



# Unified landslide hazard assessment using hurdle models: a case study in the Island of Dominica

Erin Bryce<sup>1</sup> · Luigi Lombardo<sup>2</sup> · Cees van Westen<sup>2</sup> · Hakan Tanyas<sup>2</sup> · Daniela Castro-Camilo<sup>1</sup>

Accepted: 13 April 2022 / Published online: 2 June 2022  
© The Author(s) 2022

## Abstract

Climatically-induced natural hazards are a threat to communities. They can cause life losses and heavy damage to infrastructure, and due to climate change, they have become increasingly frequent. This is especially the case in tropical regions, where major hurricanes have consistently appeared in recent history. Such events induce damage due to the high wind speed they carry, and the high intensity/duration of rainfall they discharge can further induce a chain of hydro-morphological hazards in the form of widespread debris slides/flows. The way the scientific community has developed preparatory steps to mitigate the potential damage of these hydro-morphological threats includes assessing where they are likely to manifest across a given landscape. This concept is referred to as susceptibility, and it is commonly achieved by implementing binary classifiers to estimate probabilities of landslide occurrences. However, predicting where landslides can occur may not be sufficient information, for it fails to convey how large landslides may be. This work proposes using a flexible Bernoulli-log-Gaussian hurdle model to simultaneously model landslide occurrence and size per areal unit. Covariate and spatial information are introduced using a generalised additive modelling framework. To cope with the high spatial resolution of the data, our model uses a Markovian representation of the Matérn covariance function based on the stochastic partial differential equation approach. Assuming Gaussian priors, our model can be integrated into the class of latent Gaussian models, for which inference is conveniently performed based on the integrated nested Laplace approximation method. We use our modelling approach in Dominica, where hurricane Maria (September 2017) induced thousands of shallow flow-like landslides passing over the island. Our results show that we can not only estimate where landslides may occur and how large they may be, but we can also combine this information in a unified landslide hazard model, which is the first of its kind.

**Keywords** Bayesian spatial modelling · Integrated nested Laplace approximation (INLA) · Landslide hazard · Landslide area prediction · Slope unit · Stochastic partial differential equation (SPDE)

## 1 Introduction

Dominica is situated within the Atlantic hurricane belt and, as such, is highly vulnerable to high-intensity weather events, as evidenced by its long extreme-events history, stretching as far back as the Great hurricane of 1780. In

addition, the island's infrastructure and populations are concentrated along the coastal areas, particularly in the south and west, where their situation and building conditions hardly help withstand one extreme natural hazard after another. In this paper, we focus on the aftermath of hurricane Maria, which originated from a wave leaving the west coast of Africa on 12th September 2017. The wave moved westward across the Atlantic, creating deep convection, and consolidated into a tropical depression 580 nautical miles east of Barbados on the 16th September (Pasch et al. 2018; Fobert et al. 2021). By the evening of the 18th, Maria had intensified and hit the Island of Dominica twelve hours later as a category five hurricane (Pasch et al. 2018), causing a total of 9960 landslides

✉ Erin Bryce  
2313549b@student.gla.ac.uk

<sup>1</sup> School of Mathematics and Statistics, University of Glasgow, Glasgow, UK

<sup>2</sup> Department of Earth System Analysis, Faculty of Geo-Information Science and Earth Observation (ITC), University of Twente, PO Box 217, AE 7500 Enschede, The Netherlands

on the island, 8,576 of which were classified as debris slides (Van Westen and Zhang 2018). The infrastructure and transport sectors were the main avenues of the USD 930.9 million in damages incurred by Maria, with the agriculture and tourism sectors not far behind (ACAPS 2018). This prompted a plea by Dominica's prime minister for the funds to make Dominica a fully climate-resilient nation, determined to protect their island (Gibbens 2019).

The Government of the Commonwealth of Dominica, in conjunction with the Caribbean Disaster Emergency Management Agency and the Caribbean Development Bank, commissioned a post-disaster needs assessment from hurricane Maria in order to estimate the total damage, the damages per sector, and to identify recovery needs (ACAPS 2018). The reconstruction of the destroyed infrastructure was funded by a loan from the World Bank and was a part of the project "Enhancing Resilient Reconstruction in Dominica". The project promoted the idea that Dominica could limit the damage from natural hazards by improving the uptake of resilient building practices, aiming to accelerate short-term recovery and strengthen long-term resilience to climate-related risks. Landslide hazard assessment is critically valuable to this programme. It can help define land-use capability, detect areas where intervention is needed to stabilise slopes, and identify appropriate mitigation measures.

To help tackle the above challenges, this article builds a joint probability model to map the Island of Dominica by susceptibility and size of landslides. Our model is able to detect unstable areas that could potentially host large landslides, which is key to identifying regions in need of hazard mitigation. The model belongs to the class of hurdle models, i.e., it is a two-part model that specifies a Bernoulli likelihood for landslide occurrence and a Gaussian likelihood for landslide log-sizes given a positive occurrence. The observations (presence/absence for the Bernoulli likelihood and log-sizes for the Gaussian likelihood) are assumed to be conditionally independent given a Gaussian latent process that drives the trends, dependencies and non-stationary patterns observed in the data. The latent process is characterised by a linear predictor with a flexible additive structure allowing the incorporation of covariates and spatial effects. Specifically, we rely on generalised additive models (GAMs) to flexibly model the covariates' influence using fixed and random effects, which are also known as linear and non-linear effects in terms of their influence. Spatial dependence between locations is characterised using a Gaussian Process with Matérn covariance structure. The Matérn family of covariance functions is widely used in spatial statistics due to its flexible local behaviour and interesting theoretical properties (Guttorp and Gneiting 2006; Stein 2012). To allow for fast inference, we use the stochastic partial differential equation (SPDE; Lindgren

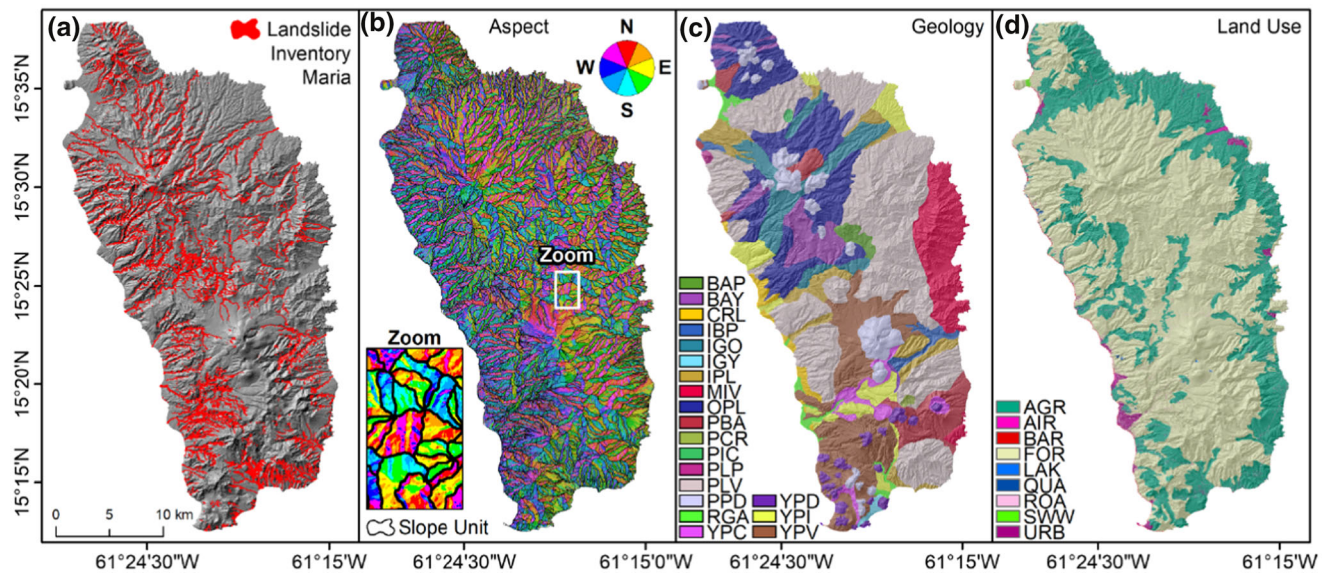
et al. 2011) approach that provides accurate Markovian representations of the Matérn covariance. Under a Bayesian framework, we assume relatively weak but highly interpretable Gaussian priors for all the model components and hyperparameters involved and fit our model using the integrated nested Laplace approximation (INLA; Rue et al. 2009). INLA uses the Laplace approximation to compute posterior distributions of interest. It is developed for the vast class of latent Gaussian models and implemented in the R-INLA library (Bivand et al. 2015). For related contributions, see Opitz et al. (2018), Castro-Camilo et al. (2019, 2021) and Lombardo et al. (2021).

Throughout this paper, we will interchangeably refer to susceptibility or probability of occurrence (see Lombardo et al. 2020; Titti et al. 2021). As for the notion of hazard, this further extends the concept of susceptibility by looking at measurable quantities of hazardous processes, such as size and frequency (Guzzetti et al. 1999). However, the literature on statistical applications in landslide science has never presented a model where the susceptibility is combined with the expected size of landslides, which we propose here for the first time.

The remainder of the paper is organised as follows. Section 2 introduces the variables considered for modelling. Section 3 outlines our methodology, describing the latent Gaussian model framework and the two likelihoods that constitute our hurdle model. We also provide details on prior distributions and inference using INLA. In Sect. 4 we summarise our results. Conclusions and a critical discussion are provided in Sect. 5.

## 2 Data description

After hurricane Maria, the Global Facility for Disaster Reduction and Recovery sponsored the CHARIM (Caribbean Handbook on Risk Information Management; [link here](#)) project, intending to build a comprehensive natural hazard assessment. As part of this assessment, a team from the University of Twente (NL) mapped a large-scale landslide inventory for the whole island (see Fig. 1a), primarily using five scenes of Pléiades satellite images with a resolution of 0.5m, dated on the 23rd September and 5th October 2017, made available through UNITAR-UNOSAT (<https://www.unitar.org/>). Additionally, a series of Digital Globe Images were used after they were gathered for the Google Crisis Response. All the images were visually inspected by image interpretation experts. As a result, landslides were mapped as polygons, separating scarp, transport and accumulation areas, and classifying the landslides according to their types.



**Fig. 1** Summary of the data used in this work. Panel **a** shows the landslide inventory over the shaded terrain. Panel **b** shows the SU delineation over the aspect map. The area marked as “Zoom” in panel

**b** highlights the consistency of SU boundaries with changes in aspect. Panels **c** and **d** report the geological and land-use maps, respectively. Acronyms definition for panels **c** and **d** can be found in Table 3

The inventory features 9960 landslides, with a cumulative planimetric area of 11.4 km<sup>2</sup>, covering 1.5 % of the Island of Dominica. To aggregate the landslide information over space, we opted for a Slope Unit partition of Dominica (see Fig. 1b). A Slope Unit (SU, hereafter) encompasses the geographic space between ridges and streamlines. In other words, they are half-basins of a given order (Amato et al. 2019) extracted by maximising areas with homogeneous aspects and are particularly suited to support landslide models for they approximate the morphodynamic behaviour of landslides well (Carrara et al. 1991). To partition the Island of Dominica into SUs, we initially used *r.slopeunits*, an open-source software developed by Alvioli et al. (2016). We have later refined them through manual editing in GIS, to obtain a total of 3335 SUs.

We use a two-step procedure to build our hurdle model at the SU scale. First, we assign to each SU the sum of all planimetric landslide areas intersecting the SU itself. Then, to build a binary dataset, we assign a landslide presence to SUs with positive landslide aggregated areas. Conversely, we assign a landslide absence to SUs unaffected by slope failures. Finally, to build a dataset of landslide sizes per SU, we extract SUs with positive landslide aggregated areas.

Covariates available for analysis are detailed in Table 1 (geographical) and Table 3 (geological) in the Appendix. They are a mixture of geographical and geological characteristics and constitute the morphology of Dominica’s landscape. All the geographical variables are continuous, while the geological and land-use types are represented as

proportions of each SU; for example, a SU could be 50% forest, 20% bare and 30% quarry.

### 3 Methods

As mentioned in Sect. 1, we aim to build a hurdle model to detect areas of higher landslide susceptibility and sizes. Our approach models the probability of observing a landslide in a SU using a Bernoulli distribution. Given that a landslide was observed, it also describes landslide log-sizes using a Gaussian distribution. We stress again that we assume that the observations (presence/absence for the Bernoulli likelihood and log-sizes for the Gaussian likelihood) are conditionally independent given a latent Gaussian structure that drives the trends, dependencies and non-stationary patterns observed in the data. Both parts of our hurdle model belong to the class of latent Gaussian models.

This section describes the general framework of latent Gaussian models and the two likelihood models. We also provide details regarding prior specification and inference based on INLA.

#### 3.1 Latent Gaussian models

Gaussian Latent Models (GLMs) are a vast and flexible class of models that are well suited for modelling spatial data (Rue et al. 2017). They admit a hierarchical representation, where observations are assumed to be conditionally independent given a latent field and a set of hyperparameters. Specifically, let **y** be the vector of

**Table 1** Summary of the Dominica dataset, including the responses and initial covariates' set

Variable	Acronym	Type	Units
Landslide size	$A_L(s)$	Continuous response	$m^2$
Landslide occurrence	$O_L(s)$	Binary response	0 = No, 1 = Yes
SU area	$SU_A$	Continuous explanatory	$m^2$
SU perimeter	$SU_P$	Continuous explanatory	m
$SU_P/SU_A$	$SU_{P/A}$	Continuous explanatory	1/m
$SU_P/\sqrt{SU_A}$	$SU_{P/\sqrt{A}}$	Continuous explanatory	Unit-less
SU maximum distance	$SU_{MD}$	Continuous explanatory	m
SU maximum distance to area ratio	$SU_{MD/A}$	Continuous explanatory	1/m
$SU_{MD}/\sqrt{SU_A}$	$SU_{MD/\sqrt{A}}$	Continuous explanatory	Unit-less
Distance to nearest stream mean and SD	D2S $\mu$ and $\sigma$	Continuous explanatory	m
Eastness mean and SD	EN $\mu$ and $\sigma$	Continuous explanatory	Unit-less
Elevation mean and SD	EL $\mu$ and $\sigma$	Continuous explanatory	masl
Northness mean and SD	NN $\mu$ and $\sigma$	Continuous explanatory	Unit-less
Planform curvature mean and SD	PLC $\mu$ and $\sigma$	Continuous explanatory	1/m
Profile curvature mean and SD	PRC $\mu$ and $\sigma$	Continuous explanatory	1/m
Relative slope position mean and SD	RSP $\mu$ and $\sigma$	Continuous explanatory	Unit-less
Slope mean and SD	SLO $\mu$ and $\sigma$	Continuous explanatory	Degrees
Topographic wetness index mean and SD	TWI $\mu$ and $\sigma$	Continuous explanatory	Unit-less

observations with components  $y(\mathbf{s})$ ,  $\mathbf{s} \in \mathcal{S}$ , where  $\mathcal{S}$  is the study region, in our case the collection of all SUs. Let  $\mathbf{x}$  with components  $x(\mathbf{s})$ ,  $\mathbf{s} \in \mathcal{S}$  be a latent Gaussian random field and let  $\theta_1$  be a set of hyperparameters. Then, the latent Gaussian field can be hierarchically formulated as follows:

$$\begin{aligned} \mathbf{y} \mid \mathbf{x}, \theta_1 &\sim \prod_{\mathbf{s} \in \mathcal{S}} \phi(y(\mathbf{s}) \mid x(\mathbf{s}), \theta_1) \\ \mathbf{x} \mid \theta_2 &\sim \mathcal{N}(\boldsymbol{\mu}_x(\theta_2), \mathbf{Q}_x^{-1}(\theta_2)) \\ \theta &= (\theta_1, \theta_2)^\top \sim \pi(\theta). \end{aligned} \quad (3.1)$$

For our hurdle modelling approach,  $\phi(\cdot)$  is either the Bernoulli probability mass function for landslide susceptibility or the Gaussian density for landslide log-sizes. The vector  $\theta_1$  contains hyperparameters associated with the likelihood model. So, for instance, if  $\phi(\cdot)$  is the Gaussian density, then  $\theta_1$  is equal to the Gaussian precision, which is the inverse of the standard deviation; see Sect. 3.2 for more details on the likelihood models. The latent Gaussian random field  $\mathbf{x}$  has the role of describing the trends and underlying dependence in the data. It is assumed to have a Gaussian distribution with mean vector  $\boldsymbol{\mu}_x(\theta_2)$  and precision matrix (the inverse of the covariance matrix)  $\mathbf{Q}_x(\theta_2)$ , both of which are controlled by the vector of hyperparameters  $\theta_2$ , which account for variability and length or strength of dependence.

The above hierarchical representation can be further simplified by assuming that  $y(\mathbf{s})$  only depends on a linear predictor  $\eta(\mathbf{s})$ . The linear predictor for our hurdle model

has an additive structure with respect to some fixed and random effects and a term that accounts for the spatial dependence between SUs. Specifically, our linear predictor can be written as

$$\eta(\mathbf{s}) = \alpha + \sum_{m=1}^M \beta_m w_m(\mathbf{s}) + \sum_{k=1}^K f_k(z_k(\mathbf{s})) + u(\mathbf{s}), \quad \mathbf{s} \in \mathcal{S}, \quad (3.2)$$

where  $\alpha$  is an intercept and  $(w_1(\mathbf{s}), \dots, w_M(\mathbf{s}))^\top$  are a subset of the covariates detailed in Tables 1 and 3 with fixed coefficients  $\boldsymbol{\beta} = (\beta_1, \dots, \beta_M)^\top$ . The functions  $\mathbf{f} = \{f_k(\cdot), \dots, f_K(\cdot)\}$  are random (or non-linear) effects defined in terms of a set of covariates  $(z_1(\mathbf{s}), \dots, z_K(\mathbf{s}))^\top$ . For our hurdle model, the specific form of the functions  $f_k(\cdot)$  is that of a Gaussian random walk of order 1 (RW1), defined over a binned version of our covariates. Specifically, for any continuous covariate  $\mathbf{w}_k = (w_k(\mathbf{s}_1), \dots, w_k(\mathbf{s}_{|\mathcal{S}|}))^\top$ , let  $\mathbf{z}_k = (z_{k,1}, \dots, z_{k,L_k})^\top$  be a discretisation of  $\mathbf{w}_k$  into  $L_k$  equidistant bins. Then,  $f_k(\mathbf{z}_k)$  is a Gaussian RW1 with precision parameter  $\tau_{0,k}$  if

$$f_k(z_{k,j}) - f_k(z_{k,j-1}) \sim \mathcal{N}(0, \tau_{0,k}^{-1}), \quad j = 2, \dots, L_k. \quad (3.3)$$

The precision parameter  $\tau_{0,k}$  controls the strength of dependence among neighbouring covariate bins, or in other words, the level of smoothness of the random walk. Finally, the term  $u(\mathbf{s})$  in (3.2) is a zero-mean Gaussian field with a stationary Matérn covariance function (Matérn

1986) with marginal variance  $\sigma^2 > 0$ , range  $r > 0$  and fixed smoothness parameter  $\nu$ . Following Vandekog et al. (2021), we set  $\nu = 1$  to reflect the rather rough nature of the underlying physical process.

The link between the representations in (3.2) and (3.1) is as follows: if  $\boldsymbol{\eta} = (\eta(\mathbf{s}_1), \dots, \eta(\mathbf{s}_{|\mathcal{S}|}))^\top$  and we assume that  $(\boldsymbol{\alpha}, \boldsymbol{\beta}^\top)^\top$  has a Gaussian prior, then the joint distribution of  $\mathbf{x} = (\boldsymbol{\eta}^\top, \boldsymbol{\alpha}, \boldsymbol{\beta}^\top, \mathbf{f}^\top)^\top$  is Gaussian and yields the latent field  $\mathbf{x}$  in the hierarchical representation 3.1; see, e.g., Rue et al. (2017) for more details.

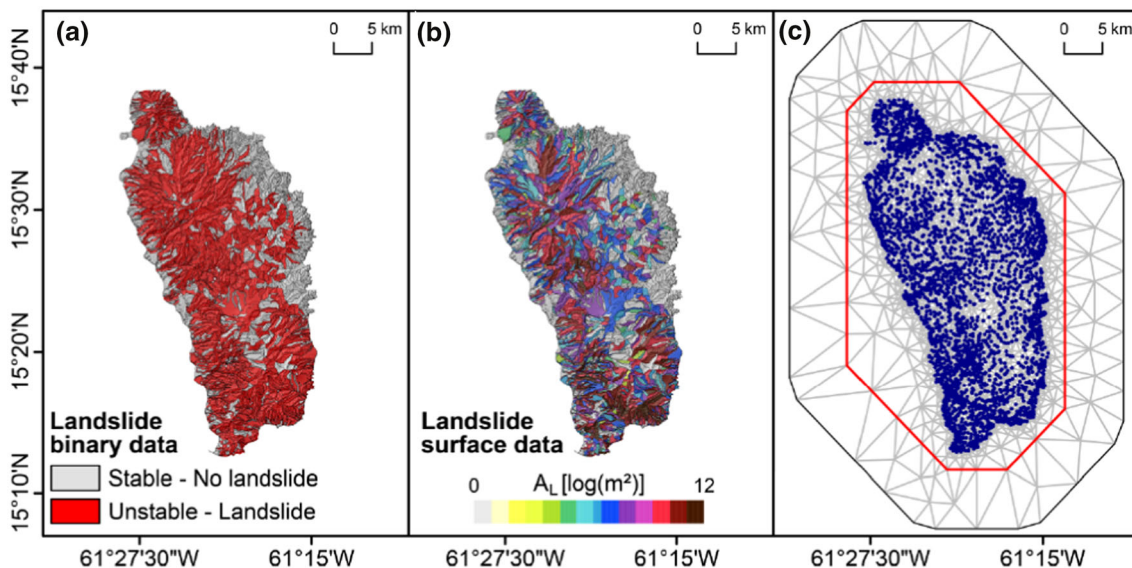
### 3.2 The likelihood models

We model landslide susceptibility with a Bernoulli distribution. Specifically, using the notation from Sect. 3.1, we have that  $y(\mathbf{s}) = O_L(\mathbf{s}) \in \{0, 1\}$  and  $\phi(y(\mathbf{s}) | \eta_{\text{Bern}}(\mathbf{s})) \equiv \text{Bern}(p(\mathbf{s}))$ , where  $p(\mathbf{s}) = \text{Pr}\{O_L(\mathbf{s}) = 1\}$ ; see Fig. 2a for the spatial distribution of  $O_L(\mathbf{s})$ . Note there are no hyperparameters for the Bernoulli likelihood, i.e., the vector  $\boldsymbol{\theta}_I$  is empty. The probability  $p(\mathbf{s})$  is related to the linear predictor  $\eta(\mathbf{s})$  through the logit link, so  $p(\mathbf{s}) = \exp\{\eta_{\text{Bern}}(\mathbf{s})\} / (1 + \exp\{\eta_{\text{Bern}}(\mathbf{s})\})$ .

Landslide size distribution is positively skewed, with extremely large values elongating the right tail. In cases like this, it is standard practice to use a monotonous transformation such as the natural logarithm to obtain a roughly symmetric, Gaussian-like distribution (see Lombardo et al. 2021). Given that a landslide has occurred in the SU  $\mathbf{s} \in \mathcal{S}$ , we model its log-size using a Gaussian distribution. Using the notation from Sect. 3.1, we have that  $y(\mathbf{s}) = \log\{A_L(\mathbf{s})\} | O_L(\mathbf{s}) = 1$  and  $\phi(y(\mathbf{s}) |$

$\eta_{\text{Gauss}}(\mathbf{s}), \boldsymbol{\theta}_1) \equiv \mathcal{N}(\eta_{\text{Gauss}}(\mathbf{s}), \tau^{-1})$ , where  $\boldsymbol{\theta}_1 = \tau$ . Figure 2b shows the spatial distribution of  $\log\{A_L(\mathbf{s})\}$ . The linear predictor is linked to the Gaussian mean via the identity link. The global precision hyperparameter  $\tau$  (reciprocal of the standard deviation) determines the concentration of all values  $y(\mathbf{s})$  around their mean  $\eta(\mathbf{s}), \mathbf{s} \in \mathcal{S}$ .

The linear predictors  $\eta_{\text{Bern}}(\mathbf{s})$  and  $\eta_{\text{Gauss}}(\mathbf{s})$  follow the general equation (3.2). Nonetheless, their specific form depends on the influence of the covariates in Tables 1 and 3 over the landslide susceptibilities and sizes. Therefore, we conduct variable selection using a stepwise forward procedure for most of the covariates in both parts of our hurdle model. This procedure was based on numerical techniques such as the Deviance Information Criterion and the Watanabe-Akaike information criterion (DIC and WAIC, respectively; Meyer 2014; Gelman et al. 2014), and graphical methods such as the probability integral transform (PIT; Gneiting et al. 2007) and fitted versus observed plots. Due to their definition or interpretability, some covariates were not tested this way. Instead, they were included or excluded based on expert opinion. Specifically, expert advice was considered for land-use types and lithology types (included linearly as they are represented as a proportion of a SU and therefore, their sum is constrained to 1); mean Eastness and mean Northness (included linearly as they are complementary measurements representing the sine and cosine of the aspect of a SU); mean slope (included non-linearly based on previous analysis, see, e.g., Tanyaş et al. 2022); and the standard deviations of all covariates (included linearly since non-linear standard deviations lack reasonable interpretation).



**Fig. 2** a Observed landslide presence/absence data; b observed landslide planimetric area, aggregated at the SU level as the sum of all landslides; c Triangulation mesh (gray) with SU centroids (blue) used

to discretise Dominica Island and fit the spatial effect  $u(\mathbf{s})$  in (3.2). The inner boundary (red) delimits the island, whereas the extension to the outer boundary (black) avoids possible boundary effects

All the covariates related to the perimeter were excluded due to collinearity issues. The selected covariates and the specific form they enter into the linear predictor equation for each model are detailed in Table 2.

### 3.3 Prior specification

Here we describe the form of the third layer of the hierarchical representation in (3.1), i.e., the prior distributions for the likelihood hyperparameters  $\theta_1 = \tau$  (the Gaussian precision) and the hyperparameters of the model components in the linear predictor,  $\theta_2 = (\alpha, \beta, \tau_0, \sigma^2, r)^\top$ . Here,  $\alpha$  is a vector of length two that contains the intercept for both likelihood models,  $\beta$  is a vector of length 48 that contains the fixed effects for both likelihood models and  $\tau_0$  is a vector of length ten that contains the RW1 precision parameters for both likelihood models; see Table 2. The vectors  $\sigma^2$  and  $r$  have length 2 and contain, respectively, the marginal variances and range parameters of the Matérn covariance of  $u(\mathbf{s})$  in (3.2) for both likelihoods.

Non-informative priors are a common choice when little expert knowledge is available. We use this approach to define prior distributions over the fixed effects and the intercepts. Specifically, we chose a zero-mean Gaussian prior with a precision of 0.001 for all fixed effects and intercepts. Prior information with different strengths can also be defined using the penalised complexity (PC) prior approach (Simpson et al. 2017). This procedure penalises excessively complex models by placing an exponential prior on a distance to a simpler base model, which helps to stabilise the estimation. Priors then shrink model components toward their base models, thus preventing over-fitting. For the precision  $\tau$  of the log-size observations, we set a weakly informative prior such that the probability of observing a standard deviation  $(1/\tau)$  larger than the empirical standard deviation for the response is 0.01. For the precision parameters of the random walks of order 1, we set weak prior distributions where the probability that the standard deviations  $(1/\tau_{0,k})$  corresponding to  $SU_A$ ,  $SU_{MD}$  and  $TWI$  are greater than 5, 0.5 and 0.1, respectively is 0.01. For the precision parameters of the remaining random walks of order 1, we set relatively weak prior

distributions where the probability that the standard deviations of the random walks are greater than the empirical standard deviation of the response is 0.01. Due to our lack of prior knowledge, we argue that this choice made sense since, a priori, the effects on the linear predictor can only be interpreted relative to the likelihood and the implicit scaling in the likelihood. For both Matérn range parameters, we guided our selection using the empirical variogram and set a prior distribution where the probability that the range is smaller than 25km is 50%. Finally, for the Matérn marginal variances, we set a prior distribution where the probability that the variance is larger than 0.25 is 50%.

### 3.4 Inference with INLA

In a Bayesian framework, the interest lies in the joint posterior distribution of unknown parameters and hyperparameters. In the context of GLMs (see (3.1)), this joint posterior distribution can be written as

$$\begin{aligned}
 p(\mathbf{x}, \boldsymbol{\theta} \mid \mathbf{y}) &\propto \pi(\boldsymbol{\theta}) \phi(\mathbf{x} \mid \boldsymbol{\theta}_2) \prod_{\mathbf{s} \in \mathcal{S}} \phi(y(\mathbf{s}) \mid x(\mathbf{s}), \boldsymbol{\theta}_1) \\
 &\propto \pi(\boldsymbol{\theta}) \mid \mathbf{Q}_x(\boldsymbol{\theta}_2) \mid^{\frac{1}{2}} \\
 &\exp \left\{ \frac{-1}{2} (\mathbf{x} - \boldsymbol{\mu}_x(\boldsymbol{\theta}_2))^T \mathbf{Q}_x(\boldsymbol{\theta}_2) (\mathbf{x} - \boldsymbol{\mu}_x(\boldsymbol{\theta}_2)) \right. \\
 &\left. + \sum_{\mathbf{s} \in \mathcal{S}} \log \phi(y(\mathbf{s}) \mid x(\mathbf{s}), \boldsymbol{\theta}_1) \right\}.
 \end{aligned}
 \tag{3.4}$$

The main goals of the Bayesian inference are to obtain from (3.4) the marginal posterior distribution for each element of the linear predictor and the hyperparameters vector, i.e.,

$$\begin{aligned}
 p(x(\mathbf{s}) \mid \mathbf{y}) &= \int p(\boldsymbol{\theta} \mid \mathbf{y}) p(x(\mathbf{s}) \mid \boldsymbol{\theta}, \mathbf{y}) d\boldsymbol{\theta}, \quad p(\theta_j \mid \mathbf{y}) \\
 &= \int p(\boldsymbol{\theta} \mid \mathbf{y}) d\boldsymbol{\theta}_{-j},
 \end{aligned}
 \tag{3.5}$$

where  $\mathbf{y}$  is the collection of response values across all the SUs and  $\theta_j$  is the  $j$ -th component of the hyperparameter vector  $\boldsymbol{\theta}$ .

**Table 2** Summary of selected covariates for both likelihood models

Likelihood	Fixed effects	Random effects
Bernoulli	$SU_{MD}, SU_{MD/A}, SU_{MD/\sqrt{A}}, D2S\mu, D2S\sigma, EN\mu, EN\sigma, NN\mu, NN\sigma, EL\sigma, PLC\mu, PLC\sigma, PRC\sigma, RSP\mu, RSP\sigma, SLO\sigma, TWI\mu, TWI\sigma, RGA, YPC, YPV, YPI, BAY, IGO, IPL, IGY, OPL, PBA, PPD.$	$SU_A, EL\mu, SLO\mu, PRC\mu.$
Gaussian	$SU_{MD/A}, SU_{MD/\sqrt{A}}, D2S\sigma, EN\mu, EN\sigma, NN\mu, NN\sigma, EL\sigma, PLC\mu, PLC\sigma, PRC\mu, PRC\sigma, RSP\mu, RSP\sigma, SLO\sigma, TWI\sigma, YPV, BAY, OPL.$	$SU_A, SU_{MD}, D2S\mu, EL\mu, SLO\mu, TWI\mu.$

INLA uses numeric techniques to approximate the integrals in (3.5), with an additional step to approximate the marginal posterior distribution for the spatial effect  $u(\mathbf{s})$  in (3.2). Recall that  $u(\mathbf{s})$  is a zero-mean Gaussian field with a stationary Matérn covariance function. Therefore, computation of the posterior marginal distribution requires the inversion of an  $n \times n$  matrix, which becomes computationally demanding for large  $n$ . To speed up this step and make inference feasible, INLA uses the stochastic partial differential equation (SPDE) approach of Simpson et al. (2017) to approximate the Matérn covariance function with numerically convenient sparse precision matrices. Specifically, the SPDE approach constructs an approximation to the precision matrix (defined as the inverse of the covariance matrix) and finds an SPDE whose solutions have desired covariance and precision structures. Lindgren et al. (2011) show how to find these solutions by discretising the study region with a triangulation mesh and representing the stochastic process as a sum of basis functions multiplied by coefficients. They also show that these coefficients form a Gaussian Markov random field, for which methods for fast computation of precision matrices already exist (see, e.g., Krainski et al. 2018 for more details on the SPDE approach). The mesh construction is key to representing the spatial process and can affect the accuracy and speed of the estimation. This step is similar to the choice of integration points on a numerical integration method (Krainski et al. 2018); therefore, we studied the sensitivity of our estimates over changes in the triangulation mesh. We found that a mesh with 4, 600 nodes is a good compromise between the accuracy and speed of the algorithm (see Fig. 2c). INLA and the SPDE approach are conveniently implemented in the R-INLA software (Bivand et al. 2015), and we use them to obtain fast and accurate approximations of the posterior distributions of all the parameters and hyperparameters involved in our models.

## 4 Application to landslide hazard assessment in the Island of Dominica

This section presents the results of our hurdle model in terms of the statistical findings and the implications for landslide hazard assessment.

### 4.1 Hurdle model: main results

Figure 3 shows the posterior means and corresponding 95% credible intervals of the fixed effects for the Bernoulli and Gaussian models. The selected covariates show relatively moderate positive and negative influences on landslide log-sizes and occurrences. The extent to which a covariate, significant or not, contributes to the model can be

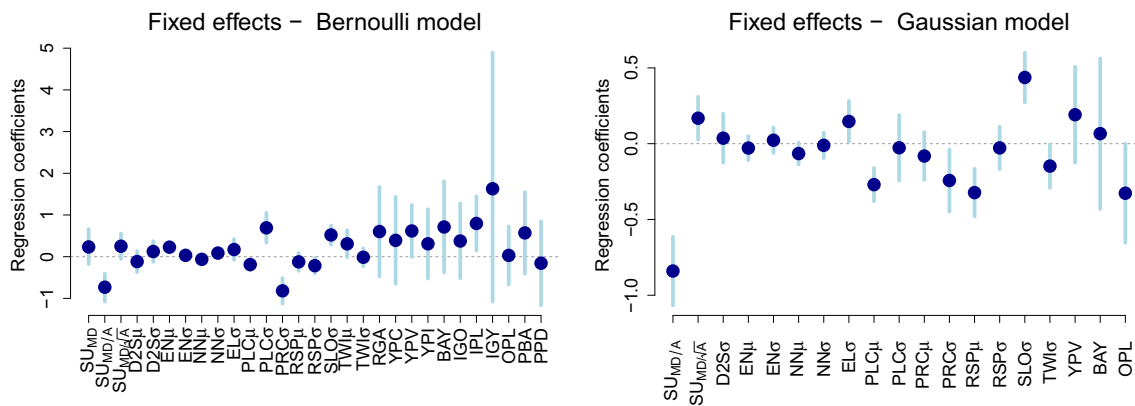
summarised by the size of the posterior mean regression coefficient. For the Bernoulli model, out of the 29 covariates used linearly, eight appeared to be significant;  $SU_{MD/A}$ ,  $EN\mu$ ,  $PLC\mu$ ,  $PLC\sigma$ ,  $PRC\sigma$ ,  $RSP\sigma$ ,  $SLO\sigma$  and  $IPL$ , with mean regression coefficients of  $-0.731$ ,  $0.228$ ,  $-0.190$ ,  $0.693$ ,  $-0.815$ ,  $-0.217$ ,  $0.521$  and  $0.793$ , respectively. From here, the contribution becomes less distinguishable.

For the Gaussian model, out of 19 covariates used linearly, eight appeared to be significant, which indicates that the model is 95% certain of the role (either positive or negative) of the given covariate with respect to the landslide log-size. We can see from the plot in Fig. 3 that the significant covariates are  $SU_{MD/A}$ ,  $SU_{MD/\sqrt{A}}$ ,  $EL\sigma$ ,  $PLC\mu$ ,  $PRC\sigma$ ,  $RSP\mu$ ,  $SLO\sigma$  and  $TWI\sigma$ , with mean regression coefficients of  $-0.835$ ,  $0.165$ ,  $0.146$ ,  $-0.269$ ,  $-0.244$ ,  $-0.322$ ,  $0.439$  and  $-0.148$ , respectively. It is important to notice that even if an effect is not significant, the effect size of the corresponding covariate may be large. Therefore, non-significant effects do not imply that the model is not influenced by the corresponding covariates (Lombardo et al. 2021).

Figure 4 displays the posterior means and corresponding 95% credible bands of the random effects for both models. We can see the highly non-linear influence of most of these covariates on landslide log-sizes and occurrences. For example, looking at the plot for landslide susceptibility, we can see that  $SU_A$  has a moderate non-linear effect, with a positive effect peaking at approximately  $0.67 \text{ km}^2$ .  $EL\mu$  displays a slight concave effect, with a positive effect on landslide susceptibility between 0 and  $0.4 \text{ m}/1000$ . We can also see that  $PRC\mu$  has a relatively linear and mild effect over landslide occurrences, while  $SLO\mu$  is among the most significant non-linear effects with very narrow credible intervals. As expected, steeper slopes are more at risk.

For landslide sizes, we can see that  $SU_A$  has a moderate non-linear effect, with a negative effect for smaller areas and an increasing and eventually positive effect for larger areas.  $SU_{MD}$  seems to have a mild concave and then convex effect on landslide log-size before becoming relatively linear.  $D2S\mu$  is relatively linear when considering the credible bands, with negative effects at small distances.  $EL\mu$  has an overall negative effect on landslide size except for the range  $0\text{--}0.4 \text{ m}/1000$ .  $SLO\mu$  steepness is again among the most significant effects with very narrow credible intervals. As expected, larger slopes—those above  $30^\circ$ —are more at risk. Finally,  $TWI\mu$  has a convex relation with landslide log-size.

Figure 5 shows the posterior mean estimates at the mesh nodes of the spatial fields for both models. We can notice that the spatial field covers almost the entirety of the island for the Bernoulli model, except for the north-eastern



**Fig. 3** Posterior means (dots) of fixed linear effects (except the intercept) with 95% credible intervals (vertical segments) for the Bernoulli and Gaussian models. The horizontal grey dashed lines indicate no contribution to the landslide occurrence or size, respectively

coastlines. This is well in line with observations of Maria passing over Dominica and discharging such an amount of rainfall that most of the shallow material draping over the island has become prone to fail. The same cannot be observed in the case of the spatial field estimated for the Gaussian model. In this case, large landslides appear to share a high degree of spatial dependence in the centre of the island.

To assess the ability of the Bernoulli model to classify landslides correctly, the left panel of Fig. 6 shows the receiver operating characteristic (ROC) curve. It is constructed by plotting the true positive rate (TPR, also called sensitivity) versus the false positive rate (FPR, also calculated as 1-specificity). The TPR boils down to the number of unstable SUs that have been correctly classified divided by the total number of unstable SUs. As for the FPR, this measure is calculated as the number of misclassified stable SUs divided by the total number of stable SUs. The input to the Bernoulli model has a dichotomous nature, whereas the output is a probability. Therefore, to match the output to the input, one needs to use a probability threshold. The way a ROC curve is then constructed is done by selecting a large number of these thresholds and storing the TPR and FPR at each iteration. The greater the area under the ROC curve (AUROC), the better the model and its classification abilities (Zou et al. 2005). The Bernoulli model has an AUROC of 0.927, so we can safely affirm that the model does an excellent job distinguishing between both classes.

To assess the goodness of fit of the Gaussian model, the middle and right panels of Fig. 6 show a histogram of the probability integral transform (PIT) values and a plot of observed versus predicted values, respectively. PITs are commonly used to assess model calibration, i.e., the statistical consistency between the predictive distribution and the observations (Gneiting et al. 2007). If a model is well-calibrated, then the observations should be

indistinguishable from a random draw from the model. For a large number of observations, the PITs histogram serves as a tool to empirically check for uniformity. As expected, large landslide sizes seem to be underestimated by the Gaussian model.

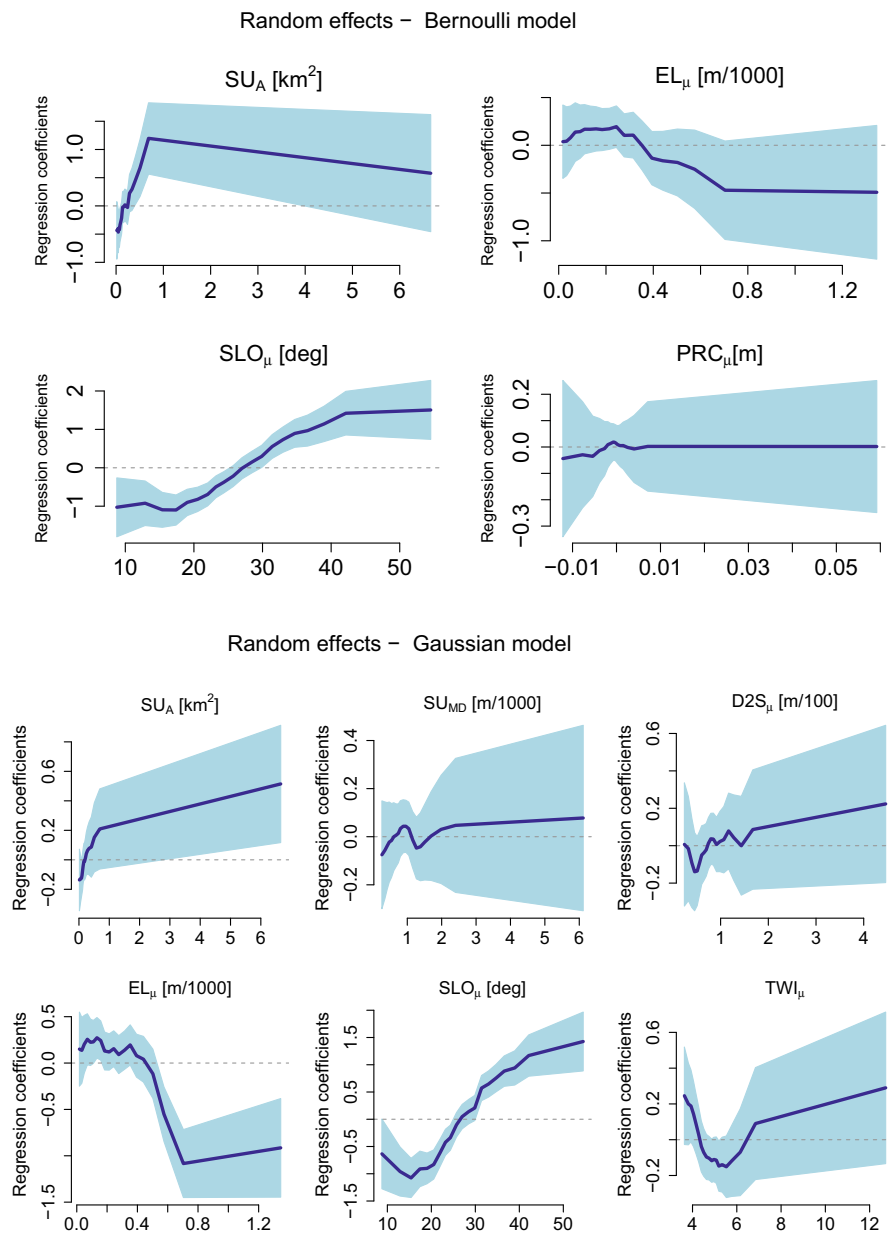
Nonetheless, the PITs in Fig. 6 are not too far from the way a histogram of uniform numbers might look, and it seems fair to assume that our model is decently calibrated. The observed versus predicted values plot shows a relatively good performance for moderate landslide sizes, although there seems to be a greater bias for landslide sizes below the first quartile.

#### 4.2 Interpretation of the covariates' role

The advantage of a statistical model over other data-driven approaches is that the association between dependent and independent variables can be clearly interpreted. This is particularly useful for examining the geomorphological reasonability of a given model. Below, we will present a few examples of our interpretation for those covariates that behave close to our assumptions from the mechanical perspective. In this case, the interpretation should cover a dual aspect of our hurdle model, both from the landslide occurrence and size perspectives. As previously mentioned, Fig. 3 reports the posterior distribution of the regression coefficients estimated for the Bernoulli and the Gaussian models. There, among the lithological classes, ignimbrites (IGO, IPL, IGY) positively contribute to increasing the mean probability of landslide occurrence, with only IPL being significant among the three. From an interpretative standpoint, we recall that ignimbrite is a pyroclastic flow deposit largely comprised of pumice with subordinated ashes. Thus, a positive contribution to the landslide occurrence is reasonable because its mineral structure is prone to weathering, and it is well known to promote slope instabilities (e.g., Chigira and Yokoyama 2005). An



**Fig. 4** Posterior means (dark blue curves) of random effects with 95% pointwise credible intervals (light blue shaded area) for the Bernoulli and Gaussian models



analogous situation can be seen in gravel and alluvium materials (RGA). These are unconsolidated deposits, inherently susceptible to slope instabilities anytime the landscape evolution has set them to drape over steep topographies.

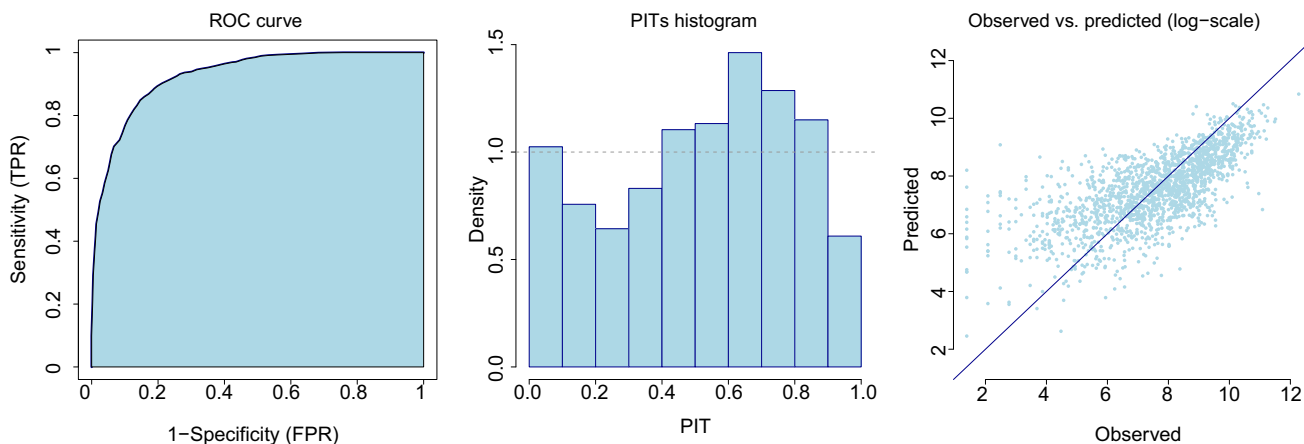
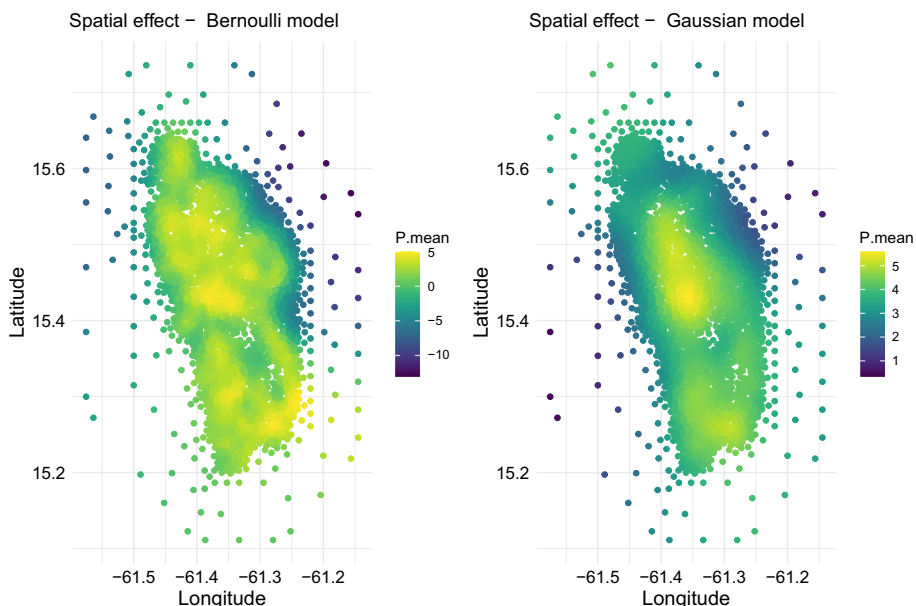
Interestingly enough, most of the lithotypes selected for the Bernoulli model do not appear in the Gaussian case. There, Older Pleistocene volcanics are also reported with a significant and negative contribution to the estimated landslide sizes. Irrespective of the Bernoulli or Gaussian framework, no land-use class plays a role in explaining the landslide occurrence or size.

Another point in common between the Bernoulli and the Gaussian models is the role of  $SU_{MD/A}$ . This covariate

expresses how elongated a given SU is. The larger the value, the more stretched the SU appears, whereas the smaller the value, the more rounded the SU is. Thus, the posterior mean negative value reported for both models may indicate that narrow SUs are not only less prone to fail, but also the lesser availability of material does not allow for large landslides to be generated.

As for the non-linear effects shown in Fig. 4, there we can observe two noteworthy behaviours.  $SU_A$  appears to influence both the Bernoulli and Gaussian models with a negative effect for very small SUs, while the contribution becomes increasingly positive for larger SUs. Similarly to the elongation/roundness ( $SU_{MD/A}$ ) effect mentioned above, this can be interpreted as if larger SUs can host

**Fig. 5** Posterior means of the spatial field for both models



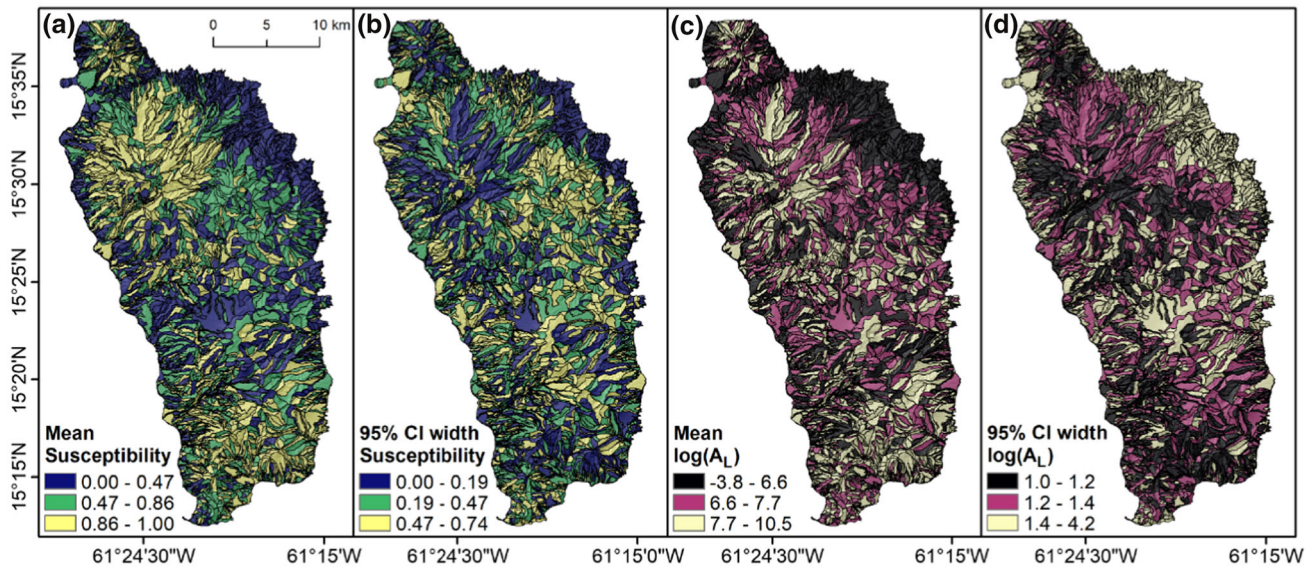
**Fig. 6** ROC curve for the Bernoulli model and PIT Histogram and observed versus fitted plot (in log-scale) for the Gaussian model

more and larger landslides compared to very small SUs. There, the conditions required for a failing mechanism to occur are much more unlikely to manifest simply because of the smaller extent. The other interesting effect corresponds to  $SLO\mu$ . In both models, the slope steepness behaves as a sigmoidal function. This is likely because no shallow flow-like landslide can occur at low steepness values. Conversely, at medium steepness values, we experience a sudden increase in the effect of this covariate, which is to be expected, up to the point where we reach an asymptotic level or even a decrease. A decrease can be interpreted as if very steep topographic conditions cannot host soil, which is washed away by normal erosional processes. Thus, the absence or near-absence of soil implies that no landslide can manifest or that, at best, a small one will mobilise the thin detrital layer draping over the stable bedrock.

**4.3 Landslide hazard components**

Figure 7 presents our findings in terms of posterior means and width of the 95% credible intervals (CI) for landslide susceptibility and sizes (in log-scale) across the island. There, large portions of the areas characterised by higher landslide probability (central north and southern coastlines) coincide with areas where larger landslide sizes are expected. Conversely, the north-eastern coast seems consistently small in landslide size and probability of occurrence.

From the uncertainty plots, we can see that areas with higher occurrence probability (0.86 to 1) have overall small 95% CI width, implying that these areas are well estimated with relatively small uncertainty. We can also see that areas with moderate occurrence probability (0.47 to 0.86) display higher uncertainty, while areas with the smallest



**Fig. 7** Map of Dominica showing posterior mean and width of the 95% credible interval for landslide susceptibility (panels (a) and (b)) and landslide log-size (panels (c) and (d))

occurrence probability (0 to 0.47) show a mixture of small and moderate uncertainty. Large landslide sizes are estimated with relatively small uncertainty, and the higher uncertainties seem to occur where the landslides are small to moderately sized.

#### 4.4 Unified landslide hazard

The Bernoulli model addresses the island’s susceptibility under extreme conditions, such as those induced by Maria. It estimates how prone a given location is to host slope failures, but it is blind to how large landslides may become once they initiate and propagate downhill. To compensate this limitation, the Gaussian model provides information on the expected size of landslides per slope unit, again under extreme meteorological stress. But once more, the model is blind to which slope was effectively prone to fail. So far, we have implemented and presented the results of the two models independently from each other. This framework is already more than enough to satisfy the requirements of the most accepted landslide hazard definition (see Guzzetti et al. 1999), taking aside the temporal dimension. Nevertheless, much more can be done to provide for the first time a unified version of the landslide hazard for the purely spatial context. Below we explain how we can provide a new and unified (data-driven-specific) landslide hazard assessment by combining the two elements of our proposed hurdle model.

Our motivation to provide a unified framework stemmed from the need to provide end-users with spatially distributed information regarding how likely each slope would be to release specific landslide sizes if expected to be

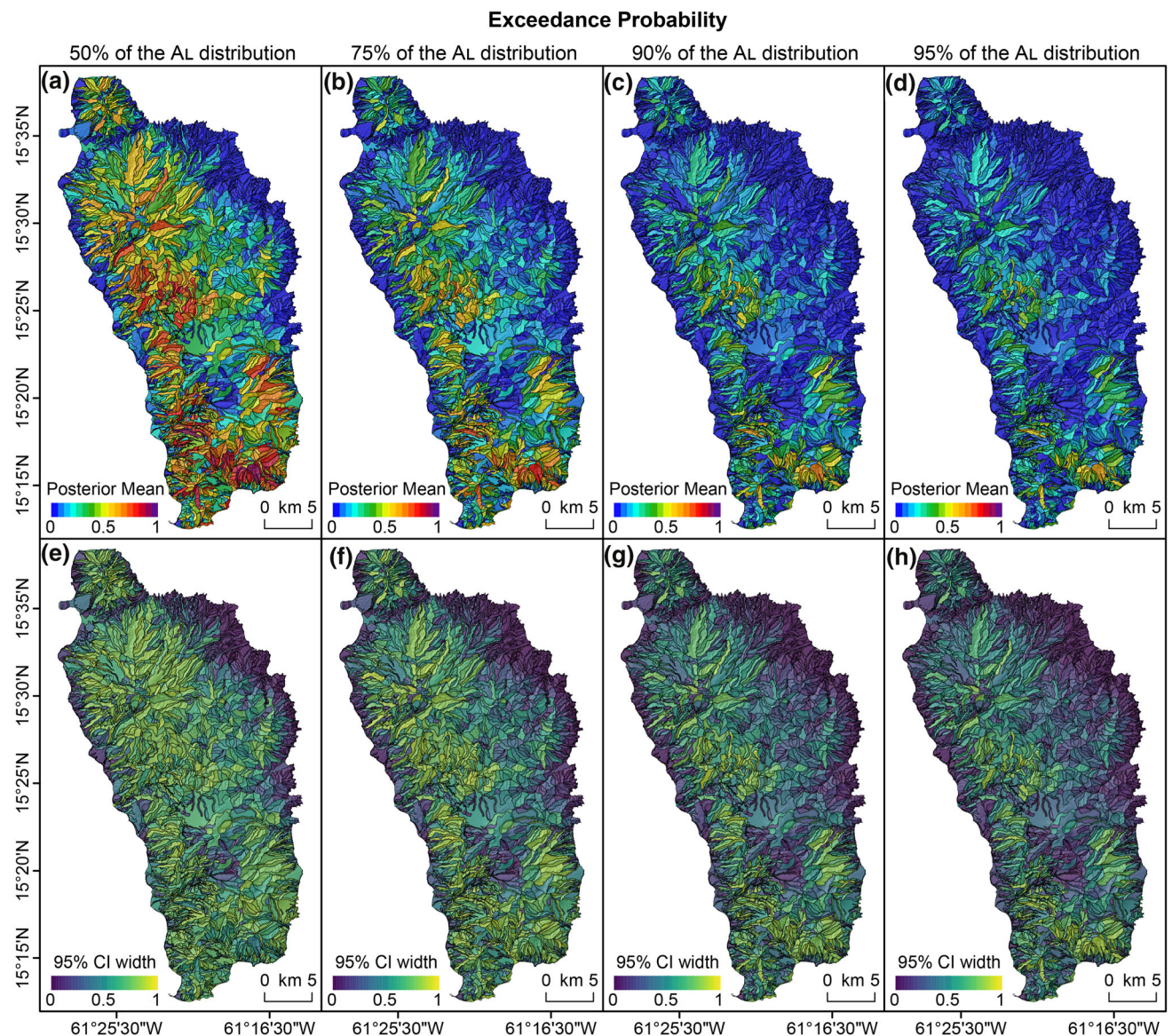
unstable under analogous extreme conditions to those brought by Maria. Small landslides should carry the least hazard; thus, they should be of little interest. Conversely, as the landslide size increases, the expected hazard should proportionally follow. So, to quantify the hazard for moderate and relatively large landslides within the range of observed landslide sizes, we use the two parts of our hurdle model to compute the probability of observing landslide sizes above the empirical 50%, 75%, 90% and 95% quantiles. Specifically, we compute:

$$\Pr(\log\{A_L(s)\} > u) = \Pr(\log\{A_L(s)\} > u \mid O_L(s) = 1) \Pr(O_L(s) = 1), \tag{4.1}$$

for  $u = \hat{F}_{\log(A_L)}^{-1}(0.5), \hat{F}_{\log(A_L)}^{-1}(0.75), \hat{F}_{\log(A_L)}^{-1}(0.90)$  and  $\hat{F}_{\log(A_L)}^{-1}(0.95)$ , where  $\hat{F}_{\log(A_L)}$  is the empirical cumulative distribution function of landslide log-sizes. Note that this procedure can be performed for any landslide size  $u$  of interest. We here focus on the empirical 50%, 75%, 90% and 95% quantiles to illustrate how our model can be used to predict the probability of exceeding medium to large landslides. The exceedance probabilities  $\Pr(\log\{A_L(s)\} > u)$  in (4.1) and their uncertainties can be easily computed using posterior samples from both the Gaussian and Bernoulli models. Specifically, we follow a Monte Carlo (MC) procedure and generate  $N = 500$  samples from the posterior predictive distributions (PPDs) of the Gaussian and Bernoulli models. The PPDs account for uncertainty in the data and the model fitting. We then compute empirical estimates of the probabilities in the right side of (4.1) to obtain one estimate of  $\Pr(\log\{A_L(s)\} > u)$ .

Finally, we replicate this procedure  $M = 1000$  times to obtain MC mean estimates and confidence intervals for  $\Pr(\log\{A_L(s)\} > u)$ , computed as the 2.5% and 97.5% quantiles of the MC estimates. Figure 8 shows plots of the exceedance probabilities and the width of the associated 95% confidence intervals for the four quantiles detailed above. It is interesting to observe how the exceedance probability of the landslide sizes changes from one hazard map to the other. As one should expect, the slopes prone to release median landslide sizes are quite numerous, and their number decreases towards larger landslides. To briefly touch on risk perspectives (although not explicitly integrated into our hurdle model), landslides greater than

90% of the landslide area distribution are particularly likely to occur in the southernmost sector of the island. There, the village of Berekua appears to be potentially vulnerable. The same situation can be seen slightly northwestward for the much larger settlement of Roseau. We stress here that this type of consideration would not be possible in the simple binary case, where the corresponding susceptibility map highlights most of the island as unstable (see Fig. 7a). Note that the confidence intervals in Fig. 8 measure the accuracy of the Monte Carlo approximation and should not be interpreted as a measure of the dispersion in the posterior predictive distribution of the exceedance probabilities.



**Fig. 8** Maps displaying exceedance probability estimates (top) and a measure of their uncertainty (bottom) for landslide sizes above the 50%, 75%, 90% and 95% empirical quantiles. The computations were

based on  $N \times M = 5 \times 10^5$  predictive posterior samples of the Gaussian and Bernoulli models

## 5 Discussion and conclusion

A complete landslide hazard assessment should address three components: where, when (or how frequently) and how large landslides may be. The three components mentioned above have always been addressed separately in the scientific literature produced so far (for non-physically-based models). In this work, we combined two of the three, leaving the temporal characteristic out of the scope because of the limited access to long time series of landslide inventories triggered in response to hurricanes within the Island of Dominica. Despite this limitation, the model we propose represents a substantial improvement with respect to traditional presence/absence susceptibility models because those are blind to the actual threat that a landslide may pose, depending on its size.

We consider this achievement a stepping stone for further experimentation in the hope that one day this unified hazard line of research may impact international guidelines for disaster risk reduction. In this work, we focused on the landslide inventory generated by Maria because of the large sample size and the availability of data on a wide variety of geographical and geological variables, with no missing values. However, further effort can be made to implement an extension in space-time of our hurdle framework. In this line, we are already investigating the combination of the hurricane Maria inventory with the landslide inventory mapped after hurricane Erika in 2015. Although the lack of more temporal replicates inhibits the applicability of complex spatio-temporal models, the statistical characterisation of both processes could provide an exciting perspective and additional means of validation on landslide occurrence and size probabilities across the island. Furthermore, studying the similarities and differences in the values for landslide size and landslide occurrence could be a promising start towards an even more comprehensive hazard map of the island.

Another potential improvement would be carried out by extending this hurdle model toward different modes of slope failures and propagation. Currently, we only model shallow flow-like landslides that either started as debris slides or flows and that generally evolved into debris flows due to the high water content of the moving mass. However, extensions of our model could be implemented to distinguish various landslide classes, including deep-seated ones, which have a completely different failure mechanisms and propagation behaviour.

Ultimately, we also envision the integration of actual rainfall data. Unfortunately, in the specific case of Maria, no reliable rainfall data is available. However, other inventories in data-rich conditions could help integrate

rainfall into the modelling strategy, which in turn could enable testing for a near-real-time application of our hurdle model.

To promote repeatability and reproducibility, the codes and data to fit our hurdle model are freely available from <https://github.com/BryceErin/ULHADominica>. We hope that this will help users utilise and extend the hurdle modelling approach to other geographic contexts, landslide types and triggers.

## Appendix

Table 3 shows acronyms for thematic variables considered in our hurdle model.

**Table 3** Summary of the Dominica geological and land-use variables

Category	Acronym	Class
Geology	BAP	Block ash on Pliocene
	BAY	Block ash flow on young Pleistocene
	CRL	Conglomerate and raised limestone
	IBP	Ignimbrite block ash Pliocene
	IGO	Ignimbrite on Old Pleistocene
	IGY	Ignimbrite on young Pleistocene
	IPL	Ignimbrite on Pliocene volcanics
	MIV	Miocene volcanics
	OPL	Older Pleistocene volcanics
	PBA	Pleistocene apron of block and ash
	PCR	Pleistocene craters
	PIC	Pliocene craters
	PLP	Pliocene Pelean domes
	PLV	Pliocene volcanics
	PPD	Pleistocene Pelean domes
	RGA	Recent river gravel and alluvium
	YPC	Young Pleistocene craters
	YPD	Young Pleistocene Pelean domes
	YPI	Young Pleistocene ignimbrites
	YPV	Young Pleistocene volcanics
Land-use	AGR	Agriculture
	AIR	Airport
	BAR	Bare
	FOR	Forest
	LAK	Lake
	QUA	Quarry
	ROA	Roads
	SWW	Swamp and wetlands
URB	Urban	

**Acknowledgements** Erin Bryce carried out the analyses under the main supervision of Daniela Castro-Camilo and the secondary supervision of Luigi Lombardo; the three have also produced the scientific illustrations and written the manuscript. Daniela Castro-Camilo and Luigi Lombardo have devised the research. Cees van Westen and Hakan Tanyas have contributed to the preprocessing phase by providing the covariates data, manually mapping the landslide inventory and generating the slope unit partition (a mix between automatic delineation and manual correction). Finally, we would like to thank the three anonymous reviewers for their insightful comments.

**Funding** This work was supported by the Additional Funding Programme for Mathematical Sciences, delivered by EPSRC (EP/V521917/1) and the Heilbronn Institute for Mathematical Research.

## Declarations

**Conflict of interest** The authors declare that no conflict of interest applies to data acquisition, analysis and interpretation presented in this article.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

## References

- ACAPS (2018) Dominica; the impact of hurricane Maria. [https://www.acaps.org/sites/acaps/files/products/files/20180131\\_acaps\\_disaster\\_profile\\_dominica\\_v2.pdf](https://www.acaps.org/sites/acaps/files/products/files/20180131_acaps_disaster_profile_dominica_v2.pdf)
- Alvioli M, Marchesini I, Reichenbach P, Rossi M, Ardizzone F, Fiorucci F, Guzzetti F (2016) Automatic delineation of geomorphological slope units with r. slopeunits v1. 0 and their optimization for landslide susceptibility modeling. *Geosci Model Dev* 9(11):3975–3991
- 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. *Eng Geol* 260:105237
- Bivand R, Gómez-Rubio V, Rue H (2015) Spatial data analysis with R-INLA with some extensions
- Carrara A, Cardinali M, Detti R, Guzzetti F, Pasqui V, Reichenbach P (1991) GIS techniques and statistical models in evaluating landslide hazard. *Earth Surf Process Landforms* 16(5):427–445
- Castro-Camilo D, Huser R, Rue H (2019) A spliced Gamma-Generalized Pareto model for short-term extreme wind speed probabilistic forecasting. *J Agric Biol Environ Stat* 24(3):517–534
- Castro-Camilo D, Mhalla L, Opitz T (2021) Bayesian space-time gap filling for inference on extreme hot-spots: an application to Red Sea surface temperatures. *Extremes* 24(1):105–128
- Chigira M, Yokoyama O (2005) Weathering profile of non-welded ignimbrite and the water infiltration behavior within it in relation to the generation of shallow landslides. *Eng Geol* 78(3–4):187–207
- Fobert M-A, Singhroy V, Spray JG (2021) InSAR monitoring of landslide activity in Dominica. *Remote Sens* 13(4):815
- Gelman A, Hwang J, Vehtari A (2014) Understanding predictive information criteria for Bayesian models. *Stat Comput* 24(6):997–1016
- Gibbins S (2019) This Caribbean island is on track to become the world's first 'hurricane-proof' country. *Natl Geographic Sci*
- Gneiting T, Balabdaoui F, Raftery AE (2007) Probabilistic forecasts, calibration and sharpness. *J R Stat Soc Ser B* 69(2):243–268
- Guttorp P, Gneiting T (2006) Studies in the history of probability and statistics XLIX on the Matérn correlation family. *Biometrika* 93(4):989–995
- 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):181–216
- Krainski ET, Gómez-Rubio V, Bakka H, Lenzi A, Castro-Camilo D, Simpson D, Lindgren F, Rue H (2018) Advanced spatial modeling with stochastic partial differential equations using R and INLA. CRC Press, Boca Raton
- Lindgren F, Rue H, Lindström J (2011) An explicit link between Gaussian fields and Gaussian Markov random fields: the stochastic partial differential equation approach. *J R Stat Soc Ser B* 73(4):423–498
- Lombardo L, Opitz T, Ardizzone F, Guzzetti F, Huser R (2020) Space-time landslide predictive modelling. *Earth Sci Rev* 209:103318
- Lombardo L, Tanyas H, Huser R, Guzzetti F, Castro-Camilo D (2021) Landslide size matters: a new data-driven, spatial prototype. *Eng Geol* 293:106288
- Matérn B (1986) Spatial variation, vol 36 of lecture notes in statistics Meyer R (2014) Deviance information criterion (DIC). *Statistics reference Online*, Wiley, pp 1–6
- Opitz T, Huser R, Bakka H, Rue H (2018) INLA goes extreme: Bayesian tail regression for the estimation of high spatio-temporal quantiles. *Extremes* 21(3):441–462
- Pasch RJ, Penny AB, Berg R (2018) National hurricane center tropical cyclone report: Hurricane Maria. National Oceanic And Atmospheric Administration and the National Weather Service, pp 1–48
- Rue H, Martino S, Chopin N (2009) Approximate Bayesian inference for latent Gaussian models by using integrated nested Laplace approximations. *J R Stat Soc Ser B* 71(2):319–392
- Rue H, Riebler A, Sørbye SH, Illian JB, Simpson DP, Lindgren FK (2017) Bayesian computing with INLA: a review. *Annu Rev Stat Appl* 4:395–421
- Simpson D, Rue H, Riebler A, Martins TG, Sørbye SH (2017) Penalising model component complexity: a principled, practical approach to constructing priors. *Stat Sci*, pp 1–28
- Stein ML (2012) Interpolation of spatial data: some theory for kriging. Springer, New York
- Tanyas H, Hill K, Mahoney L, Fadel I, Lombardo L (2022) The world's second-largest, recorded landslide event: lessons learnt from the landslides triggered during and after the 2018 Mw 7.5 Papua New Guinea earthquake. *Eng Geol* 297:106504
- Titti G, van Westen C, Borgatti L, Pasuto A, Lombardo L (2021) When enough is really enough? On the minimum number of landslides to build reliable susceptibility models. *Geosciences* 11(11):469
- Van Westen C, Zhang J (2018) Tropical cyclone Maria. Inventory of landslides and flooded areas UNITAR Map Product ID 2762
- Vandeskog SM, Martino S, Castro-Camilo D, Rue H (2021) Modelling short-term precipitation extremes with the blended generalised extreme value distribution. *arXiv preprint arXiv:2105.09062*
- Zou KH, Resnic FS, Talos I-F, Goldberg-Zimring D, Bhagwat JG, Haker SJ, Kikinis R, Jolesz FA, Ohno-Machado L (2005) A global goodness-of-fit test for receiver operating characteristic curve analysis via the bootstrap method. *J Biomed Inform* 38

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.