Integreating geological and seismological data in point process models for seismical analysis

Marianna Siino, Giada Adelfio

Dipartimento di Scienze Economiche, Aziendali e Statistiche, Universitá degli Studi di Palermo

Wageningen, 9th May 2017

Contents

Earthquakes

What is it?

- it is a sudden movement of the earth lithosphere
- unpredictable natural disaster

Main causes (excluding human activities):

- volcanic activity

Earthquakes

What is it?

- it is a sudden movement of the earth lithosphere
- unpredictable natural disaster

Main causes (excluding human activities):

- volcanic activity
- tectonic activity
  (events occur along the boundary plates and active faults)

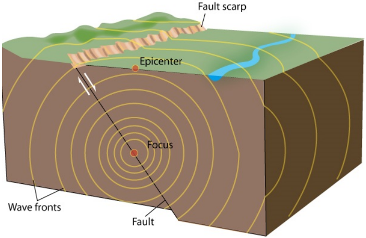


Figure: Seismic waves from a focus of an earthquake

Earthquakes

DATA + **STATISTICAL ANALYSIS** $\rightarrow$ give supports to policy decisions with

HAZARD MAPS

characterise the seismicity

PREDICTIVE MAPS

show the probability of occurrence considering the time

Earthquakes

DATA + **STATISTICAL ANALYSIS** $\rightarrow$ give supports to policy decisions with

HAZARD MAPS

characterise the seismicity

PREDICTIVE MAPS

show the probability of occurrence considering the time

Statistical models based on **Point process theory** (Illian et al.; 2008)

Earthquakes

DATA + STATISTICAL ANALYSIS → give supports to policy decisions with

HAZARD MAPS

characterise the seismicity

PREDICTIVE MAPS

show the probability of occurrence considering the time

Statistical models based on **Point process theory** (Illian et al.; 2008)

The most used is the **ETAS model** (Ogata, et all.; 2006):

- it explains induced activity of the phenomena
- input information: spatio-temporal coordinates
- DOES NOT include the effect of COVARIATES (external environmental variables, such as geological information in the study area)

Research goal

Integreation in point process models earthquake data and external geological information

↓

Research goal

Integration in point process models earthquake data and external geological information

↓

- to have more accurate risk/predictive maps
- to investigate and quantify dependencies of the seismicity on the covariates

Research goal

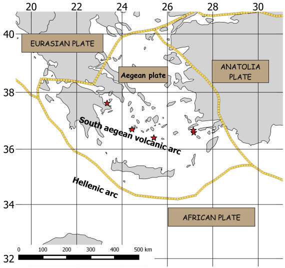
Integration in point process models earthquake data and external geological information

↓

- to have more accurate risk/predictive maps
- to investigate and quantify dependencies of the seismicity on the covariates

Application
Study area: Greece (the most seismic European-Mediterranean region)

Siino, M., et al (2016). Spatial pattern analysis using hybrid models: an application to the hellenic seismicity, Stochastic Environmental Research and Risk Assessment



Data

Seismic catalog

Seismic catalog

It contains:

- time
- focal parameters: latitude, longitude and depth
- magnitude

Seismic catalog

It contains:

- time
- focal parameters: latitude, longitude and depth
- magnitude

DATE (yyyy-month-dd)			TIME (GMT)	LAT. (N)	LONG. (E)	DEPTH (km)	MAGNITUDE (Local)
2017	MAY	6	14 57 42.4	39.08	23.24	9	3.4
2017	MAY	6	15 09 13.7	39.54	26.08	11	2.5
2017	MAY	6	16 55 53.5	39.54	23.25	11	1.6
2017	MAY	6	17 55 23.1	39.08	23.26	11	1.4
2017	MAY	6	18 09 04.8	38.46	23.49	9	1.7
2017	MAY	6	18 58 11.7	38.45	23.48	18	1.2
2017	MAY	6	20 47 40.9	41.29	23.52	11	1.4
2017	MAY	6	20 48 40.3	38.38	20.46	12	0.9
2017	MAY	6	21 32 24.8	40.19	20.57	10	1.9
...	...	...	...	...	...	...	...

Seismic catalog

Quality aspects:

- Accuracy of focal parameters
  - a volume is represented by a point
  - depends on the magnitude and number of seismic stations

Seismic catalog

Quality aspects:

- Accuracy of focal parameters
  - a volume is represented by a point
  - depends on the magnitude and number of seismic stations
- Catalog completeness
  - M_c, the minimum magnitude above which all earthquakes within a certain region are reliably recorded (Mignan and Woessner; 2012)

Seismic catalog

- **Source:** Hellenic Unified Seismological Network, (HUSN) - Institute of Geodynamics *(http://www.gein.noa.gr/en/seismicity/earthquake-catalogs)*
- **Period:** 1964 - 2017
- **Number of stations:** 150

Seismic catalog

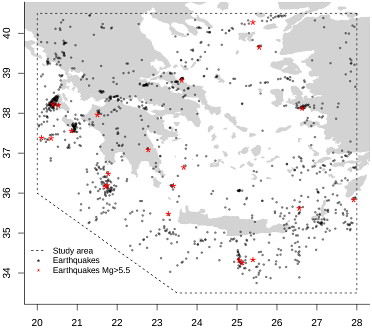
- **Source:** Hellenic Unified Seismological Network, (HUSN) - Institute of Geodynamics *(http://www.gein.noa.gr/en/seismicity/earthquake-catalogs)*
- **Period:** 1964 - 2017
- **Number of stations:** 150
- M_c**:** 2.4 (data from 2005 to 2014)
- HUSN part of the European Integrated Data Archive *(http://www.orfeus-eu.org/data/eida/)*

Seismic catalog

- **Selected subset:** 1105 events with magnitude $\geqslant 4$ between 2005 and 2014

Seismic catalog

- **Selected subset:** 1105 events with magnitude $\geqslant 4$ between 2005 and 2014
- **Main features:**
  - Spatial inhomogeneity
  - Cluster behaviour
  - Multi-scale interactions (clustering changes with the distances)
  - Random shifting of coincident points (Baddeley et al.2015)

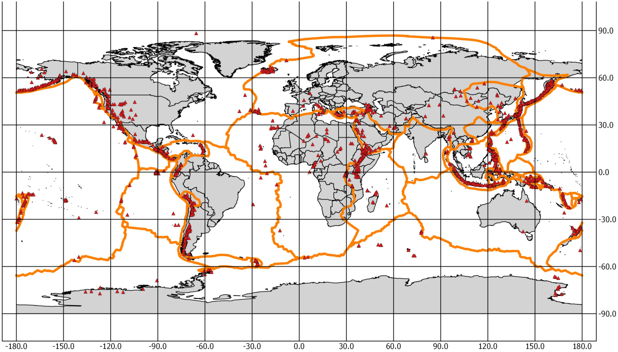


Data

Seismic catalog	Geological information

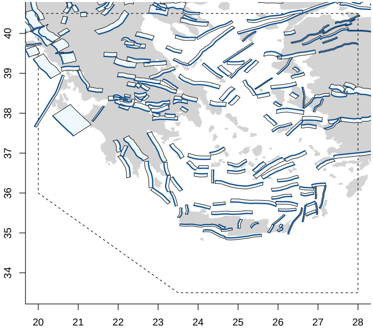
Plate boundaries and volcanoes

- Updated digital model of plate boundaries (https://github.com/fraxen/tectonicplates) (Bird; 2003)
- Global Volcanism Program database (http://volcano.si.edu) (Siebert and Simkin; 2014)

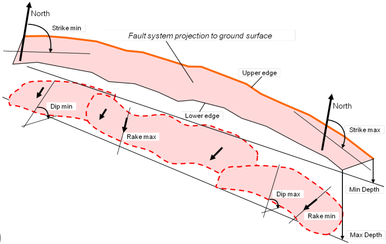


Faults

Greek Database of Seismogenic Sources, GreDaSS (Caputo et al.; 2013)



(a)

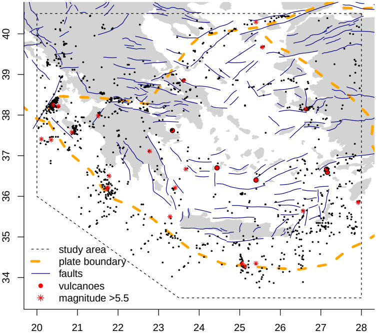


(b)

Figure: (a) Composite seismogenetic sources (b) Main geometric (strike, dip, width, depth) and kinematic (rake) parameters that charcaterised a composite source.

Composite source → complex fault system with several aligned individual seismogenic sources that cannot be separated

Research questions



With a model, we aim:

- to investigate and quantify dependencies on the covariates (faults, plate boundaries and volcanoes)
- to describe the multiscale interaction structure
- to estimate an intensity map

Methodology

Spatial point process

- $W \subset \mathbb{R}^2$ is the study window
- $\mathcal{X}$ is a process, a random countable subset of W
- $\boldsymbol{\upsilon} = \{\mathbf{u}_i\}_{i=1}^{n}$ spatial coordinates of catalogue data where $i = \{1, \ldots, 1105\}$
- $\boldsymbol{\upsilon}$ is a realisation of $\mathcal{X}$

First- and second-order properties

1. **First-order property**

$$\rho(\mathbf{u}) = \lim_{|d_{\mathbf{u}}| \to 0} \frac{E(N(d\mathbf{u}))}{d\mathbf{u}}$$

is the expected number of events in an infinitesimal region $d\mathbf{u}$ that contains the point $\mathbf{u}$

First- and second-order properties

1. **First-order property**

$$\rho(\mathbf{u}) = \lim_{|d_{\mathbf{u}}| \to 0} \frac{E(N(d\mathbf{u}))}{d\mathbf{u}}$$

is the expected number of events in an infinitesimal region $d\mathbf{u}$ that contains the point $\mathbf{u}$

- Also called intensity function
- For a homogeneous process $\rho(\mathbf{u}) = k$
- For an inhomogeneous process, $\rho(\mathbf{u})$ can be estimated as a function of underlying environmental variables $(Z(\mathbf{u}))$

First- and second-order properties

2. **Second-order property**

$$\rho^{(2)}(\mathbf{u}_i, \mathbf{u}_j) = \lim_{|d\mathbf{u}_i| \to 0, |d\mathbf{u}_j| \to 0,} \frac{E(N(d\mathbf{u}_i)N(d\mathbf{u}_j))}{|d\mathbf{u}_i||d\mathbf{u}_j|}$$

- Relationship between numbers of events in pairs of disjoint subregions with centres in $\mathbf{u}_i$ and $\mathbf{u}_j$

First- and second-order properties

2. **Second-order property**

$$\rho^{(2)}(\mathbf{u}_i, \mathbf{u}_j) = \lim_{|d\mathbf{u}_i| \to 0, |d\mathbf{u}_j| \to 0,} \frac{E(N(d\mathbf{u}_i)N(d\mathbf{u}_j))}{|d\mathbf{u}_i||d\mathbf{u}_j|}$$

- Relationship between numbers of events in pairs of disjoint subregions with centres in $\mathbf{u}_i$ and $\mathbf{u}_j$
- Different types of interactions:

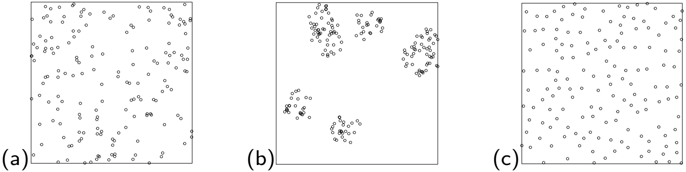


Figure: (a)complete spatial randomness (**CSR**), (b) cluster or (c) inhibitive interaction

First- and second-order properties

2. **Second-order property**

$$\rho^{(2)}(\mathbf{u}_i, \mathbf{u}_j) = \lim_{|d\mathbf{u}_i| \to 0, |d\mathbf{u}_j| \to 0,} \frac{E(N(d\mathbf{u}_i)N(d\mathbf{u}_j))}{|d\mathbf{u}_i||d\mathbf{u}_j|}$$

- Relationship between numbers of events in pairs of disjoint subregions with centres in $\mathbf{u}_i$ and $\mathbf{u}_j$
- Different types of interactions:

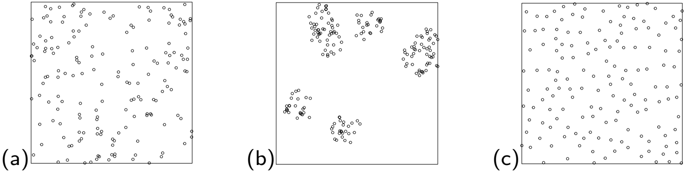


Figure: (a)complete spatial randomness (**CSR**), (b) cluster or (c) inhibitive interaction

- **Multiscale interaction**: the second-order property changes with distance

Hybrid of Gibbs models

- $\mathcal{X}$ is defined by a probability density $f(\boldsymbol{\upsilon})$ where $\boldsymbol{\upsilon} = \{\mathbf{u}_i\}_{i=1}^{n}$
- There are several types of Gibbs processes:
  Poisson process, Strauss process, pairwise interaction processes (Baddeley et al.; 2015)

Hybrid of Gibbs models

- $\mathcal{X}$ is defined by a probability density $f(\boldsymbol{\upsilon})$ where $\boldsymbol{\upsilon} = \{\mathbf{u}_i\}_{i=1}^n$
- There are several types of Gibbs processes:
  Poisson process, Strauss process, pairwise interaction processes (Baddeley et al.; 2015)
- Given m unnormalized densities $f_1(), f_2(), \ldots, f_m()$, the hybrid density is (Baddeley et al.; 2013):

$$f(\boldsymbol{\upsilon}) = f_1(\boldsymbol{\upsilon}) \ldots f_m(\boldsymbol{\upsilon})$$

Hybrid of Gibbs models

- $\mathcal{X}$ is defined by a probability density $f(\boldsymbol{\upsilon})$ where $\boldsymbol{\upsilon} = \{\mathbf{u}_i\}_{i=1}^{n}$
- There are several types of Gibbs processes:
  Poisson process, Strauss process, pairwise interaction processes (Baddeley et al.; 2015)
- Given m unnormalized densities $f_1(), f_2(), \ldots, f_m()$, the hybrid density is (Baddeley et al.; 2013):

$$f(\boldsymbol{\upsilon}) = f_1(\boldsymbol{\upsilon}) \ldots f_m(\boldsymbol{\upsilon})$$

- For example, the density of m Geyer components (with interaction ranges $r_1, \ldots r_m$ and saturation parameters $s_1, \ldots, s_m$) is

$$f(\boldsymbol{\upsilon}) = \prod_{j=1}^{m} \prod_{i=1}^{n} \underbrace{\beta(\mathbf{u}_i)}_{\text{Trend part}} \underbrace{\gamma_j^{min(s_j, t(\mathbf{u}_i, \boldsymbol{\upsilon} \backslash \mathbf{u}_i; r_j))}}_{\text{Several interaction parameters } \gamma_j}$$

Different interaction structure for different interaction distances r_i.

Analysis

Steps

R packages: **spatstat (Baddeley et al.; 2005), maptools, rworldmap**

Steps

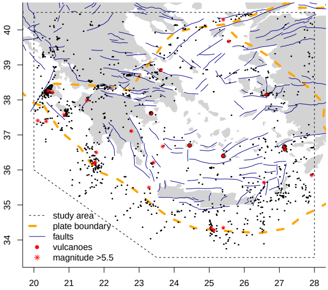
R packages: **spatstat (Baddeley et al.; 2005), maptools, rworldmap**

1. Load catalogue data and shape-files in WGS84
   (`readShapeSpatial`, `spTransform`)

Steps

R packages: **spatstat (Baddeley et al.; 2005), maptools, rworldmap**

1. Load catalogue data and shape-files in WGS84 (`readShapeSpatial, spTransform`)
2. Define the study window and the spatial point pattern (`owin, ppp`)



Steps

3. Compute raster data:
   distance to the plate boundary (D_{pb}), to the nearest volcano (D_v) and to the nearest fault (D_f)
   (`distmap`)

Steps

3. Compute raster data:
   distance to the plate boundary (D_{pb}), to the nearest volcano (D_v) and to the nearest fault (D_f)
   (`distmap`)

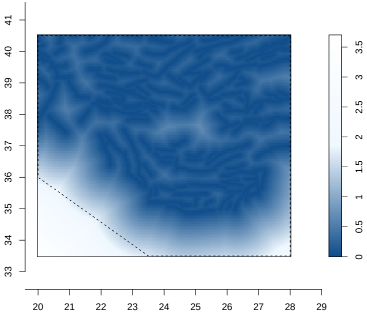


Figure: $D_f(\mathbf{u})$ distance to the nearest fault

Steps

3. Compute raster data:
   distance to the plate boundary (D_{pb}), to the nearest volcano (D_v) and to the nearest fault (D_f)
   (`distmap`)

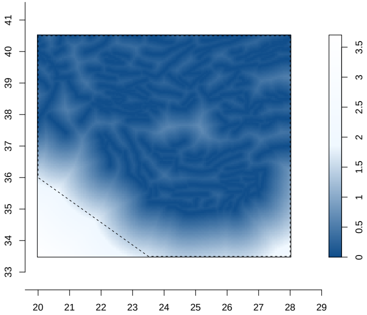


Figure: $D_f(\mathbf{u})$ distance to the nearest fault

4. Descriptive analysis (`Kest, rhohat, berman.test, ...`)

Steps

5. Estimation and model selection of inhomogeneous: (`ppm`)

Poisson models   Hybrid of Gibbs models

Steps

5. Estimation and model selection of inhomogeneous: (ppm)

Poisson models Hybrid of Gibbs models

$$\rho(\mathbf{u}) = exp\{\beta_0 + g(\mathbf{u};\boldsymbol{\beta}) + h(D_v, D_{pb}, D_f; \boldsymbol{\alpha}) + HybridComponent(\gamma_j)\}$$

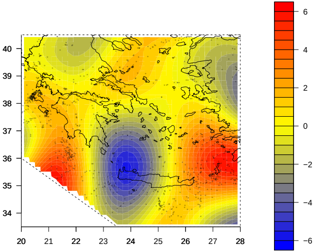
	Description
$g(\mathbf{u};\boldsymbol{\beta})$	Function of the spatial coordinates
$h(D_v, D_{pb}, D_f; \boldsymbol{\alpha})$	Function of the spatial covariates (D_v, D_{pb}, D_f)
HybridComponent	if it is null, **inhomogeneous Poisson model** otherwise **Hybrids of Geyer processes** that depends on γ_1, γ_2, ...
$\{\boldsymbol{\beta}, \boldsymbol{\alpha}, \gamma_1, \gamma_2, ...\}$	Parameters to estimate

Steps

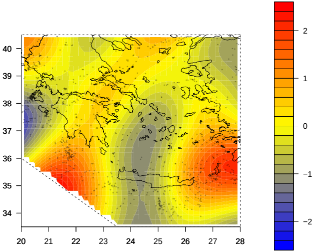
6. Model diagnostic based on: AIC, deviance, analysis of the spatial residuals, residual K and G-functions
   (`AIC`, `anova`, `diagnose.ppm`, `Kres`, `Gres`)

Steps

6. Model diagnostic based on: AIC, deviance, analysis of the spatial residuals, residual K and G-functions
(`AIC`, `anova`, `diagnose.ppm`, `Kres`, `Gres`)



(a) Raw residuals for the selected inhomogenous Poisson model

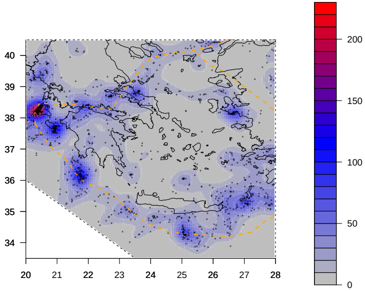


(b) Raw residuals for the selected hybrid model

- For the hybrid model, smaller range and reduction of the spatial trend of the smoothed raw residuals

Steps

7. Estimation of the spatial intensity: (`predict`)



- **Final selected model:**
  - Hybrid model of 4 Geyer processes describes adequately the cluster mutiscale structure
  - Increasing the distance to the nearest fault and to the plate boundary the intensity decreases.
  - The distance to the nearest volcano (D_v) is not significant $\rightarrow$ volcanic Hellenic arc area mostly characterised by microseismic activity.

Remarks and future works

Remarks and future works

- Integrate in point process models earthquake data and external geological information is a **new field of research**

Remarks and future works

- Integrate in point process models earthquake data and external geological information is a **new field of research**
- For Greek data, using Hybrid of Gibbs models, we describe both:
  - spatial inhomogeneity depending of geological covariates
  - multi-scale interaction between points
- Limits of the analysis:
  - spatial analysis, no prediction
  - we could consider other available information: magnitude, meta data related to the geological information,...

Remarks and future works

- **DATA**
  - Nowadays, there are several open-access catalogues of earthquakes
  - For geological information, more difficult to find all the related datasets to describe an area

Remarks and future works

- **DATA**
  - Nowadays, there are several open-access catalogues of earthquakes
  - For geological information, more difficult to find all the related datasets to describe an area

- **METHODOLOGY**
  - We are considering spatio-temporal point process models, such as **log-Gaussian Cox process** (Diggle et al.,2013; Siino et al., 2016) and **ETAS model**, adding geological covariates
  - Using a spatio-temporal models, it is possible to set up a surveillance setting and produce **predictive maps**

Remarks and future works

- **DATA**
  - Nowadays, there are several open-access catalogues of earthquakes
  - For geological information, more difficult to find all the related datasets to describe an area

- **METHODOLOGY**
  - We are considering spatio-temporal point process models, such as **log-Gaussian Cox process** (Diggle et al.,2013; Siino et al., 2016) and **ETAS model**, adding geological covariates
  - Using a spatio-temporal models, it is possible to set up a surveillance setting and produce **predictive maps**

- **ANALYSIS**
  - It would be interesting make the results available to a wider audience
  - Integrate the results into a statistical environmental risk maps for natural disasters
  - For example, Nicolis; 2015 visualises on web and mobile devices the outputs of the ETAS model of Chile seismicity (Environmental Smart Cities)

Thanks for you attention.

References

Baddeley, A.; Rubak, E., Møller, J. (2011) Score, Pseudo-Score and Residual Diagnostics for Spatial Point Process Models. Statistical Science, 26, 613-646.

Baddeley, A.; Turner, R.; Mateu, J., Bevan, A. (2013) Hybrids of Gibbs Point Process Models and Their Implementation. Journal of Statistical Software, 55, 1-43.

Baddeley, A., Rubak, E. and Turner, R. (2015). Spatial Point Patterns: Methodology and Applications with R. London: Chapman and Hall/CRC Press.

Bird, P. (2003). An updated digital model of plate boundaries. Geochemistry, Geophysics, Geosystems, 4(3).

Caputo, R., Chatzipetros, A., Pavlides, S. and Sboras, S. (2013). The greek database of seismogenic sources (gredass): state-of-the-art for northern greece. Annals of Geophysics, 55(5).

Diggle, P. J.; Moraga, P.; Rowlingson, B.; Taylor, B. M.,(2013) Spatial and spatio-temporal Log-Gaussian Cox processes: extending the geostatistical paradigm. Statistical Science, Institute of Mathematical Statistics, 28, 542-563.

Diggle, P. J. (2013). Statistical Analysis of Spatial and Spatio-Temporal Point Patterns. CRC Press.

Illian, J.; Penttinen, A.; Stoyan, H., Stoyan, D. (2008) Statistical Analysis and Modelling of Spatial Point Patterns John Wiley Sons, 70.

Mignan, A. and Woessner, J. (2012). Estimating the magnitude of completeness for earthquake catalogs, community online resource for statistical seismicity analysis, doi: 10.5078/corssa-00180805.

Nicolis, O. (2015). Environmental SmartCities: statistical mapping of environmental risk for natural and anthropic disasters in Chile. In GRASPA15 Conference, Bari (IT), 15-16 June 2015. Università degli studi di Bergamo.

Ogata, Y.; Zhuang, J. (2006). Space–time ETAS models and an improved extension. Tectonophysics, 413(1), 13-23.

Siebert, L. and Simkin, T. (2014). Volcanoes of the world: an illustrated catalog of holocene volcanoes and their eruptions. Smithsonian Institution, Global Volcanism Program Digital Information Series, GVP-3.

Siino, M.; Adelfio, G.; Mateu, J.; Chiodi, M., D'Alessandro, A. (2016) Spatial pattern analysis using hybrid models: an application to the Hellenic seismicity. Stochastic Environmental Research and Risk Assessment

Siino, M.; Mateu, J., Adelfio, G.(2016) Spatio-temporal log-Gaussian Cox processes on earthquake events. In Proceedings of the Eight International Workshop on Spatio-Temporal Modelling (METMA8)