POLITECNICO DI MILANO
Corso di Laurea Magistrale in Ingegneria Matematica
Scuola di Ingegneria Industriale e dell'Informazione

A class of geostatistical methods to predict and simulate seismic ground motion fields: from a univariate to a functional approach

Supervisor: Dr. Alessandra Menafoglio
Co-supervisor: Dr. Sara Sgobba

Tesi di Laurea di:
Filippo Lentoni
Matricola 878653

Anno Accademico 2017-2018

Abstract

Earthquakes have been constantly studied due to their devastating effects in terms of loss of human life and economic damages. In order to limit these effects and investigate the nature of this phenomenon, in literature, the generation of shaking fields has been proposed based on empirical models which predict the ground motion (Ground Motion Prediction Equation, GMPE). The variance related to the prediction error of these models and the spatial covariance of their residuals have been also studied, with the purpose of understanding if GMPEs are able to describe exhaustively the phenomenon and finding the possible corrective terms. Thanks to several records of waveforms collected during the seismic sequence of Emilia in 2012, it was possible to model the prediction error of a GMPE specific for the Northern part of Italy with a fully non-ergodic approach which identifies a systematic corrective term in the residuals for the median prediction of the model.

The aim of this thesis is to build shaking fields through the combination of the prediction of the aforementioned GMPE, the prediction and the simulation of its corrective term and of its residual uncertainty with both the univariate and the functional approaches to the geostatistics. The application of this last approach to Applied Seismology is innovative and allows to provide predictions and joint stochastic simulations of many intensity measures through computational and modeling efforts comparable to the ones of the multivariate approach. The thesis shows that for some relevant intensity measures the performances of the functional approach on the predictions are comparable or better than the ones of the univariate approach, thus providing more complete and more robust results.

Sommario

I terremoti sono da sempre oggetto di studio per via degli effetti devastanti che provocano in termini di perdite di vite umane e danni di natura economica. Al fine di limitare tali effetti e studiare la natura di questi fenomeni, in letteratura è stata proposta la generazione di mappe di scuotimento del suolo basate su modelli empirici predittivi del moto del terreno (Ground Motion Prediction Equation, GMPE). Sono inoltre state indagate la varianza associata all'errore di predizione di tali modelli e la covarianza spaziale dei relativi residui con lo scopo di capire se le equazioni GMPE descrivano esaustivamente il fenomeno e di studiarne eventuali termini correttivi. Le molteplici registrazioni di forme d'onda effettuate nella sequenza dell'Emilia nel 2012 hanno permesso la modellizzazione dell'errore di predizione di una GMPE specifica per il Nord Italia con un approccio pienamente non ergodico, che individua nei residui della regressione un termine sistematico correttivo della predizione mediana del modello GMPE.
L'obiettivo della tesi è realizzare mappe di scuotimento tramite la combinazione di tre elementi: la predizione della sopraccitata GMPE, la predizione e la simulazione del suo termine correttivo e dell'incertezza residua ad esso associata, con un approccio dapprima univariato quindi funzionale alla geostatistica. L'applicazione di metodi di analisi funzionale alla sismologia applicata è innovativo e consente di fornire previsioni e simulazioni stocastiche congiunte di molteplici misure d'intensità con uno sforzo modellistico e computazionale comparabile a quello del caso multivariato. La tesi mostra che le performance dell'approccio funzionale sulla previsione di alcune misure d'intensità rilevanti sono comparabili o migliori dell'approccio univariato, fornendo quindi risultati più completi e più robusti.

Contents

List of Figures

List of Tables

Introduction

How much does it cost to save a human life through the seismic adaptation of existing buildings? What are the factors that most significantly influence the seismic motion? With which criteria can the seismic hazard of a region or a site be measured? What is the probability that the peak ground acceleration exceeds a certain intensity? These are just some of the big queries that seismology tries to answer.

The prediction of ground acceleration when an earthquake occurs is essential to answer such questions. The probability of exceeding a certain level of ground motion for a given earthquake scenario is generally computed through the use of Ground Motion Prediction Equations (GMPE). These prediction models are linear regression models whose accuracy increases if the features used (e.g. magnitude, distance from the epicenter, etc.) are the most suitable to describe the phenomenon.

Earthquakes are very complex phenomena to be modeled and a linear prediction model is often unable to capture all the complex interactions between wave propagation and the path in which it propagates, resulting in a high variance of prediction error.

In order to take into account these complex interactions, a further analysis of prediction residuals must be developed to identify the presence of systematic and repeatable contributions that correct the prediction. If the new prediction, given by the sum of the model and the corrective term, has a lower prediction error, it means that an additional deterministic component of the phenomenon has been identified.

A non-ergodic approach is the most effective for determining the corrective term but it is applicable to residuals only when a large number of seismic recordings is available. Using this approach, the Istituto Nazionale di Geofisica e Vulcanologia (INGV) has calculated the corrective term and the variance of the residuals related to the spectral acceleration for different natural periods of oscillation. This quantity allows to compute and model the seismic action for structural response assessment. Since the residuals and the remaining aleatory variability are provided with coordinates, their spatial correlation has been investigated in order to highlight differences in the behavior at low and high periods and predict their value in new positions of the space. For this purpose, it has been adopted the univariate approach to geostatistics that models the spatial covariance of the corrective term considering each period independently.

The corrective term is modelled as the sum of a deterministic term and a random one. Variograms have been used to find the spatial covariance while the prediction has been computed through ordinary kriging, in case the deterministic part is constant in the space, or universal kriging when the deterministic part depends on the spatial coordinates (Cressie 1993). In order to consider the variance related to these predictions, Gaussian sequential simulation has been used. It enables to simulate a new value of the spectral acceleration taking into account the correlation with the values already simulated. To take into account also the correlation between the different spectral periods, multivariate geostatistics can be applied (Chilès et Delfiner, 1999).

However, it presents some limits such as the possibility to compute a restricted number of spectral periods simultaneously.

In this study, in order to overcame the aforementioned limit, for the first time a functional geostatistics approach has been implemented in applied seismology to study the corrective term of the GMPE and the related variance at different periods jointly. This approach is motivated by the fact that the spectral acceleration is a function of the natural period of oscillation of the building. This approach combines Functional Data Analysis (FDA, Ramsay and Silverman, 2005) techniques with the ones of multivariate geostatistics. Indeed, FDA allows to approximate a collection of discreet observations through a smooth function and to project these on a new reference system whose basis is made of the main principal directions (i.e. Functional Principal Components). In this way, the function is represented through a vector of coordinates representing the function in the new reference system.

Instead, the geostatistics multivariate approach, by means of the cokriking, allows us to predict the vector of coordinates in new positions of the space and thus to create the complete curve of both the corrective terms and the variance through the linear combination of the functions of the basis.

The application of the Functional Geostatistics approach to Applied Seismology is a turning point since it allows to provide predictions and joint stochastic simulations of all the periods by means of computational and modeling efforts comparable to the ones of the multivariate approach.

The thesis is organized as follows: Chapter 1 presents a review of literature regarding the GMPE and the ergodic assumption. Chapter 2 describes the case-study and the dataset. Chapter 3 describes the univariate approach to the geostatistical analysis of intensity measures. Chapter 4 describes the Functional geostatistics approach. Chapter 5 shows models comparison and the results.

Chapter 1

State Of The Art

The aim of this Chapter is first to introduce the reader to the key concepts of seismology and engineering seismology we have employed in the thesis and secondly to explain how the previous studies have inspired us and what are the innovations of our analysis in comparison to the other ones.

1.1 PGA and SA

The peak ground acceleration (PGA), in a specific location, is defined as the amplitude of the largest peak acceleration recorded by a local accelerogram during an earthquake (Douglas 2003).
In order to introduce the acceleration response spectrum (SA), we need to describe the dynamic equation of a Single Degree Of Freedom System (SDOF). The SDOF is made up of a mass m, a spring with stiffness k and a dashpot with a coefficient of viscous damping c.

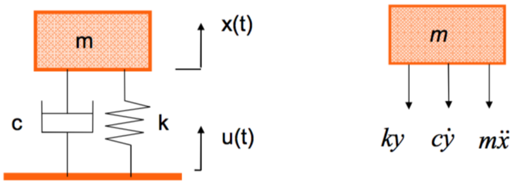

Figure 1.1: Single degree of freedom system.

Let $u(t)$ be the absolute displacement of the support at time t, $x(t)$ the absolute displacement of the oscillator and $y(t) := x(t) - u(t)$ the relative displacement of the oscillator with respect to support.
Then using Newton's second law we obtain the dynamic equilibrium equation:

$$m\ddot{x}(t) + c\dot{y}(t) + ky(t) = 0 \qquad (1.1)$$

and by substituting $x(t) = y(t) + u(t)$

$$m\ddot{y}(t) + c\dot{y}(t) + ky(t) = -m\ddot{u}(t) \tag{1.2}$$

finally

$$\ddot{y}(t) + 2\omega_n\zeta\dot{y}(t) + \omega_n^2 y(t) = -\ddot{u}(t) \tag{1.3}$$

where:

- $\omega_n = \sqrt{\frac{k}{m}}$ is the natural frequency of the oscillator
- $\zeta = \frac{c}{2m\omega_n} = \frac{c}{c_{cr}}$ is the damping ratio

The acceleration response spectrum is defined as:

$$SA(T_n, \zeta) = max_t|\ddot{x}(t)| \tag{1.4}$$

By fixing the parameter ζ and the seismic wave base acceleration, $\ddot{u}(t)$ the acceleration response spectrum becomes a function only of the natural period T_n. Finally we can introduce the relation between SA and PGA.

$$SA(T_n = 0) = PGA \tag{1.5}$$

The equality is derived by observing that when T_n=0 then w_n=0 because $T_n = 2\pi w_n$. But at the same time $w_n = \sqrt{\frac{k}{m}}$ so k=0.
In Figure 1.1 we observe that if k=0 than the mass m and the support move together and also the relative displacement $y(t) = 0$ and $x(t) = u(t)$. By substituting $x(t) = u(t)$ into equation 1.4 we obtain

$$SA(T_n = 0, \zeta) = max\left|u\ddot{(t)}\right| \tag{1.6}$$

But by definition $PGA = max|u\ddot{(t)}|$ so eq(1.5) holds true.

1.2 Ground Motion Models (GMM)

Ground motion models are commonly used in seismology to predict the probability distribution of the ground-motion intensity at a specific site due to a particular earthquake event. These models are often obtained through the regression on observed ground-motion intensities and are fitted using either the one-stage mixed-effects regression algorithm proposed by Abrahamson and Youngs (1992) or the two- stage algorithm of Joyner and Boore (1993). Ground-motion models were originally treated as fixed-effects models that take the following form:

$$y_{e,s} = \mu_{e,s} + \Delta \tag{1.7}$$

Where $y_{e,s}$ is the natural logarithm of the ground-motion parameter (such as PGA or SA) observed at site s during earthquake e, $\mu_{e,s}$ is the mean ground

motion (in log terms) predicted by the GMPE (linear predictor function of magnitude, distance, style of faulting, site conditions and other exogenous variable) and Δ is the noise term that captures all the other factors which influence the response variable and that are not considered by the deterministic part of the model.
The mixed-effects model differs from the fixed-effects model in its interpretation of the error term Δ as the sum of between event residuals δB_e and $\delta W_{e,s}$ within event residuals:

$$y_{e,s} = \mu_{e,s} + \delta B_e + \delta W_{e,s} \tag{1.8}$$

$\delta W_{e,s}$ and δB_e are zero-mean, independent, normally distributed random variables with standard deviations τ and ϕ, respectively.
The Between event residual describes the average source effects and it is influenced by some factors such as stress drop and variation of slip in space and time which are not taken into account by regression models based only on magnitude, style of faulting, and the depth of the source.
The Within event residual (which describes Azimuthal variations in source, path, and site effects), is function of elements like crustal heterogeneity, deeper geological structure, and near-surface layering which are not explained by a distance metric and a site-classification based on the average shear-wave velocity (Villani et Abrahamson 2015).
The standard deviation τ of the between event term describes the earthquake-to-earthquake variability while the within-event standard deviations ϕ describes the record-to-record variability.

Since the between-events and within-event residuals are uncorrelated, the total standard deviation σ can be expressed through the following formula:

$$\sigma = \sqrt{\tau^2 + \phi^2} \tag{1.9}$$

1.3 Ergodic And Non-ergodic Assumptions

Definition 1 (Anderson and Brune(1999)). *An ergodic process is defined as a random process in which the distribution of the random variable in space is assumed to be the same as the distribution of the same variable at a single point when sampled over time.*

This means that the ground-motion uncertainty computed from a global dataset (i.e.,including various sites and various sources for multiple events) is assumed to be the same as the variability at a single site (Rodriguez-Marek et al. 2013). When an ergodic assumption is made, there is a large overestimation of the aleatory standard deviation σ. The key to reduce the standard deviation of the model is identifying those components of ground motion variablility at a single site that are repeatable rather than purely random, so that these may be removed from the aleatory variability and transferred to the quantification of the epistemic uncertainty (Al Atik 2010).

Using the terminology introduced by Al-Atik et al. (2010) and Villani et al. (2015), we can identify three types of residuals decomposition: the fully ergodic, the partially ergodic and the fully non ergodic. The Fully ergodic assumption is made when the empirically based ground motion models are developed to compensate a lack of data (Anderson and Brune, 1999) and the partially-ergodic approaches refer to single-station σ models and have a standard deviation that is more representative of the variability of the ground motion observed at a single site (Lin et al., 2011; Rodriguez-Marek et al., 2011).
In the fully ergodic approach the prediction value $\overline{\mu}_{e,s}$ and the total standard deviation $\overline{\sigma}$ are:

$$\overline{\mu}_{e,s} = \mu_{e,s} \tag{1.10}$$

$$\overline{\sigma} = \sqrt{\tau^2 + \phi^2} \tag{1.11}$$

In order to introduce the partially non ergodic approach and the fully non ergodic approach, we have to split the between-event residuals and the within event residuals as follows.
The between event residuals in region r, $\delta B_{e,r}$, can be considered as the sum of the systematic regional difference in the median source terms and the aleatory source variability terms:

$$\delta B_{e,r} = \delta L2L_r + \delta B_{0,er} \tag{1.12}$$

The $\delta L2L_r$ term can be estimated if we have several recordings from a single source region r and it is computed as

$$\delta L2L_r = \frac{1}{NE_r} \sum_{e=1}^{NE_r} \delta B_{e,r} \tag{1.13}$$

in which NE_r is the number of earthquakes in region r.

The within-event residuals can be seen as

$$\delta W_{e,s} = \delta S2S_s + \delta WS_{e,s} \tag{1.14}$$

in which $\delta S2S_s$ can be interpreted as the systematic average site correction term for station s (site-to-site residual) and $\delta WS_{e,s}$ is called within-site residual.
In order to compute the average site term, we have to calculate the average within event residual at a site s over all the events observed at that specific site.

$$\delta S2S_s = \frac{1}{NE_s} \sum_{e=1}^{NE_s} \delta W_{e,s} \tag{1.15}$$

in which NE_s is the number of earthquakes recorded at site s.
We can also split the within-site residuals in order to highlight an average path term:

$$\delta WS_{e,s} = \delta P2P_{s,r} + \delta W_{0,es} \tag{1.16}$$

in which $\delta P2P_{s,r}$ is the mean path term from sources in region r to site s. The path term is the mean within-site residual for a given source-site pair:

$$\delta P2P_{s,r} = \frac{1}{NE_{s,r}} \sum_{e=1}^{NE_{s,r}} \delta WS_{e,s,r} \tag{1.17}$$

In the partially non-ergodic approach (Rodriguez-Marek et al. 2011)

$$\overline{\mu}_{e,s} = \mu_{e,s} + \delta S2S_s \tag{1.18}$$

$$\overline{\sigma} = \sqrt{\tau^2 + \phi_{WS,s}^2} \tag{1.19}$$

with

$$\phi_{WS,s} = \sqrt{\frac{1}{NE_s - 1} \sum_{e=1}^{NE_s} \delta WS_{e,s}^2} \tag{1.20}$$

The fully non-ergodic approach can be implemented following Villani and Abrahamanson(2015):

$$\overline{\mu}_{e,s} = \mu_{e,s} + \delta S2S_s + \delta P2P_{s,r} + \delta L2L_r \tag{1.21}$$

$$\overline{\sigma} = \sqrt{\tau_{0,r}^2 + \phi_{0,sr}^2} \tag{1.22}$$

where:

$$\tau_{0,r} = \sqrt{\frac{1}{NE_r - 1} \sum_{e=1}^{NE_r} \delta B_{0,er}^2} \tag{1.23}$$

$$\phi_{0,sr} = \sqrt{\frac{1}{NE_{sr} - 1} \sum_{e=1}^{NE_{sr}} \delta W_{0,esr}^2} \tag{1.24}$$

In this study, starting from two dataset developed by Lanzano et al. (2017) following a full non ergodic assumption, we investigate the spatial distribution properties of the corrective term defined as $\delta L2L + \delta P2P + \delta S2S$ and $\sigma_{0,sr}$ (eq 1.22) in order to predict and simulate their values in new locations and at different periods.

1.4 Modelling the spatial dependence among residuals

In the past, several researchers, by adopting a fully ergodic approach, have developed models to study the spatial dependence of between and within event residuals.
For example a regression algorithm for mixed effects models considering the spatial correlation of residuals was developed by Jayaram and Baker (Jayaram and Baker 2010) while the correlation of ground motion parameters such as PGA, the peak ground velocity PGV, and the PSA responses at two different sites was elaborated by Katsuichiro Goda, Hong and Atkinson (Goda and Hong 2008, Goda and Atkinson 2010).
A great contribution to this framework has been made by Park et al. (2007) who explored the site-to-site correlation of IMs and demonstrated its use in seismic hazard. The literature findings indicate that the spatial intraevent correlation depends on the natural vibration periods and on the separation distance.
Through these analysis, based on an univariate approach, it was possible to evaluate the correlations between residuals of spectral accelerations at the same spectral period at two different sites.
In 2012 Loth and Baker introduced a new multivariate approach to model Within-event residuals based on cross-correlation between residuals of spectral accelerations at different periods and at different sites, which is employed for example to develop model for risk assessment of a portfoglio of buildings with different fundamental periods (Loth and Baker 2012).
In our study, instead of modelling the spatial dependence of within event term, we model the spatial dependence of the corrective term $(\delta L2L + \delta P2P + \delta S2S)$ and $\sigma_{0,sr}$ previously defined in the fully non-ergodic assumption.
As the aforementioned studies, we will adopt both an univariate and a multivariate approaches but through different algorithms with respect to the past that we will explain in the next chapter.

Chapter 2

Study Framework and Dataset

2.1 Study Area

Our Model has been applied to the Po Plain, located in the Northern part of Italy, since this area presents some unique features, such as the availability of dataset of seismic records measured by stations located in sites which have the same soil classification, particularly suitable for the development and the validation of our study.

Lots of these records were recorded during the 2012 Emilia Sequence because after the first mainshock temporary seimsic stations were placed and the stations were triggered by the aftershock making available a huge dataset of recordings. Epicenters and faults of the main events of Emilia sequence are reported in Figure 2.1 and the details in the Table 2.1.

Day	latitude	longitude	depth[km]	M$_L$	M$_W$
2012/05/20	44.896	11.264	9.5	5.9	6.1
2012/05/29	44.842	11.066	8.1	5.8	6.0

Table 2.1: Features of 1st and 2nd Emilia earthquake. M_L is the Ricther magnitude and M_W is the moment magnitude.

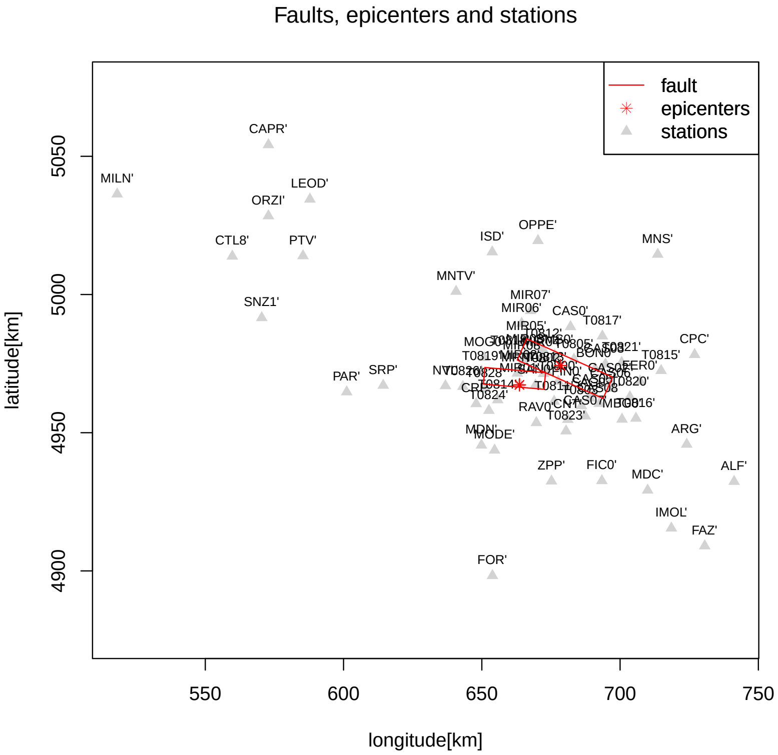

Figure 2.1: Stations, fault and epicenters.

The Po Plain is "a sedimentary basin with variable sedimentary coverage filled by Plio-Quaternary marine and continental deposits, whose thickness ranges from a few tens of meters at the top of buried anticlines up to about 8 km in the Eastern part of the basin towards the Adriatic sea" (Bigi et al. 1992). Another important factor is the proximity to the area of the northern Apennines. "The northern Apennines frontal thrust system is composed of a pile of NE-verging tectonic units that have developed as a consequence of the Cenozoic collision between the European plate and the Adria plate" (Boccaletti et al., 2004).The terminal part of this system, corresponding to the Po Plain, is made up of a complex system of thrust faults and folded arcs, called Monferrato, Emilia and Ferrara-Romagna, from West to East, which locally generated structural highs (Paolucci et. al 2015).

In Figure (2.2) we show a schematic representation of both the Northern Apennines and the Po Plain.

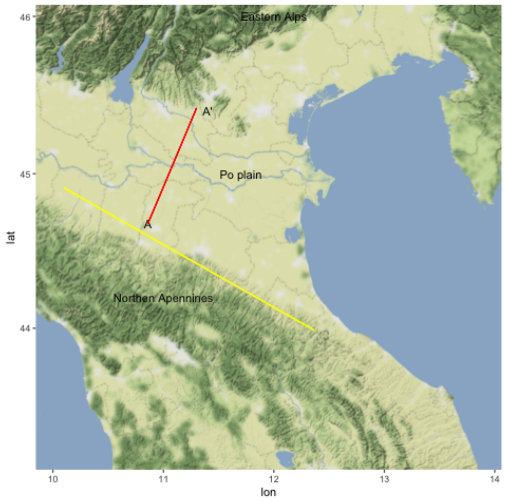

Figure 2.2: Location of the geological cross-section(red) and the northern Apennines frontal thrust(yellow).

As a result of previous researches, some features related to the ground-motion characterization of the Po Plain area have come to light:

- The reflection of S waves,in correspondence with the Moho discontinuity, produces the increase of PGA at a distance between 70 km and 200 km (Bragato at al. 2011).
- When the boby waves are trapped in the basin, surface waves are generated (Basin effects). Surface waves mainly determine the seismic signal at period greater than 2 seconds (Luzi et al. 2013).
- The presence of a privileged direction of amplification of the surface waves during the main event of the 2012 Emilia seismic sequence (Paolucci et al. 2015).

Some of the factors, which mainly influenced the near-source ground motion during the the main event of the 2012 Emilia seismic sequence, are reported in Figure 2.3. with a particular attention to the relation between the buried topography and the generation of surface waves created by the irregular geological configuration(Paolucci et al. 2015).

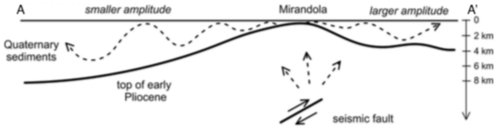

Figure 2.3: Surface waves generation (Paolucci et al.,2015).

The peculiar structure of the geological cross-section A-A' passing through the Ferrara-Romagna folder arc (Figure 2.3) will be crucial to understand by the physical point of view the results of our analysis.

2.2 Ground Motion Prediction Equation

The Ground Motion Prediction Equation (GMPE) employed to construct the datasets introduced in Section 2.3 is a model specifically tailored by Lanzano et al.,(2016) for the Northern Italy able to predict the geometric mean of horizontal response spectral accelerations in the period range 0.01-4s.

This GMPE by Lanzano et al. (2016) has the functional form:

$$\log_{10} Y = a + F_M(M) + F_D(R, M) + F_{sof} + F_S + F_{bas} \tag{2.1}$$

Y is the geometrical mean of the horizontal components of PGA (expressed in cm/s^2) and SA (in cm/s^2) for 24 periods in the range 0.04–4 s with damping ξ=5%.

Members of the model:

- a is the offset.
- $F_D(R, M)$ represents the distance function.
- $F_M(M)$ is the magnitude scaling.
- F_S concerns the site amplification.
- F_{sof} is the style of faulting.
- F_{bas} is the basin-effects correction.

R (in km) is the distance, M is the magnitude(if M is missing we adopt M_w or M_L).
M_L called Ricther magnitude or local magnitude, is determined at short distances and it is homogeneously determined for small earthquakes up to saturation at about M_L = 7.0, M_w is the moment magnitude which depends on the size of the source and the slip along the fault (Douglas 2003).

The distance function has equation:

$$F_D(R, M) = [c_1 + c_2(M - M_r)] \log_{10} \frac{R}{R_h} \tag{2.2}$$

where

- c_1 and c_2 are the attenuation coefficients.
- M_r is a reference magnitude fixed to 5.0.
- R is either the hypocentral distance R_{hypo} (distance to the hypocenter of the earthquake, i.e. the distance to the rupture's starting point (Douglas 2013)) or the distance computed as $\sqrt{R_{JB}^2 + h^2}$.
  R_{JB} (Joyner–Boore distance) is defined as the distance to the surface projection of the rupture plane of the fault (Douglas 2013).

- h is the pseudodepth coefficient.
- R_h is a hinge distance thet takes into account changes in the attenuation rate.

Since attenuation depends on both the geologic domain and distance ranges, equation 2.2 has been modified introducing the index j:

- j=1 if the site is located in PEA(central Po Plain or eastern Alps). and $R \leq R_h$
- j=2 if the site is located in PEA and $R > R_h$.
- j=3 if the site is located in NA (Northen Apenines)and $R \leq R_h$.
- j=4 if the site is located in NA and $R > R_h$.

The new form of equation 2.2 is:

$$F_D(R, M) = [c_{1j} + c_{2j}(M - M_r)] \log(\frac{R}{R_h}) \qquad j = 1...4 \tag{2.3}$$

The R_h is set to 70 km. (Douglas et al. (2003) assumed the same value for the Po Plain area and for central Italy, respectively).
To distinguish among the sites located in PEA and NA, a geographic separation has been defined through the linear equation $LAT_{ref} = 0.33LON_s + 48.3$, in which LON_s is the station longitude and LAT_{ref} is the reference latitude, expressed in decimal degrees.
Positive differences between station latitude LAT_s and LAT_{ref} identify the PEA sites, and negative differences identify the NA ones.

The magnitude function has the form

$$F_M(M) = b_1(M - M_r) + b_2(M - M_r)^2 \tag{2.4}$$

The M_r parameter is the reference magnitude, defined in equation (2.2).

The term F_{sof} in equation (2.1) is needed in order to take into account different styles of faulting and is given by

$$F_{sof} = f_j E_j \quad j = NF, TF, UN \tag{2.5}$$

The coefficients f_j in the equation are estimated during the analysis, E_j are dummy variables representing different style of faulting:

- normal (NF)
- thrust (TF)
- unspecified (UN)

The term F_s in equation (2.1) represents the site effect and it is defined as follows

$$F_s = s_j S_j \quad j = A, B, C \tag{2.6}$$

In which S_j are dummy variables for the three EC8 site classes A, B, and C. Through regression coefficients s_j are estimated.

The term F_s has been introduced to model the bias observed at short periods in the analysis of ITA10 residuals in the range 30-100 km. This bias is mainly ascribed to C class stations in the Po Plain.

The basin-effect term in equation (2.1) is defined as

$$F_{bas} = \delta_{bas} \Delta_{bas} \tag{2.7}$$

in which δ_{bas} is the coefficient to be determined during the analysis and Δ_{bias} is a dummy variable equal to 1 if the site is located in the middle of a basin and 0 else (Lanzano et al.2016).

2.3 Dataset

In the analysis we use two datasets which have dimension 71 × 25 where the i-th row refers to the i-th station (whose distribution in space can be observed in figures 2.3 and 2.4) and the j-th column to the j-th period in the range 0.01-4s. The first dataset collects the value of the corrective term ($\delta L2L + \delta P2P + \delta S2S$) while the second one collects the values of $\sigma_{0,sr}$ computed with a fully non ergodic approch.
These two datasets are subsets of systematic components and σ terms computed by Lanzano et al.,(2017) for the North of Italy using as reference GMPE the aforementioned model.
The subsets refer to a selection of more than 2200 records (supplied by ESM "Engineering Strong-Motion" and Itaca "Italian Accelerometric Archive"; Luzi et al.,2016; Pacor et al.,2011) and 71 accelerometric stations located in the Po Plain which satisfy the condition of belonging to the C1 class.
A station belongs to C1 class if V_{S30} is between 160-360 m/s and the station is located in the deepest part of the Po Plain or in smaller basins in the Apennines(Lanzano et al. 2016).
We expect to observe 2D-3D complex site effects for these station at the edges of the basin due to the surface waves which imply soil amplification at frequencies lower than 1 Hz.
These 71 stations are distributed in an Area which is defined in the latitude range 44.2-45.6 N and in the longitude range 9.23-12,04 E for an area of about 37500 km^2. All the records selected refer to events which have epicenters in the ZS912 reported in red in Figure(2.4). ZS9 is a seismic source model (consistent with the CPTI04 parametric catalogue), based on historical earthquakes, instrumental seismicity, active faults and their seismogenic potential, and seismotectonic

evidence from earthquakes since 1998 (Meletti et al. 2008). The model is composed of 36 zones (ZS912 is the 12-th zone) where earthquakes with $M_w \geq 5$ are expected and the probability that an earthquake with M_w up to 5 may occurs anywhere outside this seismogenic zones is very low (Meletti et al. 2008).

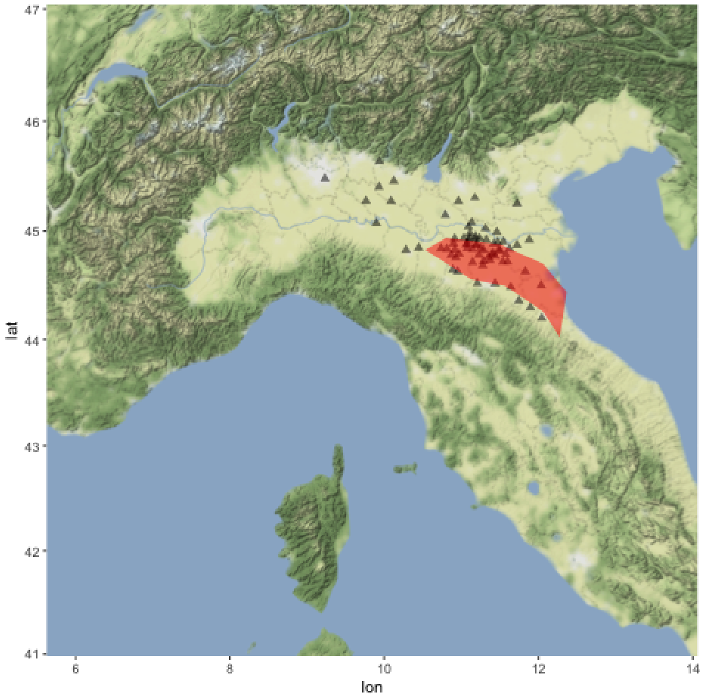

Figure 2.4: Stations and ZS912(red).

The number of available records was crucial because the decomposition of between event residuals and within event residuals in the systematic source term $\delta L2L_r$, the path term $\delta P2P_{s,r}$ and the site correction $\delta S2S_s$ can be applied when empirical datasets are sufficiently populated to estimate each single contribution (Lin et al. 2011).

Chapter 3

A univariate approach to the geostatistical analysis of intensity measures

3.1 Modelling the Covariance of the Random Field

In this Chapter we describe the univariate methods applied in our analysis. Each intensity measure will be modelled through a random field $Z_{\boldsymbol{s}}, \boldsymbol{s} \in D$.
The random field is a collection of variables having the form

$$Z_{\boldsymbol{s}} = m_{\boldsymbol{s}} + \delta_{\boldsymbol{s}} \tag{3.1}$$

$m_{\boldsymbol{s}}$ is called drift and is the deterministic part of the variable, $\delta_{\boldsymbol{s}}$ is the random component.

Definition 2 (Chilès et Delfiner, 1999)**.** *Process $\{Z_s, s \in D\}$ is said second-order stationary if the following conditions hold:*

- $E\left[Z_{\mathbf{s}}\right] = m$ *for all* $\mathbf{s}$ *in* D
- $Cov\left(Z_{\mathbf{s_i}}, Z_{\mathbf{s_j}}\right) = E\left[\left(Z_{\mathbf{s_i}} - m\right)\left(Z_{\mathbf{s_j}} - m\right)\right] = C\left(\mathbf{h}\right)$ *for all* $\mathbf{s_i}$, $\mathbf{s_j}$ *in* D*,* $\mathbf{h} = \mathbf{s_i} - \mathbf{s_j}$.

  Function C is called covariogram

Definition 3 (Chilès et Delfiner, 1999)**.** *Process $\{Z_s, s \in \ D\}$is said intrisically stationary if*

- $E\left[Z_{\mathbf{s}}\right] = m$ *for all* $\mathbf{s}$ *in* D
- $Var\left(Z_{\mathbf{s_i}} - Z_{\mathbf{s_j}}\right) = E\left[\left(Z_{\mathbf{s_i}} - Z_{\mathbf{s_j}}\right)^2\right] = 2\gamma\left(\mathbf{h}\right)$ *for all* $\mathbf{s_i}$, $\mathbf{s_j}$ *in* D*,* $\mathbf{h} = \mathbf{s_i} - \mathbf{s_j}$.

The function γ is called semivariogram and the function 2γ variogram.
The relation between semivariogram and covariogram is

$$\gamma\left(\mathbf{h}\right) = C\left(\mathbf{0}\right) - C\left(\mathbf{h}\right) \tag{3.2}$$

Definition 4. *An intrinsic stationary process $\{Z_{\mathbf{s}}, \mathbf{s} \in D\}$ is said isotropic if its variogram is isotropic,i.e,*

$$Var\left(Z_{\mathbf{s_i}} - Z_{\mathbf{s_j}}\right) = 2\gamma\left(h\right) \quad h = ||\boldsymbol{h}||$$

otherwise it is said anisotropic.

When the structure of the covariance is homogenous over all the directions, isotropy is verified. To investigate isotropy, directional variograms are employed.

The variogram is characterized by the nugget, the sill and the range.
The sill of the semivariogram is defined as

$$\tau^2 + \sigma^2 = \lim_{h \to \infty} \gamma\left(h\right)$$

where τ^2 is the nugget effect and σ^2 is said partial sill.
The existence of a finite limit indicates that the process is second-order stationary, featured by a variance $C\left(0\right) = \tau^2 + \sigma^2$.
τ^2 is called nugget and it is defined as $\lim_{h \to 0} \gamma\left(h\right)$.
The range R of a semivariogram is the value where it reaches the sill:

$$\gamma\left(R\right) = \tau^2 + \sigma^2 \tag{3.3}$$

The semivariogram range quantifies the range of influence of the process: for distance greater than the range, two elements of the process are uncorrelated. The variogram range can be infinite if the sill does not exist (indication of non-stationarity) or if the sill is reached asymptotically.
Given a dataset $Z_{s_1}, ..Z_{s_n}$, under the stationarity assumption, the sample semivariogram is computed as

$$\widehat{\gamma\left(h\right)} = \frac{1}{2\left|N\left(h\right)\right|} \sum_{(i,j) \in |N(h)|} \left[Z_{\mathbf{s_i}} - Z_{\mathbf{s_j}}\right]^2$$

where $N\left(h\right) = \{(i,j) : \|\mathbf{s_i} - \mathbf{s_j}\| = h\}$ and $|N\left(h\right)|$ is its cardinally.
After sample estimation, we can fit the sample variogram through a parametric valid model. In particular in our analysis we'll use and compare the following models:

- Exponential model:

$$\gamma\left(h\right) = \begin{cases} \sigma^2 \left(1 - e^{\frac{-h}{a}}\right) & h > 0 \\ 0 & h = 0 \end{cases}$$

  where a, $\sigma \in \mathbb{R}$. The sill is σ^2, the range is infinite, but one can define the practical range as $\widetilde{R} = 3a$. $\widetilde{R}$ satisfies $\gamma(\widetilde{R}) \sim 95\%\sigma^2$. In this thesis with a slight abuse of notation we will name *range* the *practical range* of the exponential model.

- Spherical model

$$\gamma(h) = \begin{cases} 0 & h = 0 \\ \sigma^2 \left\{ \frac{3}{2}\frac{h}{a} - \frac{1}{2}\left(\frac{h}{a}\right)^3 \right\} & 0 < h < a \\ \sigma^2 & h \geq a \end{cases}$$

  with a, $\sigma \in \mathbb{R}$. a is the range, σ^2 the sill.

The parameters are estimated following the weighted least squares criterion,i.e., looking for the parameters which minimize

$$\sum_{k=1}^{K} \frac{1}{w_k} \left(\widehat{\gamma(h_k)} - \gamma(h_k; \theta) \right)^2 \tag{3.4}$$

where w_k, $k = 1, .., K$ are set to the number of couples $N(h_k)$ within each class.

3.1.1 Drift Estimation

When the process is not stationary the mean m_s of Model 3.1 is modelled as

$$m_{\boldsymbol{s}} = \sum_{l=0}^{L} a_l f_l(\boldsymbol{s}) \tag{3.5}$$

and we assume residuals stationarity. We can apply drift estimation to perform our geostatistical analysis on the residuals.
Given $Z_{s_1}, ...Z_{s_n}$, we want to estimate the model parameters $a_0, ...a_{L+1}$ such that

$$\boldsymbol{Z} = \boldsymbol{F}\boldsymbol{a} + \boldsymbol{\delta} \tag{3.6}$$

Where $\boldsymbol{F}$ is the design matrix and $\boldsymbol{\delta}$ the residuals vector characterized by an unknown covariance structure $\boldsymbol{\Sigma}$. If we knew the $\boldsymbol{\Sigma}$, we could employ the Generalized Least Square (GLS) estimator to estimate $\boldsymbol{a}$.
This is found by minimizing

$$\boldsymbol{(Z - Fa)'\Sigma^{-1}(Z - Fa)} \tag{3.7}$$

over $\boldsymbol{a} \in R^p$. The GLS estimator $\boldsymbol{a}^{GLS}$ has the form

$$\boldsymbol{a^{GLS} = (F\Sigma^{-1}F)^{-1}F\Sigma^{-1}Z} \tag{3.8}$$

Through the following iterative algorithm we can jointly estimate γ and $\boldsymbol{a}$ via GLS and avoid the problem of unknown $\boldsymbol{\Sigma}$.
Let $\boldsymbol{z} = (z_{\boldsymbol{s_1}}, ., z_{\boldsymbol{s_n}})$ be a realization of the non stationary random field $Z_{\boldsymbol{s}}$, $\boldsymbol{s} \in D, D \subseteq R^d$:

1. Estimate the drift vector $\boldsymbol{m}$ through the OLS method ($\widehat{\boldsymbol{m}}^{OLS} = F(F^T F)^{-1} F^T \boldsymbol{z}$) and set $\widehat{\boldsymbol{m}} = \widehat{\boldsymbol{m}}^{OLS}$

2. Compute the residual estimate $\widehat{\boldsymbol{\delta}} = (\widehat{\delta}_{\boldsymbol{s_1}}, ., \widehat{\delta}_{\boldsymbol{s_n}})$ by difference $\widehat{\boldsymbol{\delta}} = \boldsymbol{z} - \widehat{\boldsymbol{m}}$

3. Estimate the semivariogram γ of the residual process $\{\delta_{\boldsymbol{s}}, \boldsymbol{s} \in D\}$ from $\widehat{\boldsymbol{\delta}}$ first with the empirical estimator and then fitting a valid model.
   Derive from γ the stimate $\widehat{\Sigma}$ of Σ

4. Estimate the drift vector $\boldsymbol{m}$ with $\widehat{\boldsymbol{m}}^{GLS}$, obtained from $\boldsymbol{z}$ using $\widehat{\boldsymbol{m}}^{GLS} = F(F^T\Sigma F)^{-1}F^T\Sigma^{-1}\boldsymbol{z}$

5. Repeat 2-4 until convergence has been reached.

3.1.2 Geostatistical analysis of the corrective term and $\sigma_{0,sr}$

As first step of our geostatistical analysis, we plot the spatial distribution of our data. In Figure 3.1 it's possible to observe how the 71 stations of the dataset are distributed (UTM coordinates).

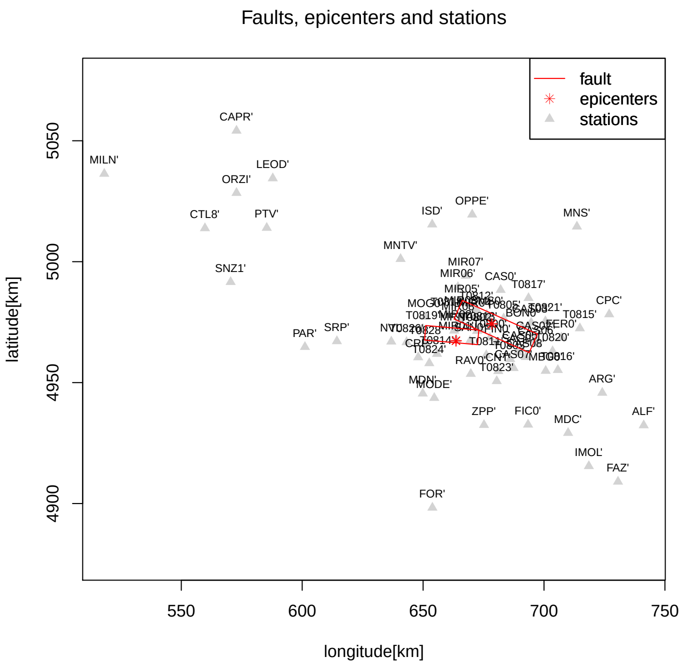

Figure 3.1: Distribution of stations.

The area of study cover a square of side 200 km (a.k.a. bounding box). We fix the cut-off distance equal to 100 km that is half of the side of the bounding box for both the corrective terms $\delta L2L + \delta P2P + \delta S2S$ and $\sigma_{0,sr}$.
The cut-off distance is the maximum distance on which the variogram is fitted. In order to understand how the period impacts on the spatial correlation of the corrective term, we develop our analysis on both the short and the long periods by considering the PGA and $T = 4s$.

3.1.3 Analysis of corrective term

The bubbleplot in Figure 3.2 shows at the same time the position and the values of the corrective term for the Peak Ground Acceleration

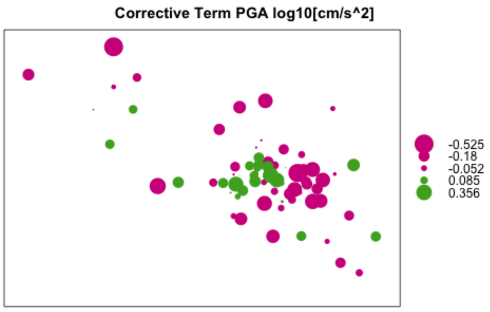

Figure 3.2: Bubbleplot PGA.

The PGA bubbleplot doesn't reveal a strong evidence of a spatial trend in the data. We assume the drift $m_{\mathbf{s}}$ constant and unknown.
Since there isn't a strong spatial trend, it is reasonable to assume that the process is intrinsic stationary and to compute both the variogram and the directional variograms in order to investigate the isotropic assumption.
The directional semivariograms in Figure 3.3 are quite similar in both the range and the sill. We thus consider the process $Z_{\mathbf{s}}$ as isotropic and we can compute the global sample variogram reported in Figure 3.4.

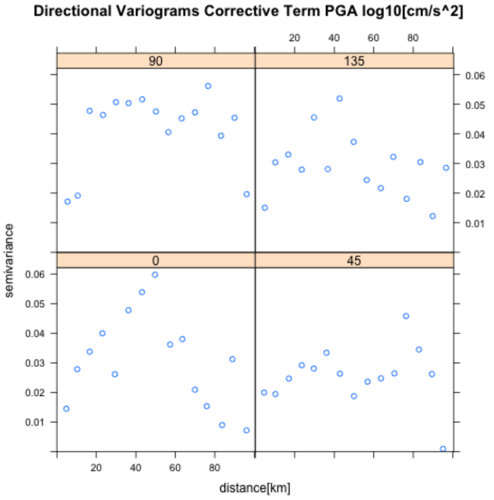

Figure 3.3: Directional variograms.

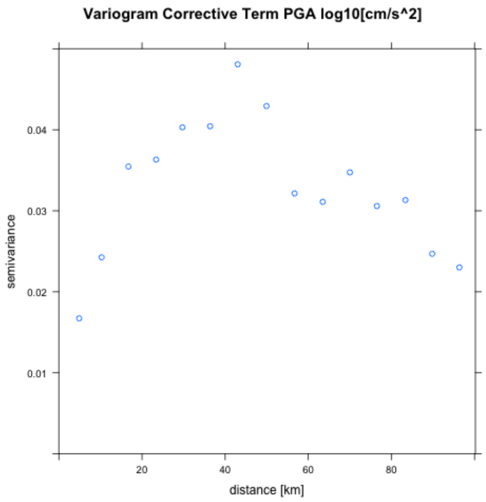

Figure 3.4: Sample variogram.

The linear behaviour in the origin suggests the application of the spherical model or the exponential one to fit the sample variogram. We choose the best fitting model through Leave-One-Out cross-validation using, as metric, the mean square prediction error and the mean square prediction error divided by the variance of prediction (prediction with high variance are weigthed less). Here, prediction at a new location $\boldsymbol{s}$ is made via the Best Linear Unbiased Predictor $\hat{Z}_{\boldsymbol{s}}$ which will be used in the following (see Section 3.2):

$$MSE = \frac{1}{N}\sum(Z_s - \widehat{Z_s})^2 \quad M_2 = \frac{1}{N}\sum\frac{(Z_s - \widehat{Z_s})^2}{var(\widehat{Z_s})} \tag{3.9}$$

Model	MSE	M_2
Exponential	0.0269	1.0799
Spherical	0.027	1.115

Table 3.1: Model Comparison.

As we can observe in Table 3.1, the exponential model gets better performance so we decide to use it to fit the sample variogram. We compute the values of the parameter of the model by minimizing weighted least squares.

In Table 3.2 we report the values of the parameters and in Figure 3.5 the sample variogram fitted with the exponential model.

Nugget	Psill	Range
Exponential	0.0393	32.85
Spherical	0.027	11.15

Table 3.2: Exponential model parameters.

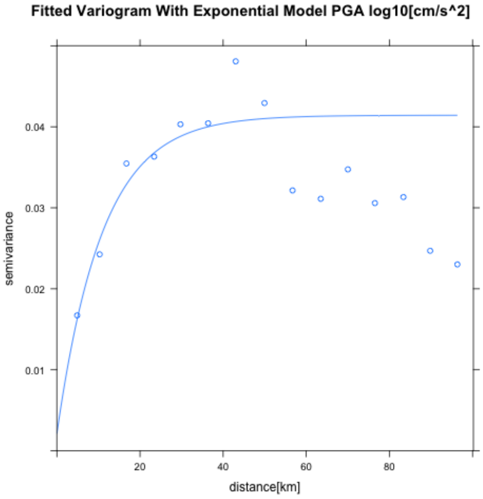

Figure 3.5: fitted variogram.

We can now compute the covariance for two random variables of the process $z_{\boldsymbol{s_i}}, z_{\boldsymbol{s_j}}$ thanks to the equation

$$\gamma(h) = C(0) - C(h) \tag{3.10}$$

where $C(h) = Cov(z_{\boldsymbol{s_i}}, z_{\boldsymbol{s_j}})$ with $h = s_i - s_j$ and $C(0) = lim_{h \to \infty} \gamma(h)$.
We repeat the previous geo-statistical analysis for the corrective term at period $T = 4$. First we plot the data with bubbleplot.

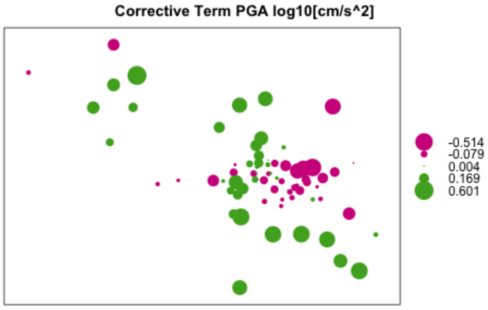

Figure 3.6: Bubbleplot for corrective term at $T = 4s$.

As you can observe in Figure 3.6, contrary to what we observed for PGA, the result shows a positive drift (green bubble) from South-Est toward North-West in the bottom left of the bubbleplot, a positive drift from North to South in the center (green bubble) and a negative drift (purple bubble) from East toward

West.

Then we want to understand what happens if we don't care of the drift and we model the process $\{Z_{\boldsymbol{s}}\}$ as intrinsic stationary.
A first warning is captured by the directional sample variograms in Figure 3.7, since in different directions the variograms show different behaviours.

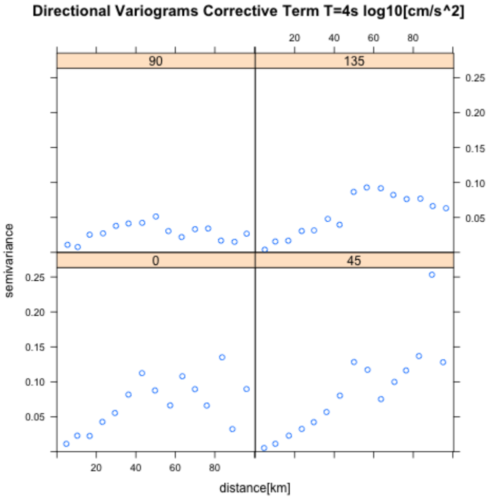

Figure 3.7: Directional variograms.

If we try to model a unique sample variogram and to fit it with an exponential model, we obtain a range greater than the cut-off distance that is an indication that, at the scale of observation, the process cannot be considered as stationary.

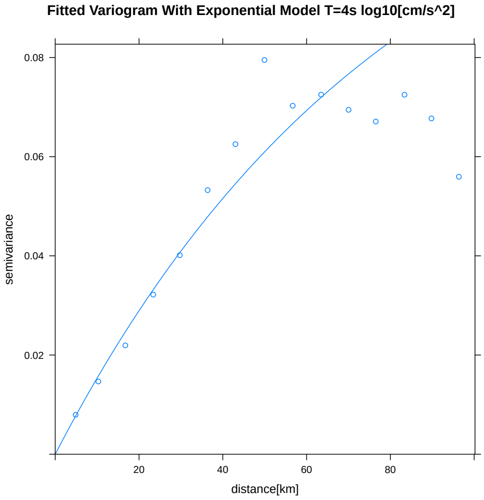

Figure 3.8: Sample variogram fitted.

We can't assume the intrinsic stationary assumption so it means that $m_{\boldsymbol{s}}$ is a function of the $\boldsymbol{s}$ coordinates. The idea is that the random field can be seen as the sum of a deterministic surface $m_{\boldsymbol{s}}$ and an aleatory second-order stationary component $\delta_{\boldsymbol{s}}$. If we can learn the surface representing the drift, then we can model the covariance of the residuals δ_s through the variogram. We try to model the drift with polynomial surfaces of degree 1 to 4:

1. $m_s = \beta_0 + \beta_1 x + \beta_2 y$
2. $m_s = \beta_0 + \beta_1 x + \beta_2 y + \beta_3 x^2 + \beta_4 y^2 + \beta_5 xy$
3. $m_s = \beta_0 + \beta_1 x + \beta_2 y + \beta_3 x^2 + \beta_4 y^2 + \beta_5 xy + \beta_6 x^2 y + \beta_7 y x^2 + \beta_8 x^3 + \beta_9 y^3$
4. $m_s = \beta_0 + \beta_1 x + \beta_2 y + \beta_3 x^2 + \beta_4 y^2 + \beta_5 xy + \beta_6 x^2 y + \beta_7 y x^2 + \beta_8 x^3 + \beta_9 y^3 + \beta_{10} x^4 + \beta_{11} y^4 + \beta_{12} x y^3 + \beta_1 3 x^3 y$

The first two curves are too simple to catch the real drift. Indeed, when the residuals δ_s are estimated as $\delta_s = Z_s - m_s$ and we fit the variogram with an exponential model, we still obtain a range greater than the cut-off distance. With surfaces 3 and 4 this problem seems overcome. In the latter case, we are able to apply the aforementioned algorithm for drift estimation and compare their performance with LOO Cross-validation based on two different metric, i.e., MSE and M_2. Since we obtain similar results for these two cases, we adopt the simplest model which is the number 3. In Figure 3.9 we can observe the sample variogram of the residuals fitted with the exponential model. In Table 3.3 we report the comparison between exponential and spherical models and in Table 3.4 the values of the parameters.

Model	MSE	M_2
Exponential	0.028	1.426
Spherical	0.03	1.54595

Table 3.3: Model Comparison.

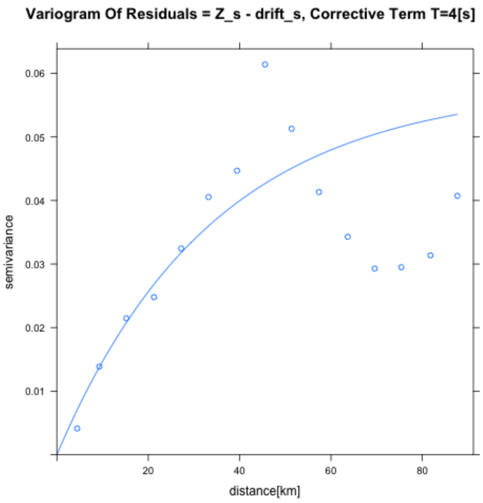

Figure 3.9: Residual variogram.

Nugget	Psill	Range
0.00	0.055	98.55

Table 3.4: Exponential model parameters.

The correlation coefficient $\rho(h)$ in case of exponential model is

$$\rho(h) = exp(-h/a) \quad a = \frac{range}{3} \tag{3.11}$$

We can compare the shapes of the correlation coefficients, as function of h, for both PGA and $T = 4s$.

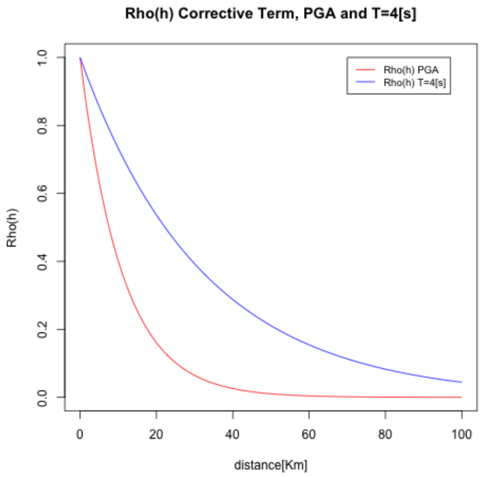

Figure 3.10: Comparison of correlation coefficients.

The random field of the PGA looses the correlation faster than $T = 4s$. Indeed, the range of PGA is lower than the one related to the period $T = 4s$.
The correlation of spectral intensity measures is period-dependent because short period waves tend to be more affected by the heterogeneities of the propagation path, thus resulting correlated at a shorter scale than long period ground motions (Zerva et Zervas, 2002; Bradley, 2014).

3.1.4 Analysis of $\sigma_{0,sr}$

We repeat the statistical analysis, already performed for the corrective term dataset, also for the variance term $\sigma_{0,sr}$ working again with PGA and $T = 4s$.
In Figure 3.11 and Figure 3.12 the Bubbleplot doesn't show an evident non-costant drift and the directional variograms are quite similar in all directions so we can assume that the random field is second-order stationary and isotropic.

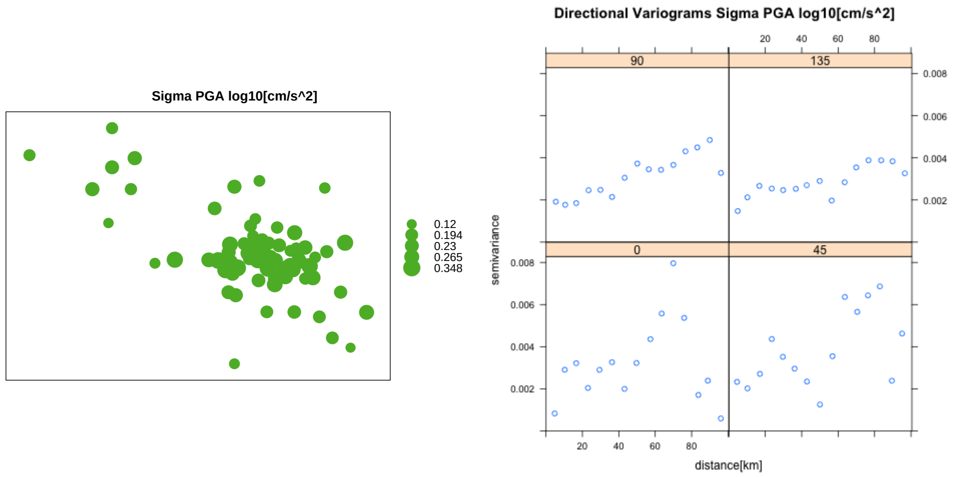

Figure 3.11: Bubbleplot and directional variogram PGA.

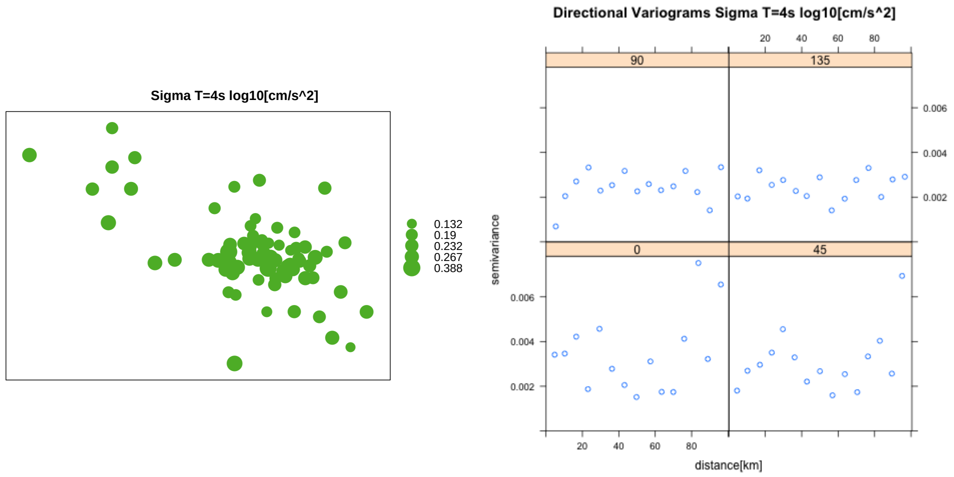

Figure 3.12: Bubbleplot and directional variogram $T = 4s$.

The exponential model, again, gets a better MSE with LOO Cross Validation than the spherical one, as reported in Table 3.5 and Table 3.6.

Model	MSE	M_2
Exponential	0.0002739	1.2
Spherical	0.002762	1.23

Table 3.5: Model Comparison PGA.

Model	MSE	M_2
Exponential	0.000234	1.008
Spherical	0.002531	1.14

Table 3.6: Model Comparison $SA(T = 4s)$.

So the sample variogram is fitted through an exponential model whose parameters are reported in Tables 3.7 and Table 3.8.

Nugget	Psill	Range
0.0009	0.0021	37.2

Table 3.7: Exponential model parameters PGA.

Nugget	Psill	Range
0.00	0.0029	12

Table 3.8: Exponential model parameters $SA(T = 4s)$.

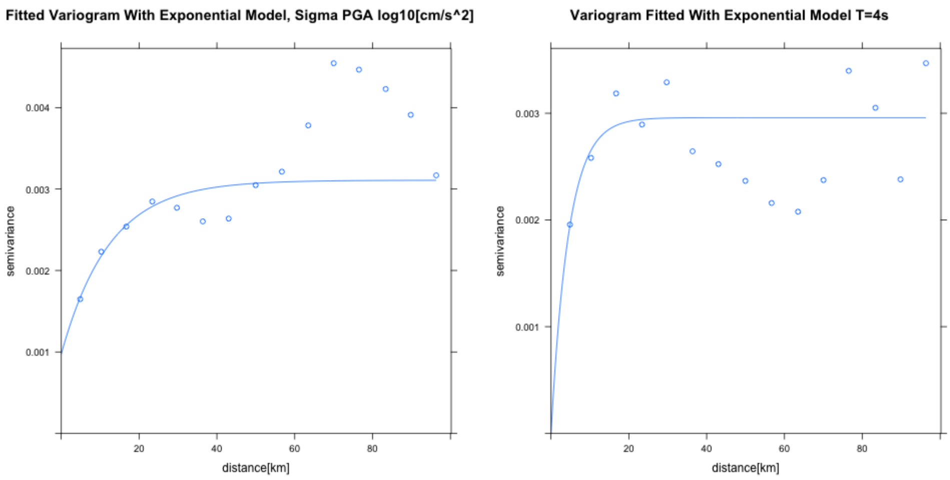

Figure 3.13: Variogram of PGA and on the left and variogram of $SA(T = 4)$ on the right fitted with the exponential model.

At the end, we compute the $\rho(h)$ of the random field PGA and $\mathrm{SA}(T = 4)$ in order to compare how the correlation decreases as function of the period. Contrary to what we observed for the corrective term, the covariance function of PGA is now generally above the one corresponding to $\mathrm{SA}(T = 4)$ (i.e., the range for the PGA is higher than that of $\mathrm{SA}(T = 4)$); however, the two curves are closer than ones in the case of corrective term.

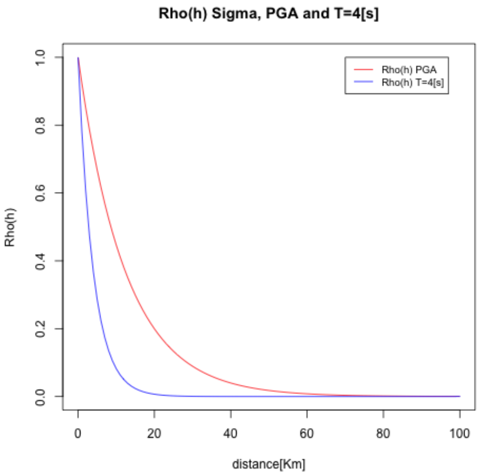

Figure 3.14: Comparison of correlation coefficients of PGA and SA($T = 4s$).

3.2 Kriging

3.2.1 Introduction

After modelling the covariance of the random field associated to different periods of the corrective term and of $\sigma_{0,sr}$, we want to predict their values in a new location. The kriging predictor linearly combines the data in the available locations and predicts the values of the random field in an arbitrary location as $Z_{s_0} = \sum_{i=1}^{n} \lambda_i Z_{s_i} = \lambda^T \mathbf{Z}$, where the weights are found according to the field covariance structure (computed in Section 3.1). We can distinguish among 3 types of kriging:

1. Simple: Simple kriging is employed for random fields with known drift. The weights are found by solving:

$$\Sigma\lambda = \sigma_{\mathbf{0}}$$

where

$$\Sigma = \left[Cov\left(Z_{\mathbf{s}_i}, Z_{\mathbf{s}_j}\right)\right], \sigma_{\mathbf{0}} = \left[Cov\left(Z_{\mathbf{s}_i}, Z_{\mathbf{s}_0}\right)\right]$$

2. Ordinary: Ordinary kriging is employed for random fields with constant unknown mean. The optimal weights are solution of the system

$$\begin{pmatrix} \Sigma & 1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} \lambda \\ \beta \end{pmatrix} = \begin{pmatrix} \sigma_{\mathbf{0}} \\ 1 \end{pmatrix}$$

where β are Lagrange multiplier.

3. Universal: Universal kriging is employed for random fields with variable drift. The Universal kriging predictor is obtained from

$$\begin{pmatrix} \Sigma & F \\ F^T & \mathbf{0} \end{pmatrix} \begin{pmatrix} \boldsymbol{\lambda} \\ \boldsymbol{\beta} \end{pmatrix} = \begin{pmatrix} \boldsymbol{\sigma_0} \\ \boldsymbol{f_0} \end{pmatrix}$$

where $\boldsymbol{f_0} = [f_l\,(s_0)]$,$F = [f_l\,(s_i)]$ is the design matrix, and $\boldsymbol{\beta} = (\beta_0, .., \beta_l)$ is the vector of Lagrange multipliers accounting for the constrains:

$$\sum_{i=1}^{n} \lambda_i f_l\,(s_i) = f_l\,(s_0)\,, l = 1, ..., L.$$

For more details about kriging see Chilès et Delfiner (1999).

3.2.2 Kriging for the corrective terms

The random field of the corrective term of PGA has been modelled in the previous section considering a constant and unknown drift. We can employ the ordinary kriging predictor to estimate the random field on a grid of 70000 points covering the Area of study. In Figure 3.15 we report the prediction and the related variance and in Figure 3.16 we zoom on the zone of the faults in which the rectangle on the left represent the projection of the fault planes on the surface of the second main events of 2012 Emilia seismic sequence and the one on the right of the first main event.

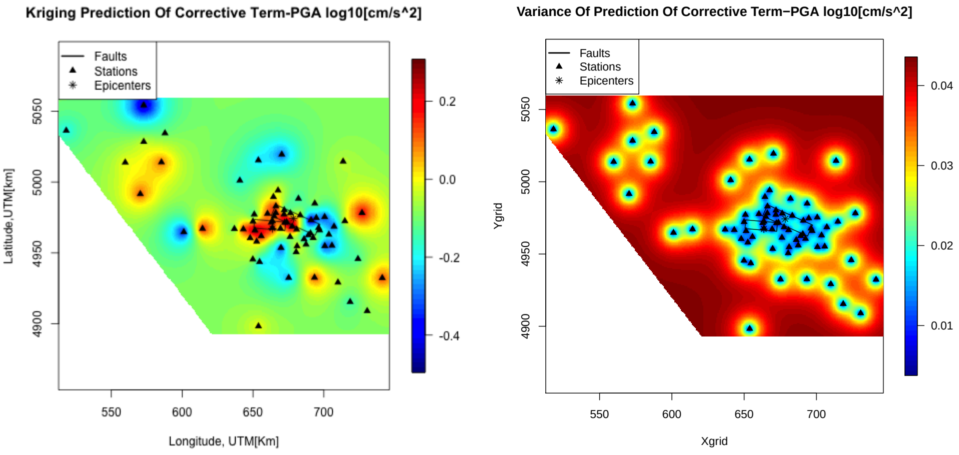

Figure 3.15: Ordinary kriging prediction and variance for PGA.

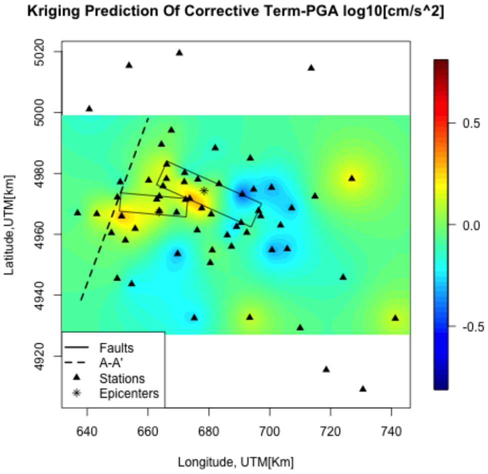

Figure 3.16: Ordinary kriging prediction in the fault zone. The colour scale has been modified to highlight local differences.

We can observe a strong amplification of ground motion in the fault zone near the cross section, whereas smaller variations are observed far from the faults. The corrective term, for period $T = 4s$, was modelled in the previous section through a random field with non constant drift. Note that the value of the Universal kriging Predictor in a new location coincides with the sum of the drift estimated via GLS and evaluated in that location and the prediction of the residual δ_s via Simple kriging (see Chiles et Delfiner, 1999). In Figure 3.17 we report in the left panel the values of the drift and in the right panel the Simple kriging prediction of the residuals whereas Figure 3.18 shows the Universal kriging prediction(the sum of previous maps) and its related variance. Finally Figure 3.19 shows a zoom on the faults.

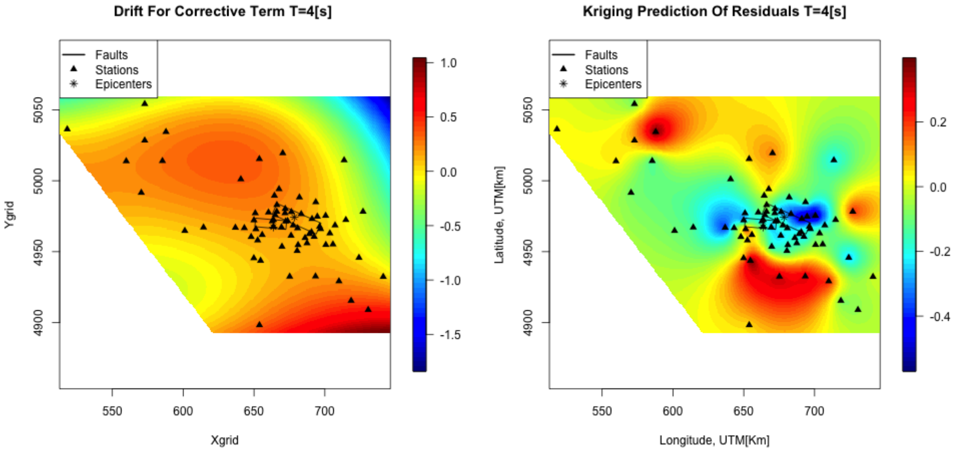

Figure 3.17: Drift for corrective term (left panel) and simple kriging prediction of the residuals (right panel).

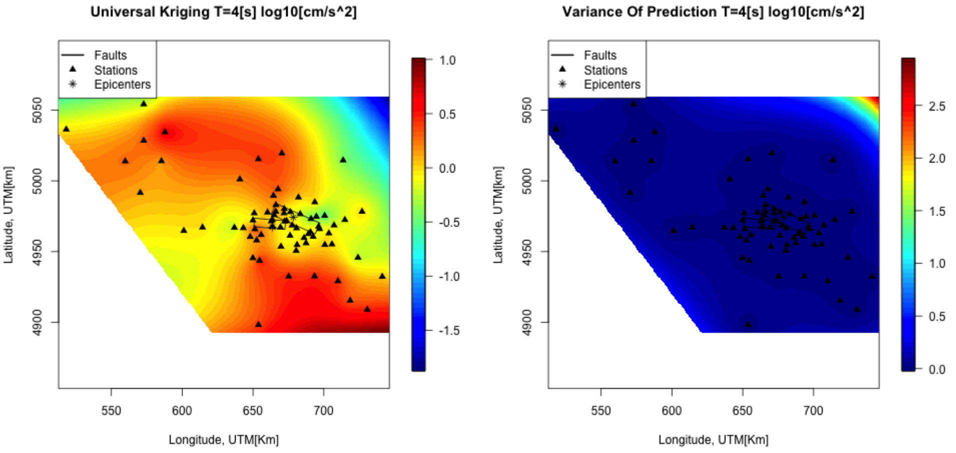

Figure 3.18: Universal kriging prediction and variance of corrective term at $T = 4s$.

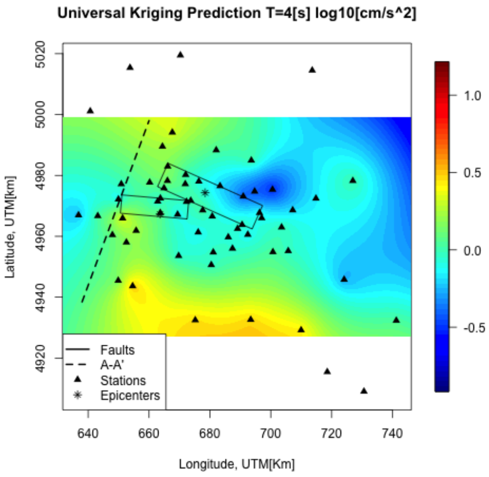

Figure 3.19: Universal kriging prediction (zoom). The colour scale has been modified to highlight local differences.

As already observed for PGA, in the fault zone we have a strong amplification of the motion. This ground motion amplification could be linked to the complex geological structure of the cross-section A-A'. Surface waves, which are generated in direction A-A' where the thickness of the Quaternario sediment cover decreases, mainly determine the seismic signal at period greater than 2 seconds. This element is too complex to be captured by the ground motion prediction equations. Additionally we can observe a strong amplification in the South of the Area which corresponds to the Appenninic area. This amplification is not observed in PGA.

3.2.3 Kriging for $\sigma_{0,sr}$

The random field of $\sigma_{0,sr}$ has been modelled as an isotropic process with constant and unknown drift for both PGA and $T = 4s$ so we can use ordinary kriging to predict the values on the grid. We report the maps for the prediction and variance in Figure 3.20 and 3.21, and a zoom on the faults in Figures 3.22.

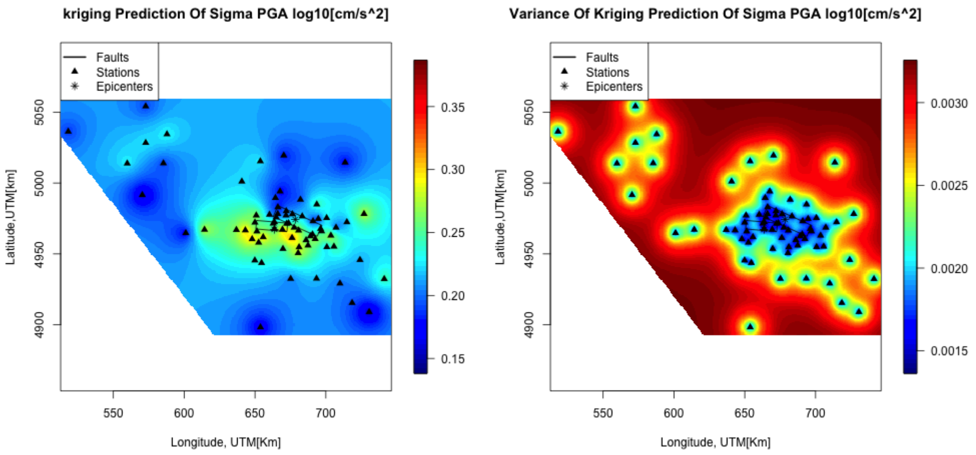

Figure 3.20: Ordinary kriging prediction and variance for the PGA.

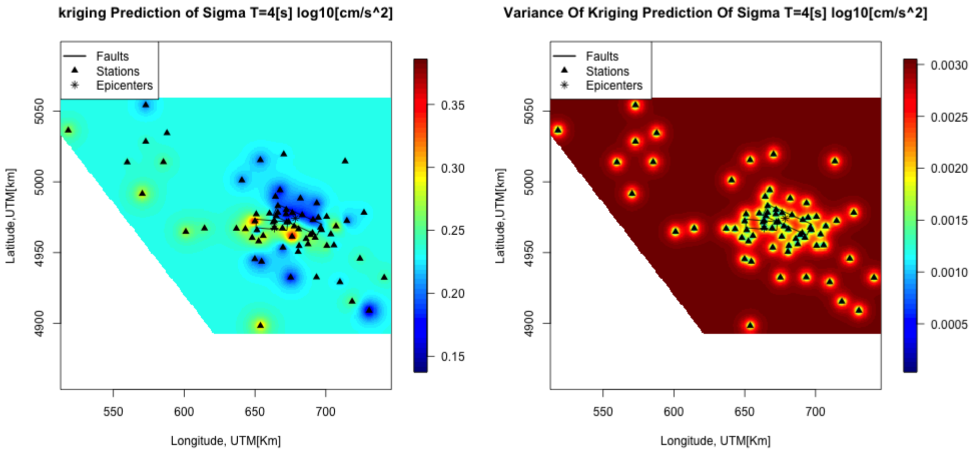

Figure 3.21: Ordinary kriging prediction and variance of $SA(T = 4s)$.

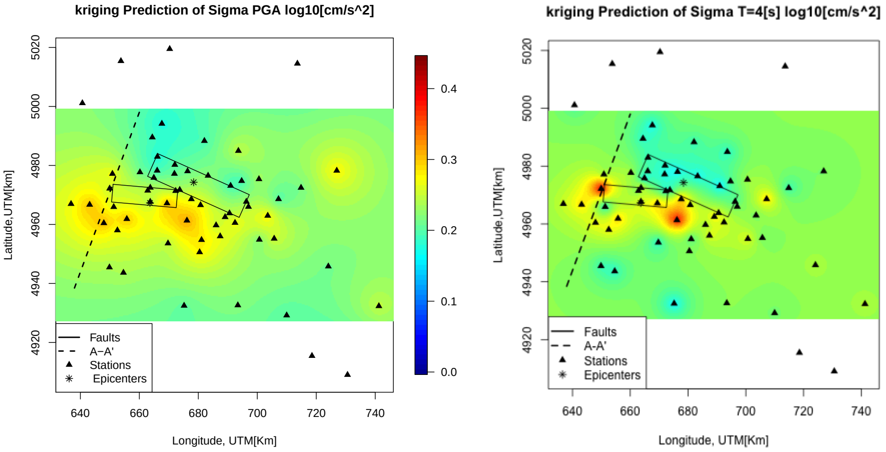

Figure 3.22: Ordinary kriging prediction of PGA and SA($T = 4s$), zoom on the faults. The colour scale has been modified to highlight local differences.

Comparing the PGA and $T = 4s$ kriging predictions for $\sigma_{0,sr}$ we observe that on the whole grid PGA has lower values but on the fault zone PGA and $T = 4s$ are very similar. We have a greater amplification from East to West below the faults zone.

3.3 Conditional Simulation

3.3.1 Introduction

The kriging prediction represents the average scenario. We would like to take into account the variability of the range of scenarios compatible with the data to understand what the best and worst case scenarios could be in a certain location. In order to consider the variance related to the prediction, we could simply add to the prediction the realization of a Gaussian noise centered in zero and with variance the kriging variance. However, this would neglect the spatial dependence among close locations. A better way to include the variance is represented by the Conditional Simulation.
Here we summarize the main concepts of the conditional simulation in the Gaussian case, by following (Chilès et Delfiner, 1999), with a particular reference to Sequential Gaussian Simulation.
Let $\boldsymbol{Z} = (Z_1, Z_2, ..., Z_M, .., Z_N)'$ be a vector collecting the random variables Z_i of the random field and suppose that we know the realization of the subvector $(Z_1 = z_1, Z_2 = z_2, ..Z_M = z_m)$.
Then the conditional distribution of $\boldsymbol{Z}$ given $Z_i = z_i, i = 1, ..., M$, can be factorized in the form

$$Pr\left\{z_{M+1} \leq Z_{M+1} < z_{M+1} + dz_{M+1}, ..., z_n \leq Z_N < z_N + dz_N | z_1, ..z_M\right\} =$$

$$Pr\left\{z_{M+1} \le Z_{M+1} < z_{M+1} + dz_{M+1}|z_1, ..z_M\right\} \times$$
$$Pr\left\{z_{M+2} \le Z_{M+2} < z_{M+2} + dz_{M+2}|z_1, ..z_M, z_{M+1}\right\} \times ...\times$$
$$Pr\left\{z_n \le Z_N < z_N + dz_n|z_1, ..z_M, z_{M+1}, ..Z_{N-1}\right\}$$

Therefore, we can simulate the vector $\boldsymbol{Z}$ sequentially by randomly selecting Z_i from the conditional distribution and including the outcome z_i in the conditioning dataset for the next step.
Once we have fixed a grid, the sequential simulation algorithm, following a stochastic path through the grid, repeats the following step (Bivand et al., 2013):

1. Computing the parameters of the conditional distribution based on both the original data and the values previously sampled
2. Sampling a new value
3. Adding the value to the dataset
4. Reaching a new location of the grid following the random path

At every new simulation (i.e. at every step of the Algorithm) the computational effort of point 1 increases and it takes more time to compute a new simulation. To avoid this problem, it is possible to set a maximum number of neighbourhoods (in our case 40) with respect to which computing the conditional distribution(Bivand et al. 2013). Here below we report the computation of parameters of point 1 in case of Gaussian random field recalling the definition of the Multivariate Normal Distribution and its conditional law(Chilès et Delfiner, 1999).

Definition 5 (Johnson et Wicherin)**.** *A p-dimensional normal density for the random vector $\boldsymbol{Z} = [Z_1, ...Z_p]'$ has the form:*

$$f(\boldsymbol{z}) = \frac{1}{2\pi^{\frac{p}{2}}|\Sigma|^{\frac{1}{2}}} e^{(\boldsymbol{z}-\boldsymbol{\mu})'\Sigma^{-1}(\boldsymbol{z}-\boldsymbol{\mu})/2}$$

where $-\infty < z_i < \infty, i = 1, 2, .., p$, $\boldsymbol{\mu}$ is a p×1 vector and $\boldsymbol{\Sigma}$ is a p×p positive definite matrix.
We shall denote this p-dimensional normal density by $\boldsymbol{Z} \sim \mathcal{N}_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$

Property 1. *Let $\boldsymbol{Z}=\begin{bmatrix}\boldsymbol{Z_1}\\ \boldsymbol{Z_2}\end{bmatrix}$ be distributed as $N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ with :*

$$\boldsymbol{\mu} = \begin{bmatrix}\boldsymbol{\mu_1}\\ \boldsymbol{\mu_2}\end{bmatrix}, \quad \boldsymbol{\Sigma} = \begin{bmatrix}\boldsymbol{\Sigma_{11}} & \boldsymbol{\Sigma_{12}}\\ \boldsymbol{\Sigma_{21}} & \boldsymbol{\Sigma_{22}}\end{bmatrix}, \quad |\boldsymbol{\Sigma_{22}}| > 0$$

Then the conditional distribution of $\boldsymbol{Z_1}$, given that $\boldsymbol{Z_2} = \boldsymbol{z_2}$, is normal and has mean

$$\boldsymbol{\mu_1} + \boldsymbol{\Sigma_{12}}\boldsymbol{\Sigma_{22}}^{-1}(\boldsymbol{z_2} - \boldsymbol{\mu_2})$$

and covariance

$$\mathbf{\Sigma_{11} - \Sigma_{12}\Sigma_{22}{}^{-1}\Sigma_{21}}$$

The Gaussian random fields assumption commonly adopted in geostatistics has been introduced in the geostatistical applications by Alabert and Massonat(1990) (Chilès et Delfiner, 1999).

In the seismology framework, other studies focused on the spatial correlation within event residuals adopted the same assumption. The assumption has been introduced in the applied seismology framework by (Park et al.,(2007)); Verros et al (2017) applied successive conditional simulation in order to estimate the within event residual, Baker et al. (2007) simulated the global residuals for loss estimation and Bradley et al. (2014) applied the conditional simulation based on the Gaussian distribution of within event residuals to develop a PGA map for the 2010-2011 Cantebury earthquakes.

In the next section we show the results of our conditional simulation applied to both the corrective term and $\sigma_{0,sr}$.

3.3.2 Sequential simulation for corrective term

Figure 3.23 represent the comparison between a simulated scenario and the kriging for the corrective term of PGA in the same scale.

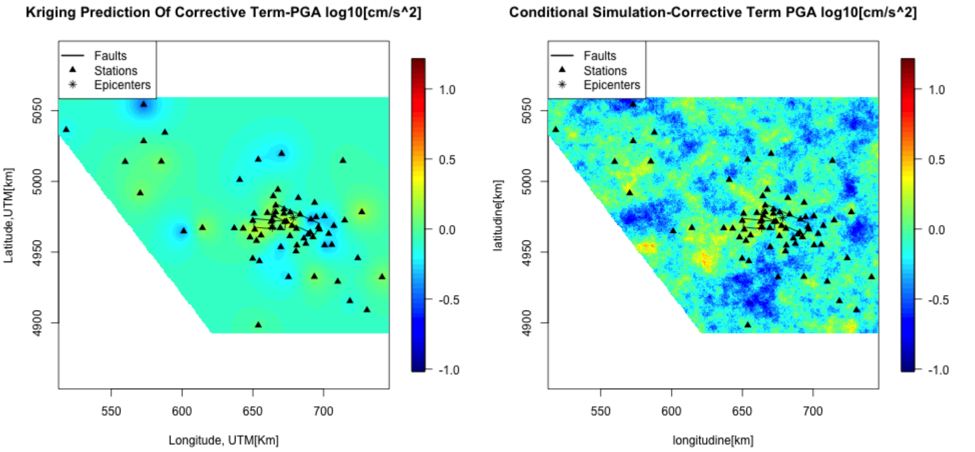

Figure 3.23: The kriging prediction and the conditional simulation for PGA.

We observe that, in the conditional simulation, maps are more rough than the ones of the kriging prediction; these scenarios differ from each other but we can also observe a common trends in the fault zone. We can consider a simulation and kriging prediction focused on the fault zone as reported in Figure 3.24.

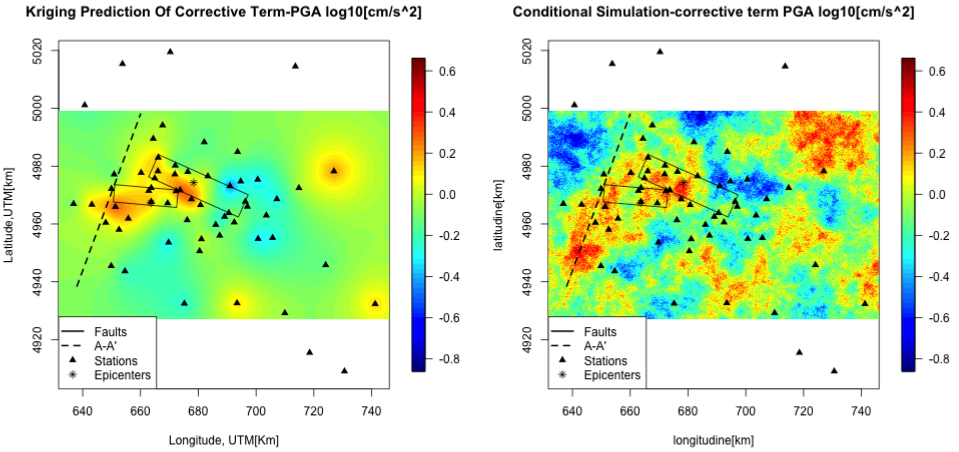

Figure 3.24: The kriging prediction and the conditional simulation, faults zone for PGA. The colour scale has been modified to highlight local differences.

We observe a strong amplification of the ground motion near the faults and the cross section A-A' and that the simulated amplifications reach higher values and cover a largest area in the fault zone than the kriging prediction. The kriging prediction represents the average scenario, this imply that in zone of greater amplification than kriging, there will be simulations with a smaller amplification and, as consequence, the worst case scenario (strong amplification) and the best case scenario (deamplification) are very different near the fault zone. As for PGA, we report the sequential simulation for period $T = 4s$. In Figure 3.25 we can observe the conditional simulations and the universal kriging prediction in the same scale.

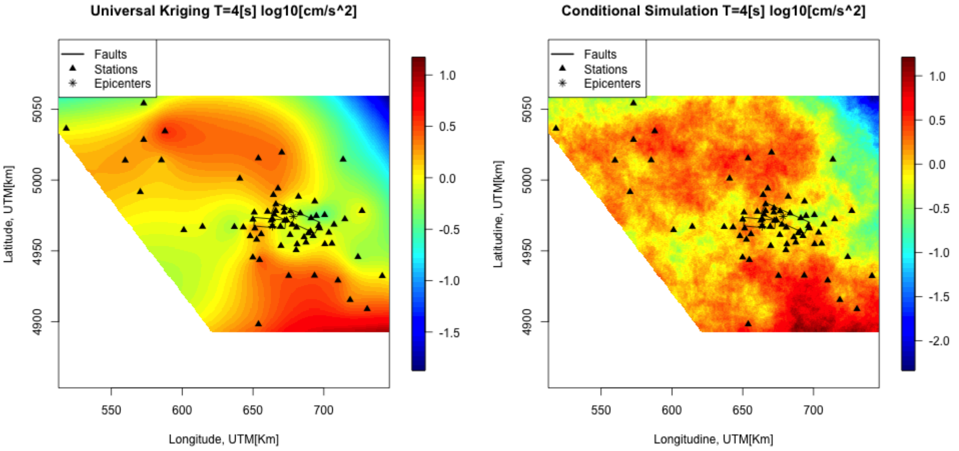

Figure 3.25: Comparison between the average scenario and the simulation T=4s.

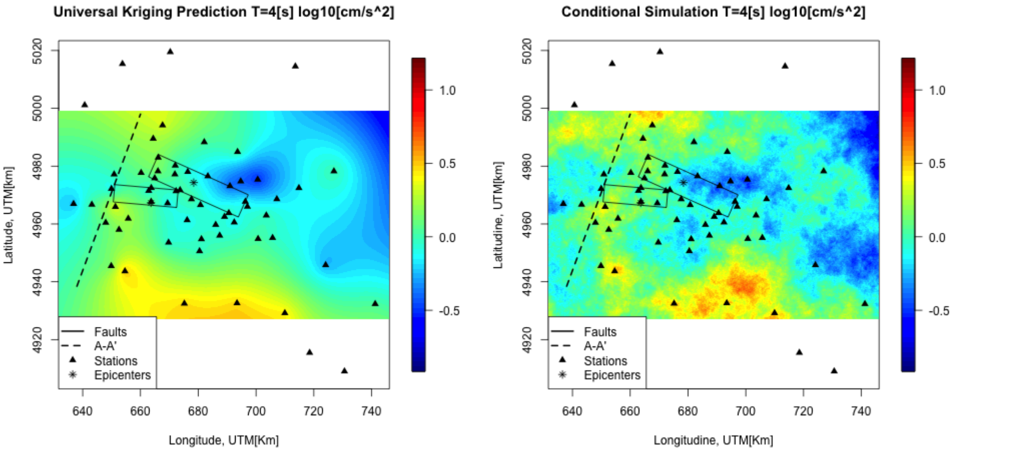

Figure 3.26: The kriging prediction and the conditional simulations for SA(T=4s), faults zone. The colour scale has been modified to highlight local differences.

The simulations assign to the three stations belonging to Appenninies, in the north of the map, higher values than the universal kriking prediction. Stronger Amplification with respect to the average scenario could be consequence of complex 2D and 3D site effects due to the presence of surface waves generated at the basin edges, with remarkable soil amplification at frequencies smaller than 1HZ(Lanzano et al.,2016).

3.3.3 Sequential simulation for $\sigma_{0,sr}$

The variable $\sigma_{0,sr}$ represents the aleatory uncertanty of the GMPE prediction. If the corrective terms and the GMPE captured all the systematic effects that influence the ground motion intensities in a specific location, then $\sigma_{0,sr}$ in that location would be very low. This entails that for the Area, in which we have greater values of $\sigma_{0,sr}$, we are not modelling all the complex sources and the propagation effects.

Let's consider first the PGA. In Figure 3.27 in the left panel we can observe the ordinary kriging prediction of $\sigma_{0,sr}$ and in the right we add the uncertainty in the kriging prediction through the conditional simulation:

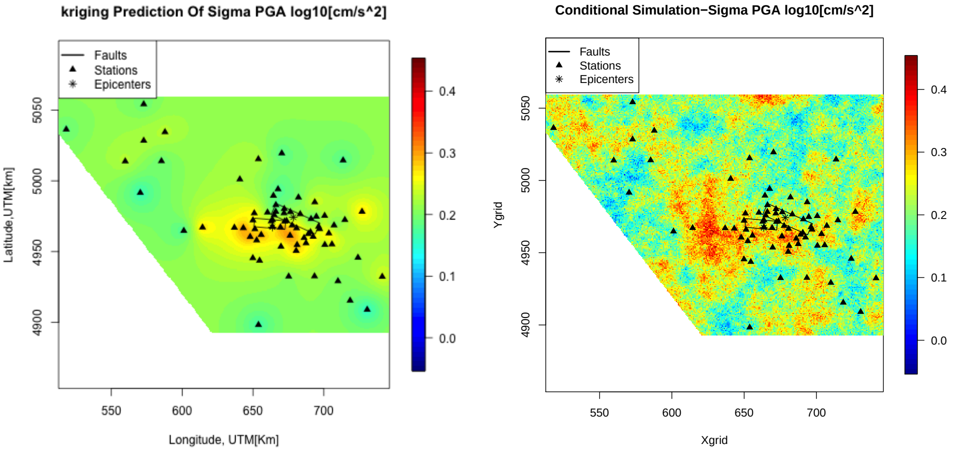

Figure 3.27: The kriging prediction and the conditional simulation.

Under the two faults, we notice a lighter area from East to West in which σ takes the greatest values. This trend indicates that in this area the model doesn't perform as well as in the other ones.
With a focus on the fault zone, it is easier to observe the trend.

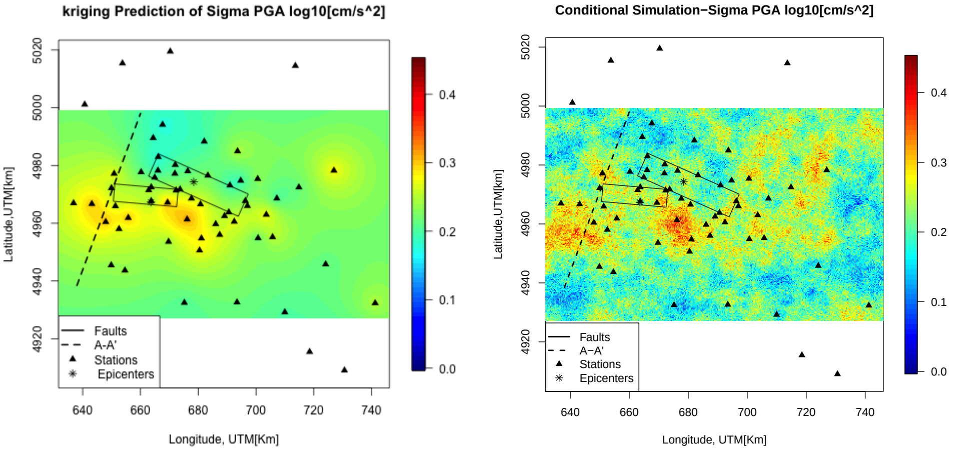

Figure 3.28: The kriging prediction and the conditional simulation, faults zone. The colour scale has been modified to highlight local differences.

In Figure 3.28 the aforementioned spatial trend is clearly visible both in the kriging prediction and in the simulation. The simulation higlights that $\sigma_{0,sr}$ could be even bigger in the zone around the faults and near the cross section A-A'. We report the same plot for $T = 4$ s.

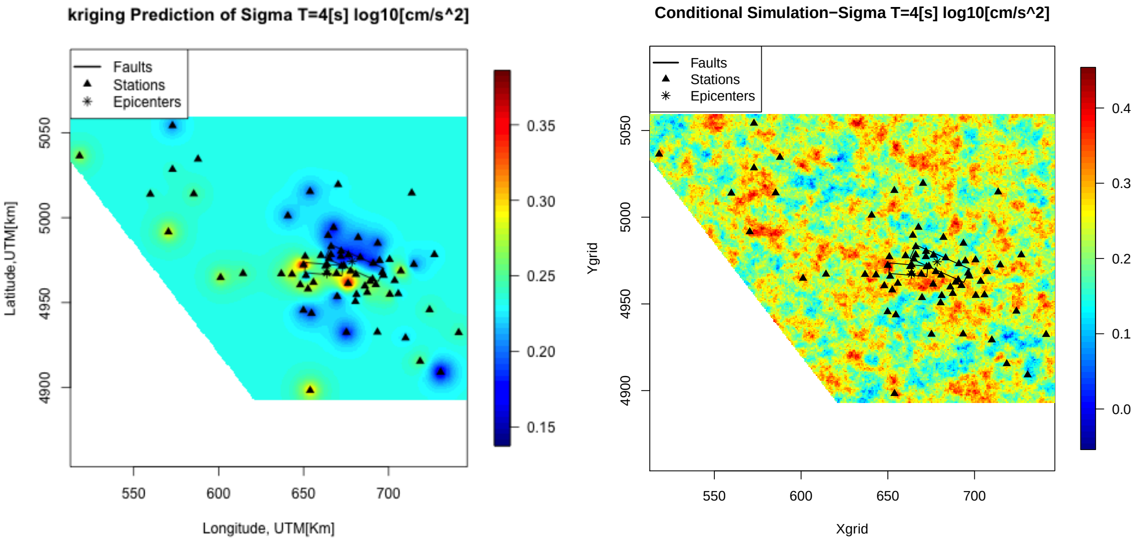

Figure 3.29: The kriging prediction and the conditional simulation.

The first result we observe is that the distribution is much more scattered and there isn't a clear zone where sigma has greater values. This observation is confirmed by the zoom on the fault area.

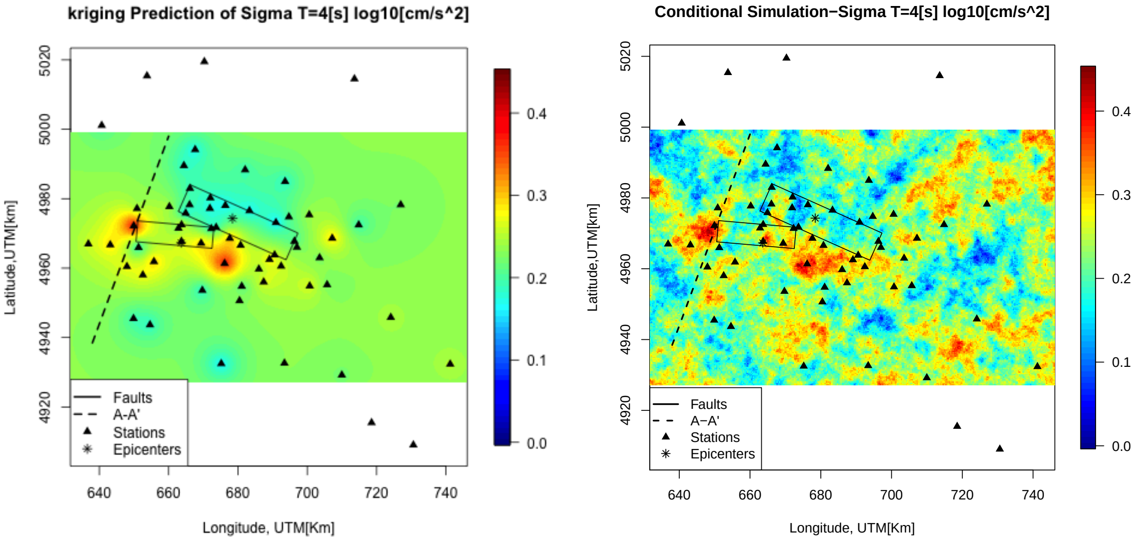

Figure 3.30: The kriging prediction and the conditional simulation, faults zone. The colour scale has been modified to highlight local differences.

3.3.4 Comparison of different periods in the faults zone

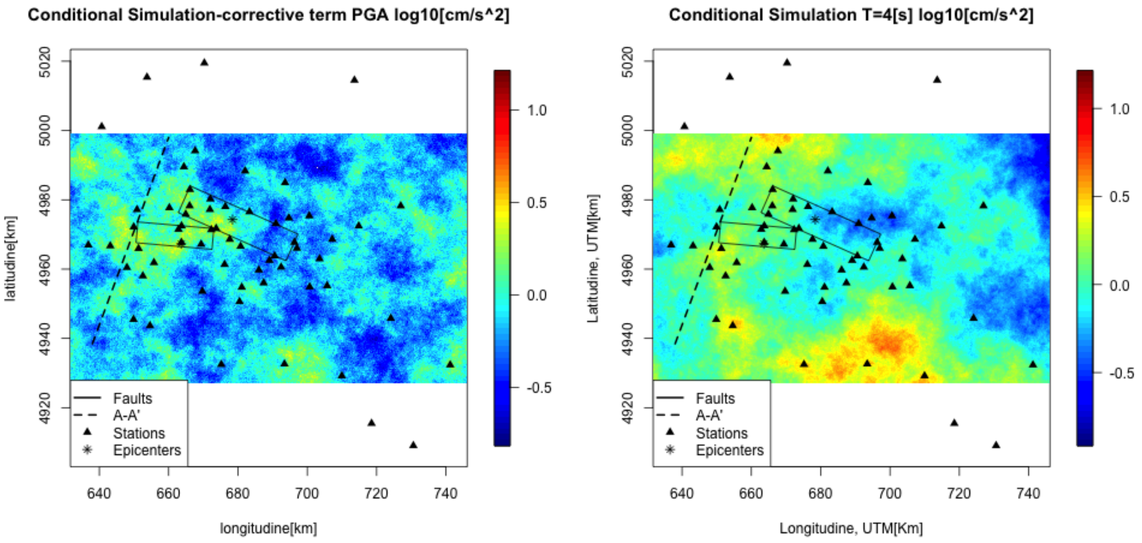

Figure 3.31: Comparison of conditional simulation, corrective term, PGA in left panel and Sigma SA(T=4s) in the right panel.

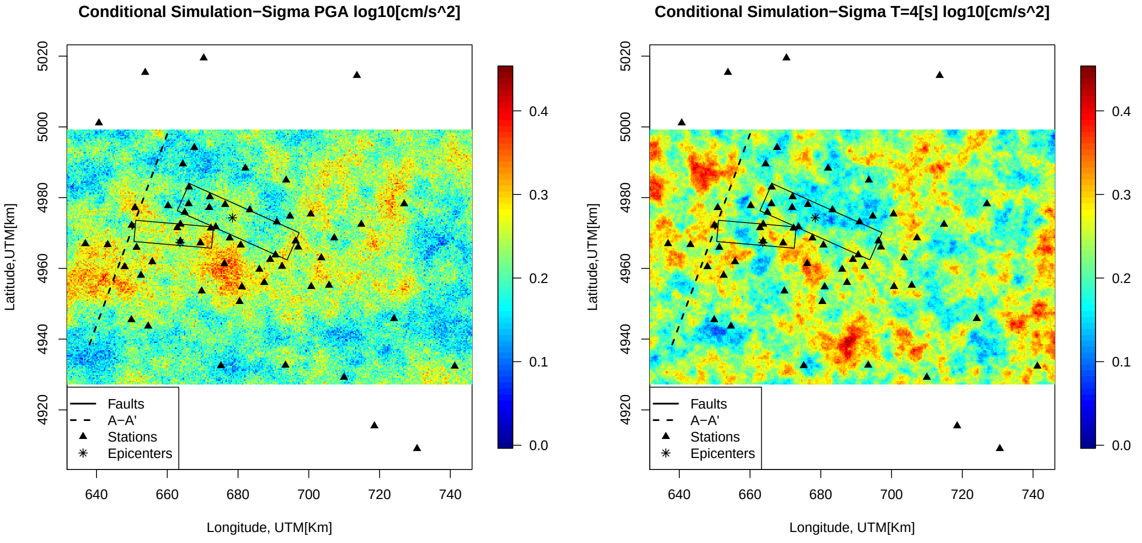

Figure 3.32: Comparison of conditional simulation, $\sigma_{0,sr}$ PGA in left panel and $\sigma_{0,sr}$ SA(T=4s) in the right panel.

The comparison between the corrective term at PGA and $T = 4s$ highlights that for both periods we can observe the amplification of the ground motion in the direction A-A' and that in the Appenninic area the amplification is much more stronger for $T = 4s$ than for PGA.

Chapter 4

Functional geostatistics for the joint analysis of intensity measures over a range of periods

4.1 State Of The Art

In the previous section we have considered four scalar random fields, namely the corrective term for PGA and SA($T = 4$), and $\sigma_{0,sr}$ for PGA and SA($T = 4$).
These models, unfortunately, don't take into account the correlation among SA at different periods as they only describes the correlation of SA at the same periods in different stations through different kriging techniques.
We are not taking advantage of all the available information. We could think to employ a multivariate approach and consider a vector random field $\boldsymbol{Z}_s$, which are vectors collecting all the twenty five periods of our dataset in different locations s.
This multivariate framework has been already applied by Loth (Loth and Baker, 2012) for the whithin event residuals, to model the spatial correlation of the spectral acceleration at six natural periods.
In our case, the number of periods is too high to employ the cross-covariograms. Indeed, fitting the variogram and cross-variogram models jointly for all periods would require to each period to have the same range and this is not feasible when one deal with vectors of twenty five components. Even if the fitting is possible, we would have to deal with the sequential simulation of 25 periods which needs a huge computational effort.
For this reason, we adopt a functional geostatistics approach instead of the multivariate framework. Functional geostatistics is a subfield of Object Oriented Spatial Statistics, a collection of techniques, algorithms and methods focused on the analysis of high dimensional and complex data. Typically, spatial distributions of curves and surfaces are the objects of the analysis (Menafoglio et al. 2017). The use of this approach to the problem we are considering would

allow to smooth the data, to manage their intrinsic noise and to compute, in a prediction or simulation setting, the value of SA at any arbitrary period in the range PGA and $T = 4s$ even if that period is not present in the dataset. Indeed, the model allows to recreate the entire shape of the corrective term and σ at all periods between PGA and $T = 4s$.

In the context of simulation of stochastic processes of functional data spatially distributed, we here follow the approach of Menafoglio et al. (2016), who propose to employ functional principal component analysis (FPCA) first, and then kriging and stochastic simulation of the FPC scores. An application to real data of the decomposition of the curves into a functional basis in order to move from a problem of kriging in infinite dimensional space to cokriging on basis coefficients is also proposed by Nerini et. al, (2009). In the next section we propose an application of these methods to seismology. This is a turning point for the development of seismic methods not based on a single period but on the joint consideration of all periods.

For more details about functional geostatistics see Delicado et. al,(2009).

4.2 Functional Data Analysis

4.2.1 Introduction

Given a set of discrete measured values $y_1, ., y_n$ representing the images (or the images plus a noise) of a certain function $x(t)$ evaluated in $t_1, ., t_n$ the first step of an analysis in the framework of Functional Data Analysis is to reconstruct the curve $x(t)$. This operation can be done through interpolation or by smoothing. In the first case we want to find a function able to pass through all y_i, in the second one we consider y_i as noisy data so that we want a function which passes only close to the points in order to avoid to fit the error.

4.2.2 Representing functions by basis functions

A basis for a functional space is a collection of functions ϕ_k $k = 1, ..., K$ such that each function of that space can be represented as the weighted linear combination of those functions ϕ_k

$$x(t) = \sum_{k=1}^{K} c_k \phi_k(t) \tag{4.1}$$

or in vector notation:

$$x(t) = \boldsymbol{c}'\boldsymbol{\phi} \tag{4.2}$$

where $\boldsymbol{\phi}$ is the column vector collecting the functions $\phi_k(t)$ in the basis.

Through K we can control the degree of smoothing the data y_j. We aim to use the smallest number K of functions ϕ_k able to reflect the features of the data.

4.2.3 Spline functions

Here we want to describe a spline function of order m on the interval $[a,b]$ with knots (or breakpoints) τ_l ,$l = 1, ..., L+1$.
In order to construct a spline function on the interval $[a,b]$, the first step is to separate the interval $[a,b]$ in L subintervals in some arbitrary points τ_l ,$l = 1, ..., L+1$ called knots or breakpoints. τ_1=a and $\tau_{L+1} = b$ are the external breakpoints. On each of these subintervals the spline function is a polynomial of order m where the order of a polynomial, by definition, is the number of his coefficients (e.g $y = ax + b$ has order 2) while the degree is the highest power of x (e.g $y = ax + b$ has order 1).
Adjacent polynomials join up smoothly at the breakpoint which separates them. The polynomials are constrained to be equal at their junction so the spline is a continuous function on $[a,b]$. In addition, all the derivates up to the $(m-2)th$ must also match up at these junctions.
The degree of freedom of the spline is define as the sum of the order(m) and the number of interior points $(L-1)$ and it represents the flexibility in the fit of the spline.

4.2.4 The B-spline basis for spline functions

A basis for the space of splines of order m on interval $[a,b]$ with knots τ is a collection of functions $\phi_k(t)$ which satisfies:

- each $\phi_k(t)$ is a spline of order m on $[a,b]$ and knot τ.
- every linear combination of $\phi_k(t)$ is a spline function.
- every spline function of order m, knots τ on [a,b] can be expressed as a linear combination of these $\phi_k(t)$

.

The basis can be created in several ways and the most popular one was introduced by De Boor (2001). The last important property to consider is that the space of Spline of order k on $[a,b]$ with knots τ is contained in the space of the Splines on the same interval, with the same knots and order $m+1$.

4.2.5 Smoothing Functional Data By Least Squares

We consider our data y_i as the sum of a deterministc model $x(t)$ and a noise ε:

$$y_j = x(t_j) + \varepsilon_j \tag{4.3}$$

and we aim to learn the function $x(t)$. First we can see the function as the linear combination of a vector of weights and the functions of a basis

$$x(t) = \sum_k^K c_k \phi_k(t) = \boldsymbol{c}'\boldsymbol{\phi} \tag{4.4}$$

where the vector $\boldsymbol{c}$ of length K contains the coefficients c_k. Now we need to estimate the coefficients able to minimizing the sum of square errors which is defined as

$$SMSSE(\boldsymbol{y}|\boldsymbol{c}) = (\boldsymbol{y} - \boldsymbol{\Phi c})'(\boldsymbol{y} - \boldsymbol{\Phi c}) \tag{4.5}$$

where $\boldsymbol{\Phi}$ is the n (number of knots) × K (number of functions in the basis) matrix containing the values $\phi_k(t_j)$. By setting the derivative of $SSE(\boldsymbol{y}|\boldsymbol{c})$ with respect to $\boldsymbol{c}$ at zero we obtain:

$$2\boldsymbol{\Phi\Phi' c} - 2\boldsymbol{\Phi' y} = 0 \tag{4.6}$$

Finally solving the equation with respect to $\boldsymbol{c}$, we can find $\widehat{\boldsymbol{c}}$ that minimizes the least square criterion and the values of the target variable estimated by the model $\widehat{\boldsymbol{y}}$

$$\widehat{\boldsymbol{c}} = \boldsymbol{(\Phi'\Phi)^{-1}\Phi' y} \tag{4.7}$$

$$\widehat{\boldsymbol{y}} = \boldsymbol{\Phi}\widehat{\boldsymbol{c}} = \boldsymbol{\Phi(\Phi'\Phi)^{-1}\Phi' y} \tag{4.8}$$

4.2.6 Smooting functional data with a roughness penality

The spline smoothing method, previously described, is able to find the function that minimizes the sum of square erros.

$$\sum [y_j - x(t_j)]^2 \tag{4.9}$$

If the function has sufficient degrees of freedom,by minimizing SSE, it tends to interpolate all the points. However, on the one hand we want to ensure a good fit of the data on the other hand we don't want to fit the noise. In order to avoid interpolation, we penalize the functions $x(t)$ that are too much locally variable and manifest a rapid local variation. First we have to define the function roughness.

4.2.7 Roughness

The curvature of a function x(t) in t is defined as the square of the second derivative $[D^2x(t)]^2$. Note that straight lines, as we expected, have zero curvature accordingly with this definition. The roughness is defined as the integral over the domain of the curvature(or the square of the L2 norm of the second derivative):

$$PEN_2(x) = \int [D^2x(s)]^2 ds \tag{4.10}$$

Functions x(t) which manifests rapid local variation in the first derivative will have high $PEN_2(x)$ value. Instead of minimize SSE, to avoid overfitting, we'll look for function x(t) able to minimize $PENSSE_\lambda$ define as

$$PENSSE_\lambda(x|y) := SSE + \lambda PEN_2(x) \tag{4.11}$$

In this framework a famous theorem (De Boor, 2002) ensures that the function that minimize these quantity is a cubic spline with knots in $t_j, j = 1, ..., n$, where $(t_1, y_1)...(t_n, y_n)$ are the observed data. Then, as basis, we could employ a B-spline basis of order four and same knots. The method is often called cubic spline smoothing. As we did for SSE, we compute the coefficients $\widehat{\boldsymbol{c}}$ of the model and the predicted values. First we expess $PEN_2(x)$ in vector notation as

$$PEN_2(x) = \boldsymbol{c'Rc} \tag{4.12}$$

where

$$\boldsymbol{R} = \int D^m \boldsymbol{\Phi(s)} D^m \boldsymbol{\Phi(s)'} ds \tag{4.13}$$

and $\boldsymbol{\phi(t)}$ is the column vector collecting the basis functions evaluated in t

$PENSSE_\lambda$ can be rewritten as:

$$PENSSE_m(\boldsymbol{y}|\boldsymbol{c}) = \boldsymbol{(y - \Phi c)'(y - \Phi c) + \lambda c' R c} \tag{4.14}$$

To find the value $\boldsymbol{c}$ that minimizes the quantity compute the derivative and set equal to 0

$$-2\boldsymbol{\Phi' y} + \boldsymbol{\Phi\Phi c} + \boldsymbol{\lambda R c} = 0 \tag{4.15}$$

By solving the equation with respect to $\boldsymbol{c}$ we find $\widehat{\boldsymbol{c}}$:

$$\widehat{\boldsymbol{c}} = (\boldsymbol{\Phi'\Phi} + \lambda \boldsymbol{R})^{-1} \boldsymbol{\Phi y} \tag{4.16}$$

4.3 Application to data

First we want to study the corrective term. To each of the 71 stations we have linked a vector of 25 components which represents SA at 25 different periods, from PGA to $T = 4s$.

In Figure 4.1 we plot simultaneously all the 71 station. Each vector is reported as a collection of segments obtained by linking the single values observed in that specific station.

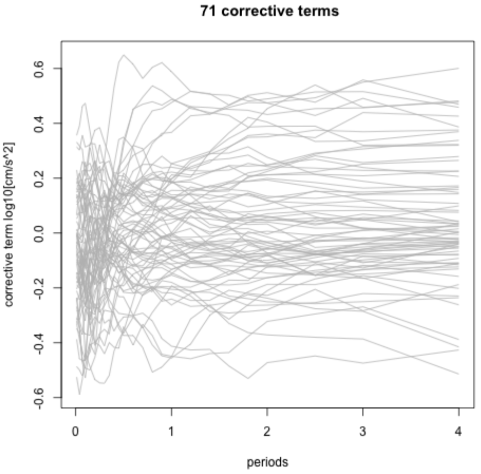

Figure 4.1: The original corrective terms at different periods, for different stations.

Then we have to to find, for each station, the curve that best fit the corrective terms as function of the period. We follow the aforementioned smoothing spline method. The first step of the algorithm is to fix an order four B-spline basis of K=25 splines:

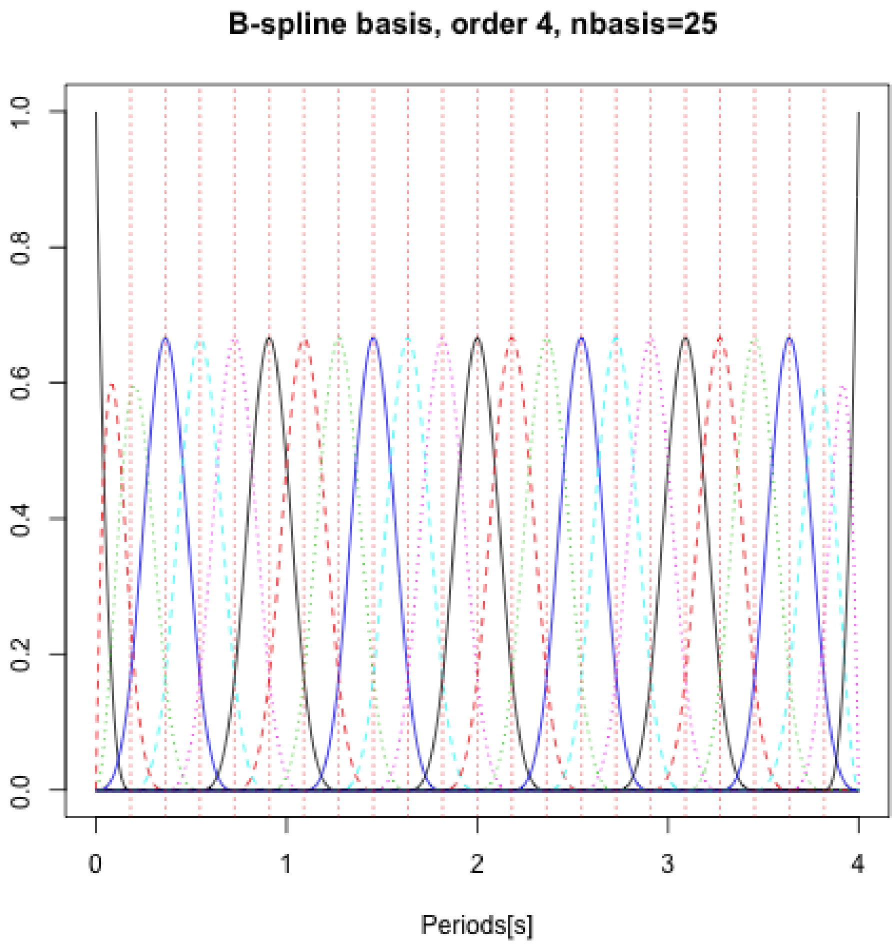

Figure 4.2: B-spline basis.

We choose the B-spline instead of a Fourier basis because we don't have evidence of periodicity in our data while the order four is a consequence of the De Boor's Theorem (Ramsay et Silverman, 2005).

We fix K, the number of Splines of the basis, equal to 25 which is the number of knots even if it's better to use a small value for K since, by increasing K, we increase both the degrees of freedom of the spline and the possibility of overfitting. However, we solve the overfitting problem by minimizing PENSSE. In this way, the PEN_λ term will penalize the functions which are too much locally variable and that manifest rapid local variation.

The value of λ is chosen between different values through the Generalized Cross Validation. Once the value of λ is fixed, we compute the coefficients $\boldsymbol{c}$ as already

explained in the last section.
We report the smoothing splines and the original data for some stations located in the Appenninic Area and in the faults zone(Figures 4.3).

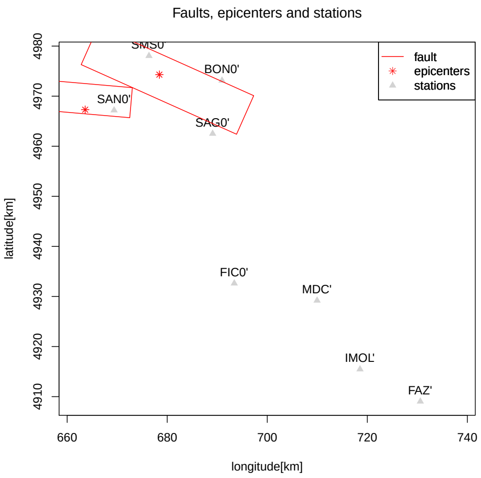

Figure 4.3: Stations located in the Appenninic Area and in the faults zone.

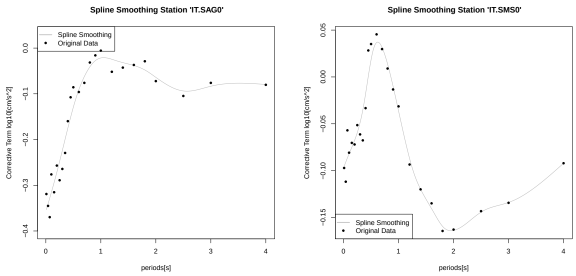

Figure 4.4: Smoothing spline for Stations SAG0 and SMS0.

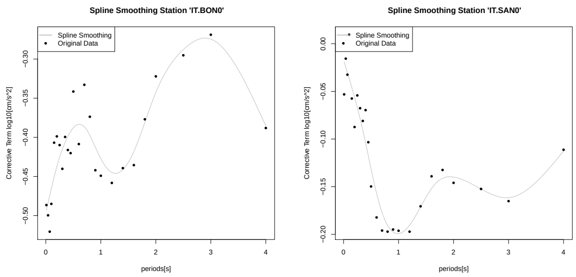

Figure 4.5: Smoothing spline for stations BON0 and SAN0.

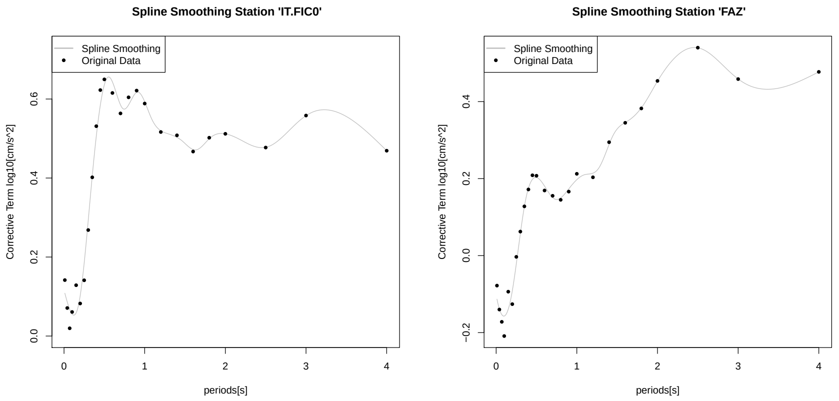

Figure 4.6: Smoothing spline for stations FICO,FAZ.

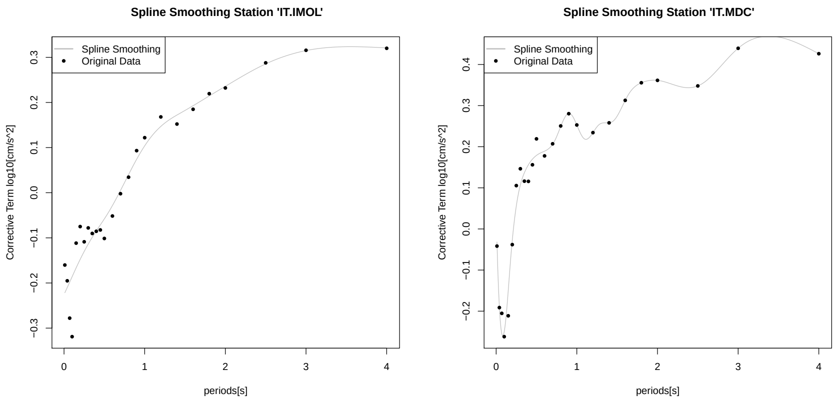

Figure 4.7: Smoothing spline for stations MOLI,MDC

As expected, the splines don't interpolate the data due to the roughness penalization. In this way we are reducing the noise affecting the data. In the group of splines, for stations located in the faults zone, we can observe that SMS0 and SAN0 have similar shapes, the global maximum and minimum of SAN0 are the same of SMS0 but shifted of one toward left. In the Appenninic group a common patter is clearly recognizable and the spline value of FAZ,IMOL,MDC increase moving toward longer periods. We can also observe that FICO and SAGO are very similar in the shape and close on the map. These pictures tell us that there could be a connection between shape of the curve and its location.

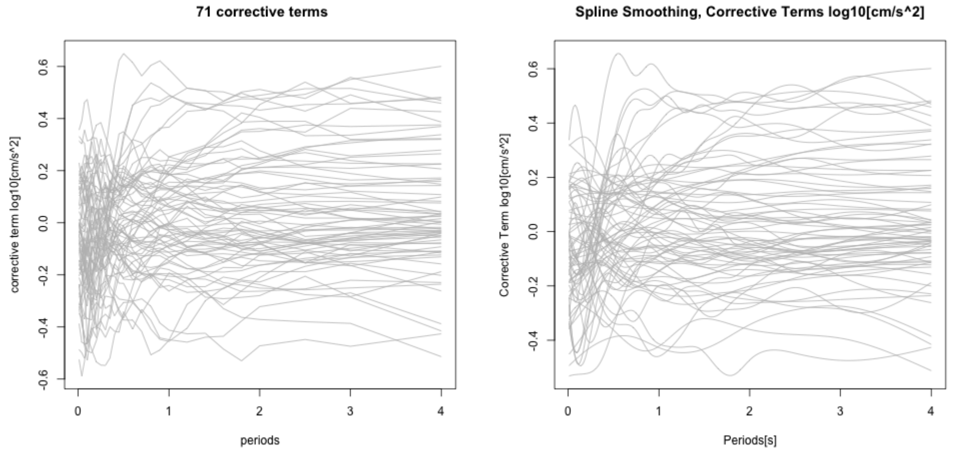

Figure 4.8: Original data on the left and on the right the smoothing splines.

$\sigma_{0,sr}$ is processed in the same way and with the same basis. In Figures 4.9, 4.10, 4.11, 4.12, 4.13 we report the original data and the smoothing splines.

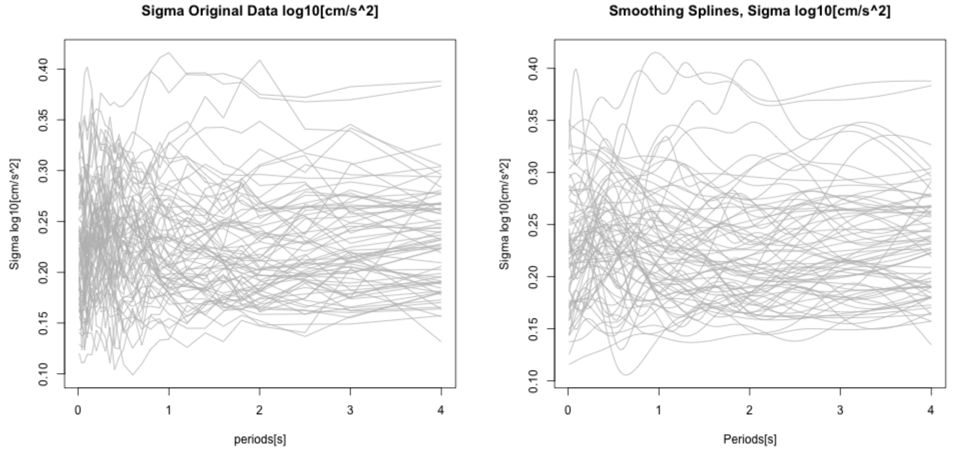

Figure 4.9: Original data on the left and on the right the smoothing splines.

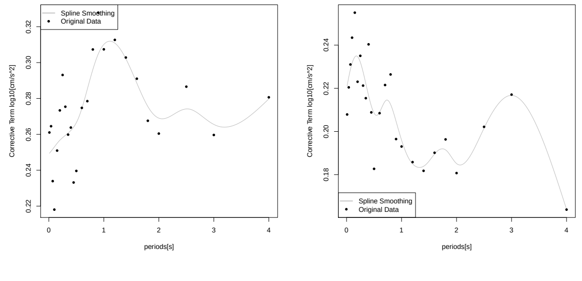

Figure 4.10: Smoothing spline for Stations SAG0 and SMS0 ($\sigma_{0,sr}$).

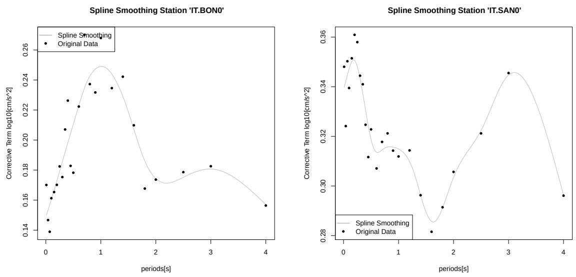

Figure 4.11: Smoothing spline for Stations BON0,SAN0 ($\sigma_{0,sr}$).

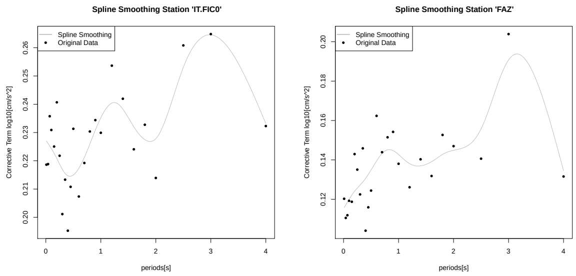

Figure 4.12: Smoothing spline for Stations FICO, FAZ ($\sigma_{0,sr}$).

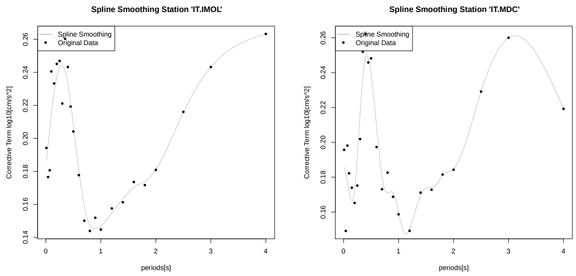

Figure 4.13: Smoothing spline for Stations MOLI, MDC ($\sigma_{0,sr}$).

4.4 Principal Component Analysis

4.4.1 PCA for multivariate data

The Principal components analysis provides a way of looking at the covariance structure of data that can reveal new relations among data. Its general objectives are dimensionality reduction and interpretation (Johnson et Wichern, 2007).
Let $\boldsymbol{x_1}, ..\boldsymbol{x_N}$ be the observed data. A Principal Component Analysis generally proceed according to the following steps:

1. substract to each $\boldsymbol{x_i}$ the mean $\overline{\boldsymbol{x}} = \frac{1}{N}\sum \boldsymbol{x_i}$

2. find $\boldsymbol{\varepsilon_1}$ such that
$$f_{i1} = \boldsymbol{\varepsilon_1}'\boldsymbol{x_i} \tag{4.17}$$

have the largest variance

$$\frac{1}{N}\sum_{i} f_{i1}^{2} \tag{4.18}$$

under the condition

$$\|\boldsymbol{\varepsilon_1}\|_2^2 = 1 \tag{4.19}$$

3. at the m step find vector $\boldsymbol{\varepsilon_m}$ such that

$$f_{im} = \boldsymbol{\varepsilon_m}'\boldsymbol{x_i} \tag{4.20}$$

have the largest variance

$$\frac{1}{N}\sum_{i} f_{im}^{2} \tag{4.21}$$

and satisfies

$$\|\boldsymbol{\varepsilon_m}\|_2^2 = 1 \tag{4.22}$$

$$\boldsymbol{\varepsilon}_k'\boldsymbol{\varepsilon}_m = 0, \quad \forall k < m \tag{4.23}$$

The $f_{i,m}$ are called scores and the $\boldsymbol{\varepsilon_m}$ is the $m-th$ principal direction.

4.4.2 Defining PCA for functional data

When we move from vector data to functional data the vectors $\boldsymbol{\beta}$ and $\boldsymbol{x}$ become functions $\beta(s)$, $x(s)$. and the inner product is defined as:

$$\int \beta(s)x(s)ds \tag{4.24}$$

and the L^2 Norm as

$$\sqrt{\int x(s)^2 ds} \tag{4.25}$$

Now the PCA for functional data, called Functional Principal Component Analysis (FPCA, Ramsay and Silverman, 2005) can be obtained by replacing the functions, inner product for functions and L^2 norm for functions to the aforementioned procedure for vector data. Note that before doing the FPCA it's necessary to center the functions substracting the cross mean

$$\overline{x_n}(t) = \frac{1}{N}\sum_{i=1}^{N} x_i(t) \tag{4.26}$$

4.4.3 Eigenanalysis

In order to find the principal directions previously defined we have to solve the variance maximization problem by finding eigenvalues and eigenvectors of the eigenequation:

$$\boldsymbol{V\varepsilon} = \rho\boldsymbol{\varepsilon} \tag{4.27}$$

where V is the covariance matrix.
The eigenvector $\boldsymbol{\varepsilon}$, with the highest eigenvalues ρ, is the first principal direction and ρ represents the maximized variance; the second principal direction has the second highest eigenvalues and so on.
In the functional case, first we define the covariance function $v(s,t)$

$$v(s,t) = N^{-1}\sum_{i=1}^{N} x_i(s)x_i(t) \tag{4.28}$$

then we introduce the eigenequation

$$\int v(s,t)\varepsilon(t)dt = \rho\varepsilon(s) \tag{4.29}$$

Finally to retrieval the form (27), we define an operator V as follow

$$V\varepsilon = \int v(\cdot,t)\varepsilon dt \tag{4.30}$$

We can express the operator V as the product of matrix and vector as

$$V\varepsilon \approx \boldsymbol{VW}\widetilde{\boldsymbol{\varepsilon}} \tag{4.31}$$

where the matrix $\boldsymbol{V}$ contains the values $v(s_j, s_k)$, $\widetilde{\boldsymbol{\varepsilon}}$ collects the values $\varepsilon(s_j)$ and the matrix $\boldsymbol{W}$ is a diagonal matrix collecting the weights that activate the trapezoidal rule to compute integrals:

$$\int f(s)ds \approx h[\frac{f(s_1)}{2} + \sum_{j=2}^{n-1} f(s_j) + \frac{f(s_n)}{2}] \tag{4.32}$$

We are ready to introduce the functional eigenequation in matrix form:

$$\boldsymbol{VW}\widetilde{\boldsymbol{\varepsilon}} = \rho\widetilde{\boldsymbol{\varepsilon}} \tag{4.33}$$

Furthermore the eigenvectors must satisfy

$$\widetilde{\boldsymbol{\varepsilon}}'_{\boldsymbol{m}}\boldsymbol{W}\widetilde{\boldsymbol{\varepsilon}}_{\boldsymbol{m}} = 1 \quad \forall m \tag{4.34}$$

We can rewrite the equation 31 as

$$\boldsymbol{W}^{\frac{1}{2}}\boldsymbol{V}\boldsymbol{W}^{\frac{1}{2}}\boldsymbol{u} = \rho\boldsymbol{u} \tag{4.35}$$

Where $\boldsymbol{u} = \boldsymbol{W}^{\frac{1}{2}}\widetilde{\boldsymbol{\varepsilon}}$ and $\boldsymbol{u}'\boldsymbol{u} = 1$

Then the procedure to find eigenvectors is:

- Choose n, the w_j and the s_j
- Compute the eigenvalues ρ_m and eigenvectors $\boldsymbol{u}_m$ of $\boldsymbol{W}^{\frac{1}{2}}\boldsymbol{V}\boldsymbol{W}^{\frac{1}{2}}$
- Compute

$$\widetilde{\varepsilon} = \boldsymbol{W}^{-\frac{1}{2}}\boldsymbol{u}_m$$

The most relevant eigenvectors (in terms of captured variance) can be used to create the basis of a new space on which we can project the original functions. In this way we can move from a virtually infinite-dimensional space to a lower dimensional space with a limited loss of information. In the next section we are going to show how dimensionality reduction can be very usefull in our framework.

4.5 PCA Application

In Figure 4.14, we can observe the first 5 principal functional directions for the corrective term.

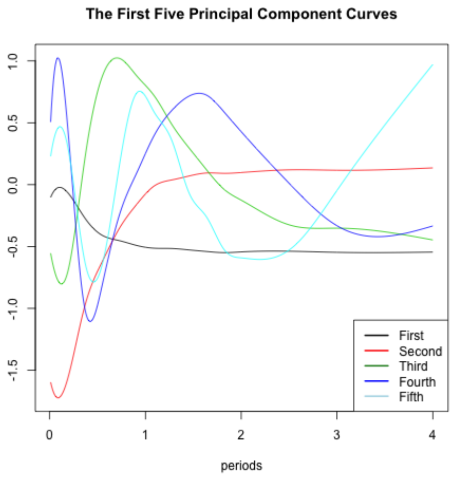

Figure 4.14: Principal Component curves for corrective term.

We would like to project the smoothing splines previously computed from the space generated by the B-spline basis to the space with principal functional directions as basis. We want to reduce the size of the space while retaining as much of the information as possible. A way too see how much information we retain is to look at the cumulative variance. In Figure 4.15 we report how the cumulative variance captured increases by adding a functional direction to the space of projection.

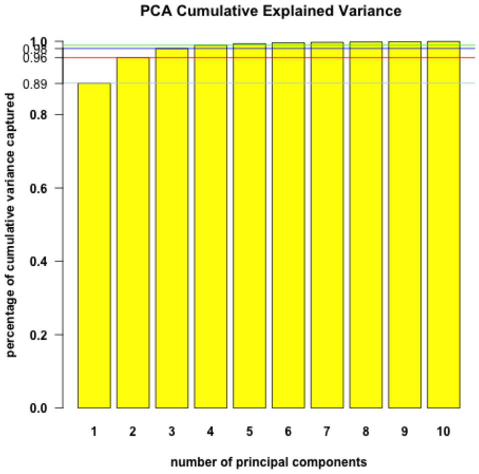

Figure 4.15: Cumulative variance.

It seems that the first three/four components are sufficient to catch most of the variability. The boxplot of scores, which are the projection of the smoothing splines on the principal directions, seems to highlight a not negligible variance till the fourth principal direction.

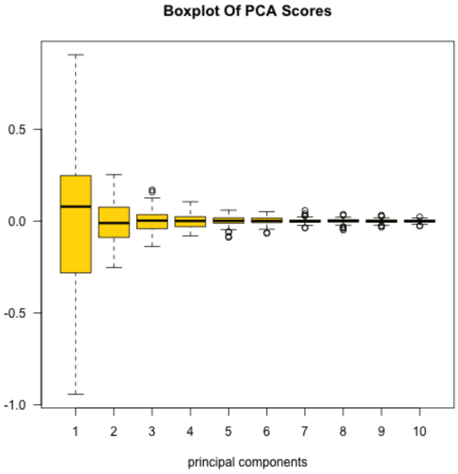

Figure 4.16: Boxplot of scores.

We now want to understand if four eigenfunctions are enough to get a good approximation of the original smoothing splines. In the Figures 4.17 and 4.18 we compare the original smoothing splines and their projections by employing PCA with two, three and four scores.

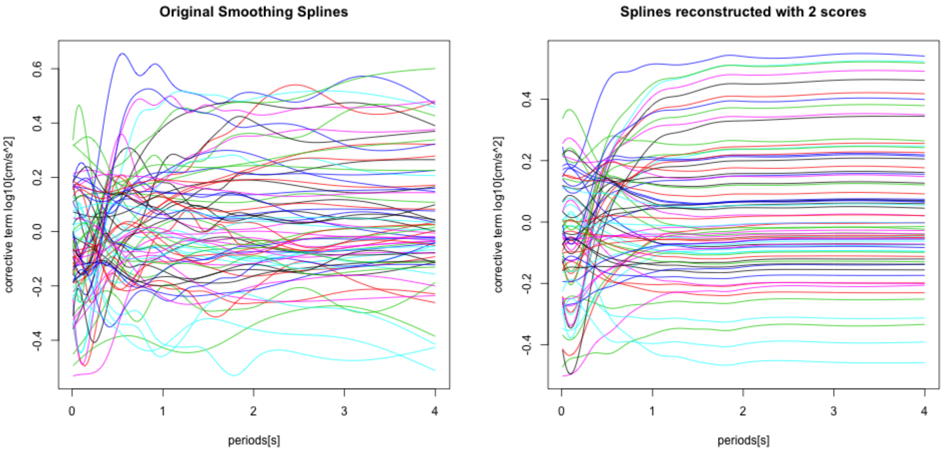

Figure 4.17: Smoothing spline projected on different spaces.

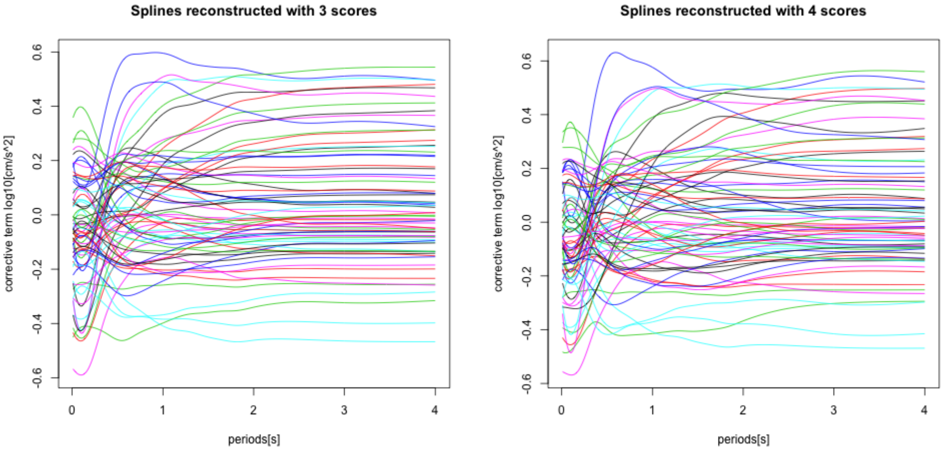

Figure 4.18: Smoothing spline projected on different spaces.

Two scores are not enough to reconstruct the splines while with three or four scores we get a much better approximation. The fourth score helps us to describe the change in the concavity as we can see in the blue spline with the highest values.

We project the stations of the Appenninic and the faults zone in the space generated by the first four eigenfunctions:

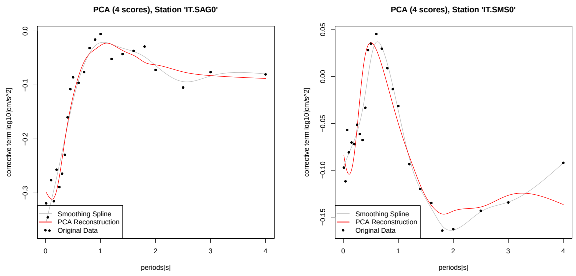

Figure 4.19: Smoothing spline projection for stations SAG0 and SMS0.

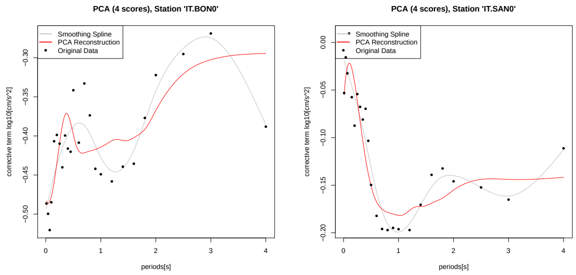

Figure 4.20: Smoothing spline projection for station BON0 and SAN0.

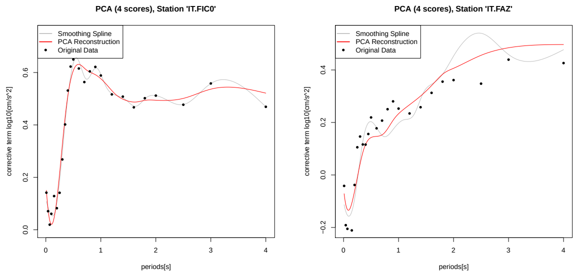

Figure 4.21: Smoothing spline projection for stations FIC0 and FAZ.

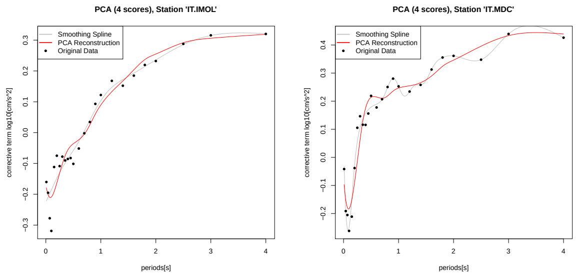

Figure 4.22: Smoothing spline projection for stations IMOL and MDC.

We can repeat the same procedure for $\sigma_{0,sr}$.

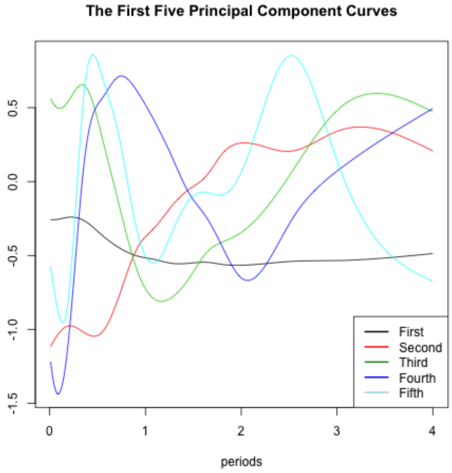

Figure 4.23: Principal Component curves for corrective term.

To understand the number of informative PCA, we look at the barplot for the cumulative varinace:

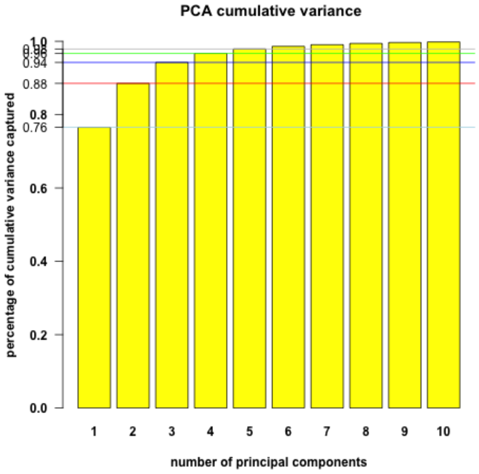

Figure 4.24: Cumulative variance

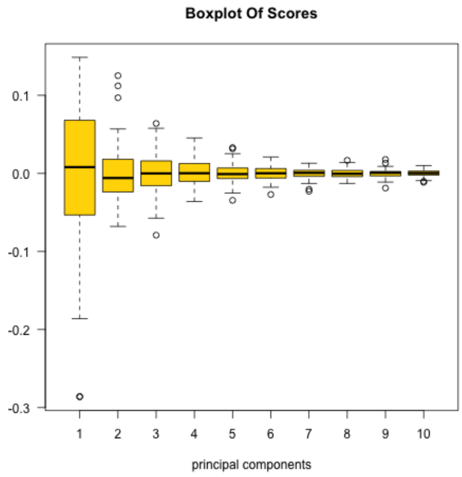

Figure 4.25: Boxplot of scores

Here, the right number of functions for the basis could be greater than five. Of course, we know that if we increase the number of functional directions, we get a better approximation but, as we will see in the next section, it is crucial to employ only the minimum number of informative principal directions to improve the computational performance. In the Figure 4.26 we compare the original smoothing splines and their projections by employing PCA with five scores.

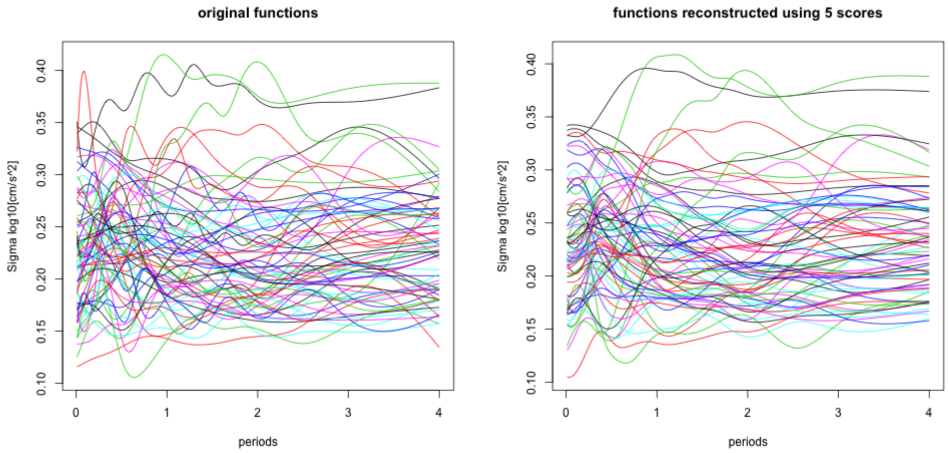

Figure 4.26: Original smoothing spline and smoothing spline projected.

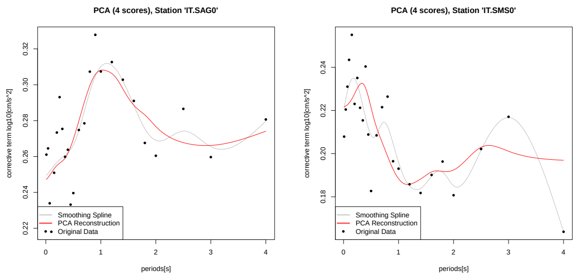

Figure 4.27: Smoothing spline projection for stations located in the faults zone.

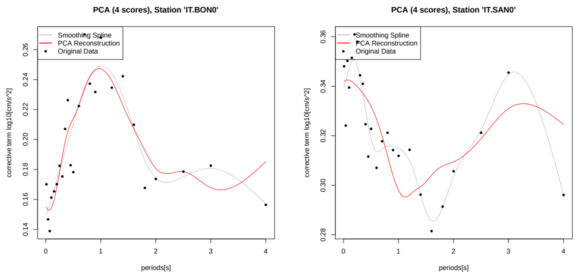

Figure 4.28: Smoothing spline projection for stations BON0 and SAN0

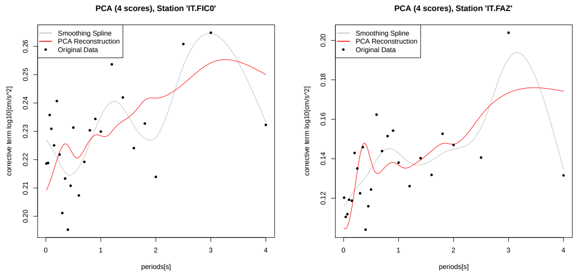

Figure 4.29: Smoothing spline projection for stations FIC0 and FAZ

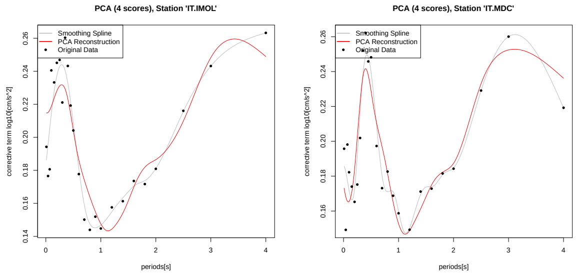

Figure 4.30: Smoothing spline projection for stations IMOL and MDC

We can observe that the red curves obtained with PCA are a good approximation of the original smoothing splines and only in FIC0 the PCA curve seems to have a different behaviour when the concavity of the function changes suddently. By adding scores, we are able to catch this behaviour. However, in order to avoid the overfitting, we don't want to employ models too complex since our data could be affected by noise and models with a large number of degree of freedom would fit that noise.

4.6 Multivariate Sequential Simulation

Through the FPCA we are able to create a bijection between the smoothing spline $g_{\boldsymbol{s}}$ in station $\boldsymbol{s}$ and a vector of scores

$$\boldsymbol{sc_s} = (sc_{1,\boldsymbol{s}}, sc_{2,\boldsymbol{s}}, sc_{3,\boldsymbol{s}}, sc_{4,\boldsymbol{s}})'$$

where the i-th score represents the projection of the curves on the i-th functional principal direction f_i:

$sc_{1,\boldsymbol{s}} = \int g_{\boldsymbol{s}f_1}$ $\quad sc_{2,\boldsymbol{s}} = \int g_{\boldsymbol{s}f_2}$ $\quad sc_{3,\boldsymbol{s}} = \int g_{\boldsymbol{s}f_3}$ $\quad sc_{4,\boldsymbol{s}} = \int g_{\boldsymbol{s}f_4}$

In this way, instead of working with functions, we can reduce the dimensionality of the problem and we can work with vectors of 4 components. Given the 71 vectors of 4 scores, we would like to simulate, in a new location $\boldsymbol{s_0}$, the corresponding vector of four scores $\boldsymbol{sc_{s_0}} = (sc_{1,\boldsymbol{s_0}}, sc_{2,\boldsymbol{s_0}}, sc_{3,\boldsymbol{s_0}}, sc_{4,\boldsymbol{s_0}})$ and then by a linear combination of the vector and the FPCA basis $\{f_1, f_2, f_3, f_4\}$ we can reconstruct the shape of the function of the corrective term versus the period in that location

$$g_{\boldsymbol{s_0}} = \sum_{i=1}^{4} f_i sc_{i,\boldsymbol{s_0}} \tag{4.36}$$

The conditional simulation requires to estimate the covariance matrix of the multivariate random field. The covariance can be recovered from the sample variogram, as in the scalar random field, but in the multivariate framework we need to introduce the concepts of the cross-covariance and the cross-variogram. First we can observe that, by construction, each of the four score random field has zero mean.

The cross-covariance functions (Chiles et Delfiner,1999) of a p-dimensional(in our case p=4) stationary random field $\boldsymbol{Z(x)} = (Z_1(\boldsymbol{x}), ..., .Z_p(\boldsymbol{x}))'$ with mean vector $m(\boldsymbol{x}) = (m_1(\boldsymbol{x}), ..., m_p(\boldsymbol{x}))'$(in our case $\boldsymbol{m(x) = 0} \quad \forall \boldsymbol{x}$) are defined by

$$C_{i,j}(h) = E[Z_i(x) - m_i][Z_j(x+h) - m_j] \tag{4.37}$$

If we permute the variables, we get another cross-covariance

$$E[Z_j(x)Z_i(x+h)] = C_{ji}(h) \tag{4.38}$$

note also that

$$C_{ij}(h) = E[Z_i(x)Z_j(x+h)] = E[Z_j(x+h)Z_i(x)] = C_{ji}(-h) \tag{4.39}$$

and in general

$$C_{ij}(-h) \neq C_{ij}(h) \tag{4.40}$$

The total covariance matrix will have the form:

$$\Sigma = \begin{bmatrix} \boldsymbol{C_{11}} & \boldsymbol{C_{12}} & .. & \boldsymbol{C_{1p}} \\ \boldsymbol{C_{21}} & \boldsymbol{C_{22}} & .. & \boldsymbol{C_{2p}} \\ .. & & & \\ \boldsymbol{C_{p1}} & \boldsymbol{C_{p2}} & .. & \boldsymbol{C_{pp}} \end{bmatrix} \tag{4.41}$$

where $\boldsymbol{C_{ii}}$ collects the covariance of the i-th component of the vector in different locations. In the univariate random field case

$$\Sigma = \boldsymbol{C_{11}} \tag{4.42}$$

As previously defined the $\boldsymbol{C_{ij}}$ terms collects the cross-covariance of the i-j components of the vector at different locations.
As we did in the univariate case, first we create the sample variogram, secondly we fit the sample variogram with a valid model and finally we retrieve the covariance structure of the field. To construct the Cross-covariance we use the Cross-variogram. Under the condition $E[Z_i(x+h) - Z_i(x)] = 0 \quad \text{for} \quad i = 1, ., p$ the cross-variogram, introduced by Matheron (1965), has the form

$$\gamma_{12}(\boldsymbol{h}) = E[(Z_1(\boldsymbol{u}+\boldsymbol{h}) - Z_1(\boldsymbol{u}))(Z_2(\boldsymbol{u}+\boldsymbol{h}) - Z_2(\boldsymbol{u}))] \tag{4.43}$$

and the relationship with the cross-covariance is

$$\gamma_{i,j}(h) = C_{i,j}(0) - \frac{1}{2}[C_{i,j}(h) + C_{i,j}(-h)] \tag{4.44}$$

4.6.1 Application for Correttive Term and $\sigma_{0,sr}$

We compute the sample cross-variogram and we fit the exponential model. For the corrective term, we use a basis of four principal components curves so we have a vector random field of four components. Instead, for $\sigma_{0,sr}$, we use five scores:

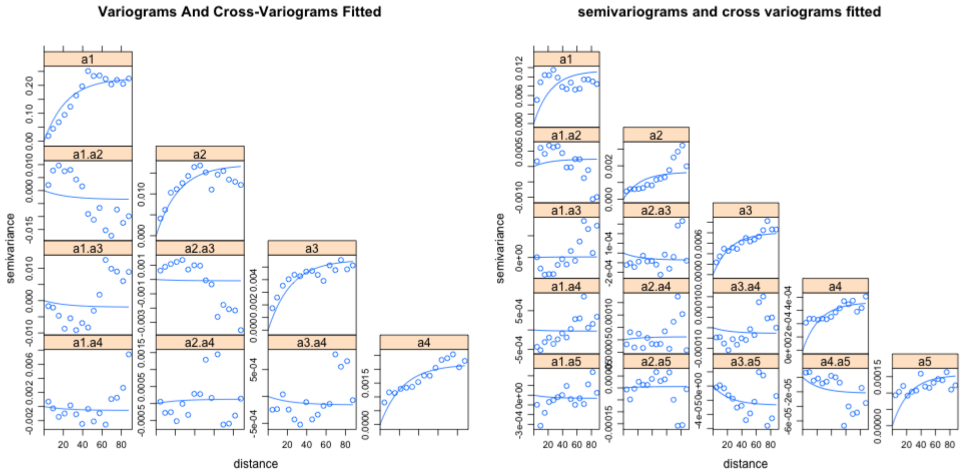

Figure 4.31: Corrective term(on the left) and $\sigma_{0,sr}$ (on the right) variograms and cross-variograms fitted with the exponential model.

From the variogram we estimate the cross-covariance and performs conditional simulation of the vectors on the grid of 70000 points.

From the simulated scores, we can reconstruct the curves in all the 70000 locations and from each locations we can find from the curves the value of the PGA and the corrective term/$\sigma_{0,sr}$ valutated at $T = 4s$. Both for $\sigma 0, sr$ and the corrective term, we compare the simulations for PGA and $T = 4s$ on the whole grid and on the faults.

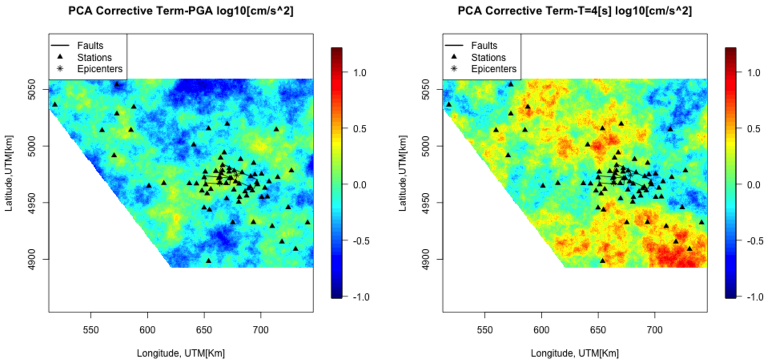

Figure 4.32: Sequential simulation of corrective term PGA(on the left) and $T = 4s$ (on the right).

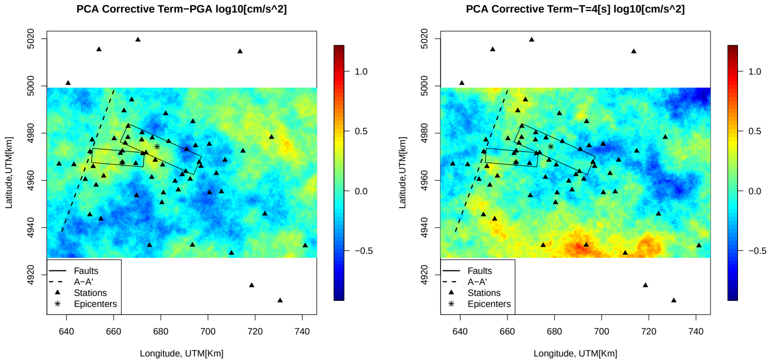

Figure 4.33: Sequential simulation of corrective term PGA (on the left) and $T = 4s$ (on the right), zoom on the faults.

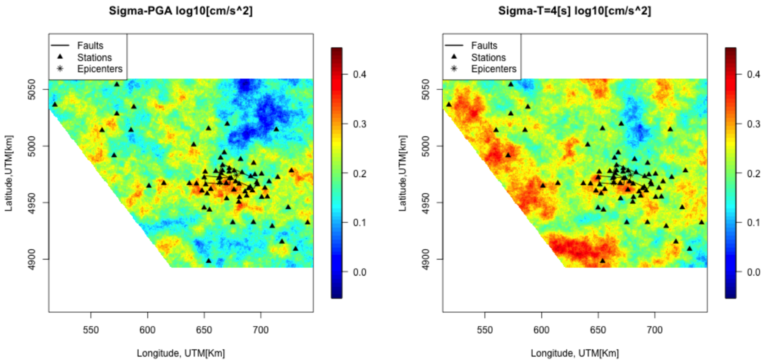

Figure 4.34: Sequential simulation of $\sigma_{0,sr}$ PGA (on the left) and $T = 4s$ (on the right).

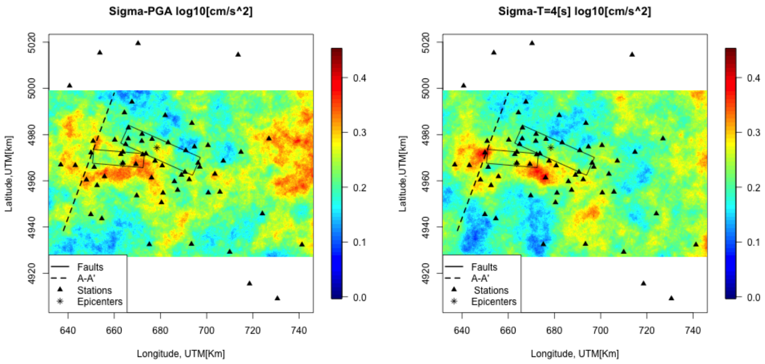

Figure 4.35: Sequential simulation of $\sigma_{0,sr}$ PGA (on the left) and $T = 4s$ (on the right), zoom on the faults.

Chapter 5

Model validation and testing

5.1 Model Comparison

In the previous chapter we have computed the kriging prediction for both the corrective term and $\sigma_{0,sr}$ at PGA and SA(T=4s) through the univariate approach and the functional one.
Now we want to compare the two models through the leave one out cross validation.
We report the performances of the two models adopting as metric the mean square error (MSE) and we take into account also its variance (VAR).

$$MSE = \frac{1}{N} \sum_{s \in S} (Z_s - \widehat{Z_s})^2 \tag{5.1}$$

$$VAR = \frac{1}{N-1} \sum_{s \in S} ((Z_s - \widehat{Z_s})^2 - MSE)^2 \tag{5.2}$$

Where S is the set of the stations.
The performances of the univariate approach in terms of MSE are reported in Table 5.1. and its variance in Table 5.2

	$\sigma_{0,sr}$	corrective term
PGA	0.0027	0.0269
$T = 4[s]$	0.0023	0.1496

Table 5.1: Univariate geostatistics performances, MSE.

	$\sigma_{0,sr}$	corrective term
PGA	$9.9 * 10^{-6}$	0.00144
$T = 4[s]$	$1.99 * 10^{-5}$	0.628

Table 5.2: Univariate geostatistics performances, VAR

The performances of the functional approach in terms of MSE are reported in Table 5.3 and its variance in Table 5.4.

	$\sigma_{0,sr}$	corrective term
PGA	0.0025	0.05287
$T = 4[s]$	0.0026	0.0281

Table 5.3: Functional geostatisctics performances, MSE.

	$\sigma_{0,sr}$	corrective term
PGA	$1.5 * 10^{-5}$	0.005
$T = 4[s]$	$2.3 * 10^{-5}$	0.0029

Table 5.4: Functional geostatistics performances, VAR.

We can observe that the two methods have comparable MSE when they are applied to $\sigma_{0,sr}$. For the corrective term, the MSE of the univariate approach for PGA is slightly better than the MSE of the functional approach. Conversely the functional approach shows a better MSE in case $T = 4s$ in which the univariate random field of the corrective term is assumed to have drift non-constant in space. Nevertheless, we have to consider that only the functional approach provides the prediction and the stocastic simulation of the whole spectrum and not only for this specific intensity measures (PGA and $T = 4s$).

5.2 Test

A leave-one-out cross-validation analysis highlights that modeling the spatial covariance of the corrective term for PGA with the univariate and the functional statistics provides very similar results while the best model in the case of T=4s is the functional one. We can compute the shaking fields based on the GMPE prediction for the first main event of the Emilia sequence (2012-05-20, $M_W = 6.1$).
For fifty five stations we have the values of PGA and $SA(T = 4)$. By computing the MSE between the intensity measures observed in the aforementioned stations and the closest point of a grid, defined as a collection of 70000 points where our GMPE has been evaluated, we test the prediction performance of models obtained with:

- functional kriging based on the simple cokriging of the FPCA scores (see Chapter 4) (evaluated in the PGA and $T = 4s$).
- ordinary kriging for the univariate case PGA (see chapter 3).

In Table 5.5 we report the MSE values for the GMPE and for a model obtained by adding the GMPE to the kriging prediction of the PGA computed with an univariate approach. MSE_2 in table 5.5 is computed taking into account all the stations available, the MSE_1 doesn't consider the station closest to the fault.

	GMPE[cm/s^2]	Model [cm/s^2]
MSE_1	59.82	56.93
MSE_2	171.24	661.51

Table 5.5: MSE for the model (univariate) and GMPE at PGA.

The station closest to the fault is dramatically overestimated by the two models and it significantly changes the value of the MSE. In the Table 5.6 we report the MSE for a predictive model of PGA, obtained by adding the GMPE to the kriging prediction of the PGA computed with the functional geostatistics approach. The model has a lower MSE in both the cases. We still observe that if we include the station closest to the fault, the MSE dramatically increases.

	GMPE[cm/s^2]	Model[cm/s^2]
MSE_1	59.82	56.32
MSE_2	171.24	101.56

Table 5.6: MSE for the model (functional) and GMPE at PGA.

In table 5.7 we compare the GMPE predictions of SA(T=4s) and the predictions of a model obtained by adding the GMPE to the kriging prediction of the SA(T=4s) computed with the functional geostatistics approach,. The MSE_1 In table 5.7 is obtained by taking into account all the stations in the region and we observe that the GMPE for SA(T=4s) has better performance. The station close the fault is well estimated by both the GMPE and the model, therefore we don't need to report MSE_2.

	GMPE[cm/s^2]	Model[cm/s^2]
MSE_1	16.22	21.56

Table 5.7: MSE for the model(functional) and for the GMPE at $T = 4s$.

Finally, we report the simulated shaking fields, for the first main event of the Emilia sequence computed as the sum of the GMPE, the corrective term simulated through the functional approach and a realization of a gaussian random field uncorrelated in space with zero mean and variance $\sigma_{0,sr}$. Also the variance $\sigma_{0,sr}$ has been simulated through the functional approach. We report in Figure 5.1 the simulated shaking fields for PGA and for SA(T=4s) and in Figure 5.2, 5.3 on the left panel, the GMPE and, on the right panel, the simulation of the shaking fields, with the colour scale saturated in the upper part, in order to compare it to the GMPE.

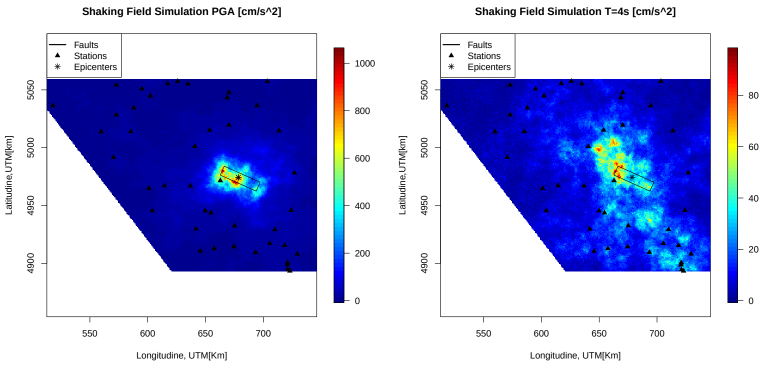

Figure 5.1: Shaking fields for PGA and SA(T=4s).

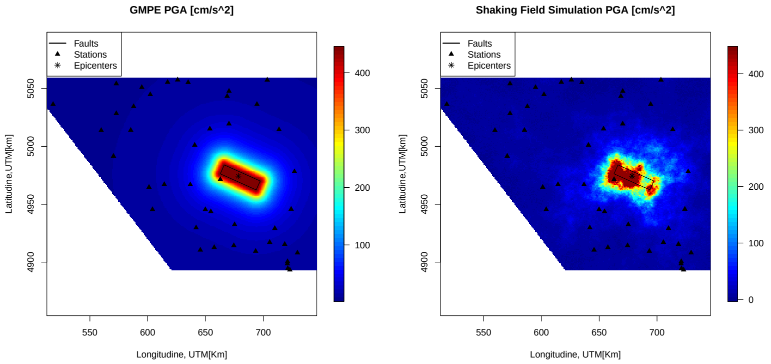

Figure 5.2: GMPE and shaking field for PGA, same scale.

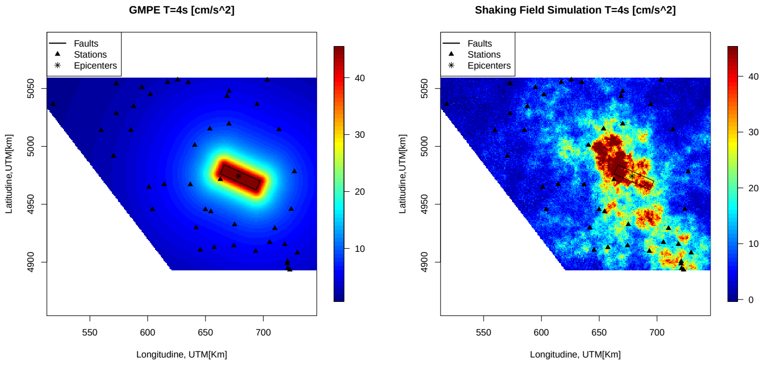

Figure 5.3: GMPE and shaking field for SA(T=4s), same scale.

We can observe that the simulation of the shaking fields presents greater values than the ones of the GMPE in the fault zone, where the epicenter of the event is located and the GMPE assigns a constant value to all the points. For SA(T=4s) we can observe a strong amplification in the area of the Appenninies that could be consequence of complex site effects due to the presence of surface waves.

Chapter 6

Conclusion

The corrective term and $\sigma_{0,sr}$ for PGA and $SA(T = 4s)$ have been investigated in order to study the spatial correlation and to predict their value in new locations.

First, through the univariate approach, it has been observed that the corrective term of PGA becomes uncorrelated faster than the one of $SA(T = 4s)$. The statistical results have been evaluated also considering the features of ground motion characterization specific for the area of study. S waves, which are generated by the trapping and conversion of the body waves in the thick sedimentary cover, dominate the seismic signals at periods longer than two seconds and determine the increase of correlation distance at long periods.

Since SA is function of the period of oscillation T, it has been possible to represent our data as a curve and to study their spatial covariance taking into account the inter-period correlation through FDA methods. We have approximated data through smoothing spline and we have recreated the curves for the corrective term of the GMPE and $\sigma_{0,sr}$ through the prediction and the simulation of the scores calculated by the Functional Principal Components Analysis. In this way, we have reproduced possible shaking scenarios for the corrective term and $\sigma_{0,sr}$ of PGA and $SA(T = 4s)$. The prediction performance of the univariate and the functional approaches have been compared by means of cross-validation.

Cross validation indicates that the univariate and the functional approach have similar results for $\sigma_{0,sr}$ that don't depend on the period. For what concerns the corrective terms, the univariate approach presents slightly better results at low periods while the functional approach presents much better results for SA(T=4s). This difference in performance could be connected to the drift estimation problem of the corrective term when we apply the univariate approach at T=4s. In conclusion, considering both the results and the availability of joint stochastic simulation of several periods, the functional approach is the best one.

Bibliography

Abrahamson, N., Youngs, R.R. (1992). A stable algorithm for regression analyses using the random effects model. Bulletin of the Seismological Society of America, 82:505-510.

Al-Atik, L., N.A. Abrahamson, J.J. Bommer, F. Scherbaum, F. Cotton and N. Kuehn (2010). The variability of ground-motion prediction models and its components. Seismological Research Letters, 81(5):794–801.

Anderson, J.G.and Brune J.N. (1999). Probabilistic seismic hazard assessment without the ergodic assumption. Seismological Research Letters, 70:19–28.

Bigi, G., Bonardi, G., Catalano, R., Cosentino, D., Lentini, F., Parotto, M., Sartori, R., Scandone, P., Turco, E. (1992). Structural model of Italy 1:500,000, CNR Progetto Finalizzato Geodinamica.

Bivand, R.S., Pebesma, E., Gòmez-Rubio, V. (2013). Applied Spatial Data Analysis with R.

Boccaletti, M., Bonini, M., Corti, G., Gasperini, P., Martelli, L., Piccardi, L., Severi, P., Vannucci, G. (2004). Seismotectonic map of the Emilia Romagna Region, Emial-Romagna Region - SGSS and CNR-IGC, S.EL.CA Florence

Boore, D.M, JoYner, W.B, Fumal, T.E. (1993). Estimation of response spectra and peak accelerations from Western North American earthquakes: an interim report.

Bradley, B.A. (2014). Site-specific and spatially-distributed ground-motion intensity estimation in the 2010–2011 Canterbury earthquakes. Soil Dynamics and Earthquake Engineering, 61-62, 83-91.

Bragato, P.L., Vuan, A., Massa, M., Saraò, A. (2011). Moho Reflection Effects in the Po Plain (Northern Italy) Observed from Instrumental and Intensity Data. Bulletin of the Seismological Society of America, 5:2142-2152.

Chilès, J-P. and Delfiner, P. (1999). Geostatistics: Modeling Spatial Uncertainty.

Cressie, N. (1993). Statistics for Spatial Data.

Delicado, P., Giraldo, R., Comas, C., Mateu, J. (2010). Statistics for spatial functional data. Environmetrics, 21(3-4):224-239.

Douglas, J. (2003). Earthquake ground motion estimation using strong-motion records: a review of equations for the estimation of peak ground acceleration and response spectral ordinates. Earth-Science Reviews, 61:43-104.

Goda, K. and Atkinson, G. M. (2010). Intraevent Spatial Correlation of Ground-Motion Parameters Using SK-net Data. Bulletin of the Seismological Society of America, 100(1):3055–3067.

Goda, K. and Hong, H. P. (2008). Spatial Correlation of Peak Ground Motions and Response Spectra. Bulletin of the Seismological Society of America, 98(1):354–365.

Jayaram, N. and Baker, J.W. (2010). Considering Spatial Correlation in Mixed-Effects Regression and the Impact on Ground-Motion Models. Bulletin of the Seismological Society of America, 100:3295–3303.

Lanzano, G., D'Amico, M., Felicetta, C., Puglia, R., Luzi, L., Pacor, F., Bindi, D. (2016). Ground motion prediction equations for region specific probabilistic seismic hazard analysis. Bulletin of the Seismological Society of America, 106(1):73–92.

Lanzano, G., Pacor, F., Luzi, L., D'Amico, M., Puglia, R., Felicetta, C. (2017). Systematic source, path and site effects on ground motion variability: the case study of Northern Italy. Bulletin of Earthquake Engineering, 15(11):4563–4583.

Lin, P.S., Chiou, B., Abrahamson, N., Walling, M., Lee, C.T., Cheng, C.T. (2011). Repeatable source, site, and path effects on the standard deviation for empirical ground-motion prediction models. Bulletin of the Seismolological Society of America, 101:2281–2295.

Loth, C. and Baker, J.W. (2012). A spatial cross-correlation model for ground motion spectral accelerations at multiple periods. Earthquake Engineering and Structural Dynamics, 42(3):397–417.

Luzi, L., Pacor, F., Ameri, G., Puglia, R., Burrato, P., Massa, M., Augliera, P., Franceschina, G., Lovati, S., Castro, R. (2013). Overview on the strong-motion data recorded during the May-June 2012 Emilia Seismic Sequence. Seismological Research Letters, 84(4):629-644.

Luzi, L., Puglia, R., Russo, E., D'Amico, M., Felicetta, C., Pacor, F., Lanzano, G., Ceken, U., Clinton, J., Costa, G. et al. (2016). The Engineering strong-motion database: a platform to access Pan-European accelerometric data. Seismological Research Letters, 87(4):987–997.

Matheron, G. (1965). Principles of geostatistics. Econ. Geol. 58:1246-1266.

Meletti, C., Galadini, F., Valensise, G., Stucchi, M., Basili, R., Barba, S., Vannucci, G., Boschi, E., (2008). A seismic source zone model for the seismic hazard assessment of the Italian territory. Tectonophysics, 450(1):85–108.

Menafoglio, A., Guadagnini, A., Secchi, P. (2016). Stochastic Simulation of Soil Particle-Size Curves in Heterogeneous Aquifer Systems through a Bayes space approach. Water Resources Research, 52:5708-5726.

Menafoglio, A., Secchi, P. (2017). Statistical analysis of complex and spatially dependent data: a review of Object Oriented Spatial Statistics. European Journal of Operational Research, 258(2):401-410.

Nerini, D., Monestiez, P., Manté, C. (2016). Cokriking for spatial functional data. Journal of Multivariate Analysis, 101(2):409-418.

Paolucci, R., Mazzieri, I., Smerzini, C. (2015). Anatomy of strong ground motion: near-source records and 3D physics-based numerical simulations of the Mw 6.0 May 29 2012 Po Plain earthquake, Italy, 203:2001–2020.

Pacor, F., Paolucci, R., Luzi, L., Sabetta, F., Spinelli, A., Gorini, A., Nicoletti, M., Marcucci, S., Filippi, L., Dolce, M. (2011). Overview of the Italian strong motion database ITACA 1.0. Bulletin of Earthquake Engineering, 9:1723–1739.

Park, J., Bazzurro, P. and Baker, J.W. (2007). Modeling spatial correlation of ground motion Intensity Measures for regional seismic hazard and portfolio loss estimation. Applications of Statistics and Probability in Civil Engineering, 1-8.

Ramsay, J.O. and Silverman, B.W. (2005). Functional data analysis.

Rodriguez-Marek, A., Cotton, F., Abrahamson, N., Akkar, S., Al Atik, L., Edwards, B., Montalva, G.A., Dawood, H.(2013). A model for single-station standard deviation using data from various tectonic regions. Bulletin of the Seismolological Society of America, 103:3149–3163.

Verros, S.A., Wald, D.J., Worden, C.B., Hearne, Ganesh, M. (2017). Computing spatial correlation of ground motion intensities for ShakeMap. Computers Geosciences, 99:145-154.

Villani, M. and Abrahamson, N. (2015). Repeatable site and path effects on the ground-motion sigma based on empirical data from Southern California and simulated waveforms from the CyberShake platform. Bulletin of the Seismological Society of America, 105(5):2681–2695.

Zerva A. and Zervas, V. (2002). Spatial variation of seismic ground motions: An overview. Applied Mechanics Reviews, 55(3):271-297.