*Proc. of the 9th International Symposium on Geotechnical Safety and Risk (ISGSR)*
*Edited by* Zhongqiang Liu, Jian Dai and Kate Robinson
*Published by* Research Publishing, Singapore. 
doi: 10.3850/GRF-25280825_isgsr-051-P259-cd

ISGSR2025

# BAYESIAN MODELING OF RAINFALL-INDUCED LANDSLIDES

Carlo Zaccardi*, Luigi Ippoliti, Pasquale Valentini
*Department of Economics, University "G. d'Annunzio" of Chieti-Pescara, Italy*
**Corresponding author. E-mail: carlo.zaccardi@unich.it*

Giovanna Vessia
*Department of Engineering and Geology, University "G. d'Annunzio" of Chieti-Pescara, Italy*

Marco A. Rodríguez
*Department of Environmental Sciences, Université du Québec à Trois-Rivières, Canada*

Alexandra M. Schmidt
*Department of Epidemiology, Biostatistics and Occupational Health, McGill University, Canada*

Landslides are a serious geologic hazard, posing substantial and increasing risks, and resulting in widespread disruptions. Correctly assessing the relationship between terrain conditions and landslides is fundamental to understand these instability-triggering factors, and to implement effective monitoring campaigns. In this work, we study rainfall-induced landslide events occurred from 2002 and 2020 in the Peri-Adriatic area of the Abruzzo region (Italy). We propose a zero-inflated Poisson (ZIP) model to estimate spatial landslide susceptibility while accounting for the presence of excess zeros in the data. Thanks to the probabilistic framework, it is possible not only to estimate the spatial susceptibility but also to locate the areas that are likely to have under-reported events. Results show that the most susceptible areas are in the vicinity of river basins and in high coastal areas. The probability of not having landslides generally increases as the distance from the Adriatic Sea increases.

*Keywords*: Bayesian, Landslides, Under-reporting, Zero-inflated Poisson, Zero-inflation.

## 1. Introduction

Hydrogeological instability poses a serious threat to communities all over the world. Due to its climatic, geological, and morphological characteristics, the Italian territory is naturally exposed to landslides and floods, (Trigila et al., 2021), and the consequences of such events can sometimes be disastrous, not only for the environment but also for the population and the economy (Schuster & Highland, 2001; Garcia-Delgado et al., 2022). Therefore, correctly evaluating landslide susceptibility (i.e., the likelihood of landslides occurring in a given area) is of utmost importance for monitoring and planning purposes.

There are mainly three approaches that can be adopted to estimate the landslide susceptibility, namely empirical approaches, physically based models, and probabilistic approaches (Mondini et al., 2023). Empirical approaches rely entirely on the expertise of the investigator, who define rainfall thresholds above which a landslide is likely to occur. The results are not quantitative and are difficult to reproduce. Physically based models simulate the hydrological process of rainfall infiltration through the soil but are applicable to areas of limited extent.

The probabilistic approach is based on a statistical model, which can be applied to study populations of landslides over large spatial domains while accounting for the uncertainty associated with the estimates. Recently, several authors proposed statistical models to create susceptibility maps, see, e.g., Loche et al. (2022) and Reichenbach et al. (2018), and Fang et al. (2024) and Lombardo et al. (2020) for spatio-temporal data. Vessia et al. (2020) proposed a geostatistical approach to draw maps of rainfall thresholds for landslides initiation.

In this paper, we study spatial susceptibility in the Abruzzo region (Italy), which is among those at the highest landslide risk, according to the landslide hazard zonation from Trigila et al. (2021). Given that the number of zeros (i.e., units with no landslide records) is larger than expected under a Poisson distribution, the data is said to be *zero-inflated*, and the statistical model should account for this fact. Therefore, we propose a zero-inflated Poisson (ZIP) model (Lambert, 1992). Zero-inflated models are widely used in epidemiological literature (see, e.g., Agarwal et al., 2002; Fernandes et al., 2009; Osei et al., 2022). Existing zero-inflated models for landslide applications (Ferrer et al., 2024; Frigerio Porta et al., 2021; Ramos-Scharrón et al., 2022) consider mapping units such as grid cells or administrative boundaries, which do not account for the morphology of the study region. The proposed approach, instead, uses the slope unit as the mapping unit. Finally, the ZIP model is also useful to identify those areas that are likely to have under-reported events.

## 2. Study Area and Data Description

The study area is situated in the Peri-Adriatic basin of the Abruzzo region and comprises the provinces of Teramo, Pescara and Chieti (see Figure 1a). The area's extension is about 5782 km$^2$, and the elevation ranges from 0 m to 2793 m. The landscape is dominated by hill areas, with marly-clayey reliefs and low slope values, and foothill areas, made up of sandstones and mudstones with higher relief energy and slope values. A more detailed description of the study area can be found in Vessia et al. (2020).

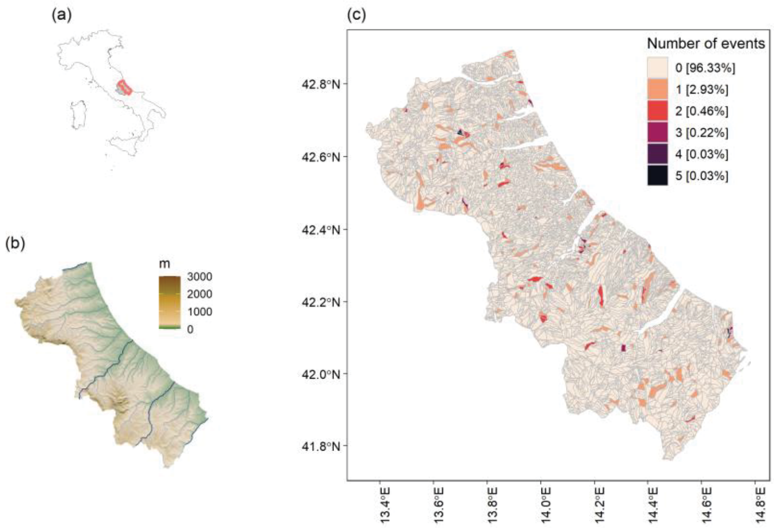


Figure 1. (a) Map of Italy with Abruzzo region in grey; the study area is defined by the red border. (b) Topographic elevation map of the study area. (c) Map showing SU delineation and number of landslide events, with frequencies of each count in square brackets.

We collected landslide information from three distinct databases, namely FraneItalia (Calvello & Pecoraro, 2018), ITALICA (ITAlian rainfall-induced LandslIdes CAtalogue, Peruccacci et al., 2023), and the database in Vessia et al. (2020). Only shallow landslides with known geographic location are considered, and the Areal Landslide Events (ALE) in the FraneItalia database are excluded due to their great geographic uncertainty. Therefore, this work is based on 329 slope failures that occurred between 2002 and 2020.

As an appropriate terrain mapping unit, we use the slope unit (SU), as defined in Alvioli et al. (2016). This is a subdivision of the terrain that maximizes homogeneity within the unit and heterogeneity between units (Lombardo et al., 2020). We use the SU delineation as calculated in Alvioli et al. (2020), where flat areas (such as those along the riverbeds – see Figure 1b) are already removed. Hence the study area is divided into $n$=6899 SUs. These are depicted in Figure 1c, where the number of landslides per SU is also reported. It can be noticed that no slope failures have been recorded for the great majority (>96%) of SUs.

To model landslide susceptibility, we use a set of explanatory variables (covariates) from Loche et al. (2022). They describe the morphometry of the study area (elevation, slope, aspect, curvature, and topographic wetness index), its lithological and pedological signal (bulk density, depth to bedrock, weight percentage of clay and silt particles, distance to stream), and the surface area of the SUs. Mean and standard deviation of these quantities are used as covariates in the statistical model. They are available at https://geomorphology.irpi.cnr.it/tools/slope-units, along with the SU delineation.

## 3. Modeling Approach

Let $Y_i$ denote the landslide count (i.e., the outcome variable) registered in the $i$-th SU, then a Poisson regression model is specified as follows, for $i = 1, \dots, n$ (see, e.g., Gelman et al., 2014):

$$Y_i | \mu_i \sim Poi(\mu_i) \tag{1}$$

$$\log(\mu_i) = O_i + \boldsymbol{x}_i' \boldsymbol{\beta} \tag{2}$$

where $\mu_i$ is the intensity, $O_i$ is an offset term corresponding to the natural logarithm of the surface area of the $i$-th unit, $\boldsymbol{x}_i = (x_{1i}, \dots, x_{Ki})'$ represents a $K$-dimensional vector of covariates, with associated effects $\boldsymbol{\beta}$.

The Poisson distribution assumption should be avoided for susceptibility mapping purposes, mainly because of the presence of excess zeros in the data and the limited ability of the Poisson distribution to account for over-dispersion. Over-dispersion indicates the situation where the variance is greater than the mean of the distribution. In the data under study, the mean number of slope failures is 0.05, which is smaller than its variance (0.08). Moreover, an overdispersion test indicates that a Poisson model would not fit well to the data ($\chi^2$=8890, *p-value*<$10^{-3}$). Alternatives such as the Negative Binomial distribution or the Poisson-Log-Normal mixture model can be considered to account for this issue. However, complete knowledge about the population of landslides in each area is still required, i.e. the landslide inventory must collect *all* events that occurred in the past. Although we observe SUs with no events, it does not mean that slope failures did not occur. It only means that it was not recorded in the inventory (*under-reporting*).

Therefore, we assume that $Y_i$ is a ZIP random variable, namely $Y_i \sim ZIP(\mu_i,\ p_i)$, whose distribution is given by a mixture between a Poisson distribution with parameter $\mu_i$ and a Bernoulli distribution with parameter $p_i$, for $i = 1, \dots, n$ (Agarwal et al., 2002):

$$Pr^{ZIP}(Y_i|\mu_i, p_i) = \begin{cases} p_i + (1-p_i)\, Pr^{POI}(0|\mu_i) & Y_i = 0 \\ (1-p_i)\, Pr^{POI}(Y_i|\mu_i) & Y_i > 0 \end{cases}, \quad (3)$$

where $Pr^{POI}(\cdot\,|\mu_i)$ denotes the Poisson probability function with mean $\mu_i$.

The preceding specification allows the ZIP model to account for over-dispersion and to distinguish between two kinds of zeros, namely *structural zeros* and *sampling zeros* (Osei et al., 2022). Structural zeros occur (with probability $p_i$) in the absence of landslides, while sampling zeros (occurring with probability $(1-p_i)\, Pr^{POI}(0|\mu_i)$) can be interpreted as those events that have not been reported. Note that $Pr^{ZIP}(Y_i|\mu_i, 0)$ is simply the Poisson distribution, $Pr^{POI}(Y_i|\mu_i)$.

In other words, ZIP models have two components to model landslide occurrence. The first component models landslide intensity, so $\log(\mu_i)$ is a linear function of covariates $x_i$ as shown in Eq. (2). The second component is dichotomous and models the probability of absence of landslides in unit $i$, $p_i$ (thus $1-p_i$ is the probability of presence of landslides). More specifically:

$$\text{logit}(p_i) = \log\left(\frac{p_i}{1-p_i}\right) = O_i + \boldsymbol{z}_i'\boldsymbol{\gamma}\,, \quad (4)$$

where the covariates $\boldsymbol{z}_i = (z_{1i}, \dots, z_{Li})'$ could be either (a subset of) the $\boldsymbol{x}_i$'s or a different set of variables. In this work, $\boldsymbol{x}_i$ contains all the covariates outlined in Section 2, while $\boldsymbol{z}_i$ contains information on landslide-triggering factors, namely elevation (which is a proxy for total annual precipitation) and slope.

We adopt the Bayesian paradigm to make inference on all the parameters of the proposed model. In summary, if $\boldsymbol{Y} = (Y_1, \dots, Y_n)'$, Bayesian inference aims to summarize the *posterior* distribution, $\Pr(\boldsymbol{\theta}|\boldsymbol{Y})$, of the vector of model parameters, $\boldsymbol{\theta}$, given the data, $\boldsymbol{Y}$. Following Bayes' theorem, the posterior distribution is proportional to the *likelihood* function, $\Pr(\boldsymbol{Y}|\boldsymbol{\theta})$, times the *prior* distribution, $\Pr(\boldsymbol{\theta})$. Markov chain Monte Carlo (MCMC) constitutes a versatile tool for sampling from $\Pr(\boldsymbol{\theta}|\boldsymbol{Y})$. The posterior mean of the parameter vector, $\text{E}(\boldsymbol{\theta}|\boldsymbol{Y})$, is usually considered as a Bayesian point estimate for $\boldsymbol{\theta}$. For a thorough discussion on Bayesian inference, see, for instance, Gelman et al. (2014).

In the proposed model, $\boldsymbol{\theta} = (\boldsymbol{\beta}', \boldsymbol{\gamma}')'$, and prior distributions on $\boldsymbol{\beta}$, $\boldsymbol{\gamma}$ are chosen to be independent but informative, following the procedure described in Agarwal et al. (2002). Results are based on 20,000 samples from the posterior distribution that were saved after a burn-in period of 10,000 samples. Convergence of model's chains is monitored visually through trace plot inspection and using usual statistics (Gelman et al., 2014).

## 4. Results

For the $i$-th SU, the susceptibility is estimated as the probability of having at least one landslide, that is $S_i = Pr^{ZIP}(Y_i > 0|\mu_i, p_i) = (1-p_i)(1-e^{-\mu_i})$. This quantity can be computed at each iteration of the MCMC sampler, and Figure 2a shows its posterior mean (the uncertainty can also be quantified and will be shown in an extended version of this paper). The map highlights that high coastline areas are generally the most susceptible. Additionally, SUs very close to river basins also show high susceptibility.

As seen in Section 3, the ZIP model classifies two categories of zeros. Figure 2b maps the posterior mean of $p_i$, that is the probability of having a structural zero in the $i$-th SU. We say that $p_i$ is a *marginal* probability because the observed landslide count is not considered. This quantity seems higher in the inland territories than coastal areas.

## 5. Discussion and Conclusions

In this paper, a ZIP model is proposed to estimate spatial susceptibility maps. This model should be preferred over a standard Poisson regression approach if the data is over-dispersed and zero-inflated. The proposed method is

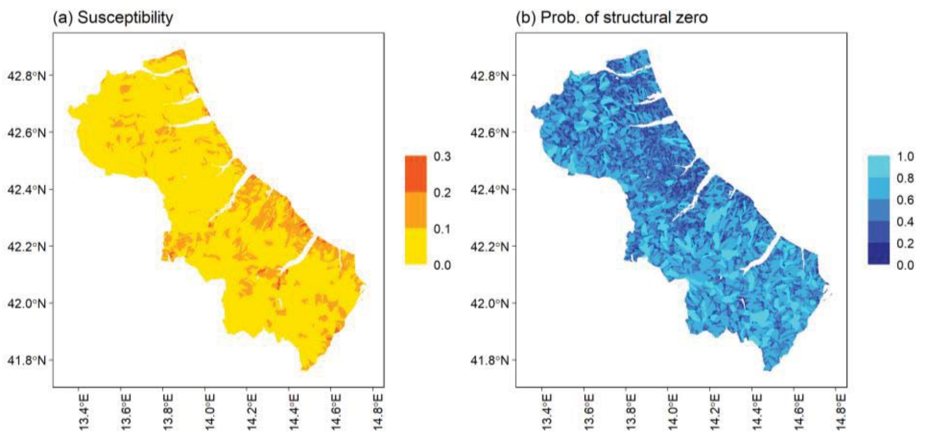


Figure 2. Maps showing posterior means of (a) the susceptibility, $S_i$, and (b) the probability that a zero is structural, $p_i$.

applied to a database collecting the population of landslides for $n$=6899 SUs of the Abruzzo region. The spatial susceptibility map suggests that elevated SUs near the coast and river basins are highly susceptible.

The ZIP model can be seen as a generalization of the Poisson regression model, as it is possible to estimate not only the susceptibility of each SU, but also the marginal probability of absence of landslides (i.e., of having a structural zero). Figure 2b, however, does not consider the observed landslide count in the SUs. Hence, it would be more informative to estimate the conditional probability of presence of landslides *given* no landslides were observed for some SUs. The higher it is, the more likely it is that the SU is under-reporting (Fernandes et al., 2009; Osei et al., 2022). These conditional probabilities might be used to locate those areas where monitoring efforts should be intensified. This aspect will be addressed in an extended version of this work, which will also introduce spatial random effects to account for latent environmental factors affecting landslide presence and susceptibility (Fernandes et al., 2009).

**Acknowledgement**

This research is funded by the European Union - NextGenerationEU, research project PRIN 2022 PNRR SLIDE - Stochastic Modeling of Compound Events (P2022KZJTZ- CUP D53D23018920001) and research project Bando a Cascata HPC (CN0 0000013, CUP H93C22000450007).

**References**

Agarwal, D. K., Gelfand, A. E., & Citron-Pousty, S. (2002). Zero-inflated models with application to spatial count data. *Environmental and Ecological Statistics*, *9*(4), 341–355. https://doi.org/10.1023/A:1020910605990

Alvioli, M., Guzzetti, F., & Marchesini, I. (2020). Parameter-free delineation of slope units and terrain subdivision of Italy. *Geomorphology*, *358*, 107124. https://doi.org/10.1016/j.geomorph.2020.107124

Alvioli, M., Marchesini, I., Reichenbach, P., Rossi, M., Ardizzone, F., Fiorucci, F., & Guzzetti, F. (2016). Automatic delineation of geomorphological slope units with r.slopeunits v1.0 and their optimization for landslide susceptibility modeling. *Geoscientific Model Development*, *9*(11), 3975–3991. https://doi.org/10.5194/gmd-9-3975-2016

Calvello, M., & Pecoraro, G. (2018). FraneItalia: A catalog of recent Italian landslides. *Geoenvironmental Disasters*, *5*(1), 13.

Fang, Z., Wang, Y., Van Westen, C., & Lombardo, L. (2024). Landslide hazard spatiotemporal prediction based on data-driven models: Estimating where, when and how large landslide may be. *International Journal of Applied Earth Observation and Geoinformation*, *126*, 103631. https://doi.org/10.1016/j.jag.2023.103631

Fernandes, M. V., Schmidt, A. M., & Migon, H. S. (2009). Modelling zero-inflated spatio-temporal processes. *Statistical Modelling*, *9*(1), 3–25. https://doi.org/10.1177/1471082X0800900102

Ferrer, J. V., Samprogna Mohor, G., Dewitte, O., Pánek, T., Reyes-Carmona, C., Handwerger, A. L., Hürlimann, M., Köhler, L., Teshebaeva, K., Thieken, A. H., Tsou, C., Urgilez Vinueza, A., Demurtas, V., Zhang, Y., Zhao, C., Marwan, N., Kurths, J., & Korup, O. (2024). Human Settlement Pressure Drives Slow-Moving Landslide Exposure. *Earth's Future*, *12*(9), e2024EF004830. https://doi.org/10.1029/2024EF004830

Frigerio Porta, G., Bebbington, M., Xiao, X., & Jones, G. (2021). A Statistical Model for Earthquake And/Or Rainfall Triggered Landslides. *Frontiers in Earth Science*, *8*, 605003. https://doi.org/10.3389/feart.2020.605003

Garcia-Delgado, H., Petley, D. N., Bermúdez, M. A., & Sepúlveda, S. A. (2022). Fatal landslides in Colombia (from historical times to 2020) and their socio-economic impacts. *Landslides*, *19*(7), 1689–1716.

Gelman, A., Carlin, J. B., Stern, H. S., & Rubin, D. B. (2014). *Bayesian data analysis* (3rd ed.). Chapman and Hall/CRC.

Lambert, D. (1992). Zero-Inflated Poisson Regression, with an Application to Defects in Manufacturing. *Technometrics*, *34*(1), 1.

Loche, M., Alvioli, M., Marchesini, I., Bakka, H., & Lombardo, L. (2022). Landslide susceptibility maps of Italy: Lesson learnt from dealing with multiple landslide types and the uneven spatial distribution of the national inventory. *Earth-Science Reviews*, *232*, 104125. https://doi.org/10.1016/j.earscirev.2022.104125

Lombardo, L., Opitz, T., Ardizzone, F., Guzzetti, F., & Huser, R. (2020). Space-time landslide predictive modelling. *Earth-Science Reviews*, *209*, 103318. https://doi.org/10.1016/j.earscirev.2020.103318

Mondini, A. C., Guzzetti, F., & Melillo, M. (2023). Deep learning forecast of rainfall-induced shallow landslides. *Nature Communications*, *14*(1), 2466. https://doi.org/10.1038/s41467-023-38135-y

Osei, F. B., Stein, A., & Andreo, V. (2022). A zero-inflated mixture spatially varying coefficient modeling of cholera incidences. *Spatial Statistics*, *48*, 100635. https://doi.org/10.1016/j.spasta.2022.100635

Peruccacci, S., Gariano, S. L., Melillo, M., Solimano, M., Guzzetti, F., & Brunetti, M. T. (2023). The ITAlian rainfall-induced LandslIdes CAtalogue, an extensive and accurate spatio-temporal catalogue of rainfall-induced landslides in Italy. *Earth System Science Data*, *15*(7), 2863–2877. https://doi.org/10.5194/essd-15-2863-2023

Ramos-Scharrón, C. E., Arima, E. Y., & Hughes, K. S. (2022). An assessment of the spatial distribution of shallow landslides induced by Hurricane María in Puerto Rico. *Physical Geography*, *43*(2), 163–191.

Reichenbach, P., Rossi, M., Malamud, B. D., Mihir, M., & Guzzetti, F. (2018). A review of statistically-based landslide susceptibility models. *Earth-Science Reviews*, *180*, 60–91. https://doi.org/10.1016/j.earscirev.2018.03.001

Schuster, R. L., & Highland, L. M. (2001). *Socioeconomic and environmental impacts of landslides in the western hemisphere*.

Trigila, A., Iadanza, C., Lastoria, B., Bussettini, M., & Barbano, A. (2021). *Dissesto idrogeologico in Italia: Pericolosità e indicatori di rischio* (356/2021). ISPRA.

Vessia, G., Di Curzio, D., Chiaudani, A., & Rusi, S. (2020). Regional rainfall threshold maps drawn through multivariate geostatistical techniques for shallow landslide hazard zonation. *Science of The Total Environment*, *705*, 135815. https://doi.org/10.1016/j.scitotenv.2019.135815