Numerical recipes for landslide spatial prediction by using R-INLA: a step-by-step tutorial

Luigi Lombardo^{1,2*}, Thomas Opitz³, Raphael Huser¹

April 24, 2018

Abstract

The geomorphological community typically assesses the landslide susceptibility at the catchment or larger scales through spatial predictive models. However, the spatial information is conveyed only through the geographical distribution of the covariates. Spatial dependence, which allows capturing similarities at neighboring sites that are not directly explained by covariate information, is typically not accounted for in the landslides literature, whilst such spatial models have become commonplace in the geostatistical literature. Here we explain step by step how to rigorously model and predict activations of debris flow based on an adequate statistical model by using the R-INLA library from the statistical software R in the context of a real multiple landslide event. This chapter follows the analysis of [Lombardo *et al.* \(2018a\)](#) with a few modifications; it is written in a tutorial style to provide the geomorphological community with a hands-on guide to replicate similar analyses in R. While our focus here is on implementation and computing, more details about the underlying statistical theory, modeling and estimation can be found in [Lombardo *et al.* \(2018a\)](#). Our modeling approach deviates fundamentally from the commonly-used regression models fitted to binary presence/absence data. Specifically, we use a Bayesian hierarchical Cox point process model to describe landslide counts at high resolution (i.e., at the pixel level), while capturing spatial dependence through a latent spatial effect defined at lower resolution over slope units. Our point process modeling approach allows us to derive the distribution of aggregated landslide counts for any areas of interest. Crucially, the latent spatial effect represents the unexplained but spatially structured component in the data when the linear or nonlinear effects of covariates are removed. Thus, in the case of sparse raingauge or seismic networks, we suggest using the latent spatial effect to uncover the trigger distribution over space. In particular, for landslides triggered by extreme precipitation, the meteorological stress can play a dominant role with respect to the covariates that are typically introduced in predictive models; therefore, accounting for the trigger in modeling may dramatically improve the performance of landslide prediction.

Keywords: Cox point process, Integrated Nested Laplace Approximation (INLA), Landslide susceptibility, Landslide intensity, R-INLA, Slope unit, Spatial point pattern

¹King Abdullah University of Science and Technology (KAUST), Computer, Electrical and Mathematical Sciences and Engineering (CEMSE) Division, Thuwal 23955-6900, Saudi Arabia.

²King Abdullah University of Science and Technology (KAUST), Physical Sciences and Engineering (PSE) Division, Thuwal 23955-6900, Saudi Arabia.

³INRA, BioSP, 84914, Avignon, France

1 Introduction

Landslides are geomorphological processes that contribute to the natural evolution of mountainous landscapes. However, as human population grows, such physiographic contexts are increasingly occupied, thus exposing our society to significant economic (Winter *et al.*, 2016) and life (Kennedy *et al.*, 2015) losses. To assess and better manage the risk of landslide occurrences, scientists have implemented predictive models since the early 1990s (see, e.g., Carrara *et al.*, 1991), primarily aiming at recognizing *where* new mass movements might occur in the future. Identifying potentially unstable terrains is vital for decision makers, for it enables authorities to better implement land use planning programs, as well as risk management procedures (Oliveira *et al.*, 2017). However, more detailed information can complement and even improve landslide risk assessment. For example, knowing *how many* (additionally to *where*) landslides might occur in a given area is crucial to quantify risks, but the scientific community has hardly investigated this key aspect.

Over the last two decades, to estimate *where* future landslides may occur, researchers have predominantly used statistical methods, especially presence/absence regression models (Atkinson and Massari, 1998; Greco *et al.*, 2007; Pourghasemi and Rossi, 2017). Such models assume that landslides occur in a fixed area according to the Bernoulli distribution. This distribution describes the probabilities of a binary outcome, namely, if landslides are present or absent in a given mapping unit (Carrara *et al.*, 1995), which may be a single pixel, a slope unit, or an administrative unit (Van Den Eeckhaut *et al.*, 2009; Erener and Düzgün, 2012). However, when the space is partitioned into such mapping units, further information may be obtained if the landslide count per unit area is considered. Therefore, the binary 0/1 paradigm can be shifted to the more informative count paradigm described by nonnegative integer values $0, 1, 2, \dots$, and often be modeled through the Poisson distribution in statistics (see section §3.2). Such framework based on landslide counts aims to estimate the landslide *intensity* (Lombardo *et al.*, 2018a), rather than solely estimate the landslide *susceptibility*. Susceptibility loosely describes the odds of observing *at least one* landslide in a given mapping unit (Petschko *et al.*, 2014), thus neglecting the actual number of such events.

In this chapter, we follow the statistical analysis of Lombardo *et al.* (2018a), apart from some minor model modifications explained below, and put strong emphasis on its practical implementation and computing. Our approach digresses significantly from the current landslides literature; in order to model landslide counts, we use a suitable Cox point process model, which provides a probabilistic framework to describe spatial point patterns (here, landslide locations) over continuous space. This doubly stochastic model may be used to account for linear and/or nonlinear effects of covariates, as well as for unobserved spatially correlated components (e.g., the landslide trigger), directly embedded into the landslide intensity. Including a spatial effect is particularly important when the intensity of the trigger, which might be a heavy storm or an earthquake (Lee, 2014), is unobserved and strongly heterogeneous over space, such that it dominates the other fixed morphometric or thematic covariate effects. In this case, the spatial effect aids in explaining a large part of the spatial

variation in the landslide scenario, especially for regions of very low and very high trigger activity, and decoupling the effects into observed covariates and the unobserved trigger is crucial to achieve the best predictive performance. When the raingauge or seismic network is poor, we suggest using the estimated spatial random effect to uncover the trigger distribution over space, which may then be used subsequently to simulate potential future landslide scenarios. Moreover, unlike traditional regression-based approaches, which rely on a fixed and subjective partition of the study region into mapping units, the proposed point process-based framework provides a continuous space description of the landslide intensity, which can then be used to consistently derive the distribution of landslide counts for *any* sub-area of interest, *without* re-fitting the model. In other words, this global continuous space representation of the data ensures validity and self-consistency of results for distinct sub-areas.

In the geomorphological literature, so far, landslide predictive models have been mostly implemented from a frequentist perspective, rather than from Bayesian approaches (see, e.g., [Das et al., 2012](#)). However, the doubly stochastic point process model proposed by [Lombardo et al. \(2018a\)](#) described above is most conveniently embedded into a Bayesian framework, in which suitable prior distributions must be defined for all model components and hyperparameters, allowing for expert knowledge to be incorporated. Such complex Bayesian models may be fitted by computer-intensive simulations based on Markov chain Monte Carlo methods, or alternatively, by the integrated nested Laplace approximation (INLA) of posterior distributions ([Rue et al., 2009](#); [Illian et al., 2012](#); [Lindgren and Rue, 2015](#); [Simpson et al., 2016](#); [Rue et al., 2017](#); [Opitz, 2017](#); [Bakka et al., 2018](#)), which is much faster, exceedingly accurate and bypasses convergence assessment issues.

While [Lombardo et al. \(2018a\)](#) provide more details on the statistical theory, especially on model building and the estimation method, this chapter, written as a tutorial, intends to serve as a practical guide for the geomorphological community to build and fit complex Cox point process models by using INLA, a library conveniently implemented within the R package R-INLA ([Bakka et al., 2018](#)). Similarly to [Lombardo et al. \(2018a\)](#), we illustrate the proposed methodology and its implementation by analyzing the multiple occurrence regional landslide event (MORLE, [Crozier, 2005](#)) that took place on October 1, 2009, in Messina, Italy, as a result of an extreme storm ([Cama et al., 2015](#)). To illustrate our method, all morphometric covariates here are treated linearly, except the Aspect, which is assumed to be nonlinear and cyclic. A different, more complex, point process model that comprises additional nonlinear effects is fitted to the same landslide inventory in [Lombardo et al. \(2018a\)](#), although the predictive performance of the simpler model presented below is close to optimal.

This chapter is organized as follows. In §2, we introduce the dataset and available covariates, and we further detail how to compute slope units in GIS. In §3, we then explain step by step how to prepare and format the dataset to be used by R-INLA, and we illustrate how to estimate the Cox point process model including several fixed effects and a latent spatial ran-

dom effects (used to capture the signal of unobserved covariates, such as the trigger) defined over slope units. In §4 we describe how to: i) extract the results, ii) calculate the landslide intensity at pixel, slope unit and catchment resolutions, iii) derive the equivalent landslide susceptibility, iv) perform a cross-validation study, and finally v) calculate goodness-of-fit and predictive metrics, such as the receiver operating characteristic (ROC) curve and the area under the curve (AUC). The results (reported in §4) are then discussed in §5. We conclude this chapter in §6 with some perspectives on future research.

For reproducibility purposes, our code is made entirely available online on <https://github.com/ThomasOpitz/landslides-point-process>, hoping that this will trigger further research based on different landslide inventories and different contexts.

2 Dataset description and preparation

2.1 Multiple occurrence regional landslide event, Messina, 2009

On October 1, 2009, a total number of 4879 landslides of debris flow and debris avalanche types (Hungre *et al.*, 2014) were triggered when an unprecedentedly powerful storm (Cama *et al.*, 2017) hit the area around the city of Messina, Italy. During the propagation phase, most of the avalanches also evolved into debris flows due to the high water content, incurring 37 deaths and approximately 500 million Euros of damages to the infrastructure. The area affected by the storm (see Figure 1) represents the North-Easternmost sector of the island of Sicily. In this region, crystalline basement rocks (see Figure 1 of Di Stefano *et al.*, 2015) made of heavily weathered, medium to high-grade metamorphic successions (Lombardo *et al.*, 2016b) outcrop together with organic, silty loamy and loamy sandy soils (Schillaci *et al.*, 2017b,a; Lombardo *et al.*, 2018b).

To build a statistical model of such a disaster, the dataset must be constructed and prepared via three major steps:

- (i) creating a point database of landslide triggering locations (usually called “LIPs” for “Landslide Identification Points”). This is common practice in the geomorphological literature, and two approaches are mostly used: the first one uses the highest position along a polygon-shaped landslide scar (Lombardo *et al.*, 2014; Cama *et al.*, 2017); the second one uses the centroid of the landslide scar (Hussin *et al.*, 2016; Dou *et al.*, 2015). In this work, we opted for the first approach because in the case of shallow landslides, this approach appears to be the most reasonable to approximate the locations where the initial failure took place. Additionally, this method has already produced reliable results in the same area (e.g., Cama *et al.*, 2015);
- (ii) choosing a set of predictors to explain the landslide distribution. Similarly to the previous step, we selected the same covariates that appeared in other papers treating this event (see, e.g., Cama *et al.*, 2016, for more details). These covariates are either continuous (see Figure 2) or categorical (see Figure 3). Although the aspect is a

Figure 1: (a) Geographic context; (b) Area hit by the storm and local toponyms; (c) Landslide Identification Points (LIPs).

continuous covariate, we discretize it into 16 equidistant classes and treat it in our analysis as a cyclic categorical covariate with neighboring class dependence; see §3.1 and 3.2 for more details on the model and its implementation;

- (iii) selecting the mapping units to define the statistical model and to present the final results. In the present work, the data, all covariates, and our point process model are defined at the pixel resolution (approximately half a million 225m^2 pixels). The latent spatial effect, used to explain spatial correlation among nearby locations, is defined over a (coarser) slope unit partition (approximately 4 thousands). This modeling approach allows us to exploit the fine-scale resolution of the data and to precisely describe the results, while significantly speeding up computations. To illustrate this, we present

Figure 2: Continuous covariates selected for this work. The color scale for Aspect is periodic.

Figure 3: Categorical covariates selected for this work. The acronyms of single classes for each categorical predictor are listed in Appendix B of Lombardo *et al.* (2018a).

the results at the pixel, slope unit and catchment levels, though it would be possible to provide results for any other mapping unit without re-fitting the model. Details on how to generate slope units in GIS are given in §2.2.

2.2 Computing slope units in GIS

Slope units correspond to physiographic features that partition any landscape into half sub-catchments. Therefore, they can be computed at different scales, which makes the process subjective and time-consuming. Most of these issues have been recently resolved, as illustrated in Alvioli *et al.* (2016), where the authors propose an efficient automatic method to compute slope units in GIS. The authors describe a code called `r.slopeunits` (see <http://geomorphology.irpi.cnr.it/tools/slope-units>), and implemented it in GRASS GIS. The script `r.slopeunits` requires only a digital elevation model (DEM) as input. Thus, after a few parameters that control the desired output resolution are set, slope units can be automatically and objectively generated.

We report below the GRASS GIS syntax to run `r.slopeunits` in Ubuntu. Here, we assume that the DEM is saved in GeoTIFF format and named as `Elevation.tif`. We fix the initial flow accumulation to 150000m^2 , the reduction factor to 2, the minimum slope unit area to 10000m^2 , the minimum circular variance of terrain aspect within slope units to 0.25, the threshold area under which the smallest slope units are removed to 2000m^2 , and finally the maximum number of iterations for recursively subdividing slope-units to 10. For more details on the exact meaning and the effect of these parameters, see Alvioli *et al.* (2016) and the references therein. After opening the terminal and running GRASS GIS, we load the DEM of the study area, and we run the following command:

```
./r.slopeunits demmap=Elevation slumap=SUraster slumapclean=SUclean
reductionfactor=2 maxiteration=10 thresh=150000 circularvariance=0.25
cleansize=2000 areamin=10000 --0
```

The output from `r.slopeunits`, here called `SUraster`, is a raster data file. To convert it into a vectorial format, we run the following command:

```
r.to.vect input=SUraster output=SUvector type=area
```

We then save it as a shapefile called `SU.shp` in what follows. The next section describes how to build and fit a meaningful point process model for our landslide inventory using `R` and the library `R-INLA` (Martins *et al.*, 2013), and how to extract the results, in order to create spatial prediction maps of landslide intensities.

3 Point process modeling and estimation using R-INLA

3.1 Preprocessing

Before proceeding with the model construction and estimation, we need to load and prepare the data in `R` to be used subsequently by `R-INLA`.

We start by opening an `R` terminal and importing the slope unit delineation previously calculated into `R`, thanks to the `maptools` package (Bivand and Lewin-Koh, 2017):

```
> library(maptools)
> SU=readShapeSpatial("SU.shp")
```

Then, we compute the adjacency graph structure (Carnia *et al.*, 2017) determining the neighborhood relationships of slope units, which can be achieved using the `spdep` package (Bivand and Piras, 2015):

```
> library(spdep)
> nb2INLA("adjgraph.txt",poly2nb(SU,queen=F,row.names=SU$ID))
```

where the `ID` column corresponds to the numbered vector of slope unit IDs. This preparatory procedure is needed for `R-INLA` in order to implement spatial dependency across slope units using a simple conditional autoregressive (CAR) model (Besag, 1975). Alternatively, it would also be possible to capture spatial correlation through a different model based on stochastic partial differential equations (SPDE, Lindgren *et al.*, 2011), but inference then becomes more complicated. For illustration purposes, we choose here the CAR model, although it would be interesting in future research to assess the effect that **the choice between the SPDE or CAR model has on the results**, and in particular on the overall predictive performance.

We then load the landslide inventory and the covariates into `R`. The data file called `datamatrix.txt` is assumed to contain a large data matrix, where the columns contain the x - and y -coordinates of each pixel, the landslide count for each pixel, a variable `SU.ID`

identifying the slope unit that contains each pixel, and the covariate values for each pixel. We read this file in R as a data frame, making sure that missing values are encoded as NAs **such that we can remove records with missing covariate values from the dataset**, and we then rescale each continuous covariate (except the aspect treated as categorical) by subtracting the mean and dividing by the standard deviation; this helps to stabilize the estimation procedure and to easily compare the importance of continuous covariates effects. We run the following code:

```
> dataDF=read.delim("datamatrix.txt",header=T)
> dataDF[dataDF== -9999 | dataDF== -99999]=NA
> dataDF=na.omit(dataDF)
> data_scaled=dataDF
> vars2scale=c("Elevation","Dist2Fault","NDVI","Plan_Cur","Prof_Cur",
               "Slope","SPI","TWI")
> data_scaled[,vars2scale]=apply(data_scaled[,vars2scale],2,scale)
```

We then extract the landslide counts into a vector `y.count`, and to avoid numerical instabilities we calculate the area of each pixel, which is used to **fix the** areal unit of the Poisson regression model to 1m^2 by including a fixed multiplicative offset in the model (called `offset` below). We also use another constant (called `avg.global`), equal to the average number of landslides per square meter, which helps in stabilizing the estimation of the intercept by concentrating its prior distribution relatively close to this value:

```
> y.count=data_scaled$count
> n.pixels=nrow(data_scaled)
> area.pixel=rep(15^2,n.pixels)
> offset=area.pixel
> n.landslides=sum(y.count)
> avg.global=n.landslides/(n.pixels*area.pixel)
```

Next, we prepare the covariate matrix to be used by R-INLA, adding the intercept (a column of ones), and merging minor lithology classes into one single class:

```
> covar.inla=data_scaled[,c("Aspect", "Elevation","Dist2Fault","Lithology",
                           "Landuse", "Landform","NDVI","Plan_Cur","Prof_Cur","Slope",
                           "SPI","TWI","SU.ID")]
> covar.inla=cbind(intercept=1,covar.inla)
> doReplace=covar.inla$Lithology %in% c(8,14,15,16,17,18,20,21,22,23,24,29)
> covar.inla$Lithology[doReplace]=0
```

The 12 lithotypes merged into a single class outcrop in the area for few hundreds of squared meters and do not contain any landslide. Thus, their separate inclusion in the model would be meaningless and could lead to high uncertainty, which may complicate the estimation of

the model and its interpretation. However, to continuously represent the whole study region without missing values in the covariates, we have to include them, and we opted for a single class as a good compromise. Ultimately, we convert the aspect into a categorical covariate with 16 equidistant classes, **by cutting the interval $[0, 360)^\circ$ into 16 segments of length 22.5° :**

```
> covar.inla$Aspect=cut(covar.inla$Aspect,labels=F,breaks=seq(0,360,length=17))
```

The function `inla.group` within the INLA package may also be useful to split a continuous covariate into a fixed number of distinct **classes**. This covariate transformation is useful to treat the aspect non-linearly, with one distinct effect per class. **It would be possible to consider more classes for a finer representation of the aspect variable, but we have found that our choice of 16 classes provides a good compromise between modeling accuracy and computational convenience.** Moreover, while categorical variables such as the land use or lithology have typically unrelated classes, the effects of a categorical variable with an intrinsic class ordering (such as the aspect) can be assumed to have a specific correlation structure; this provides smoother results and makes estimation easier by borrowing strength across neighboring classes, as explained in §3.2.

3.2 Fitting a Cox point process model using R-INLA

Loosely speaking, a Poisson point process model assumes that for any area $\mathcal{A}$, the number $N(\mathcal{A})$ of events (in our case, landslides) occurring in $\mathcal{A}$ follows the Poisson distribution with mean $\Lambda(\mathcal{A}) = \int_{\mathcal{A}} \lambda(s) ds$, where $\lambda(s) > 0$ denotes the landslide *intensity* at location s . Mathematically, $\text{pr}\{N(\mathcal{A}) = k\} = \exp\{-\Lambda(\mathcal{A})\} \Lambda(\mathcal{A})^k / k!$, $k = 0, 1, \dots$. Intuitively, $\lambda(s)$ can be interpreted as the (rescaled) expected number of events occurring in an infinitesimal region around s ; in other words, $\lambda(s)$ reflects the “density” of landslides: the larger the intensity $\lambda(s)$, the more likely events are to occur in a small neighborhood of s . The log-Gaussian Cox point process model that we fit here further assumes that the log intensity $\log\{\lambda(s)\}$ is random and is driven by a Gaussian process, whose mean and covariance structures are determined by covariates and latent random effects. Precise mathematical details about this modeling framework may be found in Lombardo *et al.* (2018a).

We now explain how to implement and estimate such a complex point process model using R-INLA, containing fixed covariates effects, one cyclic non-linear effect (the aspect), and one latent random field defined over slope units capturing the spatial dependence structure (due in our case to the intense precipitation trigger). A slightly more complicated model with additional non-linear covariates is considered in Lombardo *et al.* (2018a).

Before fitting the model, we need to load the INLA library (see <http://www.r-inla.org/download> for the latest version), and to define the formula characterizing the linear predictor, embedded into the landslide log intensity $\log\{\lambda(s)\}$. This can be achieved by additively combining the different model components in a simple way, as follows:

```
> library(INLA)
```

```

> form=y~-1+intercept+Elevation+Dist2Fault+NDVI+Plan_Cur+Prof_Cur+Slope+SPI+TWI+
  f(Aspect,model="rw1",cyclic=T,constr=T,
    hyper=list(theta=list(initial=log(25),fixed=T))) +
  f(Landuse,model="iid",hyper=list(prec=list(initial=log(100),
    fixed=T)),constr=T) +
  f(Landform,model="iid",hyper=list(prec=list(initial=log(100),
    fixed=T)),constr=T) +
  f(Lithology,model="iid",hyper=list(prec=list(initial=log(100),
    fixed=T)),constr=T) +
  f(SU.ID,model="besag",graph="adjgraph.txt",
    hyper=list(theta=list(initial=log(1),fixed=F,prior="loggamma",
    param=c(0.25,0.25))))

```

In the code above, all continuous covariates (except the aspect) appear on the first line and are simply treated as linear fixed effects. However, **the aspect, which takes values in $[0, 360)^\circ$, is treated as an ordered, cyclic, categorical variable. We assume that the dependence structure of the random effects for its 16 classes (i.e., $[0.0, 22.5)^\circ, [22.5, 45.0)^\circ, \dots, [337.5, 360.0)^\circ$) is driven a priori by a Gaussian random walk of first order with fixed precision parameter $\tau_{\text{Asp}} = 25$.** Mathematically, if x_i denotes the effect for the i th class, one has

$$x_{i+1} - x_i \sim \mathcal{N}(0, \tau_{\text{Asp}}^{-1}), \quad i = 1, \dots, 17,$$

where $x_{17} \equiv x_1$ by convention, and where the effects are constrained to sum to zero for identifiability. This model induces correlation among nearby classes, while maintaining high flexibility. Large values of τ_{Asp} imply strong interclass correlation, and vice versa. The purely categorical, unstructured, variables (land use, landform, and lithology) are modeled as independent Gaussian random effects, with fixed precision parameter $\tau_{\text{Cat}} = 100$, and also constrained to sum to zero. Finally, the spatial random effect is assumed to be driven a priori by a zero-mean Gaussian conditional autoregressive (CAR) model (Besag, 1975) defined over slope units, with precision $\tau_{\text{CAR}} > 0$ to be estimated. If z_i denotes the effect for the i th slope unit, $N(i)$ indicates the collection of its neighbors, and $\mathbf{z}_{-i}$ is a vector containing all such effects except for z_i itself, then one has

$$z_i \mid \mathbf{z}_{-i} \sim \mathcal{N} \left(\frac{1}{|N(i)|} \sum_{j \in N(i)} z_j, \frac{1}{|N(i)|\tau_{\text{CAR}}} \right).$$

In other words, the (conditional) mean effect for the i th slope unit corresponds to the average of its neighbors, which induces spatial correlation. The closer two slope units are to each other (with respect to the graph distance), the stronger the correlation of their corresponding effects. Furthermore, larger values of τ_{CAR} imply stronger correlation, and vice versa. Thus, the estimated value of τ_{CAR} is crucial, as it controls the smoothness of the estimated latent spatial effect, and this needs to be validated with care. In the linear predictor formula above,

we further specify that τ_{CAR} has a gamma prior with rate and shape parameters both equal to 0.25, leading to a moderately informative prior distribution with a mean value of 1 and a variance value of 4. In practice, **we must very carefully assess the sensitivity of the results with respect to hyperparameters (i.e., the precision parameter values, and the parameters of all prior distributions).**

After having constructed the model formula, we create a data stack required by R-INLA:

```
> stack=inla.stack(data=list(y=y.count,e=offset),A=list(1),effects=list(covar.inla))
```

We can finally fit the desired Cox point process model using INLA by calling the function `inla` as follows:

```
> fit=inla(form,family="poisson",data=inla.stack.data(stack),
  control.fixed=list(prec=2,prec.intercept=1,mean.intercept=log(avg.global)),
  E=inla.stack.data(stack)$e,num.threads=2,
  control.predictor=list(compute=TRUE))
```

The family is defined to be "poisson", as our Cox point process model is effectively discretized in practice and fitted similarly to a classical Bayesian Poisson regression. The argument `control.fixed` defines the prior distribution assumed for the intercept and fixed effects; here, the intercept has a Gaussian prior with mean set to the average number of observed landslides per pixel (taking into account the multiplicative offset) and unit precision, while fixed effects have a zero mean Gaussian prior with precision 2. The argument `num.threads=2` specifies that two threads are used for parallel computation. As INLA is memory-intensive, it is recommended *not* to exploit too many threads simultaneously. Due to the complexity of the model involving approximately 5000 latent Gaussian variables (one for each slope unit), and the large dimension of the data with around 500000 pixels, memory requirements may exceed the capacities of standard desktop computers. The total computing time to fit our model on a machine with at least 64GB of available memory was approximately 15 hours. Choosing a simpler model structure (e.g., without spatial effect), fixing (and not estimating) hyperparameters such as the precision of the spatial effect, or using faster but less accurate approximation techniques, to be specified in the argument `control.inla` of the call to `inla`, can considerably speed up the fitting procedure.

4 Results

4.1 Estimated fixed and random effects

In this section, we show how to extract results from the `inla` output in order to produce scientific plots and maps. The function `inla` returns an object (here called `fit`) of class "inla", which is an R list containing rich information about the estimated model. The usual `summary` command has been adapted for `inla` results and gives an overview of the results:

Table 1: Summary of linear covariate effects.

Covariate	Posterior Mean	95% CI	Significant
Intercept	-12.6	$[-12.7, -12.4]$	Yes
Elevation	0.01	$[-0.15, 0.18]$	No
Distance to Fault Lines	-0.05	$[-0.13, 0.03]$	No
Norm. Diff. Vegetation Index (NDVI)	0.04	$[-0.01, 0.09]$	No
Planar Curvature	0.17	$[0.14, 0.19]$	Yes
Profile Curvature	-0.12	$[-0.16, -0.09]$	Yes
Slope	0.73	$[0.67, 0.78]$	Yes
Stream Power Index	0.02	$[-0.04, 0.08]$	No
Topographic Wetness Index	0.01	$[-0.07, 0.10]$	No

```
> summary(fit)
```

More specifically, we can extract the estimated linear (or *fixed*) effects by using the following command:

```
> fit$summary.fixed
```

The `summary.fixed` object contains information on the posterior distribution for each linear covariate effect in terms of its i) *mean*, ii) *standard deviation*, iii) *2.5% quantile*, iv) *median*, v) *97.5% quantile* and vi) *mode*. Table 1 reports, for each linear covariate included in our model, the posterior mean and associated 95% pointwise credible interval (CI) defined as the interval between the 2.5% and the 97.5% posterior percentiles. When the CI excludes zero (i.e., when the 2.5% and 97.5% percentiles have the same sign), the corresponding covariate effect is deemed to be *significant* at the 95% level. From Table 1, the model intercept, planar and profile curvatures, and the slope steepness appear to be significant, while the Normalized Difference Vegetation Index (NDVI) lacks significance by little.

We can also extract the effect of categorical covariates, which are here encoded as *random* effects in R-INLA with one random variable per category. Purely categorical covariates (land form, land use and lithology) assume stochastic independence between classes, while structured or ordinal categorical covariates (here, the aspect) assume that nearby classes are dependent; recall §3.2. Similarly to the `inla` results for continuous variables with linear coefficients (*fixed effects*), the fitted object contains information about the posterior distributions of random effects. The only difference is that this is repeated for each class of these categorical covariates. To access these results, the following commands can be used:

```
> fit$summary.random$Aspect
> fit$summary.random$Landform
> fit$summary.random$Landuse
> fit$summary.random$Lithology
```

However, depending on the number of classes used to describe categorical covariates, these `summary.random` objects may involve a considerable amount of data, which may be difficult

Figure 4: *Left*: Estimated cyclic aspect effect, represented using its posterior mean (black curve) and associated 95% pointwise CI (blue curves). *Right*: Estimated land use categorical effects, represented using the posterior means (black dots) and 95% CI for each class (vertical segments). Horizontal dashed lines at zero separates positive from negative effects.

to navigate through and to understand. To better interpret the results, it is convenient to represent the estimated random effects using graphical diagnostics and maps. For example, the posterior mean and 95% CI for the aspect (modeled here as a categorical variable with 16 correlated classes) may be elegantly displayed using the following R code:

```
> aspect=fit$summary.random$Aspect
> ylim=range(aspect[,c("mean", "0.025quant", "0.975quant")])
> xvals=(c(-1:16)+0.5)/16*360; yidx=c(16,1:16,1)
> plot(xvals,aspect$mean[yidx],type="l",lwd=3,xlab="Aspect [Deg]",
      ylab="Linear predictor",ylim=ylim,xaxp=c(0,360,9))
> lines(xvals,aspect$"0.025quant"[yidx],lwd=3,col="blue")
> lines(xvals,aspect$"0.975quant"[yidx],lwd=3,col="blue")
> abline(h=0,lty=2,lwd=3,col="gray50")
```

The left panel of Figure 4 reports the results, and shows that the aspect has a highly non-linear and strongly significant effect overall. Moreover, the a priori correlation structure between nearby classes, assumed in our model for the aspect, is clearly visible in this figure, which reveals a smooth estimated pattern across classes.

Similarly, we can plot the purely categorical covariates with independent effects. For example, we can use the following code to display the estimated effects for the land use, assuming that the R object `LUnames` contains the names for each land use class:

```
> landuse=fit$summary.random$Landuse
```

```

> ylim=range(landuse[,c("mean","0.025quant","0.975quant")])
> plot(1:13,landuse$mean,xlab="CORINE Land Use",ylab="Linear predictor",
      lwd=2,ylim=ylim+c(0,0.1),pch=19,xaxt="n")
> axis(1,at=1:13,labels=1:13)
> points(1:13,landuse$"0.025quant",lwd=2,col="blue",pch=19)
> points(1:13,landuse$"0.975quant",lwd=2,col="blue",pch=19)
> segments(1:13,landuse$"0.025quant",1:13,landuse$"0.975quant",lwd=2)
> abline(h=0,lty=2,lwd=3,col="gray50")
> text(1:13-0.3,y=landuse$mean,LUnames,srt=90)

```

The right panel of Figure 4 displays the results. The main difference between the left and right panels resides in the way results are presented: In contrast to the land use, we show the estimated aspect effect by using a continuous line, rather than disconnected points and intervals, to stress that neighboring classes are linked together and assumed to be dependent.

Figure 4 is useful to display the estimated covariate effects and their uncertainty in terms of posterior means and credible intervals; however, the spatial component is lost. A way to account for this is to show the posterior means on a map. This can be accomplished using the `join by attribute table` function available in any GIS platform. The two targets of the `join` function should correspond to the shapefile of the given categorical covariate and the posterior means extracted from the fitted R-INLA object, using a common ID field. Figure 5 shows the results using a suitable color scale that discriminates positive from negative effects, therefore clearly highlighting spatial regions that contribute to slope instability.

Ultimately, the analysis of estimated fixed and random effects can be completed by checking the latent spatial effect described in §3.2, capturing the unobserved precipitation trigger. The estimate of the precision hyperparameter τ_{CAR} equals 0.22 with fairly narrow credible interval [0.20, 0.22], as can be seen from the information returned by the command

```
> fit$summary.hyperpar
```

This hints at a moderately strong spatial dependence, which expresses that we have a relatively smooth transition between neighboring slope units in general, although relatively large discrepancies between nearby slope units are still possible. We can use the following command to extract the posterior mean (i.e., the fitted values) of the latent spatial effect for each slope unit:

```
> fit$summary.random$SU.ID$mean
```

The above command returns a vector of estimated slope unit effects ordered by slope unit ID. Similarly to categorical covariates, it is convenient to visualize the estimated spatial effect on a map using GIS: we need to join the vector to the `attribute table` of the slope units polygonal shapefile and then map it. The top left panel of Figure 6 displays the results using a symmetric color scale; it reveals a negative trend from the coastline to the catchment ridges and a strong positive contribution to landslide occurrence in the central area. Thanks to the

Figure 5: Posterior means of estimated categorical covariate effects. The chosen color scale is symmetric about zero, which therefore clearly discriminates positive effects contributing to slope instability (red) from negative effects (blue).

spatial correlation across slope units, the estimated latent spatial effect varies smoothly over space.

To visualize the uncertainty surrounding the latent spatial effect, we can compute a 95% pointwise CI by extracting the 2.5% and 97.5% posterior percentiles for each slope unit using the commands:

```
> fit$summary.random$SU.ID$"0.025quant"
> fit$summary.random$SU.ID$"0.975quant"
```

The latent spatial effect is deemed to be significant at a specific location if the credible interval for the corresponding slope unit does not include zero. The top middle and right panels of Figure 6 display the results. Although some regions with few landslide events or

Figure 6: *Top:* Estimated latent spatial effect (a) with the size of a corresponding 95% pointwise credible interval (b), and the indicator whether the spatial effect is significantly positive, negative or non-significant in specific slope units (c). *Bottom:* Estimated landslide (integrated) intensity shown at the pixel (d), slope unit (e) and catchment (f) resolutions.

at the boundary of the study region have quite wide pointwise credible intervals, the latent spatial effect appears overall to be highly significant as a whole, and thus important to capture the unobserved trigger.

4.2 Estimated landslide intensity at various spatial resolutions

As mentioned in §3.2, we interpret the point process *intensity* $\lambda(s)$ as the “density” of landslides around a specific spatial location s . The integrated intensity $\Lambda(\mathcal{A}) = \int_{\mathcal{A}} \lambda(s) ds$ corresponds to the number of landslides expected to occur in a given mapping unit $\mathcal{A}$. Unlike

classical landslide susceptibility approaches based on Bernoulli regression models, whose fits strongly rely on a predefined and fixed choice of mapping units, the intensity function is continuously defined over space, and can therefore be used directly to deduce the integrated intensity over *any* areal unit of interest *without* refitting the model. **Thus, we can consistently** describe aggregated landslide counts at various spatial resolutions. In practice, the intensity $\lambda(s)$ is estimated at the highest pixel resolution, and the integrated intensity for a region $\mathcal{A}$ can be calculated by summing up the intensities of all pixels contained in $\mathcal{A}$. Thus, from the estimated pixel intensity $\hat{\lambda}(s)$, we can therefore immediately derive the estimated slope unit intensity (i.e., sum of all pixel intensities within slope units) and the estimated catchment intensity (i.e., sum of all pixel intensities within catchments). Specifically, if $\mathcal{A}$ denotes a specific slope unit or catchment, **we compute** $\hat{\Lambda}(\mathcal{A}) = \sum_{s_i \in \mathcal{A}} \hat{\lambda}(s_i)$.

We now describe how to practically extract and compute pixel, slope unit and catchment intensities using R and GIS. From the R-INLA output, fitted values for pixels can be easily accessed through the corresponding **summary** table. However, recall that in §3.1 we rescaled the area to establish a resolution-independent spatial unit of 1m^2 in the point process model. Hence, such fitted values need first to be back-transformed to the actual mapping scale. Once the intensities have been properly corrected to pixel-based values, we can assign the corresponding spatial coordinates and generate intensity maps. The pixel intensity post-processing is summarized below, assuming that the R object `xy.map` is a matrix with two columns containing the position of each pixel centroid:

```
> pixel.intensity = fit$summary.fitted.values$mean*area.pixel
> intensity.map=cbind(xy.map,pixel.intensity)
```

Once pixel intensities are computed, we can import the `intensity.map` object as a raster in GIS and plot the corresponding intensity map. This is a basic operation in any GIS platform. For instance, in SAGA GIS (Conrad *et al.*, 2015), it can be achieved by using the **Import Grid from XYZ** tool available under `/Geoprocessing/File/Grid/Import/` setting the appropriate **Cell size**, here equal to 15m. The bottom left panel of Figure 6 shows the estimated pixel intensity. Then, once the pixel intensity is loaded, we can scale it up to other mapping units by summing all values in a given polygonal partition. This operation can be done in SAGA GIS using the **Grid Statistics for Polygon** tool under `/Geoprocessing/Shapes/Grid/Grid Values/` selecting the **SUM** option. The bottom middle and right panels of Figure 6 display the intensity maps at the slope unit and catchment resolutions, respectively. As expected, the intensity increases for coarser resolutions, and the pixel intensities reveal the finest details.

4.3 Model checking and goodness-of-fit assessment

To assess the model performance, **we should always consider several diagnostics**. Some of these are already standard in the geomorphological literature. We here list and describe four possible diagnostics related to i) the observed and predicted landslide counts (using the

Figure 7: Comparison of observed and predicted landslide counts at the slope unit and catchment scales. Red lines correspond to a perfect match, whereas the gray lines correspond to the 95% CI obtained using the theoretical quantiles of a Poisson distribution for each landslide count (i.e., ignoring posterior uncertainty in the Cox point process model of §3.2).

intensities), ii) **the derived** binary presence-absence data (using the probabilities of observing at least one versus no event within specific mapping units), iii) the smoothness of the estimated latent spatial effect, and iv) the model sensitivity to prior distributions.

Our landslide inventory contains a total number of 4879 occurrences. Therefore, as the integrated landslide intensity $\Lambda(\mathcal{A})$ for a spatial unit $\mathcal{A}$ represents the *expected* number of events occurring in $\mathcal{A}$, the total estimated intensity for the entire study area should match approximately the observed total landslide count. By summing up all estimated pixel intensities, we obtain here a total intensity of 5138, which suggests that the fit is fairly good overall (taking the uncertainty into account), although this number overestimates by 5.3% the observed count. **We can also perform such a comparison** between observed and estimated landslide counts for any other choice of spatial units, and Figure 7 displays the results for slope units and catchments. Intuitively, the model should be better calibrated at coarser resolutions, and this is what we see here: on the catchment scale, the intensities estimated by our point process model are almost identical to the observed landslide counts for all catchments, while on the slope unit scale (with smaller counts), there is slightly more variability. Nevertheless, both at the slope unit and catchment scales, the model goodness-of-fit is excellent overall. Moreover, the counts are mostly contained in a 95% CI calculated from the theoretical quantiles of the Poisson distribution (ignoring posterior uncertainty), which shows that the model appropriately captures the variability of landslide counts across space. These outstanding results for our fitted model may raise legitimate concerns about potential overfitting issues and poor predictive performances. However, as described further in §4.4, **our** cross-validation study reveals that the out-of-sample predictive performance of our model is still very good, thus removing any doubts about overfitting.

Table 2: Model predictive performance in terms of AUC at the pixel, slope unit and catchment scales.

Procedure	Pixel	Slope unit	Catchment
Model fit	0.932	0.975	0.981
Cross-validation	0.894	0.927	0.968

Secondly, we can also compare the observed and estimated presence-absence probabilities, in order to check the performance of the fitted model, as is usually performed in landslide *susceptibility* studies. The aim of this comparison is to verify whether the model appropriately distinguishes stable from unstable terrain conditions. After transforming the count data to binary data representing absences (i.e., all mapping units with no landslide) or presences (i.e., all mapping units with at least one landslide), such a comparison may be achieved with standard metrics used in the geomorphology literature, such as the receiver operating characteristic (ROC) curve and the area under the curve (AUC). Thanks to the assumed Poisson distribution of counts, the probability that *at least one* landslide occurs in the region $\mathcal{A}$ can be obtained as $1 - \exp\{-\hat{\Lambda}(\mathcal{A})\}$, where $\hat{\Lambda}(\mathcal{A})$ denotes the estimated (integrated) intensity; at the finest pixel level, we should use the estimated intensity $\hat{\lambda}(s)$, which matches $\hat{\Lambda}(\mathcal{A})$. Thus, by using the library `pROC` (Robin *et al.*, 2011), we can use the following R code to compute the ROC and AUC at the pixel level:

```
> library(pROC)
> y.presence.absence=y.count
> y.presence.absence[y.count>0]=1
> pixel.probability=1-exp(-pixel.intensity)
> pixel.ROC=roc(y.presence.absence~pixel.probability)
> pixel.AUC=as.numeric(pixel.ROC$auc)
```

Similar commands can be used for other mapping units. Table 2 reports the AUC values obtained at the pixel, slope unit, and catchment scales. The results are outstanding: the AUC reaches 0.932, 0.975 and 0.981 at the pixel, slope unit, and catchment levels, respectively, which is unprecedented in the geomorphology literature, and confirms that the model calibration is better at coarser resolutions. As further explained in §4.4, the out-of-sample predictive performance is also excellent. We explain such good results by the inclusion of the latent spatial effect, which accounts for the intense precipitation trigger that is mostly unexplained by all morphometric or thematic covariates.

Thirdly, we should investigate the spatial smoothness of the estimated latent spatial effect, to check that the latter has been properly estimated using INLA and to assess its usefulness in representing the data variability over space. As a rule of thumb, the latent spatial effect should show a relatively smooth transition from one slope unit to another. In cases where the overall pattern appears very scattered or even completely random, there are three possibilities: either i) the parameterization of the spatial model component can be

improved, **or** ii) the spatial information is already carried in the model through some of the other covariates, **or** iii) the spatial field is absent in the first place, i.e., it is **useless** to represent the spatial variability of landslide counts. In view of these considerations, we recommend careful assessment of the final estimated spatial pattern to avoid misinterpretation of the physical process behind the latent spatial effect. In our analysis of the Messina disaster, the top left panel of Figure 6 shows that the estimated latent spatial effect varies quite smoothly over space and thus, being significant overall, it plays a crucial role in the fitted model.

Finally, checking the model sensitivity to the choice of prior distributions is also a mandatory requirement for constructing a reliable Bayesian model. The formulation of a Bayesian model requires to specify the form of prior probability distributions for model components such as fixed and random effects, and for hyperparameters such as the precision parameters in our model (**governing the strength of dependence between neighboring slope units or ordinal covariate categories**). The Bayesian paradigm allows us to tailor prior distributions for including expert knowledge or empirical knowledge from the data. In R-INLA, a wide range of Gaussian prior model types are proposed for modeling random effects, while the available prior choices for hyperparameters depend on the nature of the parameter. In general, the choice of the final priors should produce good results. However, good models should also be stable. If slight changes to the priors produce strong variations in the final model output or meaningless results, we may suspect instabilities arising from inappropriate model building or non-converging optimization procedures. In particular, using prior models with a relatively large number of components together with very uninformative priors may lead to a failure of optimization procedures, and to unstable estimation results. We suggest running several tests based on different choices of priors and hyperparameters, before concluding that a given model is final.

4.4 Cross-validation study and out-of-sample predictive skill

Out-of-sample validation (or cross-validation) is a procedure, whereby information contained in the original data matrix is hidden during the fitting step and subsequently exploited to estimate the predictive power of a given model at unobserved locations. Typically, many predictive methods in R implement a `predict` function to perform such validations based on a fitted model object. In contrast, estimation and out-of-sample prediction in the Bayesian framework of INLA can be achieved in a single step by simply assigning NAs to the response variable for rows of the data matrix that will be used for cross-validation, without further modifying the structure of the code. The output contained in the fitted object includes the fitted values for both observed responses and for artificially missing response values. In our case, we propose to perform *K-fold cross-validation* as described by [Petschko et al. \(2014\)](#). Specifically, we partition the data matrix (from the file `datamatrix.txt`) into K subsections, each containing approximately the same amount of rows chosen at random. We then change the landslide counts for a given subsection to NA values, and finally refit the model (hence, simultaneously predicting the NAs), making sure the fitted R objects are

Input: Response vector `y.count`; preprocessed covariate table `covar.inla`; INLA formula `form`; number of cross-validation fits `K`; other parameters.

Output: Predicted landslide pixel intensities `pred.intensity`.

```
partSU=sample(...) #Randomly partition slope units into K subsections
partition=partSU[covar.inla$SU.ID] #Corresponding partition for pixels
fits=list() #List to store all cross-validation fits
pred.intensity=c() #Vector to store all predicted pixel intensities
for i=1 to K do
  sec.i=which(partition==i) #Pixel indices for i-th subsection
  y.count.i=y.count #Clone response vector
  y.count.i[sec.i]=NA #Hide i-th subsection for fitting
  stack.i=inla.stack(y.count.i,covar.inla,...) #Create new data stack
  fit.i=inla(form,stack.i,...) #Perform i-th INLA fit
  fits[[i]]=fit.i #Store i-th cross-validation fit

  #We now extract predicted pixel intensities for the i-th subsection:
  pred.intensity[sec.i]=fit.i$summary.fitted.values$mean[sec.i]*area.pixel
end
return Predicted pixel intensities pred.intensity.
```

(Note that predicted intensities at the slope unit and catchment scales can be easily obtained from `pred.intensity` by summing up predicted pixel intensities within specific mapping units).

Algorithm 1: K -fold cross-validation procedure using R-INLA described in pseudo-code. Some function arguments are purposely missing or simplified for presentation clarity.

stored each time. From these K fitted objects, we can then reconstruct the out-of-sample predictions for the full data set by joining together the predictions of missing counts for each of the K subsections. In many geomorphological papers (e.g., [Castro Camilo et al., 2017](#)), the number of subsections K is assumed to be large, and the final susceptibility map over the given mapping units is displayed as the mean probability value across replicates. In this chapter, we apply a slight modification to this framework: we here choose $K = 4$ and randomly select one fourth of the slope units for each subsection, where each slope unit is constrained to be drawn exactly once. This K -fold cross-validation procedure is outlined in pseudo-code in [Algorithm 1](#).

Figure 7 compares the predicted intensities (i.e., expected counts) obtained from the cross-validation procedure to the observed counts, at the slope unit and catchment scales. As expected, the within-sample prediction performance has a better score than its out-of-sample counterpart, but overall there is a fairly good agreement between observed and predicted quantities in both cases, especially at the catchment scale. Table 2 reports the AUC values based on presence-absence probabilities predicted by the cross-validation procedure. The AUC reaches 0.894, 0.927 and 0.968 for the pixel, slope unit, and catchment levels. Although these values are lower than the ones obtained for the model fit based on the full

dataset, the results confirm that our model has an outstanding predictive performance at all resolutions, and **that it** does not overfit the data.

5 Discussion

We implemented our method by modeling a disastrous MORLE exclusively made of shallow flow-like landslides. We assessed the performance in terms of fitted and predicted intensities. In agreement with the current susceptibility literature, we also repeated the same procedure with respect to probabilities of presence/absence. Figure 7 demonstrates a remarkable agreement between observed and estimated landslide counts. Specifically, the fitted intensities almost perfectly describe the original data whilst the cross-validation results show a greater uncertainty and tend to slightly overestimate counts, especially for greater values. If we take the slope unit partition as an example, we can interpret this as a minor issue from a practical standpoint, for the following reasons: i) mapping units that exhibit widespread instabilities should always undergo remediation procedures and ii) the commonly applied criterion is to consider worst case scenarios for safety reasons, which may yet occur under more extreme future climatic conditions. However, when considering the more common susceptibility concept, our model produces AUC values (see Table 2) of which even the minimum 0.894 corresponds to an outstanding predictive power according to the scale of Hosmer and Lemeshow (2000). Moreover, we here investigate the intensity *in addition* to the susceptibility, and a comparable performance is rare in the susceptibility literature. This can primarily be explained through the inclusion of the latent spatial effect in the model.

Very often, sparse raingauge networks hinder the direct inclusion of the precipitation as a covariate although the rainfall discharge is the actual cause of landslides. Here, we speculate that the latent spatial effect enables us to account for the lack of this information. Aronica *et al.* (2012) have already attempted to reconstruct the spatio-temporal evolution of the storm in Messina. However, due to the presence of only two weather stations inside the affected area (Fiumedinisi and Briga, see Figure 1), the authors were only able to recognize an approximative regional pattern, in which the cloudburst is shown to travel Northeastwards. Our approach inverts this perspective. Since we do not have any reliable data on the trigger at a sufficiently fine resolution, we seek to reveal the residual dependency in the landslide pattern, leading to a spatially structured predictor component in the spatial distribution of the MORLE, which can be interpreted as a proxy for the precipitation trigger. The term “residual” refers to the remaining spatial component after having removed the one carried by the observed covariates. The latent spatial effect extracted from the data is shown in the top left panel of Figure 6. A northeastwards trend emerges, starting from a southern average zone, followed by a central sector characterized by the maximum positive latent effect and a third northernmost area where the latent effect is minimal. Our interpretation of this pattern is that the atmospheric disturbance i) has entered the southernmost sector of the study area, ii) has released the majority of the rain over the center and iii) has gently

dissipated towards the north. This pattern clearly reappears in the intensity maps shown in bottom row of Figure 6, where, regardless of the mapping unit, the central sector is exposed to a greater number of debris flows. If we assume that the latent spatial effect does carry the information of the trigger, the dominant effect shown in the intensity maps makes sense. Intuitively, a greater rainfall amount in a given area leads to a greater number of debris flow activations. In other words, the effect of the trigger tends to become dominant when compared to the other predisposing factors. Nevertheless, investigating the covariate effects is crucial for risk management purposes because it represents the only component of the model that local authorities can modify to stabilize a given slope.

Among the *fixed* covariate effects, the planar and profile curvatures, as well as the slope steepness, are significant in our model. The positive posterior mean (0.17) of the planar curvature indicates a tendency of sidewardly convex morphologies to give rise to a greater number of landslides, whereas the negative posterior mean (-0.12) of the profile curvature increases the estimated intensities for upwardly convex morphologies. These relations show good agreement with those empirically postulated by Ohlmacher (2007) in the context of flow-like landslides. Moreover, preceding studies in the same area have statistically confirmed correlations of similar order (e.g., Lombardo *et al.*, 2015). **Convex planar and profile curvatures control runoff divergence and acceleration, respectively. We interpret such effects in terms of soil cover availability (planar) and mobilization (profile).**

Slope steepness is widely known to be a key predisposing factor for shallow landslides (e.g., Donnarumma *et al.*, 2013). Here, we normalized all covariates to have zero mean and unit variance. Therefore, a posterior mean equal to 0.73 marks a greater influence of the slope on landslides compared to that of the two curvatures.

For visualization convenience, we displayed the landform classes and outcropping lithologies (*random effects*) on a map (see Figure 5), whereas for the aspect and land use we provided both their numerical and geographical representation. The aspect contributes to increasing the landslide intensity between ESE and WNW directions with a clear SSW peak (see Figure 4). Moreover, the mapped effect of the aspect in Figure 5 quite consistently marks that for every major catchment in the area the right bank of the streams is quite stable. Conversely, the left bank directly faced the approaching storm **and was directly exposed to the incoming rainfall discharge. Therefore, the model tends to estimate a greater number of landslides for West-facing slopes.** These results agree well with other studies of the same MORLE. For instance, Trigila *et al.* (2015) reports a positive effect on landslide susceptibilities for W-, SW- and S-facing slopes.

In analogy, we report the land use effect using two different plots. The static information provided in Figure 4 highlights the positive significance of sclerophyllous vegetation and moors and heathland, and the decreasing impact on intensity carried by continuous urban fabric and mixed groves. These relations are consistent with those revealed by other research in the area (Lombardo *et al.*, 2016a; Reichenbach *et al.*, 2014) and more generally in Southern Italy (Pisano *et al.*, 2017). **Sclerophyllous vegetation and moors and heathland are typically**

sparse green areas in the Mediterranean region, thus offering a greater exposure to raindrops and subsequent material mobilization. Conversely, urban fabric includes concreted area, which contributes to complete surface runoff, whereas mixed groves are more densely packed, shielding the soils from the rain. Overall, mapping the posterior means of each land use class does not show a clear match with respect to the intensities shown in Figure 6, which suggests that the land use plays a relatively minor role in the overall model.

Ultimately, we decided to graphically interpret lithological and landform influences. For the lithology, Figure 5 shows a clear positive relation to the number of debris flows from the southernmost to the central region where FDNb or phyllites to meta-arenites (see Figure 3) outcrop. Moreover, a strong negative contribution is marked for bb (see Figure 3) or recent alluvial deposits. Very similar results were obtained by (De Guidi and Scudero, 2013) (see Table 1); in that study, the authors report (for a subset of the debris flow scenario we consider in this study) that 42.9% landslides were triggered in phyllites, and no landslide occurred in the alluvium. From an interpretative standpoint, phyllites have a foliated structure which may be responsible for the initial failure along fine-grained mica flakes acting as sliding planes. As regards alluvial deposits, those can be found along the main streams and parallel to the coastline. In both cases the slope steepness is not sufficient to promote landslide activations. Moreover, landform effects are shown in Figure 5, where the main stream channels stand out with negative posterior mean values, whereas interfluvies and especially LFC9 or midslope ridges (see Figure 3) increase the estimated number of debris flows.

The interpretation of the results so far aimed at checking similarities with other researches carried out for the same MORLE. However, some differences arise since, in most of the cases, the disaster has been solely analyzed within the epicentral (e.g., Rossi and Reichenbach, 2016) or marginal (e.g., Zini *et al.*, 2015) area of our study region. One study has modeled the whole scenario (Lombardo *et al.*, 2018a); it describes our statistical approach of investigating intensities rather than susceptibility for the first time. We here provide a comparison with respect to the model Mod3 presented in Lombardo *et al.* (2018a), albeit its greater statistical complexity may produce differences compared to the results of the present chapter. The current results show a reduced number of significant linear covariates, whereas Lombardo *et al.* (2018a) report elevation, distance to fault lines and NDVI as significant. We interpret this difference to be due to the latent spatial effect, which may have partially carried some of the spatial effect of these covariates. Surely, the elevation map (see Figure 2) and the latent spatial effect (see Figure 6) show some similarities, which provides some support for the usual approach of considering elevation as a rough proxy for the precipitation pattern. As regards the final intensity maps, the results from the two models appear to be quite similar. This consideration may justify the general approach in the geomorphological literature (taking aside few examples such as in Conoscenti *et al.*, 2016) where predictors are used as linear covariates rather than nonlinear. Nonetheless, a clear benefit of the latent spatial effect is that it may always integrate certain covariate effects not captured by the other model components, at least if such effects are relatively smooth in space.

6 Conclusion

This chapter complements the methodological paper of Lombardo *et al.* (2018a): we here provide a hands-on tutorial explaining how to practically implement a model for multiple landslide events, by adopting a drastically different approach from the one that the geomorphological community has pursued in the last three decades. We switch from the landslide *susceptibility* framework to a landslide *intensity* framework, which provides information on the number of landslide events in given mapping units, additionally to where they are located. Our proposed approach is based on a continuous-space Bayesian log Gaussian Cox point process model, which can be discretized and fitted similarly to a Bayesian Poisson regression model. Using such a model, **the landslide intensity is** easily scalable from pixels to slope units, catchments, or any other mapping units of interest, without refitting a different model. We can indeed exploit the Poisson additivity by calculating sums of intensities over subareas as a consistent approach to this problem. By contrast, in classical susceptibility modeling approaches, a separate model must be built for each choice of mapping units under consideration. Therefore, the aggregation of presence-absence probabilities, as well as the representation of covariate distributions from higher to lower spatial resolutions, are much more intricate **and subjective**.

Furthermore, we introduce the use of a latent spatial effect as a proxy for the spatial pattern of the trigger, here represented by an intense storm. **This spatially correlated model component, which we define on the slope unit level to speed up computations, is able to capture the unobserved covariates that are smoothly varying over space. Therefore, its inclusion in the model usually yield an important improvement in terms of predictive performance.** This approach might be adapted to earthquake-induced landslides, which is part of our ongoing applied research activities. The estimation of the latent spatial effect is made possible thanks to the R-INLA library, which represents the core **R package that we use**. Using the few code snippets presented above, it is possible to analyze the data and reproduce the results presented in this chapter.

The complexity of our model, however, comes with a heavy computational burden, which may prevent its application to large study areas consisting of an excessively high number of pixels and slope units. So far, analyzing **data comprised of a few** million pixels with a few thousands slope units is feasible. However, treating a greater number of fine-resolution mapping units may be cumbersome, even when using high-performance computers.

In future research, we envision to perform sensitivity studies specifically over final predictive maps. We can achieve this by keeping **the** estimated covariate effects constant while simulating new, yet unobserved, spatial patterns according to either the posterior distribution of the latent spatial effect, or fully artificial scenarios for the spatial distribution of the trigger. This would allow us to investigate a range of **possible** scenarios for producing a collection of intensity or susceptibility maps.

References

- Alvioli, M., Marchesini, I., Reichenbach, P., Rossi, M., Ardizzone, F., Fiorucci, F. and Guzzetti, F. (2016) Automatic delineation of geomorphological slope units with r. slopeunits v1.0 and their optimization for landslide susceptibility modeling. *Geoscientific Model Development* **9**(11), 3975–3991.
- Aronica, G., Brigandí, G. and Morey, N. (2012) Flash floods and debris flow in the city area of Messina, north-east part of Sicily, Italy in October 2009: the case of the Giampilieri catchment. *Natural Hazards and Earth System Sciences* **12**(5), 1295.
- Atkinson, P. M. and Massari, R. (1998) Generalised linear modelling of susceptibility to landsliding in the central apennines, italy. *Computers & Geosciences* **24**(4), 373–385.
- Bakka, H., Rue, H., Fuglstad, G. A., Riebler, A., Bolin, D., Krainski, E., Simpson, D. and Lindgren, F. (2018) Spatial modelling with R-INLA: A review. Submitted, arXiv preprint:1802.06350.
- Besag, J. (1975) Statistical analysis of non-lattice data. *Journal of the Royal Statistical Society: Series D (The Statistician)* **24**(3), 179–195.
- Bivand, R. and Lewin-Koh, N. (2017) *maptools: Tools for Reading and Handling Spatial Objects*. R package version 0.9-2.
- Bivand, R. and Piras, G. (2015) Comparing Implementations of Estimation Methods for Spatial Econometrics. *Journal of Statistical Software* **63**(18), 1–36.
- Cama, M., Conoscenti, C., Lombardo, L. and Rotigliano, E. (2016) Exploring relationships between grid cell size and accuracy for debris-flow susceptibility models: a test in the giampilieri catchment (sicily, italy). *Environmental Earth Sciences* **75**(3), 1–21.
- Cama, M., Lombardo, L., Conoscenti, C., Agnesi, V. and Rotigliano, E. (2015) Predicting storm-triggered debris flow events: application to the 2009 ionian peloritan disaster (sicily, italy). *Nat Hazards Earth Syst Sci* **15**(8), 1785–1806.
- Cama, M., Lombardo, L., Conoscenti, C. and Rotigliano, E. (2017) Improving transferability strategies for debris flow susceptibility assessment: Application to the Saponara and Itala catchments (Messina, Italy). *Geomorphology* **288**, 52–65.
- Carnia, E., Suyudi, M., Aisah, I. and Supriatna, A. (2017) A review on eigen values of adjacency matrix of graph with cliques. In *AIP Conference Proceedings*, volume 1868.
- Carrara, A., Cardinali, M., Detti, R., Guzzetti, F., Pasqui, V. and Reichenbach, P. (1991) GIS techniques and statistical models in evaluating landslide hazard. *Earth Surface Processes and Landforms* **16**(5), 427–445.

- Carrara, A., Cardinali, M., Guzzetti, F. and Reichenbach, P. (1995) Gis technology in mapping landslide hazard. In *Geographical information systems in assessing natural hazards*, pp. 135–175. Springer.
- Castro Camilo, D., Lombardo, L., Mai, P., Dou, J. and Huser, R. (2017) Handling high predictor dimensionality in slope-unit-based landslide susceptibility models through LASSO-penalized Generalized Linear Model. *Environmental Modelling and Software* **97**, 145–156.
- Conoscenti, C., Rotigliano, E., Cama, M., Caraballo-Arias, N. A., Lombardo, L. and Agnesi, V. (2016) Exploring the effect of absence selection on landslide susceptibility models: A case study in Sicily, Italy. *Geomorphology* **261**, 222–235.
- Conrad, O., Bechtel, B., Bock, M., Dietrich, H., Fischer, E., Gerlitz, L., Wehberg, J., Wichmann, V. and Böhner, J. (2015) System for Automated Geoscientific Analyses (SAGA) v. 2.1.4. *Geoscientific Model Development* **8**(7), 1991–2007.
- Crozier, M. (2005) Multiple-occurrence regional landslide events in New Zealand: hazard management issues. *Landslides* **2**(4), 247–256.
- Das, I., Stein, A., Kerle, N. and Dadhwal, V. K. (2012) Landslide susceptibility mapping along road corridors in the Indian Himalayas using Bayesian logistic regression models. *Geomorphology* **179**, 116–125.
- De Guidi, G. and Scudero, S. (2013) Landslide susceptibility assessment in the peloritani mts. (sicily, italy) and clues for tectonic control of relief processes. *Natural Hazards and Earth System Sciences* **13**(4), 949–963.
- Di Stefano, P., Favara, R., Luzio, D., Renda, P., Cacciatore, M., Calò, M., Napoli, G., Parisi, L., Todaro, S. and Zarcone, G. (2015) A regional-scale discontinuity in western sicily revealed by a multidisciplinary approach:a new piece for understanding the geodynamic puzzle of the southern Mediterranean. *Tectonics* **34**(10), 2067–2085.
- Donnarumma, A., Revellino, P., Grelle, G. and Guadagno, F. M. (2013) *Slope Angle as Indicator Parameter of Landslide Susceptibility in a Geologically Complex Area*, pp. 425–433. Berlin, Heidelberg: Springer.
- Dou, J., Yamagishi, H., Pourghasemi, H. R., Yunus, A. P., Song, X., Xu, Y. and Zhu, Z. (2015) An integrated artificial neural network model for the landslide susceptibility assessment of Osado Island, Japan. *Natural Hazards* **78**(3), 1749–1776.
- Erener, A. and Düzgün, H. S. B. (2012) Landslide susceptibility assessment: what are the effects of mapping unit and mapping method? *Environmental Earth Sciences* **66**(3), 859–877.

- Greco, R., Sorriso-Valvo, M. and Catalano, E. (2007) Logistic regression analysis in the evaluation of mass movements susceptibility: The Aspromonte case study, Calabria, Italy. *Engineering Geology* **89**(1), 47–66.
- Hosmer, D. W. and Lemeshow, S. (2000) *Applied Logistic Regression*. Second edition. New York: Wiley.
- Hungr, O., Leroueil, S. and Picarelli, L. (2014) The Varnes classification of landslide types, an update. *Landslides* **11**(2), 167–194.
- Hussin, H. Y., Zumpano, V., Reichenbach, P., Sterlacchini, S., Micu, M., van Westen, C. and Bălteanu, D. (2016) Different landslide sampling strategies in a grid-based bi-variate statistical susceptibility model. *Geomorphology* **253**, 508 – 523.
- Illian, J. B., Sørbye, S. H., Rue, H. *et al.* (2012) A toolbox for fitting complex spatial point process models using integrated nested Laplace approximation (INLA). *The Annals of Applied Statistics* **6**(4), 1499–1530.
- Kennedy, I., Petley, D., Williams, R. and Murray, V. (2015) A systematic review of the health impacts of mass earth movements (landslides). *PLoS Currents* **7**(DISASTERS).
- Lee, C.-T. (2014) Statistical seismic landslide hazard analysis: An example from Taiwan. *Engineering Geology* **182**, 201–212. Special Issue on The Long-Term Geologic Hazards in Areas Struck by Large-Magnitude Earthquakes.
- Lindgren, F. and Rue, H. (2015) Bayesian spatial modelling with R-INLA. *Journal of Statistical Software* **63**(19), 1–25.
- Lindgren, F. K., Rue, H. and Lindström, J. (2011) An explicit link between Gaussian fields and Gaussian Markov random fields: the stochastic partial differential equation approach (with discussion). *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **73**(4), 423–498.
- Lombardo, L., Bachofer, F., Cama, M., Märker, M. and Rotigliano, E. (2016a) Exploiting maximum entropy method and aster data for assessing debris flow and debris slide susceptibility for the giampilieri catchment (north-eastern sicily, italy). *Earth Surface Processes and Landforms* **41**(12), 1776–1789.
- Lombardo, L., Cama, M., Conoscenti, C., Märker, M. and Rotigliano, E. (2015) Binary logistic regression versus stochastic gradient boosted decision trees in assessing landslide susceptibility for multiple-occurring landslide events: application to the 2009 storm event in messina (sicily, southern italy). *Natural Hazards* **79**(3), 1621–1648.
- Lombardo, L., Cama, M., Märker, M. and Rotigliano, E. (2014) A test of transferability for landslides susceptibility models under extreme climatic events: application to the messina 2009 disaster. *Natural hazards* **74**(3), 1951–1989.

- Lombardo, L., Fubelli, G., Amato, G. and Bonasera, M. (2016b) Presence-only approach to assess landslide triggering-thickness susceptibility: a test for the mili catchment (north-eastern sicily, italy). *Natural Hazards* **84**(1), 565–588.
- Lombardo, L., Opitz, T. and Huser, R. (2018a) Point process-based modeling of multiple debris flow landslides using INLA: an application to the 2009 Messina disaster. *Stochastic Environmental Research and Risk Assessment* To appear.
- Lombardo, L., Saia, S., Schillaci, C., Mai, P. M. and Huser, R. (2018b) Modeling soil organic carbon with Quantile Regression: Dissecting predictors’ effects on carbon stocks. *Geoderma* **318**, 148–159.
- Martins, T. G., Simpson, D., Lindgren, F. and Rue, H. (2013) Bayesian computing with INLA: New features. *Computational Statistics & Data Analysis* **67**, 68–83.
- Ohlmacher, G. C. (2007) Plan curvature and landslide probability in regions dominated by earth flows and earth slides. *Engineering Geology* **91**(2), 117–134.
- Oliveira, S. C., Zêzere, J. L., Guillard-Gonçalves, C., Garcia, R. A. C. and Pereira, S. (2017) Integration of Landslide Susceptibility Maps for Land Use Planning and Civil Protection Emergency Management. In *Advancing Culture of Living with Landslides*, eds K. Sassa, M. Mikoš and Y. Yin, pp. 543–553. Springer International Publishing.
- Opitz, T. (2017) Latent Gaussian modeling and INLA: A review with focus on space-time applications. *Journal of the French Statistical Society* **158**(3), 62–85.
- Petschko, H., Brenning, A., Bell, R., Goetz, J. and Glade, T. (2014) Assessing the quality of landslide susceptibility maps—case study Lower Austria. *Natural Hazards and Earth System Sciences* **14**(1), 95–118.
- Pisano, L., Zumpano, V., Malek, Ž., Roszkopf, C. and Parise, M. (2017) Variations in the susceptibility to landslides, as a consequence of land cover changes: A look to the past, and another towards the future. *Science of The Total Environment* **601**, 1147–1159.
- Pourghasemi, H. R. and Rossi, M. (2017) Landslide susceptibility modeling in a landslide prone area in Mazandarn Province, north of Iran: a comparison between GLM, GAM, MARS, and M-AHP methods. *Theoretical and Applied Climatology* **130**(1), 609–633.
- Reichenbach, P., Mondini, A., Rossi, M. *et al.* (2014) The influence of land use change on landslide susceptibility zonation: the briga catchment test site (messina, italy). *Environmental management* **54**(6), 1372–1384.
- Robin, X., Turck, N., Hainard, A., Tiberti, N., Lisacek, F., Sanchez, J.-C. and Miller, M. (2011) proc: an open-source package for r and s+ to analyze and compare roc curves. *BMC Bioinformatics* **12**, 77.

- Rossi, M. and Reichenbach, P. (2016) LAND-SE: a software for statistically based landslide susceptibility zonation, version 1.0. *Geoscientific Model Development* **9**(10), 3533.
- Rue, H., Martino, S. and Chopin, N. (2009) Approximate Bayesian inference for latent Gaussian models by using integrated nested Laplace approximations. *Journal of the Royal Statistical Society: Series B* **71**(2), 319–392.
- Rue, H., Riebler, A., Sørbye, S. H., Illian, J. B., Simpson, D. P. and Lindgren, F. K. (2017) Bayesian computing with INLA: A review. *Annual Review of Statistics and Its Application* **4**, 395–421.
- Schillaci, C., Acutis, M., Lombardo, L., Lipani, A., Fantappiè, M., Märker, M. and Saia, S. (2017a) Spatio-temporal topsoil organic carbon mapping of a semi-arid Mediterranean region: The role of land use, soil texture, topographic indices and the influence of remote sensing data to modelling. *Science of The Total Environment* **601**, 821–832.
- Schillaci, C., Lombardo, L., Saia, S., Fantappiè, M., Märker, M. and Acutis, M. (2017b) Modelling the topsoil carbon stock of agricultural lands with the Stochastic Gradient Treeboost in a semi-arid Mediterranean region. *Geoderma* **286**, 35–45.
- Simpson, D., Illian, J., Lindgren, F., Sørbye, S. and Rue, H. (2016) Going off grid: Computationally efficient inference for log-Gaussian Cox processes. *Biometrika* **103**(1), 49–70.
- Trigila, A., Iadanza, C., Esposito, C. and Scarascia-Mugnozza, G. (2015) Comparison of Logistic Regression and Random Forests techniques for shallow landslide susceptibility assessment in Giampileri (NE Sicily, Italy). *Geomorphology* **249**, 119–136.
- Van Den Eeckhaut, M., Reichenbach, P., Guzzetti, F., Rossi, M. and Poesen, J. (2009) Combined landslide inventory and susceptibility assessment based on different mapping units: an example from the Flemish Ardennes, Belgium. *Natural Hazards and Earth System Sciences* **9**(2), 507–521.
- Winter, M. G., Shearer, B., Palmer, D., Peeling, D., Harmer, C. and Sharpe, J. (2016) The Economic Impact of Landslides and Floods on the Road Network. *Procedia Engineering* **143**, 1425 – 1434. Advances in Transportation Geotechnics III.
- Zini, A., Grauso, S., Verrubbi, V., Falconi, L., Leoni, G. and Puglisi, C. (2015) The rusle erosion index as a proxy indicator for debris flow susceptibility. *Landslides* **12**(5), 847–859.