Seismic Hazard with Geographically Weighted Regression

Francesco Sciarretta*
School of Statistical Sciences

2023
July

Abstract

Seismic risk is a major threat to Italy's territory, financial stability, and most importantly to the safety of its inhabitants. Over the last decades a significant number of earthquakes has occurred, resulting in considerable losses of human life and economic resources. In recent years, the topic is increasingly at the center of common debate, making light on the need to counter this risk preventally rather than correctivelly after the incalculable damage has occurred. Among the most feasible solutions there are public-private partnerships between the State and insurance companies. However, to quantify the insurability of the risk, its actuarial technical basis must be studied, starting from the probability of an earthquake with a certain intensity occurring in a certain number of years. To do this, a new type of spatial regression is implemented on the Peninsula in order to asses the hazard.

*I wish to thank Davide Biancalana, Leandro D'Aurizio, Nicholas Romeo and Francesca Serra for their precious comments and suggestions.

1 Introduction

A distinctive feature of Italy is the extremely high exposure to seismic risk: first country in Europe and eighth in the world in terms of potential damage measured as a share of GDP. Since 1950, earthquakes have wiped out 5,000 lives and caused direct damages of 108 billion euros (see Cesari and D'Aurizio [6]). Natural disasters produce direct damages, human casualties, and have a significant yet hard-to-measure indirect impact on underdevelopment. The risk to Italy's housing stock is enormous: there are 34,7 million homes valued at € 5,400 billion. The risk for Italian families is intensified by the concentration of their wealth in home ownership, where 70 percent of households own their primary residence. In the face of these hazards, the insurance demand is low: natural catastrophe policies only cover slightly more than 2 percent of homes. For this kind of risk, under-insurance is a global phenomenon caused by individuals' inability to make prudent choices and the high prices charged by insurance companies in the face of reduced demand. These issues are exacerbated in Italy by a lack of insurance culture and the trust in government involvement. Damages that occurred have been compensated through Public intervention, while insurance companies have played a marginal role. This situation exposes public finance to risks and suggests a greater diffusion of the insurance instrument that has been designed for this kind of need. Earthquake insurance demand varies according to risk awareness in the country as well as costs. The international framework demonstrates examples of collaboration between public actors and insurance companies that resulted in high levels of house insurance protection, with varying legislative solutions tailored to the various nations' degrees of economic and social development. In some countries, where insurance costs rise by the level of natural risk, some programs enable inhabitants of high-risk locations to obtain coverage at a lower charge. The topic of earthquake defense is introduced in many different ways depending on the nation in which one is operating, generally as a function of available scientific knowledge, built structure, and local seismic culture. However, there are two cornerstones without which any action to mitigate earthquake risk, and consequently to transfer it to insurance companies, is doomed to failure. The first concerns the seismic hazard of the country with which one is dealing, that is an inherent characteristic of each territory that man cannot affect. The second relates to the actual vulnerability of the housing stock, that is, its overall exposure to catastrophic events: an issue that, on the other hand, is almost exclusively concerned with the actions of

man and his institutional structures (see Coviello et al.[8]). In this regard, a logical, risk-based assessment of earthquake insurance pricing has grown in relevance in recent years. The fundamental instrument in risk-based insurance pricing is catastrophic risk modeling, which simulates seismic occurrences using vulnerability criteria to calculate a logical and fair price for the damage to a client's portfolio (see Akkar et al.[1]). Traditionally, the major output of a catastrophic model is the probability distribution of estimated seismic loss for a particular portfolio. Based on the methodology developed by Friedman in 1984 (see [12]), three key factors - hazard, exposure, and vulnerability - contribute to mutually determining the amount of losses. The hazard is a natural occurrence that creates danger with a certain probability distribution, intensity, and location. The exposure is the collection of goods that can be impacted by a threat, and the vulnerability is the destructive impact of multiple hazards over various exposures. Hazard modeling is a crucial component, and there are two ways to implement it: via the historical estimation of relevant hazard measures, or through a stochastic scenario generating component, in which natural phenomena are modeled and simulated to generate a large-scale event database. This paper adopts the first strategy, following Cesari and D'Aurizio's methodology (see [7]), where the evaluation method for the seismic-risk probability is based on publicly available datasets and is suitable for insurance purposes. In particular, they determine the hazard on the dataset from *INGV* (see Section 2), linking the exceedance probability to a Poissonian model. Subsequently, the functional link between probability and earthquake's intensity in a certain area is obtained through a linear model fixed-effects panel for each geographical location.

An essential tool for hazard characterization is the modeling of spatially correlated fields of ground motion and the uncertainties associated with these elements, particularly their correlation, which is critical in determining the boundaries of the predicted risks and losses. Therefore compared to the existing models in this paper has been considered a method based on the Geographically Weighted Regression (GWR) to estimate the seismic event probability with an intensity at least equal to a level j in a return period n (see [4]). The paper is organized as follows. Section 2 describes the seismic risk, its measurement methods, and how the hazard has been computed following the Cesari and D'Aurizio's method ([7]). Section 3 introduces the GWR and section 4 presents how to estimate the probability of seismic events with the model, with the results of the application. Section 5 provides conclusions and

further possible applications.

2 Hazard Assessment

The hazard can be evaluated as the probability of catastrophic occurrences with certain severity or as the intensity of events with a given probability in a given area and over a specific time horizon. The first seismic event measurement was an empirical estimate of the global damage caused to population and buildings, using an ordinal scale known as the Mercalli Scale (see [16]), with eleven rising levels designated from I to XI. The modification by Cancani and Sieberg of this measurement (MCS, see [19]), with twelve rising levels marked from I to XII, is the most generally used scale nowadays. In English-speaking areas, a very comparable variant is known as MMI (Modified Mercalli Intensity). A second measurement was developed by C. Richter in 1935 (see [17]), the Richter Magnitude (LM or Local Magnitude), which measures in logarithmic values base 10, the maximum amplitude of the waves generated by the earthquake. In 1979, was introduced by Hanks and Kanamori the Moment Magnitude measurement (see [13]), a modification of the LM, that allows to measure the overall energy released by the earthquake. Following that, measurements of ground shaking in various micro-zones have been introduced. Peak ground acceleration (PGA) is the most commonly used measure, frequently complemented by peak ground velocity (PGV). PGA is expressed in fractions of gravity acceleration g, whereas PGV is expressed in meters per second. Both measures are strongly correlated but weakly correlated with macro-seismic intensity measurements.

The *Italian National Institute of Geophysics and Volcanology (INGV)* publishes shakemaps for each seismic event within a few hours in Italy. It is a collection of data that includes LM measurements, a map of macro-seismic intensity recorded in MCS/MMI, and a database with PGA and PGV values for all sites on the *INGV* detection grid, composed of 16.852 points with a constant step of 0.02 degrees longitude and latitude. Moreover are defined nine exceedance probabilities in 50 years with the correspondent return period and the average number of events λ in a year[1]. The *INGV* asses in this framework (see [15]) the seismic risk, that uses to compute buildings'

[1]Exceedance probabilities:= {2,5,10,22,30,39,50,63,81}
Return periods (λ^{-1}):= {2475,975,475,201,140,101,72,50,30}

resilience to earthquake events. The methodology derives sixteen geographical distributions for the PGA for each exceedance probability and each point of the grid, each obtained by combining all levels of the following:

- different levels of completeness of the historical earthquake catalogs used (2 levels);
- different methods of determining seismic intensity (2 levels);
- different measurements of earthshaking attenuation (4 levels).

For each local distribution is given a weight that reflects the confidence level in the particular approach. The weighted 16th, 50th, and 84th percentiles are finally computed from the sixteen alternative values received for each place on the map. The median is the average assessment, with the 16th and 84th percentiles providing an optimistic and pessimistic view of the local seismic risk, respectively. The *INGV* computes then the hazard in a certain zone in a certain year through the PGA_z value that can be exceeded in 50 years with a 10% probability. Using 13-class categorization with a different colour associated *INGV* provides three maps for the three percentiles described above.

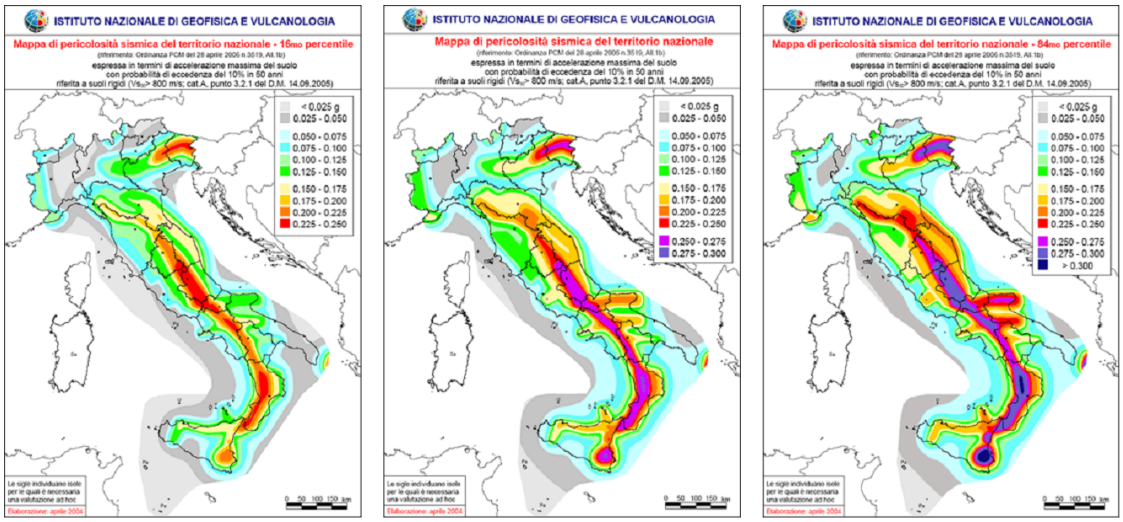


Figure 1: *Maps of Italy's seismic risk • 10% exceedance probability over 50 years • 16th, 50th and 84th percentiles*
Source: Italian National Institute of Geophysics and Vulcanology

However, the *INGV* probability formulation, is inadequate for insurance pricing purpose. A better one can be presented as follows:

Given t as time measure in $[0, T]$ and h_j, a random variable in $[0, H]$ where H is the maximum seismic intensity that occurs in a certain area and j is the type of indicator of the intensity of a seismic event. If j corresponds to the Mercalli-Cancani-Sieberg (MCS) scale, then for

$$h_j = \begin{cases} j = 1 : MCS = I, II, III, IV, V; \\ j = 2 : MCS = VI \\ j = 3 : MCS = VII \\ j = 4 : MCS = VIII \\ j = 5 : MCS = IX \\ j = 6 : MCS = X \\ j = 7 : MCS = XI, XII \end{cases}$$

Hence is defined as $p_z(h_j)$ the probability of having an event of h_j intensity in $[0, T]$ in a certain area z, with $z =(1, ..., 16.852)$.
For insurance pricing purposes, one will compute for a return period n and an intensity $h_j = \overline{h_j}$

$$p_z(\overline{h_j}) = P_z([\sum_{t=1}^{n} h_{j_t} > \overline{h_j}] \geq 1) \tag{1}$$

To compute (1) is presented the IVASS model [6]

IVASS model
The first step is estimating the MCS intensity from the corresponding PGA for all the *INGV* points. In order to do so, MCS evaluation is determined following the *INGV* approach (see Faenza and Michelini [10]). The method requires firstly an estimate of the PGV values, that have been obtained through the linear relation between PGA and PGV, estimated from an OLS model based on a collection of seismic events that occurred from 2009 to 2017. Data are extracted from the ShakeMap on the *INGV* website. The model has the following structure:

$$ln(PGV_k) = \beta_0 + \beta_1 ln(PGA_k) + \varepsilon_k \tag{2}$$

Then for each PGA result:

$$\hat{PGV}_{z;n;p_z} = \hat{c}e^{\hat{\beta}_0+\hat{\beta}_1 ln(PGA_{z;n;p_z})} \tag{3}$$

Where $\hat{c} = \dfrac{\frac{1}{T}\sum_{k=1}^{T} PGV_k}{\frac{1}{T}\sum_{k=1}^{T} e^{\hat{\beta}_0+\hat{\beta}_1 lnPGA_k}}$ is a correction factor which eliminates the bias generated by modeling the logarithm of a dependent variable.
Then one obtains a matrix of $16.852x9$ PGV estimates, where the nine estimates for every point of the grid correspond to the available exceedance probabilities. From the $INGV$ model ([10]) one can produce an estimate for h_j=MCS whenever a seismic event occurs. The model is based on two equations resulting from an orthogonal regression (see Boggs and Rogers [3]):

$$MCS_{(PGA)} = (1,68 \pm 0,22) + (2,58 \pm 0,14) log_{10}(PGA) \tag{4}$$

$$MCS_{(PGV)} = (5,11 \pm 0,07) + (2,35 \pm 0,09) log_{10}(PGV) \tag{5}$$

For each PGA and PGV (4) and (5) determine a symmetric interval of $MCS_{(PGA,PGV)}$. To determine the true value for the MCS, the model uses the distance as a credibility measure, based on a threshold value of 6 for $h_j = MCS$: between the two predictions is chosen the one more distant from 6. Are now available for each point of the grid and each exceedance probability three value for MCS.

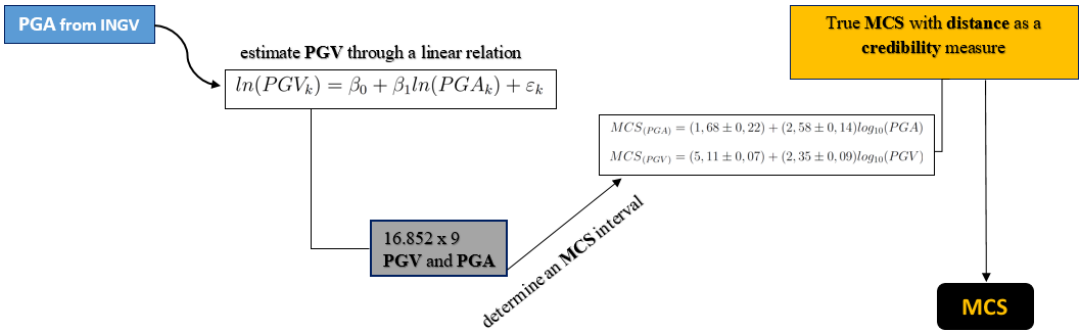


Figure 2: *MCS workflow*

Indicating with $h_{j\,z,\alpha}$ the central value of the interval, and with λ_z the yearly event frequency, where z is a generic point of the grid and α is a generic

exceedance probability, then under the Poisson specification the exceedance probability in n years is:

$$\alpha_{z,n} = P(N(n) > 0) = 1 - P(N(n) = 0) = 1 - e^{-n\lambda_z} \tag{6}$$

In order to obtain $\alpha_{z,n}$, λ_z has been modeled on h_j. The best-fitting model is log-linear (see [7]) and has the following equation:

$$ln(\lambda_z) = \beta_0 + \beta_{1,z} + \beta_2 h_{j_z} + \varepsilon_z \tag{7}$$

For an n-year horizon, from equations (6) and (7), given a generic $\overline{h_j}$, the local exceedance has the following expression:

$$\alpha_{z,n}(\overline{h_j}) = 1 - e^{-n(\hat{c}e^{\hat{\beta}_0 + \hat{\beta}_{1,z} + \hat{\beta}_2 \overline{h_j}})} \tag{8}$$

An alternative model for (7) is now presented in order to use the spatial relation between the coordinates (see Section 1).

3 Geographically Weighted Regression

Geographically Weighted Regression (see Brunsdon et al.[4]) is a technique introduced for exploring the phenomenon of regression models where the cases are geographical locations and regression coefficients couldn't remain fixed over space. Linear regression has been often used in quantitative geography papers as a way of examining the relationship between geographical factors. However, the method itself does not consider location when examining the correlations between variables.

Given the well-known components of a regression model, with X a matrix containing the set of independent variables, y a vector of dependent variables and their relation as:

$$y_i = \sum_l X_{il}\beta_l + \varepsilon_i \tag{9}$$

where β is a vector of regression coefficients and ε is a random vector whose distribution is Gaussian. This method could be applied to geographical data where each case corresponds to a location, but beyond this implicit accordance space doesn't play a role in the modeling process. Nevertheless, there might

be some circumstances where the nature of these models isn't fixed over space. This is referred to as spatial non-stationarity. Several ways for incorporating space have been considered, starting from the Cassetti expansion method (see [5]) where coefficients are expressed as explicit functions of the spatial location of the cases:

$$y_i = \sum_l X_{il}\beta_l(p_i) + \varepsilon_i \tag{10}$$

The underlying concept of the Geographically weighted regression begins from (10) and provides estimates of $\beta_l(p_i)$ for each variable l and each geographical location i, achieved by considering data for places near location p_i. This could be done by calibrating an OLS regression model based on observations whose geographical location is within the circle drawn around p_i with a certain radius r. Then the β_l are an estimate of the association between the variables in and around p_i. Evaluating β_l for each p_i one could obtain a set of estimates of spatially varying parameters.
The regression centered in each p_i could be thought as a weighted ordinary least squares regression, with observation within the circle weighted 1, others 0.

$$a_{ic} = \begin{cases} 1 & d_{ic} \leq r \\ 0 & otherwise \end{cases}$$

where d_{ic} is the distance between locations i and c. Nevertheless, the weighting function could be expressed with a continuous function, Gaussian distance for example:

$$a_{ic} = e^{\frac{-d_{ic}^2}{2b^2}} \tag{11}$$

where b will be discussed later. Alternative functions of (11) could be used, and these will be referred to as kernels K as $a_{ic} = K(d_{ic})$. Consequently, after selecting a weighting function, the estimate of β_i may be changed to produce

$$\beta_i = (X^T W_i X)^{-1} X^T W_i y \tag{12}$$

Where the matrix W_i is a diagonal matrix, whose diagonal elements correspond to the weights calibrated around point p_i.

$$W_i = \begin{bmatrix} a_{i1} & 0 & \dots & 0 \\ 0 & a_{i1} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & a_{in} \end{bmatrix}$$

Thus (12) is an array of equations with each β_i corresponding to the matrix whose elements are β_{il}. After computing each a_{lc}, the β-matrix may be calculated column by column by repeatedly using formula (12) for each i. Finally, the remaining issue in GWR is choosing b from equation (11), sometimes referred to as the kernel bandwidth, that can affect substantially the β-matrix. Least squares cross-validation is one of the methods recommended here. Assuming that for a predetermined kernel function, $\hat{y}_i(b)$ represents the predicted value of y_i from GWR as a function of b, then:

$$CV(b) = \sum_i [y_i - \hat{y}_i(b)]^2 \tag{13}$$

Hence is chosen b to minimize equation (13). Because n regressions must be fitted at each stage, choosing the bandwidth can be difficult. Alternative solutions, such as adaptive bandwidths, are available, but they are frequently more compute-intensive.
Geographically weighted regression (GWR) is then a technique used to identify non-stationarity on a map, that is when locally weighted regression coefficients deviate from their global values. It is based on the idea that the fitted coefficient values of a global model, fitted to all data, may not sufficiently capture precise local variations in data. GWR does not look for local variation in data space but rather moves a weighted window over the data, estimating one set of coefficient values at each chosen fit point. The fit points are frequently, but not always, the points at which observations were made. If the local coefficients fluctuate in space, it can be indicated that the system is non-stationary (see Bivand [2]). Discussion on the Geographically Weighted Regression are available in [11], [18] and [20], in [14] are provided examples.

4 Estimation of earthquake's probability via GWR

To compute the GWR the *spgwr* package in R is used (see [2]). Firstly is determined the bandwidth over the collection of $16.852x9$ $INGV$ points, the function finds a bandwidth for a given geographically weighted regression by optimizing a selected function. Latitude and longitude from $INGV$ grid are used as coordinates, the geographical weighting function chosen is the Gaussian and the cross-validation method is used for the optimization. In cross-validation has been scored the root mean square prediction error, where

in the *INGV* grid a minimum was found. The regression is then computed taking from (6) $ln\lambda_{z,\alpha}$ as the dependent variable and the MCS in each point and for each probability as the only independent variable:

$$ln\lambda_{z,\alpha} = \beta_{0_{z,\alpha}}(lat; lon) + \beta_{1_{z,\alpha}}(lat; lon)h_{j_{z,\alpha}} + \varepsilon_{z,\alpha} \tag{14}$$

The *gwr* function returns a Spatial Points Data Frame with fit points, weights, coefficient estimates, R-squared, and coefficient standard errors for each couple of coordinates and for the nine return periods. An intercept was found and the mean over the local R-squared values is 92.3%, and the global R-squared is 86.1%. The goodness of fit reached with this method is higher than the one obtained with equation (7).

```
Call:
gwr(dati$log_lambda ~ dati$mcs, data=dati, coords = cbind(dati$Lat, dati$Lon), bandwidth=bw,longlat=TRUE)

> summary(gwr$SDF)
Object of class SpatialPointsDataFrame
Coordinates:
            min     max
coord.x 35.3303 47.0779
coord.y  6.6915 18.9201

Number of points: 151668
Data attributes:
     sum.w        (Intercept)        dati$mcs           gwr.e                pred              localR2
 Min.   :18.70   Min.   :-4.610   Min.   :-3.2985   Min.   :-2.9974360   Min.   :-7.9666  Min.   : 0.0000014
 1st Qu.:76.97   1st Qu.: 5.560   1st Qu.:-2.2044   1st Qu.:-0.1680808   1st Qu.:-6.2857  1st Qu.: 0.9019503
 Median :81.29   Median : 7.534   Median :-1.9669   Median : 0.0254992   Median :-5.1509  Median : 0.9480184
 Mean   :80.06   Mean   : 7.694   Mean   :-1.9335   Mean   : 0.0001362   Mean   :-5.2572  Mean   : 0.9234939
 3rd Qu.:85.67   3rd Qu.:10.967   3rd Qu.:-1.7703   3rd Qu.: 0.1861022   3rd Qu.:-4.3130  3rd Qu.: 0.9856828
 Max.   :89.18   Max.   :20.481   Max.   :-0.3143   Max.   : 1.8931179   Max.   :-0.9252  Max.   : 0.9987253
```

Figure 3: *Summary gwr Spatial Data Frame from the spgwr package*

To have a concise summary of the model, one can plot over a 50-year horizon, for the nine probabilities used by the *INGV*, identified by the knots on the graph (see Figure 4), respectively:
i) the average MCS_z and the average yearly frequency λ_z;
ii) the average exceedance probability α_z and the average yearly frequency λ_z.

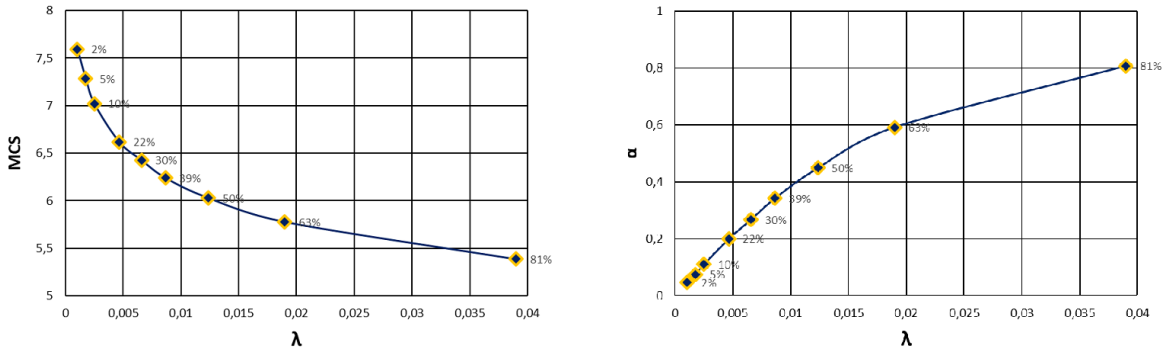


Figure 4: *Average predictions of the model over a 50-year horizon*

Figure 4 can be seen for a specific exceedance. In the first plot, for example, $j = 22\%$ corresponds to an average frequency less than $0,005$ and an average intensity greater than $6,6$. In the second plot, for the same j and the same average frequency, the average exceedance $\alpha(MCS)$ (see equation 6) is nearly $0,2$. Results are consistent with what was obtained in Cesari and D'Aurizio's (see [7]).

From an n-year horizon (see equation 6), having a λ for each point of the grid and each return period, one can determine for a fixed $h_j{=}\overline{h_j}{=}\overline{MCS}$ the local exceedance probability (see equation 1), which can be analyzed for each zone of Italy in an insurable return period i.e. 10 years. In the table included in Figure 5, the probability curve for a 10 years return period is presented for four fixed levels of hazard $\overline{h_j}{=}h_{2,\ldots,5}$, namely $MCS = VI, VII, VIII, IX$; where the meaning of the four MCS is the following:

- VI. Strong: Felt by all, and many are frightened. Some heavy furniture is moved; a few instances of fallen plaster occur. The damage is slight.
- VII. Very Strong: Damage is negligible in buildings of good design and construction; but slight to moderate in well-built ordinary structures; damage is considerable in poorly built or badly designed structures; some chimneys are broken. Noticed by motorists.
- VIII. Severe: Damage is slight in specially designed structures; considerable damage in ordinary substantial buildings with partial collapse.

Damage is great in poorly built structures. Fall of chimneys, factory stacks, columns, monuments, and walls. Heavy furniture overturned. Sand and mud are ejected in small amounts. Changes in well water. Motorists are disturbed.

- IX. Violent: Damage is considerable in specially designed structures; well-designed frame structures are thrown off-kilter. Damage is great in substantial buildings, with partial collapse. Buildings are shifted off foundations. Liquefaction occurs. Underground pipes are broken.

	Exceedance Probability Distribution														
MCS	0%	10%	20%	30%	40%	50%	60%	70%	80%	90%	100%	MEAN	STD. DEV.	MIN	MAX
VI	0,05%	2,02%	3,73%	6,38%	10,51%	17,80%	29,02%	53,72%	72,26%	87,44%	100,00%	33,24%	32,98%	0,05%	100,00%
VII	0,02%	0,47%	0,74%	1,11%	1,79%	2,89%	4,70%	7,91%	11,83%	17,70%	52,11%	6,39%	7,58%	0,02%	52,11%
VIII	0,01%	0,08%	0,15%	0,23%	0,32%	0,44%	0,65%	0,90%	1,29%	1,94%	3,77%	0,76%	0,81%	0,01%	3,77%
IX	0,00%	0,01%	0,02%	0,04%	0,05%	0,07%	0,09%	0,12%	0,15%	0,23%	0,62%	0,10%	0,10%	0,00%	0,62%

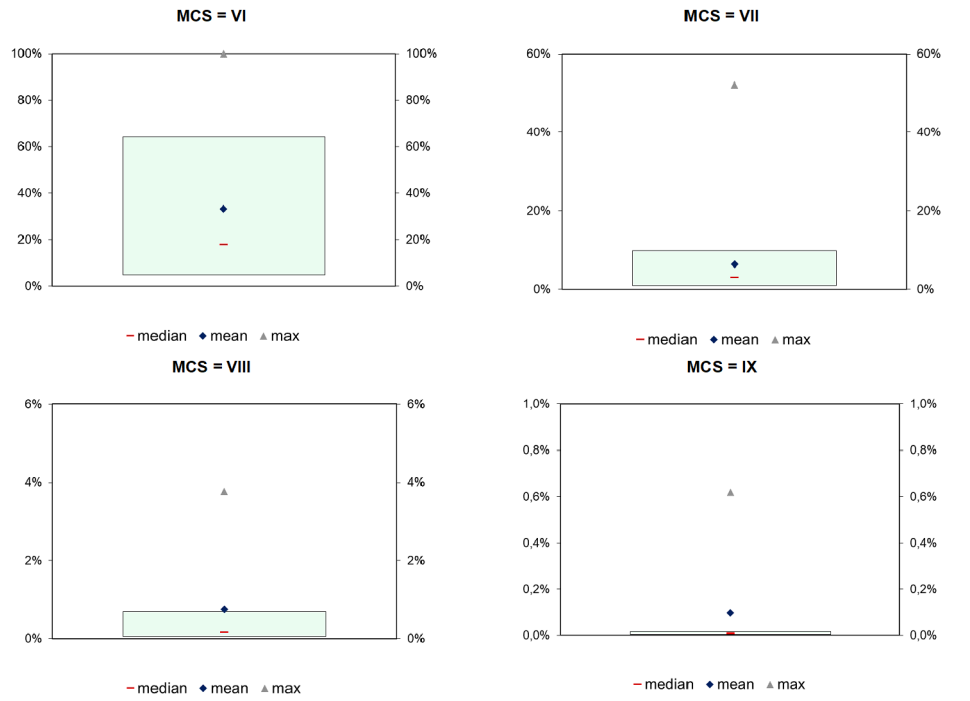
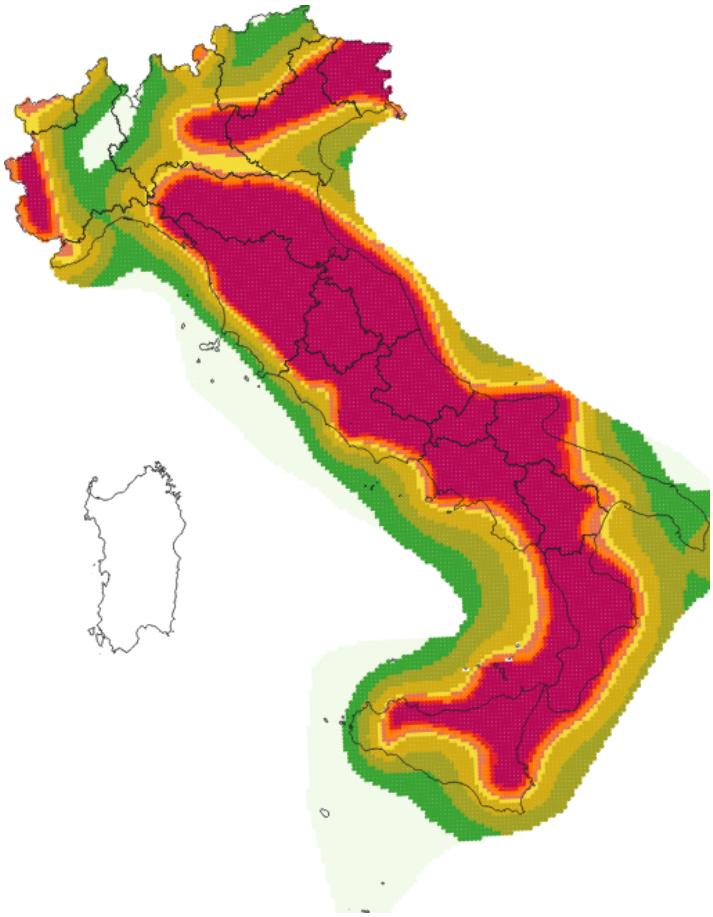


Figure 5: *Quantiles, mean, standard deviation and maximum with relative box-plot representation for MCS=VI, VII, VIII, IX*

In Figure 5 the average probability for $MCS = VI$ is $33,24\%$, considerably greater then the average for $MCS = VII, VIII, IX$, respectively equal to $6,39\%$, $0,76\%$, and $0,10\%$. Moreover, the median for $MCS = VI$ is roughly half the mean, and high values for the probability are not negligible: above the 20% of the distribution has a probability greater than 70%. As expected probabilities are lower for greater MCS, it could be seen also from the box-plots where the more MCS intensity increase, the more the probability distribution is concentrated around lesser probabilities.

Following the Ivass Risk Dashboard approach (see D'Aurizio and Sacco [9]), are defined ten risk-levels for the probability distribution for each level of intensity $\overline{MCS}$ considered. In the following pages, the risk mapping is on Italy's map for a return period of 10 years. As can be seen in the maps (see Figures 6,7,8,9), to higher intensities corresponds smaller areas. The riskier area is along the south-central Apenines with some territories in Friuli Venezia-Giulia. The charts show geographical consistency with what expected and are comparable with $INGV$ studies (see Meletti et al.[15] and Figure 1) and with the Ivass'results (see [7]). From the analysis of the maps, one can observe that more than 30 percent of Italian sites have a probability greater than 50 percent that an earthquake with an intensity equal at least to VI on the MCS scale will occur in 10 years, while in the same horizon, about a quarter have a probability over 10 percent that an earthquake of intensity at least VII will occur. The probability that earthquakes with intensity of at least VIII and IX will occur is significantly lower. The 9 percent of $INGV$ sites have a probability greater than 2 percent that an earthquake with at least VIII intensity will occur in 10 years, where the high-risked area is smaller and concentrated around the Apenines and the Friuli Venezia-Giulia region; furthermore, there are only 74 grid coordinates with a probability greater than 0,5 percent that an earthquake with at least IX intensity will occur, and the riskier area is in Sicily with the Etna's territory and the Strait of Messina. Comparing the model's results with the fixed-effects panel proposed by Cesari and D'Aurizio (see [7]), one can see higher probabilities in the GWR model for MCS equal to VI and VII, while they are lower for MCS VIII and IX. Overall, the two models are comparable and produce very similar results.

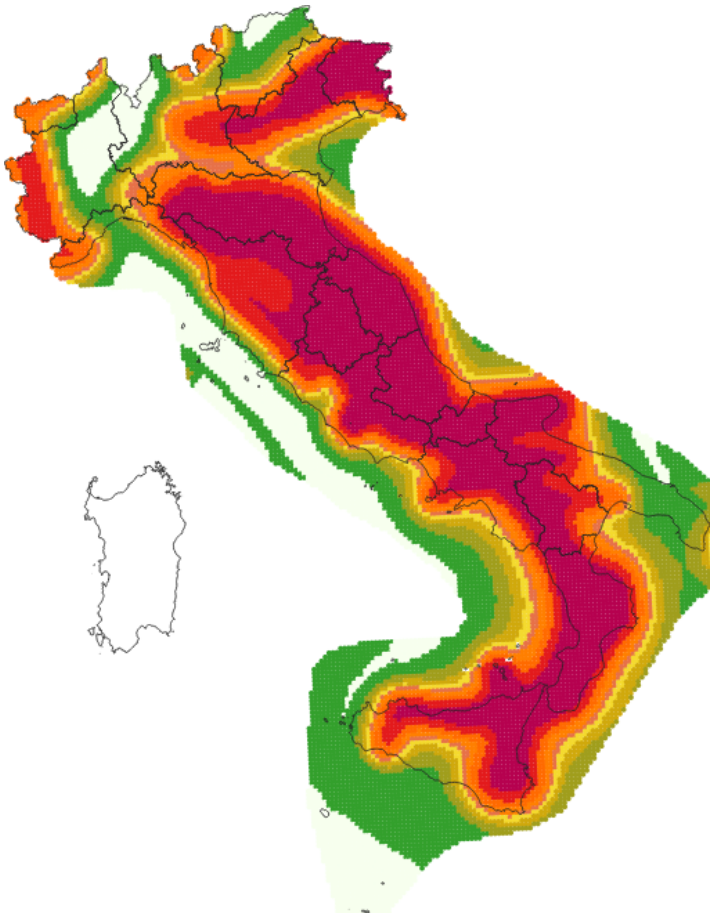
Probability risk map, VI MCS



MCS VI			
Value From	Vaue To	Risk Value	%N.Points
0,0%	2,5%	1	13,7%
2,5%	5,0%	2	11,4%
5,0%	10,0%	3	13,8%
10,0%	20,0%	4	13,2%
20,0%	25,0%	5	4,6%
25,0%	30,0%	6	3,8%
30,0%	35,0%	7	2,7%
35,0%	40,0%	8	2,0%
40,0%	50,0%	9	3,3%
50,0%	100,0%	10	31,4%

Figure 6: *Probability distribution of a seismic event of intensity equal at least to VI MCS in a return period of 10 years*

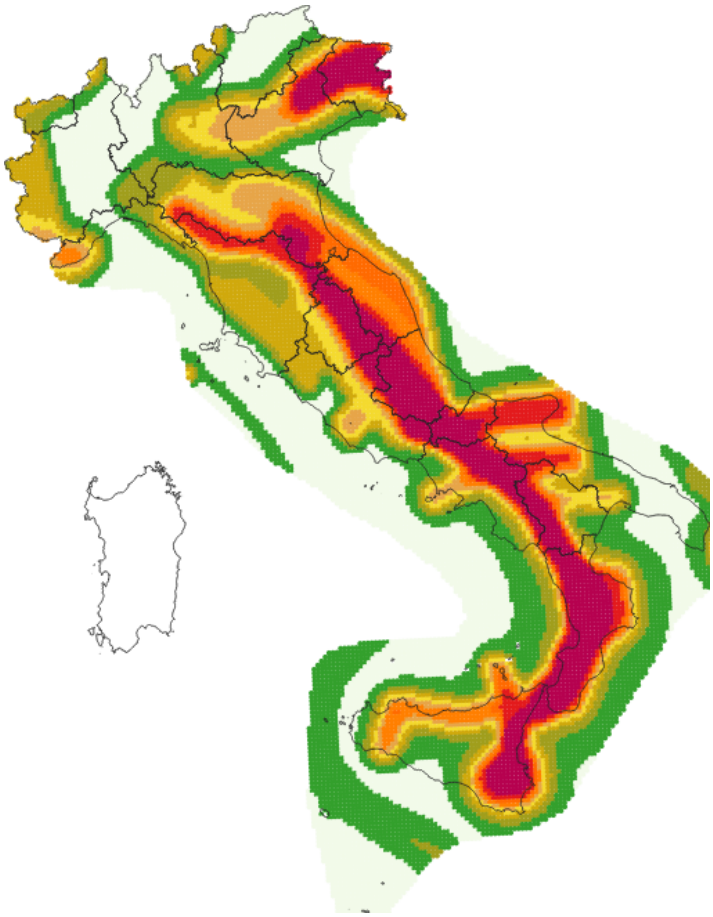
Probability risk map, VII MCS



MCS VII			
Value From	Vaue To	Risk Value	%N.Points
0,0%	0,5%	1	11,4%
0,5%	1,0%	2	16,7%
1,0%	1,5%	3	8,4%
1,5%	2,0%	4	6,0%
2,0%	2,5%	5	4,5%
2,5%	3,0%	6	3,8%
3,0%	4,0%	7	6,2%
4,0%	6,0%	8	7,6%
6,0%	10,0%	9	11,0%
10,0%	100,0%	10	24,5%

Figure 7: *Probability distribution of a seismic event of intensity equal at least to VII MCS in a return period of 10 years*

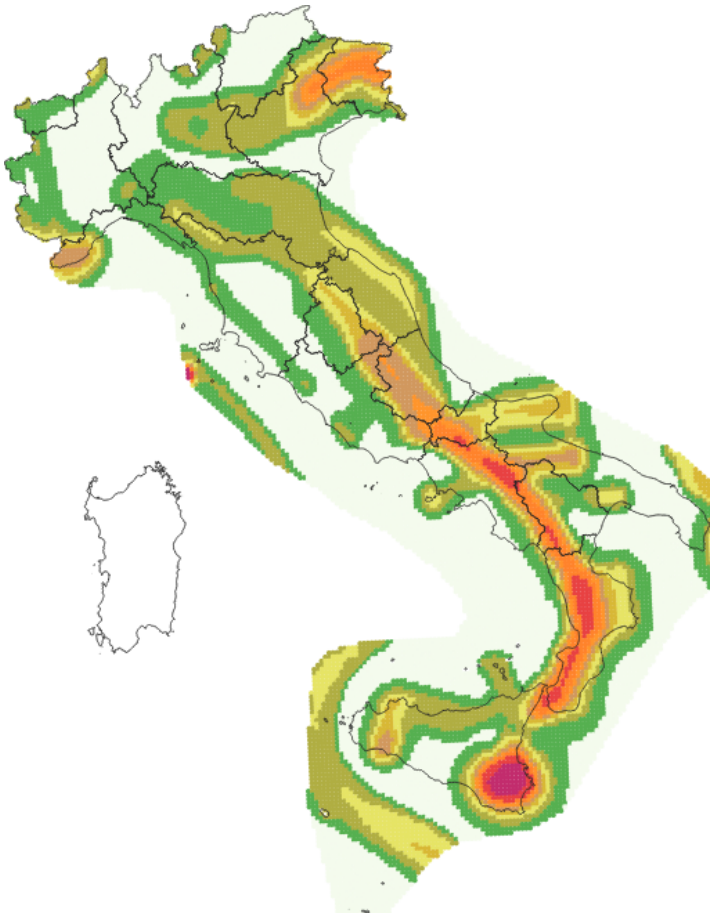
Probability risk map, VIII MCS



MCS VIII			
Value From	Vaue To	Risk Value	%N.Points
0,0%	0,2%	1	27,7%
0,2%	0,4%	2	21,0%
0,4%	0,6%	3	9,9%
0,6%	0,8%	4	8,7%
0,8%	1,0%	5	6,0%
1,0%	1,2%	6	5,3%
1,2%	1,4%	7	4,2%
1,4%	1,6%	8	3,3%
1,6%	2,0%	9	4,6%
2,0%	100,0%	10	9,4%

Figure 8: *Probability distribution of a seismic event of intensity equal at least to VIII MCS in a return period of 10 years*

Probability risk map, IX MCS



MCS IX			
Value From	Vaue To	Risk Value	%N.Points
0,00%	0,05%	1	41,0%
0,05%	0,10%	2	21,3%
0,10%	0,15%	3	17,3%
0,15%	0,20%	4	7,7%
0,20%	0,25%	5	3,8%
0,25%	0,30%	6	3,1%
0,30%	0,35%	7	2,0%
0,35%	0,40%	8	2,1%
0,40%	0,50%	9	1,2%
0,50%	100,00%	10	0,4%

Figure 9: *Probability distribution of a seismic event of intensity equal at least to IX MCS in a return period of 10 years*

For instance, in the hypothesis of one-year insurance cover, are presented the probabilities of the 5 most dangerous municipalities for $MCS = VII, VIII$

MCS VII			**MCS VIII**		
Municipalities	**Province**	**Pr** (%)	**Municipalities**	**Province**	**Pr** (%)
Zafferana Etnea	Catania	7,09	Aprigliano	Cosenza	0,383
Sant'Alfio	Catania	6,25	Marzi	Cosenza	0,382
Trecastagni	Catania	6,23	Soveria Mannelli	Catanzaro	0,381
Milo	Catania	6,11	Spezzano della Sila	Cosenza	0,380
Nicolosi	Catania	5,83	Serrastretta	Catanzaro	0,379

Table 1: Municipalities for highest earthquake probability in a year, MCS= VII and VIII

For $MCS = VII$, the first ten municipalities are in Etna's area. However, among the first 100 municipalities, above 75% are in the province of L'Aquila, the remaining in the province of Rieti, Perugia, and Frosinone. In the same time horizon, for $MCS = VIII$, the most dangerous municipalities are in the Calabrian Apennines, in the provinces of Cosenza and Catanzaro. Among the first 100 municipalities, as for the inferior degree, there are the provinces of Isernia, L'Aquila, and Rieti. The Vercelli's province is the safest in terms of probability of occurrence (respectively $0,02\%$ for MCS equal at least to VII, $0,002\%$ for MCS equal at least to VIII).
A different ranking for the municipalities is obtained by changing the MCS intensity (see Table 1). This unique result stems from the advanced framework of the GWR, which elaborates a regression in every municipality for every return period. This property could be desirable in the insurance pricing phase and differs from results obtained with the fixed-effects panel model (7).

5 Conclusions

The model presented is overall well-fitted with comforting statistical diagnostics, also results are consistent with applications already in the literature on

the same subject. GWR is suitable for the study of seismic hazard in Italy thanks in particular to the large amount of information made available by *INGV*.

Once computed the seismic hazard it is necessary to proceed in a further step to determine the seismic risk in the country: to define an appropriate vulnerability function to be applied to the country's exposure.
Also, it should be noted that such a policy if complemented with other catastrophe coverage (such as floods), thanks to the possibility of geographically diversifying the risk, the insurance premium for a natural catastrophe policy for Italian housing could be both effective and affordable, especially where renovation and remediation of the real estate of Italian households is also conducted.

References

[1] Sinan Akkar, Alper Ilki, Caglar Goksu, and Mustafa Erdik. *Advances in assessment and modeling of earthquake loss*. Springer Nature, 2021.

[2] Roger Bivand. Geographically weighted regression. *CRAN Task View: Analysis of Spatial Data*, 2017.

[3] Paul T Boggs and Janet E Rogers. Orthogonal distance regression. *Contemporary mathematics*, 112:183–194, 1990.

[4] Chris Brunsdon, Stewart Fotheringham, and Martin Charlton. Geographically weighted regression. *Journal of the Royal Statistical Society: Series D (The Statistician)*, 47(3):431–443, 1998.

[5] Emilio Casetti. Generating models by the expansion method: applications to geographical research. *Geographical analysis*, 4(1):81–91, 1972.

[6] Riccardo Cesari and Leandro D'Aurizio. Quaderno n. 13. 2019.

[7] Riccardo Cesari and Leandro D'Aurizio. From earthquake geophysical measures to insurance premium: A generalised method for the evaluation of seismic risk, with application to italy's housing stock. *Asia-Pacific Journal of Risk and Insurance*, 16(2):155–185, 2021.

[8] Antonio Coviello, Renato Somma, Adriano Giannola, and Gianluca Valensise. I rischi catastrofali azioni di mitigazione e gestione del rischio. 2021.

[9] Leandro D'Aurizio and Silvia Sacco. Quaderno n. 26. 2023.

[10] Licia Faenza and Alberto Michelini. Regression analysis of mcs intensity and ground motion parameters in italy and its application in shakemap. *Geophysical Journal International*, 180(3):1138–1152, 2010.

[11] A Stewart Fotheringham, Chris Brunsdon, and Martin Charlton. *Geographically weighted regression: the analysis of spatially varying relationships*. John Wiley & Sons, 2003.

[12] Don G Friedman. Natural hazard risk assessment for an insurance program. *Geneva Papers on Risk and Insurance*, pages 57–128, 1984.

[13] Thomas C Hanks and Hiroo Kanamori. A moment magnitude scale. *Journal of Geophysical Research: Solid Earth*, 84(B5):2348–2350, 1979.

[14] Christopher D Lloyd. *Local models for spatial analysis*. CRC press, 2010.

[15] C Meletti, V Montaldo, M Stucchi, and F Martinelli. Database della pericolosità sismica mps04. *Istituto Nazionale di Geofisica e Vulcanologia (INGV)*, 2006.

[16] G Mercalli. Intensity scales. *Bollettino della Societ Sismologica Italiana*, 8:184–191, 1902.

[17] Charles F Richter. An instrumental earthquake magnitude scale. *Bulletin of the seismological society of America*, 25(1):1–32, 1935.

[18] O. Schabenberger and C.A. Gotway. *Statistical Methods for Spatial Data Analysis*. Chapman and Hall/CRC, 2005.

[19] A Sieberg. Scala mcs (mercalli-cancani-sieberg). *Geologie der Erdbeben, Handbuch der Geophysik*, 2(4):552–555, 1930.

[20] Lance A Waller and Carol A Gotway. *Applied spatial statistics for public health data*. John Wiley & Sons, 2004.