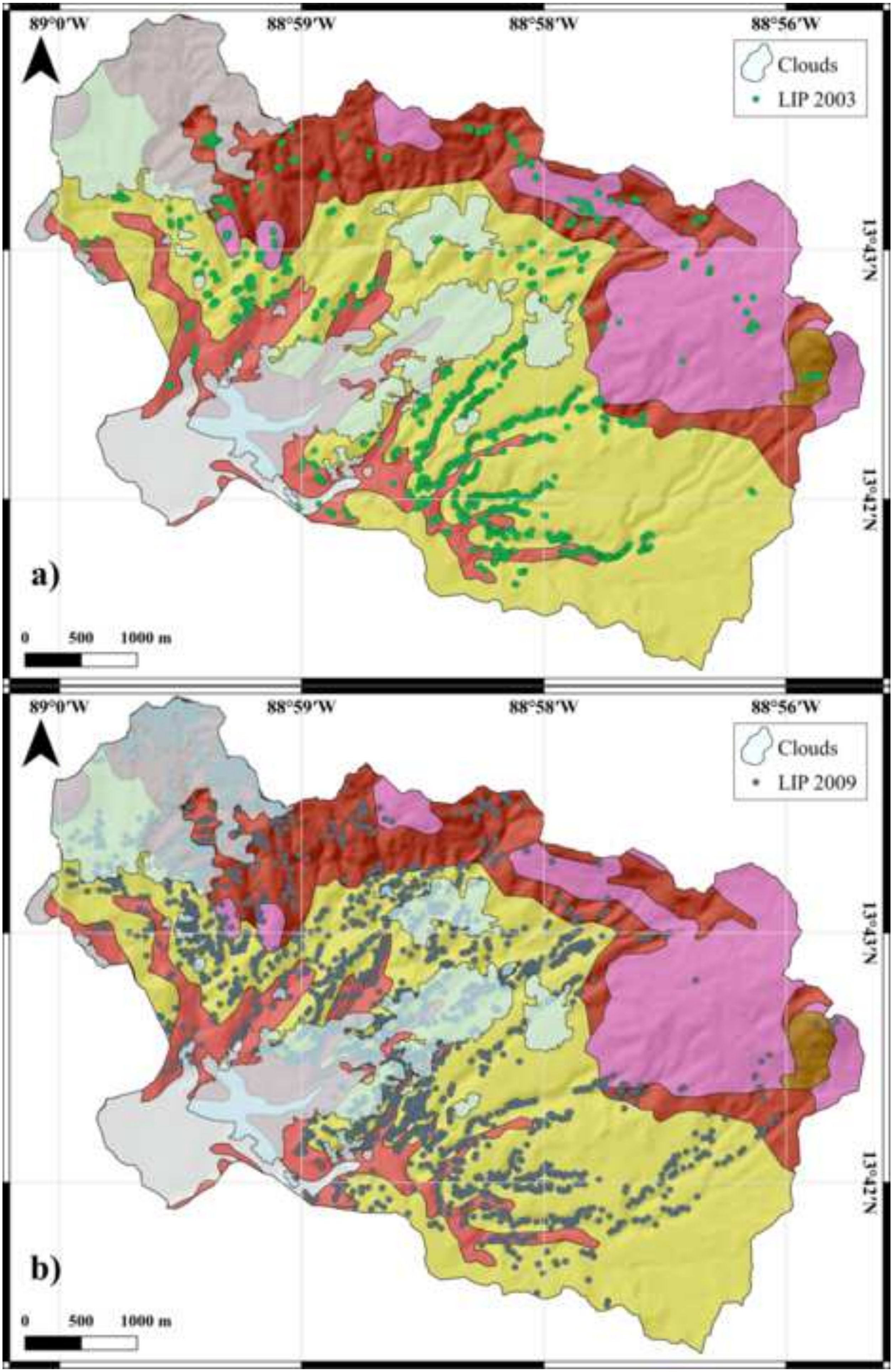
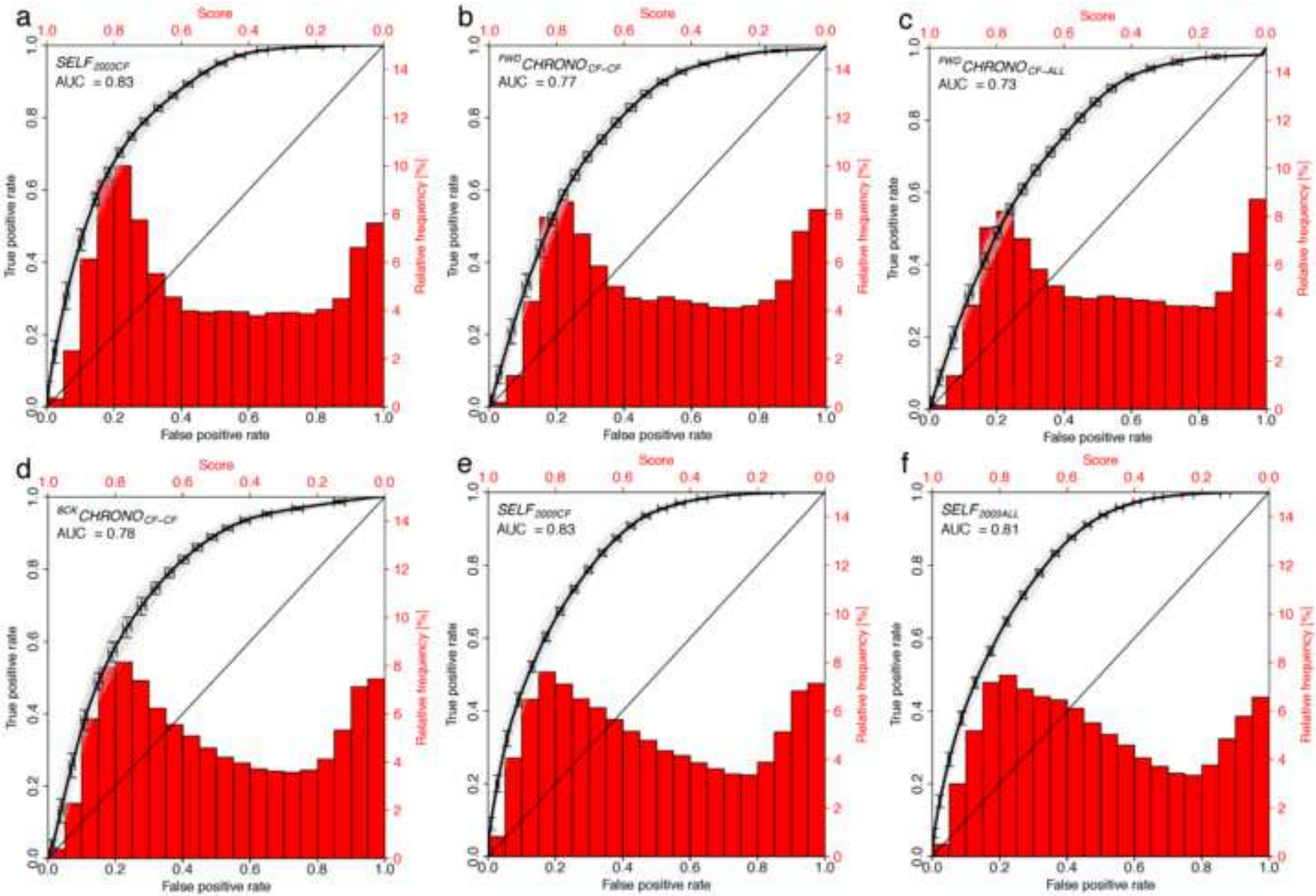


Figure7


a
$SELF_{2003CF}$
AUC = 0.83
b
$^{PWD}CHRONO_{CF-CF}$
AUC = 0.77
c
$^{PWD}CHRONO_{CF-ALL}$
AUC = 0.73
d
$^{BCK}CHRONO_{CF-CF}$
AUC = 0.78
e
$SELF_{2003CF}$
AUC = 0.83
f
$SELF_{2003ALL}$
AUC = 0.81
Score
True positive rate
False positive rate
Relative frequency [%]

Tables

Factor	*Source layer*	*Description of source parameter*	*Units*	*References*
LCL	Landform classification	Outcome of an automated procedure that recognise landforms on a gridded elevation distribution (TPI)		Wilson and Galland 2000
STP	Slope gradient	Highest first derivative of elevation	degree	Burrough and McDonell 1998
ASP	Slope aspect	Direction of steepest downwards slope from each cell to its neighbours	degree	Wilson and Galland 1996
PLN	Plan curvature	Second derivative of elevation, computed along the horizontal plane	rad/m	Zevenbergen and Thorne 1987
PRF	Profile curvature	Second derivative of elevation, computed along the direction of the highest slope gradient	rad/m	Zevenbergen and Thorne 1987
TWI	Topographic wetness index	Calculated as ln[A/tanβ], where A and β, computed on each cell, corrispond to the area of upslope drained cells and the slope gradient, respectively	m	Beven and Kirkby 1979
TRI	Terrain Ruggedness Index	Topographic height difference of a cell compared to the adjacent ones	m	Riley et al. 1999

Tab. 1 – List of the DEM-derived predictors.

Factor	*Source layer*	*Classes of the variable*
LCL	Landform classification	LCL_1 (Streams)
		LCL_2 (Midslope Drainages)
		LCL_3 (Upland Drainages)
		LCL_4 (Valleys)
		LCL_5 (Plains)
		LCL_6 (Open Slopes)
		LCL_7 (Upper Slopes)
		LCL_8 (Local Ridges)
		LCL_9 (Midslope Ridges)
		LCL_10 (High Ridges)
ASP	Slope aspect (degree)	ASP_flat (-1)
		ASP_N (North: 0-22.5|337.5-360)
		ASP_NE (NorthEast: 22.5-67.5)
		ASP_E (East: 67.5-112.5)
		ASP_SE (SouthEast: 112.5-157.5)
		ASP_S (South: 157.5-202.5)
		ASP_SW (SouthWest: 202.5-247.5)
		ASP_W (West: 247.5-292.5)
		ASP_NW (NorthWest: 292.5_337.5)
LIT	Lithology	LIT_Qf (Quaternary sedimentary deposits)
		LIT_s4 (Pyroclastics of "Tierra Blanca")
		LIT_s5b (Accumulation cones)
		LIT_c1 (Acid pyroclastics)
		LIT_c2 (Acid effusive)
		LIT_b3 (Basic-intermediate effusive rocks)
USE	Land use	USE_1 (Wood)
		USE_2 (Crop cultivation)
		USE_3 (Vegetables cultivation)
		USE_4 (Crop cultivation and pasture)
		USE_5 (Pasture cultivation)

		USE_6 (Pasture)
		USE_7 (River)
		USE_8 (Continuous urban fabric)
		USE_9 (Discontinuous urban fabric)
		USE_10 (Precarious urban fabric)
		USE_11 (Growing urban fabric)
		USE_12 (Low shrubs)
		USE_13 (Mine areas)
		USE_14 (Uncultivated areas)

Tab. 2 – List of the categorical predictors.

MOD.	VALIDATION SCHEME	CALIBRATION	VALIDATION	DATASET	REPLICATES
A	$SELF_{2003CF}$	$2003_{CF\_RND(90\%)}$	$2003_{CF\_RND(10\%)}$	10-folds cross-validation	100
B	$^{FRW}CHRONO_{CF\text{-}CF}$	$2003_{CF\_(100\%)}$	$2009_{CF\_(100\%)}$	100 (CAL X VAL)	100
C	$^{FRW}CHRONO_{CF\text{-}ALL}$	$2003_{CF\_(100\%)}$	2009_{ALL}	100 (CAL X VAL)	100
D	$^{BCK}CHRONO_{CF\text{-}CF}$	$2009_{CF\_(100\%)}$	$2003_{CF\_(100\%)}$	100 (CAL X VAL)	100
E	$SELF_{2009CF}$	$2009_{CF\_RND(90\%)}$	$2009_{CF\_RND(10\%)}$	10-folds cross-validation	100
F	$SELF_{2009ALL}$	$2009_{ALL\_RND(90\%)}$	$2009_{ALL\_RND(10\%)}$	10-folds cross-validation	100

Tab. 3 – Characteristics of the validation schemes adopted for the six susceptibility models

Variables	**MOD A (2003CF)**	**MOD E (2009CF)**	**MOD F (2009ALL)**	TYPE	
LCL_2	6	5	9	I	a
TRI	17	19	21		a
ELE	16	18	20		a
ASP_W	2	3	6		a
ASP_E	7	3	6		a
ASP_NE	12	4	2		b
USE_2	2	12	13		c
TWI	3	10	17		c
USE_4	4	17	19		c
LTL_s5b	NS	2	2	II	
PLC	NS	3	2		
PRC	NS	3	2		
ASP_NW	NS	1	2		
LIT_b3	NS	7	5		
USE_9	NS	2	4		
LCL_6	NS	1	4		
ASP_SE	NS	15	17		
ASP_S	NS	13	17		
ASP_SW	NS	6	12		

SLO	5	3	NS	III
USE_6	9	NS	8	IV
LIT_s4	1	NS	6	
LIT_c1	NS	NS	7	V
LIT_c2	NS	NS	4	
LCL_4	7	NS	NS	
USE_14	3	NS	NS	
LCL_5	NS	1	NS	

Tab. 4 – Summary of the variable importance index for the three calibrated models (NS = not selected).

MOD.	*VALIDATION SCHEME*	*ROC AUC*	*CUT-OFF*	*RECALL*		*FALL OUT*	*MISS RATE*	*PRECISION*		*ACCURACY*
		MEAN	*MEAN*	*SENS.*	*SPEC.*			*PPV*	*NPV*	
A	$SELF_{2003CF}$	0.83	0.52	0.78	0.72	0.28	0.22	0.74	0.76	0.75
B	$^{FRW}CHRONO_{CF-CF}$	0.76		0.68	0.72	0.28	0.32	0.71	0.69	0.70
C	$^{FRW}CHRONO_{CF-ALL}$	0.74		0.64	0.70	0.30	0.36	0.68	0.66	0.67
D	$^{BCK}CHRONO_{CF-CF}$	0.78	0.47	0.79	0.63	0.37	0.21	0.68	0.76	0.71
E	$SELF_{2009CF}$	0.83		0.83	0.63	0.37	0.17	0.69	0.79	0.73
F	$SELF_{2009ALL}$	0.81	0.45	0.84	0.66	0.34	0.16	0.71	0.80	0.75

Tab. 5 – Summary of the validation metrics for the six susceptibility models