UNIVERSITY OF STRATHCLYDE

Department of Civil and Environmental Engineering

Improving models of spatial correlation of earthquake ground motion to enable more informed seismic hazard and risk assessments

by

Erika Schiappapietra

Thesis submitted in fulfilment of the requirements for the degree of

Doctor of Philosophy

in

Civil and Environmental Engineering

January 2022

Glasgow, UK

Signed: Erika Schiappapietra

Date: 25/01/2022

Abstract

In an increasingly risky world, more sophisticated tools to quantify and manage seismic hazard and risk are required. In particular, the potential-loss and risk estimates of spatially-distributed infrastructures and portfolios of buildings have posed major challenges to hazard and risk analysts. Indeed, the quantification of the seismic performance of these systems requires estimates of the spatial correlation of earthquake ground motions across a region. The modelling of correlation has gained increasingly importance over the past decades, because it still represents a crucial step in the catastrophe modelling process, but there remain significant uncertainties.

This thesis attempts to advance the understanding of spatial correlation by critically investigating the factors and physical parameters that most affect the spatial variability of ground motions. A thorough literature review is provided along with analyses on large databases of recorded strong ground motion from previous earthquakes and ground-motion simulations. General outcomes suggest that spatial correlation properties are period-, regionally- and earthquake-dependent, so that a single stationarity and isotropic spatial correlation model, calibrated on heterogeneous databases including many regions and events, may not be suitable to describe the spatial variability of ground motions for the region of interest.

Key findings are used to develop customised correlation models to be included in deterministic (scenario-based) calculations to investigate to what extend spatial correlations may affect risk estimates. Besides these, probabilistic event-based results are also presented to further progress knowledge about the integration of spatial correlation and its associated uncertainties into catastrophe models. Such outcomes may constitute a starting point for the development of an upcoming earthquake catastrophe model for Italy.

Finally, the main results of this thesis can provide a primer for loss and risk modellers as well as researchers to understand, interpret and model spatial correlation for the generation of appropriate ground shaking maps and to improve hazard and risk assessments.

Acknowledgements

I would like to thank my supervisor, Dr John Douglas, for his support and technical advice throughout my PhD. I greatly benefited from his availability and punctuality, qualities that have undoubtedly made the progress of my research easier. He has always supported my research ideas and encouraged me to collaborate with international researchers, giving me the chance to grow as a researcher and pave the way for my future career.

I am also grateful to my second advisor, Dr Stella Pytharouli, for her smiling recommendations during these three years.

My PhD has benefited from the collaboration with Impact Forecasting, Aon's catastrophe model developer. My sincere thanks go to Dr Goran Trendafiloski for making this joint research project possible. A particular thank goes to Dr Svetlana Stripajová for our weekly online meetings, which have kept me motivated throughout the time of the pandemic. I would also like to thank Dr Peter Pazak for his precious help with translating our ideas into python scripts.

I am very grateful to Chiara Smerzini for her kind advice that went beyond engineering seismology matters. I am glad our paths have crossed again after so many years.

Thanks to Dr Carmine Galasso and Dr Alessandro Tarantino for carefully reviewing this thesis and their valuable suggestions.

Thanks to the group at INGV in Milan (Sara, Giovanni, Francesca and Maria) for helping me with the data and for the fruitful discussions and opportunities that have lasted since my Master's thesis.

Thanks to Professor Jack Baker's research group at Stanford University for providing helpful feedback and advice at early stages of this research.

I wish to thank the University of Strathclyde for the opportunity to undertake this experience and funding my Ph.D. I am also very grateful to SSA (Seismological Society of America) and the organisers of SECED 2019 for providing me with economic support to attend the 17WCEE and SECED2019 conferences, respectively.

A special thank to Maite for her long-distance friendship and empathy. I could not have imagined a better mate than you!

Thanks to Ema and Dani, for having given birth to Leonardo and Martino, my little nephews.

Thanks to Simone for being my family away from home.

And finally, special thanks to my mum, my grandparents and my family for encouraging me and supporting my decisions throughout these years. I could not have achieved any goal without you.

Contents

Abstract **i**

List of Figures **ix**

List of Tables **xxii**

1 Introduction **1**

1.1 Background and Motivations 1

1.2 Scope and Objectives 7

1.3 Outline of the Thesis 8

1.4 List of Publications 9

2 Modelling the spatial correlation of earthquake ground motion: Insights from the literature, data from the 2016-2017 Central Italy earthquake sequence and ground-motion simulations **11**

2.1 Introduction 12

2.2 Basic definitions 19

2.3 Databases 21

2.3.1 Central Italy earthquake sequence 21

2.3.2 Simulated ground-motions fields 22

2.4 Modelling of spatial correlation 25

2.4.1 Estimation of the within-event spatial correlation 25

2.4.2 Modelling caveats 28

2.5 Major factors influencing spatial correlation 29
2.5.1 Dependence on the estimation approach 29
2.5.2 Dependence on the fitting method 32
2.5.3 Dependence on magnitude 33
2.5.4 Dependence on period 34
2.5.5 Dependence on local site effects and geological structure 37
2.5.6 Dependence on GMPEs 40
2.6 Spatial correlation of different residual components 44
2.7 Spatial correlation of PGV, PGD and integral ground-motion intensity measures 48
2.8 Conclusions 49

3 Spatial correlation of broadband earthquake ground motion in Norcia (Central Italy) from physics-based simulations 53
3.1 Introduction 54
3.2 3D Physics-based ground motion simulations 56
3.2.1 3D Numerical modelling 57
3.2.2 ANN2BB: Broadband ground motions 58
3.2.3 Case-studies 59
3.3 Ground motion spatial correlation modelling 64
3.4 Results 68
3.4.1 Mw 6.5 Norcia earthquake simulation 68
3.4.1.1 Comparison with recordings 68
3.4.1.2 Ground motion directionality 69
3.4.1.3 Spatial correlation anisotropic properties 72
3.4.2 Dependence of spatial correlation on source effects 74
3.4.2.1 Earthquake magnitude 74
3.4.2.2 Slip distribution 76
3.4.2.3 Hypocentral locations 79
3.4.2.4 Finite fault Vs. point-source 83
3.5 Conclusions 86

4 Assessment of the uncertainty in spatial-correlation models for earthquake ground motion due to station layout and derivation method 90
4.1 Introduction 91

4.2 Spatial correlation modelling . . . 92
4.2.1 Spatial variability of within-event residuals . . . 93
4.2.2 Fitting methods for semivariogram models . . . 94
4.2.3 Maximum likelihood estimation . . . 95
4.3 Simulations set up . . . 96
4.4 The effect of bin size . . . 98
4.4.1 Least-squares regression Vs. Maximum-likelihood method . . . 101
4.5 Dependence on the number of stations . . . 101
4.5.1 Randomly simulated locations . . . 102
4.5.2 Station layouts of past earthquakes . . . 108
4.6 Discussion . . . 112
4.7 Conclusions . . . 122

5 Exploring the impact of spatial correlations in the catastrophe modelling process: a case study for Italy 125
5.1 Introduction . . . 126
5.2 Modelling spatial correlation of ground motion intensity measures . . . 128
5.2.1 Characterization of the spatial variability of univariate random functions 129
5.2.2 Stationarity and isotropy assumptions . . . 130
5.2.3 Databases . . . 131
5.2.4 Correlation models for the Mw 6.5 Norcia (Central Italy) event . . . 132
5.2.5 Correlation models for the three macro-regions in Italy . . . 136
5.3 OpenQuake engine input models . . . 141
5.3.1 Hazard component . . . 141
5.3.1.1 Scenario - based seismic hazard assessment . . . 141
5.3.1.2 Event-based probabilistic seismic hazard assessment . . . 143
5.3.2 Vulnerability component . . . 145
5.3.3 Exposure component . . . 146
5.4 Risk analysis and discussion . . . 148
5.4.1 Scenario-base loss assessment . . . 148
5.4.2 Event-based probabilistic loss assessment . . . 151
5.5 Conclusions . . . 157

6 Conclusions 160
6.1 Key findings . . . 161

6.2 Future research 164

Bibliography **167**

List of Figures

1.1 (a) Single-site seismic hazard assessment vs (b) multiple-site seismic hazard assessment. In (a) the hypothesis of independency of IMs at closely spaced sites is suitable for the estimation of the ground motion at a given site (e.g. where a building is located). By contrast, in (b) the description of the spatial variability of the ground motion is an essential element to quantify the IMs at various sites in the region. 3
1.2 Between-event and within-event residual components of a GMM. The black solid line represents a GMM, whereas the orange and green dashed lines indicate the event-corrected ground-motion medians for earthquake 1 (EQ1) and earthquake 2 (EQ2), respectively. Squares indicate the observed IMs for the two events. [Modified from Atik et al. (2010)] 4
1.3 (a) Histograms of aggregated economic loss for correlated and non-correlated ground-motion fields [modified from Wagener et al. (2016)]; (b) Ratio between disruption probabilities considering spatially-uncorrelated and spatially-correlated ground-motion fields [modified from Costa et al. (2018)]; (c) Aggregated loss curves for the masonry-wall, mid-rise, low-code type for the Firenze Administrative Province [modified from Weatherill et al. (2015)]. 6

2.1 A typical selection of spatial correlation models for PGA: B03: Boore et al. (2003); GH08: Goda and Hong (2008); GA09: Goda and Atkinson (2009); JB09: Jayaram and Baker (2009); GA10: Goda and Atkinson (2010)); EI11ESM: Esposito and Iervolino (2011) ESM database; EI11ITACA: Esposito and Iervolino (2011) ITACA database; EI12ITACA: Esposito and Iervolino (2012) ITACA database; S12: Sokolov et al. (2012); DW13: Du and Wang (2013); HM19: Heresi and Miranda (2019); HG19: Huang and Galasso (2019). 15

2.2 Epicentres (stars) of the nine $M_w \geq 5.0$ earthquakes of the sequence. Mw $\geq$ 2.5 events for one year after the Amatrice mainshock are also mapped with blue dots to better highlight the spatial extent of the sequence. [source: *http:// cnt.rm.ingv.it/*]. 22

2.3 (a) Records-distance distribution; (b) Magnitude-distance distribution; (c) Events-magnitude distribution; (d) Distribution of stations in terms of EC8 site classification. Distances are in terms of Joyner and Boore distance (R_{jb}, i.e. the distance to the surface projection of the rupture). 23

2.4 Example of median (deterministic) (a), stochastic (b) and random (c) ground-motion fields. obtained through the above-described approach. PGAs are expressed in cm/s^2. The triangles represent the stations used in the analysis. 24

2.5 Exponential model. The black solid line is the theoretical model, whereas black squares represent the experimental semivariogram. The sill and range parameters are also highlighted. 27

2.6 Correlation models proposed by Goda (2011) and Heresi and Miranda (2019) [HM19] and obtained considering data from the Central Italy sequence for PGA (a) and SA(3 s) (b). Solid lines represent the median values, whereas shadow areas indicate the event-to-event uncertainty. Black dashed line helps indicate the distances at which the correlation equals 0.05. 30

2.7 Estimated range h_0 normalized by the true value of the range $h * 0$ as a function of the number of sites (stations) using: (a) the Cressie and Hawkins (robust) estimator; the Matheron (classic) estimator; and (c) the covariance approach. Grey squares represent the mean value, whereas the solid vertical lines indicate the variability of the estimates. The red dashed line signifies a ratio of 1. 31

2.8 Comparison between the ranges obtained through both the classic and robust estimators, considering different numbers of stations. 32
2.9 Correlation models obtained through the robust (a) and classic (b) estimators. Blue curves indicate the results in which the weights used to fit the experimental data are computed based on the number of pairs and separation distance; red curves represent the results in which equal weights are applied to all the data. Shadow area represent the standard deviation of the results computed on 1000 simulations. The standard deviation of the model with weights (8.7 and 7.12 km for the robust and classic estimators respectively) is smaller than the standard deviation of the model without weights (9.02 and 7.73 km). 33
2.10 Range as a function of magnitude: (a) PGA; (b) SA(4s). Ranges are obtained by means of the robust semivariogram. We also apply a detrending processing using second-order surface models to remove long-range spatial trends. The dashed red line represents the bin size of 5 km. Any range value smaller than the bin size is indicative of non-correlation and should not be taken as an appropriate estimate. 35
2.11 Range a as function of magnitude Mw: a) PGV; b) PGD; c) CAV. Ranges are obtained by means of the robust semivariogram. The dashed red line represents the bin size of 5 km. Any range value smaller than the bin size is indicative of non-correlation and should not be taken as an appropriate estimate. 35
2.12 Range as a function of response-spectral period, T. The main studies are reported along with a model for the Central Italy seismic sequence, denoted as ‘This study ’. This model is calibrated by pooling all the data from the nine $M_w \geq 5.0$ events. The model of Jayaram and Baker (2009)) refers to their model for heterogeneous soil condition, whereas the model of Du and Wang (2013) is computed considering a V_{s30} correlation of 4.5 km. 36
2.13 Range as a function of response-spectral period, T, computed using the robust estimator for the nine $M_w \geq 5.0$ events of the sequence. In black we reported the weighted (geometric) average model, in which the weights are given based on the number of stations that recorded each event. . . . 37

2.14 (a) Range as a function of the response-spectral period, T, for within-event residuals ε_{ij} (dWes) with and without the site correction factor; (b) Correlation function in terms of PGA for within-event residuals ε_{ij} with and without the site correction factor. The black dashed line indicates a correlation of 0.05. 39
2.15 Correlation function in terms of PGA (a) and SA at T = 2 s (b). The model referred as 'This study'is computed by means of the code developed by Ming et al. (2019) using data from the nine $M_w \geq 5.0$ central Italy earthquakes. The red shadow area represents the 95% confidence intervals of the central Italy model. The black dashed line indicates a correlation of 0.05. 40
2.16 Within-event ε_{ij} (dWes) correlation models for uncorrected and corrected residuals and computed considering the GMPE by Bindi et al. (2011): (a) PGA; (b) spectral acceleration for T equals 2 s. 42
2.17 Spatial correlation values as a function of period obtained for the Central Italy region. Ad hoc refers to the GMPE calibrated on the dataset; ITA18 refers to the Lanzano et al. (2019) GMPE; K20 Erg refers to the ergodic Kotha et al. (2020) GMPE; K20 Reg refers to the region-specific Kotha et al. (2020) GMPE. 43
2.18 Sample semivariogram obtained using the GMPEs by Boore and Atkinson (2008*b*) and Bindi et al. (2011). 44
2.19 Residual plots to check if residuals show any systematic trend with predictor variables: a) η_j (δB_e) against Mw; b) $\delta S2S$ against V_{s30}; c) δWS_{ij} against R_{JB}. Green dots in b) highlight stations with a measured V_{s30} value, whereas black dots stations for which the V_{s30} is inferred from the slope, following Wald and Allen (2007). 46
2.20 Semivariograms of different residual components for PGV obtained pooling all the events considered in the analysis: a) Within-event residuals; b) Site-to-site term; c) Event- and site-corrected residuals. Black squares represent the original data, whereas red dots refer to the detrended data. 47
2.21 Range of within-event, site-to-site and event- and site-corrected term for different IMs: a) Spectral acceleration at period between 0.1 and 5 s; b) PGA, PGV, PGD, CAV and I_a. 47

2.22 Comparison among different correlation models: a) PGV; b) CAV (solid lines) and Ia (dashed lines). Black dashed line points out a level of correlation equal to 0.05. 50

3.1 Study area: (a) topography model; (b) shear-wave velocity model; (c) sediment thickness model. Yellow stars indicate the epicentre (Mw 6.5 Norcia event), whereas the black rectangles define the surface fault projection. 58

3.2 Slip distribution along the causative fault. (a) Pizzi et al. (2017); (b) HB94 model for the Mw 6.5 scenario; (c) HB94 model for the Mw 6.0 scenario; (d) HB94 model for the Mw 5.5 scenario. Yellow stars represent the hypocentral locations, whereas the largest rectangles indicate the maximum dimension of the fault ($36km \times 13km$). The smallest rectangles represent the ruptured faults, where the slip asperities (regions of very high slip) are concentred. 61

3.3 Slip distribution along the causative fault (Mw 6.5) along with the hypocentral locations of scenarios S002-S008. 62

3.4 Rupture time distribution on the causative fault for scenarios Mw6.5_S002-S008. The scenarios have the same slip distribution, but feature different hypocentral locations. The position of the hypocentre is pointed out by a yellow star. 63

3.5 Spectral acceleration at T = 2 s of the FP component: (a) Ground motion data and fitted GMMs; (b) residuals as a function of Rline; (c) QQ plot of residuals. Grey dots represent stations on the footwall for rock soil condition and the grey line indicates the corresponding GMM; Green dots represent stations on the hanging wall for rock soil condition and the green line indicates the corresponding GMM; Orange dots represent station on the hanging wall for basin soil condition and the orange line indicates the corresponding GMM. The figure shows the results obtained considering the Mw6.5_S001 scenario (30^{th} October 2016 Norcia event). 65

3.6 Experimental semivariogram and corresponding exponential and spherical models: (a) SA (T = 1.5 s) Z component; (b) SA (T = 4 s) Z component. Black dots represent the experimental semivariogram estimates for each distance bin, whereas dashed lines are the fitted models. The figure shows the results obtained considering the Mw6.5_S001 scenario (30th October 2016 Norcia event). In this example, we choose the exponential model for SA (T = 1.5 s) and the spherical model for SA (T = 4 s). 66

3.7 (a) directional semivariograms computed for different directions: 0 - NS, 45 - NE-SW, 71 - FN, 90 - EW, 135 - NW-SE, 161 - FP; (b) Variogram map. The black dashed arrow indicates the NS direction. The figure shows the results obtained considering the Mw6.5_S001 scenario (30^{th} October 2016 Norcia event). 67
3.8 Spatial variation of within-event residuals of SA (FP) at T = 4 s: (a) Raw data showing a gradual variation of the residual values; (b) detrended data featuring a patchier distribution. Yellow star indicates the epicentre (Mw 6.5 Norcia event), whereas the grey and black rectangles are the surface fault projection and the effective surface fault projection, respectively. The figure shows the results obtained considering the Mw6.5_S001 scenario (30_{th} October 2016 Norcia event). 68
3.9 (a) Experimental semivariograms obtained by using the ANN2BB approach (Predictions) and observed recordings for PGA; (b) Experimental semivariograms for SA (T = 2s); (c) Experimental semivariograms for SA (T = 5 s); (d) Spatial distribution of the seismic stations (epicentral distance < 30 km) that recorded the Norcia event. Dots are color-coded based on the V_{s30} values. The red star is the epicentre, whereas the black rectangle represents the projected fault surface. [Data are from *https://esm-db.eu/*]. 70
3.10 (a) Range as a function of the period T for different ground motion components; (b) sill as a function of the period T for different ground motion components.. 71
3.11 Ground motion maps for SA (T = 3s): (a) geometric mean component; (b) FN component; (c) vertical component; and (d) FP component. 72
3.12 Average range values, computed in the different directions, as a function of the period T. (a) Geometric mean component; (b) vertical component; and (c) FN over FP components ratio. The grey area represents the mean ± standard deviation. The FN-direction and FP-direction range values as a function of T are also reported on the graphs. 73
3.13 Same as in Figure 3.12 but in terms of sill values, computed in the different directions, as a function of the period T. 74

3.14 Directional semivariograms for SA (T = 3s) for the geometric mean component. Dots and squares represent the experimental semivariograms computed considering azimuths of 71° (FN) and 161° (FP), respectively. Solid lines represent the corresponding models. 74
3.15 Range as a function of the period T for different Mw scenarios: (a) Geometric mean component; (b) FN component; (c) vertical component; and (d) FP components. 76
3.16 Sill as a function of the period T for different Mw scenarios: (a) Geometric mean component; (b) FN component; (c) vertical component; and (d) FP components. 77
3.17 SA (T = 2s) Ground motion maps for Mw6.5_S001 (left panel) and Mw6.5_S002 (right panel) scenarios: (a) and (b) FN component; and (c) and (d) vertical component. 78
3.18 Ratio between Mw6.5_S001 and Mw6.5_S002: (a) range values; and (b) sill values. 79
3.19 Ratio of FN/FP range for the different scenarios with varying hypocentral location. 80
3.20 (a) Average and (c) standard deviation range values computed for scenarios Mw6.5_S001-S008 for the different azimuth directions; (b) Average and (d) standard deviation sill values computed for scenarios Mw6.5_S001-S008 for the different azimuth directions. 81
3.21 Ground motion maps of the GeoH spectral acceleration at 2 s [SA (T=2s)] for scenarios Mw6.5_S002-S008: (a) S003; (b) S004; (c) S005; (d) S006; (e) S007; (f) S008; (g) S002. The epicentral location is pointed out by a yellow star. The black rectangle represents the surface projection of the fault. . 82
3.22 a) Average and (c) standard deviation range values computed on scenario Mw6.5_S001-S008 for the different azimuth directions; (b) Average and (d) standard deviation sill values computed on scenario Mw6.5_S001-S008 for the different azimuth directions. Results are for the FN component. . 83
3.23 (a) Average and (c) standard deviation range values computed on scenario Mw6.5_S001-S008 for the different azimuth directions; (b) Average and (d) standard deviation sill values computed on scenario Mw6.5_S001-S008 for the different azimuth directions. Results are for the FP component. . 84

3.24 (a) Average and (c) standard deviation range values computed on scenario Mw6.5_S001-S008 for the different azimuth directions; (b) Average and (d) standard deviation sill values computed on scenario Mw6.5_S001-S008 for the different azimuth directions. Results are for the Z component. . . 85
3.25 Average range (a) and sill (b) values computed on the different azimuth directions for the M_w4.0 point-source scenario. The shadow areas represent the mean ± standard deviation. 86
3.26 Ground motion maps for Mw4.0_S001 scenario: (a) and (b) SA (T=1s) for the geoH and Z components, respectively; (c) and (d) SA (T= 4s) for geoH and Z components, respectively. The yellow star represents the epicentre. 87

4.1 Empirical and fitted semivariogram models: (a) different techniques to estimate the semivariogram parameters as introduced in this work. The solid line is the exponential fitted model, whereas squares represent the experimental semivariogram. The numbers close to the squares indicate the number of pairs used to compute the semivariances within each bin. (b) Different bin sizes to compute the experimental semivariogram. The exponential models are fitted by using the OLS approach. 95
4.2 Examples of Gaussian random fields characterized by a correlation length of 5 km (top row) and 15 km (bottom row). 97
4.3 Number of stations pairs as a function of the separation distance for different grid sizes: (a) 100 km; (b) 150 km; (c) 200 km. The black solid line is the mean value computed over the 1000 simulations, whereas the black dashed lines represent the mean ± the standard deviation. The red dashed line indicate the minimum number (30) of station pairs required per bin to obtain more robust estimates (see Schiappapietra and Douglas (2020) for further details). 99
4.4 Range as a function of the grid dimension. Solid lines represent the median values computed over the 1000 simulations, whereas dashed lines indicate the first and third quartile. OLS and REML refer to the two different approaches we use to estimate the range. The black dotted line indicates the initial range value (30 km) imposed in the simulation. 99
4.5 Bias median (a) and interquartile IQR (b) values as a function of the number of stations for different bin widths and for two different approaches (OLS, WLS). We impose an initial range value of 5 km. 100

4.6 Bias median (a) and interquartile IQR (b) values as a function of the number of stations for different bin widths and for two different approaches (OLS, WLS). We impose an initial range value of 15 km. Note that the y axis scale is different to that used in Figure 4.5. 100
4.7 Boxplots for range estimates obtained by using different bin size, different h_0 (5 and 15 km) and different number of available stations (40 and 100). 102
4.8 Bias (left panel) and Interquartile (right panel) range values as a function of the number of stations. The different rows refer to the different initial h_0 values. Different colours refer to the approaches employed in the correlation estimation. 104
4.9 Bias (left panel) and Interquartile (right panel) range values as a function of the number of stations. The different rows refer to the different initial h_0 values. Different colours refer to the approaches employed in the correlation estimation. 105
4.10 Bias (left panel) and Interquartile (right panel) range values as a function of h0. The different rows refer to the different number of stations considered. Different colours refer to the approaches employed in the correlation estimation. 106
4.11 Bias (left panel) and Interquartile (right panel) range values as a function of the number of stations. The different rows refer to the different initial h_0 values. Different colours refer to the approaches employed in the correlation estimation. 107
4.12 Boxplots of range estimates for different number of stations and h0 = 20 km: (a) 20 stations; (b) 40 stations; (c) 80 stations: (d) 100 stations . . 108
4.13 Station layouts for the four selected events: (a) ESM1 - event id IT-2012-0011; (b) ESM2 - event id ESMC-20160903_0000063; (c) NGA1 - event id 279; (d) NGA2 - event id 1011. Dots are colour-coded based on the within-event residuals (one out 1000 simulated ground-motion fields). Coordinates are normalized with respect of the epicentre of each event. 110
4.14 Median % error of the range estimates for the four different stations layouts as a function of the number of stations. (a) OLS; (b) WLS; (c) ML; (d) REML. The initial value of the range is set to 10 km. 111

4.15 IQR of the range estimates for the four different station layouts as a function of the number of stations. (a) OLS; (b) WLS; (c) ML; (d) REML. The initial value of the range is set to 10 km. 112
4.16 Median % error of the range estimates for the four different stations layouts as a function of the number of stations.. The initial value of the range is set to 20 km. We report the results obtained by using all the approaches proposed in this study (top: OLS and WLS; bottom: ML and REML). . 113
4.17 Median % error of the range estimates for the four different stations layouts as a function of the number of stations.. The initial value of the range is set to 30 km. We report the results obtained by using all the approaches proposed in this study (top: OLS and WLS; bottom: ML and REML). . 114
4.18 Median % error of the range estimates for the four different stations layouts as a function of the number of stations.. The initial value of the range is set to 40 km. We report the results obtained by using all the approaches proposed in this study (top: OLS and WLS; bottom: ML and REML). . 115
4.19 IQR of the range estimates for the four different stations layouts as a function of the number of stations. The initial value of the range is set to 20 km. We report the results obtained by using all the approaches proposed in this study (top: OLS and WLS; bottom: ML and REML) 116
4.20 IQR of the range estimates for the four different stations layouts as a function of the number of stations. The initial value of the range is set to 30 km. We report the results obtained by using all the approaches proposed in this study (top: OLS and WLS; bottom: ML and REML) 117
4.21 IQR of the range estimates for the four different stations layouts as a function of the number of stations. The initial value of the range is set to 40 km. We report the results obtained by using all the approaches proposed in this study (top: OLS and WLS; bottom: ML and REML) 119
4.22 Number of stations pairs within 12 km as a function of the number of stations for the four station layouts. The dots indicate the number of stations pairs in each simulation, whereas the solid lines represent the average number of pairs as a function of the number of stations. 120
4.23 Experimental semivariograms and theoretical models for three different events recorded by 40 stations. The number of pairs in each bin is reported close to each semivariance estimate (black dots). 121

4.24 Experimental semivariograms and theoretical models for three different events recorded by 80 stations. The number of pairs in each bin is reported close to each semivariance estimate (black dots). 121
4.25 Application of the proposed logic tree for h_0 = 20 km and number of stations equal to 60 to use within a seismic risk assessment study. We use the estimates obtained by using the REML approach. The graphs on the left-hand side correspond to n spatially correlated random fields simulated for each range value, 7 km, 20 km and 27 km, respectively. 124

5.1 p-value as a function of magnitude Mw: (a) isotropy test; and (b) stationarity test. Red dashed line indicates the 5% significance level. The results are for the Central Italy dataset and for IM = SA(T=1 s). 131
5.2 Set of Italian data: (a) Location of the epicentres as a function of Mw; and (b) magnitude-distance scatter plot. JB stands for Joyner-Boore distance. The rectangles (in dashed black) indicate the three sub-regions selected in this study: NI: Northern Italy; CI: Central Italy; and SI: Southern Italy. . 134
5.3 Spatial correlation models obtained for the Mw 6.5 Norcia (Central Italy) event. Dots and squares represent the ranges as a function of period, whereas the solid lines represent the fitted models. *Ad hoc* refers to the GMM calibrated on the Mw 6.5 Norcia observations only; ITA18 refers to the Lanzano, Luzi, Pacor, Felicetta, Puglia, Sgobba and D'Amico (2019) GMM; K20 refers to the Kotha et al. (2020) GMM. 135
5.4 (Left hand side) Computed weighted geometric mean and fitted model, and (right hand side) computed weighted standard deviation and fitted model. (Top) Northern Italy, (Middle) Central Italy, and (Bottom) Southern Italy. 138
5.5 Comparison among different correlation models: NI - Northern Italy; CI - Central Italy; SI - Southern Italy; EI12 - ITACA - Esposito and Iervolino (2012); EI12 - ESD - Esposito and Iervolino (2012); HM19 - Heresi and Miranda (2019); HG19 - Huang and Galasso (2019); S19 - Sgobba et al. (2019). 140
5.6 Spatially-correlated ground-motion maps obtained by using different correlation models: (a) No within-event spatial correlation; (b) *Ad hoc* spatial correlation model; (c) ITA18 spatial correlation model; and (d) K20 spatial correlation model. The yellow star represents the Mw 6.5 Norcia epicentre, whereas the black rectangle defines the surface fault projection. 143

5.7 (a) Hazard curves showing the return period of different ground motion levels in terms of PGA for three different cases: ground-motion fields are generated without considering correlation; ground-motion fields are generated considering correlation; ground-motion fields are generated considering correlation and its associated uncertainty. (b) ratio of ground motion levels as a function of the return period. 145
5.8 (a) Classification of the building classes according to the European exposure model. For further details on the labels, the reader should refer to Crowley et al. (2021); (b) Area of interest: residential, industrial and commercial buildings of graphs (a) and (c) are within the red rectangle; (c) Percentage of replacement costs, divided into contents and structural and non-structural components. 147
5.9 Losses (%TV computed based on different hazard models expressed in terms of percentage of total value of the portfolio: (a) Spatial correlation is not considered; (b) ad hoc correlation model; (c) ITA18 correlation model; and (d) K20 correlation model. The yellow star represents the Mw 6.5 Norcia epicentre, whereas the black rectangle defines the surface fault projection. 149
5.10 Histograms of economic losses computed based on different hazard models: (a) Spatial correlation is not considered; (b) ad hoc correlation model; (c) ITA18 correlation model; and (d) K20 correlation model. Red lines point out the mean value, whereas the red dashed lines indicate the mean ± std. 151
5.11 EP curves in terms of loss in %TV for three different cases: ground-motion fields are generated without considering correlation; ground-motion fields are generated considering correlation; ground-motion fields are generated considering correlation and its associated uncertainty. TV stands for total value of the economic portfolio (2020 SERA exposure model). In the inset, we show EP curves up to a return period of 1000 years, which is usually the range of interest for (re)insurance industry. 153

5.12 Loss ratio as a function of the return period for three different cases: ground-motion fields are generated without considering correlation; ground-motion fields are generated considering correlation; ground-motion fields are generated considering correlation and its associated uncertainty. In the inset, we show loss ratio curves up to a return period of 1000 years, which is usually the range of interest for (re)insurance industry. 154
5.13 Average annual losses (AAL) for the flat homogenous portfolio (PGA) obtained for three different cases: (a) ground-motion fields are generated without considering correlation; (b) ground-motion fields are generated considering correlation; and (c) ground-motion fields are generated considering correlation and its associated uncertainty. The yellow star represents the Mw 6.5 Norcia epicentre, whereas the black rectangle defines the surface fault projection. 156
5.14 EP curves in terms of loss in %TV for three different cases: ground-motion fields are generated without considering correlation; ground-motion fields are generated considering correlation; and ground-motion fields are generated considering correlation and its associated uncertainty. Losses are computed for flat homogeneous portfolios: (a) PGA; (b) SA at T = 0.3 s; (c) SA at T = 0.6 s; (d) SA at T = 1 s. 157

List of Tables

2.1 Main published correlation models. PGA is peak ground acceleration, PGV is peak ground velocity, SA is response spectral acceleration, Ia is Arias intensity, CAV is cumulative absolute velocity and T is spectral period. Mw is moment magnitude and RJB is Joyner-Boore distance. GMPE is ground motion prediction equation. 16

2.2 Main characteristics of the largest magnitude events of the Central Italy sequence. [source: Lanzano, Sgobba, Luzi, Puglia, Pacor, Felicetta, D'Amico, Cotton and Bindi (2019)] 21

3.1 Summary of the source parameters of the different simulated scenarios. For the Norcia event (first row), the source is taken from the source inversion model proposed by Pizzi et al. (2017), whereas HB94 refers to the kinematic source rupture generator of Herrero and Bernard (1994). 59

4.1 5%, 50% and 95% percentile values of the range estimates for different h0 and different number of stations for the approaches used in this study. Estimates are in km. 117

5.1 Coefficients of the equations fitted to the ranges for the three GMMs (equation 5.8 and 5.9). 135

5.2 Coefficients of the equations fitted to the ranges (equation 5.10, 5.11 and 5.12) for the three different regions. 139

5.3 Overall loss in %TV (percentage of total value of the portfolio) for the four different case studies. 149
5.4 Mean and Coefficient of variation of the losses obtained from the different correlation models for two different IMs. 150

Acronyms and Symbols

ANN	Artificial Neural Network
a	sill of the semivariogram
b	range of the semivariogram
CAV	Cumulative Absolute Velocity
CDN	pre seismic Design Code
CDL	Low seismic Design Code level
CDM	Moderate seismic Design Code level
CDH	High seismic Design Code level
COV	covariance
CR	Reinforced Concrete buildings
CV	Coefficient of Variation
δB (η)	Between-event (inter-event) residual component
δW (ε)	Within-event (intra-event) residual component
δWS	event-and-site corrected residuals
$\delta S2S$	site-specific correction term
FP	Fault Parallel
FN	Fault Normal
$\hat{\gamma}$	empirical semivariogram
GeoH	Geometric Mean of the two horizontal ground motion components
GMMs (GMPEs)	Ground Motion Models (Ground Motion Prediction Equations)
h	separation distance
h_0	imposed initial range

I_a	Arias Intensity
IMs	Intensity Measures
IQR	Interquartile range
MCF	Confined Masonry buildings
ML	Maximum Likelihood
M_w	Moment Magnitude
MUR	Unreinforced Masonry buildings
OLS	Ordinary Least Squares
PGA	Peak Ground Acceleration
PGV	Peak Ground Velocity
PGD	Peak Ground Displacement
PBS	Physics-Based ground motion Simulation
PSHA	Probabilistic Seismic Hazard Assessment
Q_1	first quartile
Q_3	third quartile
REML	Restricted Maximum Likelihood
ρ_ε	within-event spatial correlation
ρ_η	between-event correlation
ρ_T	total spatial correlation
R_{line}	R_{line} distance
R_{JB}	Joyner and Boore distance
S	Steel buildings
SA	Spectral Acceleration
SD	Spectral Displacement
ϕ (σ_e)	standard deviation of the within-event residuals
σ_{TOT}	total standard deviation
τ	standard deviation of the between-event residuals
$V_{s,30}$	Average shear-wave velocity of the upper-most 30 m
WLS	Weighted Least Squares
Z	Vertical ground motion component

CHAPTER 1

Introduction

1.1 Background and Motivations

Earthquakes are among the hardest natural hazards to predict and model. They regularly produce enormous losses and fatalities around the world and they still show the largest protection gap across all natural perils. According to the statistics of the Centre for Research on the Epidemiology of Disasters (CRED, *https://www.cred.be*), there have been nearly 270,000 fatalities from earthquakes over the last decade. For instance, the 2010 Haiti event is thought to have killed about 222,500 people, making it one of the deadliest earthquakes in recorded history. The 2011 Tohoku earthquake and tsunami caused considerable economic losses, which amounted to approximately 220 Billion USD according to the Japanese Cabinet Office[1].

In an increasingly risky world, many private and public stakeholders, such as governments, search-and-rescue organizations and insurance/reinsurance companies, require more sophisticated tools to quantify and manage seismic hazard and risk. A reliable and accurate evaluation of earthquakes risks is needed to: (1) gauge the socio-economic impact of future earthquakes due, not only to direct structural damage, but also to non-structural damage, content losses and business interruption; (2) enable firms to analyse the financial implications of catastrophic events and to achieve a greater understanding of the risks their businesses have to face; (3) manage emergency response and improve disaster management;

[1]UNESCO/IOC (2012). Summary Statement from the Japan - UNESCO - UNU Symposium on The Great East Japan Tsunami on 11 March 2011 and Tsunami Warning Systems: Policy Perspectives 16 - 17 February 2012

and (4) provide decision makers with effective tools for risk mitigation and retrofitting strategies.

The assessment of seismic risk and loss estimation is generally based on three pivotal elements: (1) seismic hazard models, namely region-specific estimates of earthquake ground shaking and its spatial variability; (2) exposure datasets; and (3) vulnerability and/or fragility models, which provide the probability of damage and loss conditional on given ground motion intensity measures (IMs) (Grossi and Kunreuther, 2005; Erdik, 2017). Within the framework of seismic hazard modelling, three main issues have to be considered: locations of potential future events, frequency of occurrence and severity. Once source parameters are identified, local intensities are estimated at each location over the affected area (Grossi and Kunreuther, 2005). The deterministic and probabilistic assessment of IMs at individual sites is now routine. The ground motion due to an earthquake is usually quantified by empirical ground motion models (GMMs, also known as ground motion prediction equations, GMPEs), which provide an estimate of the ground shaking and its associated aleatory variability based on the hypothesis of independency amongst IMs at all sites (including those close-by). GMMs model the IM of interest as a lognormally-distributed random variable that is a function of explanatory predictors (e.g. magnitude, source-to-site distance and site conditions), and provide only their marginal probability distribution (Douglas and Edwards, 2016). The seismic risk and loss assessment of spatially distributed building portfolios and infrastructure systems (Figure 1.1) requires, however, the quantification of the joint probability of occurrence of IMs at multiple locations over the region of interest, which implies the need to model the spatial cross-correlation of ground motions (Weatherill et al., 2015; Costa et al., 2018; Crowley et al., 2008; Wesson and Perkins, 2001; Markhvida et al., 2018; Sokolov and Wenzel, 2019).

The spatial correlation of ground motions amongst different sites arises from: (1) a common earthquake source, (2) similar travel paths and local site conditions, and (3) sites being closely spaced with respect to the main fault asperities (Park et al., 2007). The first element of the total correlation (e.g. common source) is usually taken into account by the between-event (or inter-event) (δB) residual component provided by GMMs. Such a term represents the average shift of the observed IMs of an individual event from the median predicted by the GMM — it is common for all sites. Indeed, the ground motion caused by an earthquake may be on average larger/lower than the GMM median values obtained

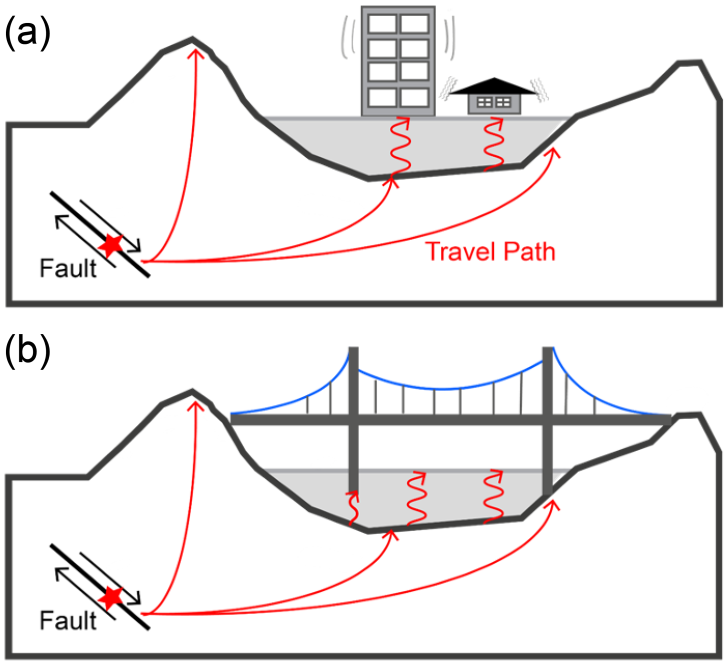


Figure 1.1: (a) Single-site seismic hazard assessment vs (b) multiple-site seismic hazard assessment. In (a) the hypothesis of independency of IMs at closely spaced sites is suitable for the estimation of the ground motion at a given site (e.g. where a building is located). By contrast, in (b) the description of the spatial variability of the ground motion is an essential element to quantify the IMs at various sites in the region.

from events of the same magnitude due to the intrinsic variability of the rupture process. The spatial correlation due to the similarity of travel path, site conditions and distance from the main slip asperities on the fault is generally accounted for by the within-event (or intra-event) (δW) residual component of the GMM. This component represents site-to-site differences of observations from the median event-corrected predictions. An illustrative scheme showing the different residual components is presented in Figure 1.2. Based on these definitions, it is understood that defining how δW varies spatially is necessary to quantify the joint probability of occurrence of IMs at multiple locations. It is also noted that the total correlation never drops to zero; despite the spatial dependency of δW decreasing as the inter-site distance increases, the δB correlation is always positive because of the commonality of the rupture process among the different sites.

In the general framework of correlation of earthquake ground motions, the spatial correlation of single IMs represents only one of the central tenets. Not only do the within-event

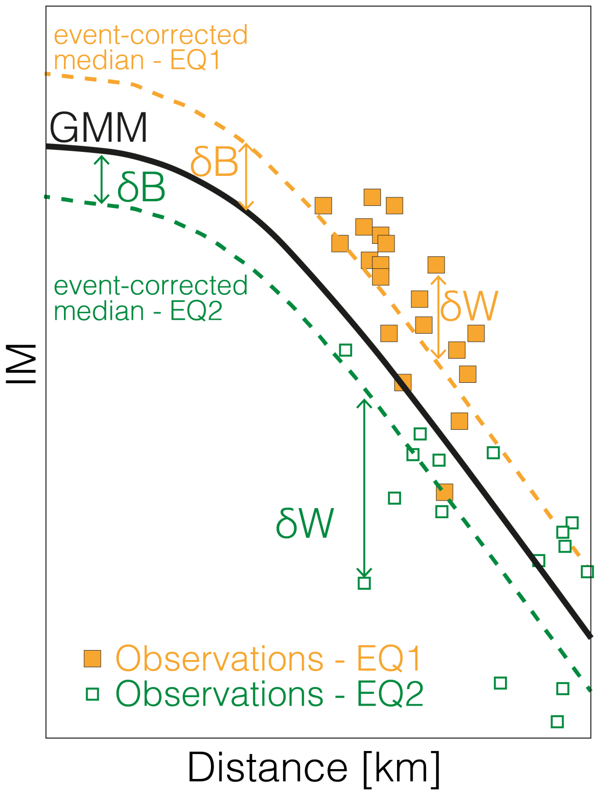


Figure 1.2: Between-event and within-event residual components of a GMM. The black solid line represents a GMM, whereas the orange and green dashed lines indicate the event-corrected ground-motion medians for earthquake 1 (EQ1) and earthquake 2 (EQ2), respectively. Squares indicate the observed IMs for the two events. [Modified from Atik et al. (2010)]

residuals feature spatial dependency, but they also exhibit cross-correlations among different IMs at the same site (e.g. Baker and Cornell, 2006; Weatherill et al., 2015; Markhvida et al., 2018; Kohrangi et al., 2021). Cross-correlation models are crucial in performance-based earthquake engineering applications, such as ground motion record selection to perform either dynamic analyses of structures or any other form of seismic response analysis. In this context, Baker (2011) and Baker and Lee (2018) as well as Bradley (2010*a*) proposed algorithms for ground motion selection (e.g. the mean conditional spectrum approach and the generalized conditional intensity measure approach), which account for the correlations among spectral parameters. To measure the joint distributions of multiple simultaneous IMs at the same site, different methodologies have been proposed: (1) Markov-type approximation (Goda and Hong, 2008); (2) linear model of co-regionalization (Loth and Baker, 2013); and (3) principal component analysis (Markhvida et al., 2018; Du and Ning, 2021).

Although considering spatial cross-correlations would certainly lead to more comprehensive regional hazard and risk assessments, this topic is beyond the scope of this thesis, which will focus only on the spatial correlation of individual ground-motion IMs.

The past twenty years have seen increasingly rapid advances in the field of spatial correlation (e.g. Boore et al., 2003; Jayaram and Baker, 2009; Esposito and Iervolino, 2011; Sokolov et al., 2012; Markhvida et al., 2018; Huang, Tarbali and Galasso, 2020; Baker and Chen, 2020; Sgobba et al., 2021). Most of the studies have assessed the spatial correlation of IMs of main engineering interest such as the peak ground acceleration (PGA) and spectral accelerations at different structural periods (SA), whereas little attention has been given to other IMs [e.g. peak ground velocity (PGV), peak ground displacement (PGD), cumulative absolute velocity (CAV)] useful for predicting damage to specific structural types and to assess earthquake-induced ground failure (e.g. Du and Wang, 2013; Foulser-Piggott and Goda, 2015; Huang, Tarbali and Galasso, 2020; Du and Ning, 2021). It is well-established that the spatial correlation of within-event residuals is affected by the inter-site distance, and that in general long-period IMs exhibit greater spatial correlation over longer distances. Despite all correlation models accounting for these aspects, significant differences among published relationships exist, meaning they predict different rates of decay of correlation with inter-site distance. Understanding the dependency of correlation on factors, such as magnitude, site conditions, ground motion frequency content and the spatial extend of faults, has been a topic of research over the past decades (e.g. Wesson and Perkins, 2001). A common consensus on these dependencies has not yet been achieved, however, and many questions are still under debate.

One of the greatest challenges in developing ground-motion correlation models is a lack of large numbers of closely-spaced recordings of previous earthquakes. This issue is partially overcome by the use of worldwide databases, which, on the other hand, may mask diagnostic spatial correlation characteristics due to dataset heterogeneities in terms of epicentral distributions, magnitude and geological conditions. In this context, physics-based ground-motion simulations (PBS) have emerged as a powerful data source (e.g. Smerzini and Pitilakis, 2018; Chen and Baker, 2019; Infantino et al., 2021). Not only do PBSs provide databases with exceptional size, especially in near-source regions, but they also facilitate sensitivity studies of factors that may affect spatial correlations by means of controlled and reproducible numerical experiments.

Several studies have examined the impact of spatial ground-motion variability on risk assessment and loss estimates at urban scales (e.g. Park et al., 2007; Weatherill et al., 2015; Wagener et al., 2016; Costa et al., 2018), as shown in Figure 1.3. Such analyses have highlighted the importance of considering the spatial correlation for a more appropriate quantification of seismic risk and estimations of potential losses. Common findings suggest that: (1) when the spatial dependency of IMs at closely-spaced sites is neglected, the losses are significantly underestimated at low probabilities of exceedance, whereas when IMs are perfectly correlated, rare losses are substantially overestimated (Park et al., 2007; Sokolov and Wenzel, 2011; Weatherill et al., 2015); and (2) the spatial correlation is an essential element to characterise the distribution of economic losses around the mean value (Wesson and Perkins, 2001; Wagener et al., 2016).

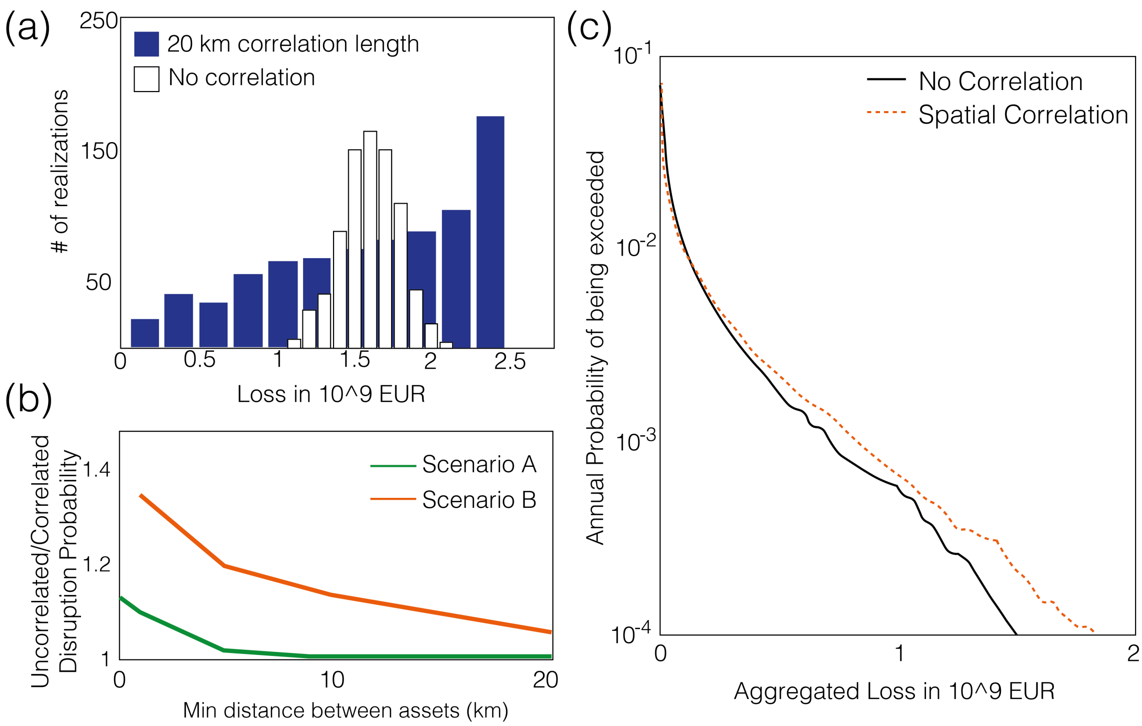


Figure 1.3: (a) Histograms of aggregated economic loss for correlated and non-correlated ground-motion fields [modified from Wagener et al. (2016)]; (b) Ratio between disruption probabilities considering spatially-uncorrelated and spatially-correlated ground-motion fields [modified from Costa et al. (2018)]; (c) Aggregated loss curves for the masonry-wall, mid-rise, low-code type for the Firenze Administrative Province [modified from Weatherill et al. (2015)].

In light of these considerations, it is evident that a proper definition of the spatial correlation of IMs is a prerequisite to provide a more accurate representation of the earthquake ground shaking and therefore enable more informed seismic hazard and risk assessments. This motivates, firstly, investigations into the main factors influencing the spatial correlation by critically reviewing already published studies and, secondly, based on this the provision of insights to guide developers and users of spatial correlation models. While there are no doubts on the relevance of spatial correlation in risk quantification, most studies have employed correlation models calibrated on either worldwide databases or regions other than the area of interest. The event-to-event and regional variability of the spatial correlation stimulates us to understand its effects on potential losses by developing custom correlation models for the region of interest as well as by including correlation-associated uncertainties.

1.2 Scope and Objectives

The aim of this thesis is to improve spatial correlation models of earthquake ground motions to enable more informed hazard and risk assessments. Therefore, this thesis attempts to:

1. critically review the state-of-art of spatial correlation modelling by discussing the impact of modelling decisions as well as by investigating the modelling elements that may influence the spatial dependency of ground motions (Chapters 2 and 4);
2. compare different methodologies to characterise the correlation structure, with the aim of understanding their main differences, advantages and limitations (Chapters 2 and 4);
3. provide evidence on the physical parameters that most affect the spatial correlation properties, which is essential to answer questions regarding the selection and development of either regionally-dependent or global models (Chapters 2 and 3);
4. form a helpful primer for scientists, engineers and loss modellers to understand, interpret and model spatial correlation in ground motion (Chapters 2,3 and 4);
5. advance the understanding of spatial correlation within earthquake catastrophe models by investigating different modelling hypotheses and including event-to-event and regional correlation-associated uncertainties (Chapter 5).

To this end, we employ synthetic ground-motion simulations, which provide controlled environments where the true correlation structure is known. The 2016-2017 Central Italy seismic sequence is chosen as a case-study because: (1) the abundance of high-quality recordings from a limited period of time and in a limited area allows quantification of the event-to-event variability for the same regional geological conditions; and (2) spatial correlations of ground motion associated with the type of faulting mechanism (normal) of these events has not been extensively addressed in literature. Finally, we take advantages of 3D PBSs for Norcia (Central Italy) to study the sensitivity of physical rupture parameters on correlation properties.

1.3 Outline of the Thesis

This thesis consists of a collection of peer-reviewed journal articles. Each chapter is therefore written in the form of stand-alone studies. Basic concepts may be repeated throughout the dissertation and notational conventions may be not strictly identical for each chapter. A complete list of acronyms and symbols can be found in the Acronysm and Symbols Section. The thesis is structured as follows:

Chapter 2 critically reviews the literature on the spatial correlation of earthquake ground motions. In particular, it provides a comprehensive overview of the current state of the art of spatial correlation modelling and identifies research gaps. Factors that may affect spatial correlations are investigated by using large databases of recorded strong ground motion from previous earthquakes and Monte-Carlo ground-motion simulations. A special focus is placed on: (1) comparing various elements of correlation models and modelling processes, (2) identifying differences that emerge from the comparison, and (3) characterizing the epistemic uncertainty in the process.

Chapter 3 extends the analysis of spatial correlation properties and explores the impact of other physical parameters, such as the source rupture process and hypocentre locations, by taking advantages of 3D PBSs. The study focuses on different ground-motion components and provides more insights into the hypothesis of isotropy and anisotropy of correlation by virtue of the unparalleled size of the simulated datasets.

Chapter 4 focuses on alternative approaches to characterize the spatial correlation of IMs. It sheds light on maximum-likelihood methods, which, in contrast to the classic semivariogram approach, directly determine the correlation structure. The impact of the number of recordings stations as well as the station layout is discussed and some guidelines are provided to account for spatial correlation uncertainty within seismic hazard and risk assessments.

Chapter 5 first develops *ad hoc* spatial correlation models for Italian macro-regions for the purposes of improving the seismic hazard and risk assessment for an upcoming earthquake catastrophe model for Italy. The models and their associated uncertainties are then implemented in deterministic and probabilistic calculations to: (1) advance the understanding of spatial correlations within the catastrophe modelling process, and (2) investigate the effects of such custom models on potential losses.

Chapter 6 summarises the major findings of this dissertation and its limitations and outlines potential directions for future research in this field.

1.4 List of Publications

The following is the full list of peer-reviewed journal articles and conference papers that resulted from this thesis.

1. **Schiappapietra**, E. & Douglas, J. (2019). *Spatial correlation of ground motions in the 2016-2017 Central Italy seismic sequence.* 2019 SECED conference, Greenwich (UK). 9-10 September 2019.
2. **Schiappapietra**, E. & Douglas, J. (2020). *Modelling the spatial correlation of earthquake ground motion: Insights from the literature, data from the 2016- 2017 Central Italy earthquake sequence and ground-motion simulations.* Earth-Science Reviews, 103139. *https://doi.org/10.1016/j.earscirev.2020.103139*
3. **Schiappapietra**, E. & Douglas, J. (2020). *Insights into spatially correlated ground motion intensity measures using Italian earthquakes.* 17^{th} World Conference in Earthquake Engineering, Sendai, Japan (postponed, 27 September - 2 October 2021).
4. **Schiappapietra**, E. & Smerzini, C. (2021). *Spatial correlation of broadband earthquake ground motion in Norcia (Central Italy) from physics-based simulations.* Bulletin of Earthquake Engineering. *https://doi.org/10.1007/s10518-021-01160-7*
5. **Schiappapietra**, E., & Douglas, J. (2021). *Assessment of the uncertainty in spatial-correlation models for earthquake ground motion due to station layout and derivation method.* Bulletin of Earthquake Engineering. *https://doi.org/10.1007/s10518-021-01179-w*

6. **Schiappapietra**, E., Stripajová, S., Pazák, P., Douglas, J., & Trendafiloski, G. (2021). *Exploring the impact of spatial correlations in the catastrophe modelling process: a case study for Italy.* Submitted to Bulletin of Earthquake Engineering.

CHAPTER 2

Modelling the spatial correlation of earthquake ground motion: Insights from the literature, data from the 2016-2017 Central Italy earthquake sequence and ground-motion simulations

Over the past decades, researchers have given increasing attention to the modelling of the spatial correlation of earthquake ground motion intensity measures (IMs), particularly when the seismic risk of spatially distributed systems and earthquake-induced phenomena is being assessed. The quantification of the seismic performance of these systems requires the estimation of simultaneous IMs at multiple locations during the same earthquake, for which the correlation between pairs of locations needs to be defined. Numerous spatial correlation models of common IMs, such as peak ground acceleration (PGA) and spectral acceleration (SA), have been published. Although the functional forms of the models are generally similar, significant discrepancies exist in terms of the rate of decay of the correlation with increasing inter-site separation distance. The main reasons for such differences lie with the selected databases, the ground-motion models used to derive the spatial correlation models, estimation approaches and regional geological conditions. Furthermore, little effort has been directed towards other IMs suitable to characterize the resulting damage to structures and predict ground failure: peak ground velocity (PGV), peak ground displacement (PGD) and spectral displacement (SD) as well as Arias intensity (I_a) and cumulative absolute velocity (CAV), to name but a few. A proper definition of the seismic action in terms of spectral displacement ordinates has progressively gained importance in performance-based seismic design, and I_a and CAV have been found to be adequate for many other earthquake engineering applications, such as evaluating the susceptibility to liquefaction and earthquake-induced landslides.

This chapter provides a comprehensive review of spatial correlation models, analysing factors that most affect the spatial dependency of IMs. We use strong-motion records from the 2016-2017 Central Italy earthquake sequence combined with ground-motion simulations to examine the influence of various factors on spatial correlation models. In particular, the dependency on: (1) the estimation method and model fitting technique; (2) the magnitude; (3) the response-spectral period; and (4) local-soil conditions, is investigated. This chapter is based on the following published article and conference proceeding (extended and updated where appropriate):

1. **Schiappapietra**, E., & Douglas, J. (2020). Modelling the spatial correlation of earthquake ground motion: insights from the literature, data from the 2016-2017 central Italy earthquake sequence and ground-motion simulations. *Earth-Science Reviews*, 203, 103139. *https://doi.org/10.1016/j.earscirev.2020.103139*
2. **Schiappapietra**, E., & Douglas, J. (2020, September). Insights into spatially correlated ground motion intensity measures using Italian earthquakes. In *17th World Conference on Earthquake Engineering.*

2.1 Introduction

Stakeholders, such as government, search-and-rescue organizations and private companies, require a reliable evaluation of the ground-motion field to assess the effects of an earthquake for more informed risk management and decision making designed to reduce economic and human losses (Park et al., 2007; Weatherill et al., 2015). The probabilistic assessment of ground-motion intensity measures (IMs) at a single site is now a well-established technique, and sophisticated methods have been developed for this aim. Traditional seismic hazard and risk analysis tools usually determine the ground motion caused by an earthquake through ground motion prediction equations (GMPEs). GMPEs provide an estimate of the ground shaking and its associated aleatory variability at a given site, considering IMs at different sites as independent. The seismic risk assessment of spatially-distributed systems, such as long bridges, water and power lifelines and portfolios of buildings, however, require not only the estimation of simultaneous IMs at multiple locations during the same earthquake, but also the quantification of the correlation structure (Goda and Atkinson, 2009, 2010; Jayaram and Baker, 2009; Esposito and Iervolino, 2011, 2012; Weatherill et al., 2015; Wagener et al., 2016; Heresi and Miranda, 2019). Indeed, understanding the spatial characteristics of the ground motion arising from similarities in the seismic wave paths and local-site effects is needed to provide a more accurate representation of ground-motion fields (Verros et al., 2017). For instance, correlation models can be used to generate spatially-correlated random fields for use in developing either scenarios for future earthquakes or retro-scenarios of past events in terms of ground motion shaking, as well as for loss estimates. Iervolino (2013), among others, discuss the importance of considering regional hazard for calculation of aggregate risk. This author compared the annual rate of exceedance of a specific IM in at least one site among several sites with the corresponding site-specific hazard assessment, demonstrating that consideration of the joint probability of occurrence in the hazard computation leads to larger values than those obtained for individual sites. Similar conclusions are found in Sokolov and Ismail-Zadeh (2016) and Sokolov and Wenzel (2019). These studies outline the discrepancies between site-specific PSHA and multiple-site PSHA: smaller within-event correlations combined with larger reference areas make the hazard assessment prone to remarkable differences in terms of annual rate of PGA exceedance in at least one site. Besides, Sokolov and Wenzel (2019) demonstrated that the fraction of the reference area in which the design ground motion level will be exceeded depends not only on the considered return period, but also on the correlation model. For instance, for a return period of 475 years, there is a 10% of probability that the design ground motion level will be exceeded at least once in 50 years in 20% and 40% of the reference area when within-event residuals are uncorrelated and perfectly correlated, respectively, with corresponding impacts on loss estimates. The importance of defining spatially-correlated ground-motion fields in seismic risk assessment was also demonstrated by Park et al. (2007) and Sokolov and Wenzel (2011). Neglecting the spatial correlation may cause a bias in loss estimates, overestimating the most likely losses and underestimating rare losses. On the contrary, overestimating the correlation may lead to the opposite result. As remarked in Park et al. (2007), the effects of including or not the spatial correlation depend also on the considered portfolio. Analogous outcomes are provided in Crowley et al. (2008), in which the authors compared the variation of the mean damage ratio of the loss model, obtained considering uncorrelated and correlated ground motion fields as well as correlated ground motion fields constrained to the recordings available at a specified number of sites. Given observations at recording stations and including proper correlation models makes the variability of the ground motions lower, and hence the variability of the losses narrower.

Over the past decade, the number of studies on spatial correlation has increased significantly. Many models have been published (Table 2.1) and common findings suggest that: (1) within-event correlation decays rapidly with increasing inter-site separation distances, and (2) ground-motion IMs associated with longer response-spectral periods have larger correlation lengths (e.g. Jayaram and Baker, 2009; Esposito and Iervolino, 2012; Wagener et al., 2016; Goda and Hong, 2008; Sgobba et al., 2019). Nevertheless, a thorough comparison among the proposed models demonstrates significant inconsistencies, which may make the assessed seismic risk prone to large uncertainties (Figure 2.1). The majority of the studies are for different regions (e.g. California, Japan and Taiwan), suggesting that the underlying database as well as regional and local-site effects are likely to play first-order roles in the observed differences. Besides, the use of either existing GMPEs or *ad hoc* GMPEs as well as different estimation approaches may contribute to different outcomes. Further analyses are thus required to draw firm conclusions on the spatial correlation structures of different IMs.

To address these issues, we use large databases of recorded strong ground motion from previous earthquakes and ground-motion simulations. In particular, we carry out a geostatistical analysis using the database of the 2016-2017 Central Italy earthquake sequence, which includes nearly 6900 records from 63 $M_w \geq 3.7$ events (and nearly 1600 records from nine events with $M_w \geq 5.0$) that occurred over a period of five months (August 2016 - January 2017). We choose this dataset because it allows: (1) removal of some of the uncertainties related to the underlying region; and (2) quantification of the variability of spatial correlation among different earthquakes when the same area is considered. Conversely, simulation of spatially-correlated ground motion fields provides a controlled environment to test the factors that most influence the correlation structure.

In this chapter our goal is to provide a well-structured review critically summarising the main findings to advance understanding of the spatial correlation of strong ground motion. We employ the outcomes from the simulations and observations to address the main issues and research gaps.

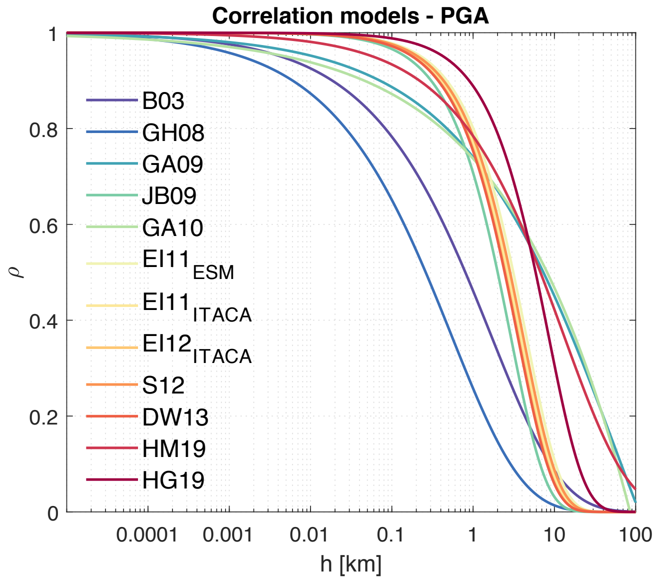


Figure 2.1: A typical selection of spatial correlation models for PGA: B03: Boore et al. (2003); GH08: Goda and Hong (2008); GA09: Goda and Atkinson (2009); JB09: Jayaram and Baker (2009); GA10: Goda and Atkinson (2010)); EI11ESM: Esposito and Iervolino (2011) ESM database; EI11ITACA: Esposito and Iervolino (2011) ITACA database; EI12ITACA: Esposito and Iervolino (2012) ITACA database; S12: Sokolov et al. (2012); DW13: Du and Wang (2013); HM19: Heresi and Miranda (2019); HG19: Huang and Galasso (2019).

Table 2.1: Main published correlation models. PGA is peak ground acceleration, PGV is peak ground velocity, SA is response spectral acceleration, Ia is Arias intensity, CAV is cumulative absolute velocity and T is spectral period. Mw is moment magnitude and RJB is Joyner-Boore distance. GMPE is ground motion prediction equation.

References	Database	Considered GMPEs	Method	Spatial Correlation Model	Ground Motion Intensity Measure
Boore et al. (2003)	M_w 6.7 1994 Northridge earthquake	-	Standard deviation of the difference of the logarithm of the PGA	Exponential	PGA
Wang and Takada (2005)	6 earthquakes recorded in Japan and Taiwan ($6.2 \leq M_w \leq 8.0$)	GMPE by Annaka et al. (1997) and by Midorikawa (2002))	Direct estimation of the sample correlation coefficient	Exponential	PGV
Goda and Hong (2008)	6 Californian earthquakes ($5.9 \leq M_w \leq 7.1$)	Ad hoc GMPE calibrated on 592 records from 39 Californian earthquake	Semivariogram and direct estimation of the correlation coefficient	Exponential	PGA and SA for T up to 3 s
	M_w 7.6 Chi Chi Earthquake	Boore and Atkinson (2008*a*)			PGA and SA at T = 0.3; 1; 3 s
Goda and Atkinson (2009)	K-NET and KiK-NET database: 106 earthquakes with $M_w \geq 5.5$ and depth ≤ 200 km	Ad hoc GMPE	Semivariogram and direct estimation of the correlation coefficient	Exponential	PGA and SA (the final model is independent of T)
Hong et al. (2009)	39 Californian earthquakes with $M_w \geq 5.0$	Ad hoc GMPE	Correlation is computed within the GMPE regression	Exponential	PGA and SA for T up to 3 s
Jayaram and Baker (2009)	6 Californian earthquakes ($5.1 \leq M_w \leq 6.7$) M_w 7.6 Chi Chi Earthquake	Boore and Atkinson (2008*b*) and Chiou and Youngs (2008)	Semivariogram (Classic estimator)	Exponential	SA for T up to 10 s
Goda and Atkinson (2010)	SK-NET, K-NET and KiK-NET database: 20 earthquakes with $5.6 \leq M_w \geq 6.8$ and $7km \leq depth \geq 70km$	GMPE by Goda and Atkinson (2009)	Semivariogram	Exponential	PGA and SA (the final model is independent of T)
Jayaram and Baker (2010)	dataset used in Campbell and Bozorgnia (2008)	GMPE by Campbell and Bozorgnia (2008)	Correlation is computed within the GMPE regression	-	PGA and SA for T up to 10 s
Sokolov et al. (2010)	TSMIP network in Taiwan: 66 earthquakes with $M_l \geq 4.5$ and depth ≤ 30 km	Ad hoc GMPE	Direct estimation of the sample correlation coefficient	Exponential	PGA

Esposito and Iervolino (2011)	Subset of ESM (M_w 5-7.6 and R_{jb} 0-100km) and ITACA (M_w 4-6.9 and R_{jb} 0-196km)	ESM: Akkar and Bommer (2010) ITACA: Bindi et al. (2010)	Semivariogram (Robust and classic estimator)	Exponential	PGA and PGV
Goda (2011)	PEER-NGA, K-NET, KiK-NET and SK-net database: 41 earthquakes with $M_w \geq 5.5$	GMPE by Boore and Atkinson (2008*a*) and by Goda and Atkinson (2009)	Semivariogram	-	SA for T up to 2 s
Esposito and Iervolino (2012)	Subset of ESM (M_w 5-7.6 and R_{jb} 0-100km) and ITACA (M_w 4-6.9 and R_{jb} 0-196km)	ESM: Akkar and Bommer (2010) ITACA: Bindi et al. (2011)	Semivariogram (Robust and classic estimator)	Exponential	SA: T up to 2 s for ITACA and 2.85 s for ESM
Sokolov et al. (2012)	TSMIP network in Taiwan: 54 earthquakes with $M_l \geq 5.0$and depth $\leq$30 km	GMPE by Sokolov et al. (2010) and by Tsai et al. (2006)	Semivariogram and direct estimation of the correlation coefficient	Exponential	PGA
Du and Wang (2013)	5 Californian earthquakes ($5.1 \leq M_w \leq 6.7$) 3 Japanese earthquakes ($6.6 \leq M_w \leq 6.8$) M_w 7.6 Chi Chi Earthquake	Campbell and Bozorgnia (2008), Campbell and Bozorgnia (2010), Campbell and Bozorgnia (2012)	Semivariogram (Robust estimator)	Exponential	SA for T up to 5 s, CAV and Ia
Sokolov and Wenzel (2013)	84 Japanese earthquakes ($4.2 \leq Mw \leq 7.4$)	GMPE by Kanno et al. (2006) and by Goda and Atkinson (2009)	Semivariogram and direct estimation of the correlation coefficient	Exponential	PGA and PGV
Bradley (2014)	2010-2011 Canterbury earthquakes	GMPE by Bradley (2010*b*, 2013)	Direct estimation of the sample correlation coefficient	-	SA for T up to 6 s
Foulser-Piggott and Goda (2015)	Subset of 203 Japanese earthquakes with more than 100 records	ad hoc GMPE calibrated on 661 Japanese earthquakes with $M_w \geq 5.0$, depth $\leq$ 150 km and $R \leq$300 km.	Semivariogram	-	Ia and CAV
Wagener et al. (2016)	8 earthquakes ($3.5 \leq M_w \leq 5.1$) recorded in the Marmara region	GMPE by Akkar and Bommer (2010)	Semivariogram	Exponential	PGA and SA for T up to 1 s
Garakaninezhad and Bastami (2017)	9 earthquakes ($5.2 \leq M_w \leq 7.6$) occurred in California, Japan and Taiwan	GMPE by Campbell and Bozorgnia (2014)	Semivariogram	-	PGA and SA for T up to 10 s

Infantino et al. (2018))	Data from ground motion simulations carried out using the numerical code SPEED (M_w 6.0 Po Plain earthquake, M_w 6.5 Volvi earthquake, M_w 7.0 Istanbul scenario, M_w 6.5 Beijing scenario	Ad hoc GMPE	Semivariogram (Classic estimator)	Exponential	PGA and SA for T up to 3.5 s
Stafford et al. (2019)	24 earthquakes occurred in the Groningen field ($2.5 \leq M_l \leq 3.6$)	GMPE by Bommer et al. (2018) in Stafford et al. (2019)	Semivariogram (Classic estimator)	Exponential	SA for T up to 1 s
Chen and Baker (2019)	Physics-based simulations from the CyberShake platform (scenarios for souther California)	-	Semivariogram and direct estimation of the correlation coefficient	Exponential	PGA and SA for T up to 10 s
Heresi and Miranda (2019)	39 well recorded worldwide earthquakes	GMPE by Boore et al. (2014)	Semivariogram	Exponential	PGA and SA for T up to 10 s
Huang and Galasso (2019)	233 Italian earthquakes ($4 \leq M_w \leq 6.9$)	Ad hoc GMPE	Correlation is computed within the GMPE regression	Exponential	PGA and SA for T up to 4 s
Ming et al. (2019)	Simulations based on 62 Italian earthquakes ($5 \leq M_w \leq 6.9$) to validate the scoring algorithm approach	Ad hoc GMPE	Correlation is computed within the GMPE regression		PGA
Sgobba et al. (2019)	29 earthquakes of the 2012 Emilia sequence	GMPE by Lanzano et al. (2016)	Semivariogram (Classic estimator)	Exponential	PGA and SA for T up to 4 s

2.2 Basic definitions

The correlation of ground-motion IMs from an earthquake includes three main elements, namely: (1) the spatial correlation among residuals of the same IM for adjacent sites; (2) the correlation among residuals of different IMs at the same location; (3) the spatial cross-correlation among residuals of different IMs for closely spaced sites (Weatherill et al., 2015). We focus on the first of the above-mentioned elements. In general, the similarity of ground-motion IMs at two different sites depends on: (1) the earthquake source; (2) the propagation path from the source to the sites and local-site effects; (3) the position of closely-spaced sites in near-source conditions with respect to the main fault asperities (Park et al., 2007). The first of these aspects is commonly accounted for by the between-event residual provided by the GMPE. GMPEs relate a ground motion IM (e.g. peak ground acceleration, PGA; peak ground velocity, PGV; peak ground displacement, PGD; or pseudospectral acceleration for 5% of critical damping, SA) to a set of explanatory variables describing the source (e.g. magnitude and faulting mechanism), the wave propagation path (e.g. distance metric and regional effects) and the site conditions (e.g. soil classification) (e.g. Douglas and Edwards (2016)). IMs are commonly modelled as lognormally-distributed random variables, through a mixed-effects approach. Therefore, GMPEs take the form:

$$\log_{10}(Y_{ij}) = \log_{10}\overline{Y_{ij}}(M, R, S, \theta) + \varepsilon_{ij} + \eta_i \tag{2.1}$$

where Y_{ij} is the IM of interest at the j^{th} site due to the i^{th} event, whereas $\overline{Y_{ij}}$ is the predicted median function of magnitude (M), distance from the source (R), local-site conditions (S) and others explanatory variables (θ). η_i is the between-event residual term, assumed as an independent, identically and normally distributed random variable with zero mean and standard deviation τ. It denotes the systematic deviation of observed IMs associated to an event with respect to the GMPE prediction and does not depend on the site. Conversely, ε_{ij} represents the independent within-event residual term, which is site dependent as it accounts for differences from the average model due to the path and local-site effects. ε_{ij} follows a multivariate Gaussian distribution, completely defined by its mean function $E[\varepsilon_{ij}]$, and covariance function, which reflects the correlation of within-event residuals:

$$COV_i(jk) = E[\varepsilon_{ij}\varepsilon_{ik}] - E[\varepsilon_{ij}]E[\varepsilon_{ik}] \tag{2.2}$$

The within-event correlation between all pairs of sites during an earthquake due to the similarity of travel path and local-site effects depends on the inter-site separation distance. Typically, it is modelled as an exponential function:

$$\rho_\varepsilon(h) = \exp(\alpha h^\beta) \tag{2.3}$$

where α and β are the model coefficients, usually inferred through a least-squares approach, and h is the separation distance.
As above-mentioned, the between-event correlation, which arises from commonality of the rupture process, is taken into account by the η_i term and is defined as the ratio between the variability components:

$$\rho_\eta = \tau^2/\sigma_T^2 \tag{2.4}$$

in which $\sigma_T^2 = \sqrt{\tau^2 + \phi^2}$ is the total standard deviation and ϕ is the standard deviation of the within-event residuals. Therefore, the total correlation among residuals for closely spaced sites, which considers both the between-event and within-event variability, is given by:

$$\rho_T(h) = \rho_\eta + \rho_\varepsilon \frac{\phi^2}{\sigma_T^2} \tag{2.5}$$

It is noted that the between-event correlation among different pairs is always positive, as a result of all sites sharing the same source rupture. As a matter of fact, the total correlation will never drop to zero, although $\rho_\varepsilon(h)$ decays to zero with increasing separation distance (Heresi and Miranda, 2019). In this study, we focus only on the within-event correlation because the within-event residual term is the only component of the total variability that varies from site to site and thus affects the spatial dependency of IMs (Stafford, 2012).

Furthermore, only one observation from a given earthquake is available for each station, making it impossible to draw any inferences from it (Webster and Oliver, 2007). Therefore, the hypothesis of second-order stationarity and isotropy are generally assumed, so that the mean function of the random variable (ε_{ij}) is constant for all sites, and the covariance $COV_i(j,k)$, and thus the correlation, depends only on the separation distance *(h)* between two sites and not on their absolute position and orientation $[\rho(\varepsilon_{ij}, \varepsilon_{ik}) = \rho(h)]$. Finally, one should be aware that the correlation is usually computed on the residual terms rather than directly on the IM. Indeed, the diverse underlying distributions of each IM value, due to different explanatory variables (such as R, S, θ), would make the assessment of the IMs correlation inappropriate (Heresi and Miranda, 2019).

2.3 Databases

2.3.1 Central Italy earthquake sequence

Starting from 24^{th} August 2016, one of the most important earthquake sequences ever recorded in Italy struck the Central Apennines between the municipalities of Amatrice and Norcia, causing widespread damage, thousands of homeless and invaluable losses for the historical heritage of the region. The first mainshock (M_w 6.0) struck on 24^{th} August 2016 at 01:36 UTC near Amatrice and it was followed, within less than an hour, by a M_w 5.4 aftershock (Chiaraluce et al., 2017). After two months, two other large earthquakes occurred: a M_w 5.9 on 26^{th} October 2016 at 19:18 UTC, near the village of Ussita and a M_w 6.5 on 30^{th} October 2016 at 06:40 UTC with an epicentre close to Norcia (Luzi et al., 2017). Four other $M_w \geq 5.0$ earthquakes occurred on 18^{th} January 2017 near the villages of Campotosto and Montereale (Figure 2.2, Table 2.2).

Table 2.2: Main characteristics of the largest magnitude events of the Central Italy sequence. [source: Lanzano, Sgobba, Luzi, Puglia, Pacor, Felicetta, D'Amico, Cotton and Bindi (2019)]

Earthquake	Event Time	Latitude [°]	Longitude [°]	Depth [km]	M_w	# Records
Amatrice	24/08/2016 01:36	42.7	13.23	8.1	6.0	158
Amatrice Aftershock	24/08/2016 02:33	42.79	13.15	8	5.3	138
Visso I	26/10/2016 17:10	42.88	13.13	8.7	5.4	158
Visso II	26/10/2016 19:18	42.91	13.13	7.5	5.9	166
Norcia	30/10/2016 06:40	42.83	13.11	9.2	6.5	159
Montereale I	18/01/2017 09:25	42.55	13.26	9.2	5.1	127
Montereale II	18/01/2017 10:14	42.53	13.28	9.1	5.5	141
Pizzoli I	18/01/2017 10:25	42.49	13.31	8.9	5.4	129
Pizzoli II	18/01/2017 13:33	42.48	13.28	10	5.0	124

Event and stations metadata are from Lanzano, Sgobba, Luzi, Puglia, Pacor, Felicetta, D'Amico, Cotton and Bindi (2019). All the events, generated by normal fault segments, were recorded by permanent and temporary networks, set up to monitor the earthquake sequence at a high resolution and to retrieve more accurate observations of the ground shaking in the near-source region (Luzi et al., 2017). In this study, we select data from 44 $M_w \geq 4.0$ well-recorded events with more than 100 recordings, recorded by 367 strong-motion stations within an epicentral distance of 200 km. The distributions of the selected data with respect to distance, magnitude and distance, magnitude and number of stations per EC8 soil classes (CEN, 2004) are summarized in Figure 2.3. The site conditions at

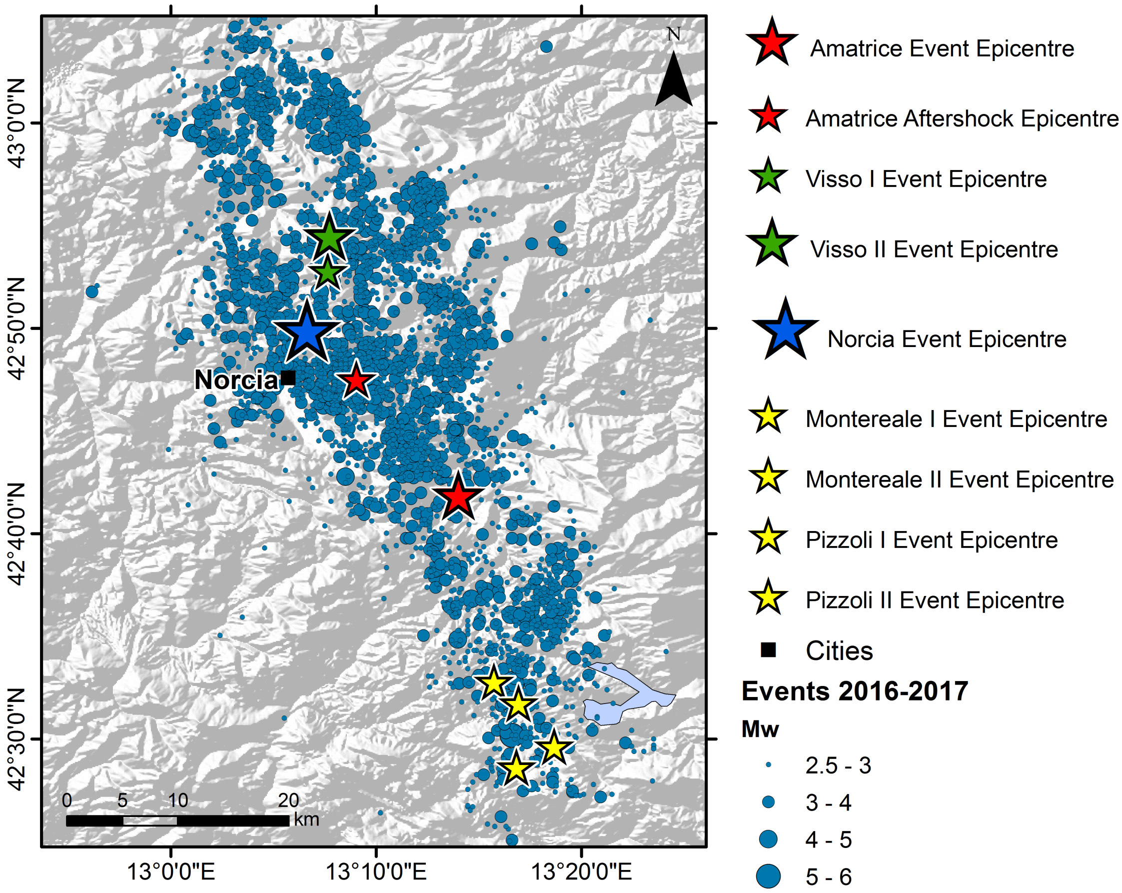


Figure 2.2: Epicentres (stars) of the nine $M_w \geq 5.0$ earthquakes of the sequence. Mw $\geq$ 2.5 events for one year after the Amatrice mainshock are also mapped with blue dots to better highlight the spatial extent of the sequence. [source: *http://cnt.rm.ingv.it/*].

each strong-motion station are expressed through the EC8 soil categories, which is based either on the average shear-wave velocity of the upper-most 30 m ($V_{S,30}$) or on the available geological information. Only about 25% of the considered stations are characterised by shear-wave velocity profiles, so that most of them are classified using geology. The majority of the selected stations are labeled as site class B, which includes very dense sand or gravel and very stiff clay deposits.

2.3.2 Simulated ground-motions fields

To simulate spatially-correlated ground-motion fields, we employ the approach described in Strasser and Bommer (2009), which accounts for both the between- and within-event

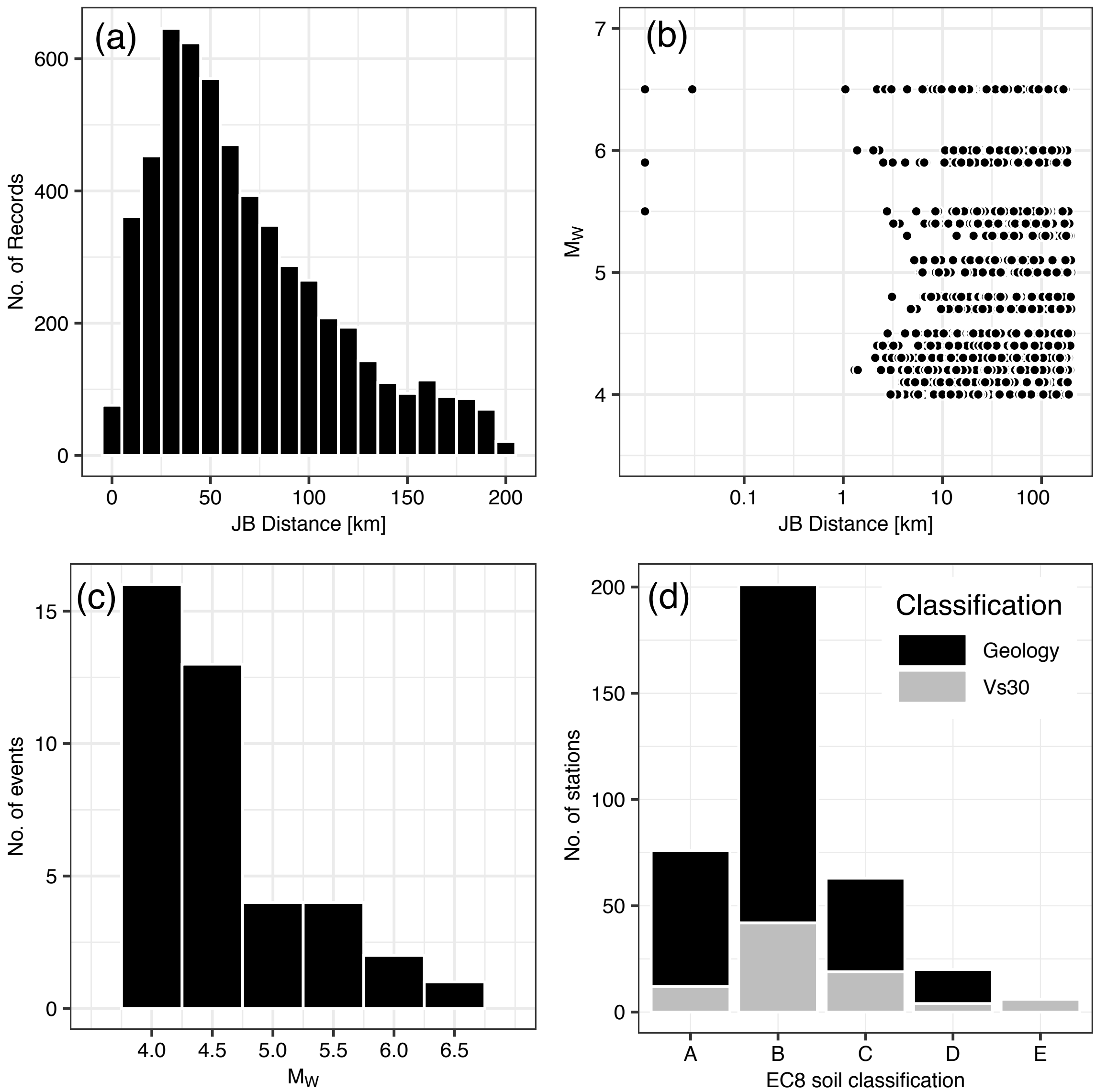


Figure 2.3: (a) Records-distance distribution; (b) Magnitude-distance distribution; (c) Events-magnitude distribution; (d) Distribution of stations in terms of EC8 site classification. Distances are in terms of Joyner and Boore distance (R_{jb}, i.e. the distance to the surface projection of the rupture).

variabilities. The former is specified through the between-event standard deviation in the IMs computation, whereas the latter is included through the computation of a random field based on a multivariate normal distribution. The procedure includes the following steps:

[1] Generation of median ground motion fields using GMPEs ($GMdet$). We employ the GMPE by Lanzano, Luzi, Pacor, Felicetta, Puglia, Sgobba and D'Amico (2019) for

shallow crustal earthquake in Italy, considering homogeneous rock site conditions (Figure 2.4a).

[2] Generation of ground motion random fields (Figure 2.4c).
The spatially correlated random field $P(h_0, \phi)$ is computed through a multivariate normal distribution characterized by an exponential correlation model with correlation length h_0 and standard deviation ϕ (see section 2.2). We choose different values of h_0 (5, 10 and 30 km) and ϕ (0.1, 0.32 and 0.5) to allow a deeper insight into the impact of these parameters on the analysis.

[3] Generation of ground-motion stochastic fields (Figure 2.4b).
This is obtained by combining steps 1 and 2:

$$GM_{sto} = 10^{[GM_{det}+P(h,\phi)]} \tag{2.6}$$

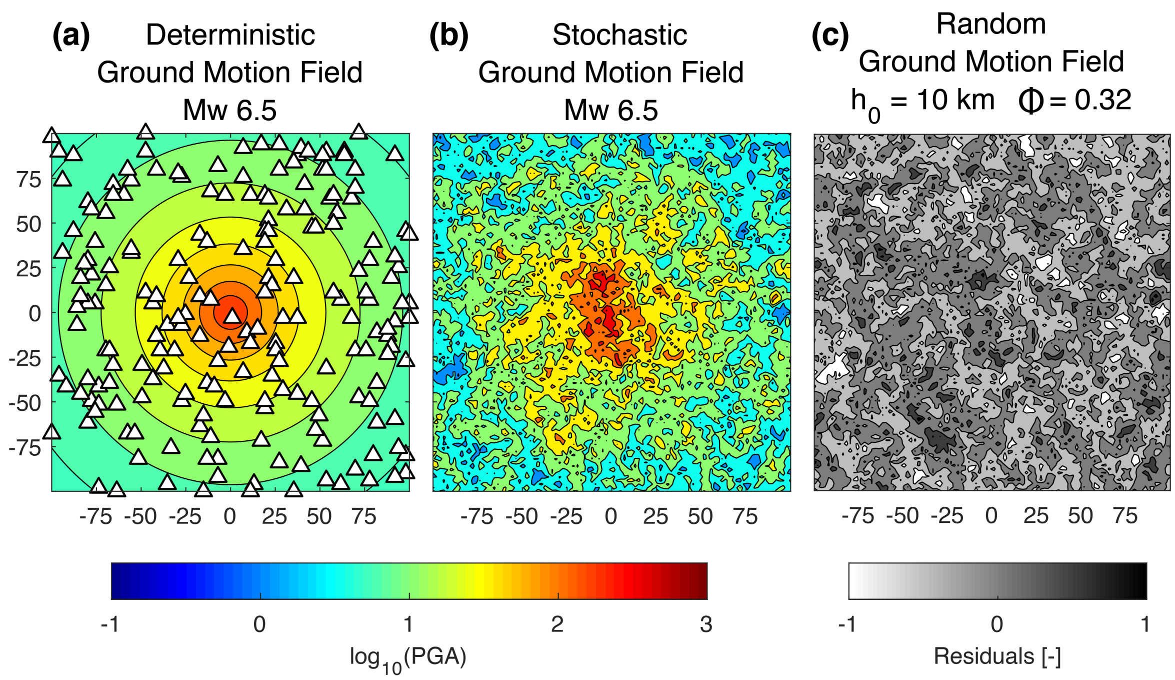


Figure 2.4: Example of median (deterministic) (a), stochastic (b) and random (c) ground-motion fields. obtained through the above-described approach. PGAs are expressed in cm/s^2. The triangles represent the stations used in the analysis.

The ground-motion fields are generated on a 200 km × 200 km grid with a 2 km resolution. Finally, we randomly locate strong-motion recording stations throughout the region and use the simulated IMs at these stations. We repeat the process 1000 times for each $h_0 - \phi$ pair to obtain stable results.

2.4 Modelling of spatial correlation

2.4.1 Estimation of the within-event spatial correlation

It is common practice to adopt an existing GMPE or an *ad hoc* study-specific GMPE to compute the within-event residuals at each site in order to assess the within-event spatial correlation $\rho_\varepsilon(h)$. This can be estimated with two different approaches, namely: (1) computing directly the covariance and the correlation coefficient (e.g. Wang and Takada (2005); Sokolov et al. (2010, 2012)), or (2) calculating the sample semivariogram, which measures the average dissimilarity between spatially distributed data (e.g. Jayaram and Baker, 2009; Esposito and Iervolino, 2011, 2012; Wagener et al., 2016; Heresi and Miranda, 2019; Sgobba et al., 2019). In the first method, the spatial correlation is estimated as:

$$\rho_\varepsilon(h) = COV(\varepsilon_{ij}, \varepsilon_{ik})/\phi^2 \tag{2.7}$$

where COV denotes the covariance function, whereas ε_{ij} and ε_{ik} are the within-event variabilities for the j^{th} and k^{th} sites during the i^{th} event with zero mean and standard deviation ϕ.

The second method computes the experimental semivariogram to represent the spatial dependency of IMs values with varying separation distance, which is defined as:

$$\hat{\gamma}(h) = \frac{1}{2}Var[\varepsilon_{ij} - \varepsilon_{ik}] \tag{2.8}$$

in which Var indicates the variance. It is common practice to adopt two different estimators to compute the sample semivariogram:

[1] Method of moments (Matheron, 1962)

$$\hat{\gamma}(h) = \frac{1}{2N(h)} \sum_{N(h)} [\varepsilon_{ij} - \varepsilon_{ik}]^2 \tag{2.9}$$

where $\hat{\gamma}(h)$ represents the empirical semivariogram and $N(h)$ is the number of pairs separated by h.

[2] Estimator proposed by Cressie (1985)

$$\hat{\gamma}(h) = \frac{1}{2}\frac{\left[\frac{1}{|N(h)|}\sum_{N(h)}|\varepsilon_{ij} - \varepsilon_{ik}|^{0.5}\right]^{4}}{0.457 + \frac{0.494}{|N(h)|}} \tag{2.10}$$

This is found to be a more robust estimator, being less sensitive to outliers (Esposito and Iervolino, 2011, 2012; Du and Wang, 2013; Oliver and Webster, 2014).

Both the covariance and the semivariogram are calculated for each pair of stations (x_j, x_k) whose inter-site spacing falls in a distance bin defined as $h - \Delta h/2 \leq |x_j - x_k| \leq h + \Delta h/2$. Esposito and Iervolino (2012) and Du and Wang (2013) suggest setting the bin size in such a way that there are at least 30 pairs in each bin. Other authors (e.g. Wagener et al., 2016) advise having at least 100 pairs per bin to have more reliable and representative estimations. Furthermore, the definition of ϕ is of primary importance. ϕ can be estimated either from the sample semivariogram at large separation distances, where the within-event residuals are assumed to be uncorrelated, or as the standard deviation of the within-event residuals for a given event (Goda and Atkinson, 2010). Alternatively, Esposito and Iervolino (2012) employ the standard deviation related to the GMPE as ϕ. However, this approach is strongly discouraged by other studies, since using a constant value of ϕ for different events might lead to a biased within-event correlation (Heresi and Miranda, 2019; Foulser-Piggott and Goda, 2015).

Parametric functions are used to fit the experimental values computed either through the sample semivariogram or covariance approaches. This allows for the spatial variation of ε_{ij} to be retrieved for any separation distance h. Examples of basic second-order stationary and isotropic models are: (1) Matern; (2) Exponential; and (3) Gaussian (Webster and Oliver, 2007). The exponential model (Figure 2.5), which usually provides the best performance, takes the form:

$$\gamma(h) = a[1 - \exp(-ch/b)] \tag{2.11}$$

where h is the separation distance, a and b are the sill and the range of the semivariogram and c is a given positive constant set to 3. The sill equals the variance of the data, whereas the range represents the distance beyond which the correlation between sites is negligible. In the case of the exponential model, the range is the distance at which $\gamma(h)$ equals 0.95 times the sill. It is worth mentioning that in some studies (e.g. Wang and Takada, 2005; Huang and Galasso, 2019) c is set equal to 1. Consequently, the range outlines the distance at which the correlation equals $\exp(-c) = \exp(-1) \cong 0.37$. The implementation of either one or the other value leads to a different meaning of the range, without however affecting the spatial correlation structure of the IM of interest. Moreover, under the hypothesis of second-order stationarity, the semivariogram and the covariance function are equivalent, so that the following relation holds (Oliver and Webster, 2014):

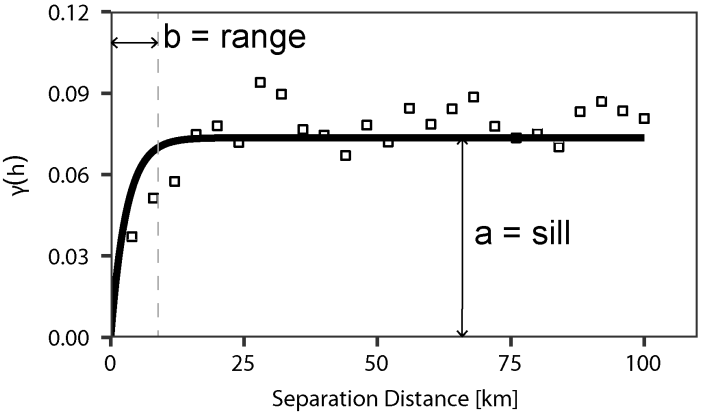


Figure 2.5: Exponential model. The black solid line is the theoretical model, whereas black squares represent the experimental semivariogram. The sill and range parameters are also highlighted.

$$\gamma(h) = \phi^2 - COV(\varepsilon_{ij}, \varepsilon_{ik}) = \phi^2[1 - \rho_\varepsilon(h)] \tag{2.12}$$

being $\rho_\varepsilon(h) = COV(\varepsilon_{ij}, \varepsilon_{ik})/\phi^2$.
Consequently, combining Eq. (2.11) and Eq. (2.12), we have that:

$$\rho_\varepsilon(h) = 1 - \frac{\gamma(h)}{\phi^2} = \exp(-ch/b) \tag{2.13}$$

which is equivalent to Eq. (2.3).

Furthermore, the hypothesis of second-order stationarity may sometimes not hold. The expected value of the random variable $E[\varepsilon_{ij}]$ may not be constant across all sites, but indeed varying depending on the location. In such cases, the semivariogram $\hat{\gamma}(h)$ increases with separation distance, without reaching a stable sill (Diggle et al., 2007). These long-range spatial trends should be removed so that small-range correlation structures can be detected. It is common practice to model spatial trends through trend surface models, namely the mean function is described by either first- or second-order polynomial functions of the geographic coordinates (Diggle et al., 2007; Oliver and Webster, 2014). Alternatively, the spatial trends may be defined as a function of any other physical parameter of the site, rather than coordinates, in order to better capture the spatial variation of the mean function (Diggle et al., 2007). Besides, the trend and the semivariogram of the residuals can be estimated all at once through the residual maximum likelihood approach, as suggested by Oliver and Webster (2014) and Papritz (2018).

Different approaches have been used in previous studies to fit the experimental data, such as the weighted least-squares and manual-fitting techniques. Jayaram and Baker (2009), among others, suggest inferring the model parameters manually so that the experimental data are better fitted at shorter separation distances. This technique should, however, be discouraged as it involves a certain degree of subjectivity that can lead to prediction errors. Regarding the least-squares approach, weights can depend on either the separation distances (e.g. increasing the weights of data at shorter distances) or the number of pairs in each bin.

2.4.2 Modelling caveats

The previously described approach suffers from some shortcomings. GMPEs (eq. 2.1) are usually developed through regression analysis using either the two-stage algorithm proposed by Joyner and Boore (1993) or the one-stage mixed-effects approach by Abrahamson and Youngs (1992), considering independent within-event residuals. The similarity of the earthquake source, path and local-site effects leads to spatially correlated within-event residuals. This issue was firstly investigated by Hong et al. (2009)), who demonstrated that the inclusion of spatial correlation in GMPE development does not affect the estimated ground motion model coefficients but it does affect the variability of the between- and within-event residuals. In particular, the between-event variance decreases, whereas the within-event component increases, with consequences for the assessment of seismic risk of spatially-distributed systems. Similarly, Jayaram and Baker (2010) adopted a different approach and drew analogous conclusions. While spatial correlation model parameters can be inferred from statistical analysis of residuals independently of GMPE regressions, caution must be applied in the estimation of ϕ and τ. Ming et al. (2019) reviewed the studies by Hong et al. (2009) and Jayaram and Baker (2010)), and pinpointed their main limitations. They proposed a new algorithm (the scoring estimation approach) that, while achieving the same results in terms of ϕ and τ, provides more statistically robust ground-motion models along with spatial correlation parameters to improve seismic hazard and risk assessments.

In addition, seismic risk analyses usually employ an average estimate of the correlation structure, derived considering heterogeneous databases, without accounting for its event-to-event variability. Goda (2011) investigated the event-to-event variability of within-event spatial correlation, studying 41 different earthquakes individually. This author found that the model by Goda and Atkinson (2010) reasonably fits the overall median tendency. However, ±0.1 units should be added to the same spatial correlation model to take into account the event-to-event uncertainty. Likewise, Heresi and Miranda (2019) quantified the event-to-event variability, comparing the within-event residuals correlation of 39 well-recorded earthquakes. They provided general equations to calculate the median and dispersion of the correlation length (range), so that the uncertainty can be easily accounted for in regional seismic risk computations.

In Figure 2.6, we compare the outcomes obtained analysing the data from the 2016-2017 Central Italy sequence (9 events with M_w greater than 5.0) with the models proposed by Goda (2011) and Heresi and Miranda (2019) along with their event-to-event uncertainty. We follow the approach described in Heresi and Miranda (2019) to compute the central tendency and the standard deviation, based on the number of stations that recorded each event. The Central Italy database features a smaller event-to-event variability because all the earthquakes nucleated within the same region. Conversely, the two other models are calibrated on worldwide datasets, leading to a larger variability in terms of correlation length. According to our analysis, a region-specific variability should be considered when performing regional seismic hazard and risk assessment to obtain more accurate results.

2.5 Major factors influencing spatial correlation

In this section we present a detailed literature review and the results from our analyses, with the aim of understanding the main factors that affect the spatial correlation of ground-motion IMs.

2.5.1 Dependence on the estimation approach

In section 2.4.1 we described two different approaches to estimate the within-event spatial correlation, namely: (1) the covariance and correlation coefficient and (2) the semivariogram. The implementation of either one or the other technique should provide similar results as the semivariogram and the covariance are equivalent for second-order

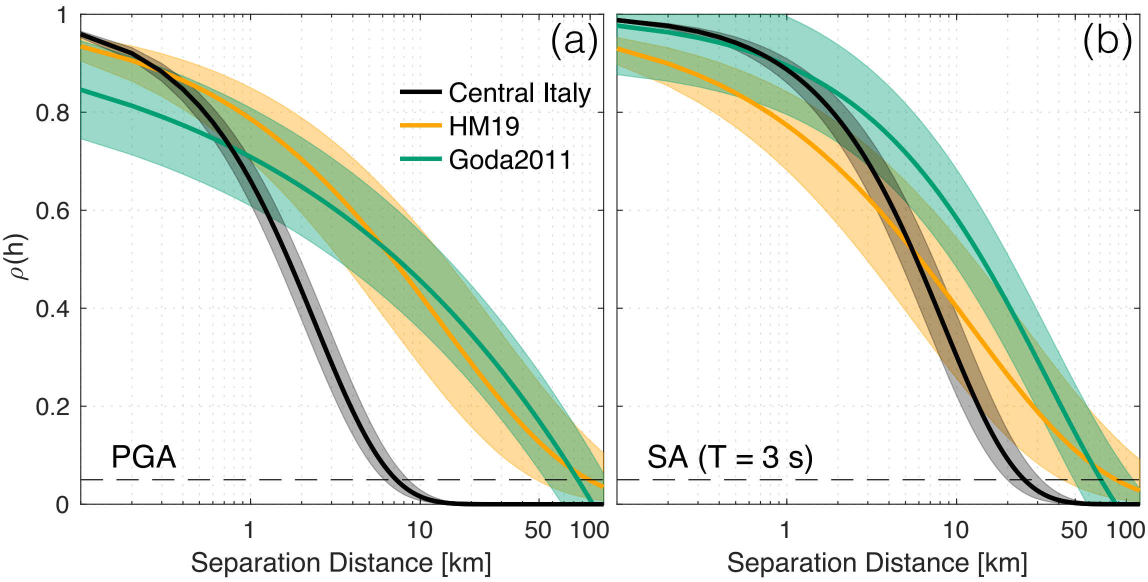


Figure 2.6: Correlation models proposed by Goda (2011) and Heresi and Miranda (2019) [HM19] and obtained considering data from the Central Italy sequence for PGA (a) and SA(3 s) (b). Solid lines represent the median values, whereas shadow areas indicate the event-to-event uncertainty. Black dashed line helps indicate the distances at which the correlation equals 0.05.

stationary random fields. Goda and Hong (2008) compared the results obtained for the M_w 7.6 Chi Chi earthquake using both the approaches and found that the computation of the covariance provides slightly different estimates. Similarly, Goda and Atkinson (2009), while characterizing the spatial correlation of Japanese data, drew the same conclusions. The correlation decays faster when the correlation coefficient is directly computed, and some inconsistent trends may occur, especially if the number of available data is inadequate. Nevertheless, they provided a general correlation model by averaging the results obtained using both approaches, as subsequently suggested by Sokolov et al. (2012) and Sokolov and Wenzel (2013).

In order to address this issue, we carry out a thorough geostatistical analysis of the within-event correlation structure using simulated spatially-correlated ground-motion fields. We compute the range implementing both the techniques and both the sample semivariogram estimators by varying the number of sites. As found by Goda and Hong (2008) and Goda and Atkinson (2009), the direct estimation of the correlation coefficient not only features slightly different results, but also a larger dispersion compared to the semivariogram approach in our simulations (Figure 2.7). Furthermore, the graph well illustrates how the availability of stations plays a crucial role in obtaining more robust outcomes: even though the mean of h_0 assumes to some extent a constant value, the uncertainty halves as the number of stations increases.

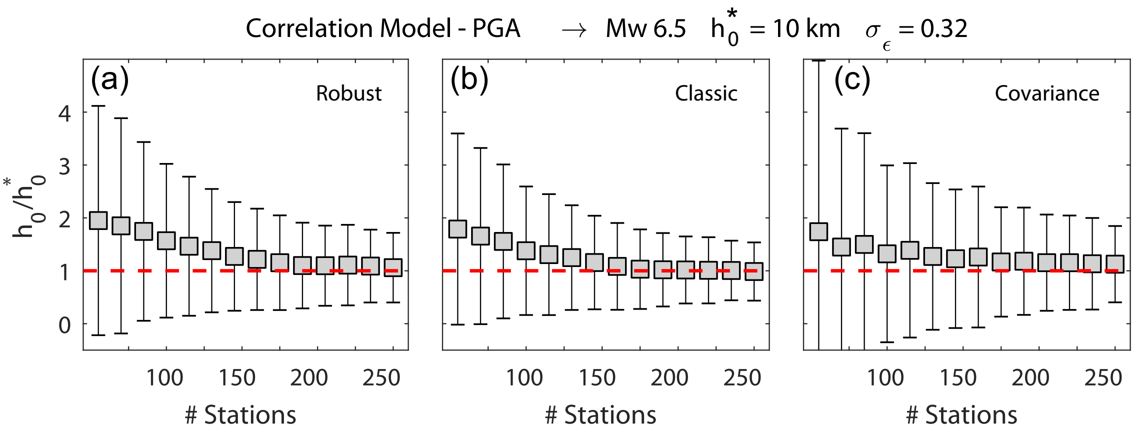


Figure 2.7: Estimated range h_0 normalized by the true value of the range $h*0$ as a function of the number of sites (stations) using: (a) the Cressie and Hawkins (robust) estimator; the Matheron (classic) estimator; and (c) the covariance approach. Grey squares represent the mean value, whereas the solid vertical lines indicate the variability of the estimates. The red dashed line signifies a ratio of 1.

Most of the proposed models use the classic estimator of Matheron (1962) to compute the sample semivariogram. The estimates obtained are variances and as such they are sensitive to outliers, which might lead to less reliable semivariograms. A statistically more robust estimator was proposed by Cressie (1985) to down-weight the effects of atypical observations (Oliver and Webster, 2014). Esposito and Iervolino (2011) and Esposito and Iervolino (2012) implemented both the estimators, concluding that there are negligible differences. Du and Wang (2013) suggested using the robust approach directly to obtain consistent estimations. Clearly, the two different estimators provide almost the same correlation distance in terms of average value, as observed in Figure 2.7 and according to Esposito and Iervolino (2011) and Esposito and Iervolino (2012). However, a more in-depth analysis suggests that: (1) the classic estimator is likely to converge faster; (2) the two different approaches do not provide always the same range, as shown in Figure 2.8. In particular, the smaller the number of considered stations, the more significant are the differences. Although our simulations do not show which is the best semivariogram estimator, Oliver and Webster (2014) demonstrated that if the data features outliers, the classic estimator diverges from the input function, in contrast to the robust estimator.

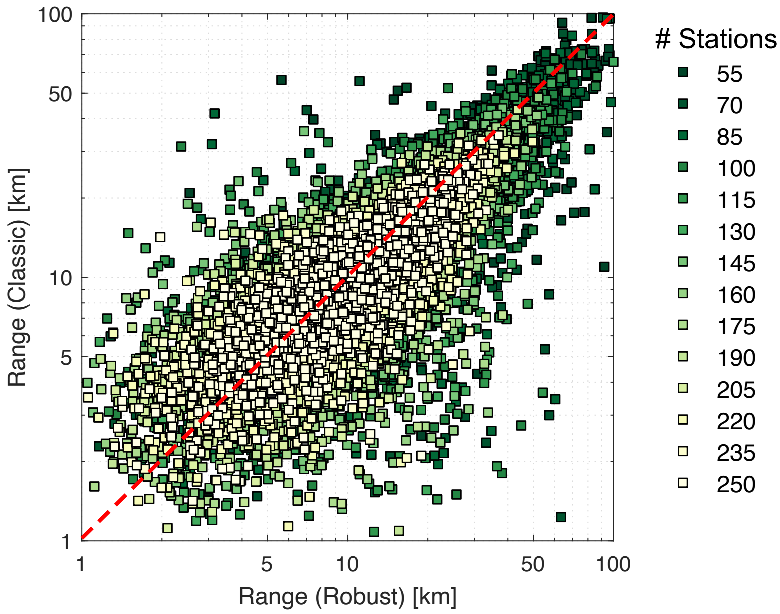


Figure 2.8: Comparison between the ranges obtained through both the classic and robust estimators, considering different numbers of stations.

2.5.2 Dependence on the fitting method

A number of methods have been proposed to fit experimental values by means of parametric models. The trial-and-error (manual fitting) and least-squares regression, among others, are the most common techniques. Jayaram and Baker (2009) recommend manually fitting the experimental semivariogram so that shorter separation distances, which are more important for engineering purposes, are weighted higher. Despite its high degree of subjectivity, this approach was chosen also by Du and Wang (2013) for its versatility in fitting the data. Esposito and Iervolino (2011) and Infantino et al. (2018) visually fitted experimental semivariograms, assuming the least-squares regression as a basis. Conversely, other studies, such as Esposito and Iervolino (2012), Sokolov et al. (2012), Wagener et al. (2016), Heresi and Miranda (2019) and Sgobba et al. (2019) adopt the weighted least-squares regression approach, using different weights. Some of the authors prioritise either the short separation distances or the number of pairs in each bin; others apply the same weight to all the experimental values, so that the model is fit equally over the full range of data. In our analysis, we opt for the weighted least squares regression, in which weights are computed based both on the number of pairs and on the separation distance, so that the impact of longer distances is minimized as they are associated with low correlations (Jayaram and Baker, 2009). However, we carry out a further analysis in which all the data are equally weighted to understand the influence of the fitting method on the correlation structure. Figure 2.9 illustrates how the different weightings lead to dissimilar results. A similar comparison can be found in Jayaram and Baker (2009), where the authors criticised the study by Wang and Takada (2005). Stafford et al. (2019) achieved similar outcomes, arguing that there is variability in the ranges due to the adopted technique. Therefore, according to Jayaram and Baker (2009), among others, we suggest prioritising models that fit the experimental data well at short separation distances, which have a significant impact on seismic risk estimates.

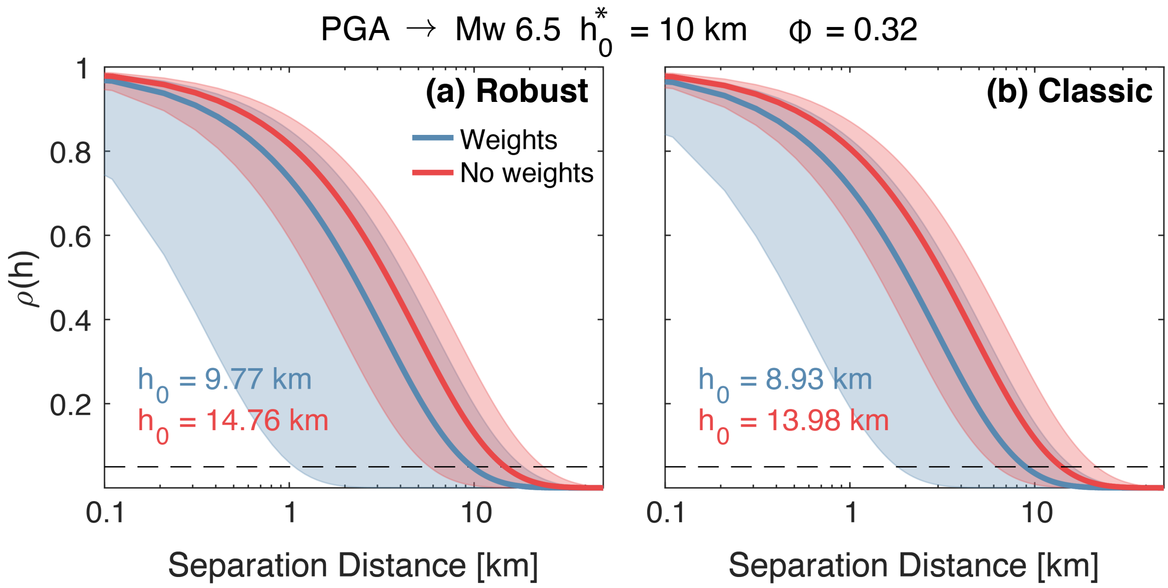


Figure 2.9: Correlation models obtained through the robust (a) and classic (b) estimators. Blue curves indicate the results in which the weights used to fit the experimental data are computed based on the number of pairs and separation distance; red curves represent the results in which equal weights are applied to all the data. Shadow area represent the standard deviation of the results computed on 1000 simulations. The standard deviation of the model with weights (8.7 and 7.12 km for the robust and classic estimators respectively) is smaller than the standard deviation of the model without weights (9.02 and 7.73 km).

2.5.3 Dependence on magnitude

The relationship between magnitude and correlation length is one of the most debated topics in studies of spatial correlation. Some studies (e.g. Sokolov et al., 2012; Sokolov and Wenzel, 2013; Foulser-Piggott and Goda, 2015) argue that the range tends to increase with increasing magnitude because moderate-to-large earthquakes feature lower frequency content and hence an additional non-random component. Conversely, other authors (e.g. Jayaram and Baker, 2009; Huang, Tarbali and Galasso, 2020) did not find any correlation between these two parameters. Heresi and Miranda (2019) performed a statistical analysis and found that only a small ratio of the variability in terms of correlation length can be explained by considering the magnitude, in particular for SA at longer periods. Garakaninezhad and Bastami (2017) showed that there is a positive correlation between magnitude and anisotropy ratio (ratio between the ranges of the direction with the largest and smallest spatial correlations, respectively) of residuals for the nine earthquakes analysed. This is valid only for PGA, however, and it does not apply to spectral accelerations.

In Figure 2.10 and 2.11 we compare the ranges obtained for each $M_w \geq 4.0$ earthquake belonging to the Central Italy seismic sequence as a function of magnitude. The results do not suggest any clear relationship (especially for shorter-period IMs) between range and magnitude, at least for this M_w interval, in agreement with the findings of Jayaram and Baker (2009). Similarly, while analysing data from 138 Italian events with $M_w \geq 4.0$, Huang, Tarbali and Galasso (2020) did not indentify any trend of the range with respect to the magnitude. It is worth noting that both Sokolov et al. (2012), Sokolov and Wenzel (2013) and Foulser-Piggott and Goda (2015) analysed a wider range of events in terms of magnitude, grouping as moderate-to-large earthquakes all events with $M_w \geq 6.0$. We believe that our analysed database represents too narrow a magnitude interval to draw sound conclusions on the relationship between magnitude and range. Besides, it is biased towards small-magnitude earthquakes, which could influence the result. Other factors should be considered to explain the variability in terms of correlation length, especially when the same seismic region is considered. Stafford et al. (2019) demonstrated that the rupture process of events of equal magnitude has a significant contribution on the variability in the range. Likewise, while studying correlation spatial patterns using physics-based ground-motion simulations, Chen and Baker (2019) found a dependence of the correlation on source effects and on the relative position of a site with respect to the fault asperities.

2.5.4 Dependence on period

Previous research (e.g. Goda and Hong, 2008; Jayaram and Baker, 2009; Esposito and Iervolino, 2012; Wagener et al., 2016; Infantino et al., 2018; Sgobba et al., 2019; Chen and Baker, 2019) has demonstrated that the spatial correlation structure is affected by the response-spectral period considered. Range and period are found to be directly proportional.

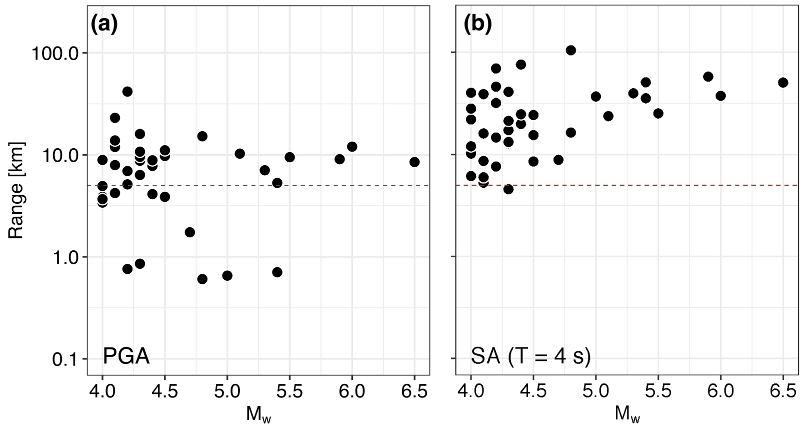


Figure 2.10: Range as a function of magnitude: (a) PGA; (b) SA(4s). Ranges are obtained by means of the robust semivariogram. We also apply a detrending processing using second-order surface models to remove long-range spatial trends. The dashed red line represents the bin size of 5 km. Any range value smaller than the bin size is indicative of non-correlation and should not be taken as an appropriate estimate.

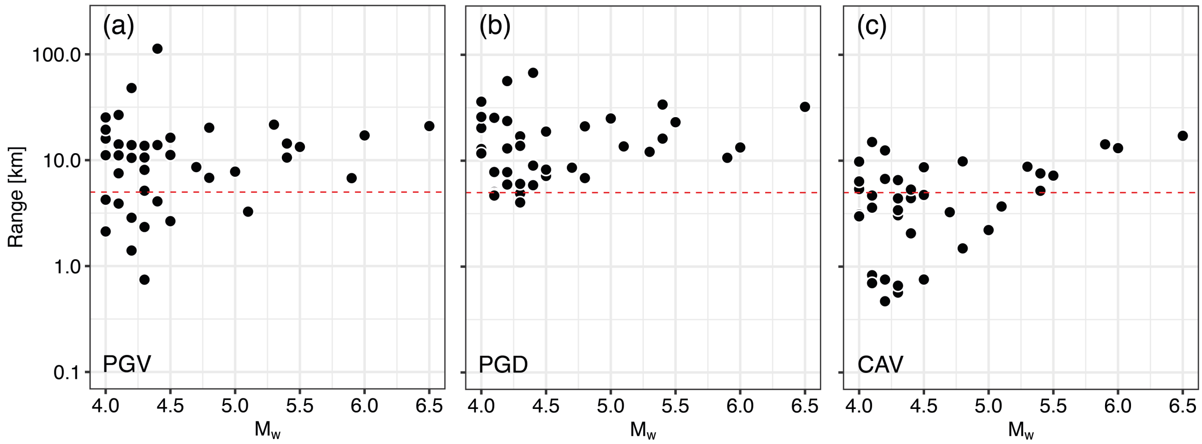


Figure 2.11: Range a as function of magnitude Mw: a) PGV; b) PGD; c) CAV. Ranges are obtained by means of the robust semivariogram. The dashed red line represents the bin size of 5 km. Any range value smaller than the bin size is indicative of non-correlation and should not be taken as an appropriate estimate.

This is consistent with studies on ground-motion coherency, a measure of similarity between waveforms and phases of two ground-motion time histories recorded at two different sites. Hough and Field (1996) demonstrated that ground motions at frequencies up to 2-3 Hz and distances up to 3-4 km are characterized by a high level of waveform coherence, which by contrast drops for shorter period ground motions. Zerva and Zervas (2002), while investigating spatial coherency, proved that small-scale heterogeneities in the travelling path strongly affect the short-period waves (short wavelengths), which therefore tend to be less coherent, in contrast to long-period waves. Figure 2.12 provides an overview of the correlation lengths as a function of period of the above-mentioned studies. All the range values in Figure 2.12 are computed as the distance at which the correlation equals 0.05 in order to compare the results. Esposito and Iervolino (2012) and Wagener et al. (2016) showed that at short periods (up to 0.5 s) it is not possible to clearly delineate a trend between range and period. Goda and Atkinson (2010) found that the correlation is clearly affected by the vibration period especially at short separation distances (h $<$ 10 km), in contrast to greater distances (h $>$ 50 km), where the models tend to converge towards similar results. Consequently, they proposed an average constant model (Figure 2.12), calibrated on the results obtained at nine different periods. To further test the hypothesis of a dependence of the correlation on period, here we present a model for the Central Italy sequence computed by pooling all the data from the nine $M_w \geq 5.0$ events.

As we summarise in Figure 2.13, we observe the increasing trend with period for all $M_w \geq 5.0$ events of the sequence taken individually, as well.

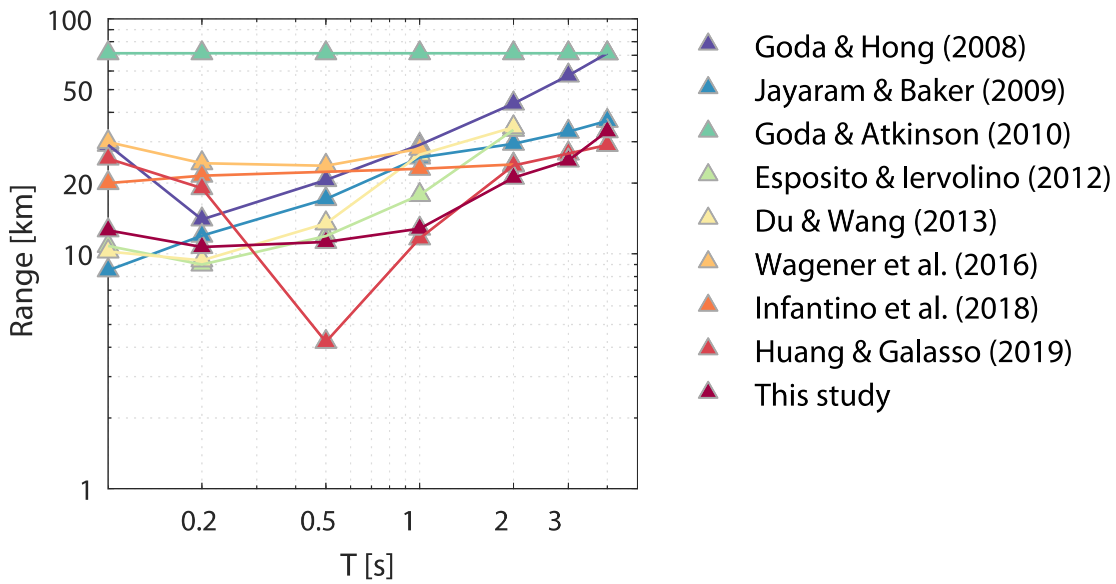


Figure 2.12: Range as a function of response-spectral period, T. The main studies are reported along with a model for the Central Italy seismic sequence, denoted as 'This study '. This model is calibrated by pooling all the data from the nine $M_w \geq 5.0$ events. The model of Jayaram and Baker (2009)) refers to their model for heterogeneous soil condition, whereas the model of Du and Wang (2013) is computed considering a V_{s30} correlation of 4.5 km.

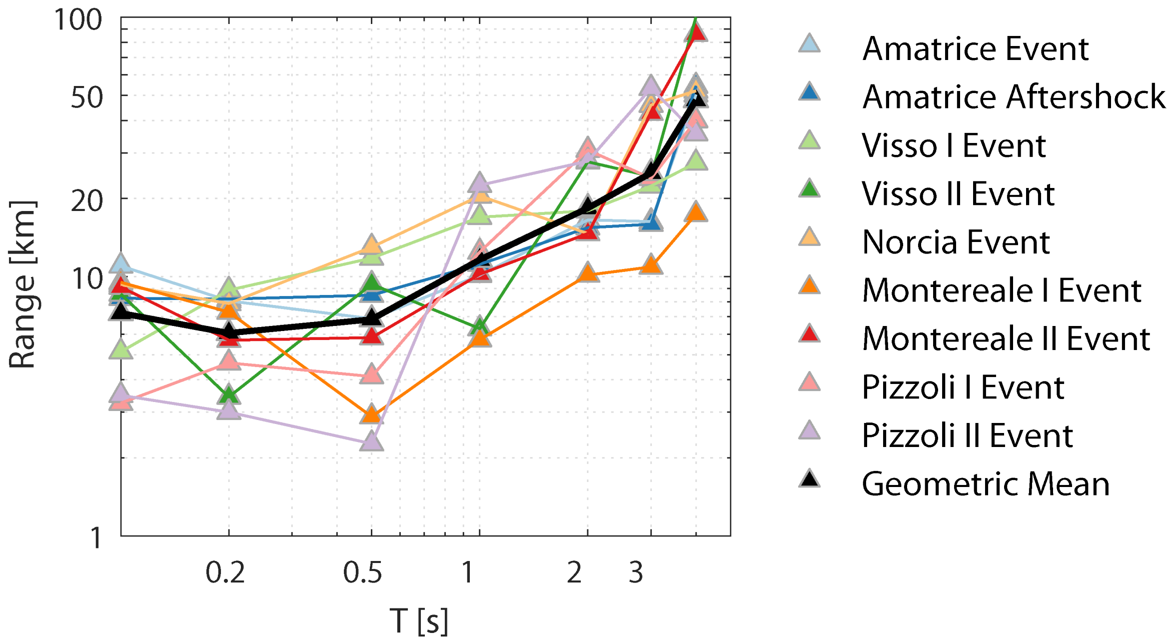


Figure 2.13: Range as a function of response-spectral period, T, computed using the robust estimator for the nine $M_w \geq 5.0$ events of the sequence. In black we reported the weighted (geometric) average model, in which the weights are given based on the number of stations that recorded each event.

2.5.5 Dependence on local site effects and geological structure

Several studies (e.g. Jayaram and Baker, 2009; Sokolov et al., 2010, 2012; Du and Wang, 2013; Infantino et al., 2018; Chen and Baker, 2019) indicated that the level of correlation is strongly influenced by regional geologic and site conditions, especially when short-period IMs are of interest. Conversely, Heresi and Miranda (2019) demonstrated that the clustering of $V_{s,30}$ values does not explain the high variability in the correlation length. To address this issue, we develop an *ad hoc* GMPE for the Central Italy sequence without accounting for different site conditions and then we introduce a site-specific correction factor to remove any dependency on local-site effects, following Sokolov et al. (2010), Sokolov et al. (2012) and Sokolov and Wenzel (2013). The within-event spatial correlation is understood as a non-random component in the residuals, which arises from neglecting a number of factors in the GMPE, such as site, path and azimuth as well as hanging-wall and footwall effects (Sokolov et al., 2012; Sokolov and Wenzel, 2013). Due to the shortage of data, we could not derive a correlation model for each site class, similarly to Sokolov et al. (2012) and Sokolov and Wenzel (2013). The following functional form is assumed:

$$\log_{10} \bar{Y}_{ij} = b_1 + b_2 M_w + (b_3 + \ b_4 M_w) \log_{10} \sqrt{R_{JB}^2 + b_5^2} \tag{2.14}$$

where $\bar{Y}_{ij}$ is the PGA or SA(T) for T equal to 0.2, 0.5, 1, 2, 3 and 4 s and R_{JB} is the Joyner-Boore distance (or the epicentral distance for those events where the fault geometry is not defined and the point-source approximation holds). b_1, b_2, b_3, b_4 and b_5 are the model coefficients inferred through the maximum-likelihood regression method by Joyner and Boore (1993). It is noted that our main goal is to estimate the residuals for correlation purposes, and not to develop a new GMPE; therefore, we keep the functional form as simple as possible. The between-event and within-event residual components are evaluated using the mixed-effects approach. In the same way, we compute the site-specific correction term $\delta S2S$, which represents the site-specific deviations from the average model calibrated considering all sites in the dataset, as explained in Kotha et al. (2018). We evaluate the sample semivariogram through equation (eq. 2.10) and we fit the values using the weighted least-squares approach (section 2.5.2) considering an exponential model (eq. 2.11). We follow the approach of Esposito and Iervolino (2011) and Esposito and Iervolino (2012) to pool data from different events and we repeat these steps including the site-specific correction term, so that the corrected IM is:

$$\log_{10} \bar{Y}_{ij,\text{corr}} = \ \log_{10} \bar{Y}_{ij} + \delta S2S_j \tag{2.15}$$

Figure 2.14 provides an overview of the within-event correlation models obtained considering or not the site-specific correction term. Similarly to Sokolov et al. (2010), Sokolov et al. (2012) and Sokolov and Wenzel (2013), the range decreases if a correction factor is included in the analysis, suggesting that neglecting site conditions might add an artificial correlation due to systematic biases in the residuals (Stafford et al., 2019) . Besides, the lower level of correlation found at longer periods (grey curve) requires further analysis, since the local-site effects influence is expected to decrease with increasing period, as demonstrated by Jayaram and Baker (2009) and Infantino et al. (2018), among others.

Furthermore, the level of correlation of $V_{s,30}$ values can be used as a proxy to represent either homogeneous or heterogeneous local-soil conditions in terms of local-site effects, since a smaller correlation indicates a more varied geological setting. Jayaram and Baker (2009) identified a positive relationship between clustering of V_{s30} values and correlation of spectral accelerations, especially at shorter periods. They proposed a model to predict the range as

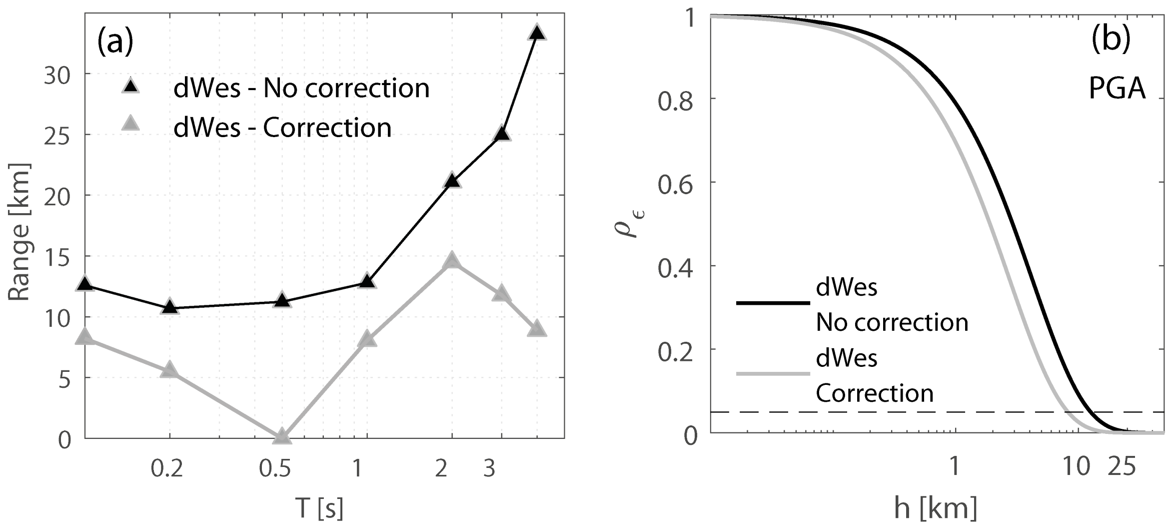


Figure 2.14: (a) Range as a function of the response-spectral period, T, for within-event residuals ε_{ij} (dWes) with and without the site correction factor; (b) Correlation function in terms of PGA for within-event residuals ε_{ij} with and without the site correction factor. The black dashed line indicates a correlation of 0.05.

a function of both the period and the $V_{s,30}$ range. Similarly, Du and Wang (2013) drew analogous conclusions, illustrating that the influence of local-site effects on the correlation of spectral accelerations drops with increasing period. Sokolov et al. (2012) provided a relationship between the PGA range and the $V_{s,30}$ correlation, after finding a dependency of these parameters for different regions in Taiwan. Garakaninezhad and Bastami (2017) found that the anisotropy ratio of residuals tends to increase as the anisotropy ratio of $V_{s,30}$ increases. However, this positive correlation is significant only for PGA, similarly to the relationship with magnitude. To further emphasis this aspect, Figure 2.15 compares the above-mentioned studies along with a model calibrated on the $M_w \geq 5.0$ events of the Central Italy sequence. We develop the latter using the code by Ming et al. (2019), which was also used in Huang and Galasso (2019) to study the correlation observed in Italian data. Two main observations are worthy of remark. Firstly, the effects of local-soil conditions on spatial correlation models are much greater at shorter periods, compared to longer periods where the different models practically overlap. Secondly, the comparison between the Huang and Galasso (2019) and Central Italy models suggests that the correlation strongly depends on the geological characteristics of the considered area, so that a single universal correlation model based on large datasets is not suitable to represent small regions. Indeed, the model of Huang and Galasso (2019) is obtained by including events that occurred in all of the Italian mainland, thus leading to an average spatial correlation model calibrated on diverse geological contexts, with variable spatial dependency of IMs. The same was also previously demonstrated by Sokolov et al. (2010, 2012) and Sokolov and Wenzel (2013) for Taiwan. Besides, this behaviour is particularly evident at shorter periods, in agreement with different studies, such as Jayaram and Baker (2009) and Du and Wang (2013), in which these authors demonstrated the impact of local-site effects as a function of period. Figure 2.12 illustrates this aspect: the Central Italy and the Huang and Galasso (2019) models nearly coincide for T>1 s.

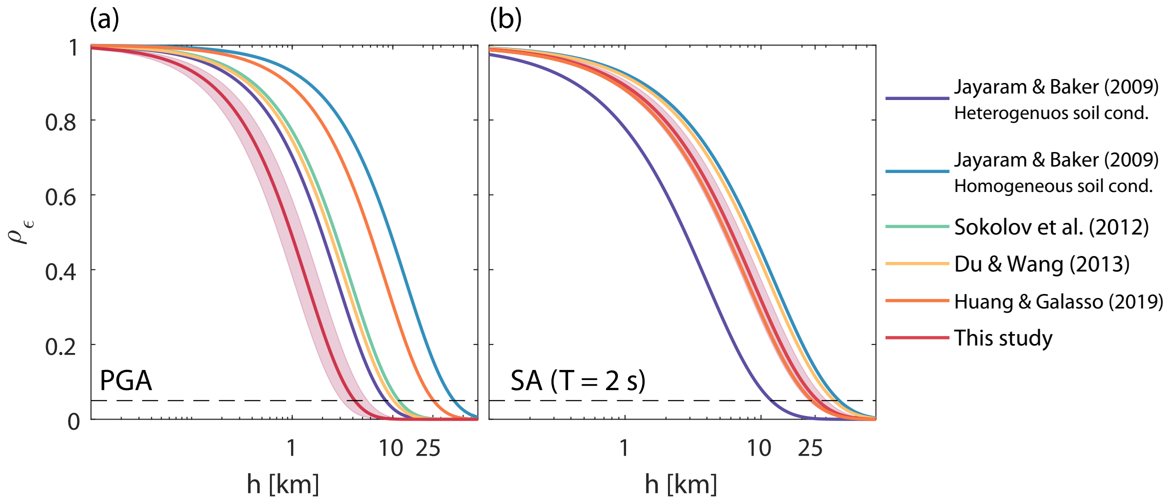


Figure 2.15: Correlation function in terms of PGA (a) and SA at T = 2 s (b). The model referred as 'This study'is computed by means of the code developed by Ming et al. (2019) using data from the nine $M_w \geq 5.0$ central Italy earthquakes. The red shadow area represents the 95% confidence intervals of the central Italy model. The black dashed line indicates a correlation of 0.05.

2.5.6 Dependence on GMPEs

To develop a correlation model either an existing GMPE or an *ad hoc* relation is used to assess the within-event residuals. Authors, such as Wang and Takada (2005), Jayaram and Baker (2009), Esposito and Iervolino (2011, 2012) and Wagener et al. (2016), employed an existing ground motion model in their analyses. Conversely, other authors such as Goda and Hong (2008), Goda and Atkinson (2009, 2010), Sokolov et al. (2010) and Infantino et al. (2018) developed a GMPE calibrated on their specific dataset. Du and Wang (2013) asserted that the chosen GMPE does not affect the correlation results. Analogously, Jayaram and Baker (2009) and Goda (2011) demonstrated that similar outcomes are achieved if different GMPEs are used. In contrast, Sokolov et al. (2010) maintained that the decomposition of the total aleatory variability into between- and within-event components is influenced by the particular GMPE and hence the spatial correlation is affected in turn, as it is calibrated on the residuals. Infantino et al. (2018) built a ground-motion model based on their databases to avoid any dependency on the GMPE. Likewise, we opt for the *ad hoc* GMPE approach and we develop a study-specific ground-motion model, as described in section 2.5.5. We also calibrate *ad hoc* models for each $M_w \geq 5.0$ event of the Central Italy sequence to obtain a specific correlation model for each earthquake. We assume the following functional form:

$$\log_{10} \bar{Y}_{ij} = b_1 + b_2 \log_{10} \sqrt{R_{JB}^2 + b_4^2} + b_3 \sqrt{R_{JB}^2 + b_4^2} \tag{2.16}$$

where b_1, b_2, b_3 and b_4 are the model coefficients inferred through a one-stage ordinary regression, which is justified as we are only using data from a single event each time. We show some of the results in Figure 2.12 and Figure 2.13.

To investigate the dependence of spatial correlation on the chosen ground-motion model, we repeat all the analyses using the GMPE provided by Bindi et al. (2011), which is also calibrated on Italian data. This model was found by Michelini et al. (2020) to be the most suitable to predict ground motions in shallow active crustal regions in Italy. Figure 2.16 presents a comparison between the correlation models obtained implementing both the *ad hoc* and Bindi et al. (2011) GMPEs. We plot the results illustrated in section 2.5.5, in which we consider both uncorrected and corrected within-event residuals. Interestingly, the Bindi et al. (2011) model provides nearly the same range as the model obtained including a site correction term for short-period IMs. At longer periods, the Bindi et al. (2011) model approaches the range value inferred without incorporating any site correction factor. We interpret these outcomes by also considering the results that we present in section 2.5.5. The Bindi et al. (2011) GMPE takes into account the local-site effects through the definition of coefficients based on the EC8 soil categories. Local-soil conditions most influence the spatial correlation of short-period IMs and thus, it is not just chance that the correlation model inferred through the Bindi et al. (2011) GMPE tends to the range obtained implementing a site-specific correction term. Conversely, the influence of site conditions at longer periods is weaker, and the Bindi et al. (2011) and the uncorrected correlation models nearly coincide.

In this regard, we also calibrate four different correlation models based on the same dataset, which includes 50 $M_w \geq 3.7$ events that occured in the Central Italy region (the majority of data belongs to both the 2009 L'Aquila and the 2016 Central Italy sequences). The models differ only for the reference GMPEs used to compute the within-event residuals, so that it is possible to investigate the effect of the GMPE on the spatial correlation.

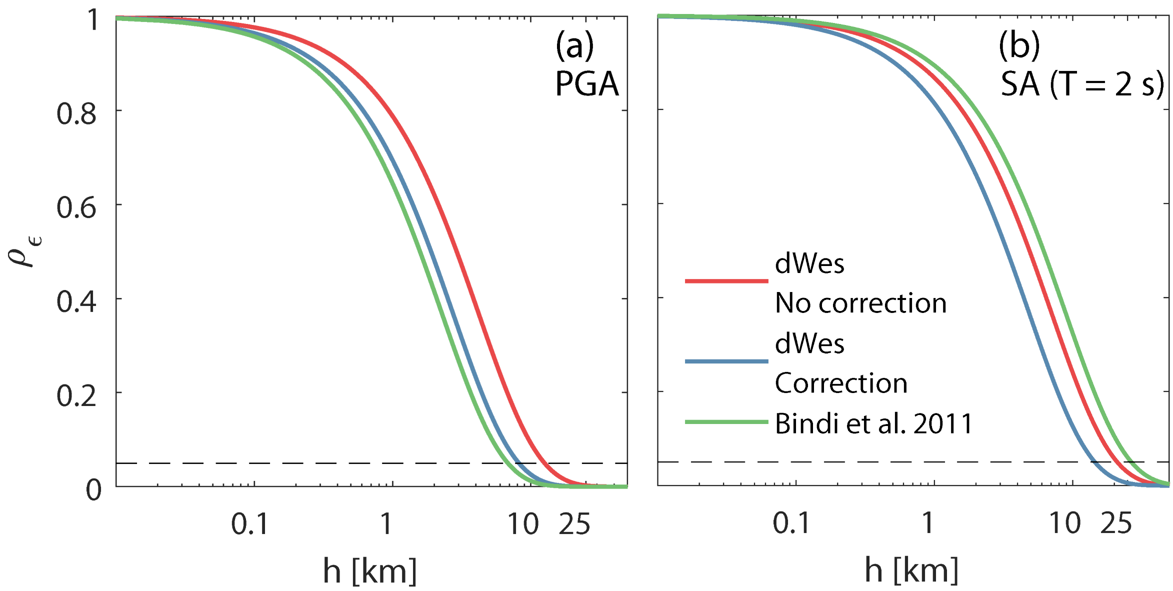


Figure 2.16: Within-event ε_{ij} (dWes) correlation models for uncorrected and corrected residuals and computed considering the GMPE by Bindi et al. (2011): (a) PGA; (b) spectral acceleration for T equals 2 s.

We develop an *ad hoc* GMPE and we choose the Lanzano, Luzi, Pacor, Felicetta, Puglia, Sgobba and D'Amico (2019) (hereinafter ITA18) and the Kotha et al. (2020) (hereinafter K20) models, which mainly differ for the working datasets. In particular, we provide different correlation properties for the K20 GMPE: (1) range values obtained by using the ergodic K20 GMPE, and (2) range values obtained by using the region-specific K20 GMPE. At intermediate and long periods, the four models converge towards very similar values, suggesting that the underlying GMPE has negligible effects on the final correlation structure. By contrast, at shorter periods ($T \leq 0.4s$), the models show a different trend. An explanation may lie with the different sensitivity of high-frequency and low-frequency ground motions to the anelastic attenuation. Indeed, Kotha et al. (2020) found larger regional differences of anelastic attenuation at short periods than at longer periods. The larger range values of the K20erg model are likely to be attributable to the faster attenuation of the Central Italy region with respect to the pan-European average (Kotha et al., 2020), which is not modelled in the ergodic K20 GMPE and therefore it manifests as apparent spatial correlation, in contrast to the region-specific K20 GMPE. The ITA18 GMPE shows a trend similar to the region-specific K20 model, demonstrating that any suitable GMPE may be used in correlation analysis with no significant differences in the final outcomes. The *ad hoc* model based on the specific dataset may better capture the attenuation of high-frequency IMs, resulting in a lower spatial correlation at short periods.

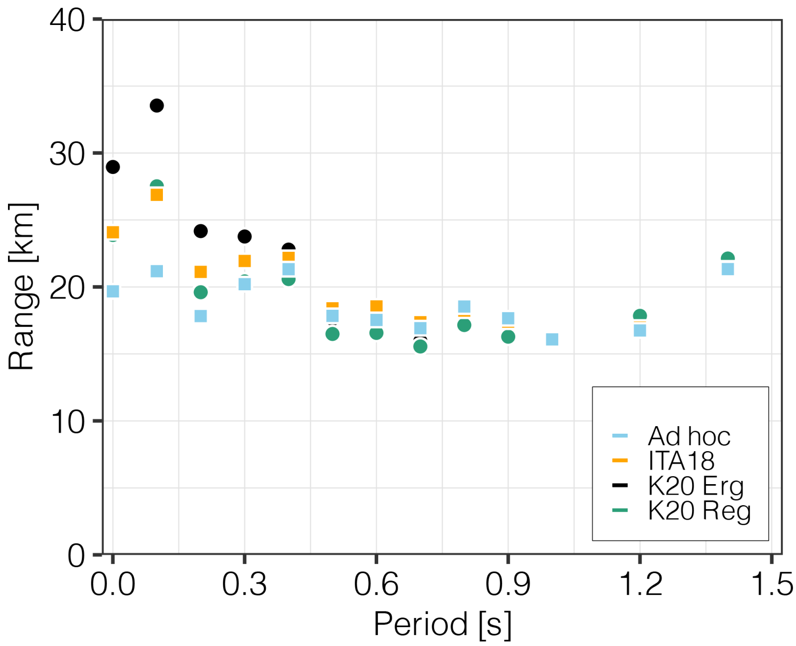


Figure 2.17: Spatial correlation values as a function of period obtained for the Central Italy region. Ad hoc refers to the GMPE calibrated on the dataset; ITA18 refers to the Lanzano et al. (2019) GMPE; K20 Erg refers to the ergodic Kotha et al. (2020) GMPE; K20 Reg refers to the region-specific Kotha et al. (2020) GMPE.

Overall, we believe that any GMPE can be used, provided that the selected ground-motion model fits the data well, so that the residuals do not exhibit biases. Systematic deviations from the predicted median might be mapped as an additional artificial correlation, which inevitably would affect the seismic hazard and risk assessments of spatial distributed systems. This behaviour is shown in Figure 2.18, in which we compare the semivariograms obtained considering both the GMPE by Bindi et al. (2011) and Boore and Atkinson (2008*b*). Clearly, the Boore and Atkinson (2008*b*) semivariogram diverges from the Bindi et al. (2011) with increasing separation distance. This might be due to the trend of the within-event residuals with distance found for the Boore and Atkinson (2008*b*) model, which reflects the inability of the GMPE to properly capture the observed IMs.

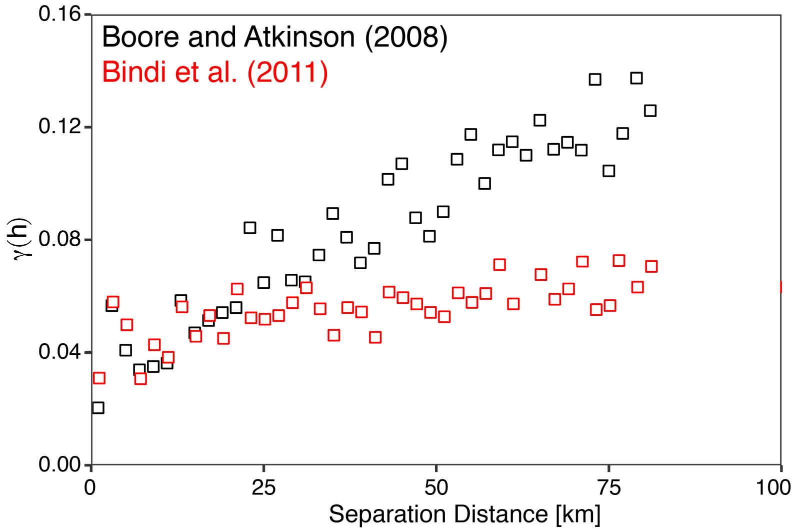


Figure 2.18: Sample semivariogram obtained using the GMPEs by Boore and Atkinson (2008*b*) and Bindi et al. (2011).

2.6 Spatial correlation of different residual components

Spatial correlation models are usually calibrated on the within-event component of residuals, obtained based on ergodic GMPEs. The ergodic assumption implies that the distribution of ground motions over time at given site is the same as their spatial distribution over all site (Lanzano et al., 2017). In the last decade many efforts have been made in order to relax the ergodic assumption, thereby leading towards a non-ergodic approach, in which the systematic and repeatable characteristics of the ground motion (e.g. site-specific and path-specific component) are specified to adjust the median prediction of a GMPE and hence significantly reduce the aleatory uncertainties (e.g. Kotha et al., 2020). Despite the importance of such approaches in case of site-specific PSHA, the assessment of spatial correlation under the non-ergodic assumption has not been widely studied yet. The major contributions in this context are by Kuehn and Abrahamson (2020) and Sgobba et al. (2019, 2021).

In this study, we decide to analyse the spatial correlation not only of the within-event term, which includes spatial variation of both local-site and travel-path effects, but also of the site-to-site term and event- and site-corrected residuals, estimated partially relaxing the ergodic assumption. We therefore derive two GMPEs with identical functional forms, but with different decomposition of residuals: (1) between-event and within-event residuals and, (2) between-event, between-site ($\delta S2S$) and event-and-site corrected residuals

(δWS_{ij}), so that $\varepsilon_{ij} = \delta S2S_i + \delta WS_{ij}$. The second model allows partially relaxing the ergodic assumption through the definition of the site-to-site term ($\delta S2S$), which includes all the local-site specific effects (Lanzano et al., 2017; Kotha et al., 2017), and thus further investigating the factors that determine the spatial dependency of IMs. We use RotD50 values (that is the median horizontal ground-motion across all nonredundant azimuths) of PGA, PGV, PGD, CAV, I_a and 5%-damped SA at 15 periods of vibration between 0.1 and 5 s; hence, the equation is:

$$\log_{10} Y_{ij} = b_1 + b_2 M + b_3 M^2 + (b_4 + b_5 M) \log_{10} \sqrt{R_{JB}^2 + b_7^2} + b_6 \sqrt{R_{JB}^2 + b_7^2} + \eta_j + \varepsilon_{ij} \tag{2.17}$$

where M is either the moment magnitude or the local magnitude and R_{JB} is the Joyner-Boore distance (i.e. the closest distance to the surface projection of the rupture). b_1 ... b_7 are the model coefficients inferred through a non-linear mixed-effect regression approach, computed using the NLMER algorithm of Bates et al. (2014) implemented in R Core Team (2019). The advantage of using such approach is twofold. Namely, it allows: (1) quantifying the between- and within-event components and further partitioning the within-event group into systematic and non-systematic components; and (2) obtaining unbiased regression for each group (event or station) which has a different number of ground motion records (Lee, 2009). It is recalled that Eq. 2.17 does not involve any site-response parameter, hence $\delta S2S$ absorbs all site-specific features and it can be considered as a proxy of the amplification function of each station (Kotha et al., 2018). Conversely, δWS_{ij} is the remaining aleatory variability, computed as the different between the within-event and site-to-site term, and it mostly accounts for path effects, being the events nucleated within the same source zone.

A visual inspection of both GMPEs residuals (Figure 2.19) suggests that the adopted functional form (Eq. 2.17) is performing well, as no significant trends with respect to the predictor variables are detected. It is noted that, for sake of brevity, we show only the plot pertinent to the partially ergodic GMPE. However, we obtain similar results for the ergodic GMPE, where η_j and ε_{ij} are plotted against M_w and R_{JB}, respectively.

Figure 2.20 compares the semivariograms obtained for the different residual components of PGV, before and after the detrending processing. Indeed, ε_{ij} and δWS_{ij} do not clearly comply with the hypothesis of second-order stationarity, as the semivariogram tends to increase without reaching a stable plateau. On the contrary, $\delta S2S$ does not reveal any spatial trend and thus we do not apply any detrending correction. Similar outcomes were found by Sgobba et al. (2021), in which the authors model the spatial correlation under the hypothesis of stationarity for the site term and non-stationarity for the path

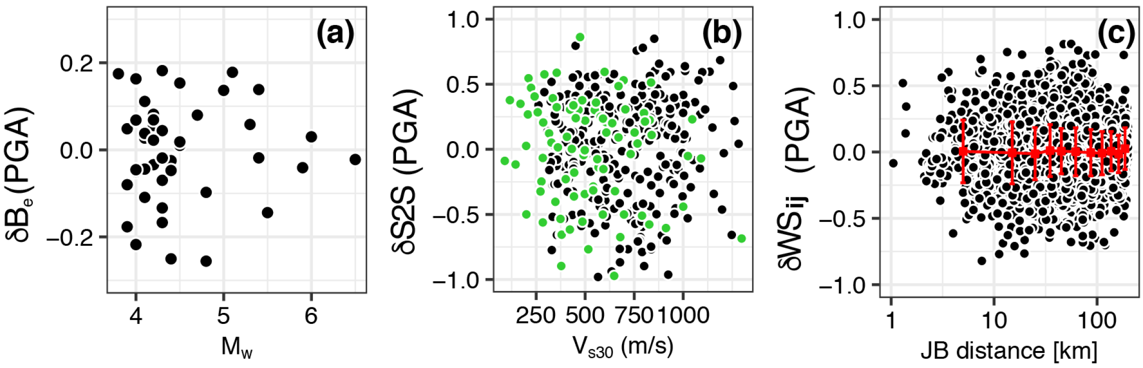


Figure 2.19: Residual plots to check if residuals show any systematic trend with predictor variables: a) η_j (δB_e) against Mw; b) $\delta S2S$ against V_{s30}; c) δWS_{ij} against R_{JB}. Green dots in b) highlight stations with a measured V_{s30} value, whereas black dots stations for which the V_{s30} is inferred from the slope, following Wald and Allen (2007).

term, while analysing data from Central Italy. Kuehn and Abrahamson (2020) employed a non-stationary covariance function, which includes a dependency on both the inter-site distance and source-to-site distance, to investigate the correlation structure of the path term. According to Kuehn and Abrahamson (2020), fitting both the spatial dependency of IMs from near and distant stations is not appropriate; indeed, closely spaced sites located near the source may show different travel paths due to small-scale heterogeneities in the rupture process, whereas seismic waves to distant sites will be along almost identical travel paths. Evidences of non-stationarity were also found by Chen et al. (2021), who proposed novel methodologies to quantify non-stationary spatial variations in strong ground motion, taking advantages of well-recorded earthquakes in New Zealand.

Figure 2.21 shows the spatial correlation results for different IMs by considering within-event and event- and site-corrected residuals as well as the site-to-site term. In agreement with Sgobba et al. (2021), the latter term does not show any correlation, with a very small range for all investigated IMs. On the contrary, the correlation distance of ε_{ij} tends to increase with increasing period, as previously observed in the literature. The spatial correlation structure is found to be affected by the response-spectral period considered: in particular, range and period are directly proportioned. δWS_{ij} shows a very similar behaviour to ε_{ij}, despite lower correlation values at longer periods. This suggests that spatial dependency of IMs is not due only to site-specific features, but also to travel-path effects and other unexplained effects, not fully captured by our GMPE. A comparison of these residual components for CAV and I_a was also investigated by Foulser-Piggott and Goda (2015), who concluded that the two terms have in general different correlation structures, and thus this variability should be accounted for in seismic hazard and risk analysis.

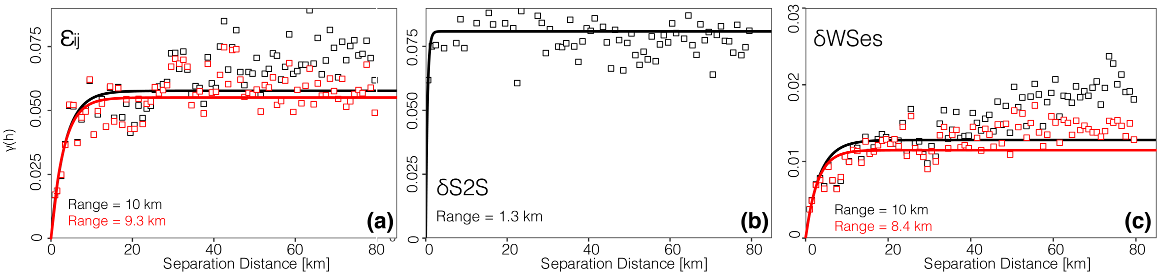


Figure 2.20: Semivariograms of different residual components for PGV obtained pooling all the events considered in the analysis: a) Within-event residuals; b) Site-to-site term; c) Event- and site-corrected residuals. Black squares represent the original data, whereas red dots refer to the detrended data.

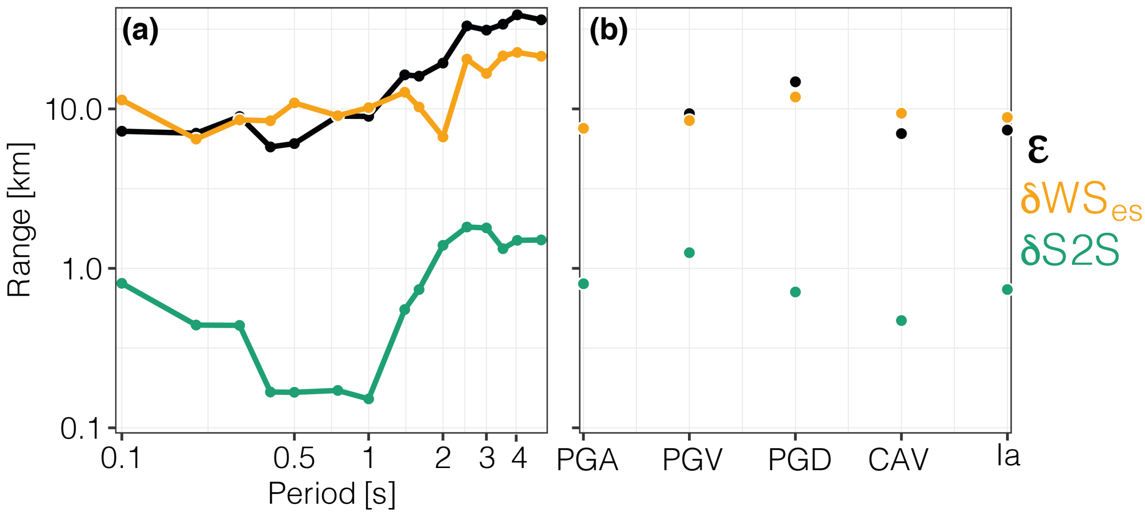


Figure 2.21: Range of within-event, site-to-site and event- and site-corrected term for different IMs: a) Spectral acceleration at period between 0.1 and 5 s; b) PGA, PGV, PGD, CAV and I_a.

2.7 Spatial correlation of PGV, PGD and integral ground-motion intensity measures

In addition to PGA and SA at different periods to define the ground-motion induced by earthquakes, other IMs turn out to be suitable to characterise the resulting damage to structures and predict ground failure: peak ground velocity (PGV) and peak ground displacement (PGD) as well as Arias intensity (I_a) and cumulative absolute velocity (CAV), to name but a few. A proper definition of the seismic action in terms of spectral displacement ordinates has progressively gained importance in performance-based seismic design (Paolucci et al., 2008), and I_a and CAV have been found to be adequate for many other earthquake engineering applications, such as evaluating the susceptibility to liquefaction and earthquake-induced landslides as well as predicting the structural damage (Foulser-Piggott and Goda, 2015; Du and Wang, 2013; Bullock, 2019; Huang, Tarbali and Galasso, 2020).

I_a is defined as the integral of the square of the acceleration time history $a(t)$ over the entire duration adjusted by a constant factor (Kramer et al., 1996):

$$I_a = \frac{\pi}{2g} \int_0^{t_{max}} a(t)^2 dt \tag{2.18}$$

where g is the acceleration of gravity. CAV is the area under the absolute acceleration time history:

$$CAV = \int_o^{t_{max}} |a\,(t)|\, dt \tag{2.19}$$

Differently from the other IMs, which are primarily related to either the amplitude or the frequency content of the ground motion, I_a and CAV implicitly reflect multiple characteristics of the time histories, including the cumulative effects of the ground motion duration (Du and Wang, 2013; Costanzo, 2018). As a matter of fact, they turn out to be more efficient to represent the cumulative potential damage due to the ground shaking (Foulser-Piggott and Goda, 2015; Huang, Tarbali and Galasso, 2020).

In Figure 2.21b, there is a clear trend of increasing of the range value from PGA to PGD. The PGD is indeed related to lower frequencies of the ground-motion, which are generally found to be correlated over longer distances, as opposed to higher frequency IMs. This result is in agreement with those presented in section 2.5.4.

In Figure 2.22, we compare our correlation models for PGV, CAV and I_a with some of the studies reported in literature. In particular, we select the models of: (1) Esposito and Iervolino (2012), based on Italian data; (2) Sokolov and Wenzel (2013), calibrated on Japanese events with $4.2 \leq M_w \leq 7.2$; (3) Wang and Takada (2005), which selected earthquakes recorded in Japan and Taiwan with $6.2 \leq M_w \leq 8$; (4) Du and Wang (2013), which developed a prediction equation for the range based on the correlation structure of $V_{s,30}$ values, using data from events occurred in California and Japan; (5) Bullock (2019), based on 172 events occured in New Zealand; (6) Huang, Tarbali and Galasso (2020), calibrated on Italian events with $M_w \geq 4.0$. Clearly, the models calibrated on different regions show a much larger correlation distance than our results. This suggests that regional and local site effects play a first-order role in defining the correlation structure. The range estimated by Esposito and Iervolino (2012) is slightly larger than that found for Central Italy. A possible explanation can lie with the extended database considered in Esposito and Iervolino (2012), which includes events that occurred in all of the Italian mainland. Similar outcomes were also found for spectral accelerations in Section 2.5.5, suggesting that a unique correlation model based on a large database is not appropriate to represent small regions. The Bullock (2019) model for CAV features a larger correlation with respect to the other studies. This behaviour is likely to be related to the different types of earthquakes considered; indeed, Bullock (2019) included not only shallow crustal, but also slab and interface events. Besides, while the I_a models show comparable correlation values among the different studies (as also observed in Huang, Tarbali and Galasso (2020)), CAV presents a much higher variability across the regions, with the Huang, Tarbali and Galasso (2020) having the lowest range. As stated in Huang, Tarbali and Galasso (2020), this result is likely a consequence of the one-stage algorithm (Ming et al., 2019) used in their study in contrast to the classical two-stage approach (as followed in this thesis). Finally, ranges estimated for the Amatrice and Norcia earthquakes are compared to those obtained by Costanzo (2018). CAV and I_a show very similar outcomes, in agreement also with Foulser-Piggott and Goda (2015). However, our results differ to those by Costanzo (2018) and the main causes can be attributable to the different estimation process.

2.8 Conclusions

In this chapter, we aimed to provide insights into the spatial correlations of earthquake ground-motion intensity measures. A number of studies, such as Park et al. (2007), Goda and Atkinson (2009), Esposito and Iervolino (2011) and Verros et al. (2017), highlighted the importance of considering such spatial correlations in seismic hazard and risk analysis,

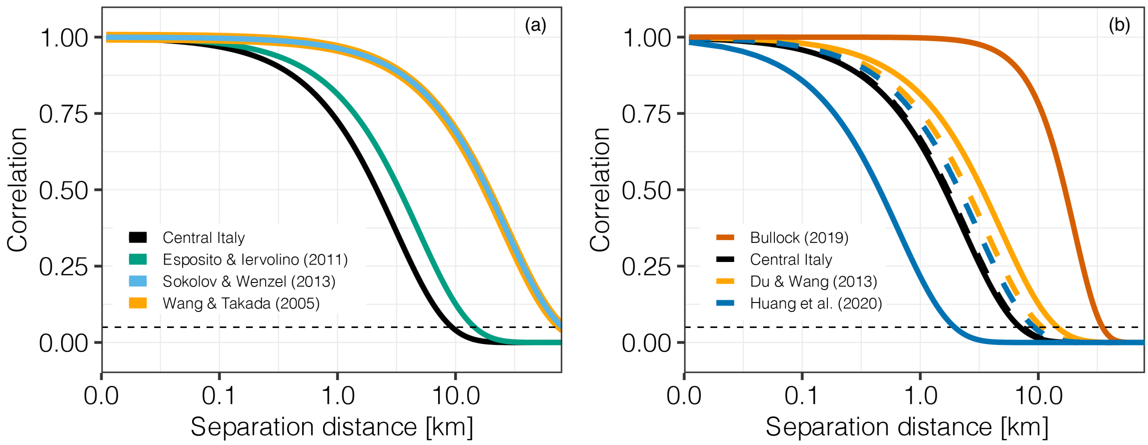


Figure 2.22: Comparison among different correlation models: a) PGV; b) CAV (solid lines) and Ia (dashed lines). Black dashed line points out a level of correlation equal to 0.05.

especially when the seismic risk of spatially distributed systems has to be evaluated. Nevertheless, the correlation models proposed over the past two decades feature significant discrepancies, which may lead to underestimation or overestimation of the assessed seismic risk. Therefore, in this article, we critically summarised the main findings of previous studies and we attempt to address the primary questions about ground-motion spatial correlation.

We showed the two main approaches to estimate the spatial correlation of ground-motion intensity measures. In principle, the implementation of either one or the other technique is expected to provide similar results, as the semivariogram and the covariance are equivalent for second-order stationary random fields. Nevertheless, in agreement with Goda and Hong (2008) and Goda and Atkinson (2009), we find that the results slightly diverge. Concerning the computation of the semivariogram, the robust estimator (Cressie) is understood to be less sensitive to atypical observations, as demonstrated by Oliver and Webster (2014). Therefore, we preferred using this estimator in contrast to the classic one throughout the analyses, although the results from our simulations did not find any significant difference between the two estimators. To fit experimental data, we advise prioritising those techniques that weigh the shorter separation distances more highly, since these distances have a strong impact on seismic hazard and risk assessments of spatially-distributed systems. At the same time, we discourage adopting a manual fitting due to the high level of subjectivity involved.

Similarly to Jayaram and Baker (2009) and Huang, Tarbali and Galasso (2020), we do not find any relationship between magnitude and correlation distance. We are aware that the analysed dataset is limited in terms of magnitude range (Mw 4.0 to 6.5), so that firm conclusions cannot be drawn on this aspect. However, we believe that magnitude itself cannot explain the large variability of correlation distance values, especially when the same region is considered. Thus, other source effects, such as directivity, azimuth and hanging wall and footwall effects, should be accounted for, as outlined in Stafford et al. (2019) and Chen and Baker (2019).

We found a positive correlation between the range and response-spectral period, as expected from the literature. Small-scale heterogeneities in the travel path tend to affect less the long-period waves, which turn out to be more correlated, compared to high-frequency waves.

We analysed the dependency of the spatial correlation on local-soil conditions, illustrating that the influence of local-site effects and regional geology is period-dependent, as demonstrated by several authors, such as Jayaram and Baker (2009), Du and Wang (2013) and Infantino et al. (2018). The comparison between our analysis of data from Central Italy and the work proposed by Huang and Galasso (2019) is significant. Firstly, it confirms the hypothesis that the dependency of spatial correlation on site conditions in terms of local-site effects decreases as the period increases. Secondly, it suggests that a single rate of decay of the correlation as a function of the inter-site separation distance is not suitable for seismic hazard and risk assessment, as the range appears to be regionally-dependent, even though all the data came from the same country (Italy in this case). We believe that region-specific spatial correlation models should be derived.

Furthermore, our results suggest that any suitable ground-motion model can be applied in correlation studies, provided that the residuals do not exhibit significant trends or biases, which might result in an apparent larger correlation. Therefore, a single correlation model could be used in seismic hazard analysis even if many different GMPEs are employed to estimate ground-motion parameters. However, as suggested by Sokolov et al. (2012), implementing a logic tree for correlation models as already incorporated in probabilistic seismic hazard assessments, should be encouraged to account for the high uncertainty in spatial correlation models.

In agreement with Kuehn and Abrahamson (2020), Sgobba et al. (2021) and Chen et al. (2021), we found that event- and site-corrected residuals, which are mostly influenced by path effects, do not comply with the hypothesis of stationarity. We got around the problem by modelling spatial trends through trend-surface models, calibrated on geographic coordinates.

However, the spatial variation of the mean function may be better calibrated using more appropriate parameters of the site and this aspect requires further analysis to draw more firm conclusions. Alternatively, the approach proposed by Kuehn and Abrahamson (2020), which directly includes the computation of a non-stationary covariance function, could be used. We also found that the analysed residual components have different degrees of spatial correlation, which may be related to the varying underlying physical processes, as suggested by Foulser-Piggott and Goda (2015). Therefore, seismic hazard and risk analysis should account for this variability.

Finally, we note that generally the models are poorly constrained at short separation distances (less than 2 km), due to a shortage of observations (Goda and Atkinson, 2010; Wagener et al., 2016). Results of ground motion simulations could provide better constraints to further advance insights into the spatial correlation of ground motions.

Data and resources

The full dataset of strong-motion waveforms of the nine $M_w \geq 4.0$ earthquakes of the 2016-2017 Central Italy sequences is available at the Engineering Strong Motion database (ESM, *http://esm.mi.ingv.it*, last accessed September 2019) and the Italian ACcelerometric Archive (ITACA, *http://itaca.mi.ingv.it*, last accessed September 2019). The full dataset of recordings of the $M_w \geq 2.5$ events of the 2016-2017 Central Italy sequences is available at *http://cnt.rm.ingv.it/* (last accessed September 2019).

Acknowledgements

We thank the research group of Dr Pacor at INGV (Istituto di Geofisica e Vulcanologia) for providing the dataset related to the $M_w \geq 4.0$ events that occurred during the Central Italy earthquake sequence. We also thank Dr Kotha and Dr Stafford for their helpful suggestions about implementing mixed-effects regressions in R. We thank Dr Ming for help with the MATLAB program to implement the approach of Ming et al. (2019). We wish to thank Professor Baker's research group at Stanford University for providing valuable feedback on our analyses. We are grateful to Impact Forecasting for fruitful discussions on hazard footprints. We thank Graeme Weatherill and an anonymous reviewer for their constructive advice on improving the manuscript. Finally, we thank the University of Strathclyde for funding the PhD during which the work was undertaken.

CHAPTER 3

Spatial correlation of broadband earthquake ground motion in Norcia (Central Italy) from physics-based simulations

Despite many researchers have aimed to identify the factors that most affect the spatial dependency of earthquake ground motions over the last 20 years, there are still open-questions to be solved, as criticised in chapter 2. The main causes lie with the shortage of a large number of closely-spaced earthquake recordings and the possibility of having data from controlled environments. In this context, physics-based ground motion simulations are likely to provide an alternative way to further advance insights into the spatial correlation of ground motions. As a matter of facts, this chapter investigates the spatial correlation of response spectral accelerations from a set of broadband physics-based ground motion simulations generated for the Norcia (Central Italy) area by means of the SPEED (SPectral Elements in Elastodynamics with Discontinuous Galerkin) software. We produce several ground-motion scenarios by varying either the slip distribution or the hypocentral location as well as the magnitude to systematically explore the impact of such physical parameters on spatial correlations. We extend our analysis to other ground-motion components (vertical, fault-parallel, fault-normal) in addition to the more classic geometric mean to highlight possible ground-motion directionality and therefore identify specific spatial correlation features. Our analyses provide useful insights on the role of slip heterogeneities as well as the relative position between hypocentre and slip asperities on the spatial correlation. Indeed, we found a significant variability in terms of both range and sill among the considered case studies, suggesting that the spatial correlation is not only period-dependent, but also scenario-dependent. Finally, our results reveal that the isotropy assumption may represent an oversimplification especially in the near-field and thus it may be unsuitable for assessing the seismic risk of spatially-distributed infrastructures and portfolios of buildings.

This chapter is based on the following published article:

1. **Schiappapietra**, E., & Smerzini, C. (2021). Spatial correlation of broadband earthquake ground motion in Norcia (Central Italy) from physics-based simulations. *Bulleting of Earthquake Engineering. https://doi.org/10.1007/s10518-021-01160-7*

3.1 Introduction

Standard tools for Probabilistic Seismic Hazard Assessment (PSHA) estimate the ground motion due to an earthquake by means of Ground Motion Models (GMMs), which provide the median and the associated aleatory variability of ground motion Intensity Measures (IMs) as a function of source, path and site parameters for a given site. As GMMs consider IMs at all sites as independent variables, they do not provide a description of the correlation structure of ground motion IMs at multiple sites. The quantification of the spatial correlation properties of ground motions is of fundamental importance for estimating risks to spatially-distributed infrastructures and portfolios of buildings. For instance, Iervolino (2013), Sokolov and Ismail-Zadeh (2016) and Sokolov and Wenzel (2019) showed the differences between site-specific and multiple-site PSHA, underlining the importance of considering regional hazard for the calculation of aggregate risk.

Despite the progressive increase of studies on spatial correlation in earthquake ground motions over the last 20 years, debates on the factors affecting spatial correlation still continue. In chapter 2, we provide a thorough literature review, critically exploring the main facets of spatial correlation that need to be addressed. Among these aspects, we report dependence on: (1) the estimation approach; (2) the fitting method; (3) the earthquake magnitude (Mw); (4) the vibration period; (5) local site-effects; and (6) ground motion prediction models. In this regard, a valuable contribution can be also found in Baker and Chen (2020), who recently propose an alternative approach to assess the uncertainty in spatial correlation models in terms of both true and estimation variability. These authors suggest that differences in terms of true variability in correlation among well-recorded and poorly-recorded events are negligible. On the contrary, the estimation uncertainty is inversely correlated to the number of available stations, as demonstrated also in chapter 2 and by Infantino et al. (2021). Moreover, spatial correlation models are usually grounded on the hypothesis of stationarity and isotropy. Namely, the correlation between any pairs of sites with equal separation distance is the same, independently of the source-to-site distance and orientation (Diggle et al., 2007; Goovaerts et al., 1997). However, different studies (e.g. Chen and Baker, 2019; Huang, Tarbali, Galasso and Paolucci, 2020; Infantino et al., 2021) demonstrate the presence of anisotropy in spatial correlation, which tends to be stronger for long-period IMs. Garakaninezhad and Bastami (2017) examined the effects of focal mechanism, $V_{s,30}$ (time-averaged shear-wave velocity in the upper 30 m), magnitude and hypocentral distance on the anisotropy ratio and angle, and reported a positive correlation between the anisotropy ratio of high frequency IMs and that of $V_{s,30}$ as well as with Mw. Similar conclusions were also drawn by Abbasnejadfard et al. (2020). While analysing correlation properties of integral IMs from the New Zealand strong motion database, Bullock (2019) found ground motions correlated over longer distances in the fault-parallel direction due to directivity effects and that such anisotropic properties were more evident in normal and reverse faulting mechanism.

Our knowledge on the factors mostly affecting ground motion spatial correlations is partially hampered by both the shortage of ground motion recordings and the possibility of having sufficient data from specific environments. In this framework, 3D Physics-Based ground motion Simulations (3D PBSs) could be employed to provide further insights into earthquake ground motion spatial correlations. 3D PBSs provide spatially-distributed estimates of region-specific ground motions, as they are based on sufficiently detailed models of the seismic source, propagation path and near-surface geology. 3D PBSs enable issues due to the lack of observations to be overcome, especially in the near-source region. They also allow investigation of the impact of factors such as local-site conditions, magnitude, slip distributions and hypocentral locations, on the spatial correlation in a detailed way that is not feasible with earthquake recordings. Infantino et al. (2021) studied the spatial correlations of broadband 3D PBSs for seven different case-studies (Po plain, L'Aquila, Marsica, Sulmona, Norcia - Italy; Thessaloniki - Greece; Istanbul - Turkey) and found a strong variability in terms of correlation lengths due to a strong dependence of the ground motion spatial variability on both local-site and rupture-propagation effects. Stafford et al. (2019) investigated the impact of source kinematics on spatial correlation using finite difference simulations, demonstrating that the rupture process of scenarios of equal magnitude contribute markedly to the correlation length variability. Likewise, Chen and

Baker (2019), taking advantage of the CyberShake database, reported supporting evidence of the impact of source and path effects on spatial correlation.

In this context, we use 3D broadband PBSs to address some of the still open questions on earthquake ground motion spatial correlation. We choose the 3D PBS of the 30^{th} October M_w 6.5 Norcia (Central Italy) earthquake as a case study (Özcebe et al., 2019). The authors generated the simulations through the spectral element code SPEED (*http://speed.mox.polimi.it*, Mazzieri et al. (2013)), which replenishes the frequency spectrum at the high frequencies by using a novel method that makes use of Artificial Neural Networks (Paolucci et al., 2018). The numerical code and its ability to reproduce ground motion spatial correlations have already been validated in different studies (e.g Stupazzini et al., 2009; Paolucci et al., 2016, 2015, 2018; Infantino et al., 2021). In addition to the 3D PBS of the M_w 6.5 Norcia event, we generate several scenarios by varying either the slip distribution or the hypocentral location to explore different spatial correlation properties, such as the dependency on the rupture process and magnitude. While relaxing the hypothesis of isotropy, we explore possible preferential directions of the ground motion spatial variability as well as the relationship of such directions with the source mechanism and frequency content of the ground motion.

Furthermore, spatial correlation models are usually calibrated on the horizontal component of the ground motion and little effort has been directed towards other components, such as the fault-normal (FN) and the fault-parallel (FP) directions as well as the vertical component (Garakaninezhad and Bastami, 2017; Infantino et al., 2021). In the near-field, the ground motion is strongly affected by both the slip distribution and fault mechanism, which can cause a significant polarization of the ground motion (Bray and Rodriguez-Marek, 2004). Somerville et al. (1997) demonstrated that the ratios between the FN and FP components of the motion are larger than 1 in the proximity of the fault due to rupture directivity effects. Consequently, we extend our studies to the vertical (Z), FN and FP components, in addition to the horizontal one (defined as the geometric mean of the two horizontal components, geoH), to identify diagnostic spatial correlation features for each component.

3.2 3D Physics-based ground motion simulations

In recent years, major advances in parallel high-performance computational resources have allowed the development of high-order numerical methods to model the seismic response of 3D heterogeneous media under realistic tectonic and geomorphological conditions. 3D PBSs embody physical and sufficiently accurate models of the seismic source, path and local-site effects to provide region-specific earthquake ground motion predictions. These estimates are, however, reliable only in the low-frequency spectrum (up to 1-1.5 Hz) due to the current limitations in: (1) computational capabilities and (2) detailed knowledge of the geological medium that would enable modelling the propagation of short wavelengths. In this study, we use the spectral element code SPEED to generate spatially distributed broadband ground motions for different scenarios. SPEED is designed to solve 3D seismic wave propagation problems in complex heterogeneous media. Its main advantage lies on its capability to generate realistic broadband ground motion waveforms by means of a novel approach based on Artificial Neural Networks (Paolucci et al., 2018).

3.2.1 3D Numerical modelling

3D numerical simulations require three main inputs: (1) a crustal velocity model, (2) a geotechnical model of local shallow geological structures (e.g. the Norcia basin), and (3) a kinematic slip distribution along the causative fault. The numerical model of the Norcia area (Figure 3.1) was produced in a previous project and the reader should refer to Özcebe et al. (2019) for further details. We note here that the geological model used in the simulations is rather simplified owing to the lack of a 3D crustal velocity model of the area under study. Nonetheless, the results of the 3D PBS for the Norcia case were compared against the available recordings and geodetic measurements to test the goodness of fit of the simulated data. While using the same crustal velocity model and geotechnical setting throughout the analysis, we implement different kinematic source models to allow for the magnitude, slip distribution and hypocentre location to vary among the considered scenarios. The full set of source parameters for the different case studies are given in section 3.2.3. The final spectral element computational domain of the Norcia area, which includes the source and velocity models as well as the surface topography (Figure 3.1a), extends over an area of $40 \times 50 \times 21$ km^3 and consists of more than 350,000 hexahedral elements of 3^{rd} order. Such a configuration allows propagating seismic waves with a maximum reliable frequency of 1.5 Hz.

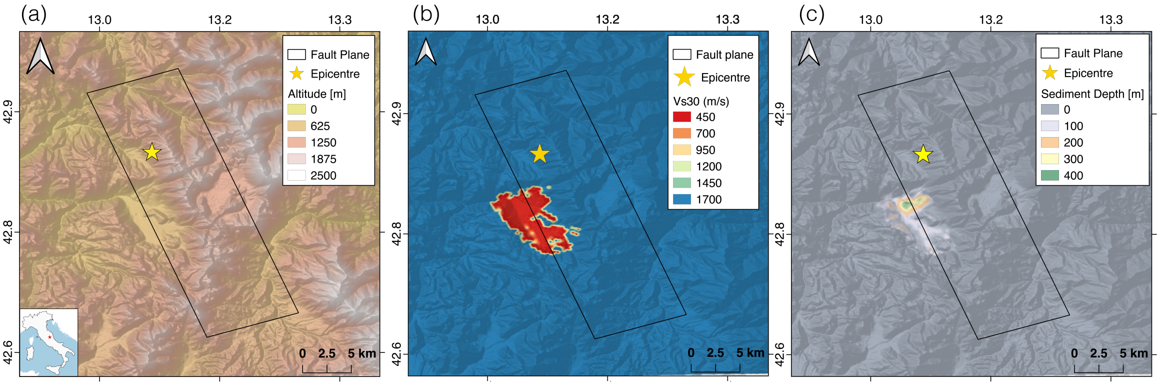


Figure 3.1: Study area: (a) topography model; (b) shear-wave velocity model; (c) sediment thickness model. Yellow stars indicate the epicentre (Mw 6.5 Norcia event), whereas the black rectangles define the surface fault projection.

3.2.2 ANN2BB: Broadband ground motions

Broadband ground motions are generated by combining the results of long-period PBSs with predictions of an Artificial Neural Network (ANN) trained on set of strong motion records. Such an approach, referred to as ANN2BB, has three main steps. First, an ANN is trained on a dataset of ground motion recordings (in our case SIMBAD v6.0, which includes also the M_w 6.5 Norcia observations, updated since Smerzini et al. (2014)) to predict the short-period (T$\leq T^*$) spectral ordinates based on the long-period ones ($T > T^*$); T^* indicates the threshold period below which ground motions from PBSs are not accurate. Second, the trained ANN is applied to estimate short-period response spectral ordinates for periods below T^*. Third, a broadband waveform is constructed by combing the long-period PBS with a linearly scaled stochastic signal, such that the corresponding response spectrum approaches the ANN target spectrum at short periods. One of the main advantages of the ANN2BB procedure, compared to standard hybrid approaches, is its ability to simulate the specificity of the ground motion of the study area, particularly at shorter periods, and to preserve the ground motion spatial correlation properties at high frequencies. This is made possible by the correlation between short- and long-periods, as retrieved from the training dataset of earthquake recordings, on which this approach is based. The reader should refer to Paolucci et al. (2018) for further details on the ANN2BB methodology and its validation tests and to Infantino et al. (2021) for systematic analyses that demonstrate the capability of this approach to portray spatial correlation features.

3.2.3 Case-studies

In the present study, we investigate the 30^{th} October 2016 M_w 6.5 Norcia earthquake simulation and ten different hypothetical scenarios obtained by varying the magnitude (from M_w 4.0 up to M_w 6.5), the kinematic slip distribution and the hypocentral location. While all the scenarios (except for the Mw4.0_S001) share the same fault geometry in the numerical model (namely a fault with a dip angle of 40° and a strike of 161°), different fault rupture scenarios are employed. A summary of the different source parameters is given in Table 3.1.

Table 3.1: Summary of the source parameters of the different simulated scenarios. For the Norcia event (first row), the source is taken from the source inversion model proposed by Pizzi et al. (2017), whereas HB94 refers to the kinematic source rupture generator of Herrero and Bernard (1994).

Case study	ID	Mw	Epicentre	Depth (km)	V_{rup} (m/s)	Rise time (s)	Source	Fault rupture width (km)	Fault rupture length (km)
Norcia event	S001	6.5	42.84°N 13.11°E	6.3	1700	Randomized around 0.7	Pizzi et al. (2017)	9	24
Hypothetical	S002	6.5	42.84°N 13.11°E	6.3	1700	Randomized around 0.5	HB94	11	21
Hypothetical	S003	6.5	42.89°N 13.15°E	2.5	1700	Randomized around 0.5	HB94	11	21
Hypothetical	S004	6.5	42.81°N 13.19°E	2.5	1700	Randomized around 0.5	HB94	11	21
Hypothetical	S005	6.5	42.73°N 13.23°E	2.5	1700	Randomized around 0.5	HB94	11	21
Hypothetical	S006	6.5	42.87°N 13.07°E	8.3	1700	Randomized around 0.5	HB94	11	21
Hypothetical	S007	6.5	42.79°N 13.11°E	8.3	1700	Randomized around 0.5	HB94	11	21
Hypothetical	S008	6.5	42.71°N 13.15°E	8.3	1700	Randomized around 0.5	HB94	11	21
Hypothetical	S001	6	42.76°N 13.15°E	6.7	1700	Randomized around 0.5	HB94	9	12
Hypothetical	S001	5.5	42.89°N 13.07°E	7.8	1700	Randomized around 0.5	HB94	6	5
Hypothetical	S001	4	42.84°N 13.11°E	6.3	-	0.4	Point source	-	-

Regarding the Mw6.5_S001 scenario (simulation of the Norcia event), preliminary analyses have been carried out by Özcebe et al. (2019) to test different source slip distributions from available fault inversion studies. We adopt the slip distribution (Figure 3.2a) along with the fault geometry and hypocentral location proposed by Pizzi et al. (2017), as it is found to be the best performing in a comparison between observations and predictions. Other parameters, such as rupture velocity and rise time, are determined through sensitivity analyses to assess the impact of the modelling assumptions on the final outcomes. All the other scenarios (except Mw4.0_S001) assume the kinematic source model proposed by Herrero and Bernard (1994) (referred to as HB94), according to the target magnitude and fault type (Figure 3.2 b,c,d). For each HB94 source model, the rise time is randomized around a mean value of 0.5 s; however, scenarios from S002 to S008 share the same rise time field so that the hypocentral location is the only varying parameters. Figure 3.3 shows the slip distribution along the causative fault together with the hypocentral locations of scenarios S002-S008, whereas Figure 3.4 presents the distribution of the rupture time. Such hypocentral configurations are intended to investigate rupture propagation effects on the ground motion spatial correlation. Finally, the Mw4.0_S001 scenario assumes a double-couple point-source, with a hypocentre located at the same location as scenarios Mw6.5_S001 and Mw6.5_S002. The aim of such a configuration is two-fold: (1) to explore the dependence of the spatial correlation on the magnitude, considering a broad magnitude interval, and (2) to investigate the effect on the spatial correlation of an extended fault compared to a point source.

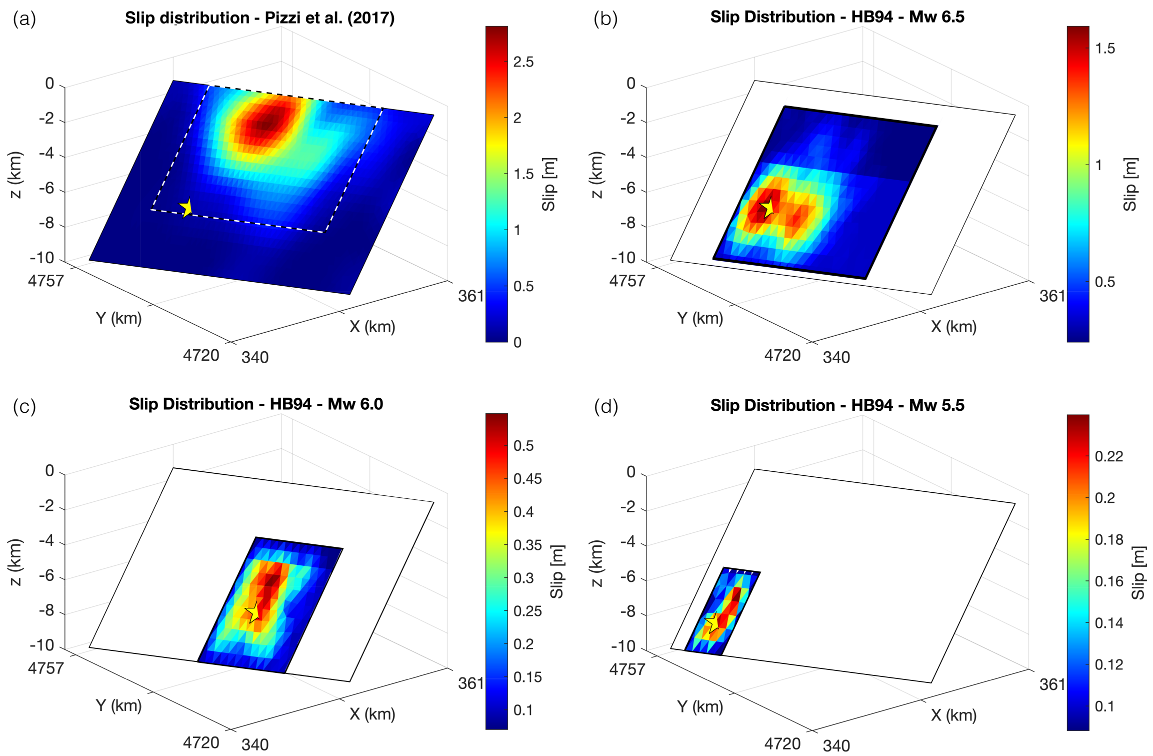


Figure 3.2: Slip distribution along the causative fault. (a) Pizzi et al. (2017); (b) HB94 model for the Mw 6.5 scenario; (c) HB94 model for the Mw 6.0 scenario; (d) HB94 model for the Mw 5.5 scenario. Yellow stars represent the hypocentral locations, whereas the largest rectangles indicate the maximum dimension of the fault ($36km \times\ 13km$). The smallest rectangles represent the ruptured faults, where the slip asperities (regions of very high slip) are concentred.

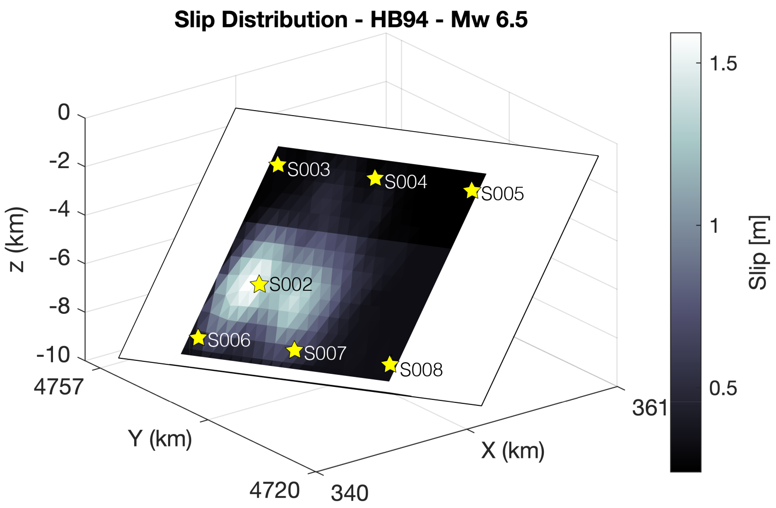


Figure 3.3: Slip distribution along the causative fault (Mw 6.5) along with the hypocentral locations of scenarios S002-S008.

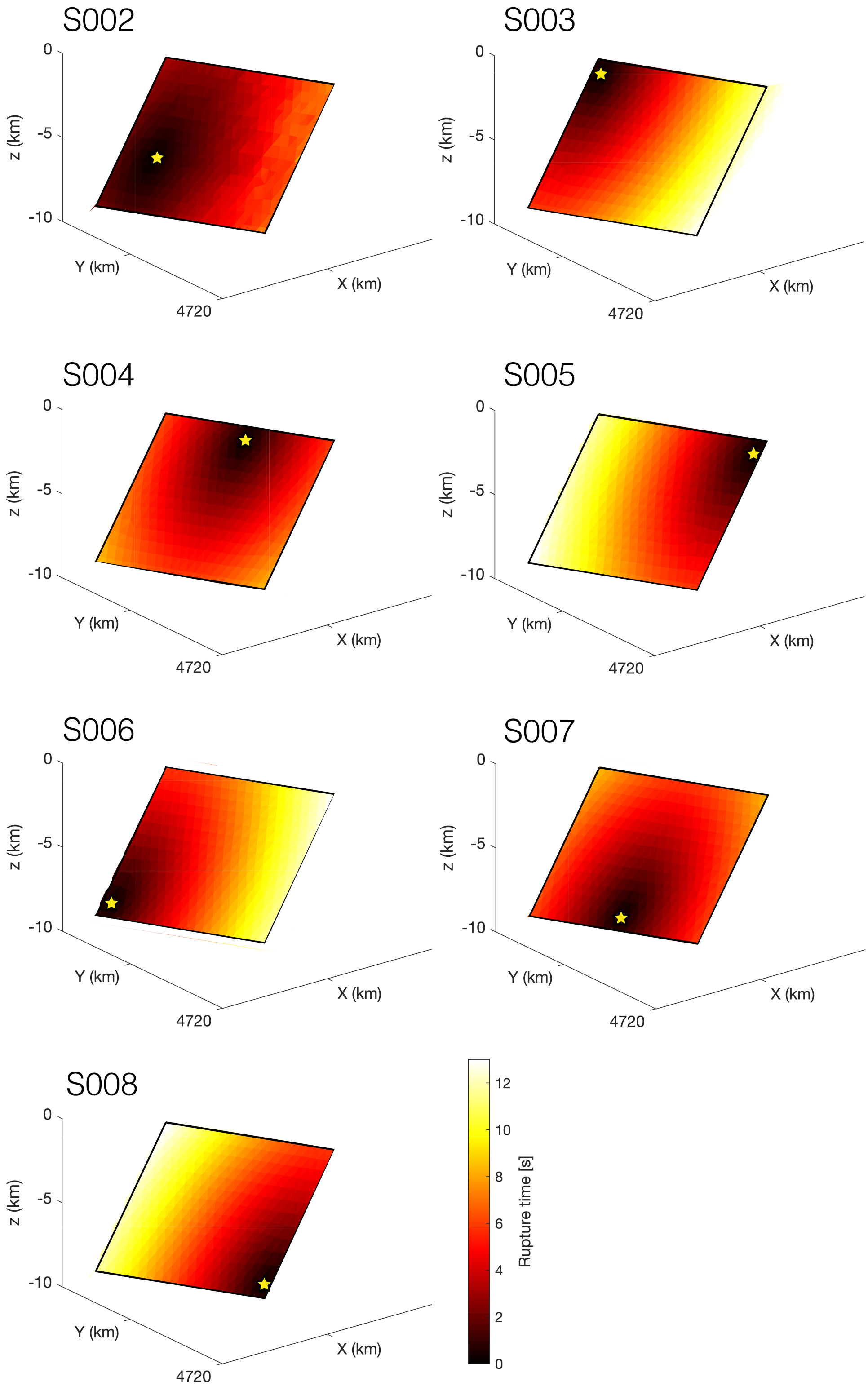


Figure 3.4: Rupture time distribution on the causative fault for scenarios Mw6.5_S002-S008. The scenarios have the same slip distribution, but feature different hypocentral locations. The position of the hypocentre is pointed out by a yellow star.

3.3 Ground motion spatial correlation modelling

In geostatistical analysis, the most common tool adopted to represent the correlation structure between spatially distributed IMs is the experimental semivariogram, which measures the average dissimilarity of a pair of random variables separated by an inter-site distance h. The semivariogram is defined as:

$$\hat{\gamma}(h) = \frac{1}{2}\mathrm{Var}[\varepsilon_{ij} - \varepsilon_{ik}] \tag{3.1}$$

where *Var* indicates the variance and ε_{ij} is the within-event residual representing the misfit between the observed ground motion at an individual site $j(k)$ due to an earthquake and the event-specific average GMM caused by path and local-site effects. In this study, we calibrate data-driven GMMs based on the simulated datasets for each ground motion component and different IMs, such as the peak ground acceleration (PGA) and 5%-damped spectral accelerations (SA) at 16 oscillator periods between 0.1 and 5 s. We perform a sensitivity test on the GMM functional form to both find the most physically representative average model and thus avoid bias. Similar analyses where carried out by Infantino et al. (2021) for the Po Plain (Italy) case-study. These authors demonstrated how the calibration of an appropriate GMM is a crucial step to both obtain robust estimates of the correlation model parameters and avoid the application of further techniques (e.g., the detrending) to get stable semivariograms. We thus adopt the following model:

$$\log_{10} \bar{Y}_{ij} = b_1 - b_2 \log_{10}(R_{line} + b_3) + b_4 \log_{10}\left(\frac{V_{s,30}}{800}\right) + b_5 HW \tag{3.2}$$

where $b_1 \ldots b_5$ are the model coefficients inferred through a one-stage non-linear regression; HW is a dummy variable introduced to specify if the site is located on the hanging wall (HW = 1) or on the footwall (HW = 0) side of the fault; Rline is the closest distance from the surface fault projection of the segment at the top edge of the rupture plan (Hashemi et al., 2015). In contrast to the standard distance metrics used for GMMs, such as the Joyner-Boore and rupture distances, we prefer using the Rline, as it is found to be more efficient (i.e. lower dispersion) in describing the near-source ground motion (e.g. Paolucci et al., 2016). We introduce the HW term to simply account for near-field effects due to both the slip distribution and directivity, which otherwise would require a more complex functional form. We assess the goodness-of-fit of the GMMs through visual inspection of the residuals, computed as the logarithm ($\log_{10}$) misfit between observations and predictions, with respect to the predictor variables and a Quantile-Quantile (QQ) plot of the residuals. An example is shown in Figure 3.5.

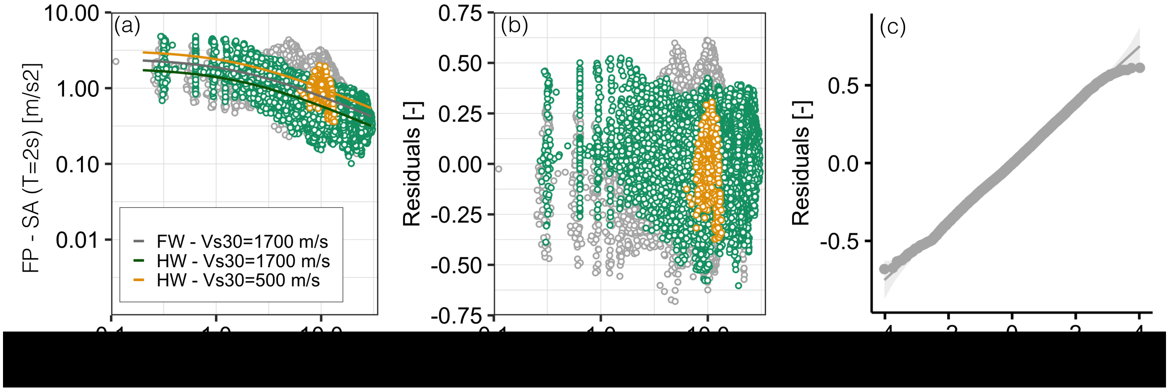


Figure 3.5: Spectral acceleration at T = 2 s of the FP component: (a) Ground motion data and fitted GMMs; (b) residuals as a function of Rline; (c) QQ plot of residuals. Grey dots represent stations on the footwall for rock soil condition and the grey line indicates the corresponding GMM; Green dots represent stations on the hanging wall for rock soil condition and the green line indicates the corresponding GMM; Orange dots represent station on the hanging wall for basin soil condition and the orange line indicates the corresponding GMM. The figure shows the results obtained considering the Mw6.5_S001 scenario (30^{th} October 2016 Norcia event).

To assess the spatial correlation structure on the residual fields, the main steps are: (1) computing the experimental semivariogram (Eq. 3.1) through either the method of moments (Matheron, 1962) or the estimator proposed by Cressie (1985), (2) choosing a parametric function (e.g. exponential and spherical models) to fit the sample semivariogram, and (3) estimating the parameters of the correlation models, namely the sill (i.e. the variance of the data) and the range (i.e. the distance beyond which the spatial dependency between pairs of sites is negligible) through a least-squares regression (Figure 3.6). In the present study, after having shown that both the estimators provide comparable outcomes, we use the classic estimator based on the method of moments to compute the experimental semivariogram:

$$\hat{\gamma}(h) = \frac{1}{2N(h)} \sum_{N(h)} \{\varepsilon_{ij} - \varepsilon_{ik}\}^2 \tag{3.3}$$

where $\hat{\gamma}(h)$ represents the empirical semivariogram and $N(h)$ is the number of pairs of sites with an inter-site spacing given by h. We select a maximum separation distance of 25 km and an inter-site bin width of 1 km, which results in an average value of $N(h)$ of about 2 million in the first 10 bins and thus in well-constrained semivariances. We opt for an exponential function to model the correlation structure, as it is the most widely adopted and best performing functional form in seismology applications. However, we also implement the spherical model for some IMs because both the visual inspection of the fits and the sum of the squared errors reveals a better performance compared to the exponential model (Figure 3.6b).

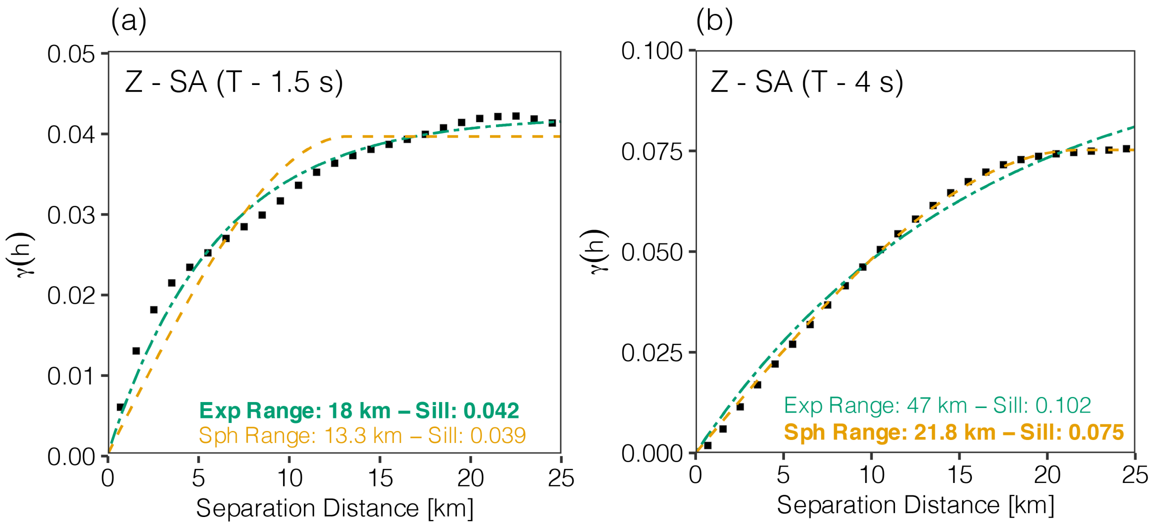


Figure 3.6: Experimental semivariogram and corresponding exponential and spherical models: (a) SA (T = 1.5 s) Z component; (b) SA (T = 4 s) Z component. Black dots represent the experimental semivariogram estimates for each distance bin, whereas dashed lines are the fitted models. The figure shows the results obtained considering the Mw6.5_S001 scenario (30th October 2016 Norcia event). In this example, we choose the exponential model for SA (T = 1.5 s) and the spherical model for SA (T = 4 s).

In most articles, it is assumed that the correlation is isotropic and hence its properties are only a function of distance and not of direction. However, this hypothesis may not always hold, and the pattern of spatial variability may vary with direction. The anisotropy can be either geometric or zonal. In the first case, the directional semivariograms feature the same sill, but different range depending on the relative orientation between pairs of sites; in the latter case, the sill varies with direction (e.g. Goovaerts et al., 1997; Oliver and Webster, 2014). Either directional semivariograms or variogram maps can be used to investigate anisotropic properties of the spatial correlation, as shown in Figure 3.7. Directional semivariograms are computed for each pair of sites whose inter-site spacing falls not only in a specified distance range but also in a precise azimuth bin. We investigate common directions, such as NS and EW, and fix an azimuth tolerance of only 10°. This was made possible by the wealth of data obtained from the simulations. A variogram map (see Figure 3.7b) displays the semivariance contribution of each pair considering simultaneously all distance and azimuth bins. Concentric contour plots imply an isotropic variation, whereas patchier distributions reflect anisotropic patterns in correlation. Any cross section of the map corresponds to the more familiar one-dimensional semivariogram. Alternatively, nonparametric tests (e.g. Garakaninezhad and Bastami, 2017; Huang, Tarbali, Galasso and Paolucci, 2020) can be employed to quantitatively assess anisotropy.

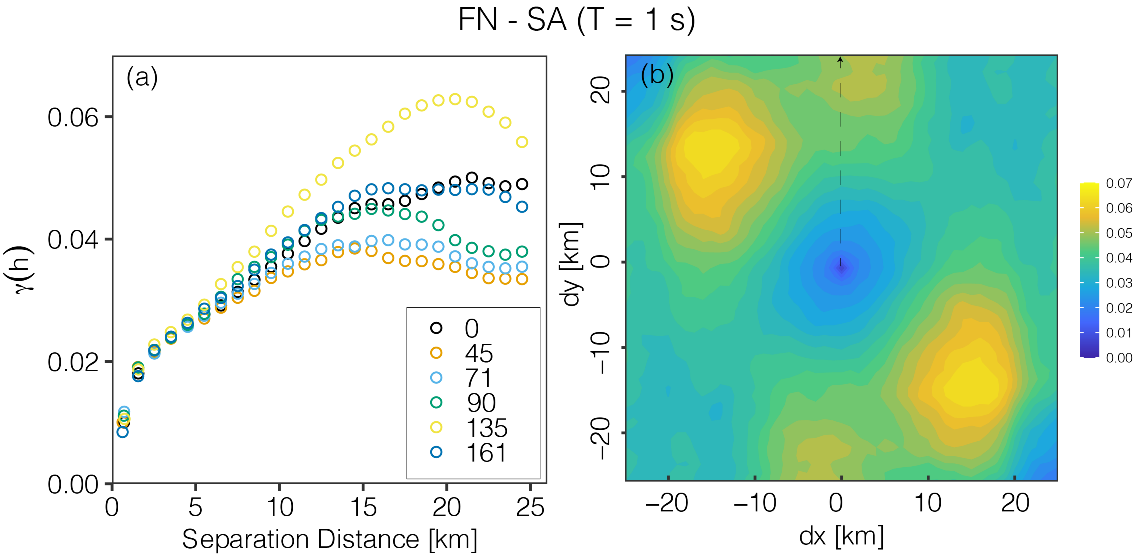


Figure 3.7: (a) directional semivariograms computed for different directions: 0 - NS, 45 - NE-SW, 71 - FN, 90 - EW, 135 - NW-SE, 161 - FP; (b) Variogram map. The black dashed arrow indicates the NS direction. The figure shows the results obtained considering the Mw6.5_S001 scenario (30^{th} October 2016 Norcia event).

Furthermore, the hypothesis of stationarity may not be valid due to specific source and path effects, as suggested by Chen and Baker (2019) and Infantino et al. (2021). In such situations, the semivariogram tends to increase indefinitely without levelling off at a constant sill value and the random variables (e.g. within-event residuals) exhibit a gradual variation in space (Figure 3.8a) that masks potential small-range correlation structure. Trend surface models are commonly adopted to model spatial trends and thus remove large-scale correlation patterns, leading to a patchier distribution of the random variable (Figure 3.8b) (Oliver and Webster, 2014). We implement such an approach in all those cases where the GMM is not able to sufficiently describe the ground motion median behaviour and capture specific features, such as rupture propagation and path effects. Alternatively, nonparametric isotropic GMMs calibrated as a function of few explanatory variables by using other regression techniques, such as LOESS (locally estimated scatterplot smoothing), could be used to potentially avoid this problem.

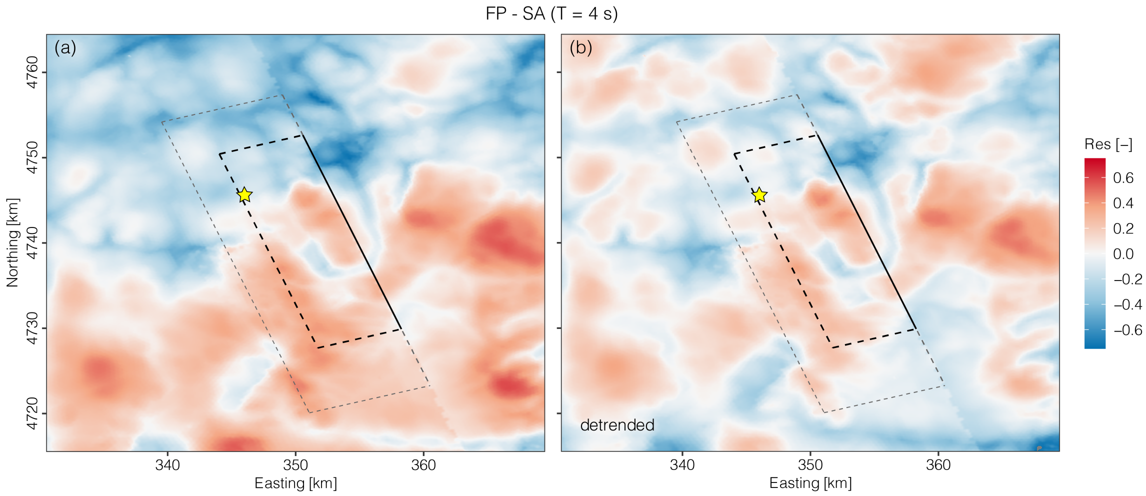


Figure 3.8: Spatial variation of within-event residuals of SA (FP) at T = 4 s: (a) Raw data showing a gradual variation of the residual values; (b) detrended data featuring a patchier distribution. Yellow star indicates the epicentre (Mw 6.5 Norcia event), whereas the grey and black rectangles are the surface fault projection and the effective surface fault projection, respectively. The figure shows the results obtained considering the Mw6.5_S001 scenario (30_{th} October 2016 Norcia event).

3.4 Results

3.4.1 Mw 6.5 Norcia earthquake simulation

3.4.1.1 Comparison with recordings

The 30^{th} October 2016 M_w 6.5 Norcia earthquake was recorded by more than 100 stations, belonging to both permanent and temporary seismic networks, within an epicentral distance of 200 km (Luzi et al., 2019). The uniqueness of this dataset lies on the availability of 36 near-source recordings at epicentral distances less than 30 km (Figure 3.9d). We compute experimental semivariograms using the data from both simulations and observations and compare them to assess the capability of 3D PBSs and ANN2BB to generate spatially correlated ground motion fields accounting for the actual correlation structure. Recordings and station-metadata are from the Engineering Strong Motion database (ESM, *https://esm-db.eu/*). Infantino et al. (2021) perform similar comparisons for the Po Plain

(Northern Italy) case study, finding a good agreement between sample semivariograms obtained from observations and simulations over a broad range of periods. As shown in Figure 3.9 (a,b,c), the Norcia case-study does not show such satisfactory agreement. We propose three possible reasons for this. First, the numerical model does not describe all the small-scale irregularities of the geo-tectonic structure, making the ground motion prone to be less variable. Indeed, the geological model used in the simulations is rather simple, due to the lack of available 3D crustal velocity models of the area (in contrast to the Po Plain case study). Second, 34 out of 36 location selected in the numerical model have a constant V_{s30} equal to 1700 m/s, whereas the ESM stations feature different V_{s30} with an average of about 650 m/s (Figure 3.9d); most of the simulated data are thus located on homogeneous hard rock that undeniably makes the ground motion more correlated. Third, source models are retrieved through finite-fault inversions and are usually calibrated on very low frequency ranges (e.g. Pizzi et al., 2017). As a consequence, source models lack that intrinsic complexity that is inherent in the real world and, in turn, the resulting ground motion features a less variability, affecting both the sill and range. Nonetheless, we note that our results show a realistic trend of the semivariogram, as the semivariances increase with increasing separation distances until they level off at a constant sill value. Therefore, such comparison supports the ability of 3D PBS along with the ANN2BB approach to reproduce the ground motion spatial correlation structure, as already corroborated by systematic analyses reported in Infantino et al. (2021).

3.4.1.2 Ground motion directionality

In the present study, in addition to the geoH horizontal component, we also present results for other components to identify diagnostic spatial correlation features. Figure 3.10 illustrates the range and the sill as a function of the oscillator period for the different ground motion components. First, we find that the spatial correlation structure in terms of the range is generally proportional to the considered period, as also observed in chapter 2. This applies to the FN and Z components, whereas the trend is less evident for the FP and geoH ones. Besides, our results show that the correlation between range and period is not well defined at shorter periods (T $<$ 0.5s), as found in other studies, such as Esposito and Iervolino (2012) and Wagener et al. (2016). Figure 3.10b suggests that the sill is similarly affected by the vibration period. This result is noteworthy, as it is common practice to work with normalized residuals and therefore the sill is not estimated, being equal to 1. Similar outcomes are provided in Sgobba et al. (2019, 2021), who estimated the range and sill of

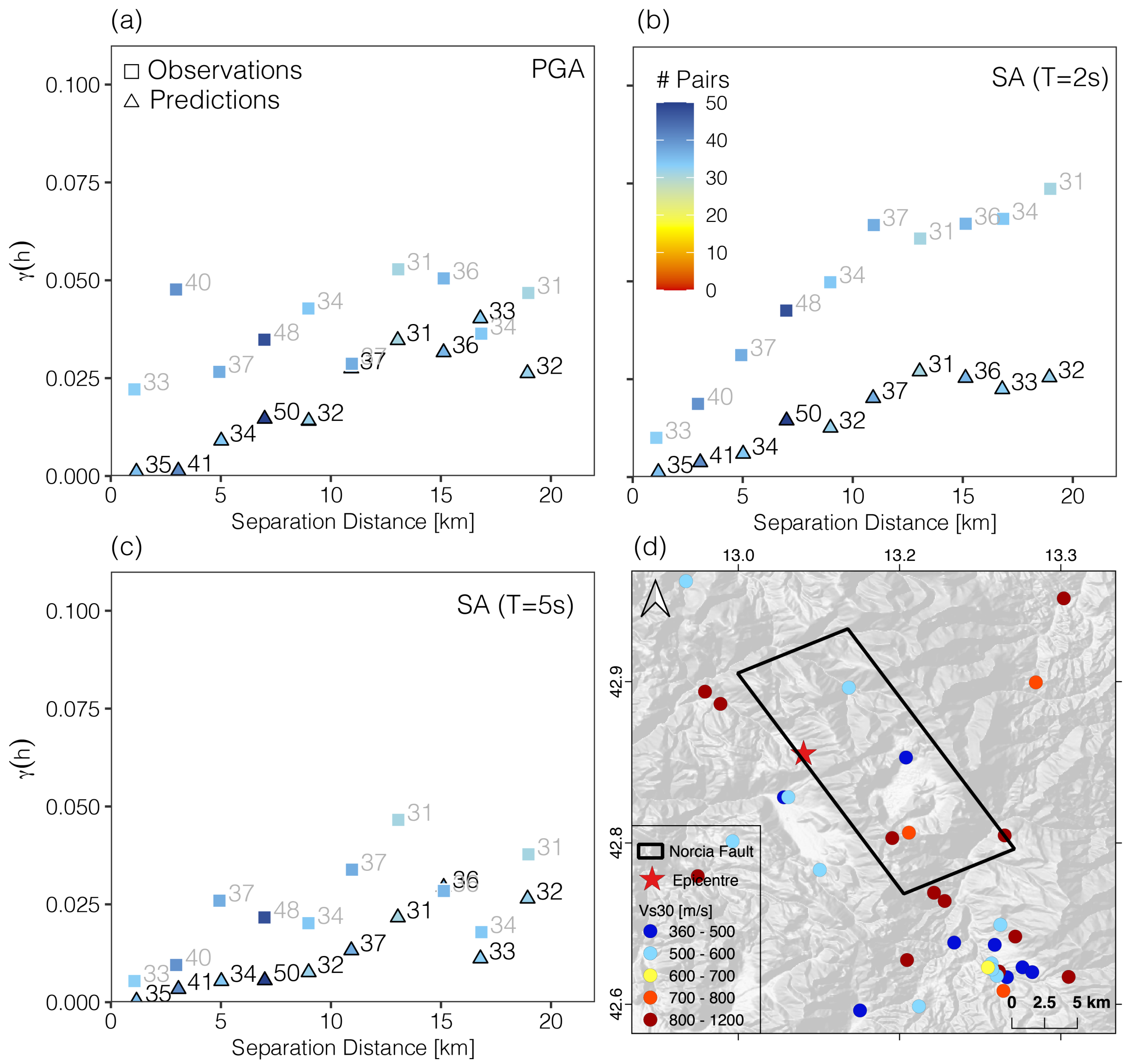


Figure 3.9: (a) Experimental semivariograms obtained by using the ANN2BB approach (Predictions) and observed recordings for PGA; (b) Experimental semivariograms for SA (T = 2s); (c) Experimental semivariograms for SA (T = 5 s); (d) Spatial distribution of the seismic stations (epicentral distance < 30 km) that recorded the Norcia event. Dots are color-coded based on the V_{s30} values. The red star is the epicentre, whereas the black rectangle represents the projected fault surface. [Data are from *https://esm-db.eu/*].

the source-, path- and site-corrective terms of the ground motion model specific for the Po Plain and central Italy regions, respectively. Secondly, the FN sill and range values tend to be larger than the FP ones over a broad range of periods, in agreement with Infantino et al. (2021), who investigated the FN/FP ratios for 6 different 3D PBSs. This is related to the fact that the FN component is typically strongly affected by directivity effects, which cause the ground motion to be more coherent and thus correlated over longer distances compared to the FP. Indeed, the normal fault mechanism along with the slip distribution, which is characterized by a slip asperity (regions of very high slip) close to the top edge of the fault (Figure 3.2a), causes a focusing of the seismic energy from the propagating rupture front along the up-dip direction. Besides, the differences in terms of range and sill rise with increasing period (T >2.5s) as a consequence of directivity effects, being predominant at longer periods. The same applies to the sill, which is strongly influenced by the slip asperities. Likewise, the vertical component shows remarkable directivity effects due to the focal mechanism, which is normal faulting with a dip of 40°. The Z component is indeed characterized by larger range and sill values across almost all periods. To better explain these outcomes, we show the ground motion maps for SA (T=3s) in Figure 3.11. Not only does the vertical component feature homogeneous maximum peak values distributed over a more extended area, but also higher variability in terms of peak values. As a consequence, the resulting ground motion is correlated over longer distances (larger range values) and is characterized by a larger sill with respect to the other components. Finally, it is noted that the spherical model provides a better performance at long periods (T = 4 and 5 s) for all the components; this leads to slightly smaller ranges compared to the exponential functional form.

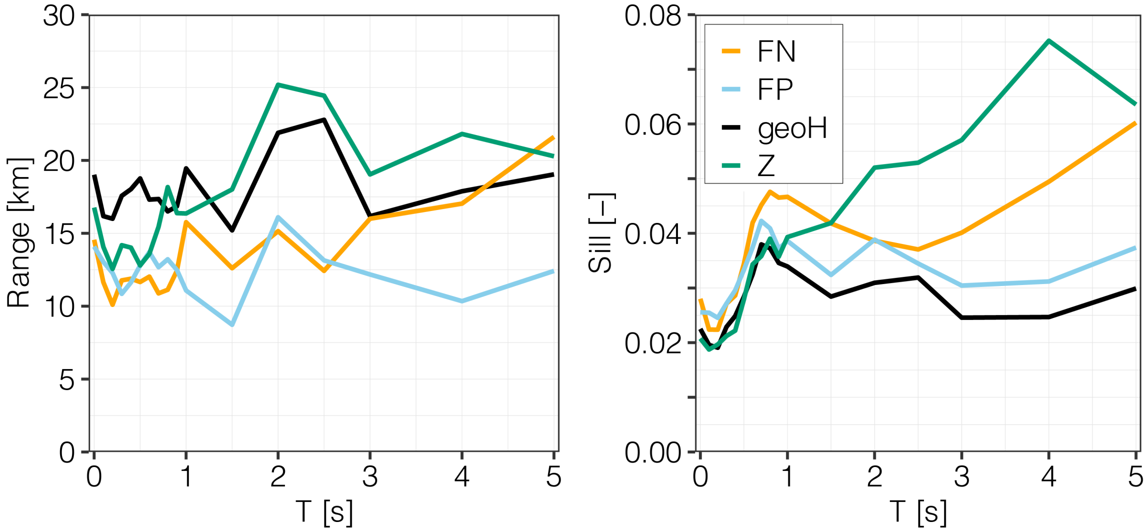


Figure 3.10: (a) Range as a function of the period T for different ground motion components; (b) sill as a function of the period T for different ground motion components..

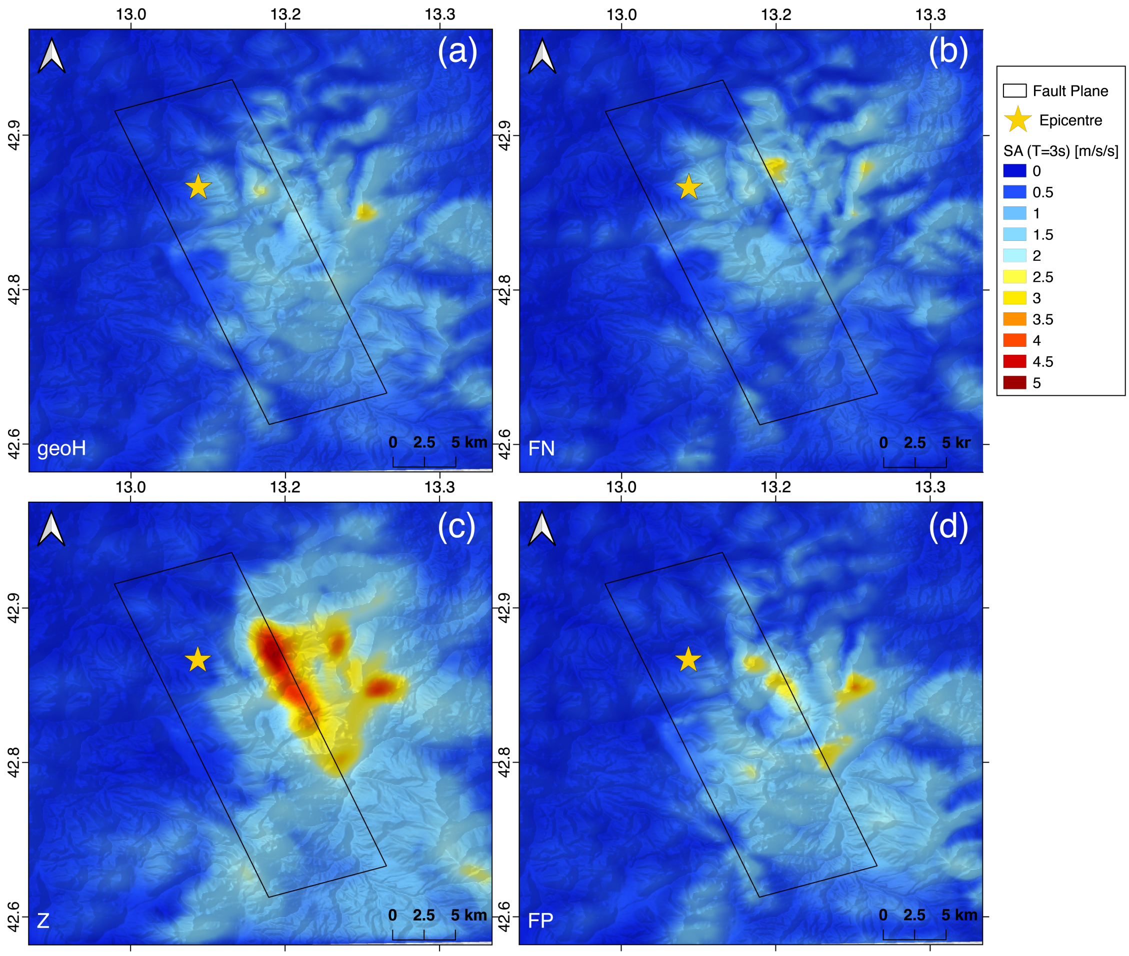


Figure 3.11: Ground motion maps for SA (T = 3s): (a) geometric mean component; (b) FN component; (c) vertical component; and (d) FP component.

3.4.1.3 Spatial correlation anisotropic properties

We relax herein the assumption of isotropy and we compute directional semivariograms for the following azimuths: (1) 0° (North); (2) 45° (North-East); (3) 71° (FN - normal to the fault strike); (4) 90° (East); (5) 135° (South-East); (6) 161° (FP - parallel to the fault strike). It is recalled that directions between 180° and 360° provide the same results because of the symmetric property of the semivariogram (Webster and Oliver, 2007). Figure 3.12 and Figure 3.13 present the range and sill, respectively, as a function of period in terms of average values computed over the different directions along with their standard deviations. There is considerable variability across all periods in terms of both sill and range, meaning that the hypothesis of isotropy is not suitable for any ground-motion component. We also show the trends in the FN and FP directions, highlighting how the former represents a rough lower bound, whereas the latter constitutes an upper bound. This is a rather remarkable but expected result. While the FN direction is mainly affected by the faulting mechanism, rupture directionality effects have a strong impact on the FP direction, leading to more coherent (larger range) and more variable (higher sill) ground motions. Such an observation is also made by Sgobba et al. (2021). While applying their novel methodology to produce non-ergodic shaking scenarios for the M_w 6.5 Norcia case-study, the authors identify a strong azimuthal dependence due to rupture directivity along the strike direction. This is likely connected to the main tectonic features of the Apennines region, where faults mostly follow a NW-SE trend. Finally, the FN over FP range ratio shows that on average the FN component is correlated over longer distances compared to the FP one due to strong directivity effects, especially at longer periods, as found for the isotropic case (Figure 3.10).

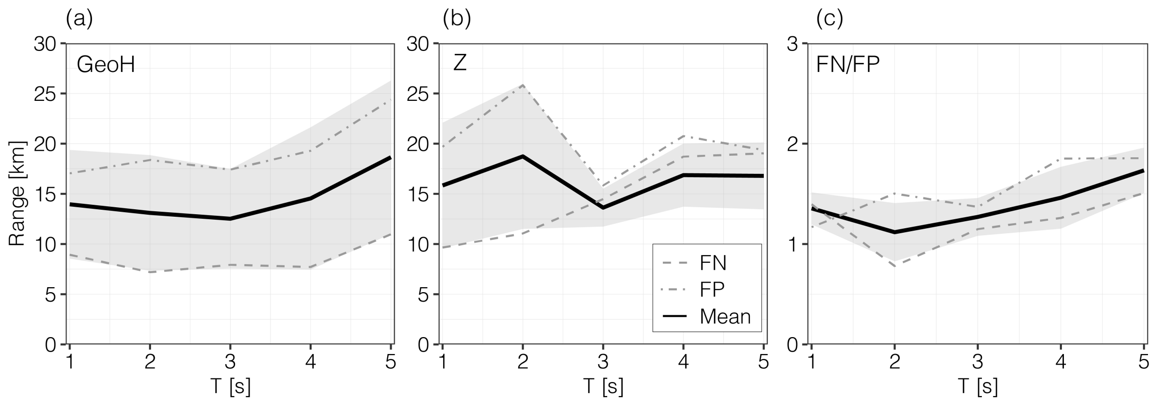


Figure 3.12: Average range values, computed in the different directions, as a function of the period T. (a) Geometric mean component; (b) vertical component; and (c) FN over FP components ratio. The grey area represents the mean ± standard deviation. The FN-direction and FP-direction range values as a function of T are also reported on the graphs.

Finally, for the sake of completeness, we present in Figure 3.14 the experimental semivariograms and the corresponding models obtained along the FN and FP directions for SA (T=3s). Clearly, the semivariogram in the FP direction has a larger sill and range compared to the one in the FN direction.

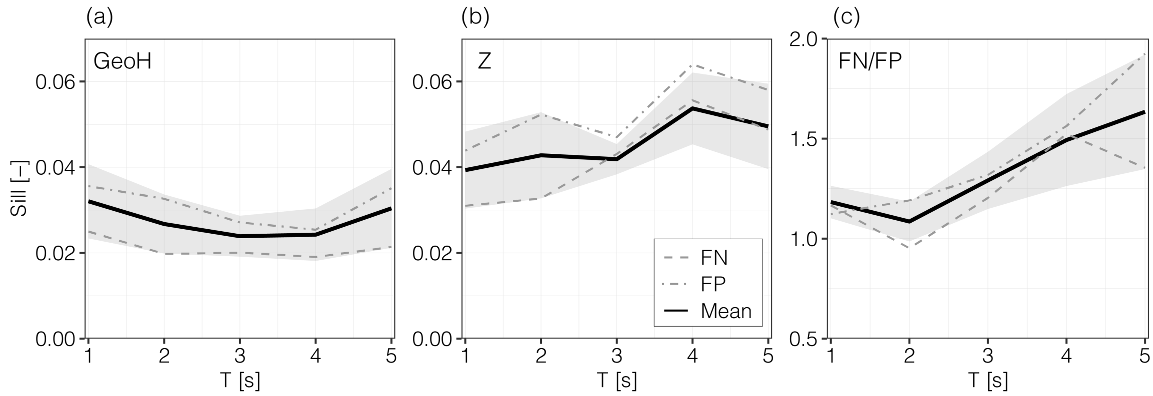


Figure 3.13: Same as in Figure 3.12 but in terms of sill values, computed in the different directions, as a function of the period T.

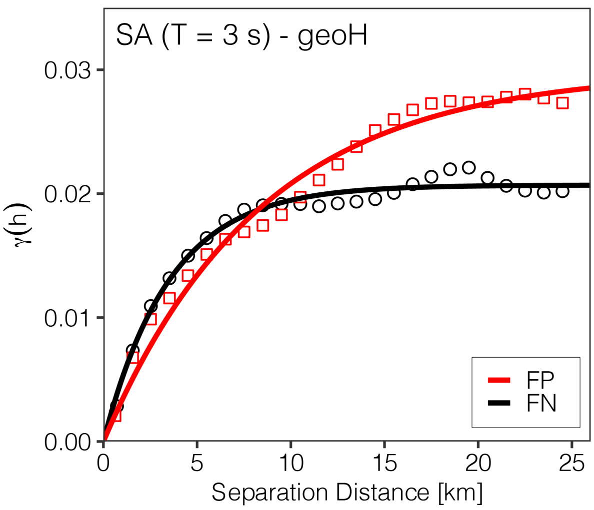


Figure 3.14: Directional semivariograms for SA (T = 3s) for the geometric mean component. Dots and squares represent the experimental semivariograms computed considering azimuths of 71° (FN) and 161° (FP), respectively. Solid lines represent the corresponding models.

3.4.2 Dependence of spatial correlation on source effects

3.4.2.1 Earthquake magnitude

The relationship between magnitude and range has been extensively studied by several authors but it is still under debate (see also chapter 2). Indeed, the database heterogeneity in terms of, for example, fault mechanism and geometry, stress drop as well as regional and local-site conditions, may contribute to the variability in the range in addition to the effect of magnitude. To deepen our knowledge on this topic, we carry out different 3D PBSs by varying the magnitude while keeping the same causative fault. Although a clearer dependency of the range on the magnitude is evident when larger M_w differences are considered (e.g. M_w 6.5 Vs. M_w 5.5 - 4.0), Figure 3.15 suggests that there is not a unique and well-defined trend between magnitude and range for any component and therefore other factors, such as the rupture process, should be considered to explain the range variability. While larger events should feature larger correlation lengths because of their lower frequency energy content, we believe that this is true only in the far-field where the point-source assumption can be deemed valid. In the near-field, the effect of the fault extension along with the intrinsic variability of the rupture mechanism are found to play a crucial role in determining the correlation. A more heterogeneous slip distribution on a larger fault may reduce the coherency of the ground motion, leading to smaller range values. At the same time, more homogeneous slip asperities on the fault plane may induce a ground motion correlated over longer distances. Such assumptions could explain the trend of the M_w 4.0 scenario; although its range values are relatively small, it provides correlation lengths as large as those of the higher-magnitude scenarios. Similar thoughts can be found in Stafford et al. (2019). Indeed, while investigating the impact of the rupture process on spatial correlations using finite difference simulations, Stafford et al. (2019) found a negative correlation between magnitude and range, with M_w 3.0 scenarios having on average a larger range compared to the M_w 6.0 ones. Besides the distribution of the slip asperities on the fault, rupture propagation phenomena coupled with the morphology of the basin as well as the topography may give rise to a higher range variability, masking the real dependency of spatial correlations on magnitude.

The sill seems to be correlated to both the slip asperities, which contribute to the intrinsic variability, and the relative position of such asperities with respect to the hypocentre. In Figure 3.16, we observe that the Mw6.5_S001 scenario generally gives the highest sill values across all periods and for any component. We do not find any particular trends with magnitude for the geoH and FP components, where the sill values are rather comparable across all the scenarios. Conversely, we note some magnitude dependency for the Z and FN components, for which the fault mechanism and directivity effects, in addition to the magnitude, play a significant role in determining the variability of the ground motion. Finally, we recall that the slip distributions of Mw6.5_S002, Mw6.0_S001, Mw5.5_S001 and Mw4.0_S001 scenarios stem from the rupture generator by Herrero and Bernard (1994), whereas the source model of Pizzi et al. (2017) was used for the Mw6.5_S001 scenario.

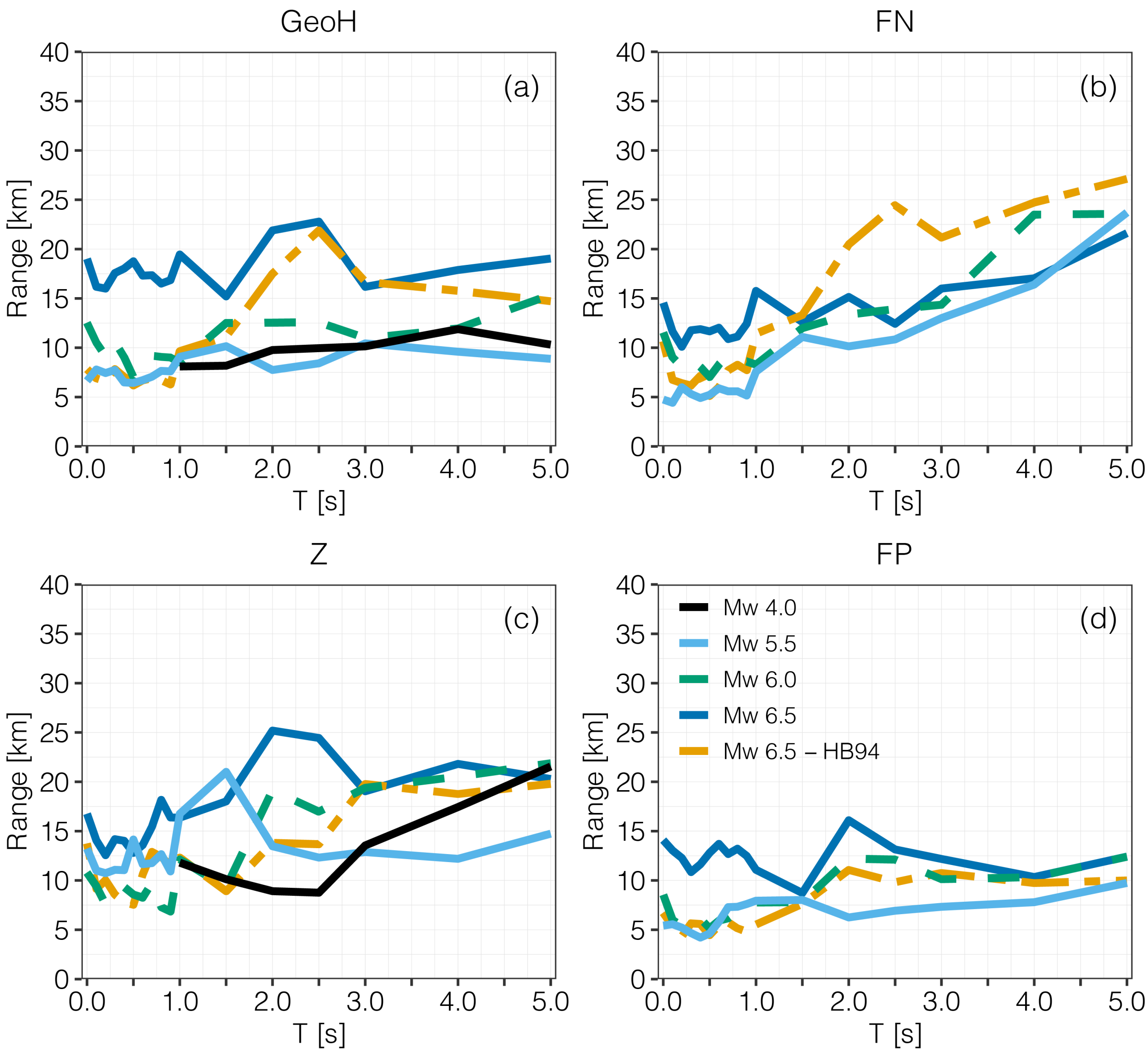


Figure 3.15: Range as a function of the period T for different Mw scenarios: (a) Geometric mean component; (b) FN component; (c) vertical component; and (d) FP components.

This may contribute to the similarity in terms of sill values of the former.

3.4.2.2 Slip distribution

We examine herein the role of the slip distribution by considering two ground shaking scenarios which, while having the same magnitude, hypocentre and causative fault, have different source models (Mw6.5_S001-S002). The impact of the source model is clearly evident in Figure 3.17, in which the maps of ground shaking are presented for SA(T=2s) and for the Z and FN components. In scenario S001, the slip distribution has a large, more

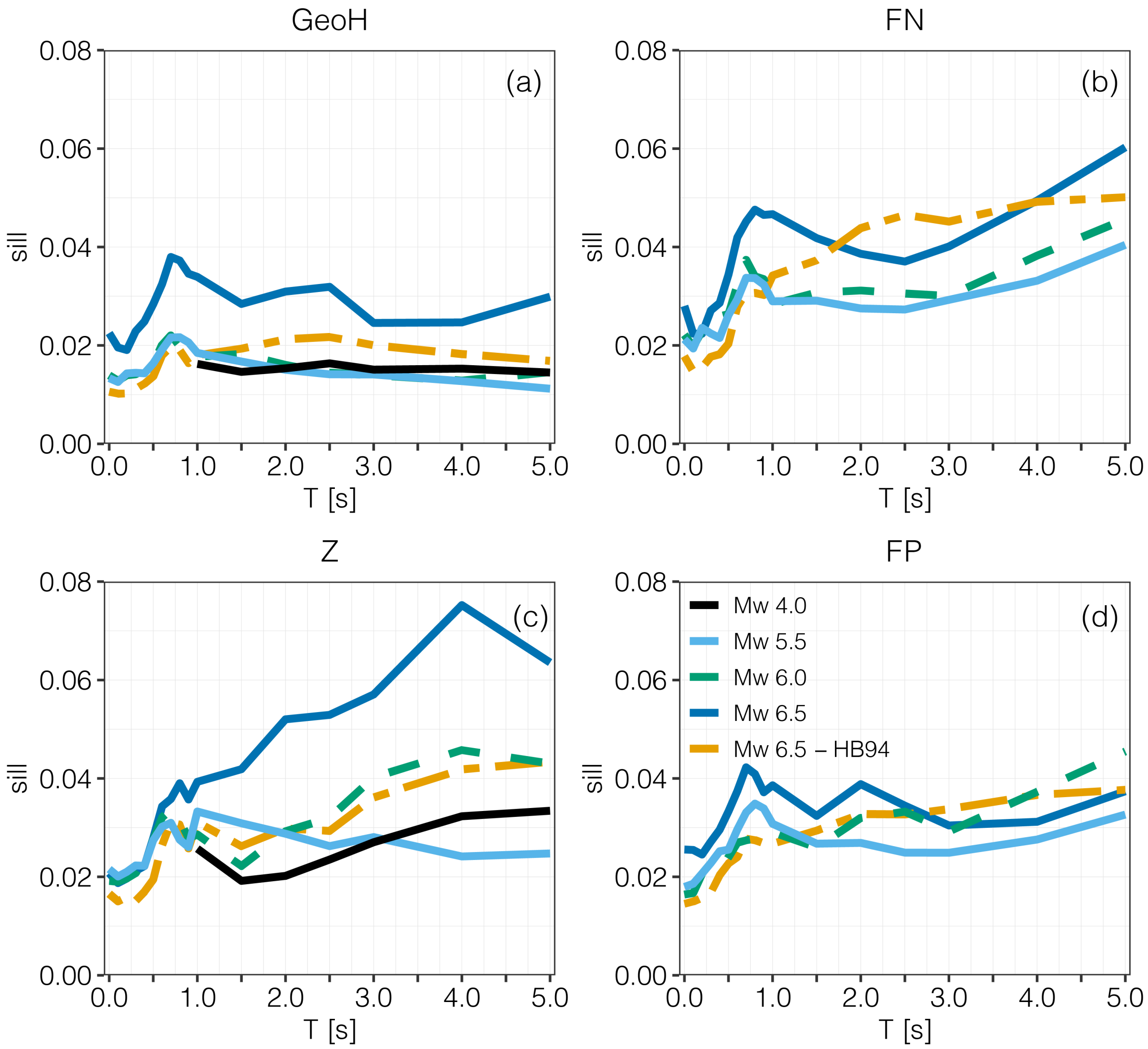


Figure 3.16: Sill as a function of the period T for different Mw scenarios: (a) Geometric mean component; (b) FN component; (c) vertical component; and (d) FP components.

homogeneous slip asperity close to the top edge of the fault, which is located along the path from the hypocentre towards the footwall side. Large peak values are thus generated at all sites located in the up-dip projection of the fault (Figure 3.17 a,c), leading to more coherent ground motions. This is further highlighted by the range and sill ratios computed as $S001/S002$ and presented in Figure 3.18. In general, scenario S001 features ground motions correlated over longer distances and with a higher variability over a broad period range for all ground-motion components, except for FN, with respect to scenario S002. The different behaviour of the FN component is evident from Figure 3.17b. The S002 scenario is characterized by a bilateral rupture, which tends to enlarge the area characterized by more homogeneous higher ground motion amplitudes. Therefore, the ground motion has a correlation structure with a larger correlation length. Our outcomes are consistent with Stafford et al. (2019), who demonstrated how the rupture process of scenarios with equal magnitudes have a pronounced impact on the range variability. Similarly, Chen and Baker (2019) and Infantino et al. (2021) identified different spatial patterns of the correlation coefficient exclusively because of slip distribution and source propagation effects.

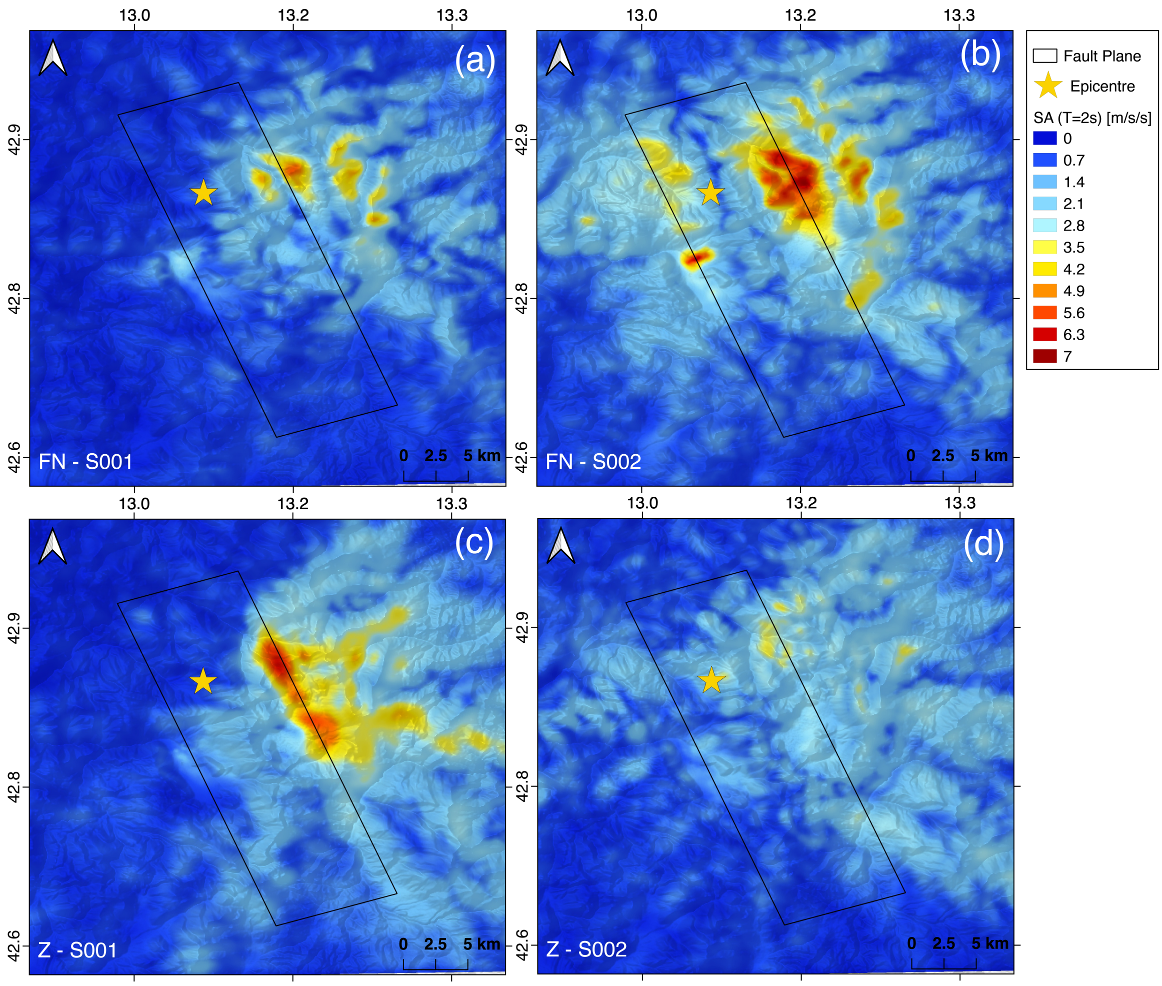


Figure 3.17: SA (T = 2s) Ground motion maps for Mw6.5_S001 (left panel) and Mw6.5_S002 (right panel) scenarios: (a) and (b) FN component; and (c) and (d) vertical component.

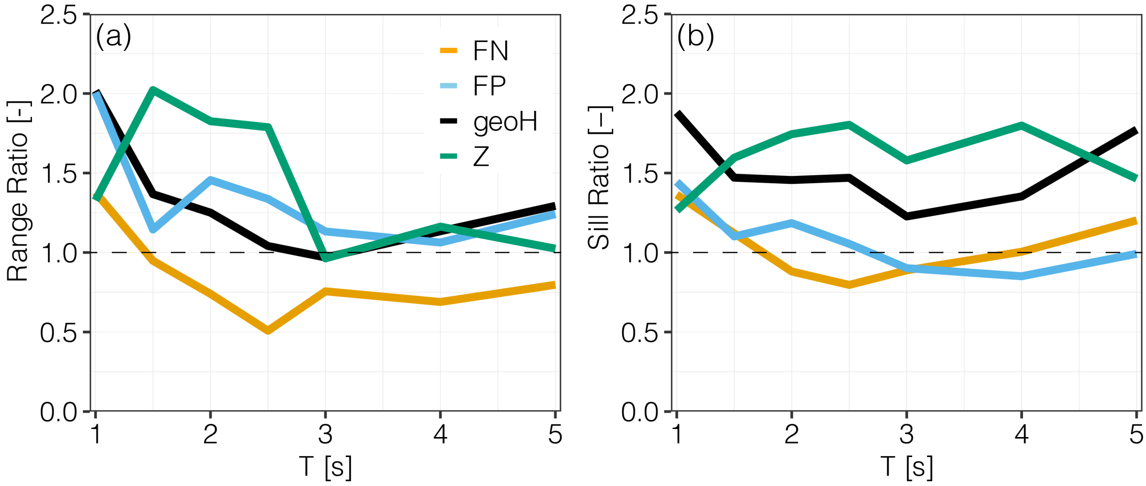


Figure 3.18: Ratio between Mw6.5_S001 and Mw6.5_S002: (a) range values; and (b) sill values.

3.4.2.3 Hypocentral locations

The relative position between slip asperities and the hypocentre is a key element that has a strong influence on the spatial correlation structure of earthquake ground motions. Here, we address the role of the hypocentral location by considering scenarios S002-S008, which differ solely in terms of their nucleation points. Figure 3.19 shows that we cannot identify a unique correlation trend between the FN and FP components. The lack of such tendency is attributable to the relative position between the slip asperity and the hypocentre which determines the size of the directivity effects. For instance, in scenarios S004, S005, S007 and S008 the path from the nucleation point to the main slip asperity follows a fault-parallel direction, causing a polarization of the FP ground motion, compared to the FN. As a result, the latter shows a correlation structure with a smaller correlation length across almost all periods. The S006 hypocentre is located close to that of scenario S002 and similarly it shows a significant directivity for the FN component. At the same time, rupture propagation effects along the FP components are marked. The combination of these effects leads to a similar correlation structure of the FN and FP in terms of range values. Finally, the S003 hypocentre is symmetrical to that of scenario S002 with respect to the main slip asperity; therefore, the rupture propagates in the opposite direction (i.e. towards the asperity) causing considerable 'backward'directivity effects and leading to a correlated FN component over longer distances compared to the FP.

We further test the effects of the hypocentre location on spatial correlation by looking at

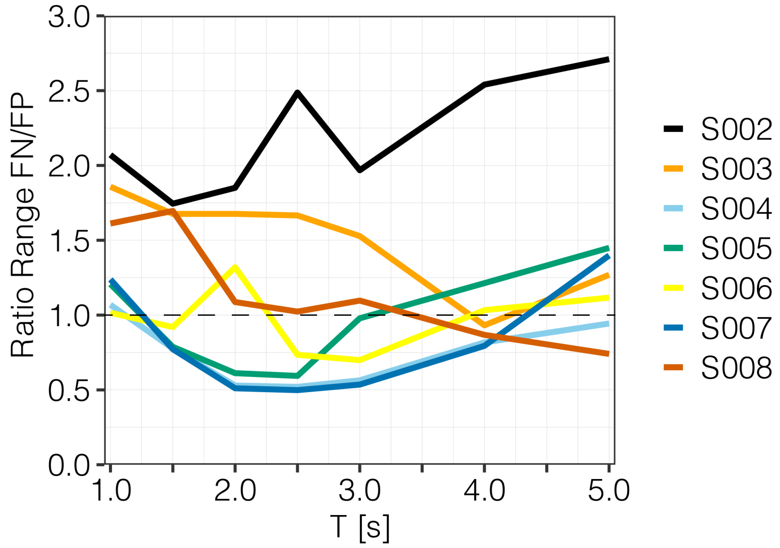


Figure 3.19: Ratio of FN/FP range for the different scenarios with varying hypocentral location.

possible anisotropy patterns. Figure 3.20a and Figure 3.20b compare the average range and sill values computed for the various scenarios for each azimuthal direction, respectively. For the sake of clarity, we plot on a different graph (Figure 3.20 c,d) the standard deviations of the range and sill values for each considered direction. Four main observations can be highlighted. First, the average range and sill values tend to increase with increasing period, independently of the azimuth. Second, the range and sill in the FP direction represent an upper bound, whereas those in the FN direction constitute a lower bound; rupture propagation effects along the FP direction prevail over directivity effects along the FN direction, thus leading to patterns of ground motion spatial variability with preferential direction along the strike of the fault. Such a phenomenon is evident in Figure 3.21, which shows the ground motion shaking of the spectral acceleration (geoH) at T = 2s for each scenario. Similar outcomes were found by Garakaninezhad and Bastami (2017) especially for strike-slip mechanisms. Third, there is higher variability in terms of both range and sill along the FP direction compared to the FN one. While the latter is mostly influenced by the focal mechanism, which is the same for all the scenarios, the former is strongly sensitive to both the rupture propagation and hypocentral location, which vary among the case-studies. As a result, the FP direction, besides having larger range and sill values, features also a larger variability due to the rupture process. Fourth, the variability across all scenarios both in terms of range and sill tends to increase with increasing periods, meaning that the impact of rupture propagation phenomena on spatial correlations is stronger at longer periods. These results refer to the geometric mean component; however, similar outcomes are obtained for the other ground-motion components, as it can be observed in Figure 3.22, 3.23 and 3.24.

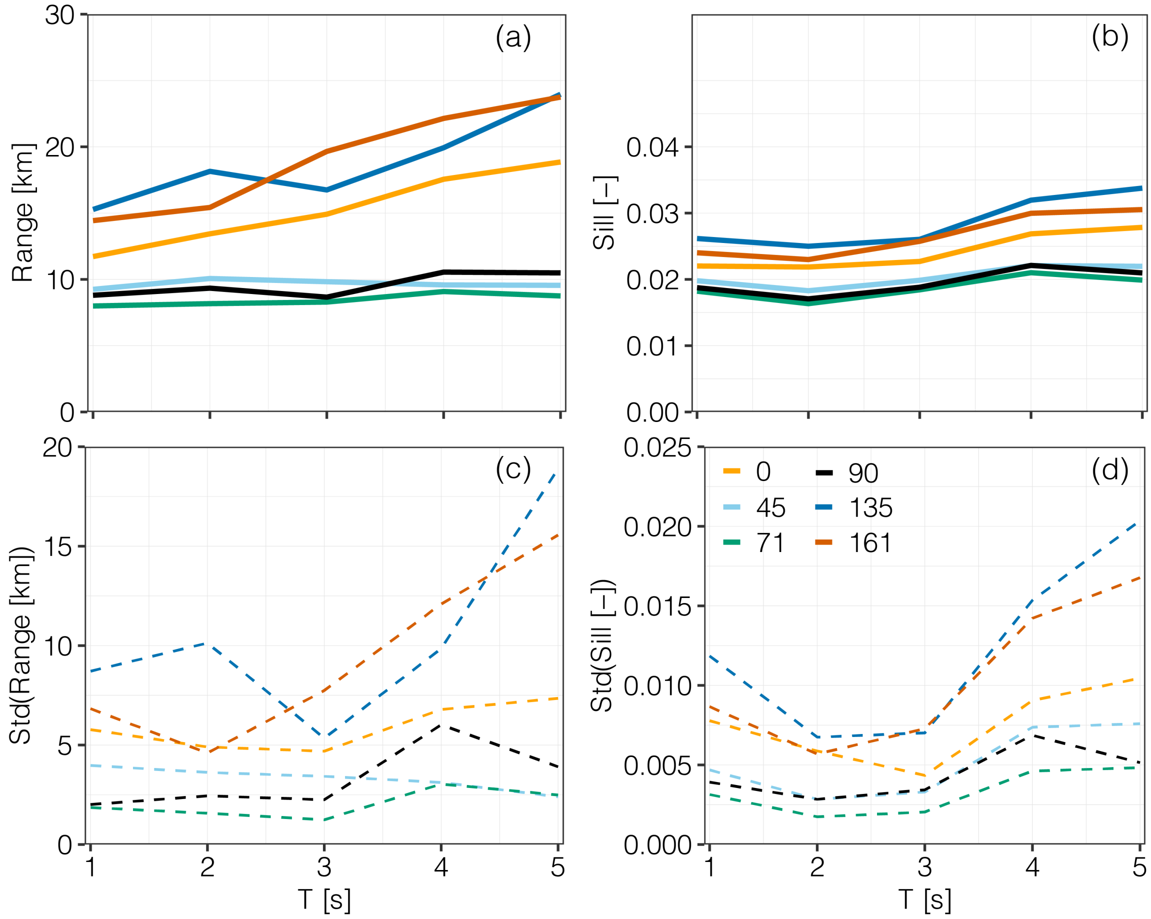


Figure 3.20: (a) Average and (c) standard deviation range values computed for scenarios Mw6.5_S001-S008 for the different azimuth directions; (b) Average and (d) standard deviation sill values computed for scenarios Mw6.5_S001-S008 for the different azimuth directions.

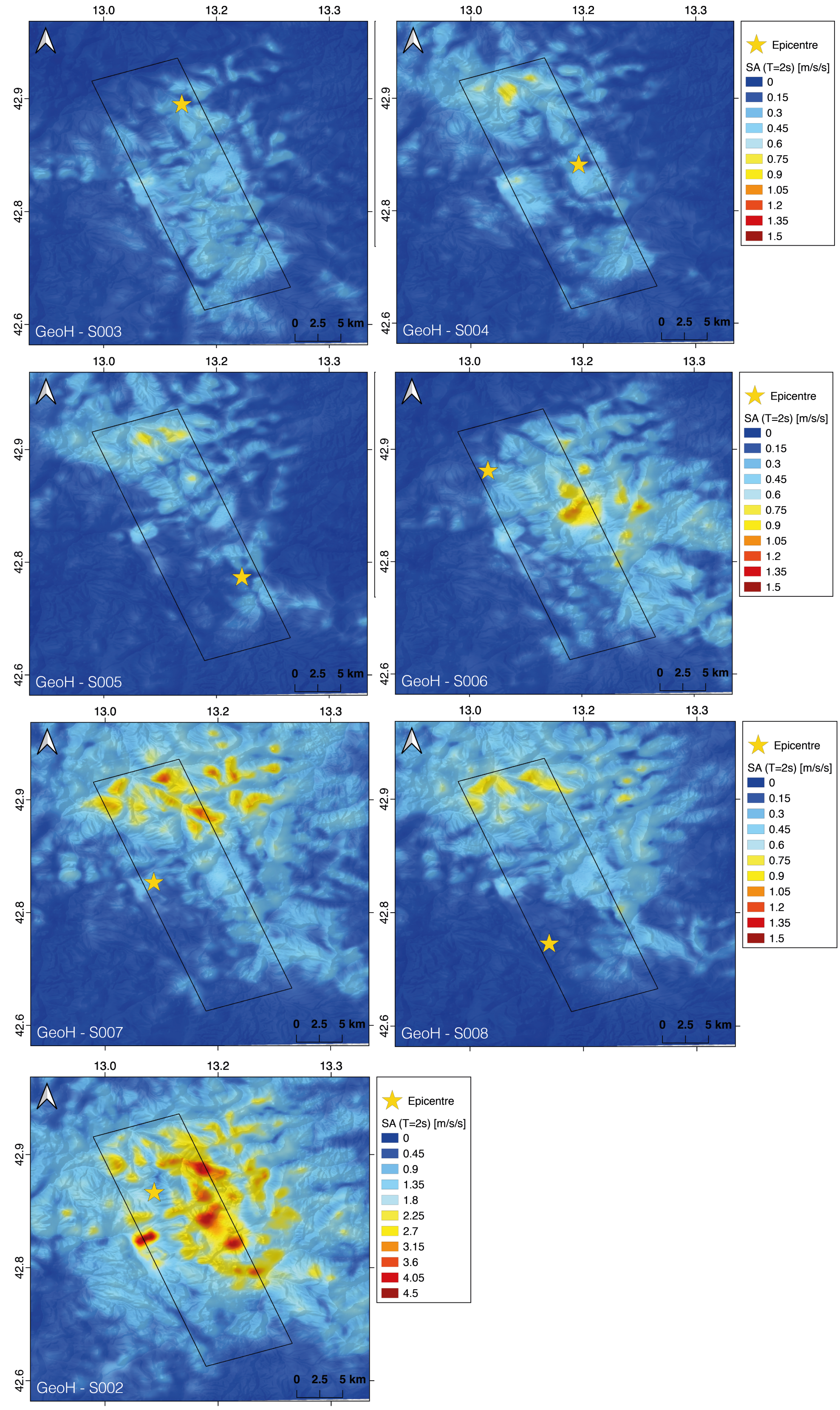


Figure 3.21: Ground motion maps of the GeoH spectral acceleration at 2 s [SA (T=2s)] for scenarios Mw6.5_S002-S008: (a) S003; (b) S004; (c) S005; (d) S006; (e) S007; (f) S008; (g) S002. The epicentral location is pointed out by a yellow star. The black rectangle represents the surface projection of the fault.

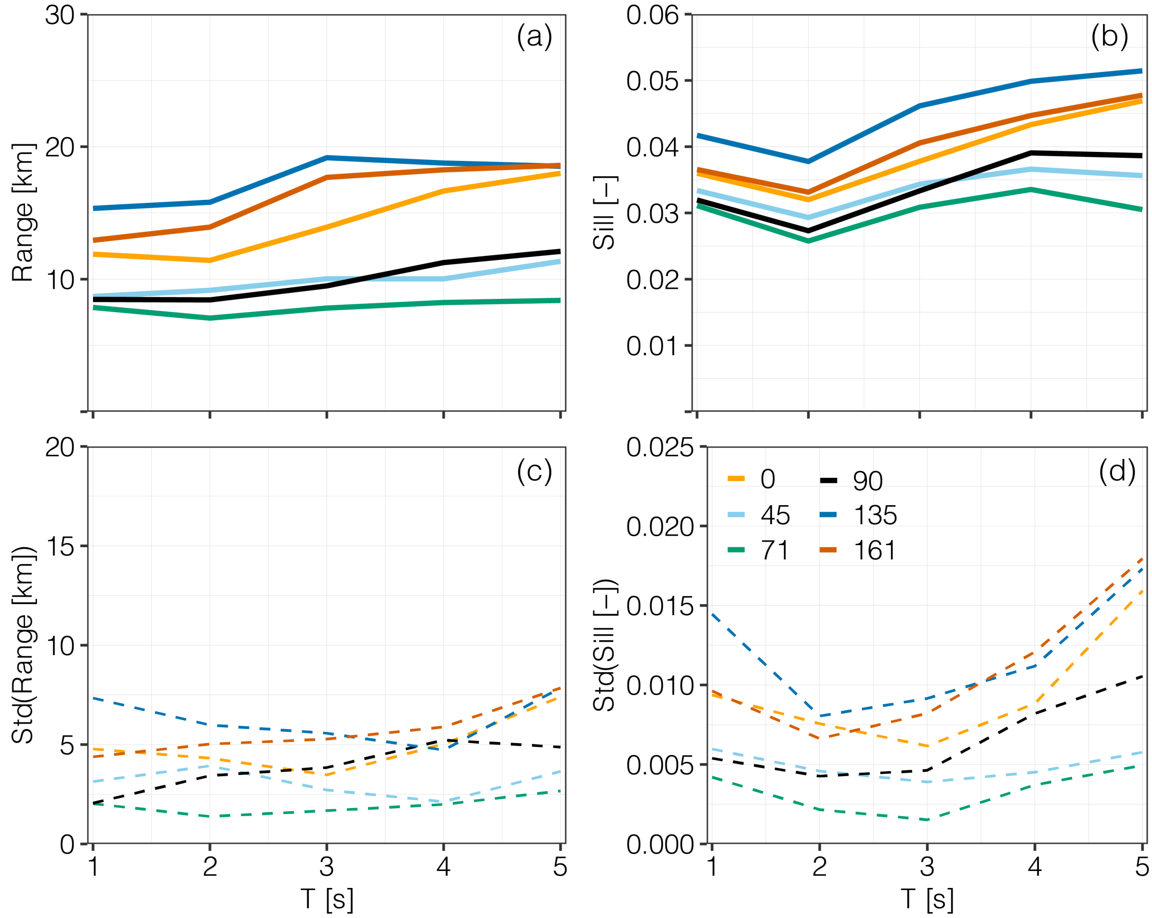


Figure 3.22: a) Average and (c) standard deviation range values computed on scenario Mw6.5_S001-S008 for the different azimuth directions; (b) Average and (d) standard deviation sill values computed on scenario Mw6.5_S001-S008 for the different azimuth directions. Results are for the FN component.

3.4.2.4 Finite fault Vs. point-source

We mentioned in section 3.4.2.1 that the effect of a finite fault can be significant in defining the correlation structure in the near-field. Here, we explore spatial correlation patterns caused by a point-source. Because of the spherical propagation from the nucleation point, we would expect a more isotropic ground motion field. However, Figure 3.25 shows significant spatial anisotropies in the correlation of the Mw4.0_S001 scenario, especially for the vertical component. Similarly to Figure 3.12 and Figure 3.13, Figure 3.25 presents

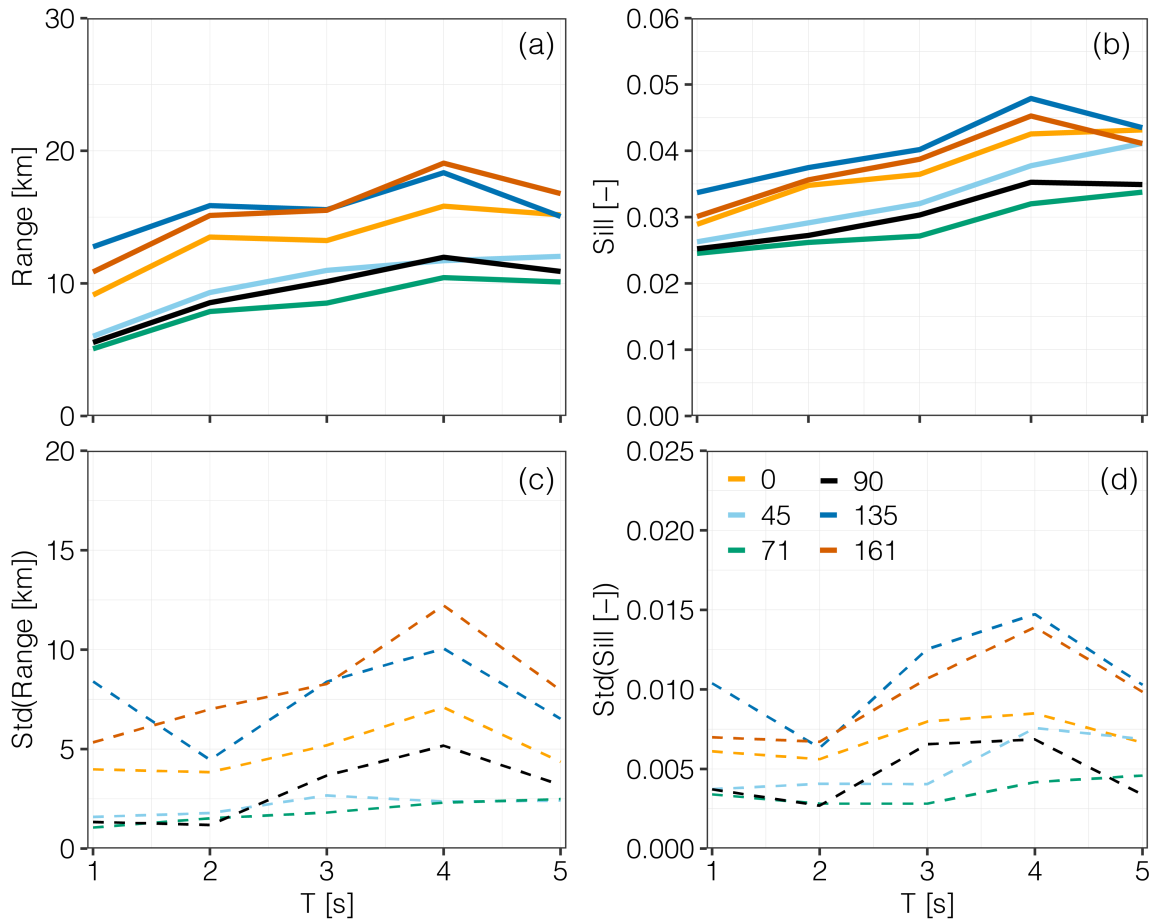


Figure 3.23: (a) Average and (c) standard deviation range values computed on scenario Mw6.5_S001-S008 for the different azimuth directions; (b) Average and (d) standard deviation sill values computed on scenario Mw6.5_S001-S008 for the different azimuth directions. Results are for the FP component.

the range and sill as a function of the period in terms of average values computed for the different directions along with their standard deviations. We believe that such anisotropic effects are likely caused by the radiation pattern, which may exhibit a distorted four-lobed pattern due to propagation in the heterogenous medium. When dealing with finite faults, this effect is also coupled with the rupture propagation on the fault plane; therefore, considering an isotropic, rather than an anisotropic, spatial correlation model may be a major simplification.

Furthermore, ground motion shaking maps (Figure 3.26) suggest that topographic

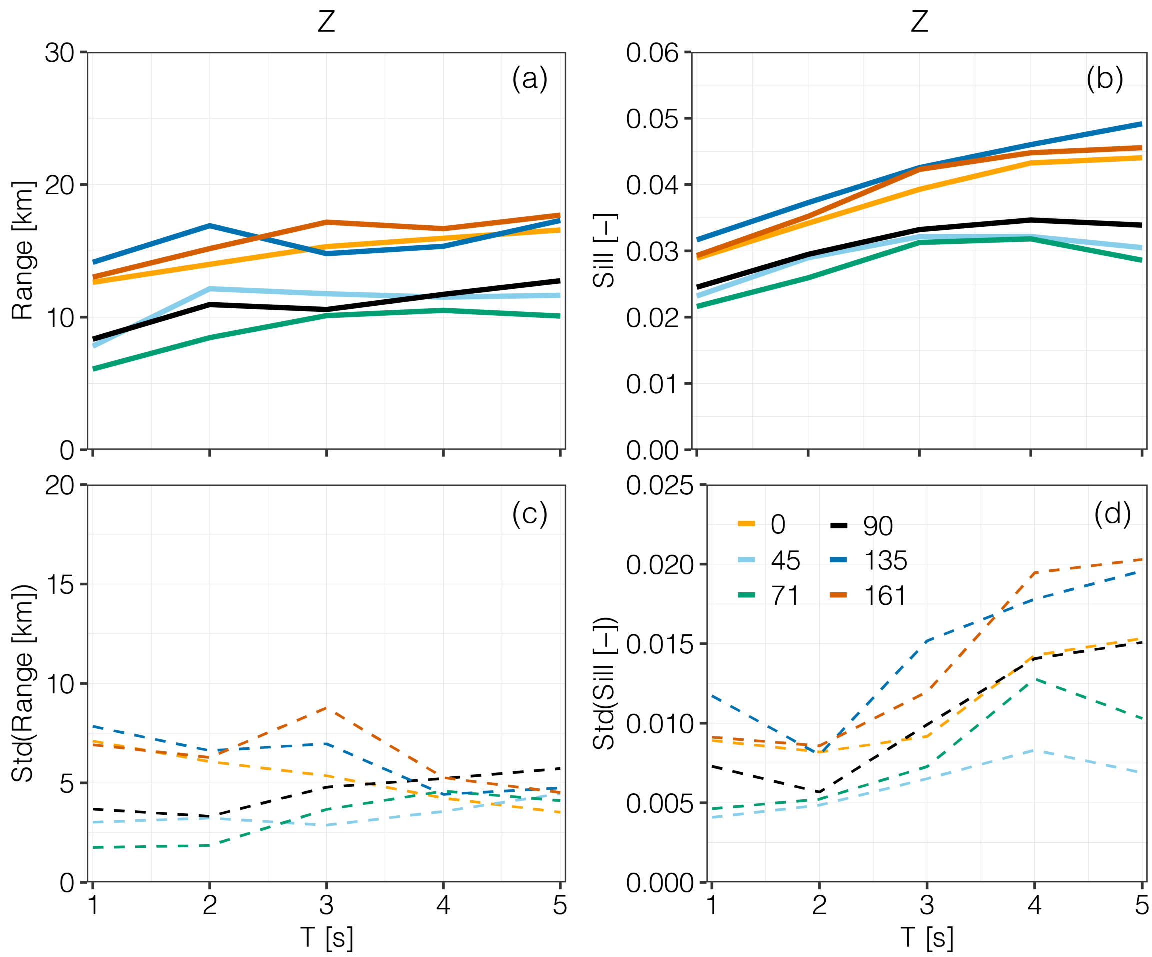


Figure 3.24: (a) Average and (c) standard deviation range values computed on scenario Mw6.5_S001-S008 for the different azimuth directions; (b) Average and (d) standard deviation sill values computed on scenario Mw6.5_S001-S008 for the different azimuth directions. Results are for the Z component.

amplifications may significantly contribute to spatial anisotropy patterns in addition to the radiation pattern. Topographic effects are indeed not negligible in the numerical model (Figure 3.1a). Finally, the sill shows a less pronounced variability, especially for the geoH component (Figure 3.25b). This is reasonable since the sill seems to be significantly coupled with the slip asperities, that contribute to the intrinsic variability of the ground motion. In this case-study, the slip is concentrated at a source-point and therefore does not feature the inherent heterogeneity of the finite fault.

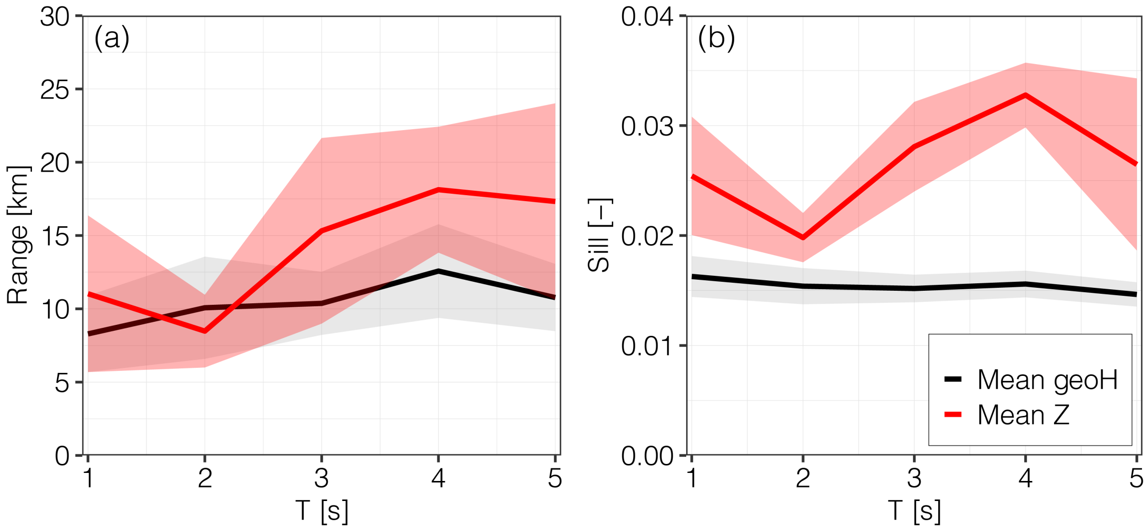


Figure 3.25: Average range (a) and sill (b) values computed on the different azimuth directions for the M_w4.0 point-source scenario. The shadow areas represent the mean ± standard deviation.

3.5 Conclusions

In this study, we explored the spatial correlation structure of response spectral accelerations from a set of broadband physics-based ground motion simulations generated for the Norcia (Central Italy) area. We use classic geostatistical tools, such as the semivariogram, to evaluate properties of the ground motion spatial correlation under both the assumptions of isotropy and anisotropy with a level of detail which could not be achieved by using recordings. Indeed, semivariance estimates are well-constrained despite the relatively small bin-width and azimuth tolerance used in computing the semivariograms.

It is widely recognized in literature that a positive correlation exists between range and period of interest. Not only are our results in agreement with such statement, but we also demonstrate that the sill also tends to increase as the period increases. This result is rather interesting, as it is common practice to work with normalized residuals so that the sill is not directly estimated.

In the near-source region, the polarization of the ground motion can be significant with larger FN and Z peak values due to directivity effects. We, therefore, investigated the spatial correlation structure of different ground motion components to identify diagnostic features. The specific case of the M_w 6.5 Norcia event simulation (scenario Mw6.5_S001) presents a FN ground motion with a larger intrinsic variability and correlated over longer

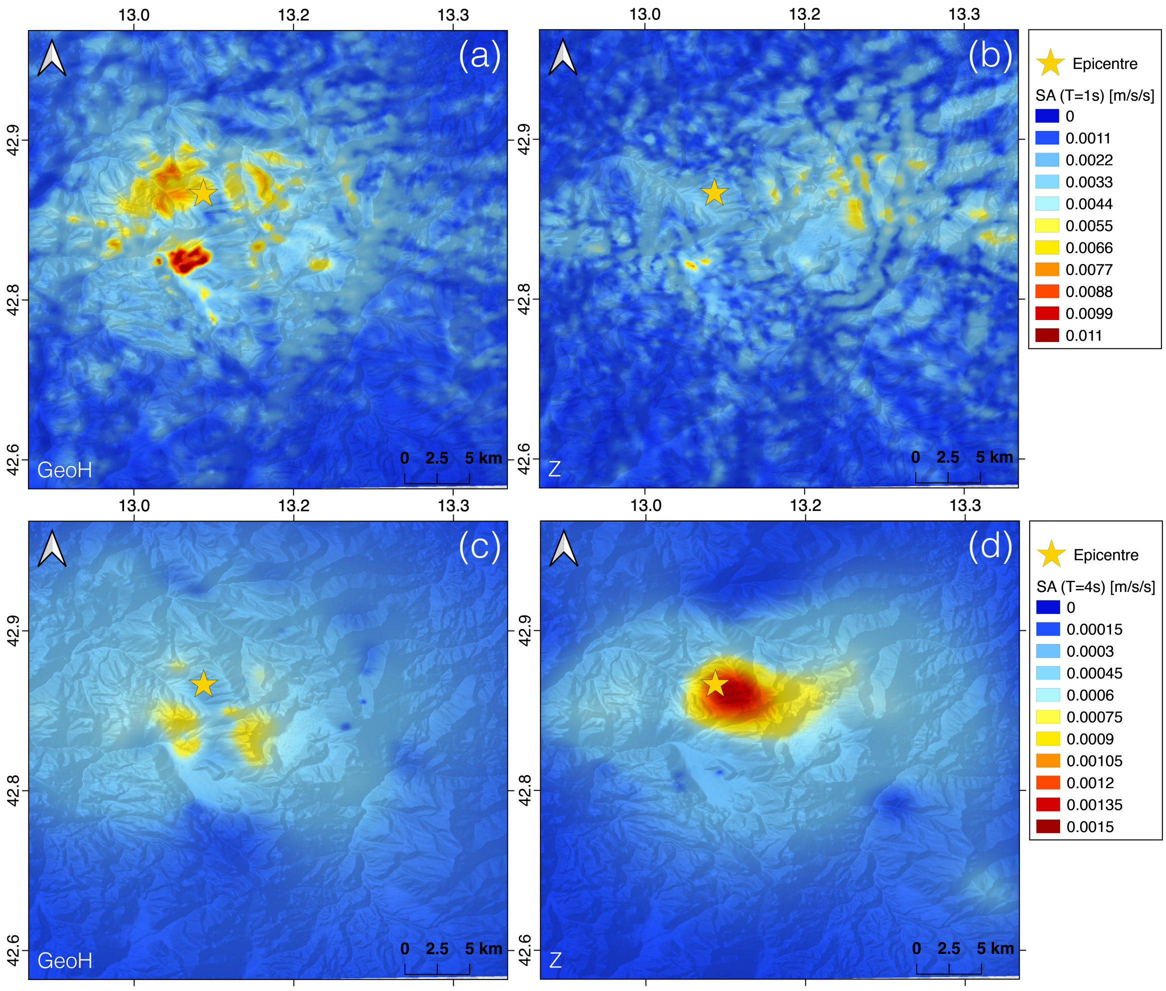


Figure 3.26: Ground motion maps for Mw4.0_S001 scenario: (a) and (b) SA (T=1s) for the geoH and Z components, respectively; (c) and (d) SA (T= 4s) for geoH and Z components, respectively. The yellow star represents the epicentre.

distances compared to the FP component, as a result of strong directivity effects. The slip distribution coupled with the normal faulting mechanism also significantly affects the Z component, which consequently shows the largest range and sill values over a broad period range. At the same time, the analysis of possible spatial anisotropy patterns reveals that the ground motion field has a strong azimuthal dependence due to rupture propagation effects along the strike direction (FP). This result is consistent with the tectonic setting of the region, as also demonstrated by Sgobba et al. (2021).

The availability of different rupture scenarios generated for the same causative fault allowed the investigation of several aspects of the spatial correlation, such as the dependence on the magnitude, slip distribution and hypocentral location. We believe that the magnitude itself cannot explain the large variability of correlation lengths among different events, especially when the same area is considered. As a matter of fact, Heresi and Miranda (2019) found that only a small ratio of the range variability can be ascribed to magnitude. Our results suggest that the rupture process, both in terms of slip distribution and hypocentral location, has a strong contribution to determine the correlation structure of ground motion IMs, in agreement with the findings of Stafford et al. (2019).

Simultaneously, we observed that the relative position between the hypocentre and the main slip asperity has a strong impact on anisotropy patterns. Not only does the FP direction show on average the largest range and sill values, but it is also characterized by a larger variance among the different scenarios (S002-S008) owing to the different rupture propagation phenomena. Conversely, the FN direction shows rather similar spatial patterns among the case-studies, being mostly affected by the faulting mechanism. Similarly, Chen and Baker (2019) and Infantino et al. (2021) suggested that non-stationarity and anisotropy in spatial correlation strongly depend on the source rupture process.

Although the comparison with recordings shows some discrepancies due to the simplifications of the numerical model, this work offers valuable insights into the spatial correlation of ground motions. Such results would likely be challenging to reach with earthquake recordings due to the shortage of data, especially in the near-source region of major earthquakes. Our outcomes suggest that the spatial correlation structure of ground motion IMs is not only period-dependent, but also scenario-dependent, so that more care should be taken when pooling data from different earthquake to calibrate a unique correlation model. Consistently with other studies (e.g Chen and Baker, 2019; Huang, Tarbali, Galasso and Paolucci, 2020; Infantino et al., 2021), our results strengthened the idea that the isotropy assumption may be a strong simplification and thus not reasonable for the seismic risk assessment of spatially-distributed infrastructure.

Finally, we can state that 3D PBSs represent a powerful tool that provide databases with an unparalleled size and richness, which allows specific features of the spatial correlation structure to be addressed. This is particularly important for those regions characterized by sparse seismic networks, for which developing a well-constrained correlation model is challenging. Further research should be undertaken to explore if the correlation properties identified in this work have a significant impact on the regional seismic risk assessment of portfolios of buildings, infrastructures and earthquake-induced phenomena.

Acknowledgements

This work has been supported both by the University of Strathclyde (Ph.D. of first author) and by the European Commission, ITN-Marie Sklodowska-Curie URBASIS-EU project, grant agreement No. 813137 (second author). We thank John Douglas for his fruitful suggestions. Finally, we thank Dr Pitarka and an anonymous reviewer for their constructive comments on a previous version of this study.

CHAPTER 4

Assessment of the uncertainty in spatial-correlation models for earthquake ground motion due to station layout and derivation method

The evaluation of the aggregate risks to spatially distributed infrastructures and portfolios of buildings requires quantification of the estimated shaking over a region. In chapter 2 and chapter 3 we analysed different factors that most affect the spatial dependency of ground motion intensity measures (e.g. peak ground acceleration), such as the estimation approach, source effects and soil conditions. A common geostatistical tool to characterize the spatial correlation is the semivariogram and over the past decades, different fitting approaches have been proposed in the geostatistics literature. A theoretically optimal approach has not yet been identified, as it depends on the number of observations and configuration layout, as mentioned in chapter 2. In this chapter, we further examine in depth such aspect and investigate estimation methods based on the likelihood function, which, in contrast to classical least-squares methods, straightforwardly define the correlation without needing further steps, such as computing the experimental semivariogram. Our outcomes suggest that maximum-likelihood based approaches may outperform least-squares methods. Indeed, the former provides correlation estimates, which do not depend on the bin size, unlike ordinary and weighted least-squares regressions. In addition, maximum-likelihood methods lead to lower percentage errors and dispersion, independently of both the number of stations and their layout as well as of the underlying spatial correlation structure. Finally, we propose some guidelines to account for spatial correlation uncertainty within seismic hazard and risk assessments. The consideration of such dispersion in regional assessments could lead to more realistic estimations of both the ground motion and corresponding losses.

This chapter is based on the following published article:

1. **Schiappapietra**, E., & Douglas, J. (2021). Assessment of the uncertainty in spatial-correlation models for earthquake ground motion due to station layout and derivation method . *Bulletin of Earthquuake Engineering. https://doi.org/10.1007/s10518-021-01179-w*

4.1 Introduction

Many authors (e.g. Iervolino, 2013; Weatherill et al., 2015; Sokolov and Ismail-Zadeh, 2016; Sokolov and Wenzel, 2019) have demonstrated the importance of considering regional hazard estimates when evaluating the aggregate risks to spatially-distributed infrastructure and building portfolios. The assessment of the seismic hazard over a geographical region requires the quantification not only of the expected ground shaking at a single location, but also how this shaking could vary over distances of a few kilometres. This variation is captured within spatial-correlation models. Spatial correlations have been increasingly studied over the last 20 years and many researchers have aimed to identify the factors that most affect the spatial dependency of earthquake ground motions. In chapter 2, we provide a thorough literature review, shedding light on the dependence of correlation on: (1) the estimation approach and fitting method; (2) earthquake magnitude; (3) structural period; (4) regional and local site-effects; and (5) ground motion prediction equations (GMPEs). Baker and Chen (2020) propose a novel approach to quantify both the uncertainty in the correlation estimation and the underlying correlation variability among different earthquakes. Further insights into the spatial correlation of ground motions are given by studies on numerical ground motion simulations. In this regard, Stafford et al. (2019), Chen and Baker (2019), Huang, Tarbali, Galasso and Paolucci (2020), Infantino et al. (2021) and Schiappapietra and Smerzini (2021) provide valuable contributions on the factors that cause the spatial dependency of earthquake ground motion to vary from case to case, with particular emphasis on the earthquake rupture process. In general, studies suggest that the spatial correlation structure is period-, regionally- and scenario-dependent.

Spatial correlation models are usually calibrated on a set of multiple events due to the shortage of ground motion observations from each single earthquake. Using data from a single event would often lead to poorly constrained correlation parameters and models that have limited applicability for future earthquakes. Although it is recognized that the correlation varies from event to event, only few studies (e.g. Goda, 2011; Heresi and Miranda, 2019) have taken into account such event-to-event correlation variability. The consideration of this dispersion in regional probabilistic risk assessment could lead to more realistic estimations of both the ground motion and corresponding losses. Baker and Chen (2020) demonstrated that the true variability in correlation estimates of poorly-recorded events does not significantly differ from that of well-recorded events and that the differences in terms of apparent total variability are exclusively due to the larger estimation uncertainty of poorly-recorded earthquakes.

In this broad framework, the research question we would like to answer is whether it is best to have a local correlation model, even though it is not well constrained, or to implement a global correlation model, characterized by a lower uncertainty but calibrated on worldwide databases? We, therefore, focus our attention on improving the estimation of correlation parameters by using alternative approaches. This study is a continuation of chapter 2 and it aims to provide guidelines for developers and users of spatial-correlation models. To achieve this goal, we use simulations of spatially-correlated ground motion fields which, as opposed to real data, provide a controlled environment where the true model is known.

Section 4.2 summaries spatial correlation modelling theory and it introduces the approaches for correlation estimation we use throughout this study. Section 4.3 describes the steps to generate spatially correlated random fields. We propose here two different studies: (1) ground-motion fields simulated on a fine grid, and (2) ground-motion fields simulated only at recording locations corresponding to those of past earthquakes. Finally, sections 4.4, 4.5 and 4.6 discuss the main results and the implications of this work.

4.2 Spatial correlation modelling

Traditional seismic hazard and risk analysis tools usually employ GMPEs to estimate the earthquake ground motion at a given site. The earthquake ground motion of interest to engineering is often the transient ground shaking that occurs during an earthquake. This ground motion is invariably evaluated in terms of one or more scalar intensity measures (IMs), such as the peak ground acceleration, the peak ground velocity and, occasionally, the peak ground displacement. The ground motion is also often expressed in terms of response spectral acceleration, which represents the maximum response of a single degree-of-freedom system of a given oscillator period and damping subject to the ground motion time-history. GMPEs provide the marginal probability distribution of the IM at a single site as a function of a set of parameters describing the earthquake source, such as the magnitude, the propagation path and local site conditions (e.g. Douglas and Edwards, 2016):

$$\log_{10} Y_{ij} = \log_{10} \bar{Y_{ij}} \left(M, R, S, \theta\right) + \varepsilon_{ij} + \eta_i \tag{4.1}$$

where Y_{ij} is the IM of interest at the j^{th} site due to the i^{th} event, whereas $\bar{Y_{ij}}$ is the predicted median function of magnitude (M), distance from the source (R), local-site conditions (S) and other explanatory variables (θ). ε_{ij} and η_i are the within-event and between-event residuals terms, respectively. ε_{ij} represents systematic deviations between observed and median predicted values due to path and local site effects, whereas η_i denotes systematic deviations associated to an event. For this reason, while ε_{ij} is site-dependent, η_i is common for all sites. Both residual terms are assumed to be normally distributed with mean zero and standard deviations φ and τ, respectively. To fully characterise ε, it is necessary to describe how the within-event residuals vary in space, namely to model the spatial dependence of ε_{ij} and ε_{ik}. Baker and Jayaram (2008) demonstrated that spatially distributed within-event residuals are jointly normally distributed. Therefore, their spatial dependence can be completely defined by the covariance matrix, which reflects their correlation structure.

4.2.1 Spatial variability of within-event residuals

In geostatistics, a common tool to describe the dependence structure of spatial distributed random variables (i.e. the within-event residuals) is the semivariogram, which measures the average dissimilarity of a pair of ε_{ij} and ε_{ik} separated by an inter-site distance h:

$$\hat{\gamma}\left(h\right) = \frac{1}{2} Var \left[\varepsilon_{ij} - \varepsilon_{ik}\right] \tag{4.2}$$

in which Var indicates the variance. The semivariogram is empirically evaluated from observations by pooling all data with a given inter-site spacing h and then using either the robust estimator proposed by Cressie (1985) or the classic method of moments proposed by Matheron (1962). Usually, the individual separation distances between pairs of observations are grouped into bins, so that the semivariances are computed for each pair of sites whose inter-site distance falls in the interval $[h - \Delta, h + \Delta]$. The hypothesis of second-order stationarity and isotropy are generally assumed due to the lack of repeated ground motion observations from the same event at a given site. Therefore, the correlation between any pairs of sites with equal separation distance is the same, independently of the source-to-site distance and orientation. Under such assumptions, the semivariogram and the correlation are equivalent and the following relation holds (Diggle et al., 2007; Oliver and Webster, 2014):

$$\gamma\left(h\right) = \ \varphi^{2} - COV\left(\varepsilon_{ij}, \varepsilon_{ik}\right) = \varphi^{2}\left[1 - \rho_{\varepsilon}\left(h\right)\right] \tag{4.3}$$

where COV is the covariance matrix and ρ_{ε} the correlation function. The reader is referred to chapter 2 for further details.

4.2.2 Fitting methods for semivariogram models

The experimental semivariogram of equation 4.2 is a discrete function, describing the spatial continuity of the random variable ε. Parametric functions are used to fit the experimental semivariogram to retrieve semivariogram models for any separation distance h. In the literature, a number of admissible models (e.g. spherical, Gaussian and exponential) exist; however, we choose the exponential function to model the correlation structure, as it is the most widely adopted (e.g. Jayaram and Baker, 2009; Esposito and Iervolino, 2012; Baker and Chen, 2020)) functional form in engineering seismology. The general form of the exponential function is:

$$\gamma\left(h\right) = a\left[1 - \exp\left(-\frac{3h}{b}\right)\right] \tag{4.4}$$

where a and b are the sill and the practical range of the semivariogram, respectively. The sill represents the variance of the random variable, whereas the practical range is the separation distance at which $\gamma\left(h\right)$ equals 95% of the sill value. An illustration of an empirical and fitted semivariogram model is presented in Figure 4.1a. Different fitting approaches have been proposed in the geostatistics literature. In general the model coefficients are chosen so that the misfit between observed and predicted values is minimised. Baker and Chen (2020) provide a useful summary of the most common techniques, such as the ordinary (OLS) and weighted (WLS) least squares, and they suggest a new weighting function to weight the values from small distance more within the fitting step. A trial-and-error (manual fitting) approach has been chosen by different authors for its versatility in fitting the data. Nevertheless, we discourage performing a visual fit due to its high degree of subjectivity.

In our analysis, we implement the R package gstat (Pebesma, 2004) to compute the experimental semivariogram and obtain semivariogram model coefficients by means of the

OLS and WLS regression techniques.

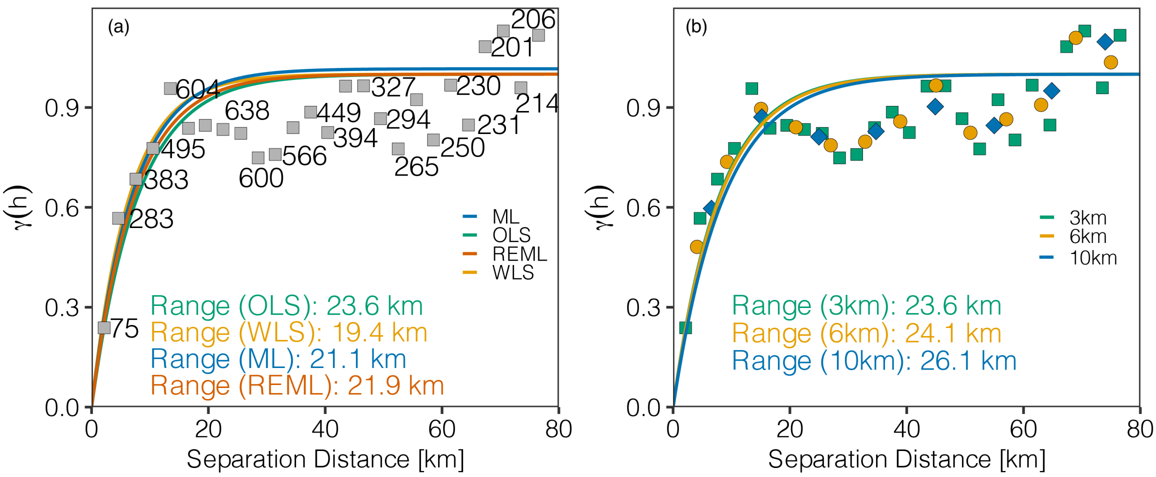


Figure 4.1: Empirical and fitted semivariogram models: (a) different techniques to estimate the semivariogram parameters as introduced in this work. The solid line is the exponential fitted model, whereas squares represent the experimental semivariogram. The numbers close to the squares indicate the number of pairs used to compute the semivariances within each bin. (b) Different bin sizes to compute the experimental semivariogram. The exponential models are fitted by using the OLS approach.

4.2.3 Maximum likelihood estimation

The above-described method of least squares to fit semivariogram models is not direct because it requires the computation of the experimental semivariogram as an intermediate step (Li et al., 2018). Moreover, the final modelling outcomes depend on different assumptions such as the variogram estimators (Oliver and Webster, 2014) and weighting functions (Baker and Chen, 2020; Schiappapietra and Douglas, 2020), and on the introduction of arbitrary parameters such as the bin size. For instance, we plot in Figure 4.1b the empirical semivariograms computed by using different bin widths and the corresponding fitted exponential models. Nonetheless, this methodology is the most widely used to determine the dependence structure of spatially distributed random variables and many geostatistical software packages [e.g. R packages geoR (Ribeiro Jr et al., 2020), gstat (Pebesma, 2004), georob (Papritz, 2020) and MATLAB functions variogram (Schwanghart, 2021*a*), variogramfit (Schwanghart, 2021*b*)] allow the user to obtain semivariogram parameter estimates easily (Li et al., 2018). On the other hand, estimation methods based on the likelihood function have increasingly gained influence in geostatistics, particularly in the presence of trends (Oliver and Webster, 2014). The parameters of the correlation structure model are directly estimated by maximising a log-likelihood function without needing further steps, such as computing the experimental semivariogram. Despite such an advantage, maximum-likelihood approaches have not been commonly used in engineering seismology. To the authors knowledge, only Ming et al. (2019) employed the maximum-likelihood method to simultaneously estimate the GMPE and correlation function coefficients. One of the main drawbacks of the maximum-likelihood estimation is that it requires the data (e.g. within-event residuals) to be normally distributed. Normality of within-event residuals has been shown to hold, at least within ±3 standard deviations of the mean (e.g. Strasser and Bommer, 2009).

In our analyses, we take advantages of different techniques, such as the Gaussian maximum likelihood (ML) and the restricted maximum likelihood (REML). In general, the model for a set of geostatistical data $Y_i = Y(x_i)\ [i = 1,\ \ldots,\ n]$ at locations x_i is defined as the following:

$$Y(x_i) = D(x_i)^T \beta + B(x_i) + \varepsilon_i \tag{4.5}$$

where $D(x_i)^T \beta$ is the spatial trend, $B(x_i)$ is the Gaussian random field with zero mean and covariance $R(h,\ \sigma^2, \alpha)$ and ε_i is an independent distribution error with zero mean and variance τ^2 (nugget effect). σ^2 and α are the sill and range parameters of the covariance function, whereas β is the vector of regression parameters of the spatial trend. For a normally distributed Y, the log-likelihood for the estimation of σ^2,α and β is defined as:

$$L\left(\beta,\ \tau^2,\ \sigma^2, \alpha\right) = \ -0.5n \log(2\pi) + \log\left\{\left|\sigma^2 R + \tau^2 I\right|\right\} + (y - D\beta)^T (\sigma^2 R + \tau^2 I)(y - D\beta) \tag{4.6}$$

Therefore, the model parameters are obtained by maximising the function $L(\beta,\ \tau^2,\ \sigma^2, \alpha)$. The reader should refer to Diggle et al. (2007) and Künsch et al. (2013) for deeper insights into likelihood-based methods. We implement both the R packages geoR and georob for the parameter estimation through maximum-likelihood approaches.

4.3 Simulations set up

We generate spatially-correlated ground-motion fields (i.e. the within-event residuals at each station location), using a multivariate normal distribution, defined by a zero mean and a covariance function, which reflects the correlation of the within-event residuals. The (unconditional) simulations of Gaussian random fields for given covariance parameters are generated by using the R package geoR. We opt for an exponential correlation model with correlation length h_0 and we choose different values of h_0 [5, 10, 15, 20, 30, 40 km] to cover the typical estimates reported from ground-motion observations. Such broad range mainly depends on magnitude, fault mechanism and source effects as well as regional and local-site conditions, even when the same seismic region is considered (e.g. chapter 2 and 3). A Monte Carlo approach is adopted to generate 1000 simulations of ground-motion residuals for each h_0 to obtain more stable and robust outcomes. Figure 4.2 illustrates two out of the 1000 simulations generated by imposing correlation lengths of 5 and 15 km, respectively.

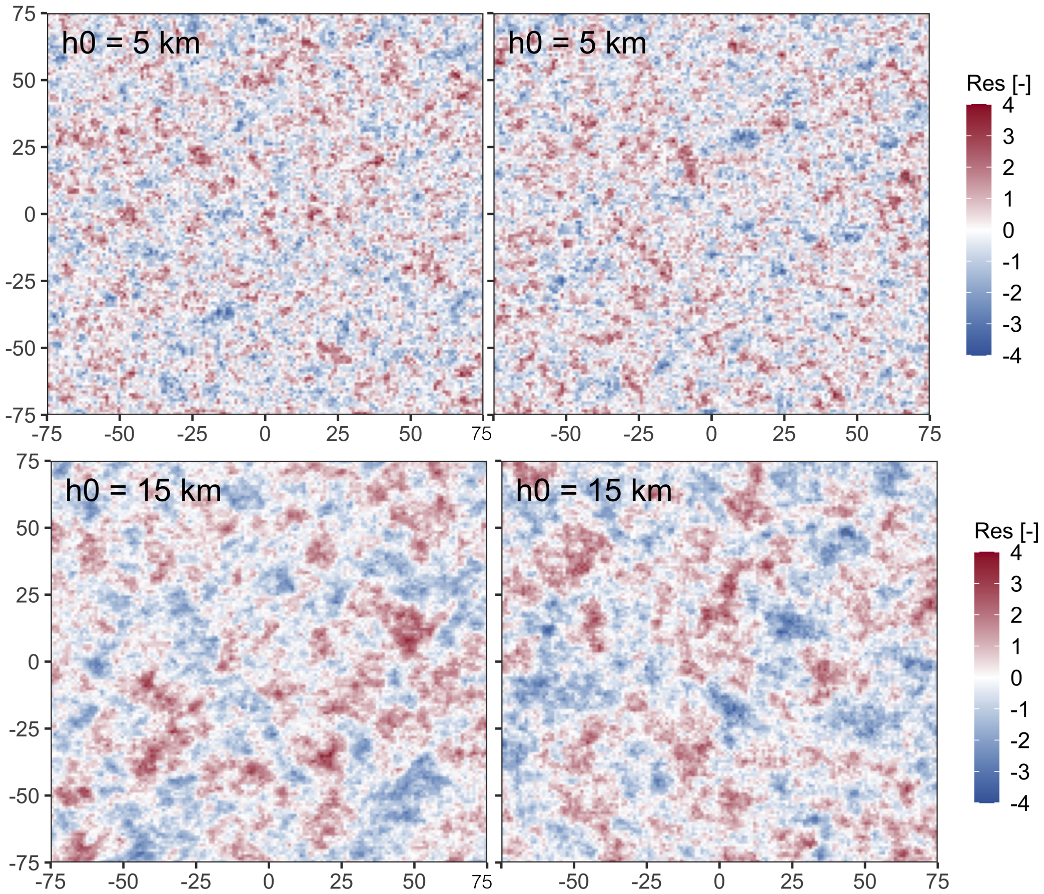


Figure 4.2: Examples of Gaussian random fields characterized by a correlation length of 5 km (top row) and 15 km (bottom row).

For each Monte Carlo simulation (i.e. for each within-event residual distribution), we follow these steps to assess the performance of the different estimation approaches and the influence of different parameters such as the bin size and number of available stations:

1. Randomly locate strong-motion recordings stations throughout the region. We select

a different number of stations [20, 40, 60, 80, 100] for each h_0 to cover the number of strong-motion stations that earthquakes are usually recorded by.

2. Estimate the empirical semivariogram and derive a semivariogram model (equation 4.4) using both least-squares regression and maximum-likelihood techniques.

3. Compare the set of range estimates with the imposed initial range h_0.

It is noted that for each h_0, the simulated ground-motion fields are the same throughout the analyses that consider different numbers of stations. Therefore, the number of available stations is the only varying parameter. At the same time, the stations layout does not change throughout the analyses for different h_0. This means that, at equal number of stations, the correlation structure is the only varying parameter among the different h_0.

The ground-motion fields are generated on a 150 km × 150 km grid with a 1 km resolution. We believe this grid dimension represents a good trade-off between computational cost and the risk of boundary effects on the results. As shown in Figure 4.3, the grid dimension affects the distance cut-off (i.e. the maximum separation distance in the semivariogram computation). In particular, we observe, as expected, a sharp drop of the number of stations at around 50 km when a grid of 100 km is used. By contrast, the number of stations pairs increases for larger separation distances when a grid of either 150 km or 200 km is chosen. Such maximum usable distance has an impact on the range estimates and their variability, particularly when larger values of h_0 are considered. As a matter of fact, we plot in Figure 4.4 the median along with the first and third quartiles values of the range estimates as a function of the grid dimension, for two different estimation approaches. The 150 km grid shows the minimum bias with respect to the initial h_0 and the lowest dispersion (as indicated by the first and third quartile) compared to the 100 km grid.

4.4 The effect of bin size

We present here some preliminary analyses performed by varying the bin size from 3 km to 10 km. For each h_0 and for an equal number of stations, the bin width is the only varying parameter. This allows the lower range value that is resolvable to be determined and to demonstrate how the bin size and range are interconnected. Figure 4.5 and Figure 4.6 show the percentage error and the interquartile range (IQR) computed by:

$$\%error = \frac{\hat{h} - h_0}{h_0} \bullet 100 \tag{4.7}$$

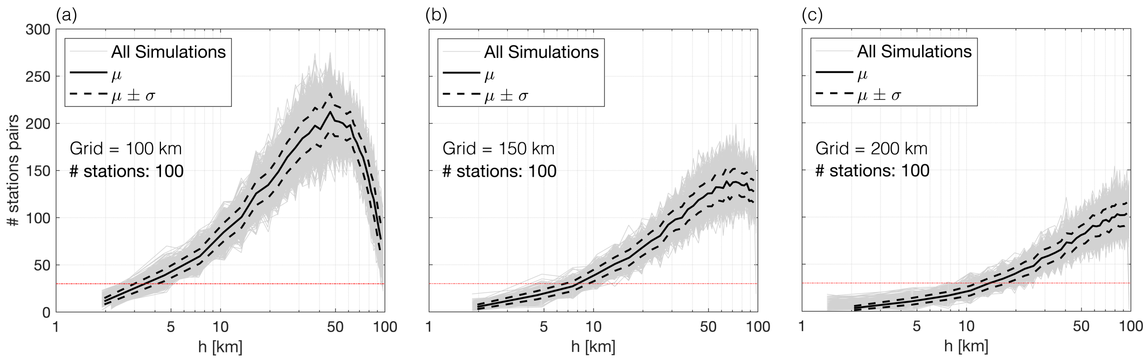


Figure 4.3: Number of stations pairs as a function of the separation distance for different grid sizes: (a) 100 km; (b) 150 km; (c) 200 km. The black solid line is the mean value computed over the 1000 simulations, whereas the black dashed lines represent the mean ± the standard deviation. The red dashed line indicate the minimum number (30) of station pairs required per bin to obtain more robust estimates (see Schiappapietra and Douglas (2020) for further details).

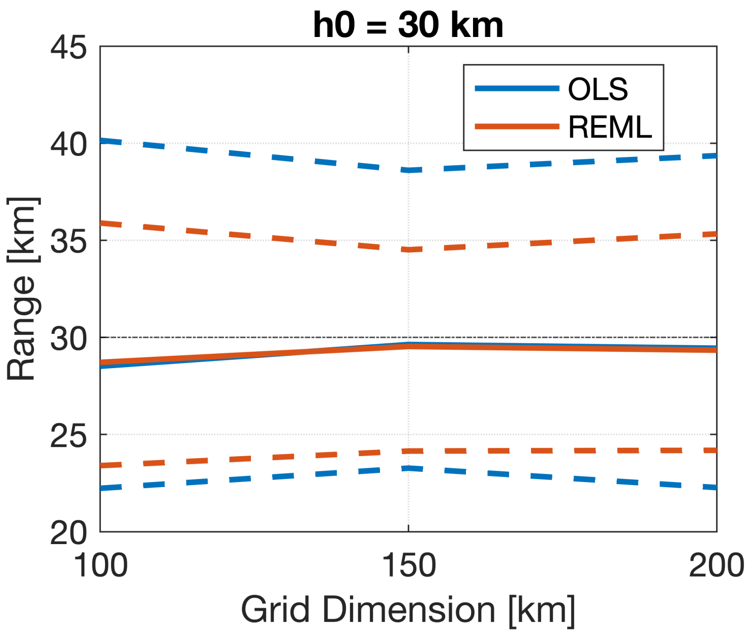


Figure 4.4: Range as a function of the grid dimension. Solid lines represent the median values computed over the 1000 simulations, whereas dashed lines indicate the first and third quartile. OLS and REML refer to the two different approaches we use to estimate the range. The black dotted line indicates the initial range value (30 km) imposed in the simulation.

$$IQR = Q_3 - Q_1 \tag{4.8}$$

where $\hat{h}$ is the range estimate for each simulated ground motion field and Q_3 and Q_1 are the third and first quartiles, respectively.

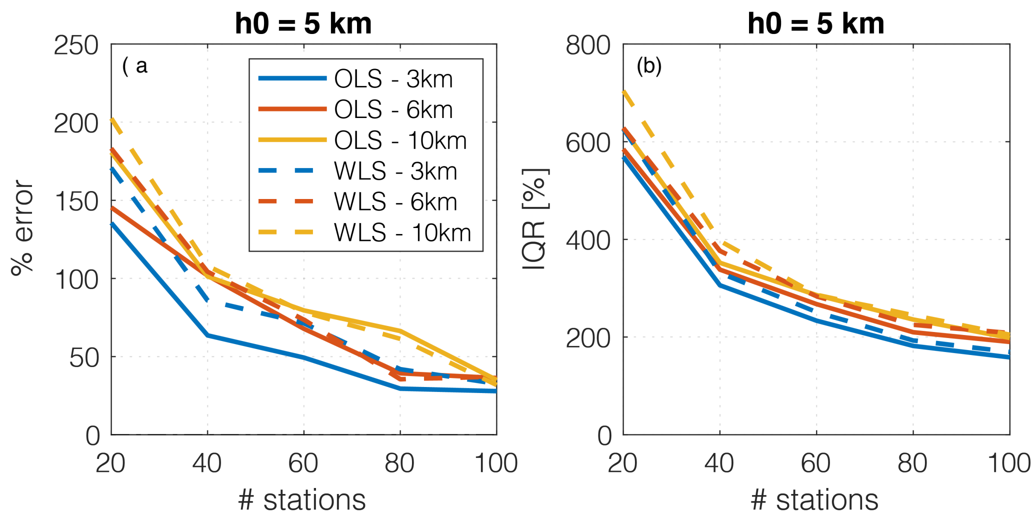


Figure 4.5: Bias median (a) and interquartile IQR (b) values as a function of the number of stations for different bin widths and for two different approaches (OLS, WLS). We impose an initial range value of 5 km.

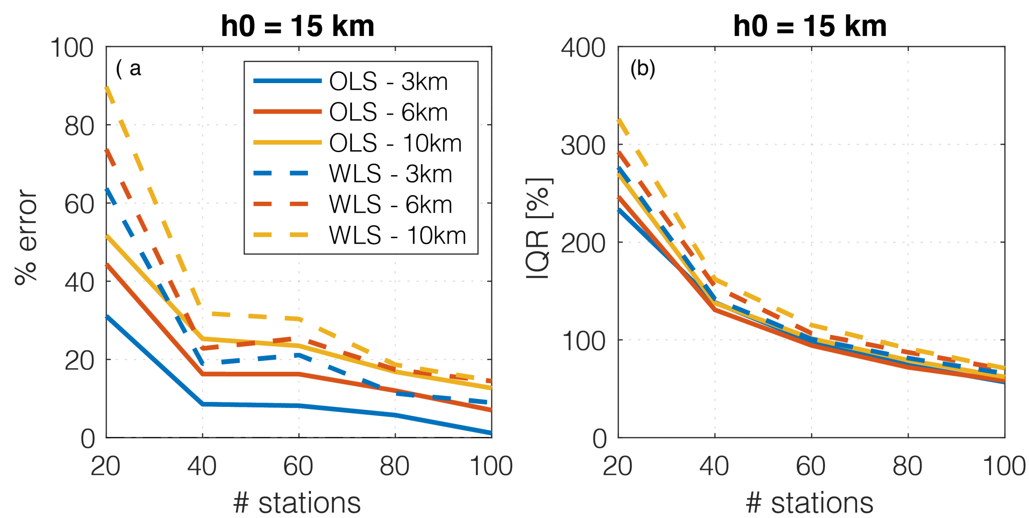


Figure 4.6: Bias median (a) and interquartile IQR (b) values as a function of the number of stations for different bin widths and for two different approaches (OLS, WLS). We impose an initial range value of 15 km. Note that the y axis scale is different to that used in Figure 4.5.

Independently of the technique employed to estimate the semivariogram coefficients (in this case, the range), the lower the bin size, the lower the bias and the variability are. Because the estimated range value tends to increase with wider bins, we believe that this apparent correlation comes exclusively from the bin size, which thus has a strong impact on the lower range that can be retrieved. Particularly, if the width is too wide, the correlation structure of ground-motion fields correlated over shorter distances may be masked. Consequently, we opt for a bin size of 3 km in the following analyses.

Such a strong dependency on bin size and other choices has motivated us to seek different approaches for the estimation of the correlation coefficients that do not depend on arbitrary parameters like the bin size, the distance cut-off and semivariogram estimator. Hereafter, we present a comparison of the outcomes obtained by means of the different techniques proposed in section 4.2, namely least-squares regression and the maximum-likelihood approach.

4.4.1 Least-squares regression Vs. Maximum-likelihood method

We carry out a comparison between range estimates obtained by means of least-squares regression (OLS and WLS) and maximum-likelihood methods (ML and REML). The results of this preliminary analysis are summarised in Figure 4.7. It is noted that for each pair of h_0 and number of stations, the bin size is the only varying parameter, so that the comparison is straightforward. Not only do the two maximum-likelihood approaches (ML and REML) provide the lowest bias and variability in the range estimates, but they also return the same outcomes regardless of the bin size. By contrast, OLS and WLS feature increasing median values as the bin becomes wider, as already shown in Figure 4.5 and Figure 4.6. We believe that such results are promising, since ML and REML do not add additional sources of uncertainty related to the choice of the bin width.

4.5 Dependence on the number of stations

To obtain robust estimates of the correlation structure from ground-motion observations, ideally one would use a large number of closely-spaced data. Although the number of earthquake recordings has dramatically increased over the last decades, seismic stations are not homogeneously distributed, making it difficult to assess the spatial correlation in regions characterized by sparse seismic networks. As already demonstrated in chapter 2 and by Baker and Chen (2020), the correlation estimation uncertainty is inversely correlated to the number of available stations. Here, we propose two different studies on the impact of the number of stations that consider both random station locations and station locations based

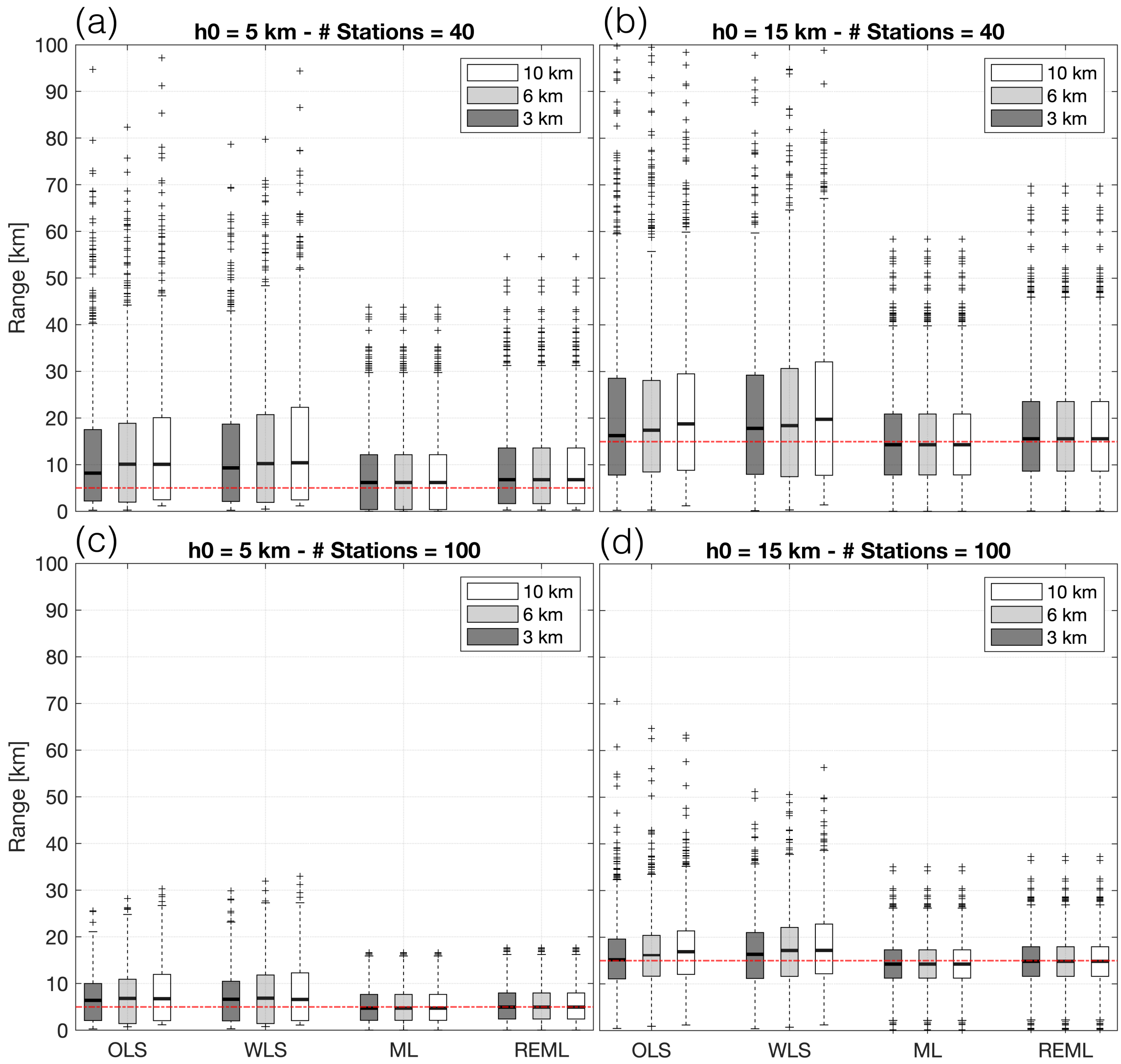


Figure 4.7: Boxplots for range estimates obtained by using different bin size, different h_0 (5 and 15 km) and different number of available stations (40 and 100).

on real networks. The main goal is to illustrate how the maximum-likelihood approaches outperform the least-squares methods especially in terms of estimation uncertainty.

4.5.1 Randomly simulated locations

We present the results of the simulations performed as described in section 4.3 for different values of the initial range h_0 [5, 10, 15, 20, 30, 40 km] and different number of stations [20, 40, 60, 80, 100]. Figure 4.8 and 4.9 illustrates the median value of the percentage error along with the IQR, which is taken as a measure of the dispersion of the range estimates.

Four main observations can be highlighted. First, independently of h_0, the maximum-likelihood approaches (ML and REML) generally show a lower dispersion compared to the least-squares methods (OLS and WLS). At the same time, ML and REML provide smaller median % error values, especially when few stations are available. This is a promising result, particularly for those regions characterized by fewer data. Second, median values tend to zero as the number of available stations increases. Similarly, the IQR decreases, halving its value as the number of stations rises from 20 to 100. Such outcomes corroborate the findings of Baker and Chen (2020) who demonstrated that the estimation uncertainty is larger for poorly-recorded events and that at least 100 stations are required to provide robust correlation estimates. Third, when a large number of stations is available both least-squares and maximum-likelihood approaches converge towards the same % error and IQR, independently of the h_0. Fourth, we observe that major differences exist among the proposed estimation techniques for smaller h_0, compared to the largest ones. We believe that smaller correlations are more difficult to detect as they require a large number of observations at very closely spaced stations. By contrast, earthquake ground motions are often recorded by a limited number of stations separated by many kilometres, with average inter-station distances in the range of 10-20 km, making it easier to measure ground motion correlated over larger distances. As a further evidence of such a trend, we plot in Figure 4.10 and Figure 4.11 the % error and the IQR as a function of h_0 for different number of stations considered. It is evident that all the approaches tend to a % error of zero when residuals are correlated over larger distances, independently of the number of stations. Besides, ML and REML provide smaller % error compared to OLS and WLS especially for poorly-recorded events and smaller h_0. Such behaviour is mainly due to the semivariogram computation, which is required in the least-squares approaches.

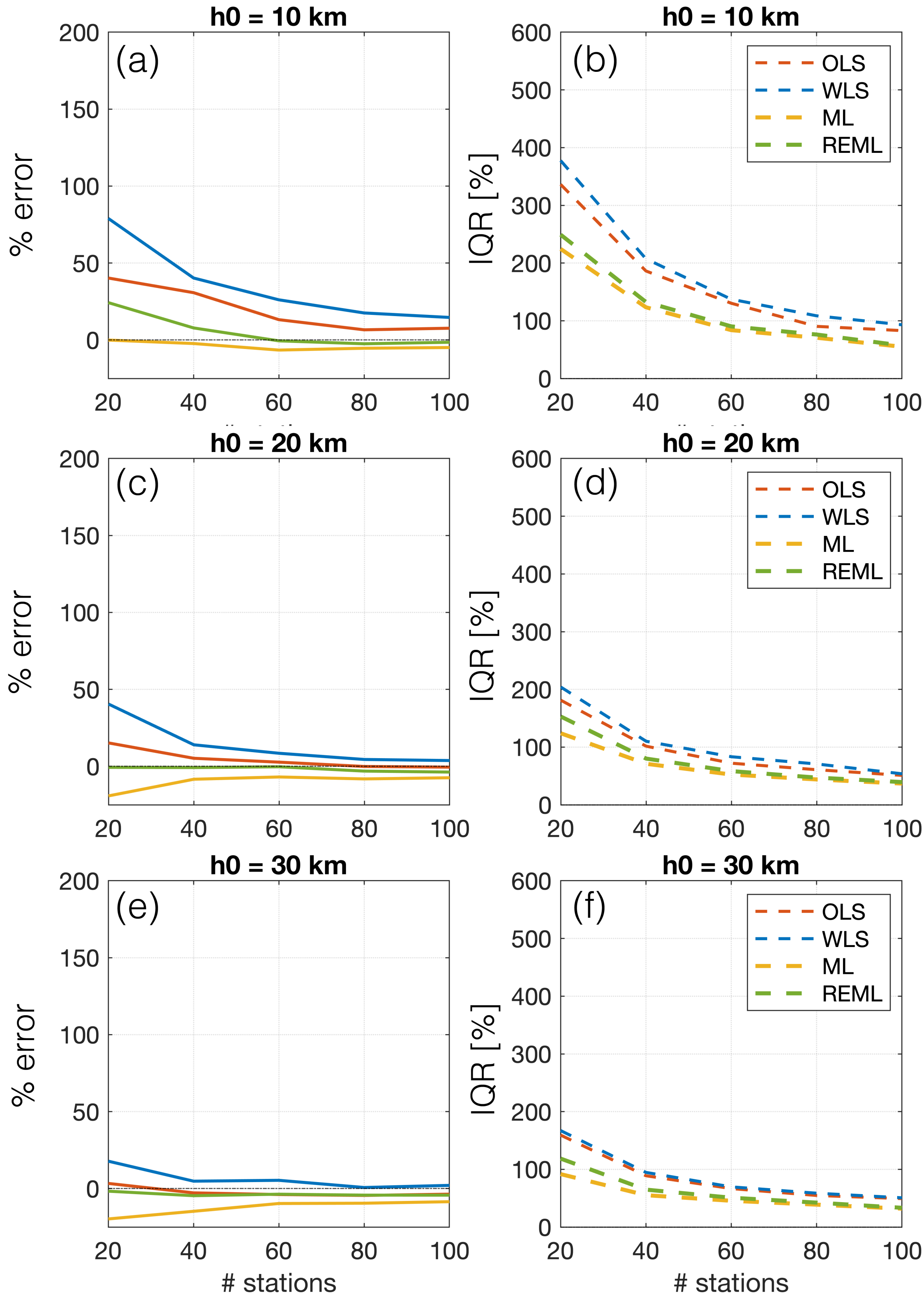


Figure 4.8: Bias (left panel) and Interquartile (right panel) range values as a function of the number of stations. The different rows refer to the different initial h_0 values. Different colours refer to the approaches employed in the correlation estimation.

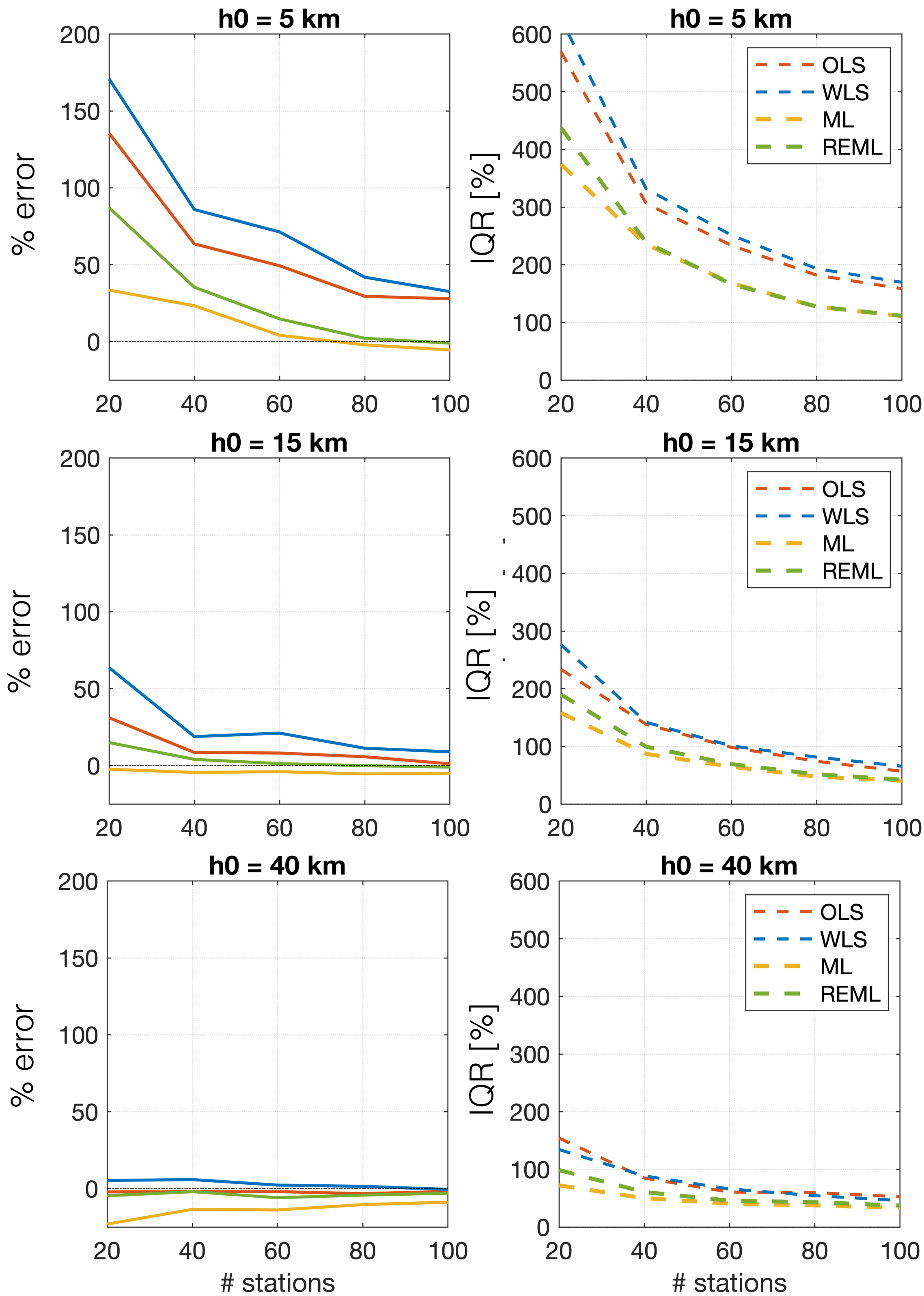


Figure 4.9: Bias (left panel) and Interquartile (right panel) range values as a function of the number of stations. The different rows refer to the different initial h_0 values. Different colours refer to the approaches employed in the correlation estimation.

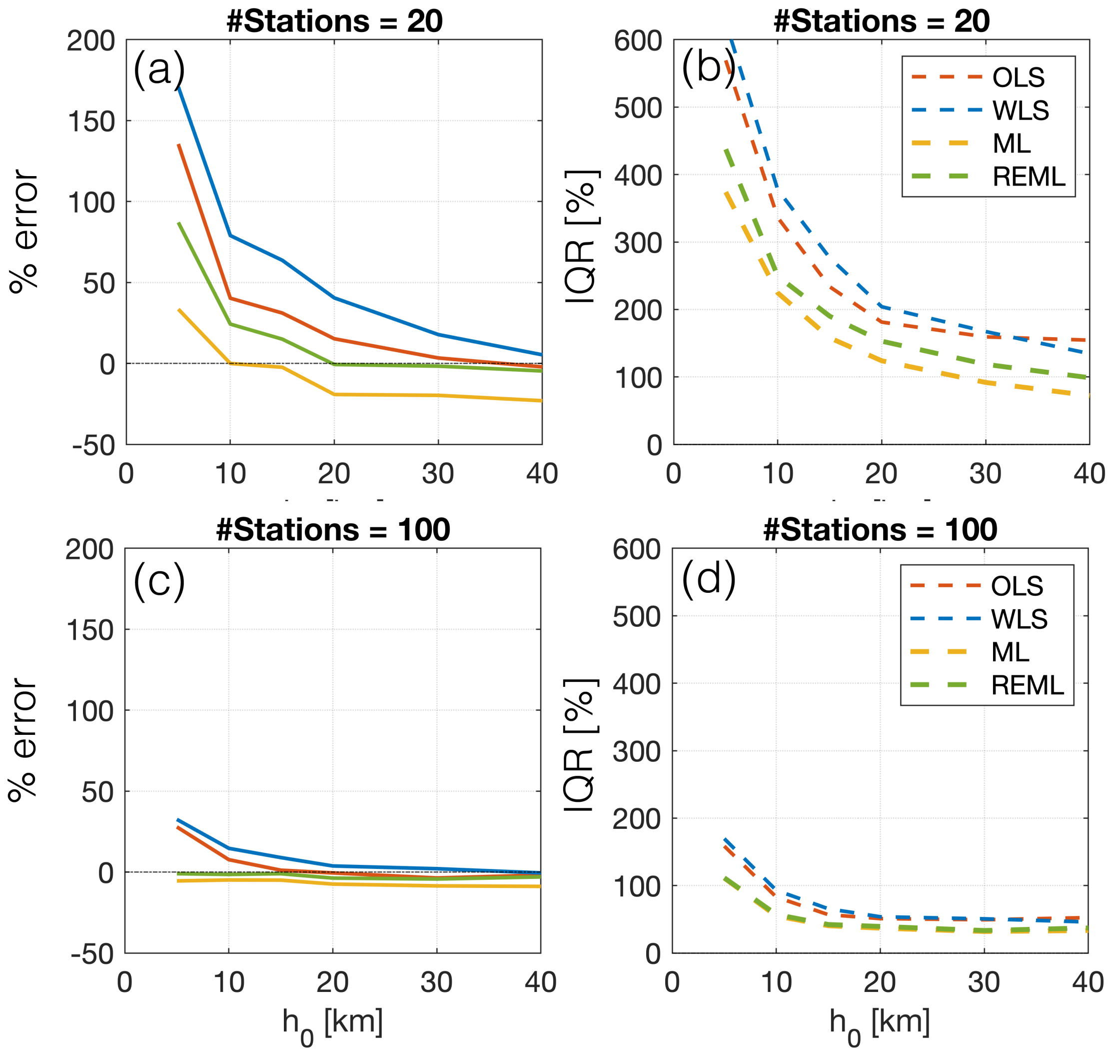


Figure 4.10: Bias (left panel) and Interquartile (right panel) range values as a function of h0. The different rows refer to the different number of stations considered. Different colours refer to the approaches employed in the correlation estimation.

Finally, we report in Figure 4.12 the boxplots of the range estimates for different number of stations and for a given h_0. Not only do the ML and REML feature narrower boxes (limited by the first and third quartiles, i.e. those used in the computations of IQR) compared to the OLS and WLS boxes, but they are also characterized by smaller whiskers and more confined outliers. The latter are defined as the estimate values that are larger than 1.5 times the IQR. Similar outcomes are obtained also for all other h_0. For the sake of brevity, we do not show all the figures here.

We recall that for an equal number of stations, the station layout is the same for different

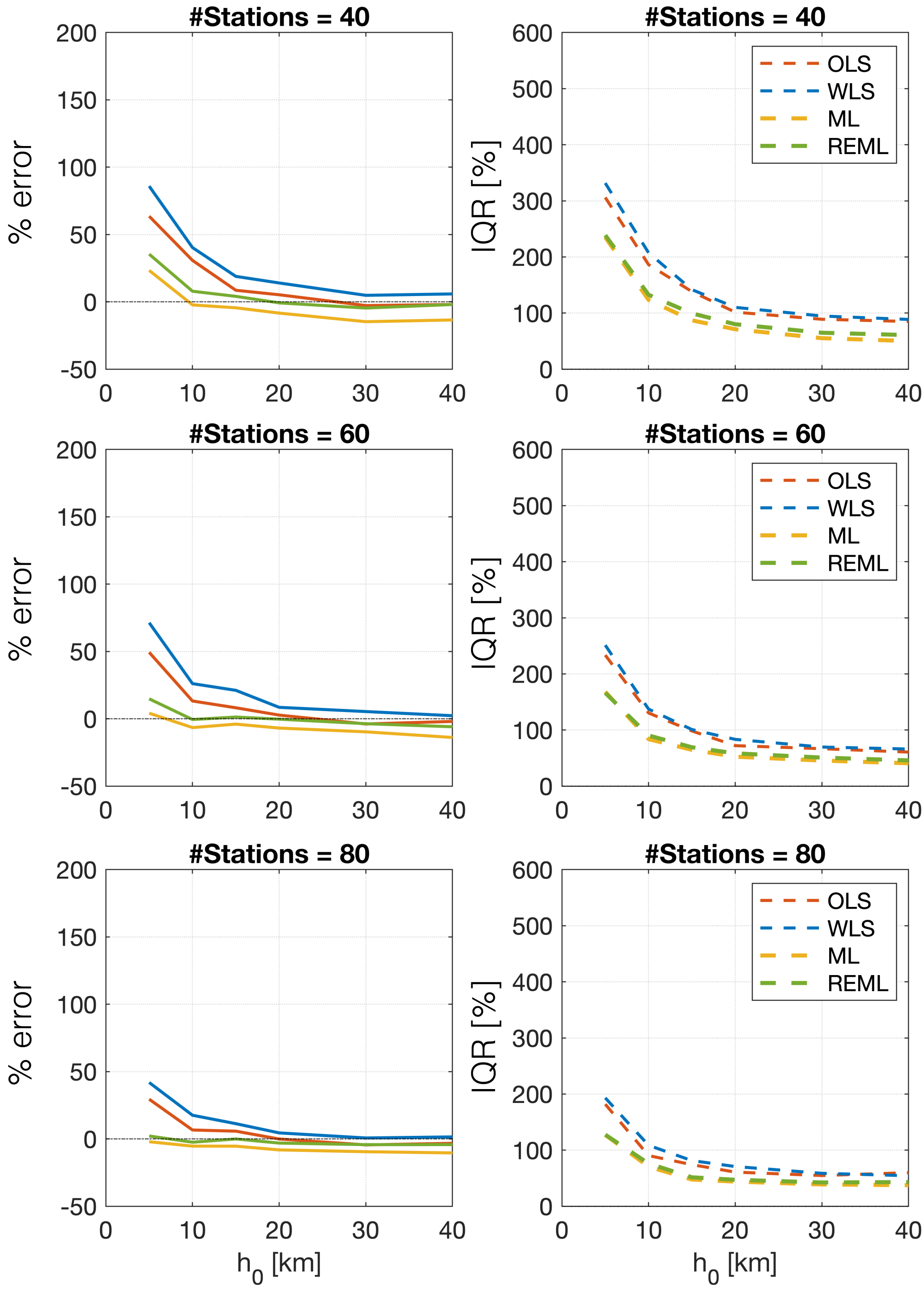


Figure 4.11: Bias (left panel) and Interquartile (right panel) range values as a function of the number of stations. The different rows refer to the different initial h_0 values. Different colours refer to the approaches employed in the correlation estimation.

h_0. Therefore, the variability among the different approaches only lies with the different correlation structure used in the simulation.

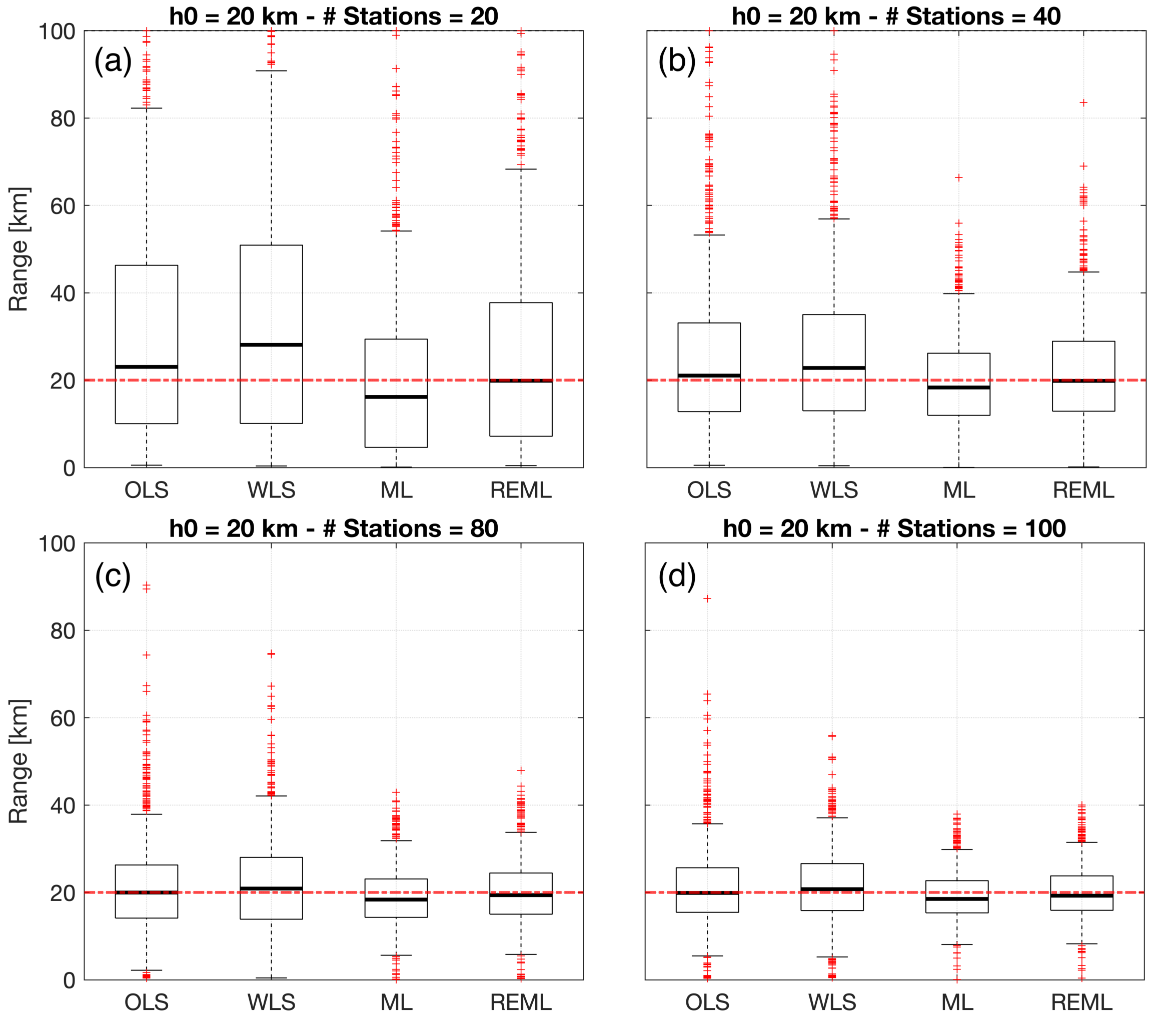


Figure 4.12: Boxplots of range estimates for different number of stations and h0 = 20 km: (a) 20 stations; (b) 40 stations; (c) 80 stations: (d) 100 stations

4.5.2 Station layouts of past earthquakes

We perform similar Monte Carlo simulations to those presented in section 4.5.1, but here the ground-motion fields are simulated only at stations that recorded past earthquakes. We take four different stations layouts as references, corresponding to four events selected within both the ESM (Lanzano, Sgobba, Luzi, Puglia, Pacor, Felicetta, D'Amico, Cotton and Bindi, 2019) and NGA-West2 (Ancheta et al., 2014) strong-motion flat-files. We consider only well-recorded earthquakes with more than 100 observations within 100 km from the epicentre. Figure 4.13 show the four station layouts: (1) ESM1 is the M_w 6.0 29^{th} May 2012 Emilia (Italy) event; (2) ESM2 is an M_w 4.3 event that occurred on 23^{rd} September 2016 in Central Italy; (3) NGA1 is the M_w 6.9 13^{th} June 2008 Iwate (Japan) event; NGA2 is a M_w 4.7 event that occurred on 18^{th} May 2009 in California. We believe such layouts are a good sample of the type of station distributions often seen in practice. We adopt a Monte Carlo approach, generating 1000 simulations of ground motion residuals for each h_0 [10, 20, 30, 40] and for different numbers of stations [20, 40, 60, 80, 100], which are randomly selected within each configuration. We note that for each correlation structure (e.g. each h_0), the only varying parameter is the number of selected stations, so that the comparison is not affected by other factors. We also imposed the same seed, which sets the starting number used to generate a sequence of random numbers. Hence, the underlying residuals distribution is the same, although the station configurations are different in each layout. This allows us to isolate the effect of the station layout.

Figure 4.14 and Figure 4.15 compare the median % error and the IQR computed for the different station layouts and $h_0 = 10km$ by using the different methodologies proposed in this study. Three main observations can be highlighted. Generally, ML and REML have the smallest median % error and they lead to lower uncertainties compared to the OLS and WLS approaches, independently of the correlation structure imposed. Such outcomes agree with the findings presented in section 4.5.1, where residuals are simulated on a fine grid and stations are randomly selected. Similar conclusions can be drawn for the other h_0 values (20, 30, 40 km), as it can be observed from Figure 4.16 to Figure 4.21.

Furthermore, ML and REML feature similar values both in terms of median and variability among the different station layouts. By contrast, OLS and WLS show a strong variability among the four layouts. This applies to all h_0 values, although differences are less pronounced for ground motions correlated over larger distances (i.e. higher h_0). We believe that such behaviour is mainly due to the semivariogram computation, whose robustness depends on both the bin size, as demonstrated in section 4.4, and the number of stations within each bin. These parameters are strictly related to the station layout so that more homogeneous station distributions would provide more reliable range estimates.

Finally, we note that generally the NGA2 configuration has the lowest median % error and IQR among all the considered layouts. We believe that this result lies with the more homogeneous and denser distribution of stations with respect to the other three layouts, which leads to more accurate range estimates. To demonstrate this, we plot in

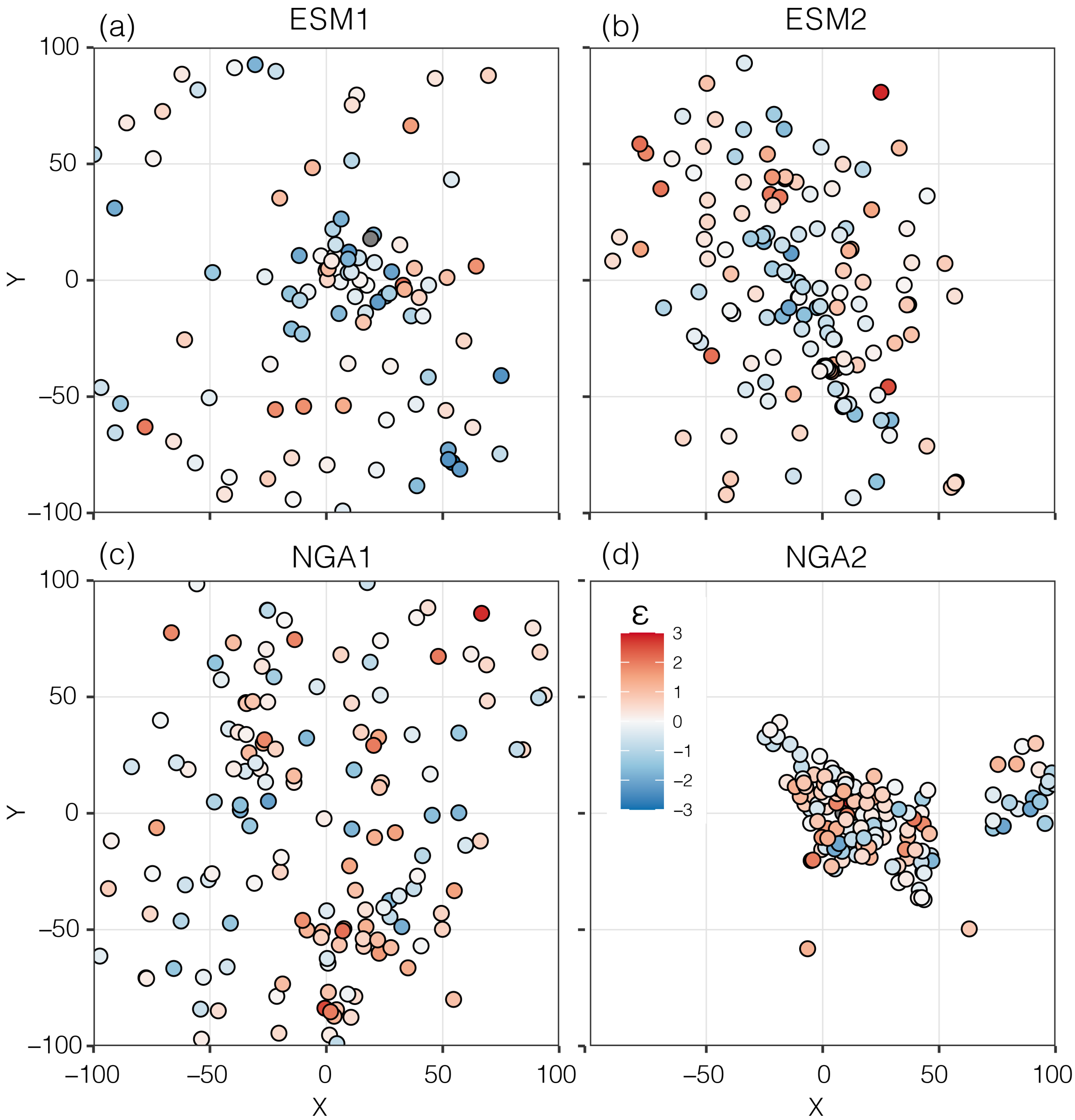


Figure 4.13: Station layouts for the four selected events: (a) ESM1 - event id IT-2012-0011; (b) ESM2 - event id ESMC-20160903_0000063; (c) NGA1 - event id 279; (d) NGA2 - event id 1011. Dots are colour-coded based on the within-event residuals (one out 1000 simulated ground-motion fields). Coordinates are normalized with respect of the epicentre of each event.

Figure 4.22 the number of station pairs within 12 km as a function of the number of stations. Systematically, the NGA2 configuration has the largest number of pairs in the first four bins ($12km/3km = 4$), independently of the number of available stations. Conversely, the NGA1 layout has the lowest number of pairs and as a consequence it features the highest median %

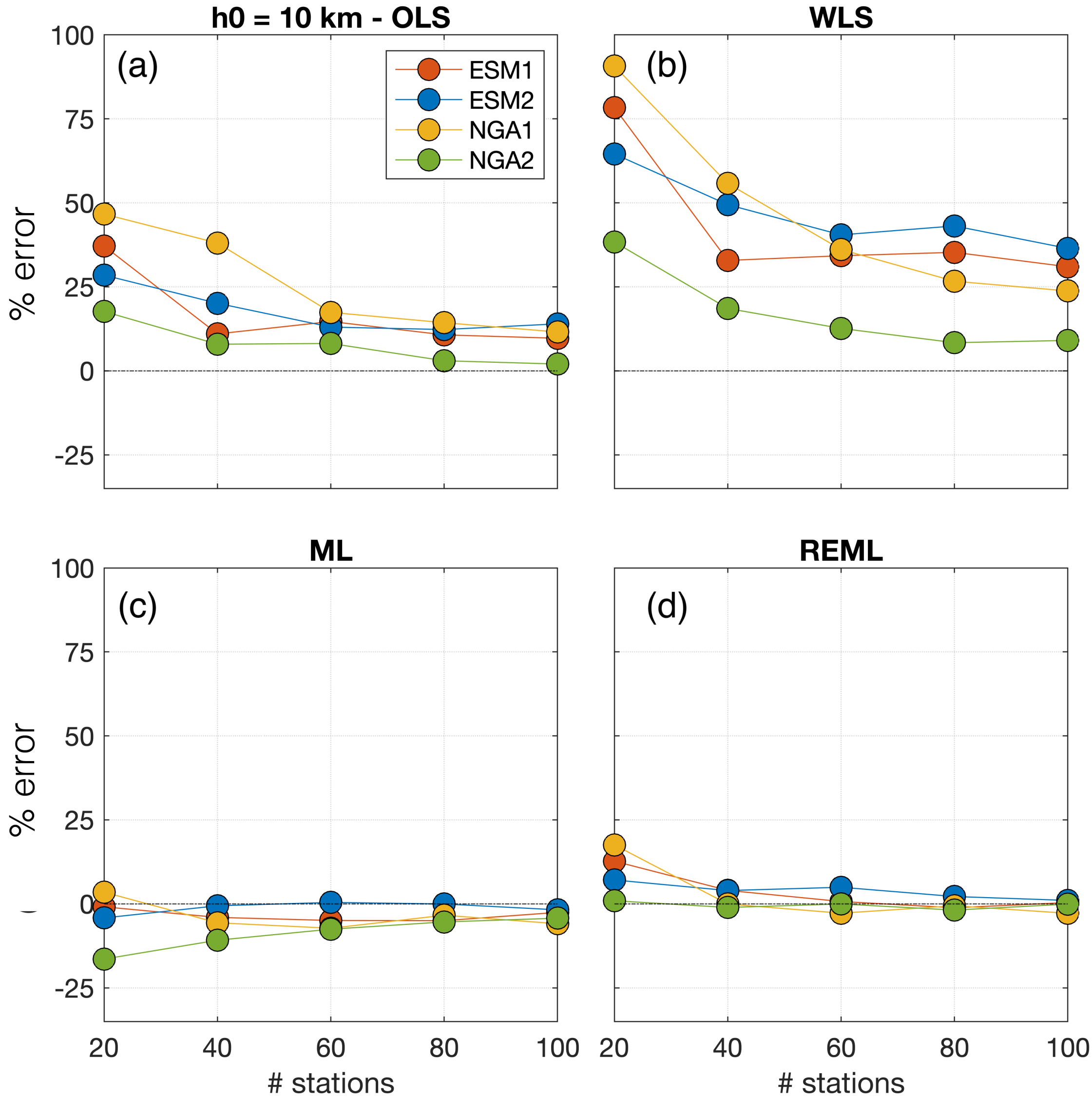


Figure 4.14: Median % error of the range estimates for the four different stations layouts as a function of the number of stations. (a) OLS; (b) WLS; (c) ML; (d) REML. The initial value of the range is set to 10 km.

error and IQR, especially when lower h_0 are considered. Such observations apply to the OLS and WLS approaches, whereas the ML and REML are not strongly affected by the different station layouts. This is a promising outcome and demonstrates how maximum-likelihood methods may outperform the least-squares approaches.

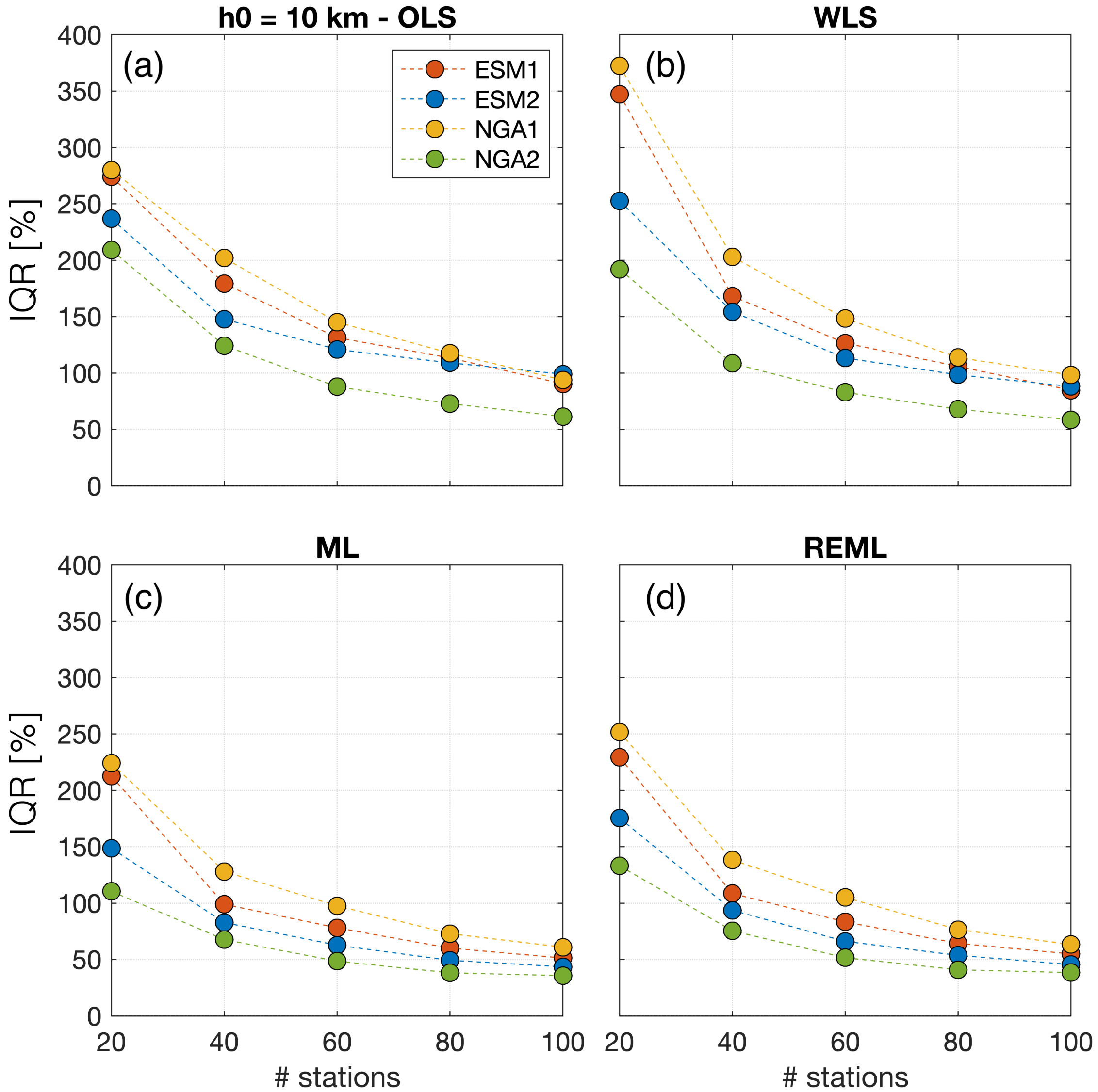


Figure 4.15: IQR of the range estimates for the four different station layouts as a function of the number of stations. (a) OLS; (b) WLS; (c) ML; (d) REML. The initial value of the range is set to 10 km.

4.6 Discussion

In this study, we show that there is uncertainty in modelling the spatial correlation and the size of this uncertainty depends on the availability of data as well as the derivation technique. Specifically, spatial correlation models for areas with limited data (e.g. regions without dense strong-motion networks and/or low seismicity) are more uncertain than those

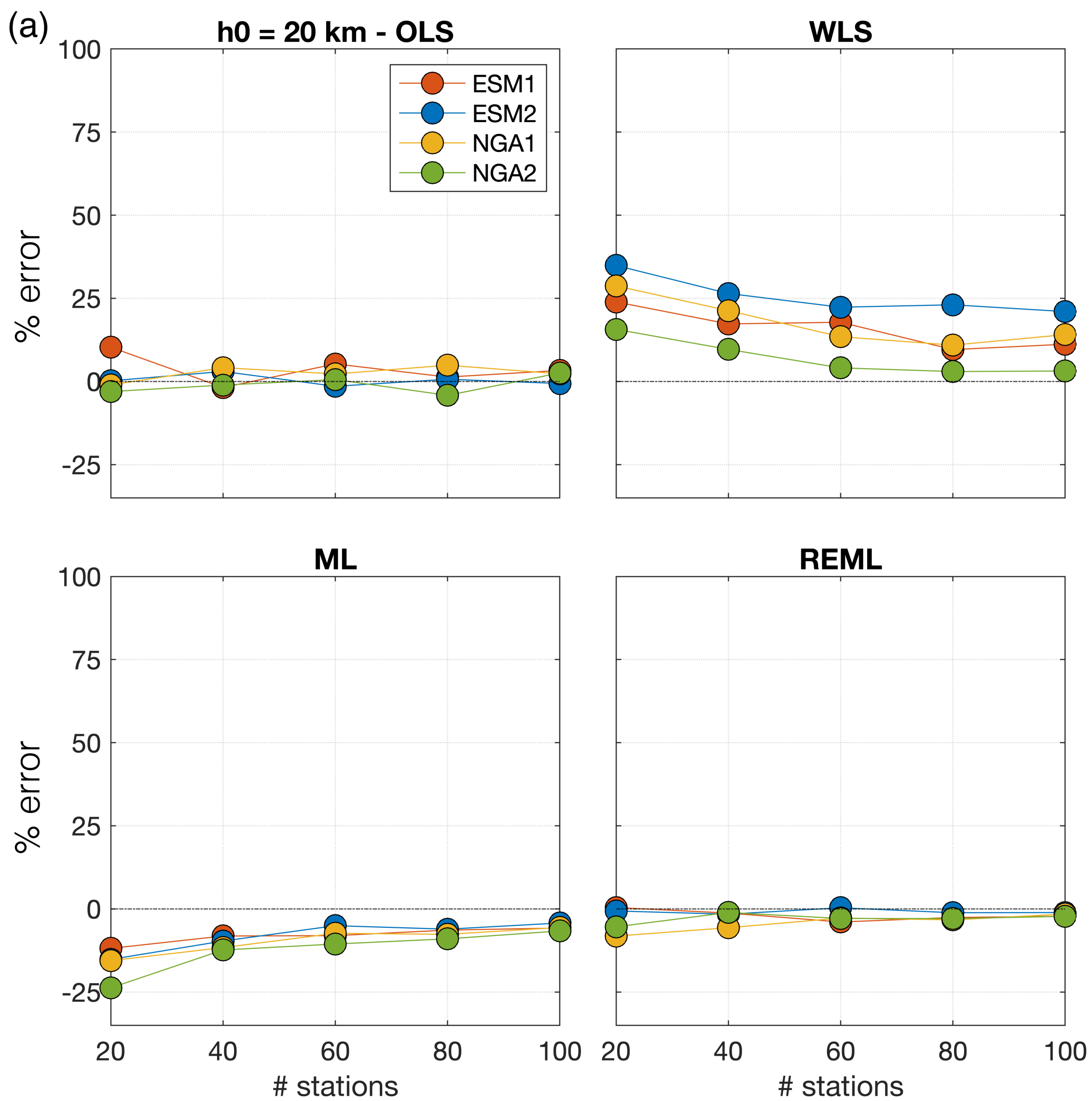


Figure 4.16: Median % error of the range estimates for the four different stations layouts as a function of the number of stations.. The initial value of the range is set to 20 km. We report the results obtained by using all the approaches proposed in this study (top: OLS and WLS; bottom: ML and REML).

with extensive observations. Consequently, regional seismic hazard models should account not only for the spatial correlation, but they should also capture its associated uncertainty, which depends on the region.

This regional-dependent uncertainty leads to the following two questions. Which correlation model or models should we use for regions with sparse observations? Is a

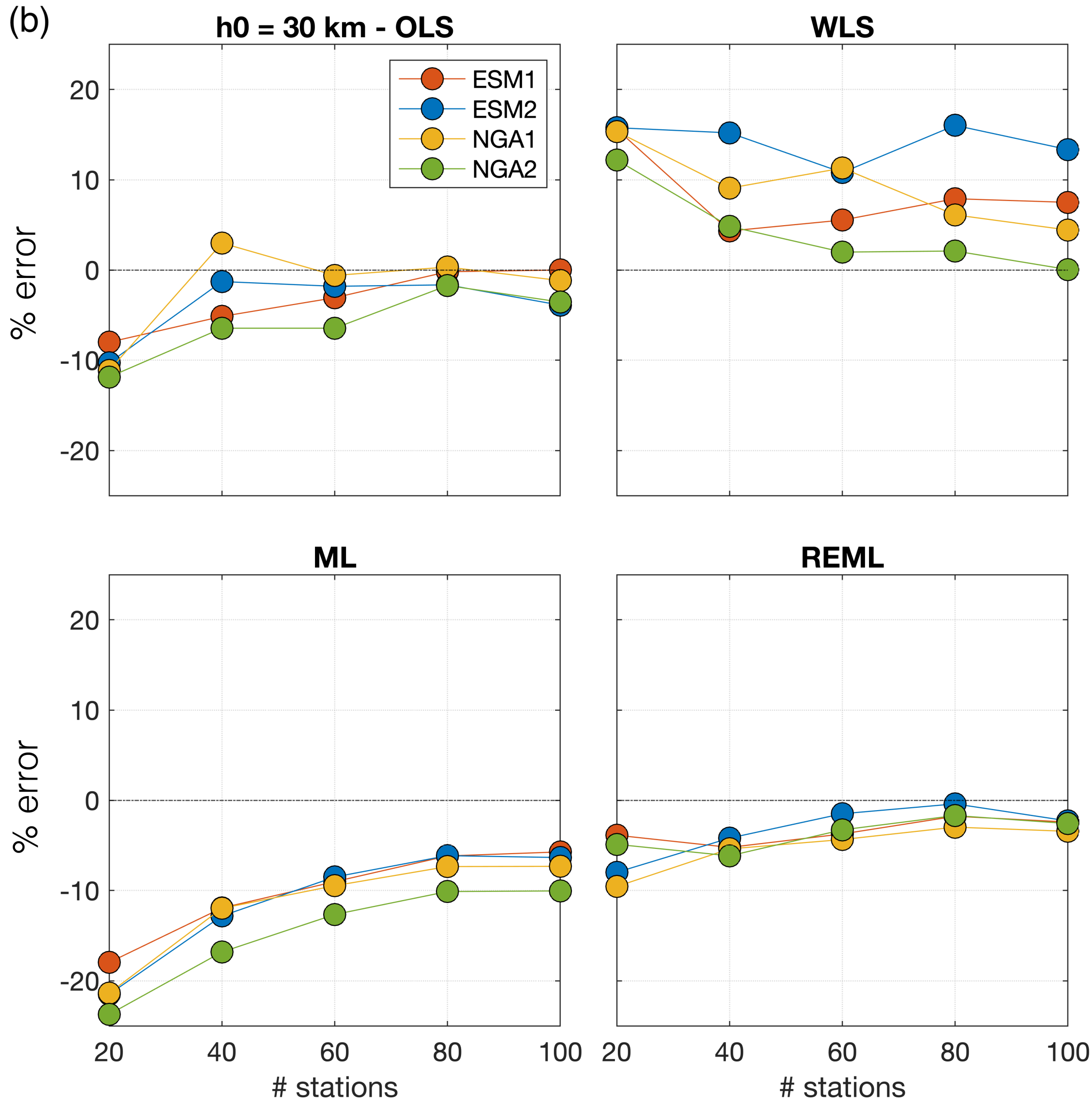


Figure 4.17: Median % error of the range estimates for the four different stations layouts as a function of the number of stations.. The initial value of the range is set to 30 km. We report the results obtained by using all the approaches proposed in this study (top: OLS and WLS; bottom: ML and REML).

global model truly able to capture the correlation of that specific area? These are similar questions faced by hazard analysts concerning the selection, modification or development of ground-motion models (e.g. Douglas, 2018).

Here, we propose some guidelines to model the spatial correlation uncertainty based on the availability of recordings. In particular, we propose a logic tree with symmetrical lower,

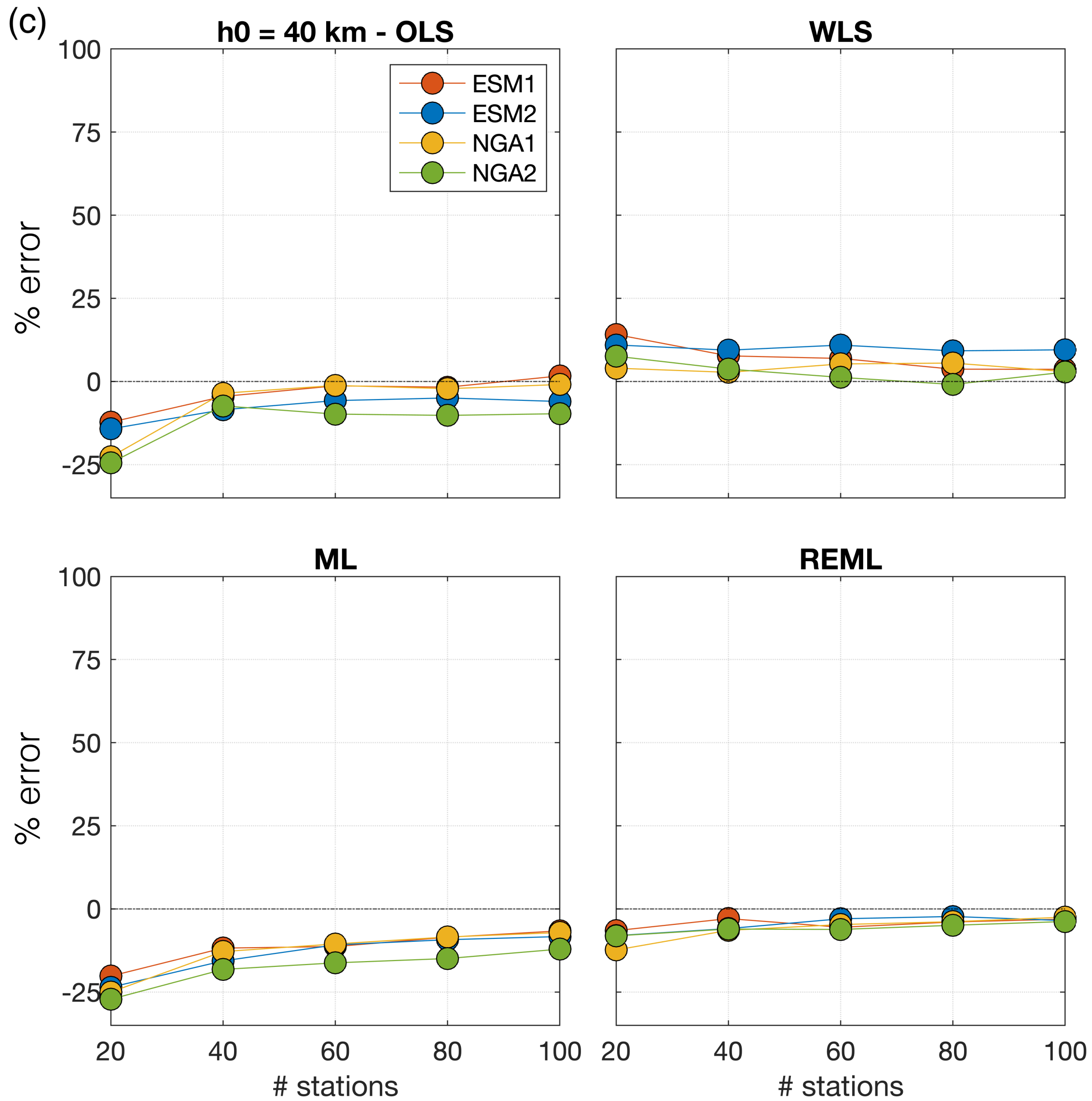


Figure 4.18: Median % error of the range estimates for the four different stations layouts as a function of the number of stations.. The initial value of the range is set to 40 km. We report the results obtained by using all the approaches proposed in this study (top: OLS and WLS; bottom: ML and REML).

middle and upper branches using a standard three-point distribution with weights equal to 0.185, 0.63 and 0.185, respectively (Keefer and Bodily, 1983). Table 4.1 reports the 5%, 50% and 95% percentile values of the range estimates as a function of both the number of stations and h_0 for the different approaches used in this study. This logic tree should capture the spread in the correlation estimates, thus leading to a more informed seismic

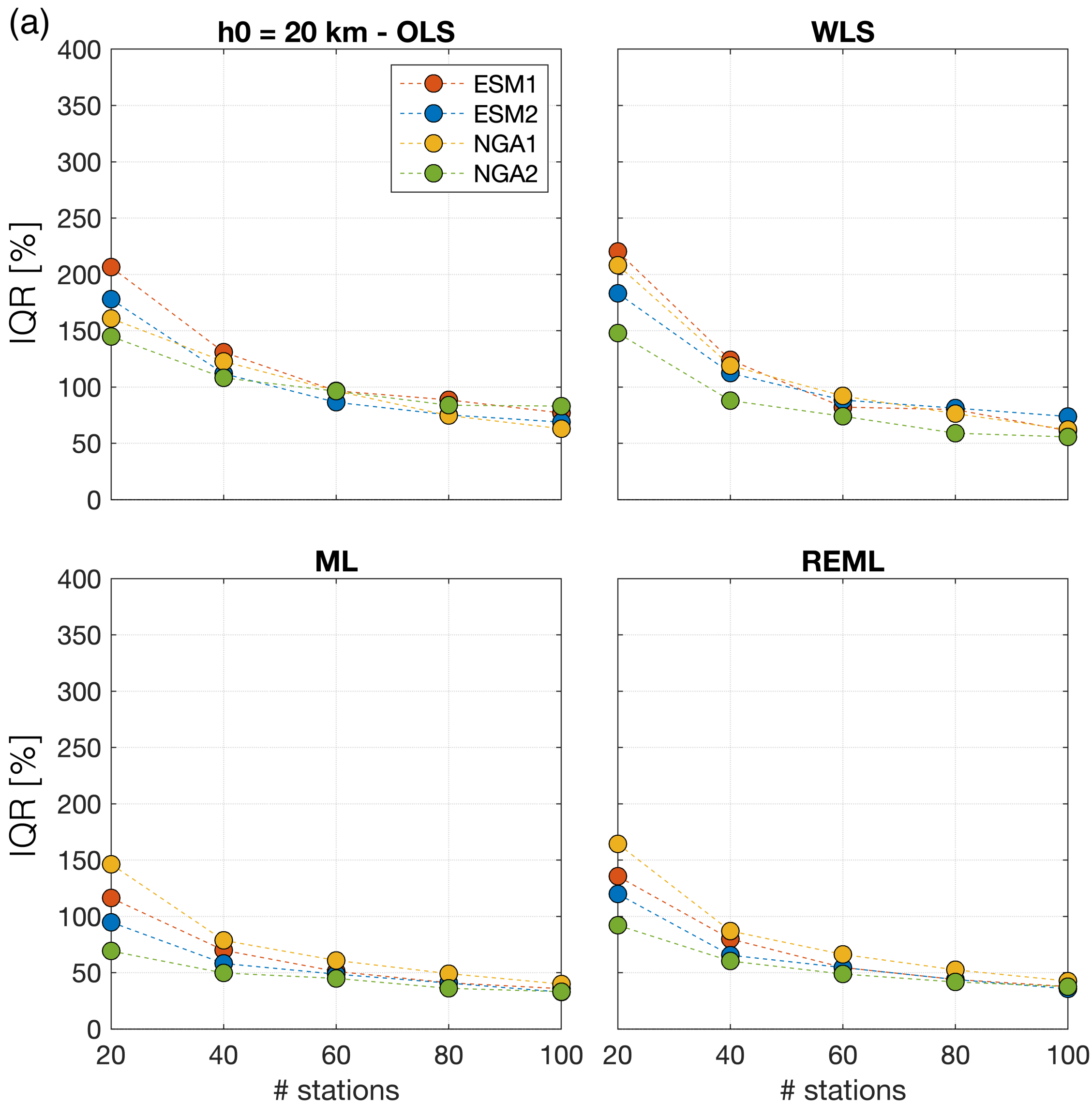


Figure 4.19: IQR of the range estimates for the four different stations layouts as a function of the number of stations. The initial value of the range is set to 20 km. We report the results obtained by using all the approaches proposed in this study (top: OLS and WLS; bottom: ML and REML)

risk assessment.

We are aware that this table does not cover all possible stations$-h_0$ combinations, but its aim is to provide a first-order estimate of the spatial correlation uncertainty that one should consider when modelling spatial correlations for both regions where observations are abundant and for those characterized by sparse recordings.

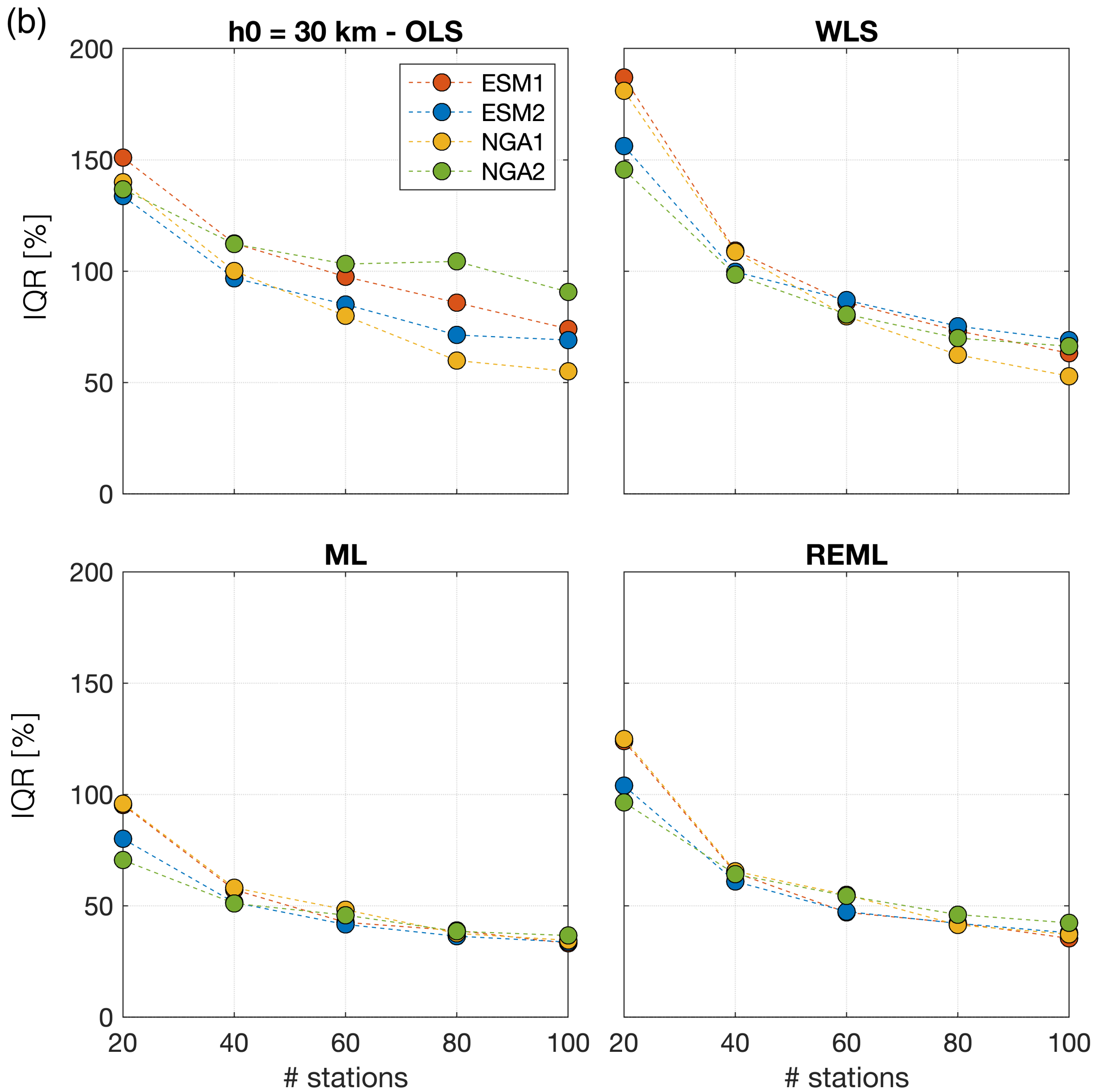


Figure 4.20: IQR of the range estimates for the four different stations layouts as a function of the number of stations. The initial value of the range is set to 30 km. We report the results obtained by using all the approaches proposed in this study (top: OLS and WLS; bottom: ML and REML)

Table 4.1: 5%, 50% and 95% percentile values of the range estimates for different h0 and different number of stations for the approaches used in this study. Estimates are in km.

		OLS			WLS			ML			REML		
h0 [km]	# Stations	5%	50%	95%	5%	50%	95%	5%	50%	95%	5%	50%	95%
5 km	20	1.3	11.8	131.7	1.2	13.5	91.0	0.2	6.7	46.5	0.9	9.4	70.6

	40	0.8	8.2	39.5	0.8	9.3	35.9	0.1	6.2	24.9	0.6	6.8	27.9
	60	0.6	7.5	26.2	0.6	8.6	26.9	0.1	5.2	18.8	0.5	5.7	21.1
	80	0.6	6.5	19.7	0.5	7.1	20.1	0.1	4.9	13.7	0.4	5.1	14.4
	100	0.6	6.4	16.3	0.5	6.6	16.6	0.1	4.7	12.0	0.4	5.0	12.6
10 km	20	1.6	14.0	319.1	1.3	17.9	154.4	0.2	10.0	48.7	1.0	12.4	65.8
	40	1.1	13.1	60.3	0.8	14.0	50.7	0.2	9.8	28.7	0.7	10.8	32.4
	60	0.8	11.3	34.3	0.7	12.6	33.1	0.2	9.3	21.6	0.8	9.9	23.5
	80	0.7	10.7	24.2	0.7	11.8	26.1	0.2	9.5	18.1	0.8	9.8	19.2
	100	0.8	10.8	22.6	0.8	11.5	24.0	2.5	9.5	17.1	2.5	9.9	17.8
15 km	20	1.6	19.7	196.5	1.4	24.6	121.6	0.3	14.6	56.5	1.1	17.2	80.4
	40	1.3	16.3	64.8	1.1	17.8	55.7	0.2	14.3	34.3	1.0	15.6	39.6
	60	1.1	16.2	41.0	0.9	18.2	41.3	1.9	14.4	28.0	2.2	15.2	30.9
	80	4.1	15.9	33.2	1.4	16.7	33.5	5.3	14.2	25.1	5.8	15.0	26.7
	100	5.2	15.2	29.0	3.4	16.3	29.1	7.2	14.3	22.5	7.4	14.8	23.9
20 km	20	2.1	23.0	175.3	1.6	28.1	110.3	0.3	16.2	53.8	1.5	19.9	74.9
	40	1.9	21.0	67.2	1.4	22.8	59.0	0.4	18.3	38.5	1.5	19.8	43.9
	60	4.9	20.5	49.2	1.7	21.7	47.3	7.2	18.6	33.7	7.7	19.9	37.6
	80	6.3	20.0	41.4	3.6	20.9	41.0	8.3	18.4	30.9	8.4	19.4	33.1
	100	9.1	19.9	36.2	6.0	20.8	36.3	11.1	18.5	28.8	11.4	19.3	30.4
30 km	20	2.7	31.0	340.9	1.7	35.3	148.3	0.4	24.1	67.4	1.8	29.5	103.3
	40	5.3	29.2	93.6	1.6	31.4	79.9	5.6	25.6	52.7	6.5	28.6	61.6
	60	10.8	28.8	70.6	7.9	31.6	66.7	11.9	27.1	47.2	12.3	28.9	53.1
	80	13.2	28.7	59.5	11.4	30.2	55.7	15.6	27.1	44.2	16.5	28.7	48.2
	100	15.4	28.9	53.8	13.4	30.6	53.7	16.7	27.4	41.6	17.5	28.7	45.1
40 km	20	3.6	39.1	578.8	2.4	42.1	207.2	0.5	30.8	78.8	2.2	38.1	123.7
	40	10.2	39.2	144.2	3.1	42.3	104.7	12.8	34.6	65.4	15.1	39.2	79.9
	60	16.7	39.2	105.4	14.7	40.9	84.9	18.9	34.4	58.8	20.1	37.6	67.2
	80	19.8	38.7	94.2	18.4	40.6	77.7	20.9	35.8	58.8	21.9	38.3	65.6
	100	21.8	39.3	80.4	20.3	39.8	72.4	24.1	36.5	57.5	25.1	38.8	63.4

We show here an example in which we compute the spatial correlation for six different earthquakes recorded by 40 (3 events) and 80 (3 events) stations. The events are selected within the ESM (Lanzano, Sgobba, Luzi, Puglia, Pacor, Felicetta, D'Amico, Cotton and Bindi, 2019) flat-file and we use the OLS and REML approaches to estimate the range. We note that the scope of this section is to discuss correlation uncertainty and not the optimal method that best estimates the range.

Figure 4.23 presents the experimental semivariograms and the theoretical models computed on the residuals of three different events recorded by 40 stations. If we look at the REML estimates, the three events feature ranges equal to 41, 15 and 5.5 km. Based on our simulations, such median estimates have the following confidence intervals, given the number of available stations: (1) [15 - 80] km, (2) [1 - 40] km, and (3) [0.6 - 28] km. At the same time, OLS provides median estimates equal to 21, 16.1 and 12.1 km, respectively, which correspond to the following confidence intervals: (1) [1.9 - 67.2] km, (2) [1.3 - 64.8] km, and (3) [1.1 - 60.3] km.

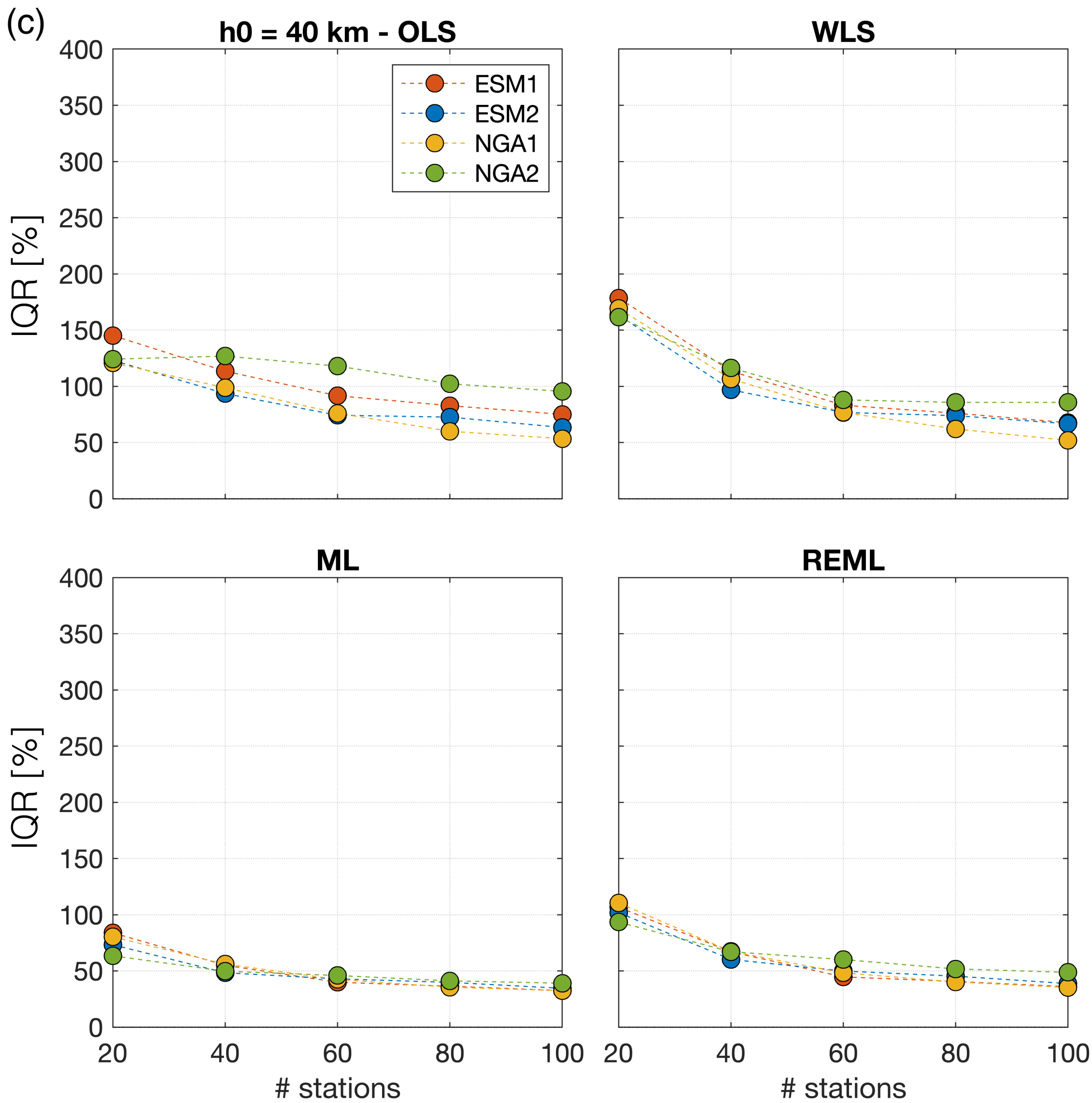


Figure 4.21: IQR of the range estimates for the four different stations layouts as a function of the number of stations. The initial value of the range is set to 40 km. We report the results obtained by using all the approaches proposed in this study (top: OLS and WLS; bottom: ML and REML)

Baker and Chen (2020) provide estimates of the correlation computed for a set of events within the NGA-West2 database (Ancheta et al., 2014) and the corresponding number of stations that recorded each event. For instance, the M_w 6.5 Big Bear-01 (1992) event has a range equal to 15.7 km computed based on 45 observations. According to Table 4.1 (WLS, h0 = 15 km, 40 stations), the range estimate is within the [1.1 - 55.7] km confidence

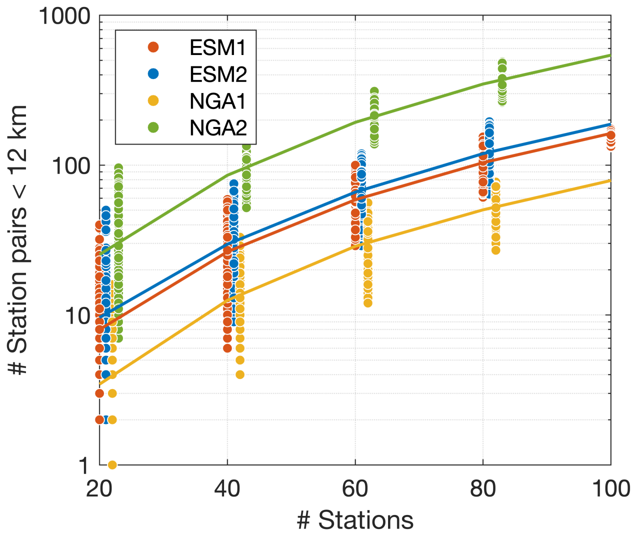


Figure 4.22: Number of stations pairs within 12 km as a function of the number of stations for the four station layouts. The dots indicate the number of stations pairs in each simulation, whereas the solid lines represent the average number of pairs as a function of the number of stations.

interval. Analogously, the M_w 6.2 Christchurch (2011) event features a range of 18.9 km evaluated on the basis of 80 stations. The corresponding confidence interval from Table 4.1 is [3.6 - 41] km. These confidence intervals appear wide but it should be recalled that these are for the 5_{th} to 95_{th} percentiles.

While the median range estimates of the considered events differ from each other, the confidence intervals overlap, suggesting that these earthquakes may, in fact, have similar correlation structures. Similar conclusions can also be drawn from Figure 4.24, where we show the results for three different events recorded by 80 stations.

These findings, while preliminary, may help us to answer the second question posed at the beginning of this section about global versus local models. A global model calibrated on data from multiple events may be suitable to describe the correlation structure of a region for which observations are currently sparse since denser datasets would provide more constrained range estimates. On the other hand, pooling data from multiple earthquakes would not enable the study of the correlation due to a particular event to be investigated and therefore it would increase the uncertainty in the estimates related to the specificity of the region of interest.

While investigating the estimation uncertainty related to a single event, in this study we have not explored multiple earthquakes uncertainties. Heresi and Miranda (2019) computed the central tendency and the dispersion of the range values from well-recorded NGA-West2 events. For instance, the spectral acceleration at T = 0.1 shows an average range of 14.3 km and a standard deviation computed on the natural logarithm of the range values equal to 0.83. We performed similar analysis for the Central Italy region, and we obtained an average range of 27.8 km and a standard deviation equal to 0.75 for the same IM. While these studies account for the event-to-event dispersion, they do not investigate the estimation uncertainty. Therefore, further work is then required to establish the fraction of uncertainty which derives from pooling data from different events.

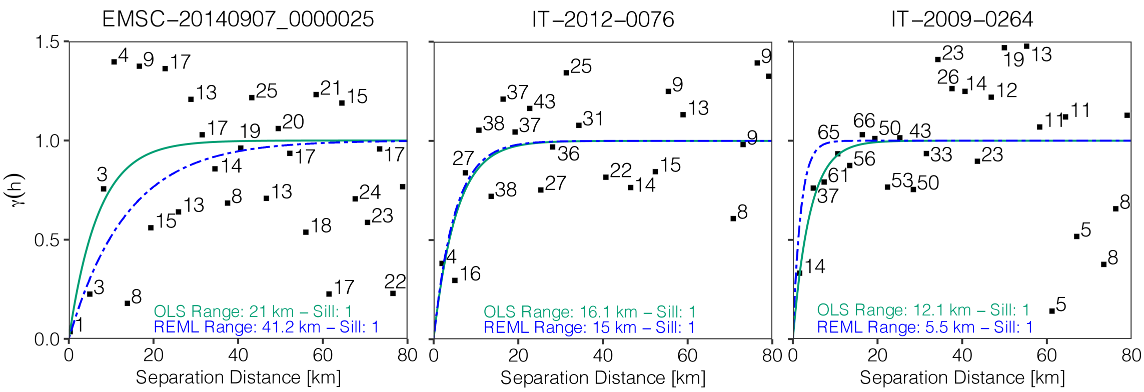


Figure 4.23: Experimental semivariograms and theoretical models for three different events recorded by 40 stations. The number of pairs in each bin is reported close to each semivariance estimate (black dots).

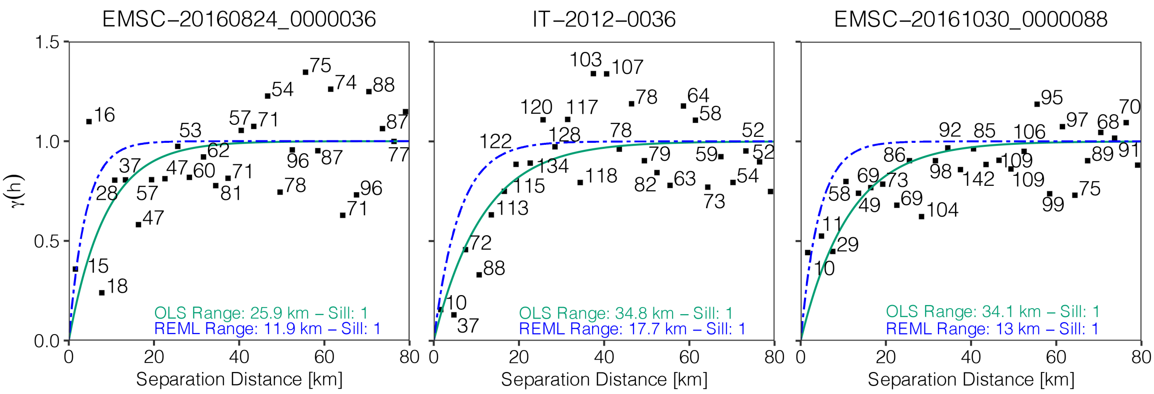


Figure 4.24: Experimental semivariograms and theoretical models for three different events recorded by 80 stations. The number of pairs in each bin is reported close to each semivariance estimate (black dots).

Finally, we report in Figure 4.25 a scheme showing the application of the proposed logic tree for a median $h_0 = 20$ km. We consider the estimates obtained by using the REML approach when 60 stations are available. Based on the correlation estimates in Table 4.1, we simulate n spatially correlated random fields for each range to use in conjunction with predicted median ground motions. The different branches (after assessing the risk, e.g. after estimating the corresponding losses for each range) would then be averaged with weights equal to 0.185, 0.63 and 0.185, respectively, to obtain the final risk assessment (e.g. resulting losses) which account for the uncertainties in the correlation estimates.

4.7 Conclusions

In this work, we introduce alternative methods to classic least-squares regression for the estimation of the correlation structure of earthquake ground motions. In particular, we employ two maximum likelihood-based approaches, namely the Gaussian maximum likelihood (ML) and the restricted maximum likelihood (REML). These have gained increasing importance in geostatistics, particularly when spatial trends exist in the data (Diggle et al., 2007; Oliver and Webster, 2014; Li et al., 2018). However, estimation methods based on the likelihood function have not been commonly used to assess ground-motion correlation. One of the main advantages of such approaches is that they are straightforward and do not require further steps for the estimation of the correlation parameters.

We first showed that ML and REML estimates do not depend on the bin size, unlike ordinary and weighted least-squares regression (OLS and WLS). Indeed, the latter requires the definition of the experimental semivariogram, whose robustness depends on both the bin width and the number of stations within each bin. Our outcomes are promising as they show how ML and REML may outperform the least-squares approaches. There is a trade-off between the bin width and the estimate robustness: wider bins include a larger number of residuals pairs, which increases the robustness of the semivariance estimates, but at the same time wider bins may mask shorter correlation lengths.

We then performed two different studies to show the dependence of the correlation on the number of available stations and on the station layout. Firstly, we carried out simulations of within-event residuals on a fine grid, varying both the h_0 and the number of stations available. Generally, ML and REML feature lower percentage errors and dispersion compared to OLS and WLS, independently of the number of stations and of the underlining spatial correlation structure (h_0). This is a rather interesting result, especially for those regions characterised by sparse strong-motion networks. Second, we carried out simulations of within-event residuals only at recording stations of past earthquakes. We chose four different station layouts, which are considered as good examples of the type of station distributions often seen in practice. Our outcomes suggest that OLS and WLS are more affected by the station configuration because their estimates are based on the computation of the experimental semivariogram. Thus, more homogeneous station layouts would provide more reliable range estimates. By contrast, ML and REML seem to be less influenced by the station layout both in terms of median percentage error and interquartile range.

This chapter intends to be a continuation of the work proposed in chapter 2, as it further analyses the dependency of correlation on different factors such as the bin size and the station configuration. We shed light on alternative approaches to characterize the spatial correlation structure of earthquake ground motions, providing useful insights for users and researchers interested in investigating ground-motion spatial correlation.

Finally, we proposed some guidelines to model the spatial correlation uncertainty based on the availability of recordings, following a logic-tree approach. The main idea is to provide hazard and risk assessment by using the median range estimate and the lower and upper bounds range values (5% and 95% confidence intervals). The resulting risk analysis accounting for the correlation uncertainties is eventually obtained by averaging all the branches with suitable weights (0.185, 0.63 and 0.185, respectively).

These findings, while preliminary, may help researchers to model the spatial correlation uncertainty that one should consider when performing regional seismic hazard and risk assessment. Application of the proposed logic-tree to a specific case study may indicate features for further developments.

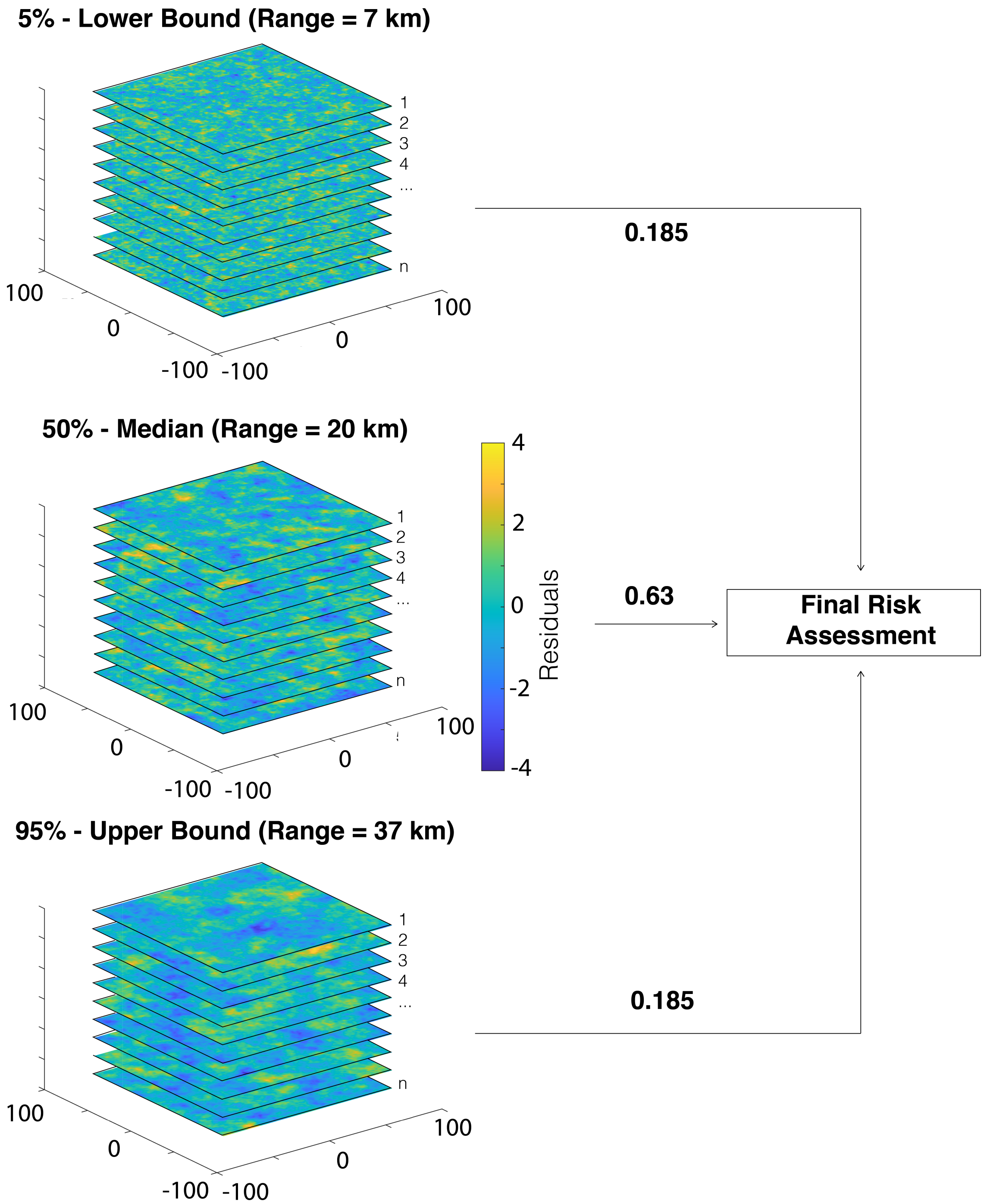


Figure 4.25: Application of the proposed logic tree for $h_0 = 20$ km and number of stations equal to 60 to use within a seismic risk assessment study. We use the estimates obtained by using the REML approach. The graphs on the left-hand side correspond to n spatially correlated random fields simulated for each range value, 7 km, 20 km and 27 km, respectively.

CHAPTER 5

Exploring the impact of spatial correlations in the catastrophe modelling process: a case study for Italy

Catastrophe models are important tools to provide proper assessment and financial management of earthquake-related emergencies, which still create the largest protection gap across all perils. Earthquake catastrophe models include three main components, namely: (1) the earthquake hazard model, (2) the exposure model and, (3) the vulnerability model. Simulating spatially distributed ground-motion fields within either deterministic or probabilistic seismic hazard assessments poses a major challenge when site-related financial protection products are required. Based on the findings of Chapter 2 and 3, here, we first develop *ad hoc* correlation models for different Italian regions (specifically northern, central and southern Italy) and then we perform both deterministic scenario-based and probabilistic event-based hazard and risk assessments in order to advance the understanding of spatial correlations within the catastrophe modelling process. We employ the OpenQuake engine for our calculations. This is an open-source tool suitable for accounting for the spatial correlation of earthquake ground-motion residuals. Our outcomes demonstrate the importance of considering not only the spatial correlation of ground motions, but also its associated uncertainty in risk analyses. Although loss exceedance probability curves for the return periods of interest for the (re)insurance industry show similar trends, both hazard and risk footprints in terms of average annual losses feature less noisy and more realistic patterns if spatial correlation is taken into account. Such results will have implications for (re)insurance companies evaluating the risk to high-value civil engineering infrastructures.

This chapter is based on the following manuscript submitted to Bulletin of Earthquake Engineering:

1. **Schiappapietra**, E., Stripajová, S., Pazák, P., Douglas, J. & Trendafiloski, G. (2021). Exploring the impact of spatial correlations in the catastrophe modelling process: a case study for Italy.

5.1 Introduction

Catastrophe models are important tools to provide proper estimations and financial management of earthquake-related emergencies, which still have the largest protection gap across all perils. The mitigation of socio-economic losses and the development of insurance and reinsurance strategies are, therefore, central targets in the assessment of the seismic risk at urban and regional scales. Earthquake catastrophe models include three main components: hazard, vulnerability and exposure (Grossi and Kunreuther, 2005). The earthquake hazard requires a proper description of the earthquake ground motion (intensity measures - IMs) and its spatial variability across the region using deterministic or probabilistic earthquake hazard models. In particular, three main issues need to be identified to achieve reliable estimates of local intensity and therefore catastrophe loss: (1) locations of potential future events, (2) their frequency of occurrence, and (3) their severity. The vulnerability component estimates the probability that structure damage will exceed various levels as a result of ground motion. It usually defines loss ratios as a function of IM values depending on material of construction, seismic design, building height, occupancy and other modifiers of the building structures. The exposure component provides geocoding of the analysed assets, i.e. (re)insurance portfolio, including asset's vulnerability modifiers, total insured value and policy conditions.

Simulating spatially-distributed ground-motion fields poses a major challenge when site-related financial protection products are required. Indeed, the loss assessment of spatially-distributed systems requires the consideration of: (1) the cross-correlation among different IMs at the same site, (2) the spatial correlation of the same IM at different sites, and (3) the spatial cross-correlation among different IMs at closely-spaced sites (Weatherill et al., 2015). Several authors have demonstrated the importance of including correlation models in risk analyses. For example, Park et al. (2007) and Sokolov and Wenzel (2011) evidenced that neglecting the spatial correlation may cause a bias in loss estimates, whose magnitude depends on the considered portfolio. Similarly, Weatherill et al. (2015) proved that the effects of including spatial cross-correlations is limited when larger footprints are considered or a building typology is predomintant in the portfolio. While investigating the losses in a district of Istanbul as a consequence of a Mw 7.2 earthquake scenario, Wagener et al. (2016) found that the loss distribution changes clearly depend on the correlation model in contrast to the mean loss, which is not affected. Wesson and Perkins (2001) also demonstrated that the spatial correlation is crucial for explaining the variance of losses to a spatially-distributed system, but not the mean loss. Costa et al. (2018) evaluated the consequences of seismic events on transportation networks and observed that the probability of disruption is higher when the spatial correlation is neglected.

It is well-established that IMs at multiple locations during the same earthquake are spatially correlated and their degree of correlation tends to be higher for closely-spaced sites and for low-frequency ground motions. In addition, several authors (e.g. Chen and Baker, 2019; Schiappapietra and Douglas, 2020; Infantino et al., 2021; Schiappapietra and Smerzini, 2021) have demonstrated that the spatial correlation of earthquake ground-motion is period−, regionally− and scenario−dependent, so that the implementation of a unique correlation model calibrated on worldwide databases may represent an oversimplification. Besides, spatial correlation models are usually inferred from a set of multiple events due to the limited number of ground-motion recordings from a single earthquake. The event-to-event correlation variability should be therefore included in regional probabilistic risk analysis to enable more realistic estimations of both the ground motion and corresponding losses (Heresi and Miranda, 2019).

In this context, we firstly calibrate different correlation models based on observations of the Mw 6.5 2016 Norcia (central Italy) earthquake. These models are used in deterministic scenario-based calculations for the same event. The primary aim of such analyses is to better investigate to what extend different input models may affect the overall economic losses. Thereafter, we perform event-based probabilistic seismic hazard assassement (PSHA) calculations in order to advance the understanding of spatial correlations within the catastrophe modelling process. Indeed, the probabilistic seismic risk assessment is usually required for decision-making and insurance/reinsurance purposes (Erdik, 2017; Kohrangi et al., 2021). To account for the event-to-event correlation variability, we follow the methodology proposed by Heresi and Miranda (2019), based on the median range value and its associated dispersion. We employ the OpenQuake engine (Silva et al., 2014; Pagani et al., 2014) for our calculations, which is an open-source seismic hazard and risk modelling tool that can account for the spatial correlation of earthquake ground-motion residuals. In particular, we develop custom spatial correlation models, and we fine-tune the already implemented approach to simulate spatially-correlated random fields to significantly reduce the computational cost and to enable whole country event-based PSHA calculations.

The goals of this study are: (1) to illustrate the impact of spatially dependent modelling on the resulting earthquake shaking losses of building portfolios or spatially-distributed infrastructures and, (2) to investigate the effects on per-event and in-location loss estimates for underwriting purposes. The results of this project, albeit only illustrative, will have implications for (re)insurance companies evaluating the risk to high-value civil engineering infrastructures.

This chapter is organized as follows. Section 5.2 describes the characterization of the spatial correlation of ground motion IMs, with particular emphasis on the custom correlation models derived for the regions of interest. Section 5.3 details the parameters and models required as input to the OpenQuake engine. We mainly focus on the seismic hazard assessment aspect, and only a brief description of the vulnerability and exposure components is given. Finally, Section 5.4 examines the outcomes of the risk analyses with the aim of illustrating the effects of spatial correlations within the catastrophe modelling process. The chapter ends with some brief conclusions.

5.2 Modelling spatial correlation of ground motion intensity measures

Empirical ground motion models (GMMs) provide an estimate of the ground shaking at an individual site as a function of explanatory variables describing the source, the propagation path and the site conditions (e.g Douglas and Edwards, 2016):

$$\log_{10}(IM_{ei}) = \mu_{IM_{ei}} + \delta B_e + \delta W_{ei} \tag{5.1}$$

where IM_{ei} is the intensity measure of interest at site i due to the event e, whereas $\mu_{IM_{ei}}$ is the predicted mean of $log_{10}(IM)_{ei}$. δB_e and δW_{ei} are the between-event and within-event residuals, respectively, which are assumed to be independent and normally distributed with zero mean and standard deviation τ and φ, respectively. δB_e represents the average shift of the observed IM of an individual event e from the median predicted by the model and it is common for all sites, whereas the various δW_{ei} account for site-to-site differences of observations from the median event-corrected predictions. GMMs express the IMs of interest as lognormally distributed random variables, providing only their marginal probability distribution at an individual site. However, for regional seismic hazard and risk assessments, one needs to quantify the joint probability of occurrence of IMs at multiple locations over the region of interest. Therefore, it is necessary to model the spatial variability of δW_{ei}, namely to define how the δW_{ei} vary in space.

5.2.1 Characterization of the spatial variability of univariate random functions

In geostatistical analysis, the most common tool adopted to describe the spatial correlation of a random function is the empirical semivariogram, which measures the average dissimilarity of a pair of random variables $z\left(x_i\right)$ and $z\left(x_j\right)$ at locations x_i and x_j, separated by an inter-site distance h:

$$\gamma = \frac{1}{2} E[z_{x_i} - z_{x_j}]^2 \tag{5.2}$$

where E[] denotes the expected value. In seismological applications, the $z\left(x_i\right)$ represent the (normalized) within-event residuals:

$$z_{x_i} = \frac{\delta W_{ei}}{\sigma_e} = \frac{log_{IM_{ei}} - \mu_{log_{IM_{ei}}} - \delta B_e}{\sigma_e} \tag{5.3}$$

in which σ_e is the sample standard deviation of the within-event residuals for earthquake e.

Due to the lack of repeated observations of ground motions at each site from a given earthquake, further simplifications are needed to estimate γ, and the hypothesis of second-order stationarity and isotropy are commonly assumed. This means that the correlation between any pairs of sites with equal separation distance is the same, independently of the source-to-site distance and orientation. Therefore, under these assumptions, equation (5.2) can be rewritten as:

$$\gamma = \frac{1}{2} E[z_{x_i} - (z_{x_i} + h)] \tag{5.4}$$

where h is the inter-site distance.

To assess the spatial correlation structure on the residuals with semivariograms, the main steps are: (1) compute the experimental semivariogram $\hat{\gamma}(h)$; (2) choose a parametric function (e.g. exponential and spherical models); and (3) estimate the correlation parameters, namely the sill (i.e. the overall variance) and the practical range (i.e. the distance beyond which the correlation is negligible) through regression analysis. In this study, the experimental semivariogram is computed based on the robust estimator proposed by Cressie (1985) and the exponential function is used to model the sample semivariogram:

$$\gamma(h) = a[1 - exp(-3h/b)] \tag{5.5}$$

where a represents the sill and b is the practical range. In the case of the exponential model, b is the distance at which 95% of the sill is reached. The sill is equal to 1 because of the normalization of the within-event residuals and therefore, the range is the only correlation parameter to be estimated. We opt for a weighted least-squares regression to fit the sample semivariogram, in which the weights depend both on the number of pairs and the separation distance, following Baker and Chen (2020):

$$w_i = N(h_i)exp(-h_i/c) \tag{5.6}$$

in which N is the number of pairs in each bin and c is a parameter that describes the decay of the weights as the distances increases. We set it equal to 5 km, based on preliminary analysis. We note that correlation parameters were also estimated through maximum-likelihood approaches in preliminary analyses. In general, the results were similar in most of the cases and therefore we favoured the method that provided a better fit and is more common. The reader should refer to chapter 2 and 4 for further details on correlation modelling.

5.2.2 Stationarity and isotropy assumptions

We use the test by Bowman and Crujeiras (2013) to verify the suitability of a stationary and isotropic model to represent the spatial correlation. The validity of the stationary assumption is determined through a test statistic which compares an estimated semivariogram based on separation distance and location with that obtained as a function of only separation distance. Analogously, the test of isotropy compares an estimated semivariogram based on separation distance and azimuth with that obtained as a function of only separation distance. If the p-value of the statistical tests is greater than 0.05, the evidence of non-stationarity and anisotropy is not statistically significant (at a 5% significance level) and therefore the assumption of stationarity and isotropy cannot be rejected.

In this study, we test the validity of such assumptions by using the sm package (Bowman and Azzalini, 2018) in the R software (R Core Team, 2019). We perform the test for each event in the database and IM considered (Peak Ground Acceleration, PGA; Spectral Acceleration, SA, at 15 periods between 0.1 and 2 s). Figure 5.1 shows the p-values of the statistical tests as a function of Mw considering the events within the Central Italy database and for SA for a period of 1 s. The majority of events have a p-value larger than 0.05, suggesting that the hypotheses of isotropy and stationarity are satisfied at a 5% significance level. We decide to keep also the events with a p-value smaller than 0.05 in our analysis after verifying that semivariograms and their associated parameters are reliable. Similar results are obtained for the other IMs and events. These figures are not shown here for the sake of brevity. In the following, we, therefore, describe the spatial correlation for the different regions through stationary and isotropic models. We note that models cannot be extrapolated to other IMs not considered in this study as we cannot guarantee the validity of the isotropy and stationarity assumptions.

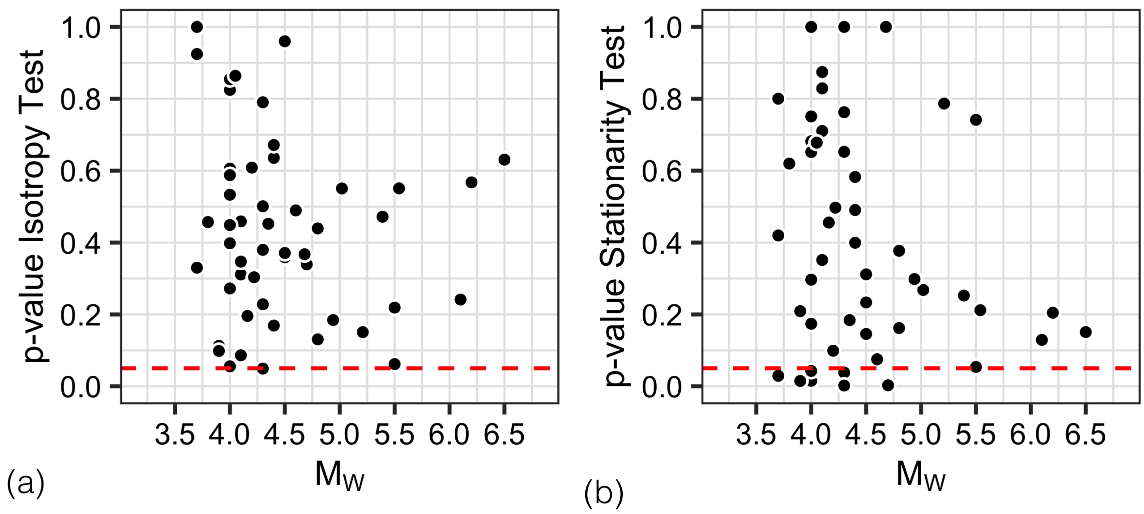


Figure 5.1: p-value as a function of magnitude Mw: (a) isotropy test; and (b) stationarity test. Red dashed line indicates the 5% significance level. The results are for the Central Italy dataset and for IM = SA(T=1 s).

5.2.3 Databases

We employ the European Strong-Motion (ESM) dataset (Lanzano, Sgobba, Luzi, Puglia, Pacor, Felicetta, D'Amico, Cotton and Bindi, 2019) to develop spatial correlation models for Italy. In particular, we apply the selection criteria proposed by Kotha et al. (2020) to select only shallow crustal earthquakes and records with an usable frequency range. Only those events with $\geq$ 40 records (within a Joyner-Boore distance $R_{JB} \leq 120$ km) in the dataset are used to guarantee more robust correlation estimates. We further subdivide the dataset into three sub-categories based on macro-regions: (1) Northern, (2) Central, and (3) Southern Italy to calibrate *ad hoc* correlation models tailored to the specific macro-region. Figure 5.2a summarises the characteristics of the three sub-datasets, whereas Figure 5.2b shows the magnitude-distance scatter plot.

The Northern Italy dataset includes 23 earthquakes (1,408 records) that mainly occurred during the 2012 Emilia seismic sequence with characteristic moment magnitudes of 4.0 - 5.0. All the events are recorded on average by 61 stations, which mainly belong to the soil class C (according to the EC8 soil classes, CEN (2004)), i.e. soft soil characterized by a V_{s30} (average shear wave velocity of the top 30 m of the soil column) between 180 and 360 m/s. The prevailing style of faulting is thrust (TF) and only a few events have normal (NF) or strike slip (SS) mechanisms.

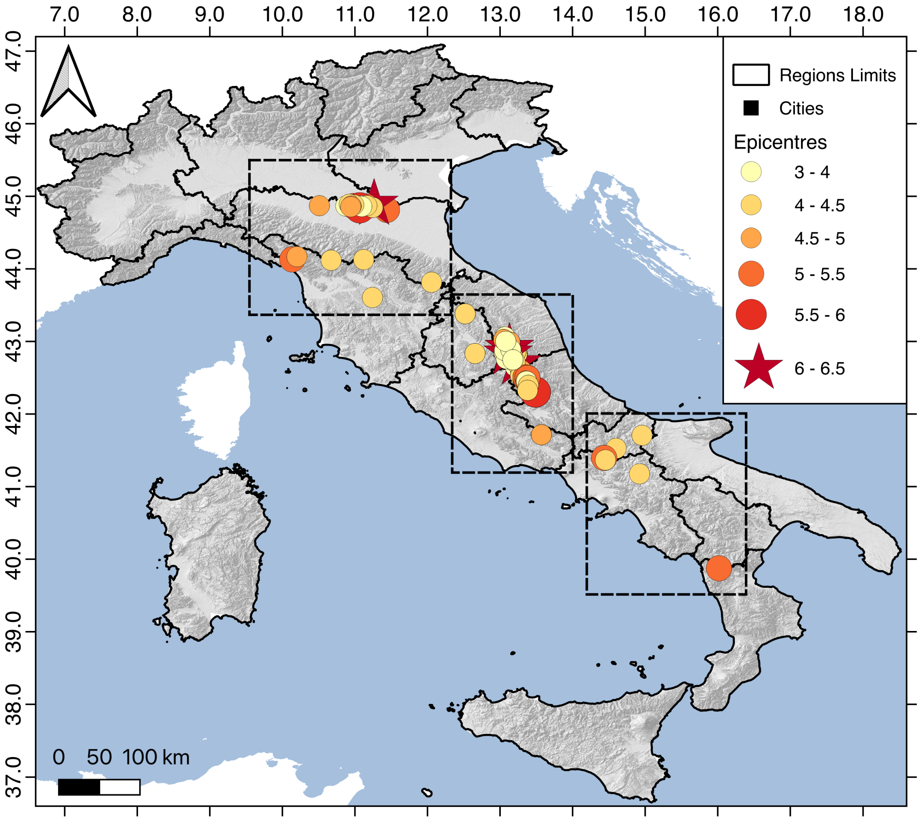
The Central Italy dataset includes 4,363 records from 50 events characterized mainly by normal-fault mechanisms (48 NF and 2 SS). The majority of data belongs to either the 2009 L'Aquila or the 2016 Central Italy sequences, and are from fault distances smaller than 70 km, as well as moment magnitudes in the range 4.0 - 5.0. The largest earthquake is the Mw 6.5 30^{th} October 2016 Norcia event. All the earthquakes are recorded on average by nearly 90 stations, which are mainly classified as soil class B, i.e. stiff soil with a V_{s30} in the range 360 - 800 m/s.

Finally, the Southern Italy dataset contains only 333 records from six events. Four out of the six events have a magnitude of 4.5 and a NF mechanism. Each event is recorded on average by 55 stations, which mainly belong to soil class B. We are aware that the Southern Italy correlation model is not well constrained due to the shortage of data in this area. However, we prefer to distinguish the three macro-regions as correlation models are found to be period$-$ and regionally$-$ dependent (see chapter 2).

5.2.4 Correlation models for the Mw 6.5 Norcia (Central Italy) event

Here, we propose three different spatial correlation models for the Mw 6.5 Norcia event obtained through the methodology explained in section 5.2.1. The models differ only by the reference GMM used to compute the within-event residuals. This makes it possible to investigate the effect of the GMM on the spatial correlation and hence on the resulting earthquake shaking losses of building portfolios. We choose the Lanzano, Luzi, Pacor,

(a)



Felicetta, Puglia, Sgobba and D'Amico (2019) (hereinafter ITA18) and the Kotha et al. (2020) (hereinafter K20) models, which mainly differ in terms of their underlying datasets. The former is calibrated on Italian data whereas the latter includes also European events. We also develop an ad hoc GMM specifically calibrated on the observations of the Mw 6.5 Norcia event. The latter takes the following form based on preliminary analyses:

$$log_{10}Y_{ij} = b_1 + b_2 log_{10}\sqrt{R^2 + b_3^2} + b_4\sqrt{R^2 + b_3^2} + b_5 log_{10}\frac{V_{s30}}{800} \tag{5.7}$$

where R is the R_{JB} distance and b_1, ..., b_5 are the model coefficients inferred through a one-stage ordinary regression, which is justified as here we are only using data from a single event.

Figure 5.3 presents the ranges obtained for PGA (T=0s) and SA at 15 periods between 0.1 and 2 s and the corresponding fitted models as a function of period. Generally, the three different models have a range that is directly proportional to the period, as previously observed in the literature (see chapter 2). At intermediate and long periods, the models converge towards very similar values, suggesting that the underlying GMM has a negligible

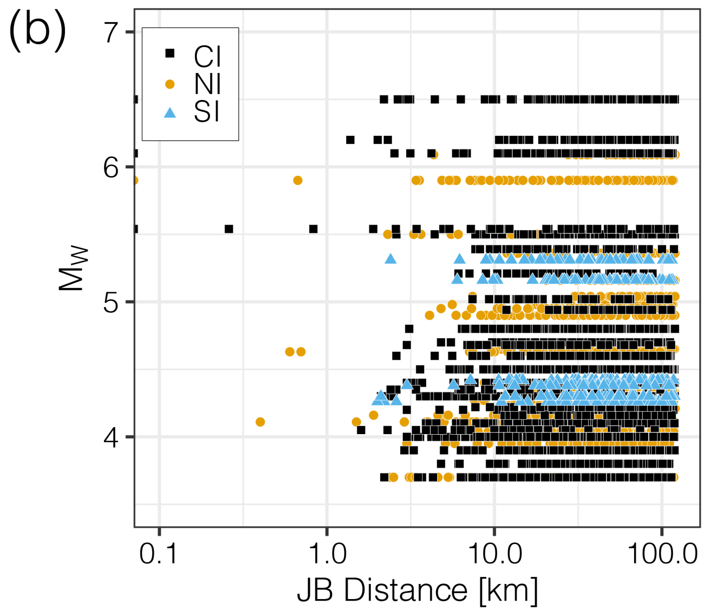


Figure 5.2: Set of Italian data: (a) Location of the epicentres as a function of Mw; and (b) magnitude-distance scatter plot. JB stands for Joyner-Boore distance. The rectangles (in dashed black) indicate the three sub-regions selected in this study: NI: Northern Italy; CI: Central Italy; and SI: Southern Italy.

effect on the final outcomes. By contrast, at short periods ($T \leq 0.4s$), the three models show different trends. An explanation may lie with the different sensitivity of high-frequency and low-frequency ground motions to the anelastic attenuation. Indeed, Kotha et al. (2020) found larger regional differences of anelastic attenuation at short periods than at longer periods. We believe that the larger ranges of the correlation model based on K20 are due to the faster attenuation of the Central Italy region with respect to the pan-European average (Kotha et al., 2020). Such strong attenuation is not modelled in the ergodic K20 GMM and therefore it manifests as apparent spatial correlation. Conversely, the ITA18 and the ad hoc GMMs better capture the anelastic attenuation of high-frequency IMs (e.g. term $b_4\sqrt{R^2 + b_3^2}$ in eq. 5.7), resulting in a lower spatial correlation at short periods.

The observed ranges are fitted either with a bilinear (ITA18 and K20) or linear (ad hoc GMM) expression to facilitate its implementation in the OpenQuake software, as follow:

$$b(T) = \begin{cases} a_0 + a_1(T - t), & T \leq t \\ a_0 + a_2(T - t), & T > t \end{cases} \tag{5.8}$$

$$b(T) = a_0 + a_1 T \tag{5.9}$$

where T is the period of interest and b is the range in km. The coefficients a_0, a_1, a_2 and t obtained through a one-stage regression are reported in Table 5.1 for the three different cases. We recall that the model cannot be extrapolated to other periods as we cannot guarantee that the hypothesis of isotropy and stationarity holds.

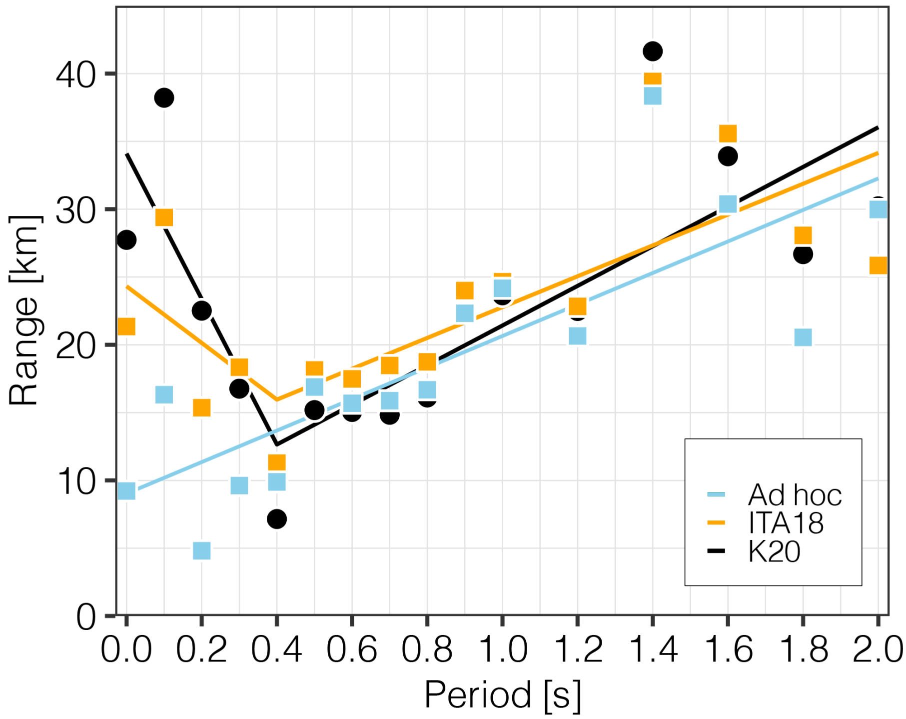


Figure 5.3: Spatial correlation models obtained for the Mw 6.5 Norcia (Central Italy) event. Dots and squares represent the ranges as a function of period, whereas the solid lines represent the fitted models. *Ad hoc* refers to the GMM calibrated on the Mw 6.5 Norcia observations only; ITA18 refers to the Lanzano, Luzi, Pacor, Felicetta, Puglia, Sgobba and D'Amico (2019) GMM; K20 refers to the Kotha et al. (2020) GMM.

Table 5.1: Coefficients of the equations fitted to the ranges for the three GMMs (equation 5.8 and 5.9).

	a_0	a_1	a_2	t
ad hoc	9.03	11.62		
ITA18	15.96	-20.89	11.37	0.40

K20	12.65	-53.62	14.62	0.40

5.2.5 Correlation models for the three macro-regions in Italy

It is recognized that the spatial correlation not only varies with geological context (e.g. Sokolov and Wenzel, 2013; Schiappapietra and Douglas, 2020), but also from event to event (Goda, 2011; Heresi and Miranda, 2019). In this study, we propose different correlation models depending on the considered region (Northern, Central and Southern Italy) and we provide an estimate of the event-to-event variability that should be accounted for to obtain more informed regional risk assessments. We follow Heresi and Miranda (2019), who proposed a simple methodology to include the event-to-event variability of spatial correlation by providing the median range and its associated dispersion. In particular, the main steps are:

1. compute the spatial correlation for each event in the database and for each considered IM, following the methodology described in section 5.2.1. We use the ITA18 GMM to compute the within-event-residuals;
2. compute the central tendency of the range for each IM as the weighted geometric mean. Weights are proportional to the square of the number of stations that recorded the earthquake;
3. compute the dispersion of the ranges for each IM as the weighted standard deviation of the natural logarithm of b;
4. compute the empirical cumulative probability distribution of b for each IM and verify that ranges are lognormally distributed through the Kolmogorov-Smirnov statistical test; and finally,
5. fit the computed weighted mean and standard deviation with simple models as a function of the period, as follows:

$$b(T) = \begin{cases} a_0 + a_1(T - t), & T \leq t \\ a_0 + a_2(T - t), & T > t \end{cases} \tag{5.10}$$

$$b(T) = a_0 + a_1 T \tag{5.11}$$

$$\sigma_{Ln(b)} = a_0 + a_1 T + a_2 T^2 \tag{5.12}$$

The fitted models for the median and variability are then used in conjunction with equation 5.5 to simulate spatially distributed ground motion fields. For each earthquake scenario in an event-based PSHA, b is sampled from a lognormal distribution with median and dispersion provided by the fitted model. By so doing, the correlation structure of the different ground motion field varies each time as would be expected in nature, rather than being fixed.

The resulting ranges and their associated variability for the three regions are presented in Figure 5.4 along with the fitted models, whereas the model coefficients are reported in Table 5.2. We note that the appropriateness of these models is measured based on standard criteria, such as the BIC (Bayesian information criteria) and AIC (Akaike's information criterion).

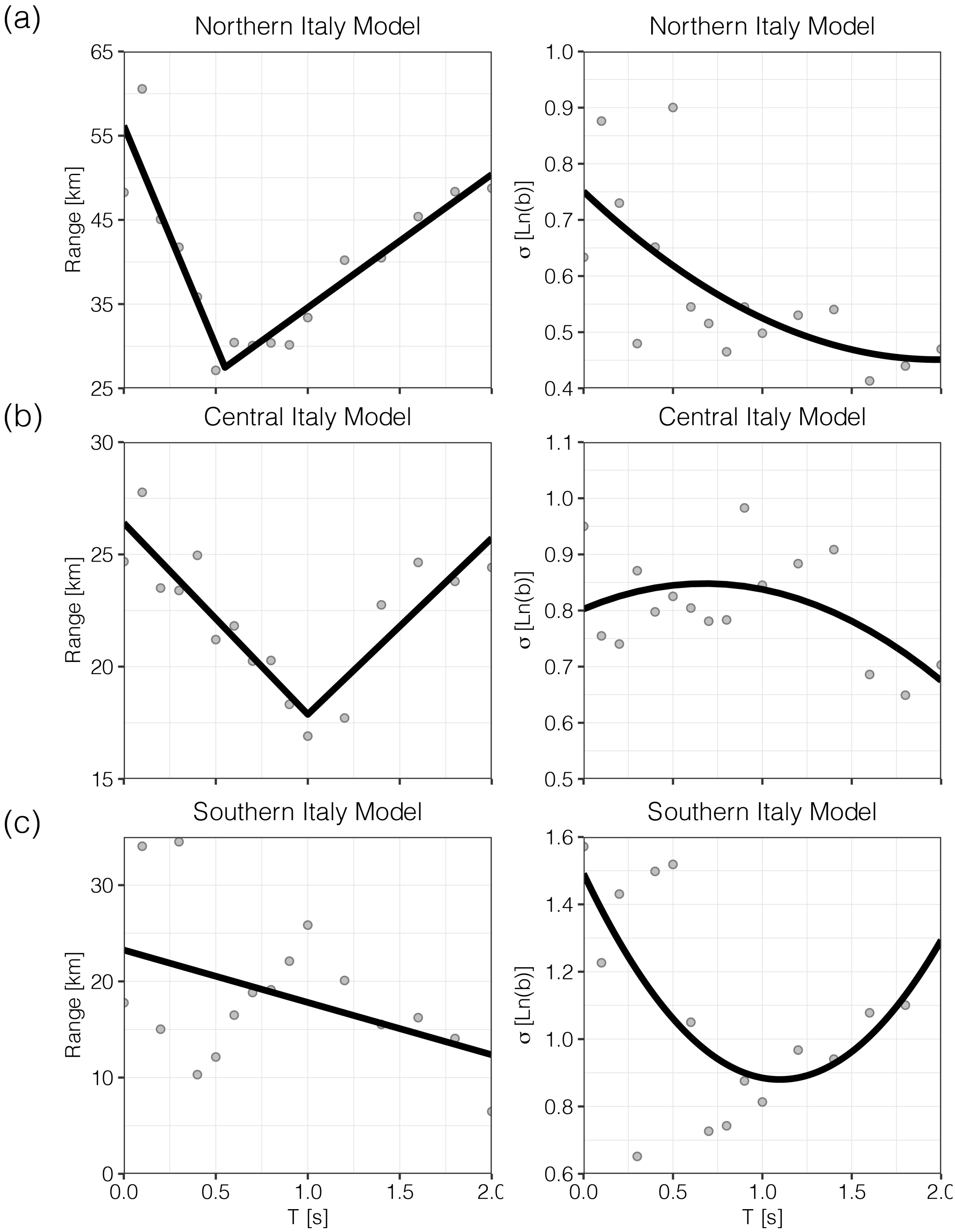


Figure 5.4: (Left hand side) Computed weighted geometric mean and fitted model, and (right hand side) computed weighted standard deviation and fitted model. (Top) Northern Italy, (Middle) Central Italy, and (Bottom) Southern Italy.

Table 5.2: Coefficients of the equations fitted to the ranges (equation 5.10, 5.11 and 5.12) for the three different regions.

Region		a_0	a_1	a_2	t
Northern Italy	b [km]	27.48	-52.20	15.81	0.55
	σ	0.75	-0.30	0.08	-
Central Italy	b [km]	17.87	-8.52	7.85	1
	σ	0.8	0.13	-0.1	-
Southern Italy	b [km]	23.25	-5.44	-	-
	σ	1.49	-1.11	0.51	-

The correlation parameters for the Northern and Central Italy regions are modelled through bilinear (range - equation 5.10) and quadratic (dispersion - equation 5.12) expressions, whereas the Southern Italy model for the range value is expressed through a simpler linear equation (equation 5.11). We note that the Southern Italy model features a higher dispersion compared to the other two regions. This is due to both the lower number of events available for this region and the smaller number of stations that on average recorded those events. The Northern Italy region has a lower variability with respect to the Central Italy model. At the same time, the former features larger ranges compared to the latter over all the considered periods, suggesting that the ground motion is on average correlated over longer distances. Such a trend is most likely due to peculiarities of the local site effects and propagation path, which may strongly affect the spatial correlation of a region, as demonstrated for instance by Jayaram and Baker (2009), Sokolov and Wenzel (2013) and Schiappapietra and Douglas (2020). The Central Italy region is indeed characterized by a high degree of heterogeneities in terms of local site effects compared to the Po plain (Northern Italy) region, where soil conditions are more homogeneous in terms of V_{s30}.

Finally, Figure 5.5 compares published correlation models with the models proposed in this study. Some observations are worthy of remark. Firstly, although Sgobba et al. (2019) provided correlation parameters of the corrective term (e.g. sum of the repeatable terms of variability due to source, path and site effects) and not of the within-event residuals, our Northern Italy model is in agreement with their results, suggesting that this region features on average ground motions correlated over longer distances. The Huang and Galasso (2019) and Esposito and Iervolino (2012)−ITACA models were developed based on events from the entire Italian territory and therefore they provide an average spatial correlation structure. While these models have different trends at shorter periods compared to those proposed in this study, differences with the Central Italy model are not significant for periods longer than 0.8s. This may likely be due to the datasets on which the correlation parameters were computed, which are dominated by normal faulting events that occurred along the Apennine chain. It is noted that Esposito and Iervolino (2012) proposed two different models based on the ITACA (*URLhttp://itaca.mi.ingv.it/ItacaNet_31/#/home*) and ESD (*http://www.isesd.hi.is*) databases, respectively. For comparison, we also show the ESD model, which is based on pan-European (and not just Italian) data. For the sake of completeness, we also show the correlation model proposed by Heresi and Miranda (2019) which is based on 39 well-recorded events from the NGA-West2 database (Ancheta et al., 2014). Generally, this study has lower ranges with respect to the models developed for Italy, suggesting that regional differences in spatial correlation cannot be neglected.

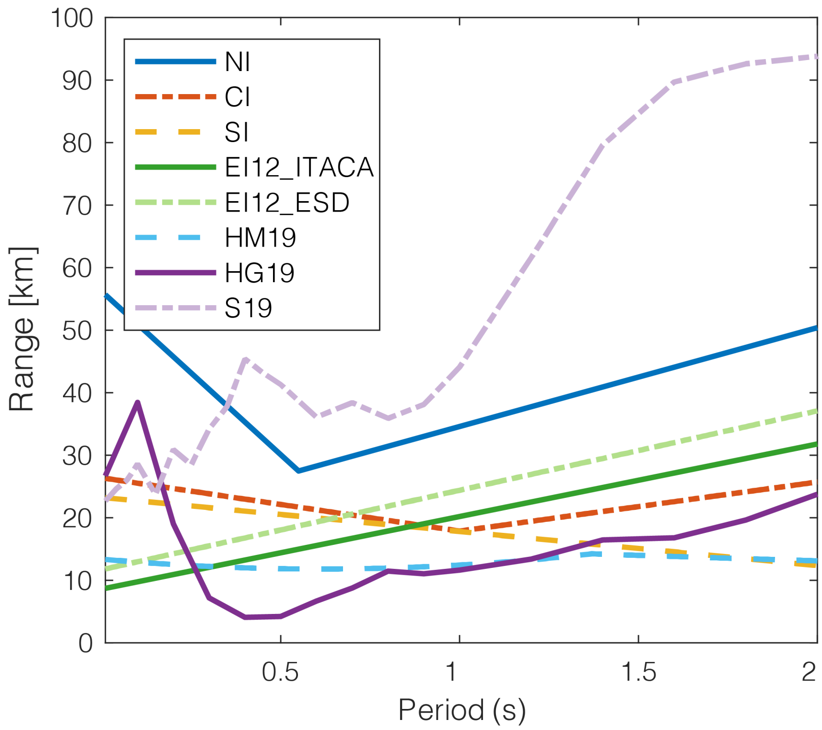


Figure 5.5: Comparison among different correlation models: NI - Northern Italy; CI - Central Italy; SI - Southern Italy; EI12 - ITACA - Esposito and Iervolino (2012); EI12 - ESD - Esposito and Iervolino (2012); HM19 - Heresi and Miranda (2019); HG19 - Huang and Galasso (2019); S19 - Sgobba et al. (2019).

5.3 OpenQuake engine input models

5.3.1 Hazard component

The OpenQuake engine is used here to perform both deterministic and probabilistic seismic hazard assessments for the region of interest. The following subsections describe the input parameters and models required for both the scenario-based PSHA and event-based PSHA calculations (Pagani et al., 2014).

5.3.1.1 Scenario - based seismic hazard assessment

The computation of ground-motion fields for a specific earthquake scenario requires three main inputs: (1) a fault rupture model that defines the location and geometry of the source, (2) a GMM, and (3) a model of the local site conditions. We consider a recent earthquake, namely the Mw 6.5 30^{th} October 2016 Norcia (Central Italy) event, as the reference to define the rupture. The following parameters (*https://esm-db.eu/#/event/EMSC-20161030_0000029*), along with the fault geometry, are used as input to OpenQuake: epicentre latitude 42.82°; epicentre longitude 13.16°; focal depth 6.8 km; Mw 6.5; rake = 95°; strike 151°; and dip 47°. We select the ITA10 GMM (Bindi et al., 2011), which is one of the best performing models for shallow active crustal regions in Italy according to Lanzano et al. (2020). We choose an independent GMM compared to the ones used in the correlation modelling, so that the correlation model is the only varying input. We calculate ground-motion fields at multiple resolution grids with $1km$ and $250m$ grid spacing. The finer grid is used for densely populated area with potential of high IMs values to properly account for local site effects. Finally, the local site conditions are taken into account following Mori et al. (2020), who derived a detailed V_{s30} map for Italy (spatial resolution of $50 \times 50m$), which also includes data from Italian seismic microzonations.

The OpenQuake engine allows the simulation of multiple spatially-correlated random fields to account for the aleatory variability in the ground motion by sampling the between- and within-event variability components from the GMM and by considering the spatial correlation of the within-event variability. To ensure the convergence of the mean and associated standard deviations of the results, we generate 1,000 ground motion fields based on the works of Silva (2016) and Costa et al. (2018). The spatial correlation is modelled as mentioned in section 5.2.4. To investigate to what extent this feature impacts the overall economic losses, we consider four different cases. First, we generate ground-motion fields without considering the spatial correlation. Thereafter, we employ the three models described in section 5.2.4. It is noted that the algorithm to simulate spatially-correlated random fields has been fine-tuned to reduce the computational cost and to enable a larger number of sites to be tested. Instead of the original covariance formulation, which is already implemented in OpenQuake, we use the randomization method of the Python package gstools (Müller and Schüler, 2021). This method represents the spatial random field through the Fourier integral (Wu and Baker, 2014) and it evaluates its discretised modes at random frequencies.

We generate spatially-correlated ground-motion fields for PGA and SA at T = 0.3, 0.6 and 1 s. These IMs are used in the vulnerability component to calculate the building damage. A building can be sensitive to vibrations of different periods depending on its height and material of construction. Figure 5.6 provides a realization of the spatially-correlated ground-motion field for PGA obtained by using the different correlation models. The hazard footprint for the case without spatial correlation appears very noisy. Indeed, the within-event residuals are randomly generated so that the IMs of interest at each site are considered as independent random variables. By contrast, the distribution of PGA values in the other cases show smoother patterns as a result of the spatial correlation, which makes closely-spaced sites likely to experience similar ground-motion levels. The comparison between the *ad hoc* and K20 models is also noteworthy. The latter has a greater range (correlation length) which makes the ground motion correlated over longer distances. Conversely, the shorter range in the *ad hoc* model results in a patchier distribution of the PGA.

The ground-motion fields are then combined with the vulnerability and exposure components to calculate the losses of the portfolio. The outcomes of the risk analysis are shown in Section 5.4

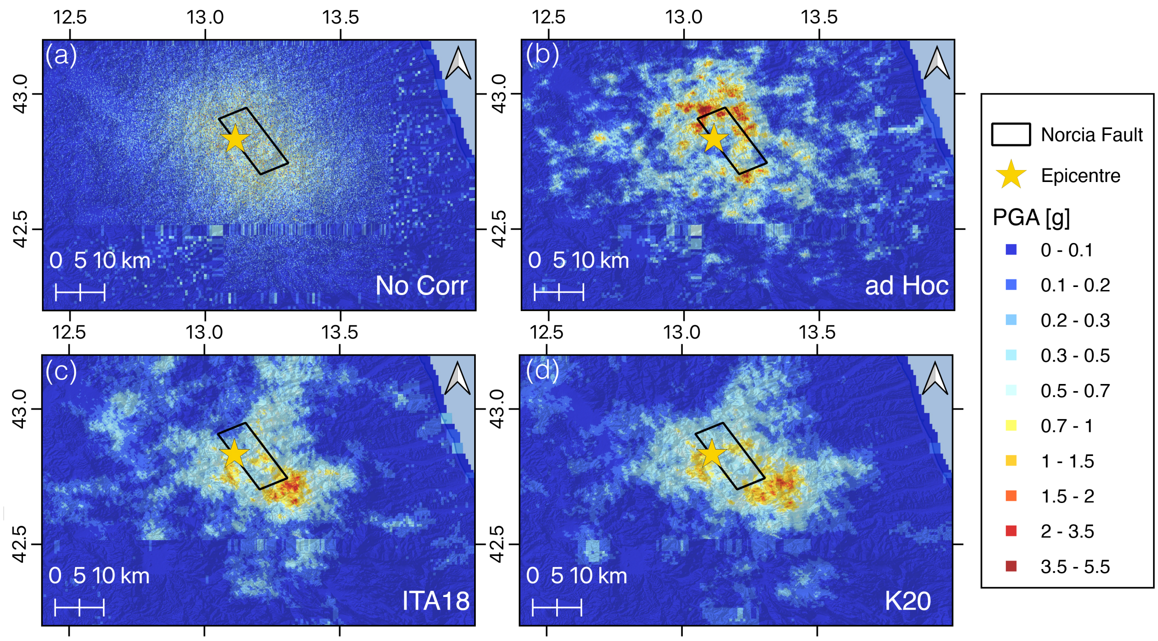


Figure 5.6: Spatially-correlated ground-motion maps obtained by using different correlation models: (a) No within-event spatial correlation; (b) *Ad hoc* spatial correlation model; (c) ITA18 spatial correlation model; and (d) K20 spatial correlation model. The yellow star represents the Mw 6.5 Norcia epicentre, whereas the black rectangle defines the surface fault projection.

5.3.1.2 Event-based probabilistic seismic hazard assessment

We use the event-based calculator from the OpenQuake engine to compute PSHA for Italy through a Monte Carlo (MC) simulation-based approach (Pagani et al., 2014). The first step generates synthetic earthquake catalogues, also called a stochastic event set, by randomly sampling all possible ruptures from the input source model (Atkinson and Goda, 2013; Musson, 2000). The events from the catalogue are then used to estimate the ground-motion IMs of interest at each site by using a list of suitable GMMs in conjunction with the model of the local site conditions. Likewise to the scenario-case, we adopt the V_{s30} map proposed by Mori et al. (2020). A logic tree of GMMs is usually employed to describe the ground-motion field with the aim of capturing the epistemic uncertainty. Finally, the IMs estimates obtained at each site for each event are rearranged so that the seismic hazard curves (annual maximum IMs as a function of the probability of exceedance) can be inferred. Musson (2000) demonstrated that for a sufficiently large number of simulations the results of the event-based PSHA are close to the outcomes of the classical PSHA, which uses the numerical integration of the total probability integral. The reader should refer to Atkinson and Goda (2013) for further details and advantages of this approach over the classical one.

In our analyses, we implement the SHARE (Woessner et al., 2015) source and ground-motion models for Italy. In order to account for spatial correlation models for three main macro-regions of Italy (see Section 5.2.5), we take into account only the source area that can contribute to the seismic hazard and risk of the region under study. With regard to the GMMs, we slightly modify the GMM logic tree to enable the modelling of the spatial correlation of the within-event residuals of the ground motion. Indeed, not all the GMMs included in the logic tree (e.g. the Cauzzi and Faccioli (2008), Toro (2002) and Campbell and Bozorgnia (2003) models) decompose the total aleatory variability into between- and within-event components. For such GMMs, we set the within-event standard deviation (φ) approximately up to 90% of the total standard deviation (σ_{TOT}) based on preliminary analyses of the ratio φ/σ_{TOT} of the other GMMs. We implement in the OpenQuake engine the spatial correlation models developed in Section 5.2.5 to evaluate the impact of such features on the risk outcomes. We perform three different tests: (1) not taking into account spatial correlation; (2) including the event-to-event spatial correlation variability as described in Section 5.2.5; and (3) considering only the median range to characterise the spatial dependency of the ground motion. Figure 5.7a shows an example of hazard curves for the Norcia city obtained for the three different case-studies. The curves represent the return period (inverse of the annual probability of exceedance) of various PGA levels. While hazard curves have the same trend at high probabilities of exceedance (short return periods) independently of the correlation structure, results tends to diverge at lower probabilities of exceedance (return periods larger than 200 years). Such trend is also highlighted in Figure 5.7b, which shows the ratio of PGA values for the three different case-studies. Note that the case in which the event-to-event spatial correlation variability is included is taken as reference.

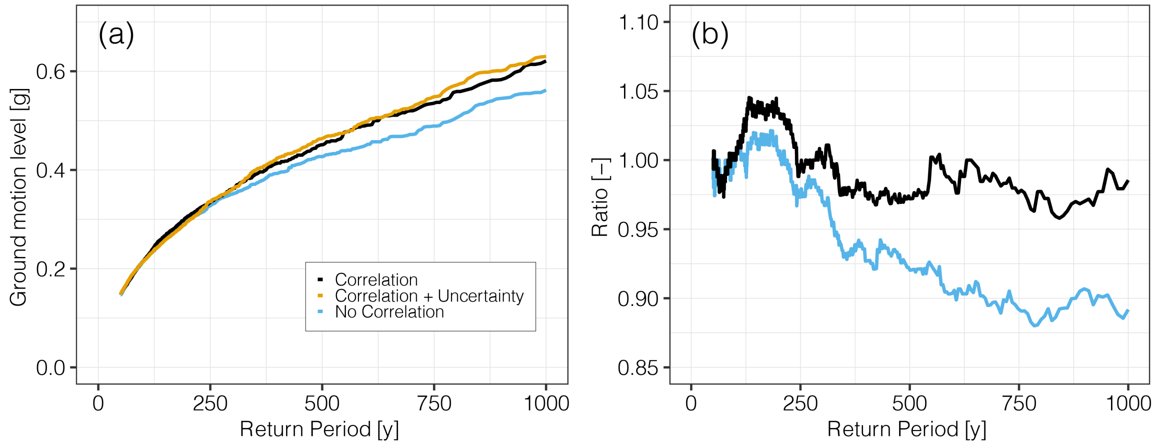


Figure 5.7: (a) Hazard curves showing the return period of different ground motion levels in terms of PGA for three different cases: ground-motion fields are generated without considering correlation; ground-motion fields are generated considering correlation; ground-motion fields are generated considering correlation and its associated uncertainty. (b) ratio of ground motion levels as a function of the return period.

5.3.2 Vulnerability component

The vulnerability component defines loss ratios as a function of IM values depending on the building structure class. In our study, we use the vulnerability functions that have been developed in the framework of the 2020 SERA (Seismology and Earthquake Engineering Research Infrastructure Alliance for Europe, *http://www.sera-eu.org/en/home/*) project. Crowley et al. (2021), Martins and Silva (2020) and Silva et al. (2020) developed about 500 functions to cover the building classes of the European exposure database (Crowley et al., 2020). The building classes for the area under study are shown in Section 5.3.3. The adopted vulnerability functions employ PGA or SA at 0.3, 0.6 and 1 s as ground motion IMs depending on the fundamental period of the different building classes. In general, high-rise structures are associated with long-period IMs, whereas low-rise buildings are associated with short-period IMs, such as the PGA or SA at 0.3 s. It is noted that, for the sake of simplicity, and to highlight the effects of different correlation models, in our analysis, we do not consider the uncertainties associated with the loss ratio. The reader should refer to Crowley et al. (2021) and Martins and Silva (2020) for further details on the calibration of the vulnerability component.

5.3.3 Exposure component

In our analysis, we implement the European exposure database (Crowley et al., 2021) for Italy, which has been developed within the SERA project. The database reports the distribution of the main residential, industrial and commercial buildings classes along with their replacement costs and numbers of occupants. Buildings are classified according to: (1) the main construction material (e.g. unreinforced masonry - MUR, confined masonry - MCF, reinforced concrete - CR, steel - S); (2) lateral load resisting system; (3) number of stories; (4) seismic design code level (e.g. pre-code - CDN, low code - CDL, moderate code - CDM, high code, CDH); and (5) lateral force coefficient used in the seismic design. The replacement cost is the value of replacing a building based on the latest building codes of the country and it is given by the sum of structural and non-structural costs as well as the cost of contents. As an example, Figure 5.8a shows the buildings per each building class according to the European exposure database, and Figure 5.8c provides the percentages of total replacement costs for the structural and non-structural components and the contents. These data refer to the area of interest of both the scenario-case and event-based calculations, as presented in Figure 5.8b. The reader should refer to Crowley et al. (2021) for further details on the derivation of the exposure database.

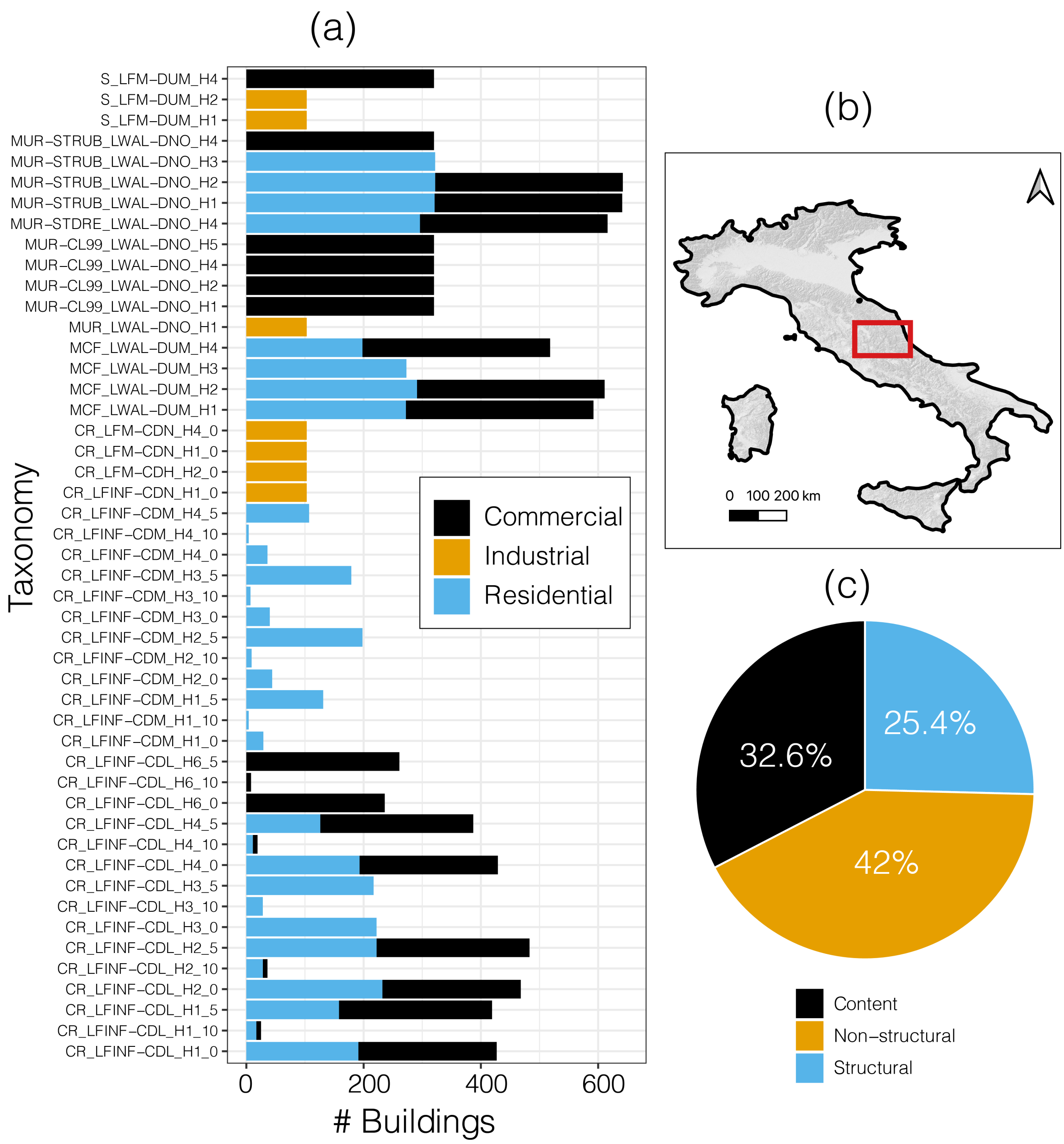


Figure 5.8: (a) Classification of the building classes according to the European exposure model. For further details on the labels, the reader should refer to Crowley et al. (2021); (b) Area of interest: residential, industrial and commercial buildings of graphs (a) and (c) are within the red rectangle; (c) Percentage of replacement costs, divided into contents and structural and non-structural components.

5.4 Risk analysis and discussion

5.4.1 Scenario-base loss assessment

In this section, we present the potential economic losses obtained by combining the vulnerability and exposure components with the hazard model presented in Section 5.3.1.1. Figure 5.9 shows the distribution of losses (average computed over the 1000 simulations) with respect to the total portfolio value (%TV). We derive mean loss maps for each of the correlation models derived for the Norcia event to highlight the potential effects of considering spatially-correlated ground-motion fields. The economic portfolio is disaggregated at the grid level, based on the population density, in contrast to the original exposure model, which assumes that all the buildings are located at the centroid of each municipality. Indeed, such configurations would mask the actual effects of using various correlation models, especially if the inter-site distance among the municipality centroids is larger than the correlation ranges. In particular, we obtain population density data for each grid cell starting from available raster (*https://www.eea.europa.eu/*) and we compute weighting factors as population density in a given cell divided by population density in the corresponding municipality. Eventually, the total economic value in each grid cell is calculated by multiplying the total economic value of the municipality by the corresponding weighting factor.

Although differences in the spatial distribution of losses are discernible among the case studies, risk calculations do not seem to be strongly affected by the specific correlation model, as also highlighted in Table 5.3.

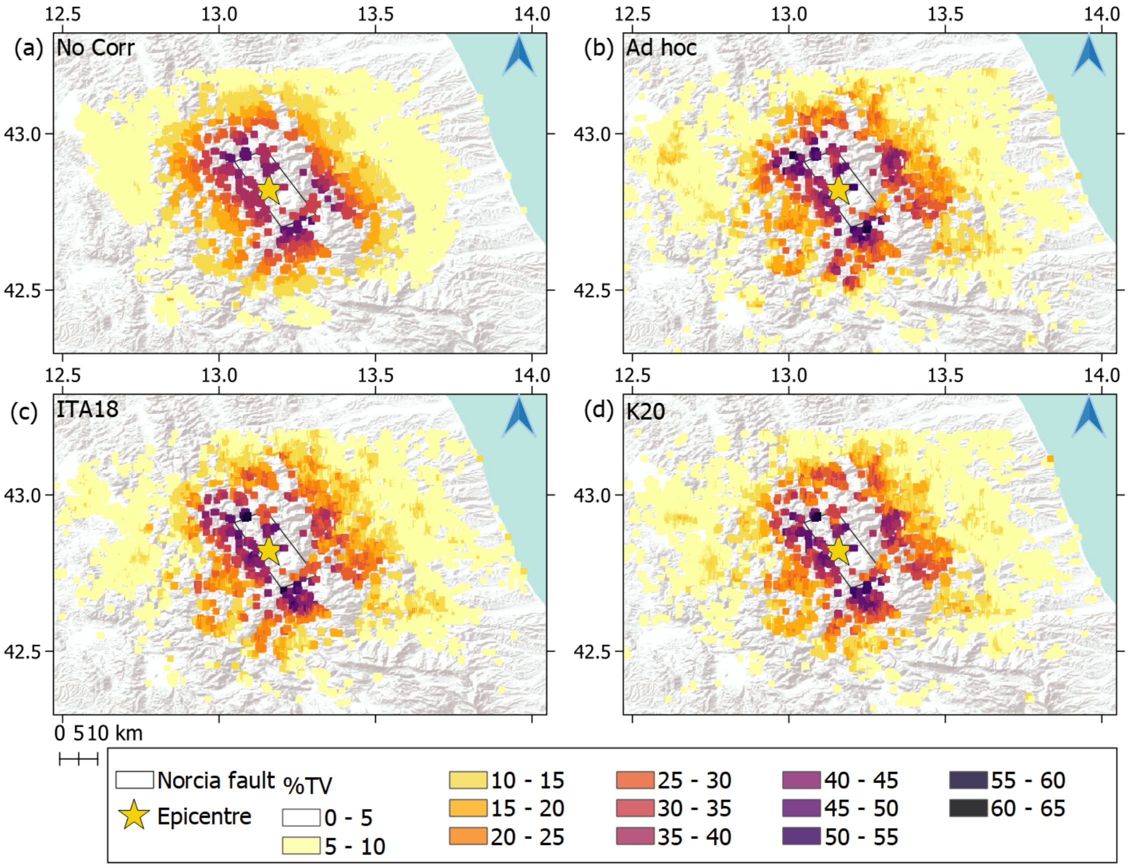


Figure 5.9: Losses (%TV computed based on different hazard models expressed in terms of percentage of total value of the portfolio: (a) Spatial correlation is not considered; (b) ad hoc correlation model; (c) ITA18 correlation model; and (d) K20 correlation model. The yellow star represents the Mw 6.5 Norcia epicentre, whereas the black rectangle defines the surface fault projection.

Table 5.3: Overall loss in %TV (percentage of total value of the portfolio) for the four different case studies.

Correlation Model	Overall Loss in %TV
No Correlation	4.56
Ad hoc	4.30
ITA18	4.63
K20	4.26

We believe that the exposure model, in terms of heterogeneity, plays a critical role in determining such results. Indeed, as shown in Figure 5.3, the correlation models tend to converge toward analogous values as the period increases. This means that if mid- and high-rise structures are predominant within the portfolio, the corresponding losses would be similar as a result of ground motions being correlated over larger distances.

Such conclusions can also be drawn from Figure 5.10, which presents the histograms of economic losses computed based on the different hazard models. We do not observe significant differences in the loss distributions, in contrast to Wagener et al. (2016), who found an increasing coefficient of variation (CV), i.e. the ratio between the standard deviation and the mean, with increasing correlation length. The CV is indeed almost the same (0.55) for all cases. In addition to the above-mentioned explanations, these diverging outcomes could be associated with the portfolio dimension. Park et al. (2007) and Weatherill et al. (2015) demonstrated that the effects of including spatial correlations are greater when smaller portfolios (with a footprint within the correlation length) are considered. In such cases, it is much more likely that ground motions are substantially higher/lower than the median values at all sites within the exposure database, compared to larger portfolio.

To further highlight the impact of spatial correlation models within the risk assessment, we therefore consider a homogeneous spatially-distributed flat portfolio for each IM (e.g. PGA and SA at 1 s). In this case, only one building type is considered at all grid points with same building cost. We compute loss distribution parameters, such as mean, standard deviation and CV (Table 5.4), for each case to identify trends. In agreement with Wagener et al. (2016), we now note that while the mean is almost constant, CV tends to increase as the correlation length increases for short-periods IMs. On the other hand, the CVs remain stable for long-period SAs. Indeed, the correlation models proposed in Section 5.2.4 converge towards comparable correlation lengths, thus resulting in similar loss distributions.

Table 5.4: Mean and Coefficient of variation of the losses obtained from the different correlation models for two different IMs.

Correlation Model	Mean (10^9) - PGA	CV - PGA	Mean (10^9) - SA 1s	CV - SA 1s
No Correlation	6.4	0.79	29.2	0.64
Ad hoc	6.1	0.80	29.6	0.65
ITA18	6.4	0.88	30	0.65
K20	6.7	0.91	29.7	0.65

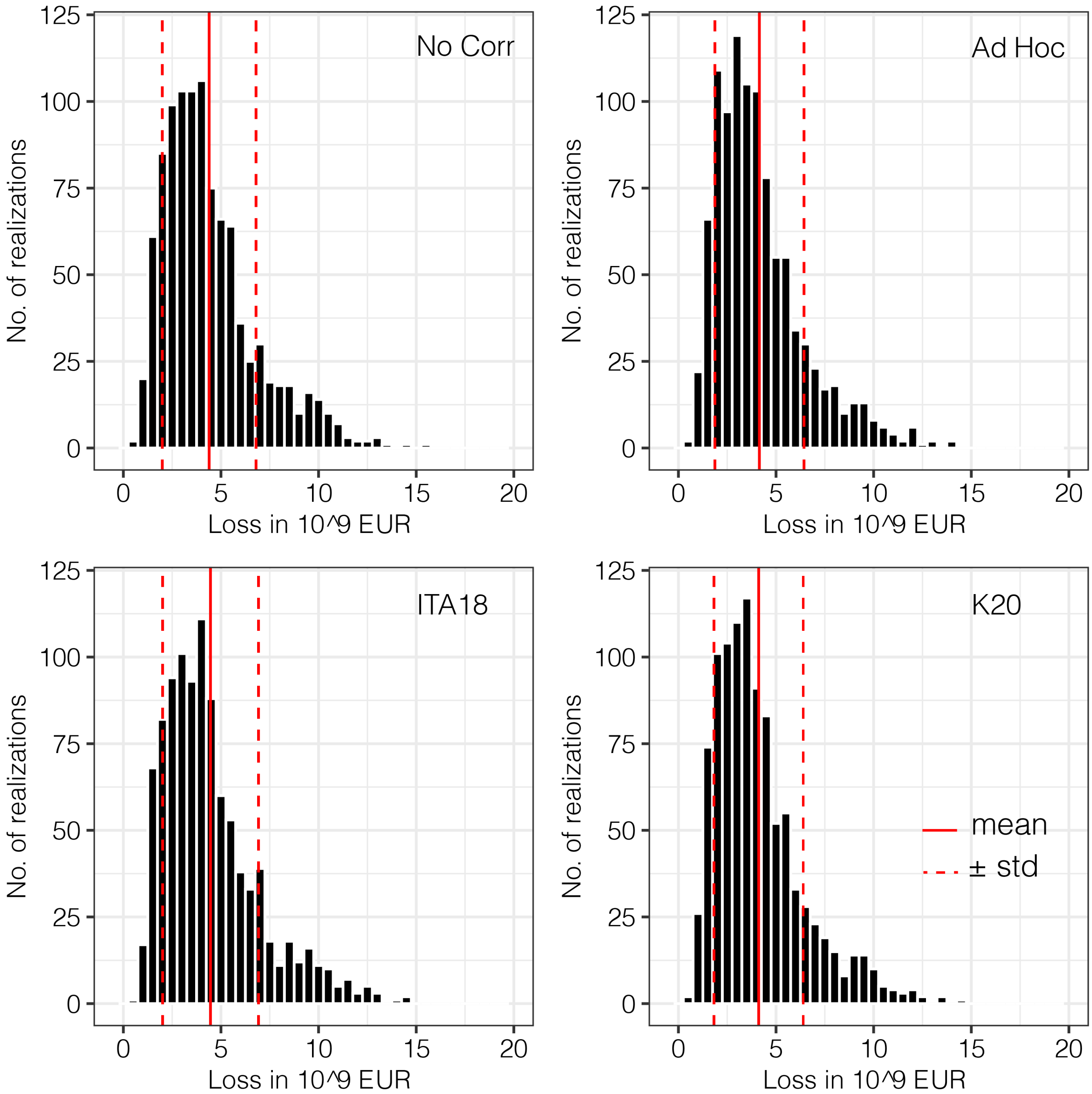


Figure 5.10: Histograms of economic losses computed based on different hazard models: (a) Spatial correlation is not considered; (b) ad hoc correlation model; (c) ITA18 correlation model; and (d) K20 correlation model. Red lines point out the mean value, whereas the red dashed lines indicate the mean ± std.

5.4.2 Event-based probabilistic loss assessment

In this section, we present the potential economic losses obtained by combining the vulnerability and exposure components with the hazard model presented in Section 5.3.1.2 We focus only on the same geographical area (Central Italy) to better understand the impact of including the spatial correlation in our analysis. Risk assessment for the whole Italian mainland will be the object of future developments. Figure 5.11 shows the exceedance probability (EP) curves in terms of losses for three different case-studies: (1) ground-motion spatial correlation is neglected; (2) ground-motion fields are generated by considering a median correlation model as presented in Section 5.2.5; and (3) event-to-event spatial correlation uncertainty is taken into account as explained in Section 5.2.5. In general, we observe greater losses at higher return period (lower annual probabilities of exceedance) when spatial correlation and its associated uncertainty are included with respect to the case in which the spatial correlation is neglected. When only a median spatial correlation is used, rare losses are slightly overestimated compared to when correlation is neglected. This trend is also highlighted in Figure 5.12, which shows the ratio of losses for the three different case-studies. Note that the case in which the event-to-event spatial correlation variability is included is taken as reference. The results are in agreement, for example, with Park et al. (2007) and Weatherill et al. (2015), even though our study does not show marked differences. An explanation may lie with the heterogeneity of the portfolio along with its dimension, which would also require the consideration of cross-correlation among different IMs to obtain more robust loss estimates. On the other hand, for return period of interest for the (re)insurance industry (up to 1000 years), EP curves show a similar behaviour, suggesting that overall losses are not strongly affected by the different correlation structures. We are aware that our outcomes are preliminary because an accurate estimation of the likelihood of observing rare and frequent losses would require the definition of the cross-correlation, especially in case of a heterogeneous portfolio. Nonetheless, such results provide useful advances to better understand the impact of including not only spatial correlation but also its associated variability. In particular, the consideration of the event-to-event spatial correlation variability leads to an overestimation of the rare losses with respect to the simple spatial correlation case, whereas a trend cannot be identified for frequent losses. We recall that when only a median correlation model is considered, the range is constant for each ground-motion field generated. Conversely, the correlation structure of the different ground-motion fields varies each time when the dispersion of the range is considered in the analysis.

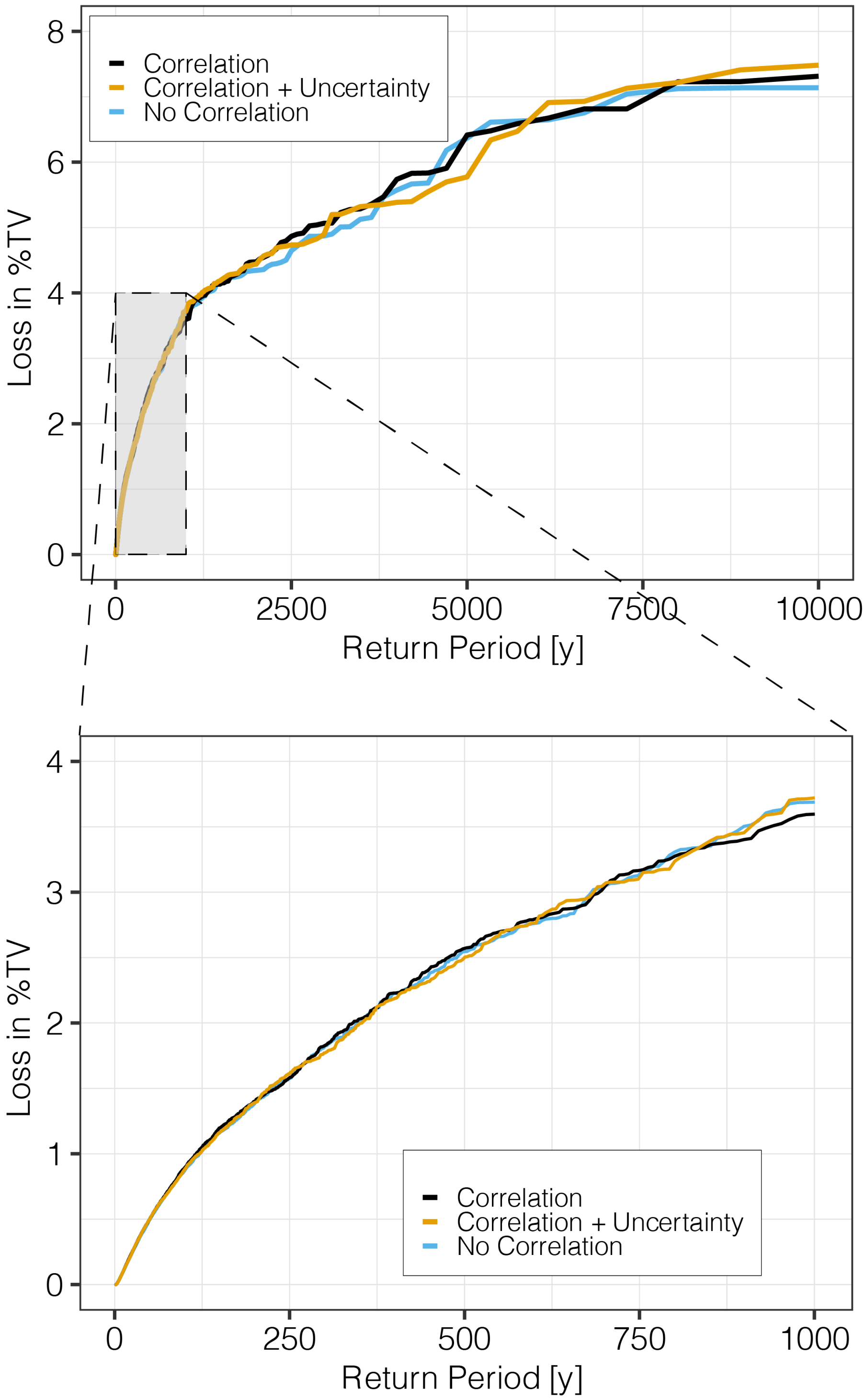


Figure 5.11: EP curves in terms of loss in %TV for three different cases: ground-motion fields are generated without considering correlation; ground-motion fields are generated considering correlation; ground-motion fields are generated considering correlation and its associated uncertainty. TV stands for total value of the economic portfolio (2020 SERA exposure model). In the inset, we show EP curves up to a return period of 1000 years, which is usually the range of interest for (re)insurance industry.

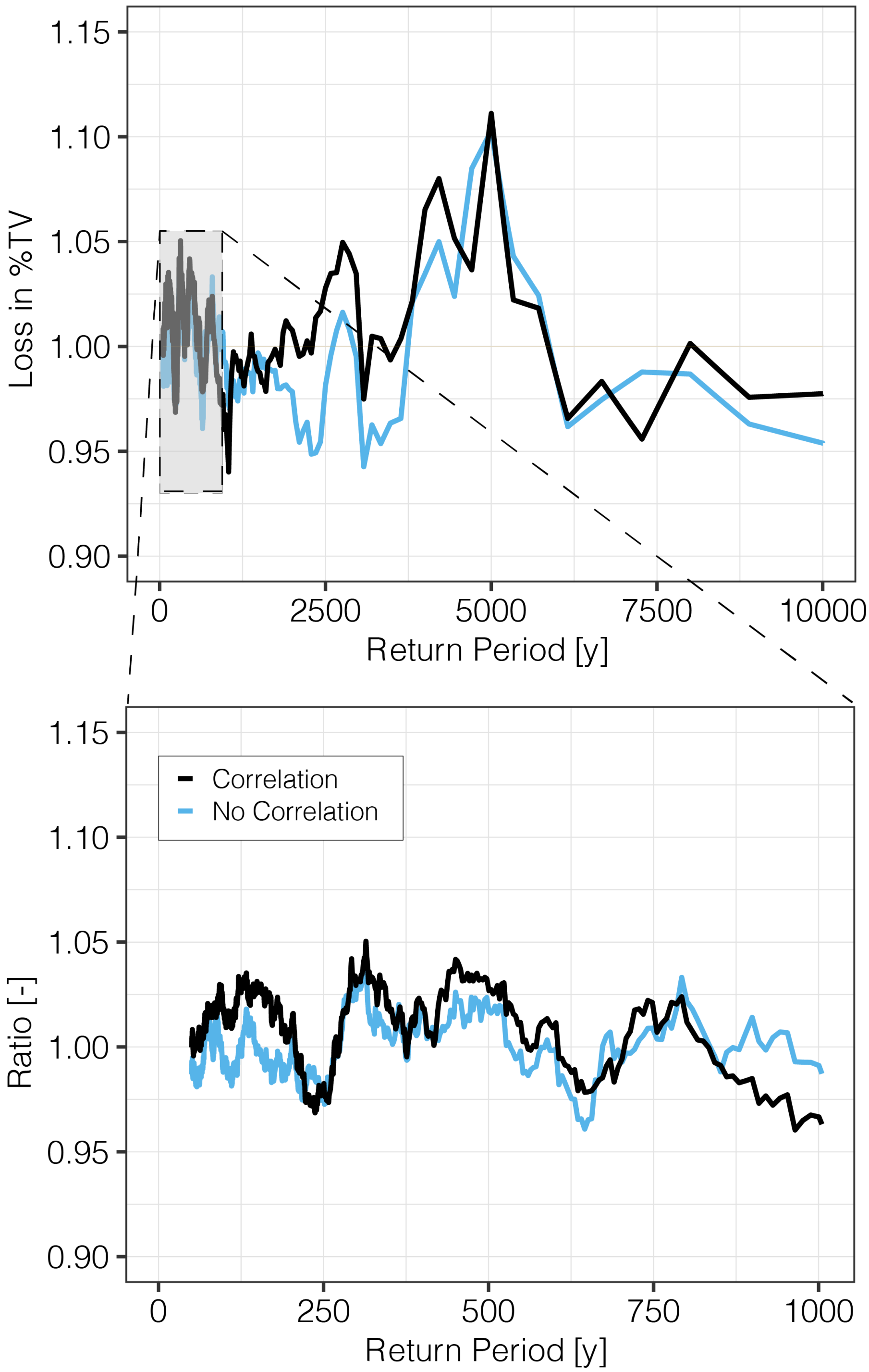


Figure 5.12: Loss ratio as a function of the return period for three different cases: ground-motion fields are generated without considering correlation; ground-motion fields are generated considering correlation; ground-motion fields are generated considering correlation and its associated uncertainty. In the inset, we show loss ratio curves up to a return period of 1000 years, which is usually the range of interest for (re)insurance industry.

To further highlight the impact of spatial correlation and its associated uncertainty within the probabilistic risk assessment, we therefore account for a homogeneous spatially disaggregated flat portfolio for different IMs, such as PGA and SA at 0.3, 0.6 and 1 s. In this case, only one building type is considered in each grid cell. Figure 5.13 compares the average annual losses (in %TV) obtained for the same three cases as considered for the SERA economic portfolio (Figure 5.11). The inclusion of spatial correlation (with and without the event-to-event variability) makes the risk footprints less noisy and more realistic with respect to the case in which spatial correlation is neglected (Figure 5.13 a). Indeed, the distribution of losses tends to be very noisy when neglecting spatial correlation as a result of the ground motion being independent at neighbouring sites. Therefore, loss estimates are also independent owing to the homogeneity of the flat portfolio. In contrast, when the correlation is considered, closely spaced sites are likely to experience similar ground-motion levels, so that the distribution of losses shows smoother patterns. Differences are also evident when comparing Figure 5.13 b and c, which is when the associated spatial correlation uncertainty is accounted for. Such outcomes demonstrate the importance of considering not only spatial correlation, but also the event-to-event spatial correlation variability in seismic risk assessments. We note that these results refer to PGA; however, similar outcomes are obtained for the other IMs.

For sake of completeness, we show in Figure 5.14 the loss EP curves obtained for the different IMs. Rare losses are generally underestimated if spatial correlation is not accounted for. This is more evident for short period IMs, such as PGA and SA at T = 0.3 s, compared to long period IMs (SA at T = 0.6 and 1 s). Indeed, the spatial correlation model for Central Italy presented in Section 5.2.5 features ranges that decreases as the period increases up to 1 s. Therefore, the effects of considering spatially-correlated ground-motion fields diminish with increasing period. Besides, we note that both frequent and rare losses are generally higher if the correlation uncertainty is accounted for with respect to the simple spatial correlation case for short period IMs. In contrast, sound conclusions are difficult to draw at longer IMs, as the EP curves show negligible differences. Further analyses are still required to advance the understanding of including the event-to-event spatial correlation variability in loss estimates; however, we believe that these findings have significant implications for providing more informed seismic risk assessments.

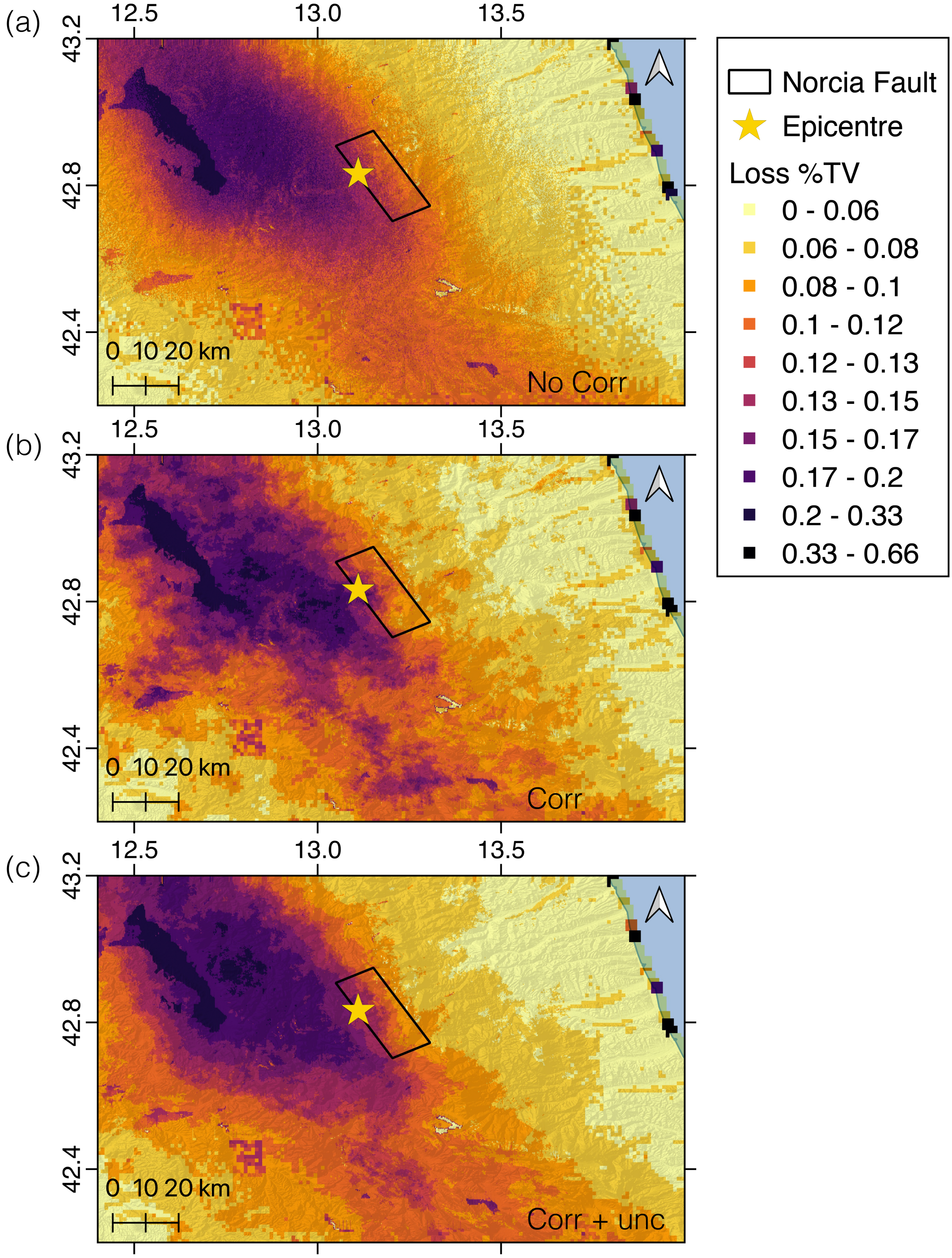


Figure 5.13: Average annual losses (AAL) for the flat homogenous portfolio (PGA) obtained for three different cases: (a) ground-motion fields are generated without considering correlation; (b) ground-motion fields are generated considering correlation; and (c) ground-motion fields are generated considering correlation and its associated uncertainty. The yellow star represents the Mw 6.5 Norcia epicentre, whereas the black rectangle defines the surface fault projection.

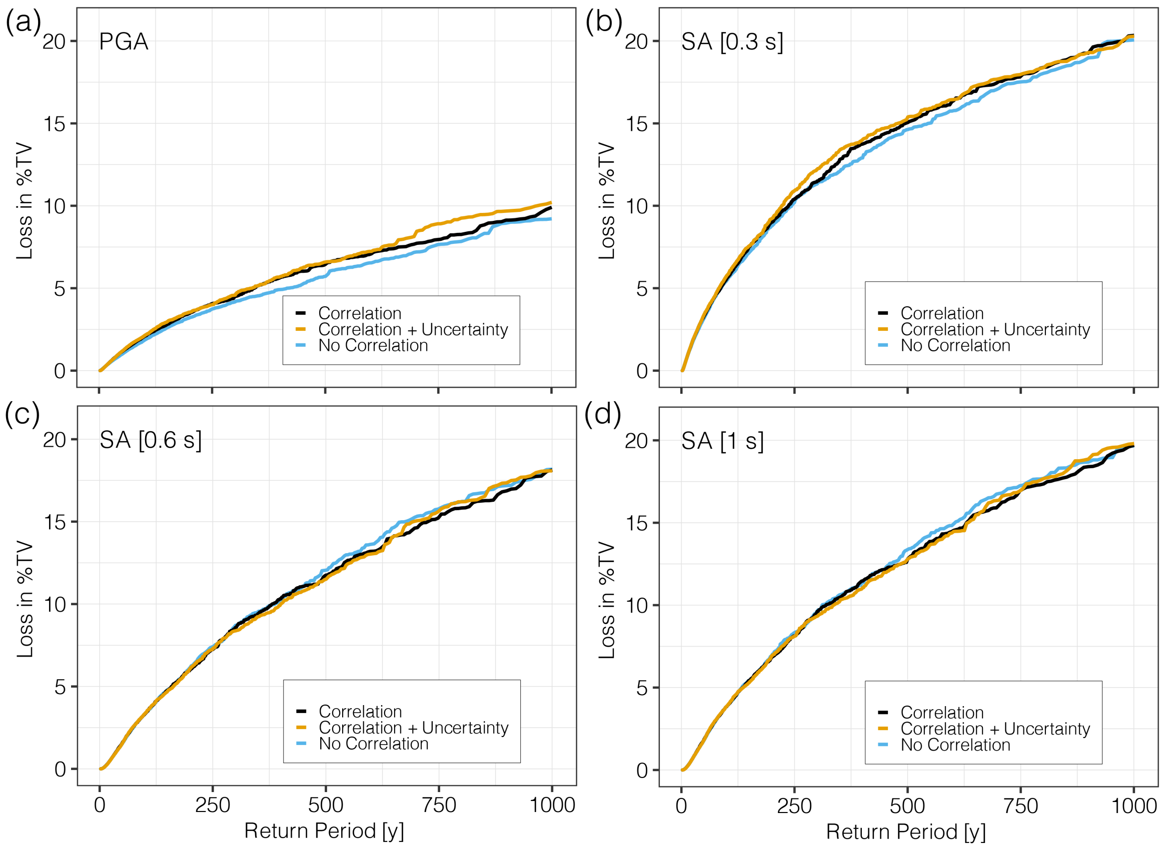


Figure 5.14: EP curves in terms of loss in %TV for three different cases: ground-motion fields are generated without considering correlation; ground-motion fields are generated considering correlation; and ground-motion fields are generated considering correlation and its associated uncertainty. Losses are computed for flat homogeneous portfolios: (a) PGA; (b) SA at T = 0.3 s; (c) SA at T = 0.6 s; (d) SA at T = 1 s.

5.5 Conclusions

In this work, we perform both scenario and event-based seismic hazard and risk calculations to advance the understanding of spatial correlations in the catastrophe modelling process. We first derive custom spatial correlation models based on the Mw 6.5 Norcia earthquake to better investigate to what extend different input models may affect the overall economic losses. Thereafter, we extend our analyses to probabilistic calculations, which are often important for decision-making and (re)insurance and underwriting purposes. In this context, we develop not only median spatial correlation models, but we also provide an estimate of the event-to-event associated dispersion to enable more realistic estimations of both the ground motion and corresponding losses.

Deterministic scenario calculations for the Norcia event suggest that the economic portfolio in terms of both heterogeneity and footprint dimension affect the impact of considering spatially-correlated ground-motion fields in risk analyses. Indeed, risk calculations for the 2020 SERA economic portfolio do not seem to be strongly affected by the specific correlation model, even though differences in the spatial distribution of losses may be spotted. To further highlight the role of different spatial correlation models, we thus assumed a flat portfolio, in which only a single building type is considered at each grid point. In agreement with Wesson and Perkins (2001) and Wagener et al. (2016), we find that the variance of losses varies according to the specific spatial correlation structure used in the generation of ground-motion fields. Therefore, these outcomes suggest that the hypotheses on which spatial correlation models are grounded (e.g. ground motion models to compute the within-event residuals) play a crucial role in the estimation of the overall potential losses for the area of interest. As a consequence, the implementation of a unique correlation model calibrated on worldwide databases may represent an oversimplification.

Event-based calculations shed light on the importance of considering not only the ground-motion spatial correlation, but also its associated uncertainty in risk analyses. Loss exceedance probability curves for both the SERA economic and flat portfolios indicate that neglecting spatial correlation leads to biases in loss estimates. In particular, smaller losses are expected at lower annual probabilities of exceedance. For return periods of interest for the (re)insurance industry (up to 1000 years), we note that EP curves show a similar behaviour, suggesting that overall losses are not strongly affected by the different correlation structures. Nonetheless, both hazard and risk footprints in terms of average annual losses feature less noisy and more realistic patterns if spatial correlation is taken into account. Therefore, this study strengthens the idea that the inclusion of spatial correlation and of its associated variability is crucial for enabling more informed risk assessments.

Although our outcomes demonstrate that spatial correlations in probabilistic seismic hazard and risk assessments improves per-event and in-location loss estimates for underwriting purposes, this study presents some caveats. When heterogeneous portfolios are of concern, not only it is important to consider the spatial correlation at closely-spaced sites, but the cross-correlation among different IMs should also be accounted for at once. Neglecting the cross-correlation would lead to independent ground-motion fields for different IMs and thus to inaccurate estimation of the likelihood of observing rare losses. The development of a novel approach to simulate spatially cross-correlated ground-motion fields indicates features for further developments of this preliminary study.

Finally, the consideration of isotropic spatial correlation models may represent a strong simplification. Although the non-parametric tests presented in Section 5.2.2 suggest that the hypothesis of isotropy is satisfied at a 5% significance level, several authors (Chen and Baker, 2019; Abbasnejadfard et al., 2020; Infantino et al., 2021; Schiappapietra and Smerzini, 2021) demonstrated that anisotropic models better capture the complex ground-motion patterns. Further modelling work should be undertaken to establish the effectiveness of including anisotropic spatial correlation structure in risk analyses.

CHAPTER 6

Conclusions

Modelling the spatial correlation of earthquake ground motions is a key ingredient in seismic hazard and risk assessment, especially when spatially-distributed building portfolios and infrastructures are of interest. The consideration of regional rather than site-specific seismic hazard has posed challenges to risk modellers and researchers, and therefore modelling the spatial variability of ground motions has gained increasing importance over the past couple of decades. Analyses of earthquake recordings have revealed that ground-motion residuals (specifically the within-event components) exhibit not only spatial correlation, but also cross-correlation among different intensity measures at the same site. The majority of spatial correlation models are based on the hypotheses of: (1) ground-motion residuals being jointly normally distributed, (2) stationarity, and (3) isotropy. While the first assumption has been demonstrated by Baker and Jayaram (2008), isotropy and stationarity are usually assumed for the sake of simplicity, especially due to the paucity of earthquake recordings. Furthermore, existing research recognises the critical role played by various elements, such as magnitude, structural period and geological conditions, on the spatial variability of ground motions; however, a systematic understanding of how such factors affect the spatial correlation was still lacking.

Therefore, this thesis focused on advancing the understanding of spatial dependency of ground motion as well as addressing different research gaps in spatial correlation modelling. The following provides a summary of the main conclusions and key contributions of this thesis along with limitations and recommendations for additional investigations.

6.1 Key findings

Throughout this thesis, we investigated the factors that most affect the spatial correlation of ground motion in order to enable more informed seismic hazard and risk assessments. In general, the outcomes suggest that spatial correlation is period-, regionally- and scenario-dependent, so that empirical correlation models calibrated on heterogeneous databases may not properly represent the spatial dependency of the ground motion in the region of interest.

In more detail, analyses of strong-motion data from the 2016-2017 Central Italy seismic sequence and 3D physics-based ground-motion simulations demonstrated that the range parameter within spatial correlation models depends on the structural period of interest, which is consistent with findings in the literature (e.g. Jayaram and Baker, 2009; Esposito and Iervolino, 2012; Wagener et al., 2016; Sgobba et al., 2019; Chen and Baker, 2019). A direct proportionality between the sill and period was also found.

The rate of decay of spatial correlation is affected by geological and local site conditions, whose effects are greater for short-period intensity measures compared to long-period measures. Homogeneous soil conditions result in a greater ground-motion correlation over longer inter-site distances, in contrast to more heterogeneous local site conditions. Many researchers (e.g Jayaram and Baker, 2009; Du and Wang, 2013; Sokolov et al., 2012; Garakaninezhad and Bastami, 2017) have accounted for such trends by considering the correlation of $V_{s,30}$ values. While employing country-specific databases to characterise the correlation structure of a region is a reasonable choice as it leads to more statistically robust estimates, our outcomes proved that such correlation properties correspond only to average trends in the data. A comparison between the Central Italy and the Po Plain regions is an instructive example. The former is characterised by shorter correlation lengths due to a higher degree of heterogeneities of the local site effects compared to the latter, where soil conditions are more homogeneous in terms of $V_{s,30}$. Similar conclusions can be inferred, for instance, from Sgobba et al. (2019), Sgobba et al. (2021) and Infantino et al. (2021) for Italy and Sokolov et al. (2012) and Sokolov and Wenzel (2013) for Taiwan.

Many researchers (e.g. Foulser-Piggott and Goda, 2015; Heresi and Miranda, 2019; Stafford et al., 2019; Kuehn and Abrahamson, 2020) have long debated the impact of magnitude on the spatial correlation. The study on the 2016-2017 Central Italy earthquake sequence did not evidence any well-defined trend between range and magnitude, although an average positive correlation may be found at longer periods. This is in agreement with previous studies (Sokolov and Wenzel, 2013; Foulser-Piggott and Goda, 2015), which stated that the correlation length tends to increase with increasing magnitude, owing to the stronger energy content at low frequencies for larger magnitude events. The analysis of 3D physics-based ground-motion simulations for the Norcia area allowed for deeper insights into the dependence of the range on magnitude. In the near-field, the spatial variability of the ground motion is strongly affected by: (1) the fault extension, (2) slip heterogeneities, and (3) the relative position of the hypocentre with respect to the main slip asperities. These factors may give rise to a higher range variability, so that the magnitude itself cannot be used as the only explanatory parameter to describe source effects. As a matter of fact, the application case-study presented in Chapter 5 suggested that the event-to-event correlation variability should be included in regional probabilistic risk analysis to enable more realistic estimations of the ground motion and corresponding losses.

In the literature, questions have been also raised about the impact on the correlation structure of the GMMs employed to compute the within-event residuals. Our results demonstrated that the range is not significantly affected by the GMM, provided that the selected models do not exhibit systematic trends in the residuals. Indeed, these biases stem from factors (e.g. anelastic attenuation) not properly accounted for by the GMM and would inevitably be reflected as artificial spatial correlation. We observed that the apparent dependence of the correlation properties on the GMM is greater for high-frequency intensity measures, such as PGA, compared to low-frequency ground motions. The different sensitivities of short- and long-period ground motions to the attenuation and local site conditions are found to be partially responsible for such trends.

In previous studies, little effort has been directed towards the characterization of the correlation structure of other ways of combining the two horizontal ground-motion components beyond the classic geometric mean of the two horizontal components and RotD50. In the near-field, the polarization of the ground motion can be significant due to phenomena such as directivity and rupture propagation effects, which may lead to diagnostic spatial correlation features. For instance, the M_w 6.5 Norcia earthquake simulations showed that the fault-normal and vertical components are characterised by a more coherent ground motion due to the effects of the slip distribution coupled with the faulting mechanism, which result in a larger correlation length with respect to the fault-parallel and geometric mean. We therefore believe that such strong variability in terms of correlation properties between different components should be accounted for in seismic hazard and risk assessments to provide a more accurate representation of ground-motion fields.

Although the least-squares regression is the preferable technique to fit experimental semivariograms rather that trial-and-error (manual fitting), it still presents several drawbacks. Some authors apply the same weight to all the experimental values, so that the model is fit equally over the full range of data, whereas others implement a weighted approach, prioritising either the number of pairs in each bin or the shorter separation distances. There is still a high degree of subjectivity in the choice of weights, in the same way as there is for the manual fitting. Moreover, additional investigations showed that the computation of experimental semivariograms depends on: (1) the bin size; (2) the number of recording stations; (3) the station location configuration; (4) the distance cut-off; and (5) the semivariogram estimator. Such dependencies motivated us to explore different methodologies to define the correlation structure of earthquake ground motions. In particular, we shed light on maximum-likelihood approaches, which directly estimate the correlation properties, without requiring intermediate steps. Our results suggest that maximum-likelihood methods may outperform the least-squares regression approaches, so that they can be considered as a valid alternative for developing spatial-correlation models in the future.

While the hypothesis of isotropy has been deemed valid in the analysis of earthquake recordings, 3D physics-based ground-motion simulations enabled us to explore possible preferential directions of the spatial variability as well as the relationship of such directions with the source mechanism and the frequency content of the ground motion. Our outcomes demonstrated that the assumption of isotropy, i.e. independence of correlation on the azimuth, may represent a strong simplification, especially in near-field regions. Non-parametric tests were also performed on the Italian dataset to verify the suitability of an isotropy model to represent the spatial correlation. We believe that the sparsity of densely earthquake observations prevented a robust assessment in this case.

Although deterministic scenario and event-based probabilistic risk calculations provided preliminary results, our analyses demonstrated the importance of considering not only the ground motion spatial correlation, but also its associated uncertainty to obtain more accurate loss estimates. For instance, both hazard and risk footprints in terms of average annual losses feature less noisy and more realistic patterns when spatial correlation is accounted for compared to the case in which it is neglected.

The region- and scenario-dependency of correlation properties poses great challenges to hazard analysts when selecting, developing or adapting spatial-correlation models for their region of interest. Therefore, this thesis can form a basis for users and researchers interested in investigating ground-motion spatial correlation. Key findings can be used by loss and risk modellers to understand, interpret and model spatial correlation to generate appropriate ground shaking maps and improve hazard and risk assessments.

6.2 Future research

Future studies and further investigations that would extend the analyses presented in this thesis are proposed in the following.

Firstly, it is understood that 3D physics-based ground-motion simulations provide a powerful tool to further advance the understanding of spatial correlation. Their main advantages are: (1) databases of unparalleled size, especially in the near-field, and (2) the possibility of carrying out sensitivity analyses to quantify the effects of physical phenomena on spatial correlation. Besides, these simulations may be used as a complementary tool to the classic ground motion prediction equations in seismic hazard and risk assessments. The inclusion of simulations from a set of rupture realizations in risk analysis would overcome several limitations, such as the hypothesis of isotropy and stationarity on which empirical spatial correlation models are usually grounded. Indeed, our outcomes along with other recent studies (Chen and Baker, 2019; Infantino et al., 2021) demonstrated that such assumptions may constitute a strong simplification. In addition, not only do such simulations inherently account for spatial correlation, they also naturally represent cross-correlation among different IMs, which is essential in the case of heterogeneous building portfolios. A comparison with the analyses presented in Chapter 5 should be instructive to better understand to which extend a proper description of the spatial variability of ground motions impact the final loss and risk estimates with respect to the standard ground motion prediction equation-based approach.

Key findings of this dissertation suggest that the spatial variability of earthquake ground motions varies depending upon the period and region as well as the specific event. This is of concern especially for those regions characterized by sparse seismic networks, for which developing a well-constrained correlation model is challenging. In addition to simulations, which however require validation, the development of methodologies, such as the backbone approach (e.g. Douglas, 2018) for ground-motion prediction, may constitute a research topic for developing spatial correlation models, especially for areas without dense strong-motion networks and/or low/moderate seismicity.

This dissertation mainly focuses on spatial correlation for single intensity measures. However, the consideration of spatial cross-correlation amongst intensity measures is also a crucial consideration in the loss estimation process, especially when heterogeneous building portfolios are of concern. Indeed, each building class is usually characterized by its own fundamental period of vibration, which is associated to specific intensity measure and therefore fragility/vulnerability model. A considerable amount of literature has been published on cross-correlation models. Nevertheless, the implementation of these models into software packages for the probabilistic assessment of seismic hazard and risk [e.g. Openquake-Engine (Pagani et al., 2014)] has posed challenges. In the same way as we fine tuned the spatial correlation approach, the inclusion of spatial cross-correlation in Openquake may indicate features for further developments.

Garakaninezhad and Bastami (2017), Abbasnejadfard et al. (2020), Infantino et al. (2021), Chen et al. (2021) and Schiappapietra and Smerzini (2021), among others, demonstrated that the hypothesis of isotropy is not generally valid. However, there has been no detailed investigation of the effects of considering anisotropic spatial correlation models on risk estimates. Preliminary work on this was recently undertaken by Abbasnejadfard et al. (2021). The authors found that isotropic models generally provide higher loss estimates at low probability of exceedance and underestimate loss values for events with a high frequency of occurrence with respect to anisotropic models. In particular, the order of magnitude of such effects depends upon the heterogeneity of soil conditions and the configuration of building portfolios. Results of Chapter 5 could be extended by including *ad hoc* anisotropic spatial correlation models for the region of interest to further investigate the impact of isotropic Vs anisotropic correlation properties. The implementation of such models should not require additional computational efforts. However, the calibration of *ad hoc* anisotropic spatial correlation models represents a great challenge owing to the dependency of anisotropy on local-site conditions, propagation path and earthquake source. Similarly to the isotropic case, uncertainties in the correlation properties should be considered to provide more realistic estimations of both the ground motion and corresponding losses.

The number of earthquake recordings has dramatically increased over the last decades, allowing relaxing the ergodic assumption in favour of a non-ergodic approach to model the ground motion. Standard approaches identify systematic and repeatable characteristics of the ground motion (e.g. source-specific, site-specific and path-specific components), which are used to adjust the median prediction of a GMM and hence significantly reduce the aleatory uncertainties. In recent years, more sophisticated approaches have been proposed. For instances, Landwehr et al. (2016) and Lanzano et al. (2021) developed new GMMs with spatially varying coefficients. Such models enable the regional variabilities to be inherently accounted for in the median estimates. An ergodic GMM coupled with a spatial correlation model should lead to similar results in terms of regionalization of ground motions. Further research could usefully compare hazard and risk estimates derived by both approaches to highlight their advantages and caveats.

Finally, the outcomes of Chapter 5 represent a starting point for the development of an earthquake catastrophe model for Italy. In particular, next steps would include: (1) the new seismic hazard model developed for Europe (*http://www.efehr.org/start/*), and (2) the development of *ad hoc* spatial cross-correlation models to achieve more reliable and robust estimates of catastrophe losses.

Bibliography

Abbasnejadfard, M., Bastami, M. and Fallah, A. (2020), 'Investigation of anisotropic spatial correlations of intra-event residuals of multiple earthquake intensity measures using latent dimensions method', *Geophysical Journal International* **222**(2), 1449–1469.

Abbasnejadfard, M., Bastami, M., Fallah, A. and Garakaninezhad, A. (2021), 'Analyzing the effect of anisotropic spatial correlations of earthquake intensity measures on the result of seismic risk and resilience assessment of the portfolio of buildings and infrastructure systems', *Bulletin of Earthquake Engineering* pp. 1–27.

Abrahamson, N. A. and Youngs, R. (1992), 'A stable algorithm for regression analyses using the random effects model', *Bulletin of the Seismological Society of America* **82**(1), 505–510.

Akkar, S. and Bommer, J. J. (2010), 'Empirical equations for the prediction of pga, pgv, and spectral accelerations in europe, the mediterranean region, and the middle east', *Seismological Research Letters* **81**(2), 195–206.

Ancheta, T. D., Darragh, R. B., Stewart, J. P., Seyhan, E., Silva, W. J., Chiou, B. S.-J., Wooddell, K. E., Graves, R. W., Kottke, A. R., Boore, D. M. et al. (2014), 'Nga-west2 database', *Earthquake Spectra* **30**(3), 989–1005.

Annaka, T., Yamazaki, F. and Katahira, F. (1997), Proposal of peak ground velocity and response spectra based on jma 87 type accelerometer records, *in* 'Proceedings, 24th JSCE Earthquake Engineering Symposium', Vol. 1, pp. 161–164.

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

Atkinson, G. and Goda, K. (2013), Probabilistic seismic hazard analysis of civil infrastructure, *in* 'Handbook of seismic risk analysis and management of civil infrastructure systems', Elsevier, pp. 3–28.

Baker, J. W. (2011), 'Conditional mean spectrum: Tool for ground-motion selection', *Journal of Structural Engineering* **137**(3), 322–331.

Baker, J. W. and Chen, Y. (2020), 'Ground motion spatial correlation fitting methods and estimation uncertainty', *Earthquake Engineering & Structural Dynamics* **49**(15), 1662–1681.

Baker, J. W. and Cornell, C. A. (2006), 'Correlation of response spectral values for multicomponent ground motions', *Bulletin of the seismological Society of America* **96**(1), 215–227.

Baker, J. W. and Jayaram, N. (2008), 'Correlation of spectral acceleration values from nga ground motion models', *Earthquake Spectra* **24**(1), 299–317.

Baker, J. W. and Lee, C. (2018), 'An improved algorithm for selecting ground motions to match a conditional spectrum', *Journal of Earthquake Engineering* **22**(4), 708–723.

Bates, D., Mächler, M., Bolker, B. and Walker, S. (2014), 'Fitting linear mixed-effects models using lme4', *arXiv preprint arXiv:1406.5823* .

Bindi, D., GFZ-Potsdam, D. G., Luzi, L. and di Milano-Pavia, I. S. (2010), 'Ground motion prediction equations (gmpes) derived from itaca. in: Project s4: Italian strong motion data base'.

Bindi, D., Pacor, F., Luzi, L., Puglia, R., Massa, M., Ameri, G. and Paolucci, R. (2011), 'Ground motion prediction equations derived from the italian strong motion database', *Bulletin of Earthquake Engineering* **9**(6), 1899–1920.

Boore, D. M. and Atkinson, G. M. (2008*a*), 'Ground-motion prediction equations for the average horizontal component of pga, pgv, and 5%-damped psa at spectral periods between 0.01 s and 10.0 s', *Earthquake Spectra* **24**(1), 99–138.

Boore, D. M. and Atkinson, G. M. (2008*b*), 'Ground-motion prediction equations for the average horizontal component of pga, pgv, and 5%-damped psa at spectral periods between 0.01 s and 10.0 s', *Earthquake Spectra* **24**(1), 99–138.

Boore, D. M., Gibbs, J. F., Joyner, W. B., Tinsley, J. C. and Ponti, D. J. (2003), 'Estimated ground motion from the 1994 northridge, california, earthquake at the site of the interstate 10 and la cienega boulevard bridge collapse, west los angeles, california', *Bulletin of the Seismological Society of America* **93**(6), 2737–2751.

Boore, D. M., Stewart, J. P., Seyhan, E. and Atkinson, G. M. (2014), 'Nga-west2 equations for predicting pga, pgv, and 5% damped psa for shallow crustal earthquakes', *Earthquake Spectra* **30**(3), 1057–1085.

Bowman, A. W. and Azzalini, A. (2018), *R package `sm`: nonparametric smoothing methods (version 2.2-5.6)*, University of Glasgow, UK and Università di Padova, Italia.
URL: *http://www.stats.gla.ac.uk/~adrian/sm/*

Bowman, A. W. and Crujeiras, R. M. (2013), 'Inference for variograms', *Computational Statistics & Data Analysis* **66**, 19–31.

Bradley, B. A. (2010*a*), 'A generalized conditional intensity measure approach and holistic ground-motion selection', *Earthquake Engineering & Structural Dynamics* **39**(12), 1321–1342.

Bradley, B. A. (2010*b*), 'Nz-specific pseudo-spectral acceleration ground motion prediction equations based on foreign models'.

Bradley, B. A. (2013), 'A new zealand-specific pseudospectral acceleration ground-motion prediction equation for active shallow crustal earthquakes based on foreign models', *Bulletin of the Seismological Society of America* **103**(3), 1801–1822.

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**, 83–91.

Bray, J. D. and Rodriguez-Marek, A. (2004), 'Characterization of forward-directivity ground motions in the near-fault region', *Soil dynamics and earthquake engineering* **24**(11), 815–828.

Bullock, Z. (2019), 'Ground motion models for arias intensity, cumulative absolute velocity, peak incremental ground velocity, and significant duration in new zealand', *Bulletin of the New Zealand Society for Earthquake Engineering* **52**(4), 193–207.

Campbell, K. W. and Bozorgnia, Y. (2003), 'Updated near-source ground-motion (attenuation) relations for the horizontal and vertical components of peak ground acceleration and acceleration response spectra', *Bulletin of the Seismological Society of America* **93**(1), 314–331.

Campbell, K. W. and Bozorgnia, Y. (2008), 'Nga ground motion model for the geometric mean horizontal component of pga, pgv, pgd and 5% damped linear elastic response spectra for periods ranging from 0.01 to 10 s', *Earthquake Spectra* **24**(1), 139–171.

Campbell, K. W. and Bozorgnia, Y. (2010), 'A ground motion prediction equation for the horizontal component of cumulative absolute velocity (cav) based on the peer-nga strong motion database', *Earthquake Spectra* **26**(3), 635–650.

Campbell, K. W. and Bozorgnia, Y. (2012), 'A comparison of ground motion prediction equations for arias intensity and cumulative absolute velocity developed using a consistent database and functional form', *Earthquake Spectra* **28**(3), 931–941.

Campbell, K. W. and Bozorgnia, Y. (2014), 'Nga-west2 ground motion model for the average horizontal components of pga, pgv, and 5% damped linear acceleration response spectra', *Earthquake Spectra* **30**(3), 1087–1115.

Cauzzi, C. and Faccioli, E. (2008), 'Broadband (0.05 to 20 s) prediction of displacement response spectra based on worldwide digital records', *Journal of Seismology* **12**(4), 453–475.

CEN (2004), 'Eurocode 8: Design of structures for earthquake resistance-part 1: general rules, seismic actions and rules for buildings', *Brussels, Comite Europeen de Normalisation (CEN)* .

Chen, Y. and Baker, J. W. (2019), 'Spatial correlations in cybershake physics-based ground-motion simulations', *Bulletin of the Seismological Society of America* **109**(6), 2447–2458.

Chen, Y., Bradley, B. A. and Baker, J. W. (2021), 'Nonstationary spatial correlation in new zealand strong ground-motion data', *Earthquake Engineering & Structural Dynamics* .

Chiaraluce, L., Di Stefano, R., Tinti, E., Scognamiglio, L., Michele, M., Casarotti, E., Cattaneo, M., De Gori, P., Chiarabba, C., Monachesi, G. et al. (2017), 'The 2016 central italy seismic sequence: A first look at the mainshocks, aftershocks, and source models', *Seismological Research Letters* **88**(3), 757–771.

Chiou, B.-J. and Youngs, R. R. (2008), 'An nga model for the average horizontal component of peak ground motion and response spectra', *Earthquake spectra* **24**(1), 173–215.

Costa, C., Silva, V. and Bazzurro, P. (2018), 'Assessing the impact of earthquake scenarios in transportation networks: the portuguese mining factory case study', *Bulletin of earthquake engineering* **16**(3), 1137–1163.

Costanzo, A. (2018), 'Shaking maps based on cumulative absolute velocity and arias intensity: the cases of the two strongest earthquakes of the 2016–2017 central italy seismic sequence', *ISPRS International Journal of Geo-Information* **7**(7), 244.

Cressie, N. (1985), 'Fitting variogram models by weighted least squares', *Journal of the international Association for mathematical Geology* **17**(5), 563–586.

Crowley, H., Despotaki, V., Rodrigues, D., Silva, V., Toma-Danila, D., Riga, E., Karatzetzou, A., Fotopoulou, S., Zugic, Z., Sousa, L. et al. (2020), 'Exposure model for european seismic risk assessment', *Earthquake Spectra* **36**(1_suppl), 252–273.

Crowley, H., Despotaki, V., Silva, V., Dabbeek, J., Romão, X., Pereira, N., Castro, J. M., Daniell, J., Veliu, E., Bilgin, H. et al. (2021), 'Model of seismic design lateral force levels for the existing reinforced concrete european building stock', *Bulletin of Earthquake Engineering* **19**(7), 2839–2865.

Crowley, H., Stafford, P. J. and Bommer, J. J. (2008), 'Can earthquake loss models be validated using field observations?', *Journal of Earthquake Engineering* **12**(7), 1078–1104.

Diggle, P., Ribeiro, P. and Geostatistics, M.-b. (2007), 'Springer series in statistics'.

Douglas, J. (2018), Capturing geographically-varying uncertainty in earthquake ground motion models or what we think we know may change, *in* 'European Conference on Earthquake Engineering Thessaloniki, Greece', Springer, pp. 153–181.

Douglas, J. and Edwards, B. (2016), 'Recent and future developments in earthquake ground motion estimation', *Earth-Science Reviews* **160**, 203–219.

Du, W. and Ning, C.-L. (2021), 'Modeling spatial cross-correlation of multiple ground motion intensity measures (sas, pga, pgv, ia, cav, and significant durations) based on principal component and geostatistical analyses', *Earthquake Spectra* **37**(1), 486–504.

Du, W. and Wang, G. (2013), 'Intra-event spatial correlations for cumulative absolute velocity, arias intensity, and spectral accelerations based on regional site conditions', *Bulletin of the Seismological Society of America* **103**(2A), 1117–1129.

Erdik, M. (2017), 'Earthquake risk assessment', *Bulletin of Earthquake Engineering* **15**(12), 5055–5092.

Esposito, S. and Iervolino, I. (2011), 'Pga and pgv spatial correlation models based on european multievent datasets', *Bulletin of the Seismological Society of America* **101**(5), 2532–2541.

Esposito, S. and Iervolino, I. (2012), 'Spatial correlation of spectral acceleration in european data', *Bulletin of the Seismological Society of America* **102**(6), 2781–2788.

Foulser-Piggott, R. and Goda, K. (2015), 'Ground-motion prediction models for arias intensity and cumulative absolute velocity for japanese earthquakes considering single-station sigma and within-event spatial correlation', *Bulletin of the Seismological Society of America* **105**(4), 1903–1918.

Garakaninezhad, A. and Bastami, M. (2017), 'A novel spatial correlation model based on anisotropy of earthquake ground-motion intensitya novel spatial correlation model based on anisotropy of earthquake ground-motion intensity', *Bulletin of the Seismological Society of America* **107**(6), 2809–2820.

Goda, K. (2011), 'Interevent variability of spatial correlation of peak ground motions and response spectra', *Bulletin of the Seismological Society of America* **101**(5), 2522–2531.

Goda, K. and Atkinson, G. M. (2009), 'Probabilistic characterization of spatially correlated response spectra for earthquakes in japan', *Bulletin of the Seismological Society of America* **99**(5), 3003–3020.

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**(6), 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.

Goovaerts, P. et al. (1997), *Geostatistics for natural resources evaluation*, Oxford University Press on Demand.

Grossi, P. and Kunreuther, H. (2005), *Catastrophe modeling: a new approach to managing risk*, Vol. 25, Springer Science & Business Media.

Hashemi, K., Mazzieri, I., Paolucci, R., Smerzini, C. et al. (2015), Spatial variability of near-source seismic ground motion with respect to different distance metrics, *in* '7th Int. Conf. on Seismology and Earthquake Engineering', pp. 1–8.

Heresi, P. and Miranda, E. (2019), 'Uncertainty in intraevent spatial correlation of elastic pseudo-acceleration spectral ordinates', *Bulletin of Earthquake Engineering* **17**(3), 1099–1115.

Herrero, A. and Bernard, P. (1994), 'A kinematic self-similar rupture process for earthquakes', *Bulletin of the Seismological Society of America* **84**(4), 1216–1228.

Hong, H., Zhang, Y. and Goda, K. (2009), 'Effect of spatial correlation on estimated ground-motion prediction equations', *Bulletin of the Seismological Society of America* **99**(2A), 928–934.

Hough, S. and Field, E. (1996), 'On the coherence of ground motion in the san fernando valley', *Bulletin of the Seismological Society of America* **86**(6), 1724–1732.

Huang, C. and Galasso, C. (2019), 'Ground-motion intensity measure correlations observed in italian strong-motion records', *Earthquake Engineering & Structural Dynamics* **48**(15), 1634–1660.

Huang, C., Tarbali, K. and Galasso, C. (2020), 'Correlation properties of integral ground-motion intensity measures from italian strong-motion records', *Earthquake Engineering & Structural Dynamics* **49**(15), 1581–1598.

Huang, C., Tarbali, K., Galasso, C. and Paolucci, R. (2020), Spatial correlation validation of 3d physics-based ground-motion simulations., *in* '17th World Conference in Earthquake Engineering'.

Iervolino, I. (2013), 'Probabilities and fallacies: Why hazard maps cannot be validated by individual earthquakes', *Earthquake Spectra* **29**(3), 1125–1136.

Infantino, M., Paolucci, R., Smerzini, C. and Stupazzini, M. (2018), Study of the spatial correlation of earthquake ground motion by means of physics-based numerical scenarios, *in* '16th European Conference on Earthquake Engineering', pp. 1–12.

Infantino, M., Smerzini, C. and Lin, J. (2021), 'Spatial correlation of broadband ground motions from physics-based numerical simulations', *Earthquake Engineering & Structural Dynamics* .

Jayaram, N. and Baker, J. W. (2009), 'Correlation model for spatially distributed ground-motion intensities', *Earthquake Engineering & Structural Dynamics* **38**(15), 1687–1708.

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**(6), 3295–3303.

Joyner, W. B. and Boore, D. M. (1993), 'Methods for regression analysis of strong-motion data', *Bulletin of the Seismological Society of America* **83**(2), 469–487.

Kanno, T., Narita, A., Morikawa, N., Fujiwara, H. and Fukushima, Y. (2006), 'A new attenuation relation for strong ground motion in japan based on recorded data', *Bulletin of the Seismological Society of America* **96**(3), 879–897.

Keefer, D. L. and Bodily, S. E. (1983), 'Three-point approximations for continuous random variables', *Management Science* **29**(5), 595–609.

Kohrangi, M., Bazzurro, P. and Vamvatsikos, D. (2021), 'Seismic risk and loss estimation for the building stock in isfahan: part ii,Äîhazard analysis and risk assessment', *Bulletin of Earthquake Engineering* **19**(4), 1739–1763.

Kotha, S. R., Bindi, D. and Cotton, F. (2017), 'From ergodic to region-and site-specific probabilistic seismic hazard assessment: Method development and application at european and middle eastern sites', *Earthquake spectra* **33**(4), 1433–1453.

Kotha, S. R., Cotton, F. and Bindi, D. (2018), 'A new approach to site classification: Mixed-effects ground motion prediction equation with spectral clustering of site amplification functions', *Soil Dynamics and Earthquake Engineering* **110**, 318–329.

Kotha, S. R., Weatherill, G., Bindi, D. and Cotton, F. (2020), 'A regionally-adaptable ground-motion model for shallow crustal earthquakes in europe', *Bulletin of Earthquake Engineering* **18**, 4091–4125.

Kramer, S. L. et al. (1996), *Geotechnical earthquake engineering*, Pearson Education India.

Kuehn, N. M. and Abrahamson, N. A. (2020), 'Spatial correlations of ground motion for non-ergodic seismic hazard analysis', *Earthquake Engineering & Structural Dynamics* **49**(1), 4–23.

Künsch, H., Papritz, A. J., Schwierz, C. and Stahel, W. A. (2013), Robust estimation of the external drift and the variogram of spatial data, *in* 'ISI 58th World Statistics Congress of the International Statistical Institute', Eidgenössische Technische Hochschule Zürich.

Landwehr, N., Kuehn, N. M., Scheffer, T. and Abrahamson, N. (2016), 'A nonergodic ground-motion model for california with spatially varying coefficients', *Bulletin of the Seismological Society of America* **106**(6), 2574–2583.

Lanzano, G., D'Amico, M., Felicetta, C., Puglia, R., Luzi, L., Pacor, F. and 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., Luzi, L., D'Amico, V., Pacor, F., Meletti, C., Marzocchi, W., Rotondi, R. and Varini, E. (2020), 'Ground motion models for the new seismic hazard model of italy (mps19): selection for active shallow crustal regions and subduction zones', *Bulletin of Earthquake Engineering* **18**(8), 3487–3516.

Lanzano, G., Luzi, L., Pacor, F., Felicetta, C., Puglia, R., Sgobba, S. and D'Amico, M. (2019), 'A revised ground-motion prediction model for shallow crustal earthquakes in italya revised ground-motion prediction model for shallow crustal earthquakes in italy', *Bulletin of the Seismological Society of America* **109**(2), 525–540.

Lanzano, G., Pacor, F., Luzi, L., D'amico, M., Puglia, R. and 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.

Lanzano, G., Sgobba, S., Caramenti, L. and Menafoglio, A. (2021), 'Ground-motion model for crustal events in italy by applying the multisource geographically weighted regression (ms-gwr) method', *Bulletin of the Seismological Society of America* .

Lanzano, G., Sgobba, S., Luzi, L., Puglia, R., Pacor, F., Felicetta, C., D'Amico, M., Cotton, F. and Bindi, D. (2019), 'The pan-european engineering strong motion (esm) flatfile: compilation criteria and data statistics', *Bulletin of Earthquake Engineering* **17**(2), 561–582.

Lee, J. (2009), Engineering Characterization of Earthquake Ground Motions., PhD thesis.

Li, Z., Zhang, X., Clarke, K. C., Liu, G. and Zhu, R. (2018), 'An automatic variogram modeling method with high reliability fitness and estimates', *Computers & geosciences* **120**, 48–59.

Loth, C. and Baker, J. W. (2013), 'A spatial cross-correlation model of spectral accelerations at multiple periods', *Earthquake Engineering & Structural Dynamics* **42**(3), 397–417.

Luzi, L., Pacor, F., Lanzano, G., Felicetta, C., Puglia, R. and D'Amico, M. (2019), '2016–2017 central italy seismic sequence: strong-motion data analysis and design earthquake selection for seismic microzonation purposes', *Bulletin of Earthquake Engineering* pp. 1–19.

Luzi, L., Pacor, F., Puglia, R., Lanzano, G., Felicetta, C., D'Amico, M., Michelini, A., Faenza, L., Lauciani, V., Iervolino, I. et al. (2017), 'The central italy seismic sequence between august and december 2016: Analysis of strong-motion observations', *Seismological Research Letters* **88**(5), 1219–1231.

Markhvida, M., Ceferino, L. and Baker, J. W. (2018), 'Modeling spatially correlated spectral accelerations at multiple periods using principal component analysis and geostatistics', *Earthquake Engineering & Structural Dynamics* **47**(5), 1107–1123.

Martins, L. and Silva, V. (2020), 'Development of a fragility and vulnerability model for global seismic risk analyses', *Bulletin of Earthquake Engineering* pp. 1–27.

Matheron, G. (1962), *Traité de géostatistique appliquée*, number 14, Editions Technip.

Mazzieri, I., Stupazzini, M., Guidotti, R. and Smerzini, C. (2013), 'Speed: Spectral elements in elastodynamics with discontinuous galerkin: A non-conforming approach for 3d multi-scale problems', *International Journal for Numerical Methods in Engineering* **95**(12), 991–1010.

Michelini, A., Faenza, L., Lanzano, G., Lauciani, V., Jozinović, D., Puglia, R. and Luzi, L. (2020), 'The new shakemap in italy: progress and advances in the last 10 yr', *Seismological Research Letters* **91**(1), 317–333.

Midorikawa, S. (2002), Attenuation relationships of peak ground acceleration and velocity considering attenuation characteristics for shallow and deeper earthquakes, *in* 'Proceedings of the 11th Earthquake Engineering Symposium, 2002'.

Ming, D., Huang, C., Peters, G. W. and Galasso, C. (2019), 'An advanced estimation algorithm for ground-motion models with spatial correlationan advanced estimation algorithm for ground-motion models with spatial correlation', *Bulletin of the Seismological Society of America* **109**(2), 541–566.

Mori, F., Mendicelli, A., Moscatelli, M., Romagnoli, G., Peronace, E. and Naso, G. (2020), 'A new vs30 map for italy based on the seismic microzonation dataset', *Engineering Geology* **275**, 105745.

Müller, S. and Schüler, L. (2021), *GeoStat - Framework GSTools*.
URL: *https://doi.org/10.5281/zenodo.1313628*

Musson, R. M. (2000), 'The use of monte carlo simulations for seismic hazard assessment in the uk'.

Oliver, M. and Webster, R. (2014), 'A tutorial guide to geostatistics: Computing and modelling variograms and kriging', *Catena* **113**, 56–69.

Özcebe, A., Smerzini, C., Paolucci, R., Pourshayegan, H., Rodríguez Plata, R., Lai, C., Zuccolo, E., Bozzoni, F. and Villani, M. (2019), On the comparison of 3d, 2d, and 1d numerical approaches to predict seismic site amplification: the case of norcia basin during the m6. 5 2016 october 30 earthquake, *in* '7th International Conference on Earthquake Geotechnical Engineering', pp. 4251–4258.

Pagani, M., Monelli, D., Weatherill, G., Danciu, L., Crowley, H., Silva, V., Henshaw, P., Butler, L., Nastasi, M., Panzeri, L. et al. (2014), 'Openquake engine: An open hazard (and risk) software for the global earthquake model', *Seismological Research Letters* **85**(3), 692–702.

Paolucci, R., Evangelista, L., Mazzieri, I. and Schiappapietra, E. (2016), 'The 3d numerical simulation of near-source ground motion during the marsica earthquake, central italy, 100 years later', *Soil Dynamics and Earthquake Engineering* **91**, 39–52.

Paolucci, R., Gatti, F., Infantino, M., Smerzini, C., Özcebe, A. G. and Stupazzini, M. (2018), 'Broadband ground motions from 3d physics-based numerical simulations using artificial neural networksbroadband ground motions from 3d pbss using anns', *Bulletin of the Seismological Society of America* **108**(3A), 1272–1286.

Paolucci, R., Mazzieri, I. and Smerzini, C. (2015), 'Anatomy of strong ground motion: near-source records and three-dimensional physics-based numerical simulations of the Mw 6.0 2012 May 29 Po Plain earthquake, Italy', *Geophysical Journal International* **203**(3), 2001–2020.
URL: *https://doi.org/10.1093/gji/ggv405*

Paolucci, R., Rovelli, A., Faccioli, E., Cauzzi, C., Finazzi, D., Vanini, M., Di Alessandro, C. and Calderoni, G. (2008), 'On the reliability of long-period response spectral ordinates from digital accelerograms', *Earthquake engineering & structural dynamics* **37**(5), 697–710.

Papritz, A. (2018), 'Tutorial and manual for geostatistical analyses with the r package georob'.

Papritz, A. (2020), *georob: Robust Geostatistical Analysis of Spatial Data*. R package version 0.3-13.
URL: *https://CRAN.R-project.org/package=georob*

Park, J., Bazzurro, P., Baker, J. W. et al. (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* pp. 1–8.

Pebesma, E. J. (2004), 'Multivariable geostatistics in s: the gstat package', *Computers and Geosciences* **30**, 683–691.

Pizzi, A., Di Domenica, A., Gallovič, F., Luzi, L. and Puglia, R. (2017), 'Fault segmentation as constraint to the occurrence of the main shocks of the 2016 central italy seismic sequence', *Tectonics* **36**(11), 2370–2387.

R Core Team (2019), *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria.
URL: *https://www.R-project.org/*

Ribeiro Jr, P. J., Diggle, P. J., Schlather, M., Bivand, R. and Ripley, B. (2020), *geoR: Analysis of Geostatistical Data*. R package version 1.8-1.
URL: *https://CRAN.R-project.org/package=geoR*

Schiappapietra, E. and Douglas, J. (2020), 'Modelling the spatial correlation of earthquake ground motion: insights from the literature, data from the 2016–2017 central italy earthquake sequence and ground-motion simulations', *Earth-Science Reviews* **203**, 103139.

Schiappapietra, E. and Smerzini, C. (2021), 'Spatial correlation of broadband earthquake ground motion in norcia (central italy) from physics-based simulations.', *Bulletin of Earthquake Engineering* .

Schwanghart, W. (2021*a*), *Experimental (Semi-) Variogram*.
URL: *https://www.mathworks.com/matlabcentral/fileexchange/20355-experimental-semi-variogram*

Schwanghart, W. (2021*b*), *Variogramfit*.
URL: *https://www.mathworks.com/matlabcentral/fileexchange/25948-variogramfit*

Sgobba, S., Lanzano, G. and Pacor, F. (2021), 'Empirical nonergodic shaking scenarios based on spatial correlation models: An application to central italy', *Earthquake Engineering & Structural Dynamics* **50**(1), 60–80.

Sgobba, S., Lanzano, G., Pacor, F., Puglia, R., D'Amico, M., Felicetta, C. and Luzi, L. (2019), 'Spatial correlation model of systematic site and path effects for ground-motion fields in northern italyspatial correlation model of systematic site and path effects for ground-motion fields in northern italy', *Bulletin of the Seismological Society of America* **109**(4), 1419–1434.

Silva, V. (2016), 'Critical issues in earthquake scenario loss modeling', *Journal of Earthquake Engineering* **20**(8), 1322–1341.

Silva, V., Amo-Oduro, D., Calderon, A., Costa, C., Dabbeek, J., Despotaki, V., Martins, L., Pagani, M., Rao, A., Simionato, M. et al. (2020), 'Development of a global seismic risk model', *Earthquake Spectra* **36**(1_suppl), 372–394.

Silva, V., Crowley, H., Pinho, R. and Varum, H. (2014), 'Development of an open-source platform for calculating losses from earthquakes', *University of Aveiro, Portugal* .

Smerzini, C., Galasso, C., Iervolino, I. and Paolucci, R. (2014), 'Ground motion record selection based on broadband spectral compatibility', *Earthquake Spectra* **30**(4), 1427–1448.

Smerzini, C. and Pitilakis, K. (2018), 'Seismic risk assessment at urban scale from 3d physics-based numerical modeling: the case of thessaloniki', *Bulletin of Earthquake Engineering* **16**(7), 2609–2631.

Sokolov, V. and Ismail-Zadeh, A. (2016), 'On the use of multiple-site estimations in probabilistic seismic-hazard assessment', *Bulletin of the Seismological Society of America* **106**(5), 2233–2243.

Sokolov, V. and Wenzel, F. (2011), 'Influence of spatial correlation of strong ground motion on uncertainty in earthquake loss estimation', *Earthquake engineering & structural dynamics* **40**(9), 993–1009.

Sokolov, V. and Wenzel, F. (2013), 'Further analysis of the influence of site conditions and earthquake magnitude on ground-motion within-earthquake correlation: analysis of pga and pgv data from the k-net and the kik-net (japan) networks', *Bulletin of Earthquake Engineering* **11**(6), 1909–1926.

Sokolov, V. and Wenzel, F. (2019), 'Areal exceedance of ground motion as a characteristic of multiple-site seismic hazard: Sensitivity analysis', *Soil Dynamics and Earthquake Engineering* **126**, 105770.

Sokolov, V., Wenzel, F. and Kuo-Liang, W. (2010), 'Uncertainty and spatial correlation of earthquake ground motion in taiwan', *TAO: Terrestrial, Atmospheric and Oceanic Sciences* **21**(6), 9.

Sokolov, V., Wenzel, F., Wen, K.-L. and Jean, W.-Y. (2012), 'On the influence of site conditions and earthquake magnitude on ground-motion within-earthquake correlation: analysis of pga data from tsmip (taiwan) network', *Bulletin of Earthquake Engineering* **10**(5), 1401–1429.

Somerville, P. G., Smith, N. F., Graves, R. W. and Abrahamson, N. A. (1997), 'Modification of empirical strong ground motion attenuation relations to include the amplitude and duration effects of rupture directivity', *Seismological research letters* **68**(1), 199–222.

Stafford, P. J. (2012), 'Evaluation of structural performance in the immediate aftermath of an earthquake: A case study of the 2011 christchurch earthquake', *International Journal of Forensic Engineering* **1**(1), 58–77.

Stafford, P. J., Zurek, B. D., Ntinalexis, M. and Bommer, J. J. (2019), 'Extensions to the groningen ground-motion model for seismic risk calculations: component-to-component variability and spatial correlation', *Bulletin of Earthquake Engineering* **17**(8), 4417–4439.

Strasser, F. O. and Bommer, J. J. (2009), 'Strong ground motions,Äîhave we seen the worst?', *Bulletin of the Seismological Society of America* **99**(5), 2613–2637.

Stupazzini, M., Paolucci, R. and Igel, H. (2009), 'Near-fault earthquake ground-motion simulation in the grenoble valley by a high-performance spectral element code', *Bulletin of the Seismological Society of America* **99**(1), 286–301.

Toro, G. R. (2002), 'Modification of the toro et al.(1997) attenuation equations for large magnitudes and short distances', *Risk Engineering Technical Report* .

Tsai, C.-C. P., Chen, Y.-H. and Liu, C.-H. (2006), 'The path effect in ground-motion variability: An application of the variance-components technique', *Bulletin of the Seismological Society of America* **96**(3), 1170–1176.

Verros, S. A., Wald, D. J., Worden, C. B., Hearne, M. and Ganesh, M. (2017), 'Computing spatial correlation of ground motion intensities for shakemap', *Computers & Geosciences* **99**, 145–154.

Wagener, T., Goda, K., Erdik, M., Daniell, J. and Wenzel, F. (2016), 'A spatial correlation model of peak ground acceleration and response spectra based on data of the istanbul earthquake rapid response and early warning system', *Soil Dynamics and Earthquake Engineering* **85**, 166–178.

Wald, D. J. and Allen, T. I. (2007), 'Topographic slope as a proxy for seismic site conditions and amplification', *Bulletin of the Seismological Society of America* **97**(5), 1379–1395.

Wang, M. and Takada, T. (2005), 'Macrospatial correlation model of seismic ground motions', *Earthquake spectra* **21**(4), 1137–1156.

Weatherill, G., Silva, V., Crowley, H. and Bazzurro, P. (2015), 'Exploring the impact of spatial correlations and uncertainties for portfolio analysis in probabilistic seismic loss estimation', *Bulletin of Earthquake Engineering* **13**(4), 957–981.

Webster, R. and Oliver, M. A. (2007), *Geostatistics for environmental scientists*, John Wiley & Sons.

Wesson, R. L. and Perkins, D. M. (2001), 'Spatial correlation of probabilistic earthquake ground motion and loss', *Bulletin of the Seismological Society of America* **91**(6), 1498–1515.

Woessner, J., Laurentiu, D., Giardini, D., Crowley, H., Cotton, F., Grünthal, G., Valensise, G., Arvidsson, R., Basili, R., Demircioglu, M. B. et al. (2015), 'The 2013 european seismic hazard model: key components and results', *Bulletin of Earthquake Engineering* **13**(12), 3553–3596.

Wu, J. and Baker, J. (2014), Ground motion modeling for risk and reliability assessment of san francisco infrastructure systems, *in* 'Proceedings of the tenth US national conference on earthquake engineering, Anchorage, Alaska'.

Zerva, A. and Zervas, V. (2002), 'Spatial variation of seismic ground motions: an overview', *Appl. Mech. Rev.* **55**(3), 271–297.