Journal of the Royal Statistical Society Series C: Applied Statistics, 2023, **72**, 1139–1161
https://doi.org/10.1093/jrsssc/qlad077
Advance access publication 13 September 2023
Original Article

ROYAL STATISTICAL SOCIETY
DATA | EVIDENCE | DECISIONS
Journal of the Statistics Society
Series C
Applied Statistics

Joint modelling of landslide counts and sizes using spatial marked point processes with sub-asymptotic mark distributions

Rishikesh Yadav[1], Raphaël Huser[1], Thomas Opitz[2] and Luigi Lombardo[3]

[1]Statistics Program, Computer, Electrical and Mathematical Sciences and Engineering (CEMSE) Division, King Abdullah University of Science and Technology (KAUST), Thuwal, Saudi Arabia
[2]Biostatistics and Spatial Processes (UR546), INRAE, Avignon, France
[3]Faculty of Geo-Information Science and Earth Observation (ITC), University of Twente, Enschede, The Netherlands

Address for correspondence: Raphaël Huser, Statistics Program, Computer, Electrical and Mathematical Sciences and Engineering (CEMSE) Division, King Abdullah University of Science and Technology (KAUST), Thuwal 23955-6900, Saudi Arabia. Email: raphael.huser@kaust.edu.sa

Abstract

To accurately quantify landslide hazard in a region of Turkey, we develop new marked point-process models within a Bayesian hierarchical framework for the joint prediction of landslide counts and sizes. We leverage mark distributions justified by extreme-value theory, and specifically propose 'sub-asymptotic' distributions to flexibly model landslide sizes from low to high quantiles. The use of intrinsic conditional autoregressive priors, and a customised adaptive Markov chain Monte Carlo algorithm, allow for fast fully Bayesian inference. We show that sub-asymptotic mark distributions provide improved predictions of large landslide sizes, and use our model for risk assessment and hazard mapping.

Keywords: Bayesian hierarchical modelling, extreme event, landslide hazard, marked point process, Markov chain Monte Carlo, sub-asymptotic modelling

1 Introduction

Landslides are a severe natural hazard worldwide and are common in mountains and hills where they can pose a severe threat to human lives, disrupt services such as water supply and destroy public and private properties, generating annual damages amounting to billions of dollars (Broeckx et al., 2020; Daniell et al., 2017; Kennedy et al., 2015). It is essential to understand the physical mechanisms triggering devastating landslides, and to assess future landslide risk in terms of various geophysical, geomorphologic, thematic, or climatic factors that can be accurately measured. Susceptibility maps are the most common ways to predict the presence and absence of landslide occurrences over a fixed geographical region. These maps often result from spatial predictive models (Brenning, 2005; Chen et al., 2017). Most of the current research focuses on mapping landslides by exploiting only geographical covariates to predict presence–absence information (Reichenbach et al., 2018). In contrast, Lombardo et al. (2018), Lombardo, Bakka, et al. (2019), Lombardo, Opitz, et al. (2019), and Lombardo et al. (2020) developed a Bayesian hierarchical model with relevant stochastic spatial dependence structure based on the intensity concept for spatial landslide prediction, where landslide counts are viewed as a spatial or spatio-temporal point pattern of log-Gaussian Cox type. They used the integrated nested Laplace approximation (INLA,

Received: November 13, 2022. **Revised:** April 3, 2023. **Accepted:** August 4, 2023

Rue et al., 2009) for the inference of such models. Recently, Opitz et al. (2022) extended such models to incorporate small-scale random variations and space-varying regression to impose certain natural physical constraints. In their framework, high-resolution mapping of landslides was made possible by introducing random effects at a physically defined lower resolution, using the so-called slope units (SUs, Amato et al., 2019). These are physiographic entities corresponding to half-sub-catchments (see Titti, Sarretta, et al., 2022). Their popularity in the landslide literature originates from the fact that two adjacent SUs face each other with opposing angles of exposition (e.g. Titti, Napoli, et al., 2022). Therefore, any landslides occurring in two different SUs are reciprocally independent. As for the covariate information, this was still used at the higher spatial resolution (pixels) in Opitz et al. (2022). However, while this flexible class of models is helpful to predict landslide occurrences and counts per mapping unit, this methodology still completely ignores the actual destructiveness of the landslides themselves, which is perhaps even more important for landslide risk assessment and country planning.

Landslide hazard, by definition, is directly linked to actual landslide size (Guzzetti et al., 1999; Tanyaş et al., 2019), be it expressed as the volume of the displaced mass, or the area of the landslide scar, or indirectly to the landslide 'diameter' and other length-to-width ratio properties (Taylor et al., 2018). Therefore, it is crucial to study the landslide size distribution jointly with landslide occurrences. Here, we shall define the landslide size by (a function of) the landslide planimetric area, which is relatively easy to measure from remotely sensed images and serves as a good proxy for the landslide volume, thanks to a well-known area–volume conversion formula. Precisely, we shall model the square root of the landslide area, which is lighter-tailed than the area itself and therefore allows for simpler statistical estimation and more realistic properties related to the existence of means and variances. Furthermore, it can be measured and intuitively interpreted on the same scale as the landslide diameter. In the literature, other attempts have already been made to model and predict landslide sizes or 'magnitudes' in addition to the modelling of landslide counts (see, e.g. Guo et al., 2017; Lombardo et al., 2021; Roback et al., 2018; Valagussa et al., 2019; Vanani et al., 2021). For example, Guo et al. (2017) used power-law relationships of the size distribution to study earthquake-induced landslides in both the Himalayan and Lesser Himalayan regions. In some related work, Vanani et al. (2021) studied the correlation between the landslide counts and sizes using a bivariate model. More recently, Lombardo et al. (2021) proposed a statistically based model to estimate the area of landslides aggregated over SUs, though they completely ignored their occurrence probability. Specifically, they proposed a Bayesian version of generalised additive models where both the maximum landslide size per SU and the sum of all landslide sizes per SU were modelled using a log-Gaussian model. In contrast, in this paper, we propose using a spatial marked point-process framework to jointly model landslide counts and sizes at high resolution, with various mark distributions strongly justified by extreme-value theory. At the data level, landslide counts are assumed to follow a Poisson distribution and to be independent conditional on their underlying intensity (i.e. the Poisson mean) modelled as a spatial process, while landslide sizes are assumed to follow a so-called 'sub-asymptotic' extreme-value distribution and to be independent conditional on their underlying median modelled through another (correlated) spatial process. The two underlying spatial processes given by the intensity of the count process and the median of the size process are spatially correlated through a shared latent process with a scaling parameter estimated from the data. Shared random-effect techniques for creating the marked point process have been used in various fields; for instance, Gao (2004) used shared random effects to jointly model the probability of disease and probability of death in longitudinal data, and Diggle et al. (2010) demonstrated the utility of a shared latent process for capturing preferential sampling. Alternatively, marked points processes known as intensity-dependent mark models are constructed by using the intensity of a non-homogeneous Poisson process to parametrise the mark process. Our approach has a similar mechanism by including a random effect that acts on both the point-process intensity and the mark distribution. Ho and Stoyan (2008) studied two variants of intensity-dependent mark models for marked Cox point processes, and Jiao et al. (2021) proposed a Bayesian marked point-process model by incorporating the intensity as a fixed effect covariate in the mark process for analysing basketball shot data.

The sub-asymptotic distributions used in this work are parametric models developed to be compliant with extreme-value theory in both tails, with a flexible transition in between; they have been

proposed in recent literature (Naveau et al., 2016; Papastathopoulos & Tawn, 2013; Yadav et al., 2021, 2022) and have the benefit to be theoretically justified by extreme-value theory, thus providing some guarantees for tail extrapolation; see also Stein (2021a, 2021b) who recently developed a constructive framework and new models for sub-asymptotic distributions with flexible behaviour in both tails. These distributions are especially more flexible than the two-parameter generalised Pareto (GP) distribution arising asymptotically for threshold exceedances (Davison & Smith, 1990). As these models are specifically designed to be flexible both in their lower and upper tails, they provide a natural framework for accurately modelling landslide size data all the way from low to high quantiles. This is important in our landslide hazard assessment context, because the largest landslides are usually the most destructive ones and need to be modelled accurately using tail-focussed models, while smaller landslides can still in some cases lead to important damages and should not be ignored. Since the GP distribution arises for the tails of these sub-asymptotic distributions, model parameters thus have an intuitive interpretation, and our framework allows us to assess 'how far' the estimated landslide size distribution is from its upper tail limit behaviour.

In other applied contexts, related approaches have been proposed for the joint statistical modelling of occurrence and size data (Cisneros et al., 2023), sometimes treating sizes as numerical 'marks' of the occurrence positions; see Pimont et al. (2021) and Koh et al. (2023) for wildfire count and burnt area modelling. In particular, assuming a marked spatio-temporal log-Gaussian Cox process (LGCP) for daily wildfire data in France, Pimont et al. (2021) and Koh et al. (2023) developed a joint Bayesian hierarchical model within the INLA framework using so-called 'split-models' for fire sizes: extreme threshold exceedances were assumed to follow the GP distribution, and moderate values followed an appropriately constructed different distribution. Similar split-models were introduced by Patel et al. (2021) to distinguish bulk and tail properties of marks in a self-exciting marked spatio-temporal point process for Afghanistan terror attacks. In contrast, our new framework avoids split-modelling and instead rely on sub-asymptotic extreme-value distributions, which has the great benefit of providing a parsimonious and unified model for the full mark distribution with continuous density, thus bypassing the tricky threshold choice and the unrealistic model structural discontinuity at the threshold. We also note that it is crucial to have a parsimonious model for the mark distribution in order to accurately model relatively small spatial marked point-process datasets. Therefore, it is important to use relatively simple distributions that are well supported theoretically, have interpretable model parameters, and remain flexible in the way extremes are captured. Sub-asymptotic extreme-value distributions are thus ideally placed in view of these elements, and in this work, we test different distributions and compare sub-asymptotic distributions with more classical choices (e.g. log-Gaussian, Gamma, etc.).

The Bayesian hierarchical framework allows for explanatory and predictive modelling, where Gaussian prior distributions at the latent level capture non-linear effects of covariates and spatial coordinates along with spatial dependence (Banerjee et al., 2014; Cressie, 1993). Our marked point processes possess latent Gaussian processes that are combinations of fixed effects, spatial random effects assumed to follow intrinsic conditional autoregressive (ICAR) priors (Besag, 1975), and independent Gaussian random effects. Our Bayesian inference procedure exploits a general and efficient Markov chain Monte Carlo (MCMC) sampling scheme, with customised updates for our proposed random-effect modelling framework. It combines standard Metropolis–Hastings updates (Metropolis et al., 1953), Gibbs sampling (Casella & George, 1992), and the adaptive Metropolis-adjusted Langevin algorithm (MALA) developed in Yadav et al. (2021) to enhance its efficiency with numerous latent Gaussian variables and hyperparameters. Although our MCMC-based inference is computationally more demanding than other approximate Bayesian inference techniques such as INLA, it has two major advantages: (i) MCMC samplers are known to provide 'virtually exact' inferences, provided Markov chains mix well and converge fast enough and the algorithm is run for a sufficient number of iterations and (ii) new types of mark distributions or model structures, currently not available (or not implementable due to conditions on model structure) in other frameworks, can be readily incorporated with only slight implementation changes.

The manuscript is organised as follows. In Section 2, we describe available data and covariates. In Section 3, we specify our Bayesian hierarchical models. Section 4 explains the Bayesian

inference scheme and contains a simulation study to validate accurate MCMC performance. In Section 5, we illustrate our new methodology on the landslide dataset from Turkey. Specifically, we fit different models with various mark distributions and compare different models through cross-validation criteria. Concluding remarks and further research directions are given in Section 6.

2 Landslide data and predictor variables

2.1 Study area and data description

The study area has irregular spatial boundaries and covers 33.2 km^2 in the municipality of Ulus (Bartin) in the Western Black Sea, Turkey, drained by the Ulus river. It lies between the following geographical coordinates: 41°34′35″N–41°34′39″N; 32°37′20″E–32°43′31″E. Figure 1 shows a topographic map of the study region. The landslide inventory is based on airborne Light Detection and Ranging (LiDAR) technology. Elevation within the study region varies from 195 to 900 m with a mean of 375 m, computed from a 1×1 m digital elevation model (DEM). The Ulus region is mostly hilly and has a maritime climate with mean annual precipitation of 1,020 mm. Landslides mostly occur as a result of intense and prolonged rainfall events (Can et al., 2005). For more details, see Görüm (2019). Figure 2 shows landslide locations (bottom left panel) and their landslide areas (m^2) in log-scale (bottom right panel) along with important predictor variables, such as the slope, detailed in Section 2.2. Regions with lower slopes tend to have lower landslide counts and sizes. Landslide counts and sizes are available at high spatial resolution (pixels) with a pixel area of 50×50 m. In total, 933 landslides are observed in the study region, where 836 pixels contain a single landslide, 47 pixels contain two landslides, and 1 pixel contains three landslides.

2.2 Covariate information

Nine morphometric covariates are available, each observed at the 50×50 m pixel level of the LiDAR survey and consisting of a series of elevation derivatives computed from a 1×1 m DEM, namely slope (Zevenbergen & Thorne, 1987), vector ruggedness measure (VRM; Sappington et al., 2007), planar and profile curvatures (planCurv and profCurv, respectively; Heerdegen & Beran, 1982), Topographic Positioning Index (TPI; De Reu et al., 2013), Topographic Wetness Index (TWI; Kirkby & Beven, 1979), local relief (LR; Stepinski & Jasiewicz, 2011), slope height (s-height; Evans, 2019), and valley depth (v-depth; Lóczy et al., 2012). Most of these covariates have already been shown to be associated with the landsliding process; see for example Lombardo et al. (2018), and references therein. Four of these covariates (slope, TWI, planCurv, and profCurv) are graphically illustrated in Figures 2 and 3. All of these covariates are at most moderately correlated, with the estimated Spearman correlation coefficient close to zero for most pairs of covariates and a maximum absolute correlation of 0.384 between LR and s-height; see the cross-correlation matrix in Table 2 in the online supplementary material.

2.3 Mapping units

We consider two types of mapping units covering the entire study region: (i) a high-resolution regular grid (contained within the irregular domain boundaries), comprising 13,275 observed squared pixels (with 4 pixels deleted in the final analysis due to some outliers in covariate values), each of which having an approximate size of 50×50 m $= 2{,}500$ m^2 and (ii) an SU (Amato et al., 2019) partition of the (irregularly shaped) domain, which is at a lower resolution than pixels but is defined in a meaningful physically based way that is relevant to the landsliding process; see, e.g. Lombardo et al. (2018). We have 355 SUs in our study region.

In our marked point-process models, landslide counts, sizes, and covariates are all expressed on the fine spatial grid resolution (pixel level), see Figure 2, whereas the latent spatial effects are defined at the coarser SU resolution to reduce the computational burden; Section 3.2 provides more details on our modelling approach. Defining random effects at high resolution (pixel level) is computationally demanding, and the strong dependence between adjacent pixels might increase numerical instabilities. Instead, SUs are attractive for their physical interpretation because they are geomorphologically independent spatial entities that provide a homogeneous response of a given slope when a landslide occurs, and they also lead to considerably faster and more stable inference because of their coarser resolution. For these reasons, we thus adopt a hybrid modelling approach,

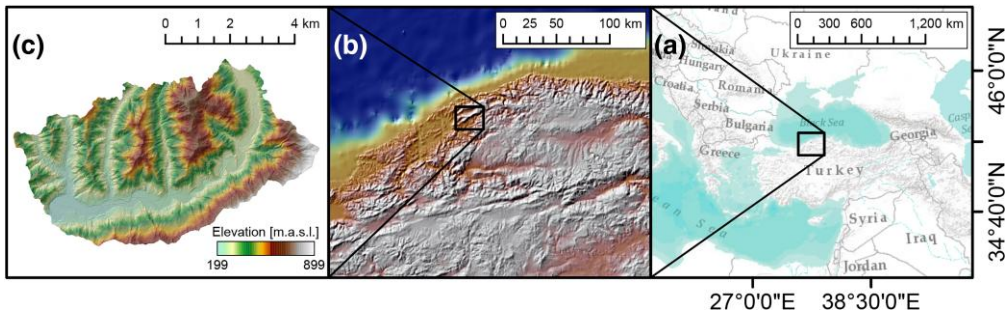

Figure 1. (a) Large geographic context; (b) zoom into the area surrounding the study zone, showing its proximity to the Black Sea in the north, responsible for summer cloudbursts, and to large fault lineaments in the south, responsible for tectonic deformation; (c) study area shown in terms of elevation superimposed on the shaded relief.

with landslide data and covariates at high pixel resolution, and random effects at lower SU resolution. An alternative modelling strategy is to aggregate landslide counts, sizes, and covariates at the SU level, and to build a joint areal model; see Section 3.4 for details. Figure 3 shows an example of covariates as well as the landslide counts and areas aggregated at the SU scale; all covariates are obtained by averaging, and counts and sizes are obtained by adding up the corresponding pixel-based values. While we here mostly focus on marked point-process models in this paper, Section 5.2 summarises results from joint areal models, and the online supplementary material provides further details and a comparison with marked point-process models.

3 Modelling framework

3.1 Log-Gaussian Cox processes

Log-Gaussian Cox processes (see Møller et al., 1998) are Poisson processes with random intensity function given by a log-Gaussian process. Thanks to their doubly stochastic structure, these processes are ideally placed for capturing overdispersion (i.e. variance larger than mean) and for modelling the spatial clustering of points due to unobserved environmental predictors. In Bayesian hierarchical modelling, latent processes are viewed as Gaussian process priors on the Poisson log-intensity function and may contain various fixed and random effects that are aggregated additively to preserve the Gaussian distribution of the linear predictor, and therefore the LGCP structure. To hierarchically define the LGCP, let $\Lambda(s)$ denote the random intensity function of the point process, and let $Y(\mathcal{B})$ denote the random counting process indicating the number of points in a Borel set $\mathcal{B} \subset \mathcal{S}$ contained in the study area $\mathcal{S} \subset \mathbb{R}^2$. At the data level, we have

$$Y(\mathcal{B}) \mid \Lambda(s) \sim \text{Poisson}\left\{\int_{\mathcal{B}} \Lambda(s)\, \mathrm{d}s\right\}. \tag{1}$$

We write $\eta(s)$ for the log-intensity function into which we additively include jointly Gaussian spatial structures via K (a priori independent) random effects X_k, $k = 1, \ldots, K$, P fixed effect covariates z_p with corresponding coefficients β_p, $p = 1, \ldots, P$, and an intercept β_0, i.e.

$$\eta(s) = \log \Lambda(s) = \beta_0 + \sum_{p=1}^{P} \beta_p z_p(s) + \sum_{k=1}^{K} X_k(s), \quad s \in \mathcal{S}.$$

Poisson processes and LGCPs are widely used in the modelling of spatial point patterns and have risen as a counterpart of Gaussian processes (used for continuous variables) for modelling discrete spatial phenomena. Recently, LGCP models have drawn attention in modelling landslide occurrences; see, e.g. Lombardo et al. (2020, 2018, 2021), and Opitz et al. (2022). In this work, we take LGCP models as a basis and in Sections 3.2–3.4, we build joint models for landslide occurrences and sizes (here defined as the square root of landslide areas), where sizes are viewed as numerical marks, each associated with exactly one point. Stochastic dependence could arise between the mark distribution of sizes and the

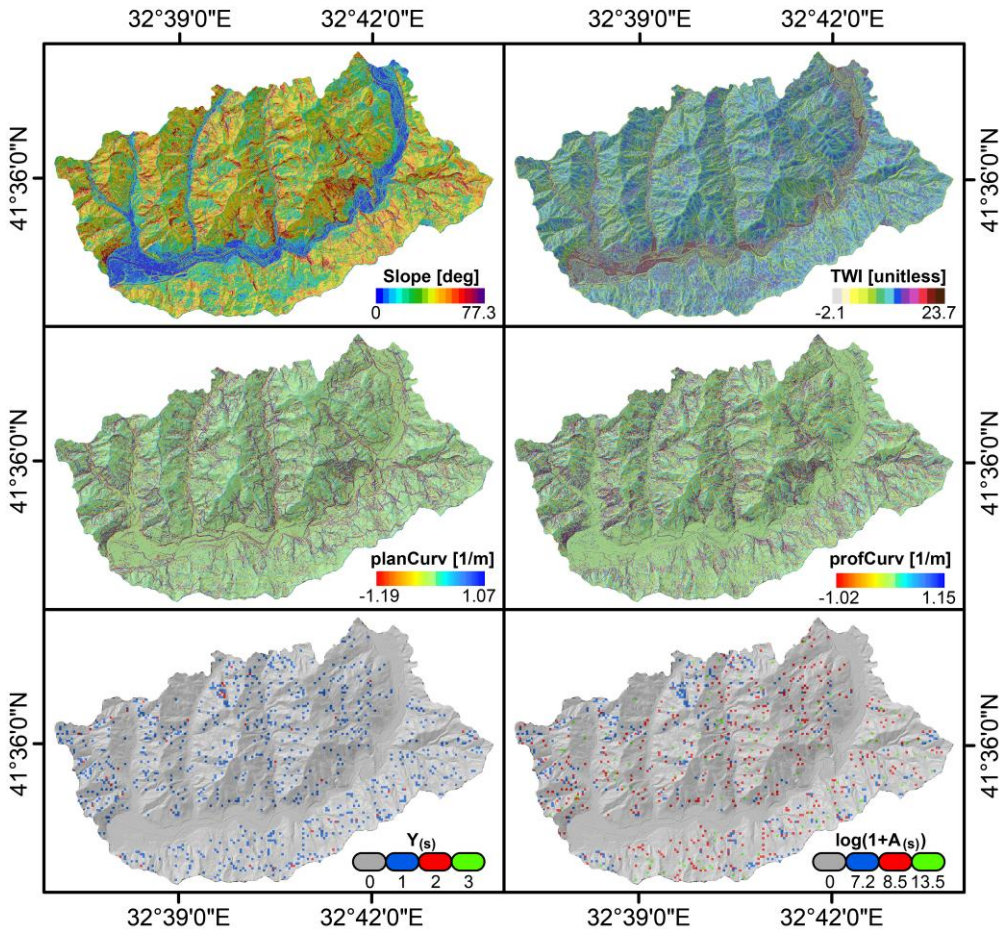

Figure 2. Example of four covariates (details in Section 2.2) at the level of the LiDAR survey: Slope Steepness (top left), Topographic Wetness Index (top right), planar and profile curvatures (middle left and right). Bottom row: number of landslides (left) in a 50 × 50 m lattice; and landslide areas (right; on log-scale) for each landslide occurrence.

random intensity function of occurrences, and our models will be designed to capture such dependence. In practice, the spatial domain $\mathcal{S}$ is often finely discretised for statistical inference on LGCP models, for instance through a fine regular grid, where the intensity function is assumed constant within each grid cell. This is the approach that we consider here using the 50 × 50 m grid of pixels covering our study domain. Therefore, given the grid cells $\mathcal{B}_i$, we can rewrite the LGCP model as a sample of conditionally independent Poisson variables $N(\mathcal{B}_i)$ with random mean that follows a spatially correlated log-Gaussian process, and which provide grid cell-based landslide counts according to equation (1). Small rectangular pixels may be replaced by more general choices of less regular and sometimes larger mapping units to discretise $\mathcal{S}$, for instance physically based SUs, though at the cost of degrading the approximation of the continuous-space point process. Such an approach is described further in Section 3.4.

3.2 Marked point-process models

Throughout the paper, the symbols n_1, n_2, L, N shall be used to denote the total number of pixels, the number of pixels affected by landslides, the number of landslides, and the number of SUs, respectively; in our application, we thus have $n_1 = 13,271$, $n_2 = 884$, $L = 933$, $N = 355$. Moreover, we use generic notation $Y(\boldsymbol{s})$ and $A(\boldsymbol{s})$ for counting and size processes, respectively, observed at the spatial location $\boldsymbol{s}$ (or within the grid cell surrounding its representative point $\boldsymbol{s}$) in the spatial domain $S \subset \mathbb{R}^2$. Hence, $Y(\boldsymbol{s}_i)$ indicates the landslide occurrences over grid cell $i \in \{1, 2, \ldots, n_1\}$ with centroid $\boldsymbol{s}_i$, while $n_2 = |\{l : Y(\boldsymbol{s}_l) > 0\}|$ and $L = \sum_{i=1}^{n_1} Y(\boldsymbol{s}_i)$. We denote the surface area of grid cell i by

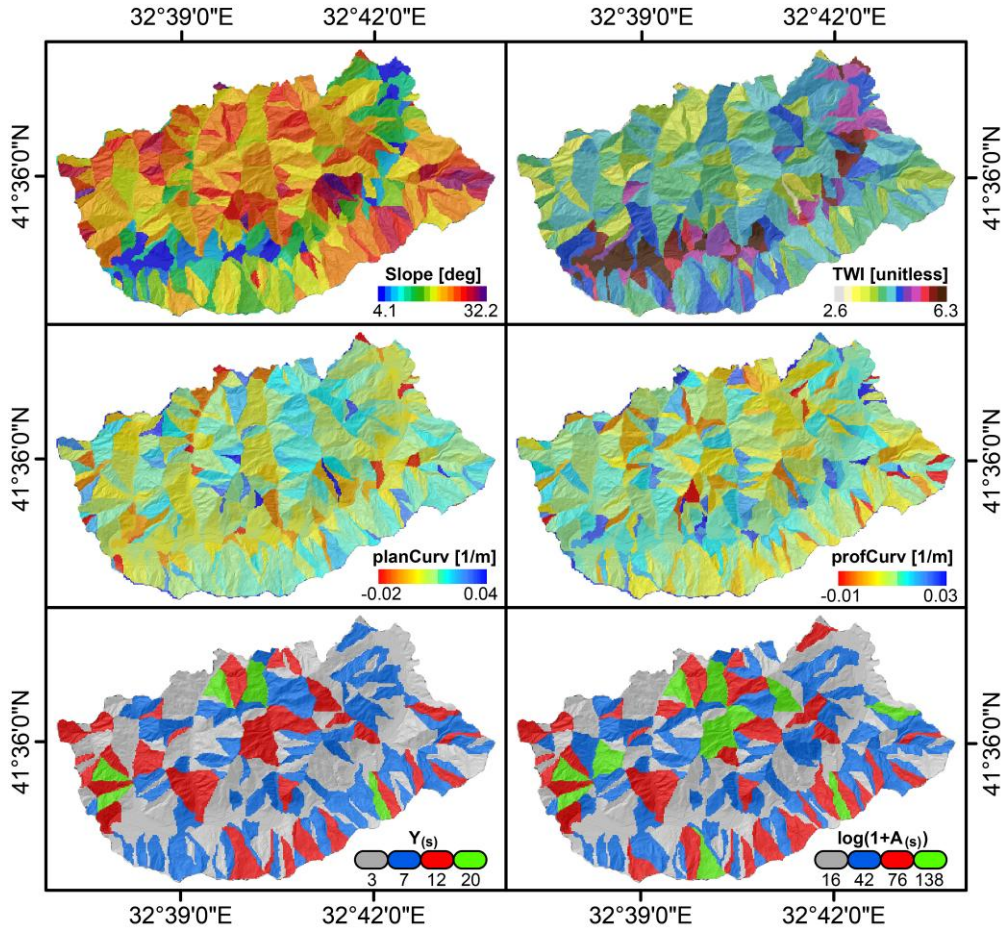

Figure 3. Example of four covariates aggregated to the mean value per slope unit: Slope Steepness (top left), Topographic Wetness Index (top right), planar and profile curvatures (middle left and right). Bottom row: number of landslides (left) and associated landslide areas (right) displayed on log-scale and aggregated at the slope unit level.

$e_i > 0$. While we set $e_i = 1$ in our marked point-process models, given that all pixels have the same area, e_i varies across i for the joint areal models (defined at SU scale) described in Section 3.4. Given a positive landslide count $Y_i = Y(\boldsymbol{s}_i) > 0$, we define $\boldsymbol{A}(\boldsymbol{s}_i) = \{A_1(\boldsymbol{s}_i), \ldots, A_{Y_i}(\boldsymbol{s}_i)\} \in (0, \infty)^{Y_i}$ as the continuous mark vector gathering the landslide sizes (i.e. square root of landslide areas) at the grid cell i. We refer to the mark distribution through its probability density f_A with parameter vector $\boldsymbol{\Theta}_A$ and median denoted by the symbol μ (with appropriate subscripts whenever needed), and Section 3.3 below details possible choices. Then, we define our general marked point-process model as follows:

$$
\begin{aligned}
&Y(\boldsymbol{s}_i) \mid \eta(\boldsymbol{s}_i) \overset{\text{ind}}{\sim} \text{Poisson}\left[e_i \exp\left\{\eta(\boldsymbol{s}_i)\right\}\right], \quad i = 1, \ldots, n_1; \\
&\eta(\boldsymbol{s}_i) = \gamma_1 + \sum_{p=1}^{P_1} \beta_{1p} z_{1p}(\boldsymbol{s}_i) + \boldsymbol{b}_1(\boldsymbol{s}_i)^T \boldsymbol{W}_1 + \varepsilon_{\boldsymbol{\eta}}(\boldsymbol{s}_i); \\
&A_j(\boldsymbol{s}_l) \mid \mu_j(\boldsymbol{s}_l) \overset{\text{ind}}{\sim} f_A[\,\cdot\,;\, \exp\{\mu_j(\boldsymbol{s}_l)\}, \boldsymbol{\Theta}_A], \quad l : Y(\boldsymbol{s}_l) > 0;\ j = 1, \ldots, Y(\boldsymbol{s}_l); \\
&\mu_j(\boldsymbol{s}_l) = \gamma_2 + \sum_{q=1}^{P_2} \beta_{2q} z_{2q}(\boldsymbol{s}_l) + \beta \boldsymbol{b}_2(\boldsymbol{s}_l)^T \boldsymbol{W}_1 + \boldsymbol{b}_2(\boldsymbol{s}_l)^T \boldsymbol{W}_2 + \varepsilon_{\boldsymbol{\mu}}(\boldsymbol{s}_l).
\end{aligned}
\tag{2}
$$

Here, the processes $\boldsymbol{W}_1$, $\boldsymbol{W}_2$ are two spatially structured random vectors, while $\varepsilon_{\boldsymbol{\eta}}(\cdot)$, $\varepsilon_{\boldsymbol{\mu}}(\cdot)$ are independent (spatially unstructured) random effects, defined at the pixel level, included in the log-intensity process $\eta(\boldsymbol{s})$, and log-median process $\mu(\boldsymbol{s})$, respectively. In particular, we assume that the spatial random effects $\boldsymbol{W}_1$ and $\boldsymbol{W}_2$ follow an ICAR prior (Besag, 1975; Besag & Kooperberg, 1995; Morris et al., 2019) defined on SUs, so that the vectors $\boldsymbol{W}_1$ and $\boldsymbol{W}_2$ are of moderate dimension $N = 355$. More explicitly, $\boldsymbol{W}_1$ and $\boldsymbol{W}_2$ follow a singular multivariate Gaussian distribution with mean vector zero and precision matrices $\tau_{w_1} Q$, and $\tau_{w_2} Q$, respectively, where $\tau_{w_1}, \tau_{w_2} > 0$ are related to the marginal precision parameters. The matrix Q is fixed for a given spatial structure and has rank $N - 1$; it can be obtained as $Q = D - A$, where $D = \text{diag}(d_1, \ldots, d_N)$ is the diagonal matrix constructed from the number of neighbours d_i of SU i, and A is the adjacency matrix between SUs. The vectors $\boldsymbol{b}_1(\boldsymbol{s}_i)$ and $\boldsymbol{b}_2(\boldsymbol{s}_l)$ project the information from the SU to the pixel level for count and size processes, respectively, and are defined as: $\boldsymbol{b}_1(\boldsymbol{s}_i) = \{b_{11}(\boldsymbol{s}_i), \ldots, b_{1N}(\boldsymbol{s}_i)\}^T$, with $b_{1r}(\boldsymbol{s}_i) = 1$ if the pixel $\boldsymbol{s}_i$ belongs to the rth SU and $b_{1r}(\boldsymbol{s}_i) = 0$ otherwise; similarly, $\boldsymbol{b}_2(\boldsymbol{s}_l) = \{b_{21}(\boldsymbol{s}_l), \ldots, b_{2N}(\boldsymbol{s}_l)\}^T$, with $a_{2r}(\boldsymbol{s}_l) = 1$ if the pixel $\boldsymbol{s}_l$ belongs to the rth SU, and $a_{2r}(\boldsymbol{s}_l) = 0$ otherwise, where l is such that $Y(\boldsymbol{s}_l) > 0$ and $l = 1, \ldots, L$. The observed variables $z_{11}(\boldsymbol{s}_i), \ldots, z_{1P_1}(\boldsymbol{s}_i)$ denote the P_1 fixed effect covariates measured at the ith location that affect the count process $Y(\boldsymbol{s})$, and similarly, $z_{21}(\boldsymbol{s}_l), \ldots, z_{2P_2}(\boldsymbol{s}_l)$ denoted the P_2 fixed effect covariates measured at the lth location (such that $Y(\boldsymbol{s}_l) > 0$) that affect the mark process $A(\boldsymbol{s})$. The parameter vectors $\boldsymbol{\beta}_1 = (\beta_{11}, \ldots, \beta_{1P_1})^T$ and $\boldsymbol{\beta}_2 = (\beta_{21}, \ldots, \beta_{2P_2})^T$ are the corresponding covariate coefficients, and γ_1, γ_2 are the corresponding intercept parameters. Thus, while the random effects $\boldsymbol{W}_1$ and $\boldsymbol{W}_2$ are defined at the SU level (low resolution), the fixed effects are still used at the pixel level (high resolution). The parameter β scales the common component $\boldsymbol{W}_1$ and controls how much information is shared from the count predictor towards the size predictor and determines the strength of interaction between the two processes. Precisely, it allows capturing positive ($\beta > 0$) or negative ($\beta < 0$) correlation. If $\beta = 0$, then $Y(\boldsymbol{s})$ is independent of $A(\boldsymbol{s})$, and the model (2) for counts becomes the standard LGCP model (1).

3.3 Sub-asymptotic mark distributions

The GP distribution is the asymptotic model arising for the excesses over a high threshold when the threshold increases towards the upper endpoint of the distribution. If the random variable Y follows a distribution F with upper endpoint $y_F = \sup\{y \in \mathbb{R} : F(y) < 1\}$, then for a wide class of distributions F, high threshold exceedances $Y - u \mid Y > u$ may be asymptotically approximated as

$$\Pr(Y - u \leq y \mid Y > u) = \frac{F(u + y) - F(u)}{1 - F(u)} \approx H_{\tau,\xi}(y) = 1 - (1 + \xi y/\tau)^{-1/\xi}, \tag{3}$$

as $u \to y_F$, where $H_{\tau,\xi}(y)$ denotes the GP distribution with scale $\tau > 0$ and shape ξ. Although the GP distribution is a natural model for exceedances above high thresholds, there are several limitations for its use in practice (see, e.g. Naveau et al., 2016; Yadav et al., 2021, for details). One such limitation is due to its asymptotic characterisation; the approximation of F by the GP distribution may be appropriate only for very high thresholds u, which results in very small datasets of exceedances for model fitting and hence larger uncertainly in the parameter estimates. Moreover, if modelling values below the threshold remains important for the application, then the GP distribution would have to be combined with another model applying below the threshold. On the other hand, the so-called 'sub-asymptotic distributions' are designed to be more flexible than the classical GP distribution and are intended to mitigate its limitations; see Naveau et al. (2016), Yadav et al. (2021), and Stein (2021a, 2021b).

We here use the generic term 'sub-asymptotic distributions' for probability distributions applying to the full range of possible values of landslide sizes but possessing high flexibility, as required by extreme-value theory, for modelling and interpretation of both the lower and upper tail properties controlled by specific parameters. These distributions provide different degrees of flexibility for separate control over bulk and tail features depending on their overall number of parameters. By carefully selecting the probability density $f_A(\cdot\,; \mu, \boldsymbol{\Theta}_A)$ used to model the mark process $A(\boldsymbol{s})$, we obtain models (2) with structurally different sub-asymptotic mark distributions; see Table 1 for

Table 1. Density functions of different mark distributions, reparametrised such that μ represents the median

Marks	Reparametrised density $f_A(x)$, $x > 0$	$\boldsymbol{\Theta}_A$	ξ_U, ξ_L	κ_U
gen-Gam	$f_{gGam}(x) = \frac{(c/\sigma^\kappa)}{\Gamma(\kappa/c)} \frac{x^{\kappa-1}}{\exp\{(x/\sigma)^c\}}$, $\sigma = \frac{\mu}{\{F^{-1}_{Gam}(0.5;\kappa/c,1)\}^{1/c}}$	$\mu>0$, $\kappa>0$, $c>0$	$0, -1/\kappa$	c
Gamma	$f_{Gam}(x) = f_{gGam}(x)$ with $c = 1$	$\mu>0$, $\kappa>0$	$0, -1/\kappa$	1
Weibull	$f_{Wb}(x) = f_{gGam}(x)$ with $c = \kappa$	$\mu>0$, $\kappa>0$	$0, -1/\kappa$	κ
l-Gamma	$f_{lGam}(x) = \frac{1}{\sigma^\kappa\Gamma(\kappa)} \frac{\{\log(x+1)\}^{\kappa-1}}{(x+1)^{1+1/\sigma}}$, $\sigma = \frac{\log(1+\mu)}{F^{-1}_{Gam}(0.5;\kappa,1)}$	$\mu>0$, $\kappa>0$	$\sigma, -1/\kappa$	0
l-Gaussian	$f_{lGau}(x) = \frac{\sqrt{\kappa}}{x\sqrt{2\pi}} \exp[-\frac{\kappa}{2}\{\log(x) - \log(\mu)\}^2]$	$\mu>0$, $\kappa>0$	$0, -\infty$	0
Burr*	$f_{Bu}(x) = \frac{(c\kappa/\sigma)(x/\sigma)^{c-1}}{\{1+(x/\sigma)^c\}^{\kappa+1}}$, $\sigma = \frac{\mu}{(2^{1/\kappa}-1)^{1/c}}$	$\mu>0$, $\kappa>0$, $c>0$	$1/(c\kappa), -1/c$	0
ext-GP*	$f_{eGP}(x) = \begin{cases} \frac{\kappa\{1-(1+\xi x/\sigma)^{-1/\xi}\}^{\kappa-1}}{\sigma(1+\xi x/\sigma)_+^{1+1/\xi}}, & \xi>0, \\ & \sigma = \frac{\mu\xi}{(1-0.5^{1/k})^{-\xi}-1} \\ \frac{\kappa\{1-\exp(-x/\sigma)\}^{\kappa-1}}{\sigma\exp(x/\sigma)}, & \xi\to 0, \\ & \sigma = \frac{\mu}{-\log(1-0.5^{1/k})} \end{cases}$	$\mu>0$, $\kappa>0$, $\xi\geq 0$	$\xi, -1/\kappa$	$\mathbb{I}(\xi=0)$
GP	$f_{GP}(x) = f_{eGP}(x)$ with $\kappa = 1$	$\mu>0$, $\xi\geq 0$	$\xi, -1$	$\mathbb{I}(\xi=0)$
Gam–Gam*	$f_{GG}(x) = \frac{(c_1/c_2)^{c_1/2}}{\alpha B(c_1/2, c_2/2)} \times \frac{(x/\sigma)^{-1+c_1/2}}{(1+c_1x/c_2\sigma)^{(c_1+c_2)/2}}$, $\sigma = \frac{\mu}{F^{-1}_f(0.5, c_1/2, c_2/2)}$	$\mu>0$, $c_1>0$, $c_2>0$	$2/c_2, -2/c_1$	0

Note. Incomplete gamma function $\Gamma(x) = \int_0^x t^{x-1}\exp(-t)\mathrm{d}t$; beta function $B(a,b) = \int_0^1 t^{a-1}(1-t)^{b-1}\mathrm{d}t$; inverse of gamma cumulative distribution function (cdf) $F^{-1}_{Gam}(x;a,b)$ for $x\in[0,1]$ with shape a and rate b; inverse cdf $F^{-1}_f(x;d_1,d_2)$ of the Snedecor's F distribution for $x>0$ and degrees of freedom d_1 and d_2; Asterisks (*) indicate three-parameter families of sub-asymptotic distributions that comply with extreme-value theory in both tails by allowing for any possible combination of positive upper tail index ξ_U and negative lower tail index ξ_L; finally, κ_U denotes the upper tail Weibull index; $a_+ = a$ if $a>0$ and 0 otherwise; $\mathbb{I}$ is the indicator function. gen-Gam = generalised-Gamma; l-Gamma = log-Gamma; l-Gaussian = log-Gaussian; ext-GP = extended generalised Pareto; Gam–Gam = Gamma–Gamma.

certain possible choices of f_A and their tail characteristics. Various joint occurrence-size models are obtained by combining different mark distributions with different specifications of the structured spatial and independent random effects; see Section 5 for details.

A broad classification of tail asymptotics is given by the value of the extreme-value tail index ξ. Positive values of $\xi>0$ indicate heavy tails with slow power-law tail decay; a value $\xi=0$ implies an exponential-type tail decay but with a large variety of possible sub-asymptotic structures; and $\xi<0$ characterises distributions with finite upper bound and polynomial tail decay towards this bound. More refined classifications are possible, see Engelke et al. (2019) for an overview. We write ξ_L and ξ_U to specify the lower and upper tail index, respectively. We consider distributions with support on $[0,\infty)$, which requires $\xi_L<0$ and $\xi_U\geq 0$. If we exclude the boundary cases of the upper tail index being 0, the GP upper ($y\to\infty$) and lower ($y\to 0$) tail behaviour may be formally defined by the following tail expansions, respectively:

$$1-F(y) = \ell^U(y)y^{-1/\xi_U}, \quad y\to\infty, \quad F(y) = \ell^L(1/y)y^{-1/\xi_L}, \quad y\to 0,$$

where $\ell^U(y)$ and $\ell^L(y)$ are slowly varying at infinity, i.e. $\ell^U(ty)/\ell^U(t)\to 1$ as $t\to\infty$, and similarly for $\ell^L(y)$. Within the particularly large class with $\xi_U=0$, both light and heavy tails are possible, and a relevant refinement is given by the Weibull tail behaviour defined by survival functions of the form

$$1-F(y) = \ell^U(y)y^\alpha\exp\left(-\gamma_{Wb}y^{\kappa_U}\right), \quad \alpha\in\mathbb{R}, \ \kappa_U,\gamma_{Wb}>0,$$

where κ_U is the upper tail *Weibull index*, with heavy tails for $\kappa_U<1$ and light tails for $\kappa_U\geq 1$; exponential tails with $\kappa_U=1$ form the boundary between these two scenarios.

Table 1 lists several mark distribution families with their reparametrised density, lower and upper tail indices, as well as the upper tail Weibull index, which we set to 0 if tails are heavier than the Weibull class. The Gamma–Gamma (Gam–Gam), extended-GP (ext-GP), and Burr distributions are particularly flexible sub-asymptotic distributions with separate parameters to control the lower tail, the bulk, and the right tail of the distribution. Another flexible example is the generalised-Gamma (gen-Gam) distribution; despite its exponential upper tail ($\xi_U = 0$), the parameters κ and c can allow for efficient modelling of moderate to large observations and of differences in tail heaviness. The log-Gamma (l-Gamma) and GP distributions are examples of parsimonious heavy-tailed models having only two parameters with positive tail index ($\xi_U > 0$), but they lack flexibility in the lower tail or the bulk. The Weibull distribution is also relatively flexible in capturing behaviour in small to large observations, but it has $\xi_U = 0$ and only two parameters, thus lacking flexibility compared to the Burr, Gam–Gam, and ext-GP models. The Gamma and log-Gaussian (l-Gaussian) distributions are examples of two classical models with low flexibility in the upper tail ($\xi_U = 0$), though their simplicity could make them statistically more robust for prediction. For our real data application in Section 5, we use all these distributions as candidate mark distributions in our models, comparing their performances on our landslides dataset.

3.4 An alternative joint model using areal data

Building upon the general formulation (2) of marked point processes, several alternative joint areal models may be constructed at coaser resolution levels (up to minor model modifications). A relevant option is to define landslide counts, sizes and observed fixed effects at low resolution on the SU scale (thus, the unit area e_i now varies for each i). In model (2), the spatial random effects $\boldsymbol{W}_1$ and $\boldsymbol{W}_2$ are already defined at the SU level. Landslide counts and sizes aggregated to SU resolution can be obtained by adding up their pixel-based values within each SU, and similarly, fixed effect covariates at SU level may be obtained by averaging them over the corresponding pixels. This alternative areal model is obtained from the general model (2) through the following changes. The index j is removed since it is always equal to one; the variables $Y(\boldsymbol{s}_i)$, $A(\boldsymbol{s}_l)$, $z_{1p}(\boldsymbol{s}_i)$, and $z_{2q}(\boldsymbol{s}_l)$ are all defined at SU level by adding up or averaging their pixel-based values. Finally, $b_{1r}(\boldsymbol{s}_i) = 1$ if $r = i$, and $b_{1r}(\boldsymbol{s}_i) = 0$ otherwise, and similraly for $a_{2r}(\boldsymbol{s}_l)$. In practice, we can model $\{Y(\boldsymbol{s}), A(\boldsymbol{s})\}^T$ as a bivariate random field defined at the SU scale, and treat $A(\boldsymbol{s}_i)$ as missing (NA) whenever $Y(\boldsymbol{s}_i) = 0$.

This model is no longer a marked point process due to its aggregation of the marks within each SU, such that information about individual points is lost. The structure of this model also differs from the approach of simply integrating predictions of model (2) over SUs, since covariates have also been aggregated in the new model. It is a joint areal model (i.e. a so-called lattice or network model) for landslide counts and sizes and could still provide valuable insights into their joint behaviour despite its lower resolution.

4 Bayesian inference

Hierarchical model constructions suggest using Bayesian inference. Our proposals belong to the class of latent Gaussian models, and hence approximate Bayesian estimation techniques, such as the INLA (Rue et al., 2009) or Max-and-Smooth (Hrafnkelsson et al., 2021) that are well adapted for this type of models, could be used at relatively high spatial resolution under mild conditions on the mark distribution. However, we consider in our work various types of mark distributions that are not available in off-the-shelf implementations (e.g. the `R-INLA` software), or which do not necessarily comply with the theoretical requirements of INLA. Hence, we instead use simulation-based inference based on MCMC sampling, combined with efficient block proposals ensuring good mixing and fast convergence of Markov chains, and thus virtually exact posterior inference for all of our proposed models.

4.1 MCMC sampler

Let $\boldsymbol{Y} = \{Y(\boldsymbol{s}_1), \ldots, Y(\boldsymbol{s}_{n_1})\}^T$ denote the random vector of counts and $\boldsymbol{\eta} = \{\eta(\boldsymbol{s}_1), \ldots, \eta(\boldsymbol{s}_{n_1})\}^T$ the corresponding random vector of log-intensities. Let $\boldsymbol{U} = \{l : Y(\boldsymbol{s}_l) > 0\}$ denote the set of indices (pixels for the model in Section 3.2, SUs for the model in Section 3.4) for which the mark process A is defined, with n_2 the length of vector $\boldsymbol{U} = (U_1, \ldots, U_{n_2})^T$. Then, $\boldsymbol{A} = \{\boldsymbol{A}(\boldsymbol{s}_{U_1})^T, \ldots, \boldsymbol{A}(\boldsymbol{s}_{U_{n_2}})^T\}^T$ is

the vector of landslide sizes, with $\boldsymbol{A}(\boldsymbol{s}_l)$ representing the vector of landslide sizes observed at the lth pixel, and $\boldsymbol{\mu} = \{\boldsymbol{\mu}(\boldsymbol{s}_{U_1})^T, \ldots, \boldsymbol{\mu}(\boldsymbol{s}_{U_{n_2}})^T\}^T$ is the log-median random vector corresponding to $\boldsymbol{A}$.

We use conjugate priors whenever possible. We assume independent Gaussian priors with mean zero and variance $\sigma^2_{\boldsymbol{\beta}_1} = \sigma^2_{\boldsymbol{\beta}_2} = \sigma^2_{\gamma_1} = \sigma^2_{\gamma_2} = \sigma^2_{\beta} = 100$ for the covariate coefficients β_{1p}, $p = 1, \ldots, P_1$, β_{2q}, $q = 1, \ldots, P_2$, intercepts γ_1 and γ_2, and sharing parameter β. We set relatively informative gamma hyperpriors for the precision parameters $\kappa_{\boldsymbol{\eta}}$, $\kappa_{\boldsymbol{\mu}}$, $\kappa_{\boldsymbol{w}_1}$, and $\kappa_{\boldsymbol{w}_2}$, where the parameters of the hyperpriors are chosen such as to place equal emphasis on the variances of the random effects with and without spatial correlation. We assume gamma hyperpriors with shape $a_{\kappa_{\boldsymbol{\eta}}} = a_{\kappa_{\boldsymbol{\mu}}} = 0.25$ and rate $b_{\kappa_{\boldsymbol{\eta}}} = b_{\kappa_{\boldsymbol{\mu}}} = 3$ for $\kappa_{\boldsymbol{\eta}}$ and $\kappa_{\boldsymbol{\mu}}$. For the precision parameters of spatial random effects, we set gamma hyperpriors with shape $a_{\kappa_{\boldsymbol{w}_1}} = a_{\kappa_{\boldsymbol{w}_2}} = 0.25$ and rate $b_{\kappa_{\boldsymbol{w}_1}} = b_{\kappa_{\boldsymbol{w}_2}} = 3/(\overline{m}0.7^2)$, where $\overline{m}$ is the average number of neighbours across the whole region (i.e. the average number of adjacent SUs to a given SU); see Bernardinelli et al. (1995) for details. For the hyperparameters in $\boldsymbol{\Theta}_A$, we set independent gamma priors with shape $a_{\boldsymbol{\Theta}_A} = 0.25$ and rate $b_{\boldsymbol{\Theta}_A} = 0.25$. We also studied the sensitivity of results to the choice of hyperpriors by fixing their parameters to several different configurations, and we observed only slight differences in posterior distributions; see Section 2.3 in the online supplementary material for details.

The hierarchical joint occurrence-size model (2) can now be fully specified as

$$
\begin{aligned}
Y(\boldsymbol{s}_i) \mid \eta(\boldsymbol{s}_i) &\overset{\text{ind}}{\sim} \text{Poisson}[e_i \exp\{\eta(\boldsymbol{s}_i)\}], \quad i = 1, \ldots, n_1;\\
\boldsymbol{\eta} \mid \gamma_1, \boldsymbol{\beta}_1, \kappa_{\boldsymbol{\eta}}, \boldsymbol{W}_1 &\sim \mathcal{N}_{n_1}(\gamma_1 \mathbf{1} + \boldsymbol{Z}_1\boldsymbol{\beta}_1 + \boldsymbol{B}_1\boldsymbol{W}_1, \kappa_{\boldsymbol{\eta}}^{-1}\boldsymbol{I}_{n_1});\\
\boldsymbol{W}_1 \mid \kappa_{\boldsymbol{w}_1} &\sim \mathcal{N}_N(0, \kappa_{\boldsymbol{w}_1}^{-1}\mathcal{Q}^{-1}); \quad (\boldsymbol{\beta}_1^T, \gamma_1)^T \sim \mathcal{N}_{P_1+1}(0, \sigma^2_{\boldsymbol{\beta}_1}\boldsymbol{I}_{P_1+1});\\
\kappa_{\boldsymbol{\eta}} &\sim \text{Gamma}(a_{\kappa_{\boldsymbol{\eta}}}, b_{\kappa_{\boldsymbol{\eta}}}), \quad \kappa_{\boldsymbol{w}_1} \sim \text{Gamma}(a_{\kappa_{\boldsymbol{w}_1}}, b_{\kappa_{\boldsymbol{w}_1}}),\\
A_j(\boldsymbol{s}_l) \mid \mu_j(\boldsymbol{s}_l) &\overset{\text{ind}}{\sim} f_A[\,\cdot\,;\ \exp\{\mu_j(\boldsymbol{s}_l)\}, \boldsymbol{\Theta}_A], \quad l : Y(\boldsymbol{s}_l) > 0,\ j = 1, \ldots, Y(\boldsymbol{s}_l);\\
\boldsymbol{\mu} \mid \gamma_2, \boldsymbol{\beta}_2, \boldsymbol{W}_1, \boldsymbol{W}_2, \beta, \kappa_{\boldsymbol{\mu}} &\sim \mathcal{N}_L(\gamma_2 \mathbf{1} + \boldsymbol{Z}_2\boldsymbol{\beta}_2 + \beta\boldsymbol{B}_2\boldsymbol{W}_1 + \boldsymbol{B}_2\boldsymbol{W}_2, \kappa_{\boldsymbol{\mu}}^{-1}\boldsymbol{I}_L);\\
\boldsymbol{W}_2 \mid \kappa_{\boldsymbol{w}_2} &\sim \mathcal{N}_N(0, \kappa_{\boldsymbol{w}_2}^{-1}\mathcal{Q}^{-1}); \quad (\boldsymbol{\beta}_2^T, \gamma_2, \beta)^T \sim \mathcal{N}_{P_2+2}(0, \sigma^2_{\boldsymbol{\beta}_2}\boldsymbol{I}_{P_2+2});\\
\kappa_{\boldsymbol{\mu}} &\sim \text{Gamma}(a_{\kappa_{\boldsymbol{\mu}}}, b_{\kappa_{\boldsymbol{\mu}}}), \quad \kappa_{\boldsymbol{w}_2} \sim \text{Gamma}(a_{\kappa_{\boldsymbol{w}_2}}, b_{\kappa_{\boldsymbol{w}_2}}),
\end{aligned}
\tag{4}
$$

where $L = \sum_{i=1}^{n_1} Y(\boldsymbol{s}_i)$, and $\boldsymbol{B}_1$ and $\boldsymbol{B}_2$ are the projection matrices of dimensions $n_1 \times N$ and $L \times N$ filled row-by-row appropriately with the N-dimensional vectors $\boldsymbol{b}_1(\boldsymbol{s}_i)$ and $\boldsymbol{b}_2(\boldsymbol{s}_l)$, respectively; $\boldsymbol{Z}_1$ and $\boldsymbol{Z}_2$ are the design matrices of dimension $n_1 \times P_1$ and $L \times P_2$ corresponding to the P_1- and P_2-dimensional vectors of covariate coefficients $\boldsymbol{\beta}_1$ and $\boldsymbol{\beta}_2$, respectively; $\mathcal{N}_d(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ denotes the d-variate Gaussian distribution with mean vector $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$; $\boldsymbol{I}_n$ denotes the n-by-n identity matrix; Gamma(a, b) denotes the gamma distribution with shape a and rate b.

We update the model parameters $\kappa_{\boldsymbol{\eta}}, \kappa_{\boldsymbol{\mu}}, \kappa_{\boldsymbol{w}_1}, \kappa_{\boldsymbol{w}_2}, \boldsymbol{\beta}_1, \boldsymbol{\beta}_2, \gamma_1, \gamma_2, \beta, \boldsymbol{W}_1$, and $\boldsymbol{W}_2$ using the standard Gibbs sampling algorithm (Casella & George, 1992) by exploiting the analytically available conditional distributions of these parameters. For hyperparameters in $\boldsymbol{\Theta}_A$, we implement a standard Metropolis algorithm (Metropolis et al., 1953). The full conditional distributions of $\boldsymbol{\eta}$ and $\boldsymbol{\mu}$ do not have closed-form expressions, and hence we use the adaptive MALA proposed in Yadav et al. (2021) to efficiently simulate from the full conditionals of both $\boldsymbol{\eta}$ and $\boldsymbol{\mu}$ in two separate blocks.

Let $\boldsymbol{W}_1 = (W_{11}, \ldots, W_{1N})^T$ and $\boldsymbol{W}_2 = (W_{21}, \ldots, W_{2N})^T$ be the random vectors of latent spatial effects having ICAR prior. We centre their joint distributions at zero and write $\kappa_{\boldsymbol{w}_1}$ and $\kappa_{\boldsymbol{w}_2}$, respectively, for marginal precisions. We rewrite their densities using the pairwise difference representation:

$$
\pi(\boldsymbol{W}_h) \propto \exp\left\{\frac{-\kappa_{\boldsymbol{w}_h}}{2}\sum_{i\sim j}(W_{hi} - W_{hj})^2\right\}, \quad h = 1, 2, \tag{5}
$$

where $i \sim j$ means that the ith and jth SUs are neighbours sharing a common boundary. The joint density of the ICAR random vectors given by equation (5) leads to a moderate computational burden. However, the pairwise difference is non-identifiable, i.e. any constant added to $\boldsymbol{W}_h$ cancels out in

terms $W_{hi} - W_{hj}$, hence the intercept terms γ_1 and γ_2 are non-estimable. To avoid the issue of non-identifiability with non-estimable intercepts, we add the constraint $\sum_{i=1}^{N} W_{hi} = 0$, $h = 1, 2$, to centre the latent spatial effects. In our MCMC algorithm, this is achieved by centring the Markov chains of both $\boldsymbol{W}_1$ and $\boldsymbol{W}_2$ at every iteration by subtracting their means, which is called hard-centring. Another way of centring the ICAR component is by assigning a Gaussian prior with mean 0 and small variance (e.g. 0.01) to the average $\sum_{i=1}^{N} W_{hi}/N$, $h = 1, 2$, hence instead of summing exactly to zero, this method 'soft-centres' the mean by keeping it close to zero (Morris et al., 2019). Yet another way of imposing the constraint is used by the INLA implementation with non-degeneracy achieved by conditioning the Gaussian vector to sum to zero.

4.2 Simulation experiment and sanity check

We conducted several simulation experiments to check the estimation and prediction performance of our MCMC sampler detailed in Section 4.1. For conciseness, we report only the results for a simulation scenario similar to the data application in Section 5, where we use a log-Gaussian size distribution for which we can further compare results across our MCMC sampler and the INLA method. We simulate data from the marked point-process model (2) for a fixed parameter configuration given below. We use the adjacency structure of the SUs in the study area presented in Section 2 to simulate the spatially correlated random effects $\boldsymbol{W}_1$ and $\boldsymbol{W}_2$ with $\kappa_{w_1} = 2$ and $\kappa_{w_2} = 2$, and we set the precisions of spatially independent effect to $\kappa_{\eta} = \kappa_{\mu} = 3$. Fixed-effect design matrices $\boldsymbol{Z}_1$ and $\boldsymbol{Z}_2$ are filled with the standardised covariates available in the landslide dataset; see Section 2.2 (slope, TWI, VRM, profCurv, planCurv, TPI, LR, s-height, v-depth), and the corresponding covariate coefficients are set to $\boldsymbol{\beta}_1 = (0.2, 0.2, 0.2, -0.2, -0.2, -0.2, 0.15, 0.15, 0.15)^T$ and $\boldsymbol{\beta}_2 = (0.15, 0.15, 0.15, -0.1, -0.1, -0.1, 0.2, 0.2, 0.2)^T$. We use the log-Gaussian distribution as the mark distribution for the size with precision (i.e. inverse variance of the underlying Gaussian) set to $\kappa = 5$. The sharing parameter is set to $\beta = 1$. The simulation set-up is exactly the same as in the data application in Section 5, and we have $n_1 = 13{,}271$ pixels, $N = 355$ SUs, $P_1 = P_2 = 9$ covariates, and $L = 933$ simulated landslides.

We use our MCMC sampler with prior distributions of hyperparameters as discussed in Section 4.1 for the inference of the model (2). We generate 250,000 MCMC samples and discard 187,500 samples as burn-in, and all summary statistics are based on the last 62,500 samples, resulting in 625 samples after thinning by a factor 100 to strongly reduce autocorrelation in MCMC series. The Markov chains (see Figure 2 in the online supplementary material) for all hyperparameters and latent parameters visibly converge after about 50,000 samples. Figure 4 compares the estimation performance of INLA and our MCMC sampler. In general, the absolute bias and standard error in the parameter estimates are comparable in both approaches, especially for fixed effect parameters. The uncertainties in hyperparameter estimates are comparable except for the precision parameter κ_{μ} where the MCMC sampler is more variable. On the other hand, INLA strongly overestimates the precision parameter ($\widehat{\kappa} = 9.207$) of log-Gaussian observations, but again this seems acceptable given that the estimation uncertainty is relatively high due to the small sample size (see Table 1 in the online supplementary material for more details).

We also assess the prediction performance of our MCMC sampler by an out-of-sample (OOS) experiment using 15-fold cross-validation, where we randomly partition the 355 SUs into 15 different sets. We repeat this step to obtain 15 separate combinations, each of which consists of 355 SUs with data at one set of SUs treated as missing, and the other 14 sets of SU data used for training and to predict data at the missing SUs. Figure 1 in the online supplementary material shows the OOS prediction performance of our MCMC sampler; its first row reports the true (left display) and predicted (right display) landslide counts, and by analogy the second row shows true and predicted landslide sizes, all of them being aggregated to the SU scale. We conclude from these plots that our MCMC sampler performs satisfactorily for predicting landslide counts and sizes jointly at unobserved SUs.

5 Data application

5.1 Results for the marked point-process models

We fit the proposed model (2) to the landslides data detailed in Section 2 and jointly model landslide counts and sizes for different choices of mark distribution; see Table 1 for the details of mark

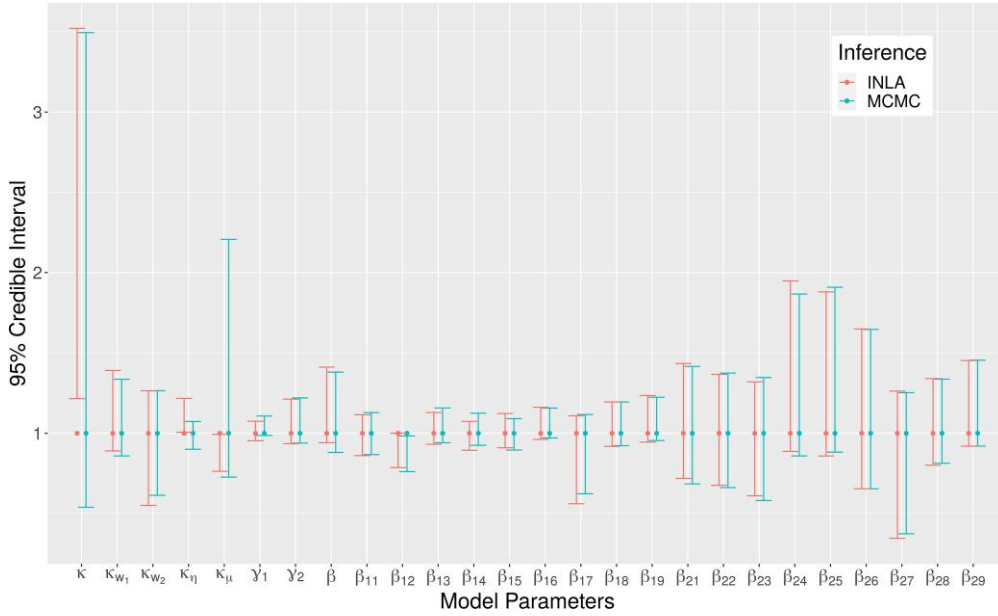

Figure 4. Comparison between posterior estimates based on INLA (first error bar for each parameter) and the MCMC sampler (second error bar for each parameter) in the simulation study. The data were simulated from the model (2). To facilitate numerical comparison and visual display, the error bars are based on the 95% credible intervals divided by the true values of the corresponding model parameters (hyperparameters and covariate coefficients).

distributions. For each mark distribution, we considered eight different sub-models possessing distinct properties while being nested within the general model (2). These models share the same general structure but have different specifications, i.e. whether or not they include the random effects ε_{η} and ε_{μ} in equation (2), and whether or not the sharing parameter β is fixed to zero (base model), corresponding to landslide occurrences and sizes being independent of each other; see Table 2 for details on the different sub-models. The full model presented in equation (2) is denoted as M_1; it includes iid random effects in both latent processes, whereas the base model M_1^0 is similar to model M_1 but with fixed $\beta = 0$. Model M_2 includes the spatially structured random effects in both latent processes but does not have any iid random effects, and model M_2^0 is the same as model M_2 but with fixed $\beta = 0$. We use all available covariates detailed in Section 2.2, along with the total area of SUs that is only included as an additional covariate for landslide sizes.

We fit all candidate models using the MCMC sampler detailed in Section 4.1, generating 100,000 posterior samples for each model. Our proposed MCMC sampler is reasonably fast and takes around 3.5 hr to run four Markov chains in parallel on a Supercomputer with an assignment of a single node and four cores to generate 100k samples for the most general model M_1 with Gam–Gam mark distribution. All summary statistics reported below are based on the last 25,000 MCMC samples after deleting the first 75,000 burn-in samples. The Markov chains (see Section 2.2 in the online supplementary material) converge fairly quickly to a stationary distribution, and appear to mix well for all hyperparameters and latent parameters after about 20k iterations; see Figures 4–7 in the online supplementary material. In particular, Figures 4 and 5 in the online supplementary material show the trace plots for four independent Markov chains with different initial values for model M_1 with Gam–Gam mark distribution in two different validation settings detailed later in this section. Despite their different initial values, all four Markov chains converge to the same stationary distribution for all the model parameters. Furthermore, the effective sample size per minute (ESS/min) is satisfactory for all model hyperparameters and latent parameters; see Table 5 in the online supplementary material. These diagnostics confirm that our MCMC algorithm is accurate and fast enough to infer our proposed model at moderate computational cost. For fitting models M_2, M_3, and M_4, we actually fit model M_1 but with very high fixed precision parameters $\kappa_{\eta} = 1{,}000$ and $\kappa_{\mu} = 1{,}000$ whenever the corresponding independent random effect is

Table 2. Candidate models that differ with respect to the inclusion of independently and identically distributed (iid) random effects (indicated through different subscripts) and to the sharing coefficient being fixed at $\beta = 0$ in the general model (2) (indicated through a superscript of 0)

Model	iid random effect ε_η in $\eta(s)$	iid random effect ε_μ in $\mu(s)$	$\beta = 0$ (base model)
M_1	Yes	Yes	No
M_2	No	No	No
M_3	Yes	No	No
M_4	No	Yes	No
M_1^0	Yes	Yes	Yes
M_2^0	No	No	Yes
M_3^0	Yes	No	Yes
M_4^0	No	Yes	Yes

absent, and their base model counterparts are fitted with additionally fixing the sharing parameter β to zero. Here, fixing $\beta = 0$ simplifies our modelling framework by removing any prior dependence between the count and size processes, such that we could run separate MCMC schemes for these two sub-models. The choice of hyperparameter priors is detailed in Section 4.1.

For comparing the goodness-of-fit and predictive performance of candidate models, we design three types of experiments, providing within-sample (WS) or OOS diagnostics. Out-of-sample analyses are here based either on a 10-fold cross-validation where we remove data to be predicted for entire SUs, or on a fivefold thinning-based approach (Leininger & Gelfand, 2017) where we remove individual landslides from the dataset, such that the training and validation data can contain landslides from the same SU. In the fivefold thinning-based OOS approach, we randomly remove 20% of landslides and their associated sizes in each fold, and predict them in our MCMC algorithm. We determine the fold allocation of each landslide by sampling from a uniform random variable for each landslide and then categorise the landslides as part of fold $i \in \{1, 2, \ldots 5\}$ if the uniform value is in $[(i-1) \times 0.2, i \times 0.2]$. This randomly divides the total number of 933 landslides into five folds of sizes 188, 183, 172, 199, 191. After estimating the point-process intensity on the remaing 4/5 of data, we re-adjust the estimated intensity by multiplying it by 5/4. All models are fitted at the pixel level with random effects defined at SU level, but we report prediction results at the SU level, which is easier to interpret and compare. Results at the SU level are obtained by aggregating the predicted counts and sizes from the pixel level. Subsequently, we use the component-wise mean of $\exp(\boldsymbol{\mu}_{\text{post}})$ and $\exp(\boldsymbol{\eta}_{\text{post}})$ as posterior predictive estimates for the landslide counts and sizes, respectively, where $\boldsymbol{\eta}_{\text{post}}$ and $\boldsymbol{\mu}_{\text{post}}$ denote posterior samples of latent parameters corresponding to counts and sizes, respectively. In WS and 10-fold OOS experiments, we directly impute the sample of $\boldsymbol{\eta}_{\text{post}}$ and $\boldsymbol{\mu}_{\text{post}}$ at the prediction locations, whereas for the thinning-based OOS approach we use the posterior samples of γ_1, γ_2, $\boldsymbol{\beta}_1$, $\boldsymbol{\beta}_2$, $\boldsymbol{W}_1$, and $\boldsymbol{W}_2$ to recreate the posterior samples of $\boldsymbol{\eta}_{\text{post}}$ and $\boldsymbol{\mu}_{\text{post}}$ using the model construction (4). For conciseness, we report only results for 10-fold OOS experiments, whereas the results for the WS and fivefold thinning-based approach are detailed in Section 2.4 of the online supplementary material.

Model M_1 outperforms models M_2, M_3, M_4 and its baseline counterpart; see Tables 6–8 in the online supplementary material for a detailed comparison across all fitted models, and Table 3 for a summary. Table 3 reports the AUC values and mean absolute errors in the 10-fold cross-validation experiment for models M_1, M_1^0, M_2, and M_2^0 with different mark distributions. The AUC metric is a performance measure for binary data, and so here we first threshold observed and predicted landslide counts and sizes to obtain binary information. The thresholds for AUC calculation are set to one for landslide counts, which assesses presence/absence of landslides, while for landslide sizes, it is set to 77.6, the average landslide size (i.e. square root of landslide area). Here the AUC metric puts focus on how well models discriminate between small and moderately large values, whereas absolute errors are more strongly influenced by extreme values due to the heavy tails in observations of counts and sizes. In Table 3, model M_1 with shared spatial effect generally outperforms its corresponding base model M_1^0 when predicting landslide counts, and it has smaller

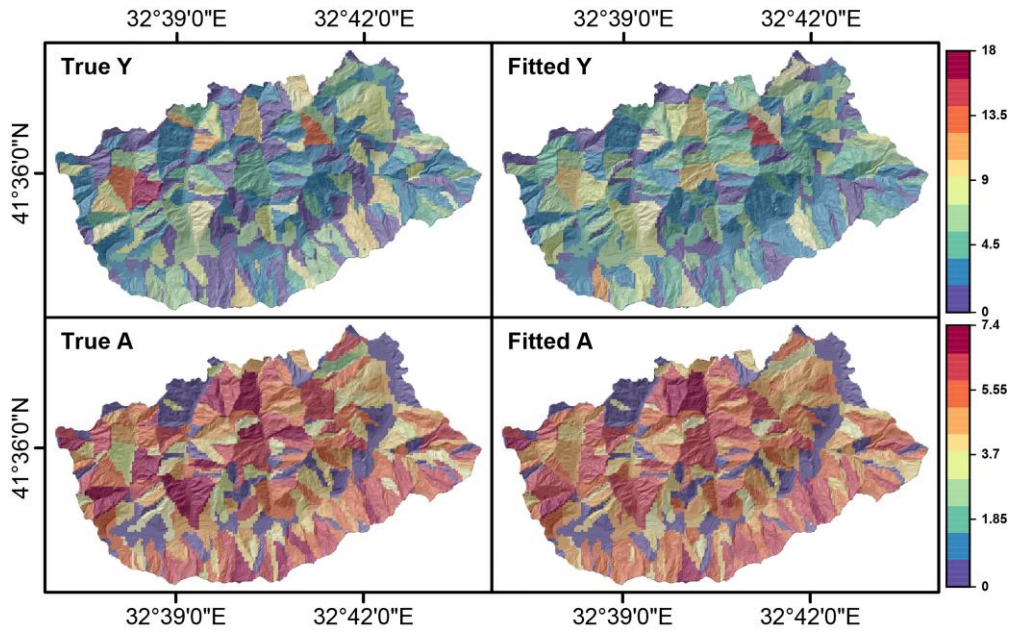

Figure 5. Fitted vs. true landslide counts and sizes for the landslides data detailed in Section 2, where the fitted data are obtained based on model M_1 with Gam–Gam marks using 10-fold cross-validation. First row: true (left) and predicted landslide counts (right). Second row: true (left) and predicted landslides sizes (right), both on log-scale.

absolute errors when predicting landslide sizes using flexible sub-asymptotic mark distributions allowing for positive upper tail index ξ_U. In contrast, model M_1^0 has often smaller absolute errors for landslide sizes when other mark distributions with lighter tails are used, e.g. Gamma, Weibull, log-Gaussian, and generalised-Gamma; however, these mark distributions have higher absolute errors compared to M_1 for predicting landslide counts. When comparing the different models (M_1, M_1^0, M_2, and M_2^0) and different mark distributions listed in Table 1, the overall best performing model is M_1, but there is no clear winner for the mark distribution. With model M_1, the Gamma distribution has the highest AUC value (0.837) for landslide sizes; the log-Gamma distribution has the highest AUC value (0.818) for landslide counts; the ext-GP distribution has the lowest absolute error (128) for landslide sizes; and the Gam–Gam distribution has the lowest absolute error (1.55) for landslide counts. The three sub-asymptotic distributions (ext-GP, Burr, and Gam–Gam) have the first, second, and fourth smallest absolute errors, respectively, for predicting landslide sizes, often much smaller than for other mark distributions; the highest (i.e. worst) absolute error is obtained with the log-Gamma (277) distribution. We also compared the different models (M_1, M_1^0, M_2, M_2^0) and different mark distributions using the continuous ranked probability score (Gneiting & Raftery, 2007) and the deviance information criterion (DIC; Gelman et al., 2014), leading to very similar interpretations; see Table 9 in the online supplementary material. Note that we have here calculated the DIC values for the count and size processes separately, though a more accurate and precise approach would consist in calculating joint DIC values by accounting for the size and count processes jointly; this could be done similarly to Zhang et al. (2017), who have defined the DIC values for the joint modelling of longitudinal and survival data by decomposing the corresponding DIC values. Overall, Model M_1 with Gam–Gam marks consistently provides a very good performance in predicting landslide counts and sizes.

In the WS setting (see Tables 6–8 in the online supplementary material), the simpler models with Gamma, Weibull and log-Gaussian marks possessing relatively light upper tails outperform the more flexible and complex models; however, in the OOS setting, we see the opposite behaviour, and flexible sub-asymptotic mark distributions, such as Gam–Gam, ext-GP and Burr models, outperform the more parsimonious classical models. Based on our cross-validation experiments, we select model M_1 with Gam–Gam mark distribution as the best model overall for jointly predicting

Table 3. Comparison of OOS predictions in the data application reporting area-under-the-curve (AUC) and mean absolute error (AbsError) diagnostics for models M_1, M_1^0, M_2, and M_2^0 with different mark distributions listed in Table 1, obtained for 10-fold cross-validation

Summary	Model	Data	f_{Bu}	f_{Gam}	f_W	f_{lGam}	f_{lGau}	f_{eGP}	f_{GP}	f_{GG}	f_{gGam}
AUC	M_1	Count	0.771	0.814	0.797	0.818	0.797	0.802	0.780	0.786	0.791
		Size	0.792	0.837	0.760	0.632	0.813	0.777	0.790	0.802	0.825
	M_1^0	Count	0.786	0.795	0.795	0.803	0.782	0.786	0.798	0.779	0.786
		Size	0.794	0.775	0.768	0.646	0.798	0.804	0.802	0.821	0.763
	M_2	Count	0.786	0.787	0.787	0.785	0.787	0.786	0.786	0.786	0.787
		Size	0.617	0.618	0.626	0.636	0.715	0.618	0.627	*NA*	0.617
	M_2^0	Count	0.787	0.786	0.786	0.785	0.786	0.787	0.786	0.785	0.787
		Size	0.622	0.622	0.642	0.645	0.809	0.626	0.651	*NA*	0.620
AbsError	M_1	Count	1.59	1.56	1.56	1.91	1.56	1.56	1.60	1.55	1.58
		Size	130	142	147	277	133	128	131	132	149
	M_1^0	Count	1.58	1.56	1.56	1.58	1.57	1.58	1.58	1.55	1.58
		Size	135	132	146	264	131	133	134	132	145
	M_2	Count	16.0	16.2	16.2	16.2	15.7	16.0	16.2	15.7	15.8
		Size	297	292	281	277	178	293	283	303	297
	M_2^0	Count	15.8	15.4	15.4	16.3	16.2	15.8	15.4	15.7	15.8
		Size	289	283	270	268	135	283	259	301	293

Note. Bold blue-coloured digits highlight the best performance among different models and different marks, and light-blue digits show the best performance for a particular choice of mark distribution. Values reported as NA (not available) are cases where AUC cannot be reported due to numerical instabilities arising with predicted values. Higher AUC and lower AbsError are better.

landslide counts and sizes at unobserved slope units. A similar interpretation follows from thinning-based cross-validation; see Table 10 in the online supplementary material. Figure 5 shows the true observed values (left column) and the posterior predictive means (right column) of landslide counts (top row) and landslide sizes (bottom row) when assuming a Gam–Gam mark distribution in the 10-fold cross-validation setting. Results look visually satisfactory, and we conclude from these plots that our proposed model accurately predicts both observed and unobserved landslide counts and sizes at the slope unit level.

Table 4 shows posterior summaries for model M_1 with the Gam–Gam mark distribution in the WS setting. The estimated precision parameters for the ICAR components are $\widehat{\kappa}_{w_1} = 0.783$ and $\widehat{\kappa}_{w_2} = 3.52$. Moreover, estimated precisions for the independent random effects are $\widehat{\kappa}_{\eta} = 2.26$ and $\widehat{\kappa}_{\mu} = 5.63$, which highlights non-negligible unstructured spatial variation in both landslide counts and sizes. The parameter β, which controls the sharing of information between models for counts and sizes, is estimated highly negative, $\widehat{\beta} = -0.444$, which shows that a small number of landslides in a particular region is positively correlated with a bigger landslide size (square root of landslide area) in the same region. The negative estimate of the correlation parameter β is highly significant with the value zero being clearly outside its credible interval. Since the spatial random effect is defined at SU resolution, a relatively higher occurrence intensity of landslides in an SU comes with generally smaller landslide sizes; that is, if there is a relatively large number of landslides in an SU, they tend to be smaller, which makes sense for physical reasons given the finite area of SUs. The estimated upper tail index is $(2/\widehat{c}_2) = 0.243$, which confirms that landslide sizes are quite heavy-tailed. Since the model is here fitted to the square root of landslide areas, the tail index for landslide areas themselves would be about twice as large. As expected, the regression coefficient for the slope steepness, $\widehat{\beta}_{1_{\text{slope}}} = 0.38$, is highly positive for landslide counts and significant (with lower bound of the credible interval far from zero); however, for landslide sizes, it turns out to be nonsignificant. The regression coefficient for VRM is negative $\widehat{\beta}_{1_{\text{VRM}}} = -0.125$, as VRM is related to the solidness of the soil. Most covariates related to the size process seem to

Table 4. Posterior median, standard error (se) and 95% credible interval lower bound (CI^-) and upper bound (CI^+) for model M_1 with Gamma–Gamma mark distribution in the within-sample setting

Parameter	Median	se	CI^-	CI^+	Parameter	Median	se	CI^-	CI^+
c_1	32.5	7.58	20.4	48.3	$\beta_{1_{\text{TPI}}}$	0.029	0.035	− 0.037	0.099
c_2	8.25	1.35	6.63	11.7	$\beta_{1_{\text{LR}}}$	0.255	0.060	0.125	0.356
κ_{w_1}	0.783	0.118	0.593	1.02	$\beta_{1_{\text{s-height}}}$	− 0.035	0.042	− 0.112	0.046
κ_{w_2}	3.52	1.02	2.18	6.09	$\beta_{1_{\text{v-depth}}}$	0.175	0.029	0.119	0.234
$\kappa_{\boldsymbol{\eta}}$	2.26	0.262	1.95	3.03	$\beta_{2_{\text{slope}}}$	− 0.041	0.024	− 0.086	0.012
$\kappa_{\boldsymbol{\mu}}$	5.63	1.11	3.86	8.35	$\beta_{2_{\text{TWI}}}$	− 0.018	0.026	− 0.064	0.036
γ_1	− 3.24	0.035	− 3.30	− 3.16	$\beta_{2_{\text{VRM}}}$	0.001	0.026	− 0.046	0.048
γ_2	4.08	0.038	4.01	4.16	$\beta_{2_{\text{profCurve}}}$	0.012	0.025	− 0.037	0.058
β	− 0.444	0.093	− 0.623	− 0.261	$\beta_{2_{\text{planCurve}}}$	− 0.041	0.028	− 0.101	0.005
$\beta_{1_{\text{slope}}}$	0.380	0.037	0.307	0.461	$\beta_{2_{\text{TPI}}}$	0.036	0.026	− 0.015	0.088
$\beta_{1_{\text{TWI}}}$	0.095	0.038	0.029	0.162	$\beta_{2_{\text{LR}}}$	0.026	0.042	− 0.046	0.111
$\beta_{1_{\text{VRM}}}$	− 0.125	0.037	− 0.206	− 0.067	$\beta_{2_{\text{s-height}}}$	0.009	0.029	− 0.052	0.061
$\beta_{1_{\text{profCurve}}}$	0.058	0.028	− 0.006	0.109	$\beta_{2_{\text{v-depth}}}$	0.038	0.028	− 0.013	0.090
$\beta_{1_{\text{planCurve}}}$	0.006	0.043	− 0.084	0.074	$\beta_{2_{\text{area-SU}}}$	0.010	0.036	− 0.059	0.082

be nonsignificant; this may be due to the rather small number of samples for the landslide size, resulting in quite large uncertainties in parameter estimates.

Finally, to illustrate the practical usefulness of our joint modelling framework for landslide hazard assessment and country planning, Figure 6 shows the mean susceptibility and mean hazard plots across four different main roads in the study region, based on the best model M_1 with Gam–Gam marks. These main roads are created with a buffer of 50 m on each side. This allows assessing the risk associated with landslide events close to roads, which pose a threat of human and economic losses. The susceptibility is here defined as the posterior probability of having at least one landslide, i.e. it is of the form $1 - \exp\{-\exp(\boldsymbol{\eta})\}$ using the Poisson assumption. In contrast, the calculation of landslide hazard includes information from both landslide counts and sizes, and we here define it as the product $\exp(\boldsymbol{\eta}) \times \exp(2\boldsymbol{\mu})$, where $\boldsymbol{\eta}$ and $\boldsymbol{\mu}$ denote the log-intensity and log-median processes associated with landslide counts and sizes, respectively. In other words, the hazard measures the propensity of a given spatial location (conditional on predictors) to experience landslides that are both likely and big. Note that the factor 2 is used to transform the landslide sizes (square root of landslide area) back to the original landslide area scale. Estimates of susceptibility and hazard are obtained by their posterior means, averaging the respective quantities defined above after replacing $\boldsymbol{\eta}$ and $\boldsymbol{\mu}$ with their post-burn-in posterior values from the MCMC output. Note that the correlation between the count and size processes assumed in our marked point process models is key to accurately computing the landslide hazard, as the latter is a function of both. The mean susceptibility plot in the left panel in Figure 6 across four different roads shows that there is a non-zero probability of landslide occurrence close to roads. The two smallest roads appear as the most exposed ones, and the longest roads have the lowest risk of landslide occurrence, certainly because most parts of the longest roads are in the flatter parts of the study region. In the mean hazard plots in the right panel of Figure 6, many similar patterns arise, but sometimes we also discern a different behaviour with opposite trends of susceptibilities and hazards (e.g. for the longest road) because the estimated sharing parameter of latent spatial fields $\widehat{\beta} = -0.444$ is negative. Our joint modelling framework is beneficial in providing estimates of susceptibility and a unified way of calculating hazard-related quantities that depend on both landslide counts and sizes, which is helpful to assess danger levels in specific sub-regions of the domain. We can also compute the aggregated hazard for the whole road network by adding up the pixel hazard over all four roads. The resulting simulation-based posterior predictive quantiles at 0%, 5%, 25%, 50%, 75%, 95%, and 100% are 58,075, 66,114, 78,122, 86,584, 95,462, 111,103, and 135,005, respectively, which

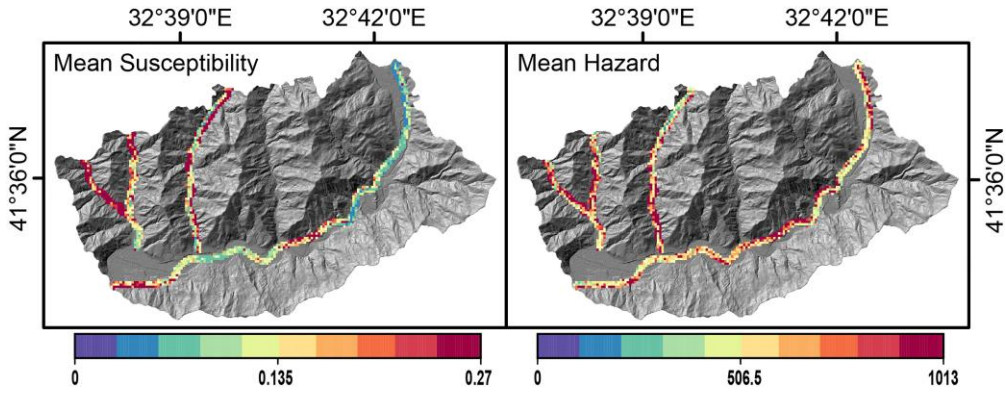

Figure 6. Estimated susceptibility (left) and hazard (right) across the four main roads (coloured) in the study regions based on the model M_1 with Gam–Gam marks.

shows that we can estimate this quantity with moderate uncertainty. Comparing such aggregated hazard quantities for different sub-regions of the study domain could also be helpful for risk mitigation and country planning, as one might only be able to stabilise the most risky slopes with a fixed budget.

5.2 Results for areal data: joint model at SU level

For comparison, we also fit the joint areal models to the data aggregated to the SU scale, as described in Section 3.4. To make inference, we use straightforward modifications of our MCMC sampler from Section 4.1. We generate 100,000 posterior samples, and all posterior inferences are based on the last 25,000 samples after deleting the first 75,000 burn-in samples.

Table 11 in the online supplementary material shows the AUC values and absolute errors for different mark distributions with models M_1, M_1^0, M_2, and M_2^0 in the WS experiment. Model M_1 has higher AUC and lower absolute errors overall for both landslide counts and sizes. Therefore, we select model M_1 with Burr mark distribution as the best model due to its overall better performance for both processes. The Weibull and log-Gaussian mark distributions also have very good and quite similar performance, especially in terms of lowest absolute error for landslide sizes. Nevertheless, by analogy with previous results from Section 5.1, we expect that Weibull and log-Gaussian distributions will perform worse when predicting unobserved landslide data in the OOS setting.

Table 12 in the online supplementary material shows summary statistics for model M_1 with Burr mark distribution. Its tail index for landslide sizes is estimated at $1/(\widehat{c\kappa}) = 0.087$, indicating heavy-tailed landslide sizes. The estimated precision of $\widehat{\kappa}_{w_1} = 0.111$ for the shared ICAR random effect $\boldsymbol{W}_1$ is rather small, which implies that the shared spatial random effect jointly explains a large part of the variability in both processes, with a highly positive sharing parameter $\widehat{\beta} = 1.44$. It therefore has positive sign opposite to the sharing coefficient for the models fitted at the pixel-level resolution in Section 5.1: locally, occurrence of several landslides typically corresponds to relatively small landslides, whereas a strong multiplicity of landslides at SU level tends to go hand in hand with larger aggregated sizes. Therefore, even if landslides tend to be smaller if there are relatively many of them in an SU, as found in the pixel-based point-process model, their aggregated size still tends to be larger overall. For instance, the size of a single large landslide in a SU will tend to be smaller than the aggregated sizes of several, relatively smaller landslides in an SU with comparable geomorphological features. Several of the covariate coefficients are highly significant, e.g. slope for counts, planCurve for counts and sizes, s-height for counts and sizes, and LR for counts and sizes. The variability of estimated covariate coefficients is often larger than for the pixel-level model, such that several covariates remain nonsignificant. This appears to be a disadvantage of modelling aggregated data directly at the SU level compared to the marked point-process models in Section 3.2, for which data and

covariate information is fully exploited at the pixel level for individual landslides rather than at the SU level.

6 Summary and conclusions

We proposed a joint spatial modelling framework allowing for stochastic interactions between landslide counts and sizes. While counts are (marginally) modelled using an LGCP, the sizes are treated as a 'mark' in a marked point-process framework, and various types of sub-asymptotic mark distributions justified by extreme-value theory are used. The high flexibility of our new models suggests that they can be more generally used to jointly model count and size processes in other contexts, e.g. for wildfire counts and burnt areas, where spatio-temporal extensions could be developed. Although LGCPs are based on the Poisson distribution for counts, the log-intensity is modelled as a multi-resolution Gaussian process, which allows us to capture both overdispersion and spatial dependence. This, combined with flexible mark distributions, provides a rich enough framework for modelling landslide counts and sizes in this case study (see also Opitz et al., 2022). More specifically, our joint model can be expressed through a hierarchical construction with Gaussian ICAR priors to capture spatial dependence at the latent level. The spatially structured interaction between the landslide count and size processes is measured by a shared ICAR prior with sharing coefficient estimated from the data. The ICAR prior is used efficiently at the SU level, and additionally, the sparse structure of ICAR priors reduces the computation burden significantly, thus opening the door to fit our models in very high spatial dimensions. Although we have used a simple linear structure for the fixed effects, a more general non-linear form such as splines may be easily adopted into our modelling framework, if required by the application.

We fitted our models using a customised MCMC method, whereby the high-dimensional latent variables $\boldsymbol{\eta}$ and $\boldsymbol{\mu}$ are simulated efficiently using the MALA. The great benefit of our MCMC implementation is that various types of mark distributions and model structures can be easily incorporated and tested, unlike other techniques such as INLA. In future research, it would be interesting to implement some of these flexible mark distributions in INLA, or to explore alternative approximate Bayesian techniques that are both fast and accurate for a wider range of models, such as the recently published Max-and-Smooth approach (see Hazra et al., 2023; Hrafnkelsson et al., 2021; Johannesson et al., 2022). Alternatively, it would also be fascinating to investigate using amortised neural Bayes estimation approaches that could provide further speed-ups thanks to advanced deep learning methods (Radev et al., 2022; Sainsbury-Dale et al., 2023).

To study a landslide inventory from Turkey, we compared and fitted models with structural differences in the bivariate latent process and the mark distributions. Models with shared random effects and sub-asymptotic mark distributions performed especially well for predicting unobserved data at the SU scale, compared to baseline models (with independent count and size processes) or models using classical mark distributions. In future research, it would be interesting to use our models in even higher spatial dimensions and to apply them to other point patterns available for very large observation windows possibly exploiting different sparse precision matrix formulations, e.g. the so-called SPDE approach (Lindgren et al., 2011) or nearest-neighbour Gaussian processes (Datta et al., 2016). Although ICAR priors perform well in our application, another future research direction includes exploring alternative spatial priors for areal data and extensions thereof, including variants of conditional (CAR) and simultaneous (SAR) autoregressive models (Ver Hoef et al., 2018), as well as the Besag–York–Mollié (BYM) model, the Leroux model, or the Dean model; see Riebler et al. (2016) and Morris et al. (2019) for more details about these variants.

Finally, another research avenue is towards spatio-temporal extensions for landslide counts and sizes or other environmental and ecological marked point patterns, and specifically to dynamic models with self-inflating or self-deflating effects of occurrences on the point-process intensity.

Acknowledgments

The authors are grateful to the reviewers and the editors for their helpful comments and suggestions that improved the quality of the manuscript.

Funding

This publication is based on the work supported by the King Abdullah University of Science and Technology (KAUST) Office of Sponsored Research (OSR) under Award No. OSR-CRG2020-4338.

Data availability

The data underlying this article cannot be shared publicly due to confidentiality reasons. The data may, however, be shared upon reasonable request to the corresponding author.

Conflict of interests: None declared.

Supplementary material

Supplementary material is available online at *Journal of the Royal Statistical Society: Series C*. The *R* code and Supplementary Material for this article are available from the GitHub repository of the first author, namely https://github.com/yadavrishikesh/Joint-Model-with_ICAR_Prior. The code replicates a simulation experiment to generate data from our model and then perform inference using our proposed Bayesian inferential technique (on the simulated data).

References

Amato G., Eisank C., Castro-Camilo D., & Lombardo L. (2019). Accounting for covariate distributions in slope-unit-based landslide susceptibility models. A case study in the alpine environment. *Engineering Geology*, *260*, 105237. https://doi.org/10.1016/j.enggeo.2019.105237

Banerjee S., Carlin B. P., & Gelfand A. E. (2014). *Hierarchical modeling and analysis for spatial data* (2nd ed.). CRC Press.

Bernardinelli L., Clayton D., & Montomoli C. (1995). Bayesian estimates of disease maps: How important are priors? *Statistics in Medicine*, *14*(21–22), 2411–2431. https://doi.org/10.1002/(ISSN)1097-0258

Besag J. (1975). Statistical analysis of non-lattice data. *Journal of the Royal Statistical Society: Series D (The Statistician)*, *24*(3), 179–195. http://dx.doi.org/10.2307/2987782

Besag J., & Kooperberg C. (1995). On conditional and intrinsic autoregressions. *Biometrika*, *82*(4), 733–746. http://dx.doi.org/10.1093/biomet/82.4.733

Brenning A. (2005). Spatial prediction models for landslide hazards: Review, comparison and evaluation. *Natural Hazards and Earth System Sciences*, *5*(6), 853–862. https://doi.org/10.5194/nhess-5-853-2005

Broeckx J., Rossi M., Lijnen K., Campforts B., Poesen J., & Vanmaercke M. (2020). Landslide mobilization rates: A global analysis and model. *Earth-Science Reviews*, *201*, 102972. https://doi.org/10.1016/j.earscirev.2019.102972

Can T., Nefeslioglu H. A., Gokceoglu C., Sonmez H., & Duman T. Y. (2005). Susceptibility assessments of shallow earthflows triggered by heavy rainfall at three catchments by logistic regression analyses. *Geomorphology*, *72*(1–4), 250–271. https://doi.org/10.1016/j.geomorph.2005.05.011

Casella G., & George E. I. (1992). Explaining the Gibbs sampler. *The American Statistician*, *46*(3), 167–174. http://dx.doi.org/10.2307/2685208

Chen W., Pourghasemi H. R., Kornejady A., & Zhang N. (2017). Landslide spatial modeling: Introducing new ensembles of ANN, MaxEnt, and SVM machine learning techniques. *Geoderma*, *305*, 314–327. https://doi.org/10.1016/j.geoderma.2017.06.020

Cisneros D., Gong Y., Yadav R., Hazra A., & Huser R. (2023). A combined statistical and machine learning approach for spatial prediction of extreme wildfire frequencies and sizes. *Extremes*, *26*(2), 301–330. https://doi.org/10.1007/s10687-022-00460-8

Cressie N. A. C. (1993). *Statistics for spatial data*. Wiley Online Library.

Daniell J. E., Schaefer A. M., & Wenzel F. (2017). Losses associated with secondary effects in earthquakes. *Frontiers in Built Environment*, *3*, 30. https://doi.org/10.3389/fbuil.2017.00030

Datta A., Banerjee S., Finley A. O., & Gelfand A. E. (2016). Hierarchical nearest-neighbor Gaussian process models for large geostatistical datasets. *Journal of the American Statistical Association*, *111*(514), 800–812. https://doi.org/10.1080/01621459.2015.1044091

Davison A. C., & Smith R. L. (1990). Models for exceedances over high thresholds (with discussion). *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, *52*(3), 393–442. https://doi.org/10.1111/j.2517-6161.1990.tb01796.x

De Reu J., Bourgeois J., Bats M., Zwertvaegher A., Gelorini V., De Smedt P., Chu W., Antrop M., De Maeyer P., Finke P., Van Meirvenne M., Verniers J., & Crombé P. (2013). Application of the topographic position index to heterogeneous landscapes. *Geomorphology*, *186*, 39–49. https://doi.org/10.1016/j.geomorph.2012.12.015

Diggle P. J., Menezes R., & Su T.-l. (2010). Geostatistical inference under preferential sampling. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, *59*(2), 191–232. http://dx.doi.org/10.1111/j.1467-9876.2009.00701.x

Engelke S., Opitz T., & Wadsworth J. (2019). Extremal dependence of random scale constructions. *Extremes*, *22*(4), 623–666. https://doi.org/10.1007/s10687-019-00353-3

Evans I. S. (2019). General geomorphometry, derivatives of altitude, and descriptive statistics. In R. J. Chorley (Ed.), *Spatial analysis in geomorphology* (pp. 17–90). Routledge.

Gao S. (2004). A shared random effect parameter approach for longitudinal dementia data with non-ignorable missing data. *Statistics in Medicine*, *23*(2), 211–219. https://doi.org/10.1002/(ISSN)1097-0258

Gelman A., Hwang J., & Vehtari A. (2014). Understanding predictive information criteria for Bayesian models. *Statistics and Computing*, *24*(6), 997–1016. https://doi.org/10.1007/s11222-013-9416-2

Gneiting T., & Raftery A. E. (2007). Strictly proper scoring rules, prediction, and estimation. *Journal of the American Statistical Association*, *102*(477), 359–378. https://doi.org/10.1198/016214506000001437

Görüm T. (2019). Landslide recognition and mapping in a mixed forest environment from airborne LiDAR data. *Engineering Geology*, *258*, 105155. https://doi.org/10.1016/j.enggeo.2019.105155

Guo C.-w., Huang Y.-d., Yao L.-k., & Alradi H. (2017). Size and spatial distribution of landslides induced by the 2015 Gorkha earthquake in the Bhote Koshi river watershed. *Journal of Mountain Science*, *14*(10), 1938–1950. https://doi.org/10.1007/s11629-016-4140-y

Guzzetti F., Carrara A., Cardinali M., & Reichenbach P. (1999). Landslide hazard evaluation: A review of current techniques and their application in a multi-scale study, Central Italy. *Geomorphology*, *31*(1–4), 181–216. https://doi.org/10.1016/S0169-555X(99)00078-1

Hazra A., Huser R., & Jóhannesson Á. V. (2023). Bayesian latent Gaussian models for high-dimensional spatial extremes. In B. Hrafnkelsson (Ed.), *Statistical modeling using latent Gaussian models—With applications in geophysics and environmental sciences*. Springer.

Heerdegen R. G., & Beran M. A. (1982). Quantifying source areas through land surface curvature and shape. *Journal of Hydrology*, *57*(3–4), 359–373. https://doi.org/10.1016/0022-1694(82)90155-X

Ho L. P., & Stoyan D. (2008). Modelling marked point patterns by intensity-marked Cox processes. *Statistics & Probability Letters*, *78*(10), 1194–1199. https://doi.org/10.1016/j.spl.2007.11.013

Hrafnkelsson B., Siegert S., Huser R., Bakka H., & Johannesson Á. V. (2021). Max-and-smooth: A two-step approach for approximate Bayesian inference in latent Gaussian models. *Bayesian Analysis*, *16*(2), 611–638. https://doi.org/10.1214/20-BA1219

Jiao J., Hu G., & Yan J. (2021). A Bayesian marked spatial point processes model for basketball shot chart. *Journal of Quantitative Analysis in Sports*, *17*(2), 77–90. https://doi.org/10.1515/jqas-2019-0106

Johannesson Á. V., Siegert S., Huser R., Bakka H., & Hrafnkelsson B. (2022). Approximate Bayesian inference for analysis of spatio-temporal flood frequency data. *Annals of Applied Statistics*, *16*(2), 905–935. https://doi.org/10.1214/21-AOAS1525

Kennedy I. T., Petley D. N., Williams R., & Murray V. (2015). A systematic review of the health impacts of mass earth movements (landslides). *PLoS Currents*, *7*. https://doi.org/10.1371/currents.dis.1d49e84c8bbe678b0e70cf7fc35d0b77

Kirkby M., & Beven K. (1979). A physically based, variable contributing area model of basin hydrology. *Hydrological Sciences Journal*, *24*(1), 43–69. https://doi.org/10.1080/02626667909491834

Koh J., Pimont F., Dupuy J.-L., & Opitz T. (2023). Spatiotemporal wildfire modeling through point processes with moderate and extreme marks. *Annals of Applied Statistics*, *17*(1), 560–82. https://doi.org/10.1214/22-AOAS1642

Leininger T. J., & Gelfand A. E. (2017). Bayesian inference and model assessment for spatial point patterns using posterior predictive samples. *Bayesian Analysis*, *12*(1), 1–30. https://doi.org/10.1214/15-BA985

Lindgren F., Rue H., & Lindström J. (2011). An explicit link between Gaussian fields and Gaussian Markov random fields: The stochastic partial differential equation approach. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, *73*(4), 423–498. https://doi.org/10.1111/j.1467-9868.2011.00777.x

Lóczy D., Pirkhoffer E., & Gyenizse P. (2012). Geomorphometric floodplain classification in a hill region of Hungary. *Geomorphology*, *147*, 61–72. https://doi.org/10.1016/j.geomorph.2011.06.040

Lombardo L., Bakka H., Tanyas H., van Westen C., Mai P. M., & Huser R. (2019). Geostatistical modeling to capture seismic-shaking patterns from earthquake-induced landslides. *Journal of Geophysical Research: Earth Surface*, *124*(7), 1958–1980. https://doi.org/10.1029/2019JF005056

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

Lombardo L., Opitz T., & Huser R. (2018). Point process-based modeling of multiple debris flow landslides using INLA: An application to the 2009 Messina disaster. *Stochastic Environmental Research and Risk Assessment*, *32*(7), 2179–2198. https://doi.org/10.1007/s00477-018-1518-0

Lombardo L., Opitz T., & Huser R. (2019). Numerical recipes for landslide spatial prediction using R-INLA: A step-by-step tutorial. In H. R. Pourghasemi, & C. Gokceoglu (Eds.), *Spatial modeling in GIS and R for earth and environmental sciences* (pp. 55–83). Elsevier.

Lombardo L., Tanyas H., Huser R., Guzzetti F., & Castro-Camilo D. (2021). Landslide size matters: A new data-driven, spatial prototype. *Engineering Geology*, *293*, 106288. https://doi.org/10.1016/j.enggeo.2021.106288

Metropolis N., Rosenbluth A. W., Rosenbluth M. N., Teller A. H., & Teller E. (1953). Equation of state calculations by fast computing machines. *Journal of Chemical Physics*, *21*(6), 1087–1092. https://doi.org/10.1063/1.1699114

Møller J., Syversveen A. R., & Waagepetersen R. P. (1998). Log Gaussian Cox processes. *Scandinavian Journal of Statistics*, *25*(3), 451–482. https://doi.org/10.1111/sjos.1998.25.issue-3

Morris M., Wheeler-Martin K., Simpson D., Mooney S. J., Gelman A., & DiMaggio C. (2019). Bayesian hierarchical spatial models: Implementing the Besag York Mollié model in Stan. *Spatial and Spatio-Temporal Epidemiology*, *31*, 100301. https://doi.org/10.1016/j.sste.2019.100301

Naveau P., Huser R., Ribereau P., & Hannart A. (2016). Modeling jointly low, moderate, and heavy rainfall intensities without a threshold selection. *Water Resources Research*, *52*(4), 2753–2769. https://doi.org/10.1002/2015WR018552

Opitz T., Bakka H., Huser R., & Lombardo L. (2022). High-resolution Bayesian mapping of landslide hazard with unobserved trigger event. *Annals of Applied Statistics*, *16*(3), 1653–1675. https://doi.org/10.1214/21-AOAS1561

Papastathopoulos I., & Tawn J. A. (2013). Extended generalised Pareto models for tail estimation. *Journal of Statistical Planning and Inference*, *143*(1), 131–143. https://doi.org/10.1016/j.jspi.2012.07.001

Patel L., Shand L., Tucker J. D., & Huerta G. (2021). 'Spatio-temporal extreme event modeling of terror insurgencies', arXiv, arXiv:2110.08363, preprint: not peer reviewed.

Pimont F., Fargeon H., Opitz T., Ruffault J., Barbero R., Martin-StPaul N., Rigolot E., Rivière M., & Dupuy J.-L. (2021). Prediction of regional wildfire activity in the probabilistic Bayesian framework of Firelihood. *Ecological Applications*, *31*(5), e02316. https://doi.org/10.1002/eap.2316

Radev S. T., Mertens U. K., Voss A., Ardizzone L., & Kȟe U. (2022). BayesFlow: Learning complex stochastic models with invertible neural networks. *IEEE Transactions on Neural Networks and Learning System*, *33*(4), 1452–1466. https://doi.org/10.1109/TNNLS.2020.3042395

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

Riebler A., Sørbye S. H., Simpson D., & Rue H. (2016). An intuitive Bayesian spatial model for disease mapping that accounts for scaling. *Statistical Methods in Medical Research*, *25*(4), 1145–1165. https://doi.org/10.1177/0962280216660421

Roback K., Clark M. K., West A. J., Zekkos D., Li G., Gallen S. F., Chamlagain D., & Godt J. W. (2018). The size, distribution, and mobility of landslides caused by the 2015 Mw7. 8 Gorkha earthquake, Nepal. *Geomorphology*, *301*, 121–138. https://doi.org/10.1016/j.geomorph.2017.01.030

Rue H., Martino S., & Chopin N. (2009). Approximate Bayesian inference for latent Gaussian models by using integrated nested Laplace approximations. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, *71*(2), 319–392. https://doi.org/10.1111/j.1467-9868.2008.00700.x

Sainsbury-Dale M., Zammit-Mangion A., & Huser R. (2023). Likelihood-free parameter estimation with neural Bayes estimators. *The American Statistician*. https://doi.org/10.1080/00031305.2023.2249522

Sappington J. M., Longshore K. M., & Thompson D. B. (2007). Quantifying landscape ruggedness for animal habitat analysis: A case study using bighorn sheep in the Mojave desert. *The Journal of Wildlife Management*, *71*(5), 1419–1426. https://doi.org/10.2193/2005-723

Stein M. L. (2021a). A parametric model for distributions with flexible behavior in both tails. *Environmetrics*, *32*(2), e2658. https://doi.org/10.1002/env.2658

Stein M. L. (2021b). Parametric models for distributions when interest is in extremes with an application to daily temperature. *Extremes*, *24*(2), 293–323. https://doi.org/10.1007/s10687-020-00378-z

Stepinski T. F., & Jasiewicz J. (2011). Geomorphons—A new approach to classification of landforms. *Proceedings of Geomorphometry*, *2011*, 109–112.

Tanyaş H., van Westen C. J., Allstadt K. E., & Jibson R. W. (2019). Factors controlling landslide frequency–area distributions. *Earth Surface Processes and Landforms*, *44*(4), 900–917. https://doi.org/10.1002/esp.4543

Taylor F. E., Malamud B. D., Witt A., & Guzzetti F. (2018). Landslide shape, ellipticity and length-to-width ratios. *Earth Surface Processes and Landforms*, *43*(15), 3164–3189. https://doi.org/10.1002/esp.v43.15

Titti G., Napoli G. N., Conoscenti C., & Lombardo L. (2022). Cloud-based interactive susceptibility modeling of gully erosion in Google Earth Engine. *International Journal of Applied Earth Observation and Geoinformation*, *115*, 103089. https://doi.org/10.1016/j.jag.2022.103089

Titti G., Sarretta A., Lombardo L., Crema S., Pasuto A., & Borgatti L. (2022). Mapping susceptibility with open-source tools: A new plugin for QGIS. *Frontiers in Earth Science*, *10*, 842425. https://doi.org/10.3389/feart.2022.842425

Valagussa A., Marc O., Frattini P., & Crosta G. (2019). Seismic and geological controls on earthquake-induced landslide size. *Earth and Planetary Science Letters*, *506*, 268–281. https://doi.org/10.1016/j.epsl.2018.11.005

Vanani A. A. G., Shoaei G., & Zare M. (2021). Statistical analyses of landslide size and spatial distribution triggered by 1990 Rudbar-Manjil (Mw 7.3) earthquake, northern Iran: Revised inventory, and controlling factors. *Bulletin of Engineering Geology and the Environment*, *80*(4), 3381–3403. https://doi.org/10.1007/s10064-021-02106-8

Ver Hoef J. M., Hanks E. M., & Hooten M. B. (2018). On the relationship between conditional (CAR) and simultaneous (SAR) autoregressive models. *Spatial Statistics*, *25*, 68–85. https://doi.org/10.1016/j.spasta.2018.04.006

Yadav R., Huser R., & Opitz T. (2021). Spatial hierarchical modeling of threshold exceedances using rate mixtures. *Environmetrics*, *32*(3), e2662. https://doi.org/10.1002/env.2662

Yadav R., Huser R., & Opitz T. (2022). A flexible Bayesian hierarchical modeling framework for spatially dependent peaks-over-threshold data. *Spatial Statistics*, *51*, 100672. https://doi.org/10.1016/j.spasta.2022.100672

Zevenbergen L. W., & Thorne C. R. (1987). Quantitative analysis of land surface topography. *Earth Surface Processes and Landforms*, *12*(1), 47–56. https://doi.org/10.1002/(ISSN)1096-9837

Zhang D., Chen M.-H., Ibrahim J. G., Boye M. E., & Shen W. (2017). Bayesian model assessment in joint modeling of longitudinal and survival data with applications to cancer clinical trials. *Journal of Computational and Graphical Statistics*, *26*(1), 121–133. https://doi.org/10.1080/10618600.2015.1117472