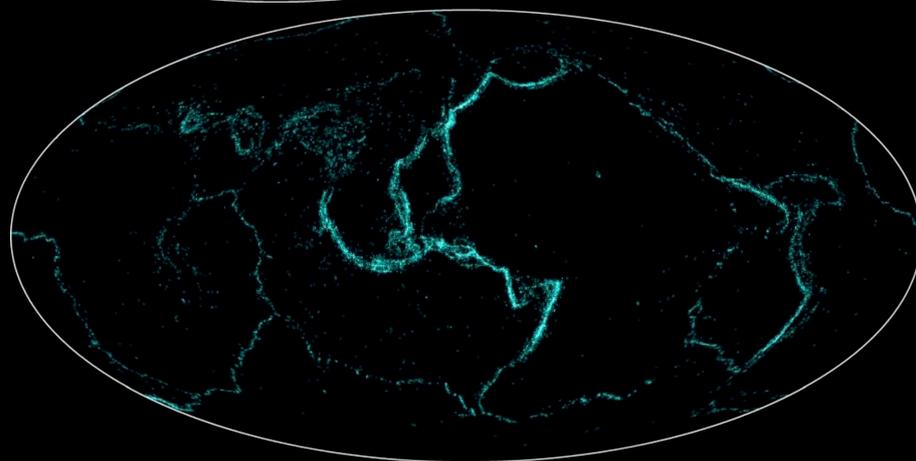
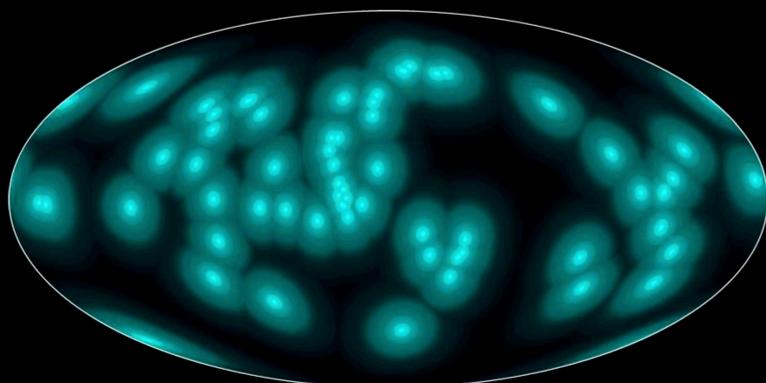


Contributions to Probabilistic Earthquake Forecasting

Álvaro
González
Gómez



Doctoral Thesis



Universidad
Zaragoza

1542



Contributions to Probabilistic Earthquake Forecasting

Álvaro González Gómez

Doctoral Thesis

Advised by:

Javier B. Gómez & Amalio F. Pacheco



Departamento de Ciencias de la Tierra

Universidad de Zaragoza

Spain

2016

This thesis is partially based on the following published articles, used with permission.

Previously unpublished material ©Álvaro González Gómez 2016, all rights reserved.

Álvaro González (2016)

The Spanish National Earthquake Catalogue: Evolution, precision and completeness.

Journal of Seismology, in press.

©Springer 2016.

Álvaro González (2010)

Measurement of areas on a sphere using Fibonacci and latitude–longitude lattices.

Mathematical Geosciences, 42 (1), 49–64.

©International Association for Mathematical Geosciences 2009.

Álvaro González; Miguel Vázquez-Prada; Javier B. Gómez & Amalio F. Pacheco (2006):

A way to synchronize models with seismic faults for earthquake forecasting: Insights from a simple stochastic model.

Tectonophysics, 424 (3–4), 319–334.

©Elsevier B.V. 2006.

Álvaro González, Javier B. Gómez & Amalio F. Pacheco (2006)

Updating seismic hazard at Parkfield.

Journal of Seismology, 10 (2), 131–135.

©Springer 2006.

Álvaro González, Javier B. Gómez & Amalio F. Pacheco (2005)

The occupation of a box as a toy model for the seismic cycle of a fault.

American Journal of Physics, 73 (10), 946–952. [Erratum: (2007) 75 (3), 286.]

©American Association of Physics Teachers 2005, 2007.

Álvaro González; Miguel Vázquez-Prada; Javier B. Gómez & Amalio F. Pacheco (2004):

Using synchronization to improve the forecasting of large relaxations in a cellular automaton model.

Europhysics Letters, 68 (5), 611–617.

©EDP Sciences 2004.

Contents

Acknowledgements	ix
Abstract	1
Resumen	3
1 Overview and goals	5
I Spatial Earthquake Forecasting	9
2 Distances between past earthquakes forecast the distance to the next one	13
2.1 Introduction	13
2.2 Design of the forecast method	16
2.3 Application to earthquake catalogues	22
2.4 Retrospective testing	26
2.4.1 Catalogues used	26
2.4.1.1 Earth	26
2.4.1.2 Southern California	27
2.4.1.3 Iberian Region	28
2.4.2 Results	30
2.5 Independent daily testing in the Collaboratory for the Study of Earthquake Predictability	42
2.5.1 Setup of the testing experiment	43
2.5.2 Adaptation of the model to CSEP standards	44
2.5.3 Model testing	46
2.5.4 Testing regions	47
2.5.4.1 California	48
2.5.4.2 Global (whole Earth)	49
2.5.4.3 Western Pacific	52
2.5.5 Overall results and comparison with a reference model	54
2.5.5.1 Reference model	57
2.6 Final remarks	60

3	The Spanish National Earthquake Catalogue: Evolution, precision and completeness	63
3.1	Introduction	63
3.2	Evolution of the catalogue	66
3.2.1	Historical era (until 1900)	67
3.2.2	First observatories (1897–1924)	69
3.2.3	Printed bulletins (1924–1995)	69
3.2.4	Civil War (1936–1939) and postwar	70
3.2.5	Systematic earthquake location (1955–present)	70
3.2.6	Instrumental era (1962–present)	70
3.2.7	First telemetred network (1978–1992)	71
3.2.8	First automatic locations (1991–1997)	72
3.2.9	The Sonseca Array (1992–present)	74
3.2.10	New automatic earthquake location, and stations from other institutions (1997–present)	76
3.2.11	Broadband network (1999–present)	76
3.2.12	Monitoring around the Itoiz and Yesa reservoirs (2005–2013)	77
3.2.13	Monitoring of seismo-volcanic activity in the Canary Islands (2004 and 2011–present)	77
3.2.14	Current practice on catalogue compilation	78
3.3	Precision of earthquake locations	79
3.3.1	Location procedures	80
3.3.2	Epicentral location	82
3.3.3	Focal depth	84
3.4	Magnitude determination and precision	87
3.4.1	M_D , Iberian Region (1910–1961)	87
3.4.2	M_D , Canary Islands (1964–1997)	88
3.4.3	m_{bLg} (1962–present)	88
3.4.3.1	m_{bLg} (1962–2002)	89
3.4.3.2	m_{bLg} (2002–present)	89
3.4.4	m_b (1997–present)	91
3.4.5	M_w (2002–present)	91
3.4.6	Evolution of magnitude values in the catalogue	92
3.5	Magnitude of completeness	94
3.5.1	Definition and influencing factors	95
3.5.2	Method used to calculate M_c	96
3.5.3	Spatial sampling	97
3.5.4	Choice of temporal periods and magnitude scales	98

3.5.5	M_c in the Iberian Peninsula and surroundings	101
3.5.6	M_c in the Canary Islands	106
3.6	Daily and weekly variations	108
3.7	Catalogue contamination by blasts	110
3.8	Discussion	114
4	Measurement of areas on a sphere using Fibonacci and latitude–longitude lattices	117
4.1	Introduction	117
4.2	Latitude–longitude lattice	119
4.3	Fibonacci lattice	121
4.3.1	Lattice construction	121
4.3.2	Lattice history	124
4.4	Area measurement	127
4.4.1	Weights in the latitude–longitude lattice	127
4.4.2	Weights in the Fibonacci lattice	128
4.5	Error assessment	128
4.6	Results	129
4.6.1	Analytical approach to the root mean squared error	133
4.7	Discussion	133
II	Temporal Earthquake Forecasting	135
5	Probabilistic temporal forecasting of large earthquakes in the Parkfield section, San Andreas Fault.	139
5.1	Introduction	140
5.2	The seismic cycle	141
5.3	Renewal models	145
5.4	The Parkfield series	147
5.5	The box model	149
5.5.1	Rules of the model	150
5.5.2	Analytical results on the model properties	151
5.5.2.1	Review of the geometric distribution	151
5.5.2.2	Mean, variance and aperiodicity of the cycles	151
5.5.2.3	Probability distribution for the duration of the cycle	155
5.5.2.4	Conditional probability	155
5.5.2.5	Pattern of loading	155
5.5.2.6	Stress shadow	157
5.5.3	Fitting the box model to the Parkfield series	157

5.5.3.1	Earthquake probabilities at Parkfield with the box model	158
5.6	Binary earthquake forecasting	160
5.6.1	The random guessing strategy as baseline	162
5.6.2	Reference strategy	162
5.7	Multi-model approach	165
5.7.1	Fits with other renewal models	165
5.7.2	Comparison of probability estimates at Parkfield	167
5.7.3	Binary forecasts	171
5.7.4	Later developments	173
6	Insights on how to synchronize models with seismic faults for earthquake forecasting	175
6.1	Introduction: Data assimilation in models of seismic faults	175
6.2	The minimalist model	180
6.2.1	Rules	180
6.2.2	Main properties	183
6.3	General scheme of binary temporal forecasting	187
6.3.1	Random guessing strategy	188
6.3.2	Reference strategy	188
6.3.3	Ideal strategy	189
6.4	Synchronization-based forecasting	190
6.5	Discussion and conclusions	198
7	Conclusions	201
8	Conclusiones	207
A	Mathematical Appendix	213
A.1	Properties of the box model	213
A.1.1	Discrete probability distribution for the duration of the seismic cycle, $P_N(n)$	213
A.1.2	Asymptotic mean of $P_N(n)$	217
A.1.3	Asymptotic standard deviation of $P_N(n)$	218
A.1.4	Asymptotic conditional probability	219
A.2	Deduction of the ideal forecasting strategy in the minimalist model	220
	Bibliography	222

To all those who love and share this small world.

Acknowledgements

Thanks are due to my advisors, Amalio and Javier, for guiding me in my initial steps in Science, and for letting me pursue my own goals. Learning from their different disciplines (theoretical Physics and Earth Sciences) helped me develop simple approaches to complex problems. Miguel Vázquez-Prada was of great professional and personal support at the beginning of our teamwork.

This research especially benefited from three stays abroad, the first one at the Center for Computational Science and Engineering (University of California at Davis), where I fine-tuned my skills on numerical modelling. I am grateful to Donald L. Turcotte and John B. Rundle for their hospitality and for letting me learn from them and their team, especially from Robert Shcherbakov, Gleb Yakovlev and Kazuyoshi Nanjo. Bruce D. Malamud contributed to make this stay possible and always offered valuable advice.

My second research stay was at the Institute of Geological and Nuclear Sciences (now GNS Science) at Lower Hutt, New Zealand. I am indebted to Pilar Villamor, Mark W. Stirling, Kelvin R. Berryman and Robert Langridge for their hospitality, friendliness, and for helping me decipher the faults' handiwork on the landscape of Aotearoa.

My last and longest stay was at the GFZ German Research Centre for Geosciences, in Potsdam, Germany. I am fortunate to have too many friendly colleagues there to be able to list them. Danijel Schorlemmer was an excellent host, and offered great insight on testing forecast models and calculating the completeness of earthquake catalogues. Further help, support and friendship was offered by many colleagues, and especially Thomas Beutin, Robert A. Clements, Sebastian Heimann, Katja Müller and Elske de Zeew-van Dalssen. Frank Roth and Simone Cesca merit special thanks for supporting this thesis for a mention of European Doctorate.

The earthquake data used were thankfully provided by diverse institutions. In particular, the work on the Spanish earthquake catalogue would not have been possible without the efforts by the colleagues and technicians of the Instituto

Geográfico Nacional. I am especially grateful to Resurrección Antón for technical support, and to Carmen del Fresno for valuable comments.

Throughout the years, several parts of my work have been directly funded by research grants of the Spanish National Government, the Regional Government of Aragón, the Mapfre Foundation and the Caja Madrid Foundation. Attendance to scientific meetings was also supported by UNESCO and the European Science Foundation. I thank them, and especially their taxpayers or clients, for this necessary funding.

The daily, automatic testing of my earthquake forecast models was thankfully made by the Collaboratory for the Study of Earthquake Predictability, at the Southern California Earthquake Centre. This initiative is funded by the United States Geological Survey, the U.S.A. National Science Foundation, the W.M. Keck Foundation and the U.S. Department of Homeland Security. Maria (Masha) Liukis always provided timely and kind technical support in Los Angeles.

I would also like to acknowledge Edward B. Saff and the Department of Mathematics of Vanderbilt University for their kind invitation to Nashville (Tennessee) to present the results about the spherical Fibonacci lattice to a mathematical audience. And Thorsten W. Becker for bringing me to Los Angeles to discuss the forecast models at the University of Southern California.

Part of the numerical simulations were calculated using the supercomputing facilities of the Institute of Biocomputing and Physics of Complex Systems of the University of Zaragoza. Thanks are due to the open-source software community for developing diverse tools used throughout this work, and particularly to Pål Wessel for kindly improving Generic Mapping Tools to suit my needs.

The parts already published of this dissertation benefited from the comments contributed by anonymous reviewers, and those of Harvey Gould, Vladimir G. Kossobokov, William I. Newman and Kristy F. Tiampo.

In the years spent in my *alma mater*, the University of Zaragoza, I have been lucky to share my time, joy and suffering with supportive friends: Patricia Acero, Luis F. Auqué, Laura Becerril, Jorge (JP) Galve, María José Gimeno, and Vicente Sánchez Cela, among many others. Antonio M. Casas offered timely institutional support when needed.

I would like to dedicate this work to the unforgettable memory of my mother, María Luisa, to my father Fabián, who has always been an example to follow, to my ever-supportive sister Ana, my mother in law Pepi and in general to all my family and friends. Especially, to María José and to our fantastic children: Muriel, Arturo and Gemma. They have always trusted me, and have upheld the virtue of loving and sharing this small world.

*Truly a tale I have that I would tell thee,
a word that I would recount to thee:
a tale of trees and a whisper of stones,
the sighing of the Heavens to the Earth,
of the Deeps to the Stars.
I understand lightning not known to the Heavens,
a tale not known to humankind,
nor yet understood by the multitudes of the Earth.**

*Did I not tell you once of the whisper of stones?
Did I not tell you once of the murmur which I could not comprehend?
I missed the message in a sudden rush of wind
Or the passing of a car with a rattle on the road.†*

* Ba'al, West Semitic god of thunder, talking in *The Myth of Ba'al*, of unknown author, written 1400–1350 years before current era. Clay tablet KTU 1.1, column III, lines 12–15, found at Ugarit, Bronze Age Canaanite city (currently Ras Shamra, Syria).

† Patrick Wallace (2000): *Dance Beneath the Modern Sun* (fragment).

Abstract

This thesis presents a set of contributions to probabilistic earthquake forecasting, both spatial and temporal.

A novel empirical method with no parameter is used to show that the distances between past earthquakes forecast, in a probabilistic way, the distance to the next earthquake. This procedure is first applied retrospectively to earthquake catalogues of the whole Earth, Southern California and the Iberian region. The resulting probabilistic maps for forecasting the locations of forthcoming earthquakes self-sharpen when updated as new earthquakes occur.

Within the Collaboratory for the Study of Earthquake Predictability, the Southern California Earthquake Center has tested the method prospectively, in an independent and automatic way, calculating daily forecast maps. The results of six years of tests for the whole Earth, California and western Pacific show that the method outperforms the one typically used as a reference.

The earthquake catalogue of the Spanish Instituto Geográfico Nacional, used here for forecast testing in the Iberian region, is analysed in detail with regard to its development, location precision, magnitude of completeness and inclusion of artificial blasts. Spatial and temporal heterogeneities, and overall progressive improvements, are highlighted.

An efficient technique is devised for measuring the areas on the map where each new earthquake is expected. The spherical Fibonacci lattice, a highly homogeneous spiral pattern, is proposed for measuring complex areas on a spherical surface by point counting. Measurement errors are greatly reduced compared to those obtained with latitude-longitude lattices.

A new stochastic model is introduced as an idealization of the seismic cycle of a fault, and applied to the series of earthquakes generated by the San Andreas Fault at Parkfield (California). Along with other, statistical renewal models, it is used to calculate time-depending probabilities for the occurrence of the next large earthquake, and to test a simple forecasting strategy.

Finally, a method is devised to try to synchronize numerical models with the seismic faults that they simulate, by forcing them to reproduce the observed

sequence of earthquakes. This procedure is able to partially synchronize models with another, stochastic one, and to forecast its largest synthetic earthquakes more efficiently than with simpler approaches.

Resumen

Esta tesis presenta un conjunto de contribuciones al pronóstico probabilista de terremotos, tanto espacial como temporal.

Un método novedoso, empírico y sin ningún parámetro, se usa para mostrar que las distancias entre terremotos previos pronostican, de manera probabilista, la distancia al terremoto siguiente. Este procedimiento se aplica primero retrospectivamente a catálogos de terremotos de toda la Tierra, el sur de California y la región Ibérica. Los mapas probabilistas resultantes, para pronosticar la localización de los siguientes terremotos, se refinan por sí solos conforme se actualizan al ocurrir terremotos nuevos.

El *Southern California Earthquake Centre*, dentro del *Collaboratory for the Study of Earthquake Predictability*, ha puesto a prueba el método de manera prospectiva, independiente y automática, calculando mapas de pronóstico diarios. Los resultados de seis años de pruebas para toda la Tierra, California y el Pacífico occidental muestran que el método supera al usado típicamente como referencia.

El catálogo de terremotos del Instituto Geográfico Nacional, usado aquí para pruebas de pronóstico en la región Ibérica, se analiza en detalle, en lo referido a su desarrollo, precisión de las localizaciones, magnitud de completitud e inclusión de explosiones artificiales. Se destacan heterogeneidades espaciales y temporales, y mejoras globales progresivas.

Se diseña una técnica eficiente para medir las áreas del mapa donde se espera cada nuevo terremoto. El retículo esférico de Fibonacci, un patrón espiral altamente homogéneo, se propone para medir áreas complejas en una superficie esférica mediante conteo de puntos. Los errores de medida se reducen ostensiblemente, comparados con los obtenidos con retículos latitud-longitud.

Un nuevo modelo estocástico se presenta como una idealización del ciclo sísmico en una falla, y se aplica a la serie de terremotos generados por la Falla de San Andrés en Parkfield (California). Junto con otros modelos estadísticos de renovación, se emplea para calcular las probabilidades de la ocurrencia del siguiente gran terremoto (dependientes del tiempo) y para probar una simple estrategia de pronóstico.

Finalmente, se diseña un método para intentar sincronizar modelos numéricos con las fallas sísmicas que éstos simulan, forzándolos a reproducir la secuencia observada de terremotos. Este procedimiento es capaz de sincronizar parcialmente unos modelos con otro, estocástico, y de pronosticar los mayores terremotos sintéticos en éste de manera más eficiente que con métodos más sencillos.

Overview and goals

This thesis presents several contributions on earthquake forecasting, and on auxiliary techniques and seismic catalogues used for it. They cover a wide spectrum of topics, and are sufficiently different from each other as to be organized in separate chapters, each one with its own introduction. This initial chapter will present a general overview and describe the pursued goals.

Part of the research presented here has already been published in six papers in international scientific journals. Instead of just copying them verbatim in this dissertation, they have been merged and expanded, with added background, insight and examples, shaping chapters 3 to 6. Sometimes these additions could not be included, because of their length, in the original publications. In other cases, they expand the research to accommodate the findings of literature published after ours. I hope that future readers will benefit from these additions.

Perhaps it is wise to start by recognizing that we cannot currently predict earthquakes, and that most likely we will not be able to do so in an exact way. And then to discuss what can be possibly achieved and whether it is useful.

A major limitation is that earthquakes have an unobservable dynamics (e.g. Rundle *et al.*, 2003), since they are generated in the interior of the Earth which (with the few punctual exceptions of deep drilling projects) cannot be observed directly. So variables of paramount importance, such as rock type, temperature, stress and elastic strain cannot be typically measured at depth. Proposed observable precursors (e.g. Uyeda *et al.*, 2011), which might inform on the occurrence of an impending earthquake, are rarely tested systematically (e.g. Blackett *et al.*, 2011) and so far not reliable for issuing formal warnings (Jordan *et al.*, 2011).

Moreover, earthquake generation involves a great range of temporal and spatial scales. Tectonic motion proceeds during millennia, at rates of a fraction of a millimetre to (at most) a few centimetres per year, while dynamic earthquake

ruptures propagate on a fault plane in seconds, at up to kilometres per second (e.g. Kanamori & Brodsky, 2004).

An additional difficulty is that the earthquake process is complex and chaotic (e.g. Turcotte, 1997). *Complex systems* are those which have many interacting parts, with properties which emerge from such interactions. Earthquakes are generated in a complex network of interacting faults. Each earthquake can trigger others by static and dynamic stress transfer, possibly even at great distances, and just by slight increases of stress (e.g. Freed, 2005; Pollitz *et al.*, 2012; Parsons *et al.*, 2014). This is the hallmark of *chaos*, a property of systems that, despite being deterministic (non-random), are very sensitive, so even a slight perturbation may modify significantly their evolution.

But even chaotic systems are predictable to some extent. The atmosphere is a paradigmatic case of a chaotic system (Lorenz, 1963), yet it admits some degree of forecasting. These forecasts are probabilistic, given that the initial conditions of the system and its inner workings are not perfectly known. Forecasting the evolution of a system, at least in a probabilistic way, is the ultimate test of how well we understand it.

The most common, and so far the most useful, earthquake forecasts are probabilistic seismic hazard assessments. They are typically based on the assumption of a Poisson process, in which earthquakes occur at a fixed frequency, randomly and independent to each other, so that their probabilities do not change with time (e.g. Cornell, 1968; Reiter, 1991).

But earthquakes have additional, non-random features which make them predictable to some extent. They do not occur everywhere with equal probability, yet cluster in space (e.g. Kagan, 2014; Kamer *et al.*, 2015), and frequently are generated by known faults (e.g. Yeats *et al.*, 1997; McCalpin, 2009). Their spatial occurrence is clearly influenced by stress changes which can be calculated to some extent (e.g. Freed, 2005). They are not equally likely along time either. For example, in aftershock sequences their probability of occurrence is orders of magnitude larger than the long-term one (e.g. van Stiphout *et al.*, 2010). And the probability of a large earthquake is expected to change slowly with time as elastic strain accumulates on the blocks of a fault (e.g. Jordan *et al.*, 2003). We are just starting to measure this process of strain accumulation on the Earth's surface with enough detail by high-precision geodesy (e.g. Tong *et al.*, 2014), a century later than proposed by Reid (1910) when first presenting the mechanism of earthquake generation.

Long-term prevention against earthquakes is clearly the most valuable tool to mitigate their impacts on society, and cannot be replaced. But understanding how earthquake probabilities change in space and time, in short and long time scales,

is already contributing complementary information valuable for decision-making (e.g. Jordan *et al.*, 2011, 2014). Rigorous testing of earthquake forecasts is necessary to provide valuable insight on the physics of the process (e.g. Schorlemmer & Gerstenberger, 2007) and rejecting those that do not pass the tests (e.g. Hardebeck *et al.*, 2008).

Forecasts of an event can be expressed in two different ways. One is to specify an explicit probability that the event will occur, such as “tomorrow it will rain here with 90% probability”. The other is to provide a single statement (a binary, *yes* or *no* forecast), such as “tomorrow it will rain here”. In this latter case a probability can be assigned in the long term, after checking the performance of such statements in a long series of events (e.g. Mason, 2003).

In this thesis, both, intertwined kinds of forecasts are proposed and tested as much as possible with the available data. Preferably with extensive catalogues of real earthquakes, and when these do not contain enough data for testing, with long synthetic catalogues generated by numerical models.

The research of this thesis is permeated by a quest for simplicity, and is divided into two parts. The first one (chapters 2, 3 & 4) focuses on spatial earthquake forecasting, related techniques and datasets, and the second one (chapters 5 & 6) on temporal earthquake forecasting.

Chapter 2 presents a novel, non-parametric method for spatial earthquake forecasting, with the goal to provide a reference model to which one can compare the merits of more complex spatial forecasts. This method is tested in a comprehensive way, both retrospectively and prospectively, with up to tens of thousands of earthquakes, for the whole Earth and in more detail for specific regions.

Chapter 3 analyses in detail the national Spanish earthquake catalogue, compiled by the Instituto Geográfico Nacional, which is one of the data bases used for the tests above. Earthquake catalogues, compiled over decades, are being analysed in progressively more sophisticated statistical ways, which demand complete and reliable data sets (e.g. Vere-Jones, 1992). This analysis of the Spanish catalogue aimed to assess such completeness and reliability in depth.

Chapter 4 presents a simple auxiliary technique, designed for measuring efficiently complex areas on a sphere, based on a homogeneous lattice of sampling points. Here it is used to measure the areas marked in the forecast maps presented in Chapter 2. After its publication (González, 2010a), it has fostered further research in fields as diverse as biochemistry, computer science, crystallography, geodesy, mathematics and industrial design.

Chapter 5 deals with the calculation of long-term earthquake probabilities in the Parkfield section of the San Andreas Fault, which has generated a well-known,

relatively regular series of earthquakes (Bakun *et al.*, 2005). The focus, as in the next chapter, is on theoretical statistical physics. It presents a simple stochastic model (González *et al.*, 2005) and applies it and other, statistical models (González *et al.*, 2006a) to calculate how the probabilities for the next large earthquake at Parkfield may change with the time elapsed since the last one.

Chapter 6 enters further in the realm of theoretical physics. Currently there are increasingly realistic numerical models of active faults (e.g. Field *et al.*, 2015a; Yoder *et al.*, 2015), but key data (stress, strain, etc.) to tune the models are absent because of the unobservable dynamics of the process. This chapter explores theoretically whether the models could be synchronized with the faults by forcing them to reproduce the series of past earthquake ruptures. This novel approach is shown to be able to synchronize simple, yet stochastic, models with another one, which allows forecasting more efficiently the synthetic earthquakes that the latter generates (González *et al.*, 2004, 2006b).

Chapters 7 and 8 present the conclusions in English and Spanish. Finally, analytical derivations regarding models used in chapters 5 and 6 are detailed in a separate mathematical appendix, followed by the bibliography.

Part I

Spatial Earthquake Forecasting

*To those with the patience to master its language, the Earth still responds.
To those with the insight to comprehend its moods and reflect on its mysteries,
the Earth still teaches.*

Frank H.T. Rhodes & Bruce D. Malamud (2008, p. xi)

Distances between past earthquakes forecast the distance to the next one

*Come forth into the light of things,
Let Nature be your teacher.*

William Wordsworth

The Tables Turned

(fragment, Wordsworth & Coleridge, 1798, p. 187).

2.1 Introduction

Earthquakes usually originate in some geographic regions, but not in others, leaving large areas of the Earth's surface devoid of seismicity. Although this is an ancient observation and one of the fundamentals of geophysics (e.g. Udías & Mezcua, 1997), it is not yet clear how well the location of past earthquakes is able to forecast the location of future ones.

Commonly (e.g. Cornell, 1968; Reiter, 1991), epicentres have been assumed to be uniformly distributed within seismically active regions (*seismogenic zones*). But the actual distribution of earthquakes in space is more complicated, as first shown by Kagan & Knopoff (1980). Epicentres are grouped (*clustered*) at all scales of observation, either in a self-similar, *fractal* way which does not depend on the scale of observation (e.g. Kagan, 2007, 2014), or in a *multifractal* way (e.g. Kamer *et al.*, 2015, and references therein), in which the degree of clustering varies in space.

This non-uniform distribution of earthquakes reflects the underlying complex geometry of the faults which generate them (e.g. Bonnet *et al.*, 2001; Kamer *et*

al., 2015, and references therein). Persistent spatial clustering of hypocentres is expected to result from the persistent stress concentrations due to the irregular geometry of the fault system (Parsons, 2008b). Exceptions aside (such as when fissures develop rapidly due to magma injection, e.g. Sigmundsson *et al.*, 2015), tectonic faults develop, change or become inactive, in long, geological timescales (e.g. Yeats *et al.*, 1997; Nicol *et al.*, 2005; McCalpin, 2009). So generally, earthquakes are expected to cluster in space persistently at short, human, timescales (Parsons, 2008b).

In principle, the spatial distribution of past earthquakes could be extrapolated to forecast where future ones will take place, but this approach has several limitations. This observed distribution appears to converge slowly towards the long-term one, in which case decades or even centuries of earthquake recording would be required to have a representative sample of earthquakes in space (Kagan, 2007, 2014). The high-quality earthquake catalogues, compiled from instrumental observations, are available only for the last decades, but large earthquakes occur infrequently, so they are yet poorly sampled. Because of this scarcity of instrumental records, in some regions there are apparent *seismic gaps* devoid of seismicity where eventually a large earthquake takes place (e.g. Scholz, 2002; Melnick *et al.*, 2012; Cassidy, 2013). And in relatively stable continental regions, it is discussed if the seismicity distributes in a persistent way (Kafka, 2007) or if it alternates between different locations along time (e.g. Liu *et al.*, 2011). So is not always safe to assume that future earthquakes will occur only precisely where the others, recently recorded, did.

More recently, smoothing kernels have been used for estimating a non-uniform, continuous spatial probability for earthquakes, calculated from the locations of earlier ones (e.g. Vere-Jones, 1992; Stock & Smith, 2002; Helmstetter & Werner, 2014; Crespo *et al.*, 2014, and references therein). These kernels are based on a variety of theoretical functional forms (such as Gaussian, power-law or inverse-biquadratic), and use at least one parameter, the *bandwidth*, which controls the degree of smoothing of the resulting map. Because of such a smoothing, some probability is assigned to sites devoid of past epicentres, to acknowledge that earthquakes can originate at locations where none did during the available record.

A good probability model for the long-term spatial distribution of earthquakes should have two qualities. First, it should be consistent with the original distribution of earthquakes used to calculate it (so those past earthquake locations would have been likely according to the model). Second, it should be able to forecast the locations of future ones (so when tested with newer earthquakes, the locations of these should also be likely according to the model). Competing kernel models have been subject to limited testing (e.g. Stock & Smith, 2002), and there are different

ways to measure the forecast performance, so the optimal model depends on the metrics used (e.g. Stock & Smith, 2002; Zechar & Jordan, 2010b). As a result, there is yet no agreement on which spatial model or smoothing kernel of seismicity works best, so different models, with different degrees of smoothing, have to be tried to check the influence on the results (e.g. Becker *et al.*, 2015), and different options are being explored in an attempt to improve the forecasts (e.g. Nanjo, 2010, 2011; Zechar & Jordan, 2010b; Helmstetter & Werner, 2014).

In order to be statistically significant, a complex method for forecasting future earthquake locations should outperform the simple idea that they will occur close to previous ones. But this null hypothesis, or *reference model*, can be formulated in different ways (e.g. Rhoades & Evison, 2004; Holliday *et al.*, 2005; Zechar & Jordan, 2008; Shcherbakov *et al.*, 2010; Kafka & Ebel, 2011). The simplest reference models have no parameter and are based on counting the frequency of past earthquakes in each region (Holliday *et al.*, 2005; Zechar & Jordan, 2008), or on considering the areas around past earthquakes, up to a distance which is explored by trial and error as to cover the location of a desired fraction of new earthquakes (Kafka & Ebel, 2011, and references therein).

This chapter presents a new method for spatial earthquake forecasting, which aims to be used as reference. It has no parameter, and is based exclusively on the empirical distribution of nearest-neighbour distances between earthquakes. Here it is shown that this distribution suffices to forecast, in a probabilistic way, at which distance the next earthquake will occur from previous ones. Section 2.2 presents the method and exemplifies it with a theoretical distribution of points in space. The application to real earthquake databases (*earthquake catalogues*) is detailed in Sec. 2.3, and tested retrospectively in an extensive way in Sec. 2.4, at global and regional scales, using the best available catalogues for each of them (whole Earth, Southern California, and Iberian region).

For testing the method prospectively, as a forecast in strict sense, it was submitted to the Collaboratory for the Study of Earthquake Predictability (Schorlemmer & Gerstenberger, 2007; Zechar *et al.*, 2010) for automatic, independent, daily testing. Section 2.5 details how the method was adapted for such tests, and the results obtained for the whole Earth, California, and western Pacific during six years. The chapter is closed by a discussion of the results, compared with those of the most frequently used reference model.

2.2 Design of the forecast method

Let us start by considering that a number n of points are located on a spatial region (Fig. 2.1). Each of them i has a *nearest neighbour*, separated by a distance d_i . With this definition there will be n nearest-neighbour distances. The number of different values for these distances may range from a minimum of 1 (if the points are, for example, distributed in a regular lattice, such as all distances between them are identical to each other) to a maximum of $n - 1$ (as there will be at least one pair of points which are the closest to each other).

With the nearest neighbour distances arranged in increasing length such as $d_1 \leq d_2 \leq \dots \leq d_n$ we can calculate the *empirical distribution function*, which here will be denoted as P , and which equals the fraction of points whose nearest-neighbour distances are $\leq d$:

$$P(d) = \frac{1}{n} \sum_{i=1}^n \mathbb{1}_{d_i \leq d} \quad (2.1)$$

where $\mathbb{1}_{d_i \leq d}$ is the *indicator function*:

$$\mathbb{1}_{d_i \leq d} = \begin{cases} 1 & \text{if } d_i \leq d \\ 0 & \text{if } d_i > d. \end{cases} \quad (2.2)$$

The empirical distribution function is a step function that jumps up by $1/n$ at each of the n data points. Conversely, we can also consider the complementary distribution, which expresses the frequency of exceedance (Fig. 2.1), that is, the fraction of distances $> d$:

$$\bar{P}(d) = 1 - P(d) = \frac{1}{n} \sum_{i=1}^n \mathbb{1}_{d_i > d}, \quad (2.3)$$

which uses the complementary indicator function $\mathbb{1}_{d_i > d} = 1 - \mathbb{1}_{d_i \leq d}$.

The basis of the proposed spatial forecasting method follows, and is illustrated in Fig. 2.1. We will assume that the next point added to the region, the n -th+1, will distribute in space similarly as the previous points did. This will imply that, once placed in the region, the distance d_{n+1} to its nearest neighbour would follow the same distribution $P(d)$ (and $\bar{P}(d)$) observed for the previous n points. That is, if the distance d_p is the percentile which corresponds to a given P , the next point added is expected to be located at $d_{n+1} \leq d_p$ from previous ones with probability P . Conversely, it is expected further away, $d_{n+1} > d_p$ with probability \bar{P} . Distance percentiles will be defined here using the so called *nearest rank method*, without interpolation, so d_p is the smallest distance such as P percent of the nearest-neighbour distances are $\leq d_p$.

The following issues are easier to visualize by considering a theoretical distribution of points, such as the classical Sierpiński's fractal gasket (Sierpiński,

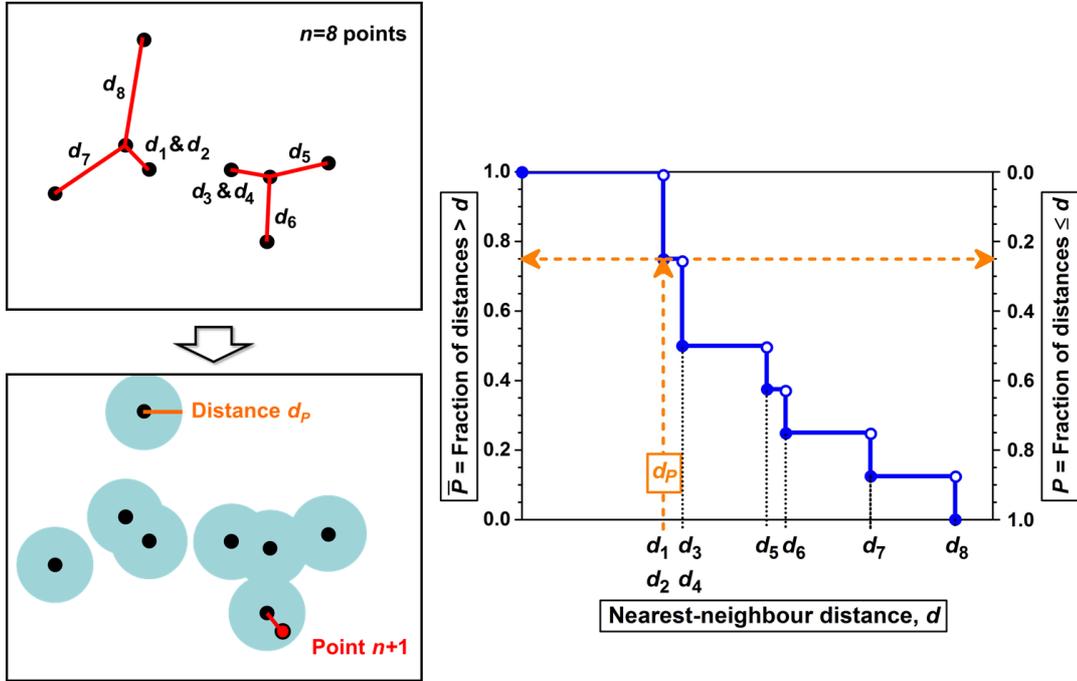


Figure 2.1: **Computation of empirical probabilities based on nearest-neighbour distances.** In the example, $n = 8$ points are already located in a region, and we wish to forecast the location of the next point $n + 1$. *Top left:* From the locations of the points we measure the distance from each point to its nearest neighbour, up to n distances denoted in increasing length, $d_1 \leq d_2 \leq \dots \leq d_n$. Some distances are repeated ($d_2 = d_3$, $d_4 = d_5$) because several points are both the nearest neighbours to each other. *Right:* There is a percentage P of nearest-neighbour distances smaller than or equal to the percentile d_p of the empirical distribution, and a percentage $\bar{P} = 1 - P$ of larger ones. *Bottom left:* The method assumes that the next point will be located with probability P within a distance $d_{n+1} \leq d_p$ of its nearest-neighbour, and further away, $d_{n+1} > d_p$ with probability \bar{P} . A larger P implies a larger d_p , which covers a larger fraction of the area. In the example, the next point is expected in the shaded area with probability 25% ($P = 0.25$).

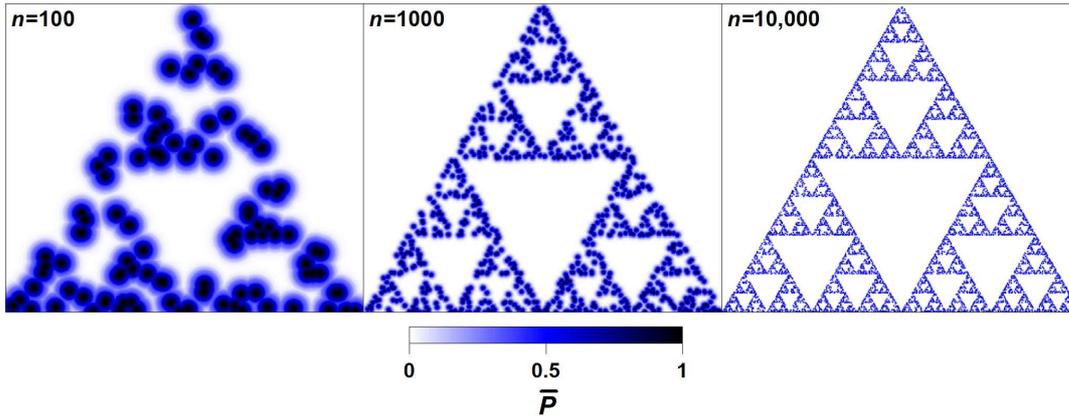


Figure 2.2: **Self-sharpening probability maps for forecasting random points on a Sierpiński gasket.** A number of points (n) have already been placed (black spots, where $\bar{P} = 1$) and are used for calculating the map, aimed at forecasting the location of the next point ($n + 1$). For each site of the map, \bar{P} is the empirical probability that the next point will be placed further away from the existing points than this site is. For example, the next point is expected with 10% probability where $\bar{P} < 0.1$, and with 90% probability where $\bar{P} \geq 0.1$.

1915, Fig. 2.2). In this example, points are placed at random on the gasket, independently to each other, using an iteration algorithm called “chaos game” (e.g. Barnsley, 2012). The maps of Fig. 2.2 show a value of \bar{P} for each site. This is obtained by measuring the distance from the site to the nearest point, and checking which \bar{P} corresponds to that distance in the empirical distribution of exceedance (Fig. 2.1).

A binary forecast map such as that of Fig. 2.1 can be built by considering any fixed probability threshold P^* . The next point $n + 1$ to be added to the map is forecast to occur with expected probability P^* in the areas of the map where $\bar{P} \geq 1 - P^*$. The higher the P^* chosen, the larger these overlapping areas around previous points will be.

The map self-sharpens as $n \rightarrow \infty$ for two reasons, related with how the distribution of \bar{P} changes (Fig. 2.3). First, this distribution becomes smoother, tending to a continuous one, since its steps become smaller. Second, since the density of points in the region increases, typically the nearest-neighbour distances will decrease. This second effect implies that, for a given value of P , the corresponding $d_p \rightarrow 0$ as $n \rightarrow \infty$.

The decay of d_p as $n \rightarrow \infty$ is related to the geometry of the underlying geometry of the figure (the *attractor*, in mathematical jargon) where the points distribute. The mean distance \bar{d} between nearest neighbours is expected to decrease when $n \rightarrow \infty$ as (Badii & Politi, 1984a, 1985):

$$\bar{d} = Kn^{-1/D}, \quad (2.4)$$

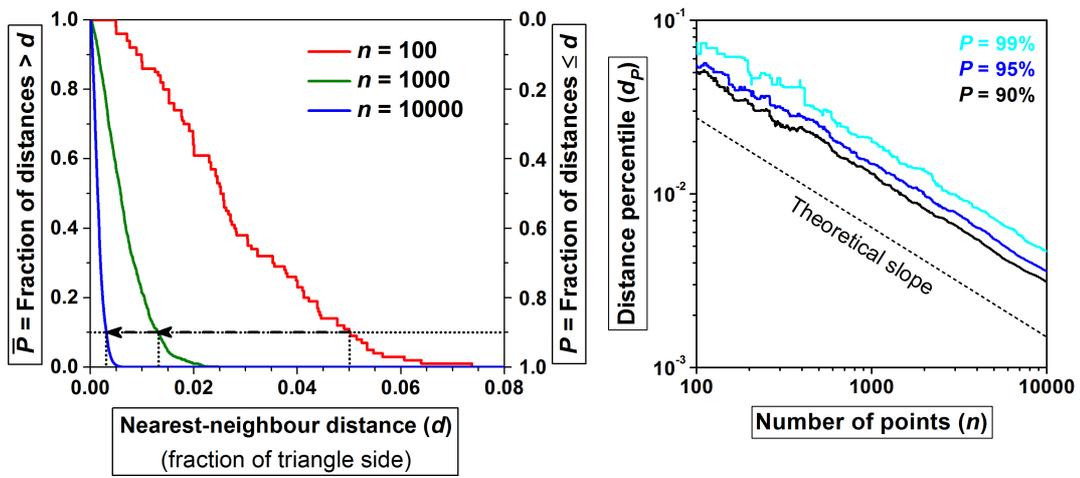


Figure 2.3: **Evolution of the probability distribution of distances in a Sierpiński gasket.** *Left:* As the number of points $n \rightarrow \infty$, nearest-neighbour distances tend to become smaller and their empirical distribution shifts (pictured for $n = 100, 1000$ and 10000 points). Any distance percentile d_P (such as for $P = 0.90 = 90\%$, marked) will decrease accordingly. *Right:* A double-logarithmic plot shows how the distance percentiles, exemplified for $P = 90\%$, 95% and 99% , decrease for increasing n . The power-law decay depends on the geometry of the distribution of points, since it is proportional to $n^{-1/D}$ where D is a dimension ($D = 1$ for a line, $D = 2$ for a uniform distribution of points on a plane, etc.). In particular, The Sierpiński gasket is a fractal, with dimension $D = \log(3)/\log(2)$, and a theoretical line with the expected power-law decay is shown.

where K is a constant, and D is the dimension of the attractor. Any percentile of the distribution of nearest-neighbour distances has dimensions of length, as \bar{d} , so from dimensional arguments, it will also decay in this way if the distribution lacks scale-dependent parameters (e.g. Barenblatt, 2003). Considering that the region (Fig. 2.1) may be a volume, the n points can distribute, among infinite options, in these configurations:

- on a set of fixed positions (a finite number of points), so $D = 0$;
- uniformly on a line, so $D = 1$,
- uniformly on a surface, so $D = 2$,
- uniformly on a volume, so $D = 3$,
- on an fractal attractor, with fractional, non-integer $D < 3$ (e.g. Turcotte, 1997; Barnsley, 2012), or
- on a multifractal attractor (which has a range of values for D , e.g. Harte, 2001). In this case there are scale dependencies and a different slope (and thus a different apparent fractal dimension) would be obtained for the power-law decay of each d_p .

Note that the more compact (or clustered) the long-term distribution of points (lower D), the faster the decay of d_p will be, resulting in a faster sharpening of the map.

Albeit the Sierpiński gasket lies on a bi-dimensional surface, its points do not fill it uniformly, but occupy only a fractional (fractal) dimension. This is larger than that of a line (so $D > 1$) but smaller than that of the embedding surface (so $D < 2$). In particular, for the Sierpiński gasket $D = \log(3)/\log(2)$, what explains well the observed decay of d_p with n (Fig. 2.3). Note that even in this theoretical, random, case, d_p has fluctuations, which are known to be intrinsic to such a kind of decay (Badii & Politi, 1984b).

The proposed forecast method will be tested with every new point added to the region. So n points and their distribution of distances will be used to forecast the location of the point $n + 1$, then these $n + 1$ points will be used to forecast the location of the point $n + 2$ and so on. No nearest-neighbour distance can be measured before the point $n = 2$ is added to the region, so it will be assumed that $\bar{P} = 0$ everywhere. The map of \bar{P} will be based on measured distances for forecasting the location of the points $n \geq 3$.

For an expected probability P^* , the *hit rate* will be the average fraction of *hits*, the points which take place at a nearest neighbour distance from the others $\leq d_{P^*}$

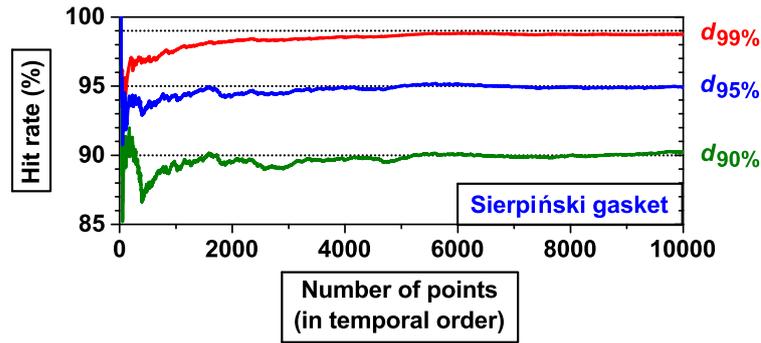


Figure 2.4: Hit rates expected (P^* , dotted lines) and observed (continuous lines) when forecasting points of a Sierpiński gasket. For each distance percentile d_{P^*} shown as example, the hit rate is the fraction of earthquakes occurred at a distance $\leq d_{P^*}$ from its nearest neighbour. That is, where $\bar{P} \geq 1 - P^*$ in the maps (Fig. 2.2). The fluctuations around the expected values (dotted lines) are relatively small after a few hundreds of points.

and are then correctly forecast (Fig. 2.1). Note that here this distance threshold changes as new points are added to the map, while other forecast methods based on nearest-neighbour distances consider instead a static map and a fixed distance around n points, and the hit rate is measured with the whole set of new points $n + 1 \dots \infty$ (Kafka, 2007; Kafka & Ebel, 2011, and references therein). We expect that, despite d_{P^*} usually decreasing as $n \rightarrow \infty$, the hit rate should be similar to P^* . So that in the long term, an expected fraction P^* of points takes place in the map where $\bar{P} \geq 1 - P^*$.

The actual hit rate for each distance threshold is very similar to the expected one, especially as n increases (Fig. 2.4). Initially, it oscillates, what is easily explained by the the initial oscillations of d_{P^*} due to the crude shape of the distribution of distances, and the few points available in the region to constrain the attractor. The hit rate for a larger P^* is intrinsically more stable, because it requires fewer trials (earthquakes) to be fulfilled. As n increases, more locations and distance information is available, the map sharpens (Fig. 2.2), and the hit rate becomes more similar to the expected one (Fig. 2.4). From this example (and other random trials not shown), it can be seen that the hit rate is already similar to the expected one after few hundred points, but a minimum of $n \sim 1000$ points seem to be required to approach the expected hit rate closely, at least for this geometrical figure.

Note that the method is memoryless, but its evaluation is not. It is memoryless because the map would look the same after n points, regardless of the temporal order in which they were added. But the hit rate (or the decay of distances percentiles and the area they cover) does depend on the specific order in which the points were added. Note also that the hit rate is measured in a pessimistic way,

since it is based in all the points, including those that occurred at the beginning of the simulation when the map was crude.

Although the method has been here exemplified with a fractal figure, it can be tested with any distribution of points added on a region. Tests made using diverse random distribution of points (such as multifractal or uniform) also yielded hit rates very similar to the expected ones.

But the map self-sharpens faster if the distributions of points is tighter (more clustered). The more disperse the distribution is, the larger will be the distances (and the areas covered by them) required to forecast a given fraction of points. This is due to Eq. 2.2, which shows that the distances decay faster for lower values of D .

For example, if the points can be located only on a fixed set of positions (dimension $D = 0$), as n increases, eventually they will start plotting on top of each other, so past point locations will be perfect predictors of future ones, and the map will tend to become perfectly sharpened.

Conversely, if points are distributed uniformly on a surface ($D = 2$), the map will never sharpen: it will have increasing detail, but to forecast a given fraction of points it will be required to consider the distances which cover exactly the same fraction of space. For example, in this uniform case, about 90% of points will take place in the 90% of the area around past points, regardless of how many points there are in the map. This is indeed the same as in a *random guess*, in which each point is expected to occur in a given percentage of area chosen randomly, so if tested in the long term, the same percentage of points will be forecast just by chance (e.g. Molchan, 1997; Zechar & Jordan, 2008).

Finally, note that the resulting map is perfectly self-consistent. That is, it would result in a perfect back-testing (*hindcasting*) of the point locations which were used to calculate it, which have the highest \bar{P} . This self-consistency would not be necessarily the case if the map had been generated by a theoretical smoothing kernel (e.g. Stock & Smith, 2002).

2.3 Application to earthquake catalogues

Earthquakes themselves may provide the largest amount of information useful for forecasting future earthquakes, since they are the only events related to the seismic process which are recorded systematically and continuously, in global and regional scales.

To test with earthquakes the forecasting method proposed in the previous section, it is necessary to use a complete earthquake catalogue (e.g. Vere-Jones, 1992), as otherwise we might obtain biased or spurious results. The magnitude

of completeness, denoted M_c , is the minimum one at which all earthquakes are detected and incorporated in a catalogue. It is heterogeneous in space and time, depending on the seismic network instrumentation, distance to the seismic stations, and many other factors, reviewed in a different chapter (see Sec 3.5). So for each catalogue we will consider a specific M_c and test the method using earthquakes with magnitude $M \geq M_c$.

Since the method is designed to forecast point locations, for each earthquake I will use its punctual representation projected on the Earth's surface. In most catalogues this location is the epicentre, while in the Global Centroid Moment Tensor Catalogue (Ekström & Nettles, 2015) it will be the projection of the centroid. The method can be extended to three dimensions considering the distances between hypocentres (or centroids), and the volume delimited within each distance threshold around them. But depth is typically the most uncertain spatial coordinate of earthquakes, and is frequently fixed when not well constrained (e.g. Bondár & Storchack, 2011, and Ch. 3). A maximum depth threshold will be used, since deeper earthquakes are less likely detected and located (that is, M_c increases with depth, e.g. Schorlemmer *et al.*, 2010b).

Typically earthquake catalogues have improved with time, due to the increasing capabilities for detecting and locating earthquakes (e.g. Ch. 3). A specific starting time will be chosen for each test, as the moment in which the catalogue is most complete (with lower M_c) and reliable (with lower uncertainties in earthquake locations). The method is expected to provide better results as more points (earthquakes) are used, so it is convenient to use a combination of starting time and M_c such as to include the maximum possible number of earthquakes.

Regional catalogues are typically more complete (have lower M_c) than global ones, allowing for more detailed, regional tests of the method in different tectonic environments. Each testing region has to be delimited precisely. It will be chosen as the geographic area in which the catalogue is most complete and epicentres are best located. Epicentres outside this area have to be disregarded, resulting in border effects, which are expected to be small, given that only nearest-neighbour distances are used (and not larger ones, between more distant earthquakes).

Choosing retrospectively a region has the possibility of biasing the results, even inadvertently (Mulargia, 2001). That is, one might be tempted to choose the specific region in which retrospectively the method works. In this section, only retrospective tests will be done (so "forecasting" earthquakes that have already occurred), in regions where the catalogues are best. In the next section, the tests will be fully prospective (actual forecasts) so any possible bias will be avoided.

In this approach, all past earthquakes with magnitude $M \geq M_c$ will be used to forecast the location of the next earthquake with $M \geq M_c$. Only later it will be checked if the results depend on the magnitude. In this way, the number of points in the map is maximized, so it can reach greater detail. Small earthquakes are expected to improve the spatial forecasts of larger ones (e.g. Kafka, 2007; Ebel, 2008; Helmstetter & Werner, 2014). Also, testing against the abundant small earthquakes, and not only against the large, infrequent ones, will provide statistically significant results using a shorter period of data.

The catalogues will not be *declustered* (trying to remove space-time clusters such as aftershock sequences or earthquake swarms), for four reasons. First, it is necessary to forecast the location of aftershocks, given that they are numerous and can be even more damaging than the mainshock (e.g. the 2011 Christchurch aftershock of the 2010 Darfield earthquake, in New Zealand, Michael, 2012). Second, declustering can be done using different techniques, which yield different results (e.g. van Stiphout *et al.*, 2012). Third, it would drastically reduce the amount of points available for analysis (e.g. $\sim 80\%$ of earthquakes in Southern California with magnitude ≥ 3 between 1984 and 2002 may be aftershocks according to Marsan & Lengliné, 2008). And fourth, the locations of past aftershocks are expected to be good predictors of the locations of future aftershocks within the same sequence (e.g. Felzer *et al.*, 2003).

All earthquakes will contribute equally to the distribution of distances (regardless of their magnitude) and the actual geometry of extended earthquake ruptures will not be explicitly taken into account. Aftershocks typically occur in the vicinity of the main shock rupture, delineating it and/or the surrounding rock damage zone (e.g. Liu *et al.*, 2003). And according to the so-called *productivity law* (Utsu, 1969), larger earthquakes typically trigger more aftershocks above the magnitude threshold M_c . So larger earthquakes will indirectly contribute more to the map because of their aftershocks, without the need of any tuning of the method.

The method will yield improved results if in the catalogue there are repeating earthquakes or foreshocks. *Repeating earthquakes* have very similar hypocentral locations, and may constitute a significant fraction of the total Earth's seismicity (Schaff & Richards, 2004; Dodge & Walter, 2015). The location of past repeating earthquakes will perfectly or almost perfectly forecast the location of future ones (so the hit rate will be non-zero even for zero or very small distances around past earthquakes). *Foreshocks* are earthquakes which occur near, and shortly before, a future earthquake (e.g. Hauksson *et al.*, 2011; Marsan & Enescu, 2012; Brodsky & Lay, 2014), although identifying them as such precursors is yet elusive (e.g. Brodsky & Lay, 2014). Given that the map will be updated as new earthquakes

happen, foreshocks (even if not formally identified as such) will raise the computed probabilities for the next earthquake to occur in their vicinity.

Distances on Earth will be measured as great-circle distances between points, using a spherical Earth model with radius of 6371.007 km, the sphere of equal area of the most frequently used Earth reference ellipsoids (Moritz, 2000; NIMA, 2000). Using a different radius would increase or decrease the distances between points proportionally and would not affect any of the forecast results. Using a spheroid model would change the values of distances only slightly (by $<0.5\%$, Earle, 2006).

To evaluate the performance of the forecast, it will be measured how much area around past earthquakes is covered within each distance threshold d_{p^*} , and the fraction of earthquakes (hit rate) which took place within such thresholds. The area within a distance around a point on a spherical surface is not a circle, but a spherical cap (e.g. González, 2010a, Ch. 4). The areas within a distance threshold $d \leq d_{p^*}$ around past epicentres are those with values $\bar{P} \geq 1 - P^*$ in the cumulative probability map. They constitute a complex set of partially overlapping spherical caps. The area of this set will be measured numerically with a point-counting technique specifically designed for this purpose, which uses the Fibonacci lattice, a quasi-optimal lattice of sampling points on the sphere (González, 2010a, detailed in Ch. 4). The lattice spacing was chosen such as there were $\sim 10^6$ sampling points within each region analysed, so areas are measured with negligible error (Ch. 4).

Each map will be valid in the time interval between the previous earthquake with $M \geq M_c$ and the next one. Such intervals between consecutive earthquakes will be calculated by the difference in the Julian dates of their origin times (e.g. Fliegel & van Flandern, 1968). The area within each distance threshold changes with time from map to map, so the temporal average of the area will be calculated, weighting the area of each map by the interval during which it is valid.

Location uncertainties tend to blur the actual spatial distribution of earthquakes, making it more uniform and less concentrated (e.g. Custódio *et al.*, 2015). So if uncertainties are large, the resulting maps calculated by the method will be less sharp, as discussed in the previous section. Location uncertainties affect especially the smallest nearest-neighbour distances. This is particularly evident when these are smaller than the typical location uncertainty, so that they become not physically meaningful and cannot be reliably related, for example, with the fractal dimension of the distribution (Meisel & Johnson, 1994). The largest nearest-neighbour distances (corresponding to largest percentiles in the distribution of P) will be least affected by location uncertainties and will be the only ones discussed below.

Where possible, high-quality, relocated, earthquake catalogues were used, in which earthquake coordinates are recalculated in a refined way, and which show

more grouped (clustered) earthquake distributions (e.g. Hauksson *et al.*, 2012). Such catalogues are not typically available in real time, but will allow to test the method with a distribution of epicentres closer to the real one, and potentially provide sharper maps.

Nearest-neighbour distances are also affected by rounding errors of the epicentral coordinates. If these are rounded too grossly for the scale of the region considered (in a too coarse fraction of a degree of latitude or longitude), eventually the epicentres would begin plotting on top of another, in a latitude-longitude grid, so the measured distances between them would tend to become zero and past earthquake locations would become perfect predictors of future earthquake locations, as discussed in the previous section. In the earthquake catalogues used below the coordinates are provided with enough precision such as to avoid this problem.

2.4 Retrospective testing

The method is tested here retrospectively with three catalogues: for the whole Earth, Southern California and the Iberian Region. Note that these tests are retrospective, since they will use earthquakes that have already occurred, but pseudo-prospective, in the sense that to forecast each earthquake only the previous ones will be used.

2.4.1 Catalogues used

For each region tested, the best available catalogues will be used (in terms of completeness and location precision), as follows.

2.4.1.1 Earth

For the whole Earth (Fig. 2.5), the definite earthquake database is the ISC Bulletin (International Seismological Centre, 2015b). It is built from data contributed by different agencies, revised, merged and used to relocate the global seismicity into a more complete catalogue with better location accuracy (Storchak *et al.*, 2000; Willemann & Storchak, 2001; Bondár & Storchack, 2011).

This refined compilation takes about two years to be accomplished, and at the time of this writing (November 2015), the reviewed bulletin is only available until January 2013. The test is performed for earthquakes with magnitude $M_c \geq 5$ and hypocentral depths ≤ 70 km occurred since 1980, for which this Bulletin is complete (i.e., the global M_c being 5.0 Woessner & Wiemer, 2005). For each earthquake, I use

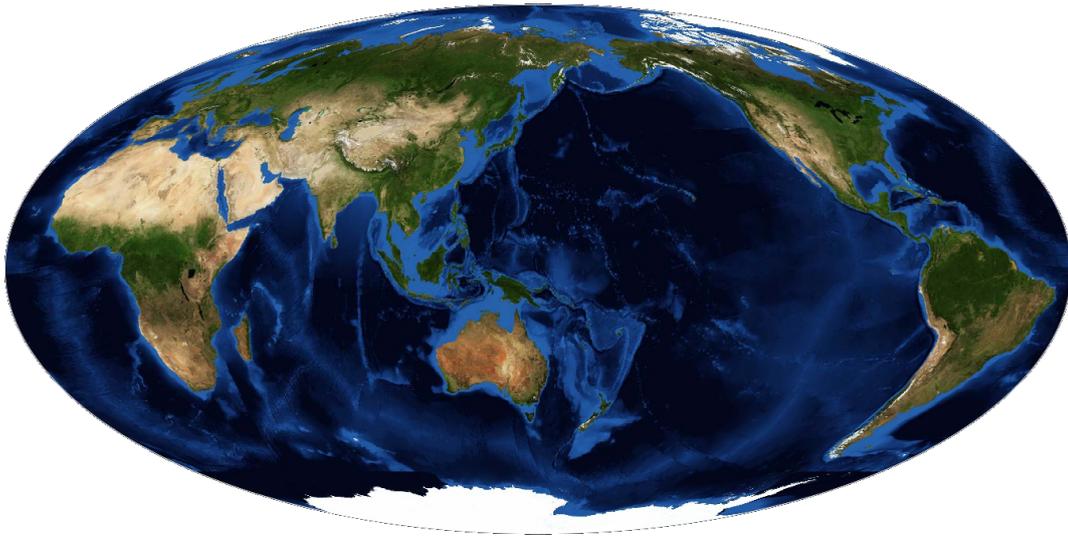


Figure 2.5: **Mollweide (equal-area) projection of the Earth used in this chapter.** It has been centered on the meridian 150°W in order to better show the belt of seismicity of the Pacific “Ring of Fire”. Real-colour satellite imagery from NASA (2002) and false-colour bathymetric shading derived from Amante & Eakins (2009).

the so-called prime (best) location and the largest magnitude associated to it. This selected data set contains 36,914 earthquakes.

2.4.1.2 Southern California

For Southern California I use the waveform-relocated catalogue of Hauksson *et al.* (2012), where earthquake locations are calculated by using a three-dimensional velocity model. Nowadays it is available online from January 1981 until December 2013 (Southern California Earthquake Data Center, 2015). This catalogue provides improved earthquake locations (showing tighter spatial clustering) than others Hauksson *et al.* (2012), and excludes events identified as quarry blasts.

I will use the same polygonal area analysed by Kagan (2007) (Fig. 2.6), where the epicentral locations have lower uncertainties (Kagan, 2007; Zaliapin & Ben-Zion, 2015). This also avoids as much as possible the retrospective selection bias when choosing the region boundaries. The absolute location error (expressed as a standard deviation) of the epicenters in this catalogue within this polygon is typically <1 km, and frequently <0.3 km (Zaliapin & Ben-Zion, 2015).

Earthquake monitoring in Southern California has improved along the decades and so the M_c of the catalogue has progressively decreased (Schorlemmer & Woessner, 2008; Hutton *et al.*, 2010). A conservative $M_c = 3$ and a maximum hypocentral depth of 30 km are chosen for this analysis, ensuring that this subset

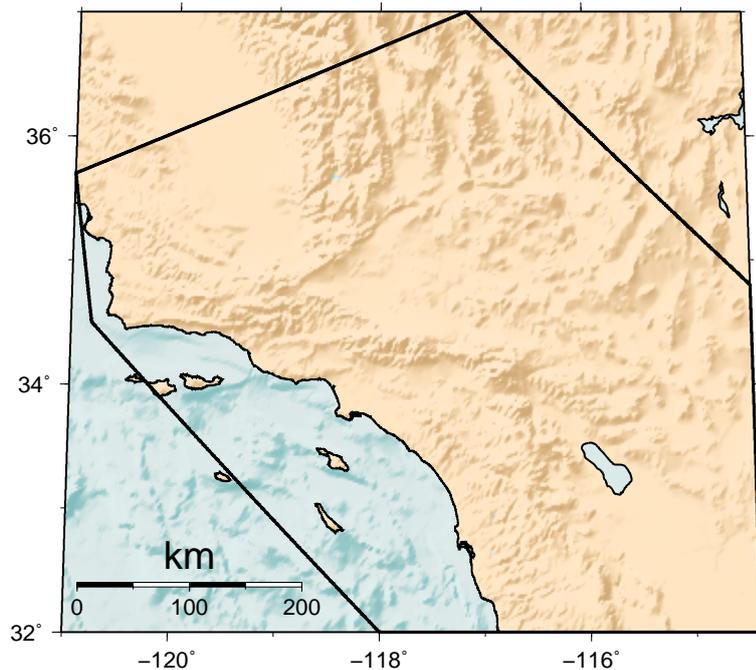


Figure 2.6: **Sinusoidal (equal-area) projection of Southern California used in this chapter.** The polygon (defined according to Kagan, 2007) delimits the area where the method was applied. Shading was derived from the topography of Amante & Eakins (2009).

of the catalogue can be deemed complete up to the border of the region and in the earlier years (Schorlemmer & Woessner, 2008). The central part of the analysed region has lower values of M_c (Schorlemmer & Woessner, 2008) and is amenable to future, more detailed, analyses. The selected data set contains 9,477 earthquakes.

2.4.1.3 Iberian Region

For the Iberian Region I will use the most complete data set available, the national Spanish earthquake catalogue, compiled by the Instituto Geográfico Nacional (2015). Chapter 3 presents an extensive analysis of this catalogue, and those results are taken into account here.

For consistency, I will use the same polygonal region (Fig. 2.7) considered in an earlier work (González, 2009), whose vertices are listed in Table 2.1. This polygon encloses the Iberian Peninsula, the Alborán Sea and Balearic Islands, where stations of the backbone seismic network are installed (Sec. 3.5.5). The epicentral location precision is substantially worse outside this polygon (Sec. 3.3).

Most of the seismicity in the region is shallow (depth ≤ 30 km), but depth values are frequently fixed or poorly determined in the catalogue (Sec. 3.3.3), so only the

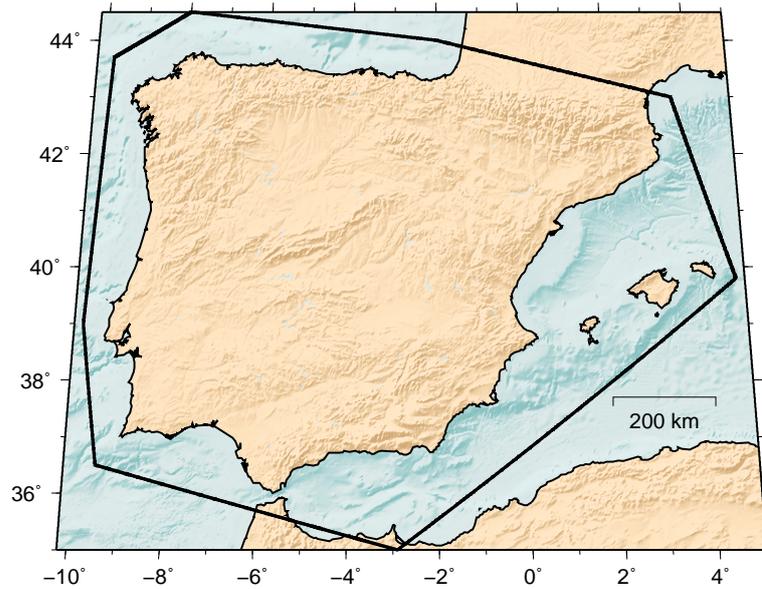


Figure 2.7: **Sinusoidal (equal-area) projection of the Iberian Region used in this chapter.** The polygon (defined in Table 2.1, according to González, 2009) delimits the area where the method was applied. Shading was derived from the topography of Amante & Eakins (2009), replaced by the bathymetry of Zitellini *et al.* (2009) where available.

few deep earthquakes (hypocentral depth > 600 km) will be excluded from the analysis.

Only the data from January 1985 onwards will be used (as by González, 2009), since before this date only a few stations were in operation and the location precision was worse (Sec. 3.3). The test will be done with events occurred until October 2015, with the catalogue as downloaded from the online server (Instituto Geográfico Nacional, 2015) on November 2nd, 2015.

The magnitude of completeness has improved (lowered) substantially along the decades (see Sec. 3.5 for details). To maximize the number of earthquakes usable, several thresholds of M_c will be used, instead of using a single, higher one:

- $M_c = 3.5$ from 1985 until October 1997. Since November 1997, improved procedures for automatic earthquake detection and location lowered M_c .
- $M_c = 3.0$ from November 1997 until February 2002. In March 2002, the formula for calculating most magnitude values in the catalogue changed.
- $M_c = 2.5$ from March 2002 until October 2015.

In Sec. 3.7 it is found that, despite the efforts of the network operators at identifying blasts, a small fraction of them are misidentified as natural earthquakes and included in the catalogue. Using a magnitude threshold of $M \geq 2.5$, as done here,

Longitude (°)	Latitude (°)
3.6	43.0
4.8	39.8
-2.9	35.0
-9.5	36.5
-10.0	39.0
-9.8	43.7
-8.0	44.5
-2.0	44.0

Table 2.1: **Coordinates of the polygon vertices used for delimiting the Iberian Region** (Fig. 2.7).

is expected to eliminate the vast majority of blasts, if not all (see Sec. 3.7 for details). The selected data set contains 4,095 earthquakes.

2.4.2 Results

The maps obtained by the forecast method are shown here in equal-area projections, to provide a reliable visual representation of the area covered within each distance (or \bar{P}) threshold. For the Earth, a Mollweide projection was used (Fig. 2.5) and for the smaller regions a sinusoidal one (Figs. 2.6 y 2.7).

These maps (as most maps in this dissertation) are plotted with Generic Mapping Tools (Wessel & Smith, 1998; Wessel *et al.*, 2013), using software scripts specifically written for this purpose, which enable automatized map updates. Example forecast maps are drawn after the first $n = 100$ and $n = 1000$ earthquakes, and after the final earthquake of the period considered, in Figs. 2.8, 2.9 and 2.10.

The maps tend to illuminate the surface projection of faults, and show increasing detail and complexity as new earthquakes are included. Unlike traditional maps of epicentres, where the symbols tend to clutter the map as more earthquakes are represented, here the probability footprint of each epicentre becomes smaller as new ones are added, allowing the map to sharpen continually. The last map for the Earth shows a very fine-grained distribution of probabilities which evidence complexity within the large-scale tectonic features. In the map for Southern California the calculated probabilities are heterogeneous along the trace of the San Andreas Fault, and others faults are also highlighted by conspicuous streaks of epicentres. The largest patches with high \bar{P} in the map are typically due to clusters of aftershocks of major earthquakes, discussed later. In the Iberian Region, some streaks of epicenters also follow known fault traces (García-Mayordomo *et al.*, 2012).

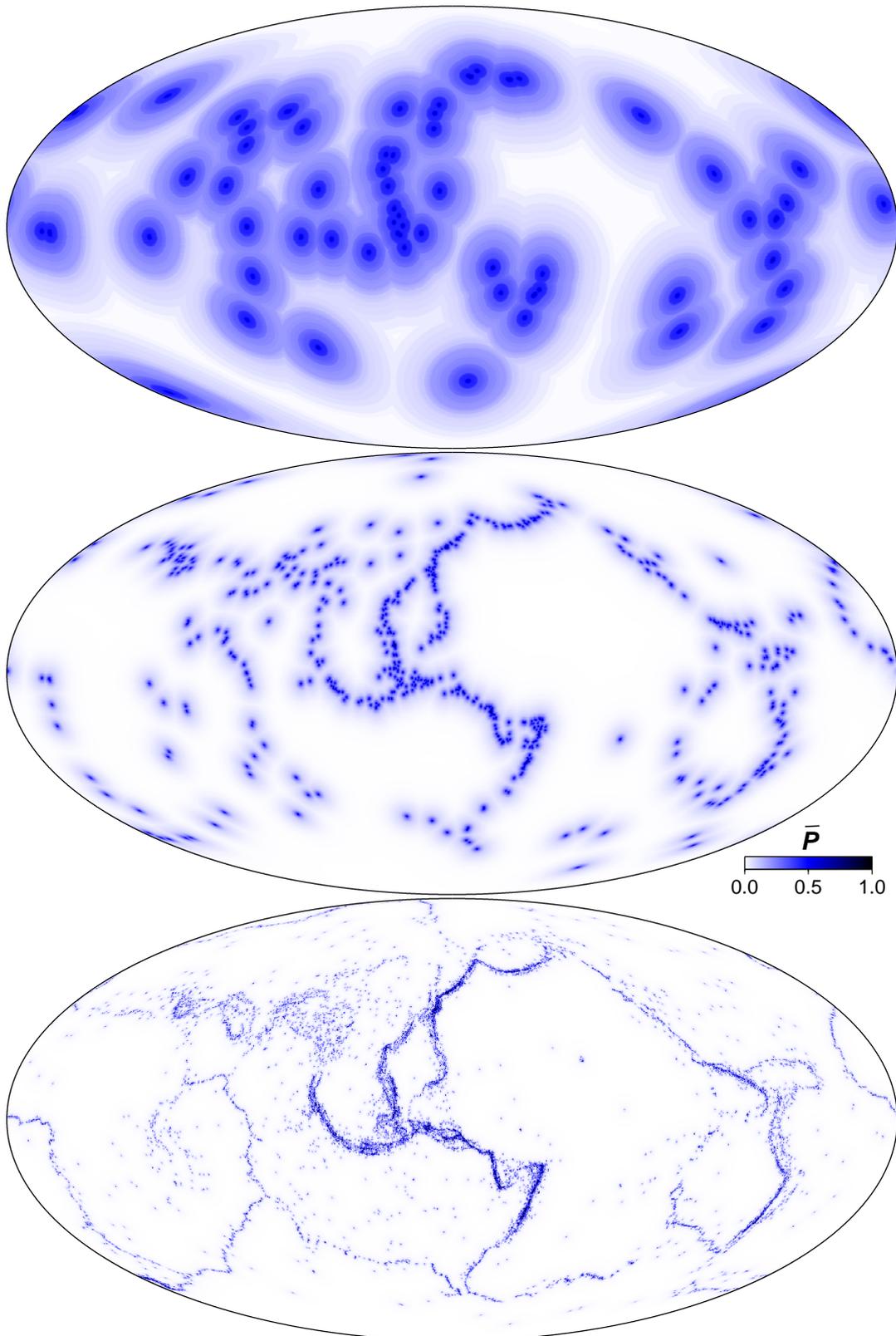


Figure 2.8: **Self-sharpening forecast maps for the whole Earth, after 100, 1000 and 36,914 earthquakes since January 1980.** For coastlines, see Fig. 2.5. *Colour scale:* Empirical probability of distance exceedance \bar{P} (Figs. 2.1). For each map, the next earthquake ($M \geq 5.0$ and depth ≤ 70 km) is expected to occur with probability P^* in the areas where $\bar{P} < P^*$, and with probability $1 - P^*$ where $\bar{P} \geq P^*$. The last forecast map is issued for the first earthquake of February 2013.

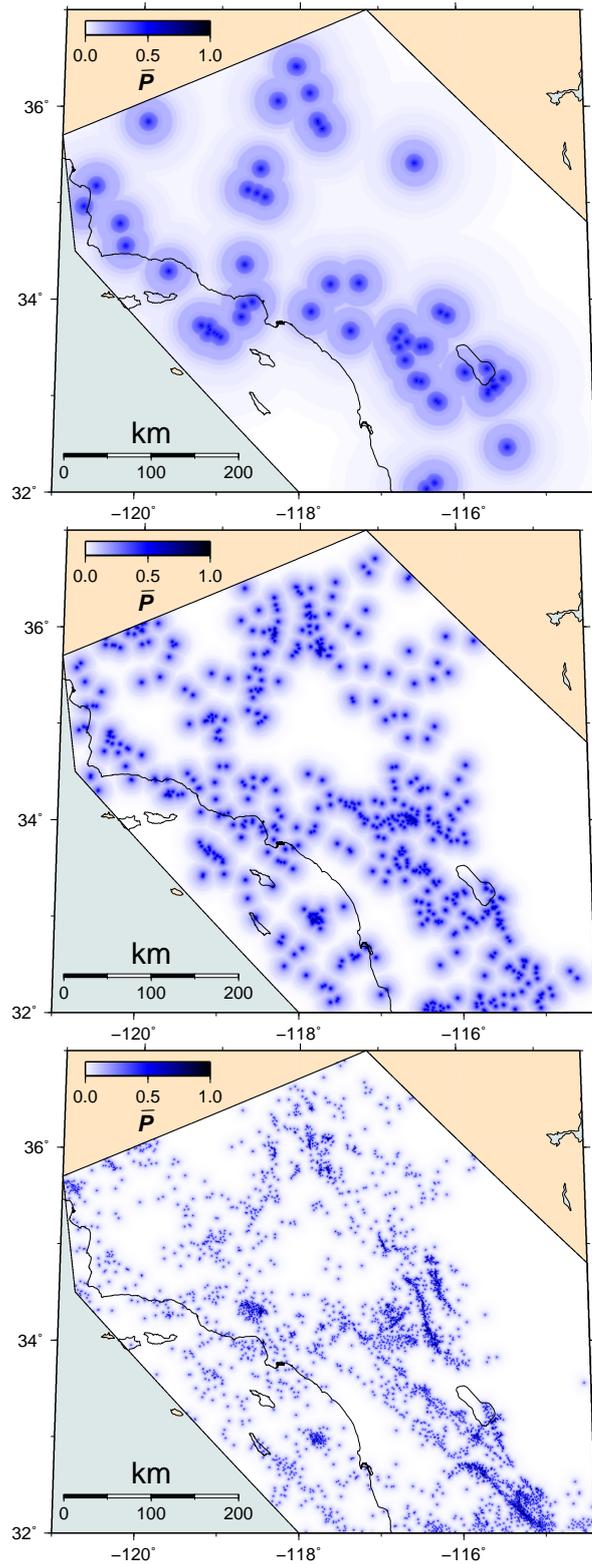


Figure 2.9: **Self-sharpening probability maps for southern California, after 100, 1000 and 9,477 earthquakes since January 1981.** Colour scale: Empirical probability of distance exceedance \bar{P} (Figs. 2.1). For each map, the next earthquake ($M \geq 3.0$ and depth ≤ 30 km) is expected to occur with probability P^* in the areas where $\bar{P} < P^*$, and with probability $1 - P^*$ where $\bar{P} \geq P^*$. The last forecast map is issued for the first earthquake of January 2014.

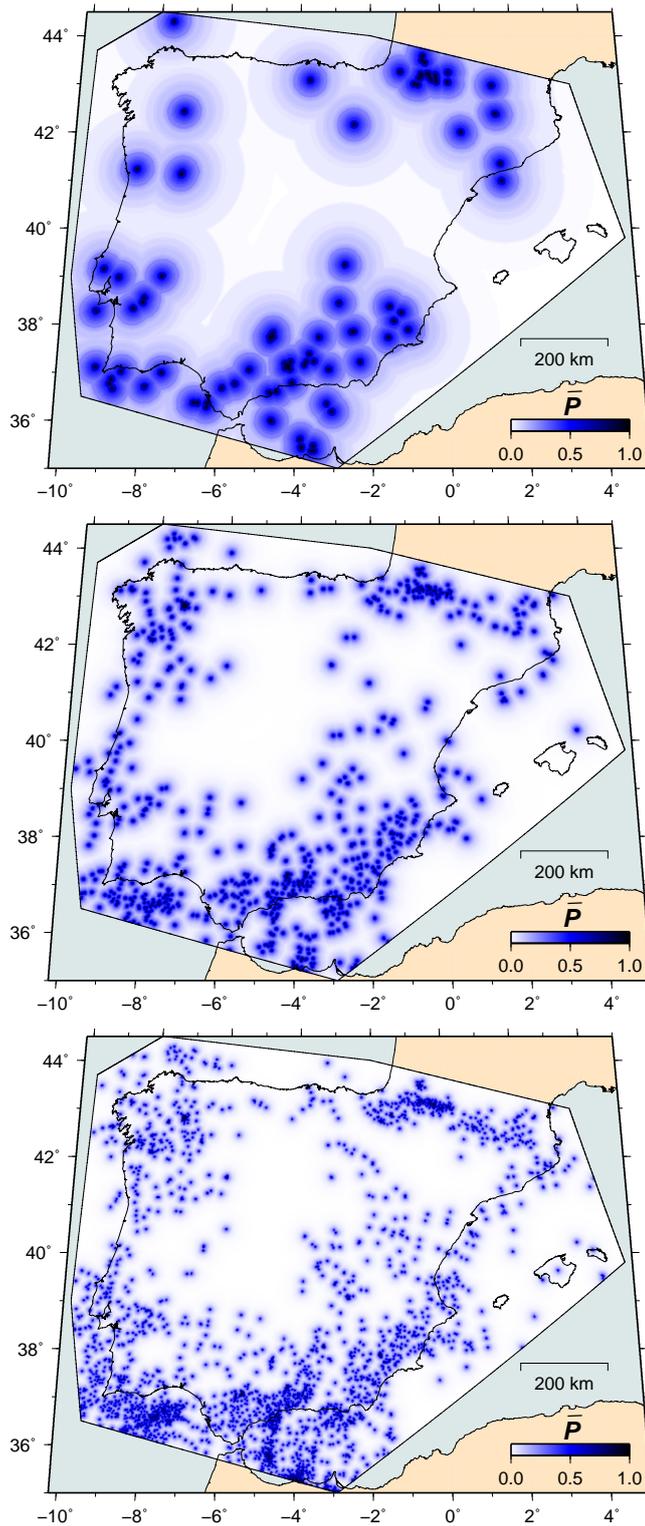


Figure 2.10: **Self-sharpening probability maps for the Iberian region, after 100, 1000 and 4,095 earthquakes since January 1985.** Colour scale: Empirical probability of distance exceedance \bar{P} (Figs. 2.1). For each map, the next earthquake is expected to occur with probability P^* in the areas where $\bar{P} < P^*$, and with probability $1 - P^*$ where $\bar{P} \geq P^*$. The last forecast map is issued for the first earthquake with $M \geq 2.5$ of November 2015.

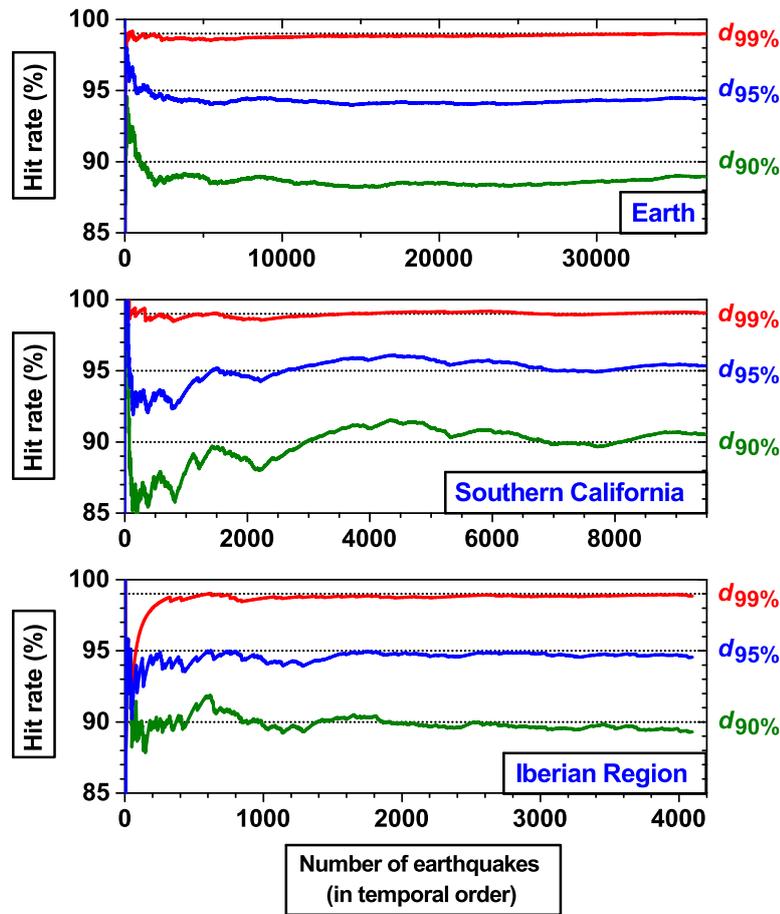


Figure 2.11: Hit rates expected (P^* , dotted lines) and observed (continuous lines) when forecasting earthquakes in the retrospective tests. The hit rate for each distance percentile d_{P^*} (Figs. 2.1 and 2.14) is the fraction of earthquakes occurred at a distance $\leq d_{P^*}$ from previous ones (where $\bar{P} \geq 1 - P^*$ in the map). The observed values closely approach the expected ones after ~ 1000 earthquakes.

The observed hit rates are equal than, or very similar to, the expected ones (Fig. 2.11). Only the performance for high percentiles of the distance distribution is plotted. Those percentiles correspond to longer distances and thus are the least affected by location uncertainties. The hit rate for a larger P^* is more stable, as for the theoretical case discussed earlier where points had no location uncertainties (Fig. 2.4).

The fractions of area covered within these distance thresholds in Southern California, chosen as a representative example, are shown in Fig. 2.12. Initially they decay faster with the increasing number of earthquakes in the map, and then more slowly, until they tend to stabilize. This might evidence a predictability limit, so that to forecast a given fraction of earthquakes a minimum area of the map might have to be covered despite the increasing information on past earthquake locations.

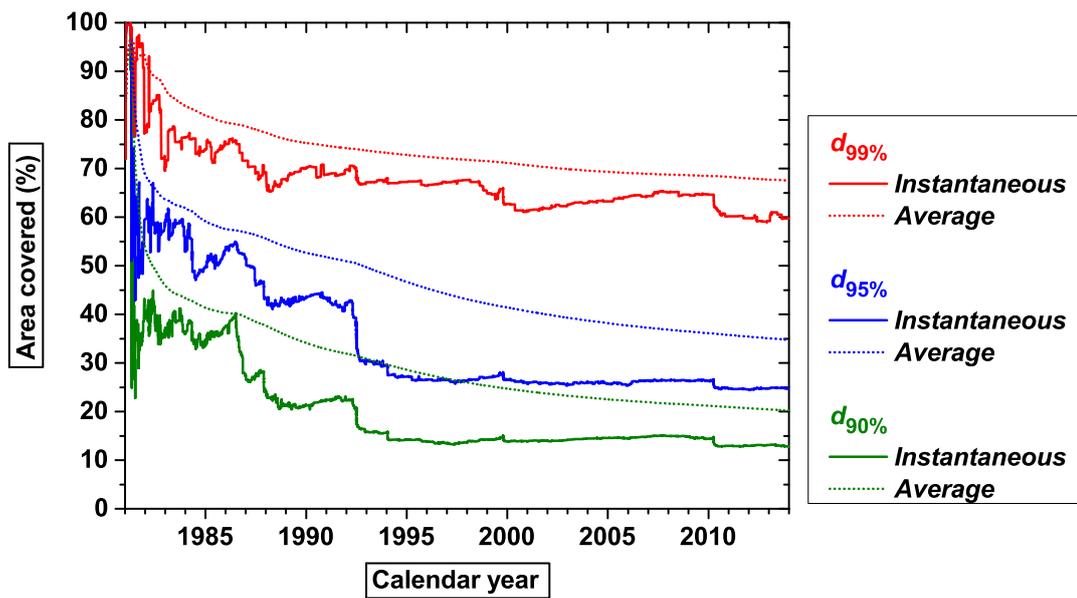


Figure 2.12: **Fraction of area covered within different distance percentiles in Southern California.** The *instantaneous* values are those for the forecast map calculated with earthquakes that had already occurred at each moment. The *average* values are the temporal running means since the beginning of the test until that moment. For example, by the end of the test, $\sim 90\%$ of earthquakes took place in the areas within a distance $d_{90\%}$ around past earthquakes, where $\bar{P} \geq 0.1$ (Fig. 2.11), and these areas have covered on average $\simeq 20\%$ of the map. Areas have been measured numerically with a homogeneous, dense lattice of sampling points (Chapter 4).

Name	Date	Longitude (°)	Latitude (°)	M	n
Superstition Hills	1987/11/24	-115.83650	32.98658	6.6	1302
Landers	1992/06/28	-116.43500	34.20183	7.3	2420
Northridge	1994/01/17	-118.56213	34.22713	6.7	3961
Hector Mine	1999/10/16	-116.27033	34.59783	7.1	5284
El Mayor-Cucapah	2010/04/04	-115.29450	32.26433	7.2	7601

Table 2.2: **Largest earthquakes ($M > 6.5$) in Southern California during the period of retrospective testing.** Coordinates and magnitudes as provided by Hauksson *et al.* (2012). The index n is the sequential number of the earthquake in the test.

The average hit rate and percent area shown are pessimistic values, compared with the expected performance of the latest map. This is because the results of the earlier, crude maps are averaged with those of the newer, sharpened ones. Typically the hit rate at the beginning is lower than expected (as in the theoretical Sierpiński example) and the performance of newer maps compensate slowly the earlier misses. At the beginning the maps have typically large areas covered under any distance threshold, which are averaged with the smaller ones which are achieved later on. For example, the average areas are much larger than the instantaneous ones covered at the end of the retrospective test in Southern California (Fig. 2.12).

These results for earthquake catalogues are affected by the correlations between events, especially by sequences of aftershocks, which occur close to each other in space and time (e.g. Zaliapin *et al.*, 2008; Zaliapin & Ben-Zion, 2015). This is a major difference with the test with uncorrelated points placed at random in the theoretical example, where any fluctuations observed are purely stochastic in origin.

The most obvious effects of correlations are observed in Southern California, related to aftershocks of earthquakes with $M > 6.5$, which are listed in Table 2.2. Aftershocks rapidly increase the number of epicentres available to calculate the map, which produces a sudden sharpening, manifested by a shrinkage of the area covered under a given distance percentile (Fig. 2.12). This is particularly noticeable at the beginning of the test period, when there are fewer earthquakes available for making the map and thus the increase of the number of events is proportionally larger. For example, the area dropped much more after the 1987 Superstition Hills earthquake than following the 2010 El Mayor-Cucapah earthquake, despite the latter had larger magnitude and generated a prolific aftershock sequence (e.g. Hauksson *et al.*, 2011). This indicates that the pattern of the map becomes progressively more stable with time, as the number of earthquakes used to make it increases.

Aftershocks are also especially well forecast by the method, as expected, what typically rises the hit rate (as shown in Southern California, in Fig. 2.11, after the events numbered in Table 2.2). This seems to be compensated by “background seismicity” occurring further away from previous events once the aftershock sequences fade. That these are effects are due to correlations between events was confirmed by reshuffling the catalogue (so that earthquakes occur at the same locations but in different, random, temporal order). In reshuffled catalogues, the hit rates are also similar to the expected ones, but their oscillations become typically smaller. It is noteworthy that in actual earthquake catalogues, despite of the correlations between earthquakes and the great variations of earthquake rates implied by aftershock sequences, the long-term hit rate deviates very little from the expected one.

The self-sharpening of the maps can be related to the geometry of the spatial distribution of epicentres. The empirical distribution of nearest-neighbour distances used for calculating the final maps in each region is shown in Fig. 2.13. The shape of these distributions (especially their tails) is different for each region.

Figure 2.14 shows the decay of the distance percentiles for $P=90\%$, 95% and 99% as n increases. When n is small these percentiles may take the same values, since they are based on the empirical distribution of distances, without interpolation. The scaling with n (expected from Eq. 2.2) is more evident only after $n \gtrsim 1000$, indicating that the distribution is poorly constrained for smaller n , as discussed for the hit rates.

Such a decay, unlike in the example of the Sierpiński gasket (Fig. 2.3), evidences a different scaling exponent (slope of the double-logarithmic plot) for different percentiles, and is consistent with Eq. 2.2 if the dimension D of the distribution of epicentres is between 1 and 2. For example, Corral (2003) calculated a fractal dimension $D = 1.6$ for the epicentral distribution in Southern California, similar to that of the scaling shown here. Alternatively, multifractal scaling (e.g. Molchan & Kronrod, 2009; Kamer *et al.*, 2015), with different D at different scales of observation, would explain the variability in the observed slopes within each region. The 99% percentile, located at the tail of the distribution, is the least reliable, and shows a more irregular decay than the others. The calculated fractal dimensions of an epicentral distribution depend on the location uncertainties, border effects due to the finite area of the geographic region, the geometry of the projection of the hypocentres (different for different fault dips), and temporal and spatial correlations between events (e.g. Eneva, 1994; Kagan, 2007, 2014). An in-depth analysis of the geometry of the epicentral distribution is out of the scope of this chapter, but the implied $1 < D < 2$ implies that the forecasts maps self-sharpen faster than if

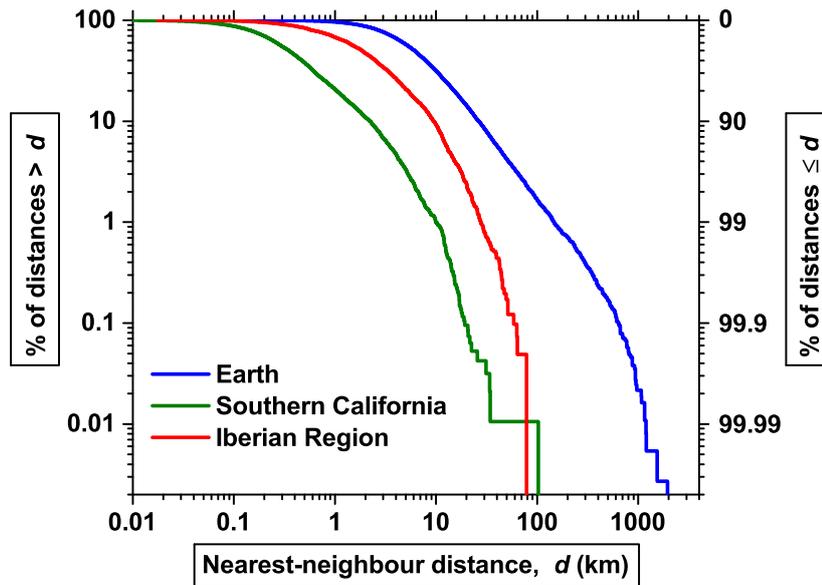


Figure 2.13: **Distribution of nearest-neighbour distances between earthquakes in each region.** It is expressed as a frequency of exceedance (complementary of the empirical distribution function). Note the double-logarithmic scale. These are the distributions used for calculating the last forecast maps in each region (bottom maps in Figs. 2.8, 2.9 and 2.10). Distance percentiles (last values in Fig. 2.14) correspond to the percentages labelled by the scale at right. The maximum distance in each distribution is the largest value shown. Note the differences in the distributions, both in shape and scale, which justify using each one for its particular region, instead of a generic one.

epicentres were distributed uniformly at random ($D = 2$) over the geographic areas considered.

The overall performance of the retrospective forecasts can be evaluated with a Molchan diagram (Fig. 2.15 Molchan, 1997, 2003, emphasizing here the hit rate instead of its complementary, the miss rate). The results shown in previous figures referred to specific distance percentiles chosen as examples. In contrast, the curve of the Molchan diagram is drawn considering all the distance percentiles, representing for each of these its hit rate versus the average space-time it covered.

There is an excellent agreement between the expected and observed hit rates for the fraction of area covered, on average (fraction of total space-time) within each distance (or \bar{P}) threshold. These results with such a high number of earthquakes, are statistically significantly better than the null hypothesis of a random guess (e.g. Zechar & Jordan, 2008). A random guessing strategy would consist in trying to forecast the location of the earthquakes choosing areas at random, and the long-term hit rate would equal the area fraction covered on average (so its results would lie on the marked diagonal). A perfect forecast would lie on the upper left corner of

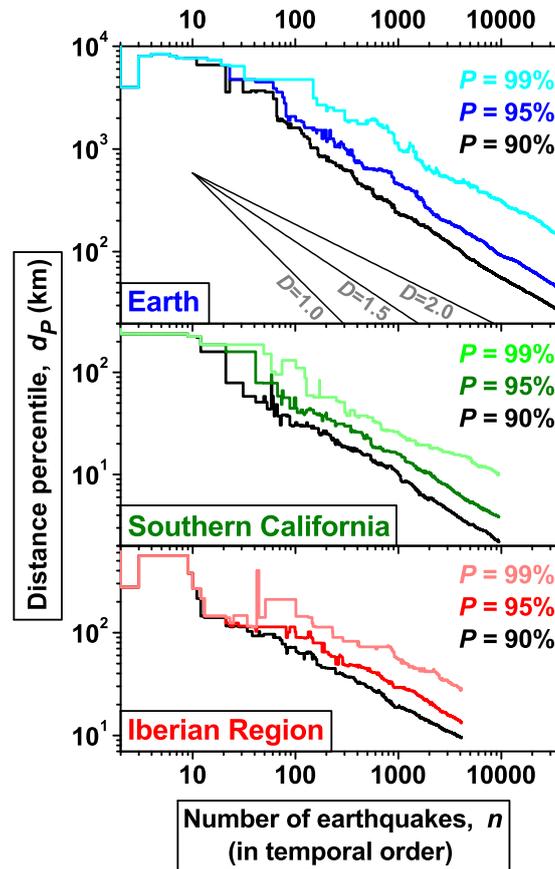


Figure 2.14: **Decay of distance percentiles as earthquakes take place in each region analysed.** The percentiles shown are for percentages $P=90\%$, 95% and 99% of the distribution of nearest-neighbour distances. The hit rates corresponding to these percentiles are shown in Fig. 2.11, and the respective area covered in Southern California in Fig. 2.12. The last values shown in each region correspond to the distance distributions in Fig. 2.13. Note the double-logarithmic scale, which is the same in all diagrams. These are shown separately to avoid overlaps. The percentiles, especially for $n \gtrsim 1000$, decay as power laws. The slopes of theoretical decays are shown for points distributed in space in dimensions $D = 1.0, 1.5$ and 2.0 . The observed scaling is consistent with a fractal (possibly multifractal) distribution of epicentres in space, with fractal dimension (or dimensions) $1 < D < 2$.

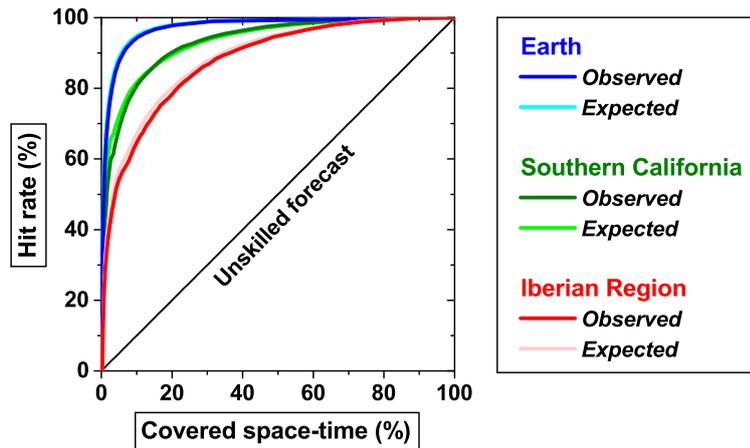


Figure 2.15: **Assessment of forecast performance and comparison with the expected one.** There is an excellent correspondence between the results expected from the method (*clear lines*) and the observed ones (*darker lines*). Each point of the curve indicates the percentage of earthquakes forecast within the distances around past epicentres which covered, on average, a given area (fraction of space-time).

the diagram, in which 100% of the future earthquake locations would be pinpointed beforehand (with 0% area covered).

The seismicity of the Earth as a whole is the most predictable spatially, since there are large areas almost devoid of earthquakes (Fig. 2.8). Southern California is an intermediate case, probably in part because the data set contains about 4 times less earthquakes than for the whole Earth, so the map cannot be as detailed. The data set for the Iberian Region contains even fewer events, and perhaps the seismicity is also more spatially diffuse, so a larger fraction of area around past earthquakes is required to reach the same hit rate than in the other cases.

No earthquake is expected to occur at a nearest-neighbour distance d larger than the maximum one d_{\max} between any pair previous earthquakes occurred until that moment. If an earthquake happens at a distance $> d_{\max}$ from its nearest neighbour (where $\bar{P} = 0$), the method fails. In that case, the whole map area has to be regarded as covered in order to forecast that location, as for the first two earthquakes ($n = 1$ and $n = 2$). This increases up to 100% the average area necessary to achieve a 100% hit rate. Such a situation may happen occasionally when $n > 2$, and especially at the beginning of the test, when n is small and the distribution of distances is poorly constrained. In these retrospective tests, most of the area ($>90\%$) was covered within a distance d_{\max} ($\bar{P} > 0$) around past earthquakes, and very few earthquakes ($n > 2$) took place beyond it (six in the whole Earth, five in Southern California and six in the Iberian Region). Some statistical tests would reject a forecast model which assigned strictly zero probability to a site where an earthquake happened,

Region	M	Total	Hits	Hit rate (expected 90%)
Earth	≥ 8.0	19	17	89%
Southern California	≥ 6.0	10	9	90%
Iberian Region	≥ 5.0	13	11	85%

Table 2.3: **Results for the largest earthquakes in each region, within the 90% distance percentile around past epicentres.** Considering the small number of events, the observed hit rates are not statistically significantly different from the expected one.

even if the model performed very well with a plethora of other earthquakes; but this is counter-intuitive and not particularly useful to assess model performance (as discussed by Taroni *et al.*, 2014). Such issues might be avoided by assigning a tiny, but not null, probability to those sites (Taroni *et al.*, 2014).

The method shows similar performance for the largest earthquakes as for the whole set (Table 2.3). Given that the largest earthquakes are scarce, I will only focus on the results for the expected hit rate of 90%, instead of a larger one, so that there will be a few worst cases to discuss below. For Southern California the observed hit rate is already 90% (Table 2.3), while in the other regions is smaller. Disregarding correlations for a moment, let us assume that each earthquake forecast is an independent Bernoulli trial (a random experiment with exactly two possible outcomes, “success” and “failure”), with expected success rate $P=90\%$. Then, using exact confidence intervals for the binomial distribution (Clopper & Pearson, 1934) we can test if the observed hit rate is too small to result from this expectation, for the (small) number of earthquakes in the sample. Indeed the expected long-term 90% success rate cannot be rejected at any level of statistical significance for the results shown in Table 2.3.

For the whole Earth, two great earthquakes ($M_{ISC} \geq 8.0$) were missed at the 90% distance threshold. One was the May 12th, 2008 Sichuan earthquake, which took place at a distance $d = 62.7$ km from the closest epicentre of a $M \geq 5.0$ previous earthquake (96.4% distance percentile). Of the total Earth’s area, 9.7% was a the time covered within that distance from previous epicentres. The second was the April 11th 2012 east Indian Ocean earthquake (e.g Pollitz *et al.*, 2012), which took place at $d = 60.1$ km (96.6% distance percentile).

For southern California the only large earthquake missed at the 90% distance threshold was Northridge, which occurred 7.096 km (95.76% distance percentile) away from its nearest previous earthquake of $M \geq 3.0$. Up to 34% of the region was within this distance threshold from previous epicentres in that moment. The epicentre of the largest earthquake (Landers), on the contrary, took place a mere

0.452 km away from the nearest one, and only 0.44% of the region was within such a small distance from previous epicentres.

For the Iberian region, the two largest earthquakes missed by the 90% distance threshold are the March 12th 1992, $M = 5.3$ in the Alborán Sea (for which the forecast map was based on less than 200 epicentres), and the August 6th, 2002 La Poca earthquake ($M = 5.0$, Benito *et al.*, 2007). The latter took place 19.6 km away from the nearest epicentre considered (90.6% distance percentile, 31% area covered at the time). The most damaging earthquake in Spain during the analysed period (Lorca, May 11th 2011, $M = 5.1$, e.g. Martínez-Díaz *et al.*, 2012) took place at just 0.4 km from the nearest epicentre, and only 0.13% of the geographic area was covered within that distance threshold.

These results illustrate a trade-off, so that in order to forecast the location of a larger fraction of earthquakes (including damaging ones) a larger distance threshold has to be used (covering a larger fraction of the map). That the overall results for large earthquakes are similar to those for small ones highlight the importance of taking small earthquakes into account for forecasting the location of larger ones (as also already pointed out, e.g. by Ebel, 2008).

2.5 Independent daily testing in the Collaboratory for the Study of Earthquake Predictability

To test independently the performance of these spatial forecasts I submitted the procedure for formal, automatic, daily testing at the Southern California Earthquake Center (Los Angeles, USA), within the Collaboratory for the Study of Earthquake Predictability (CSEP). This is an international cooperative project, devised as a continuation and expansion of the experiments for testing forecast models in California (RELM, the Regional Earthquake Likelihood Models experiment, Field, 2007; Schorlemmer & Gerstenberger, 2007). Apart from the computational testing centre in California, others are based in Europe (Schorlemmer *et al.*, 2010a), Japan (Tsuruoka *et al.*, 2012) and New Zealand (Gerstenberger & Rhoades, 2010), with prospects also in China (Mignan *et al.*, 2013). The research presented here has been the first contribution from Spain to this international initiative.

The goal of CSEP is to explore earthquake predictability by testing the forecast models in a truly prospective, automatic, and reproducible way (Schorlemmer & Gerstenberger, 2007; Zechar *et al.*, 2010). Tests are made in near-real time, independently of the researcher who submitted the model, what prevents any possibility to influence the results. In CSEP the forecasts are expressed in a

standardized way, to facilitate comparison between the models (Schorlemmer & Gerstenberger, 2007).

The proposed model has been tested during six years at the time of this writing (November 2015), for earthquakes occurred in California, the whole Earth and the western Pacific. During such a long period, enough target earthquakes occurred in each of these regions to allow statistically meaningful tests. The following subsections present first an overview of the CSEP testing experiment, then explain how the model was adapted to CSEP standards, how it is being tested, the regions and data used for the tests, and the overall results, compared to those of a reference model.

2.5.1 Setup of the testing experiment

The computational setup of each CSEP testing centre is complex and technically demanding, as detailed by Zechar *et al.* (2010). Scientists (modelers) cannot test directly the models, but have to submit their original software code for independent testing. The code must be able to read the data required (such as a specifically formatted earthquake catalogue), to compute the forecast for a pre-defined grid of cells in which each testing region is divided, and to save it in a standardized XML format. After checking that the software integrates properly in the CSEP computing environment and generates the expected (and reproducible) results, prospective testing can start. The model is tested against the occurrence of earthquakes which had not yet taken place when the software was initially submitted.

The original software for computing the forecasts with this model was written in Fortran 90 and called *Nearest* (short for “nearest-neighbour” forecast). It can be freely used and modified, under a GNU General Public Licence (version 3 or later, Free Software Foundation, 2007). Most software used by CSEP (to a large extent developed on purpose for this collaboration) is free and open-source. Including *Nearest*, it can be downloaded for personal use, either directly through the web page or under request through the contact form there included. The testing centres store all the information required to ensure reproducibility of the forecasts (source codes, original earthquake catalogues, computed forecasts and test results, Zechar *et al.*, 2010).

The maps and tests are made as in real time (for example, the forecast for each day is made daily, using the data automatically downloaded for earthquakes that occurred until the day before). The compilation of a real-time earthquake catalogue is initially preliminary, and has to be revised manually, with some delay, to correct errors (see, e.g. Sec. 3.2.14). So any real-time catalogue version is flawed to some extent and, if the tests were made with it, they would also be unreliable. Since the

goal of CSEP is to test the models rigorously, it is necessary that the catalogue is revised, and reasonably close to the final version; for this reason the maps are made and tested with a delay of one month (Schorlemmer & Gerstenberger, 2007).

Forecast maps and test results are posted automatically in the web page www.cseptestesting.org (Fig. 2.16). CSEP tests models in different regions, for different time periods (such as one day, one month, three months, a year, five years), and of several kinds, such as *alarm models* (aimed to forecast earthquake locations) or *rate models* (aimed to forecast earthquake rates –frequencies). The updated maps and results of the model presented here can currently be checked in www.cseptestesting.org/results/scec, choosing the test region and selecting the *One-day alarm models* class (Fig. 2.16).

2.5.2 Adaptation of the model to CSEP standards

In CSEP, the maps are discretized into a pre-defined grid of latitude-longitude cells (whose resolution depends on the region considered, Table 2.4) and two basic types of forecast models are being tested. In most models (e.g. Field, 2007) each cell of the map is assigned a rate (frequency) of earthquakes, for a specified magnitude and depth range. Other models, instead, are intended to forecast only the spatial distribution of future earthquakes. In these, each cell is assigned a specific value of a function (*alarm function* Zechar & Jordan, 2008). It is analogous to the term *susceptibility* used when forecasting other natural hazards (e.g. Galve *et al.*, 2011; Becerril *et al.*, 2013). The alarm function increases (not necessarily in an explicit way) with the expected likelihood of future earthquakes in that cell during the testing period (Zechar & Jordan, 2008).

The pixelization of the maps in latitude-longitude cells is not homogeneous, since the geographic area covered by each cell is larger near the equator and smaller near the poles (e.g. González, 2010a, Ch. 4). For a spherical Earth model with radius R , the area of a cell within latitudes ϕ_1 and ϕ_2 , and longitudes λ_1 and λ_2 (in degrees) is:

$$a = \frac{\pi}{180} R^2 |\sin(\phi_1) - \sin(\phi_2)| |\lambda_1 - \lambda_2|. \quad (2.5)$$

This implies, for example, that a $1^\circ \times 1^\circ$ cell bounded by the equator is > 114 times larger than one bounded by the pole. For each region I submitted to CSEP a null hypothesis model, named *Uniform*, specifying the area fraction of the testing region, a_i , covered by each cell i .

The *Nearest* model as adapted to CSEP has again no parameter. The alarm function \tilde{P}_i for each cell i was chosen as the area integral of \bar{P} over the two

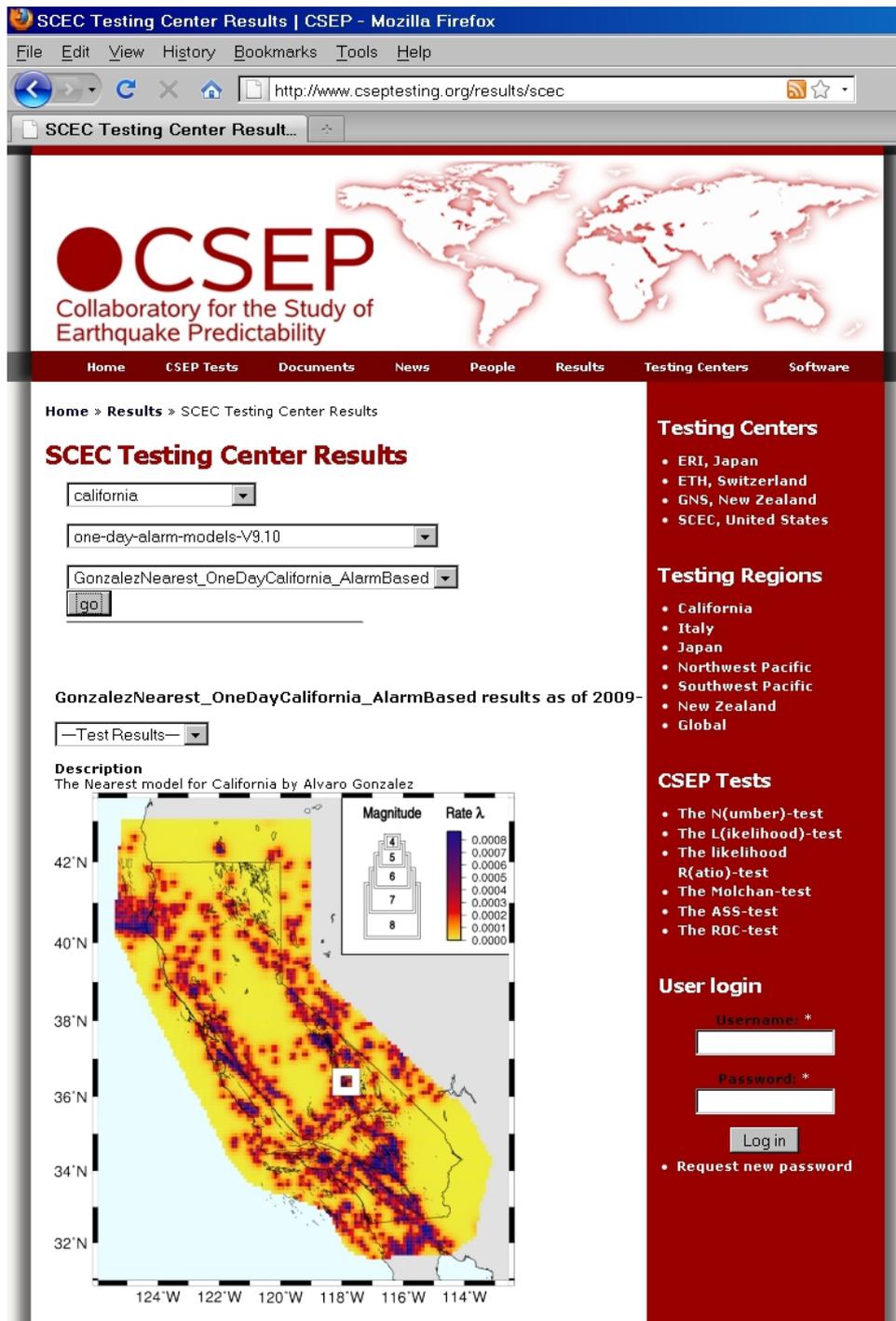


Figure 2.16: Snapshot of the CSEP web page showing a daily forecast map for California with the Nearest model. The map shown is for October 3rd, 2009. Overlapping white squares: epicentres of target earthquakes ($M \geq 3.95$) occurred (close to each other) that day and used for automatic testing of the model.

dimensions of the cell:

$$\tilde{P}_i = \iint_i \bar{P} da = \langle \bar{P} \rangle_i \times a_i, \quad (2.6)$$

where $\langle \bar{P} \rangle_i$ is the average value of \bar{P} in the cell (the empirical frequency of exceedance of nearest-neighbour distances, Sec. 2.2). With this definition, a cell tends to have a larger alarm function if it contains more past earthquakes, if it is close to them, and/or if it has larger geographic area. The maximum \tilde{P} in general will be <1 . The minimum \tilde{P} , for cells far away from previous earthquakes, may be $=0$. The binning into latitude-longitude cells thus imposes an unequal rasterization and local averaging of the map. Unlike the maps of the original method, these CSEP maps cannot self-sharpen beyond the resolution limit of the cell size.

2.5.3 Model testing

The procedure used to test alarm-based forecasts in CSEP does not depend on the absolute value of the alarm function, but on the relative one (Zechar & Jordan, 2008), so it can be rescaled for convenience. To merge the results of all days homogeneously, within each day, the value for each cell, \tilde{P}_i , in each day will be rescaled to \tilde{P}'_i by dividing by the largest \tilde{P} in the map, \tilde{P}_{max} :

$$\tilde{P}'_i = \frac{\tilde{P}_i}{\tilde{P}_{max}}. \quad (2.7)$$

This is analogous to a contrast stretch of the map image, so that the rescaled values range strictly up to 1.

CSEP routinely evaluates the performance of alarm-based models with three tests: the Molchan test, the Area Skill Score and the ROC test. The Molchan test (or error diagram, Molchan, 1997) is identical to the diagram used in the previous section (Figs. 2.15), but plotting in the ordinates the *error rate* (also called *failure rate* or *fraction of errors*), which equals one minus the hit rate. The Area Skill Score (Zechar & Jordan, 2010a) is derived from the Molchan diagram, and has been recently shown to provide inconsistent results (Molchan, 2012), so it will not be further discussed here. The ROC test (commonly used in other fields, such as atmospheric forecasting, Mason, 2003) plots the hit rate above an alarm function threshold versus the false alarm rate (percentage of cells above that threshold where no earthquake took place). The ROC results will be omitted here for brevity, since they are almost identical to those of the Molchan diagram (both diagrams would be identical in the continuum limit, with infinite cells in the map, each one with infinitesimal area, Zechar & Jordan, 2008).

The Molchan diagram for these models is calculated by considering all possible thresholds of the alarm function (Zechar & Jordan, 2008). For each threshold, it is

Testing region	M_{\min}^{target}	M_c	Maximum depth (km)	Cell side ($^\circ$)	Data since (year)	Start date (m/y)	Events
<i>California</i>	3.95	3.50	30	0.1	1984	10/2009	385
<i>Global</i>	5.95	5.95	30	1.0	1977	1/2010	563
<i>NW Pacific</i>	5.80	5.80	70	0.5	1977	1/2010	233
<i>SW Pacific</i>	5.80	5.80	70	0.5	1977	1/2010	349

Table 2.4: **Data ranges used for using and testing the forecast model in the CSEP experiment.** Cell sizes and maximum depth thresholds are fixed (pre-defined) by CSEP for all models used in each region. Data start to be used from a year when the catalogue is considered more reliable, and only earthquakes with magnitude $M \geq M_c$ are used, being M_c the magnitude of completeness. The automatic testing was done for the number of earthquakes listed in the last column, which had a minimum magnitude M_{\min}^{target} and occurred since the start date specified.

measured the fraction of geographic area covered by all cells whose function \tilde{P}' is at or above the threshold, and the fraction of the total number of target earthquakes which took place in them. In the cumulative results shown here, the fraction of area considered will be the average along all days of the testing period. A spatial forecast will be good if most earthquakes take place within the cells with largest alarm function, and if these cover a small fraction of the area. The results for the Molchan diagram will be shown together for all regions at the end of the chapter.

2.5.4 Testing regions

This section details the regions and data used for the tests, and shows some chosen examples of forecast maps. Since the method has been tested with several hundreds of events, it is impractical to discuss the results for each of them individually, so only a few selected cases will be briefly mentioned, and the major focus will be given to the overall results presented later.

Table 2.4 lists the data ranges used for calculating and testing the forecasts. In all regions, at the time of the beginning of the tests, the number of earthquakes used for calculating the first daily map was already >1000 , so it was expected that the distribution of nearest-neighbour distances was sufficiently constrained for applying the method reliably. The results presented here include the daily testing until September 30th, 2015.

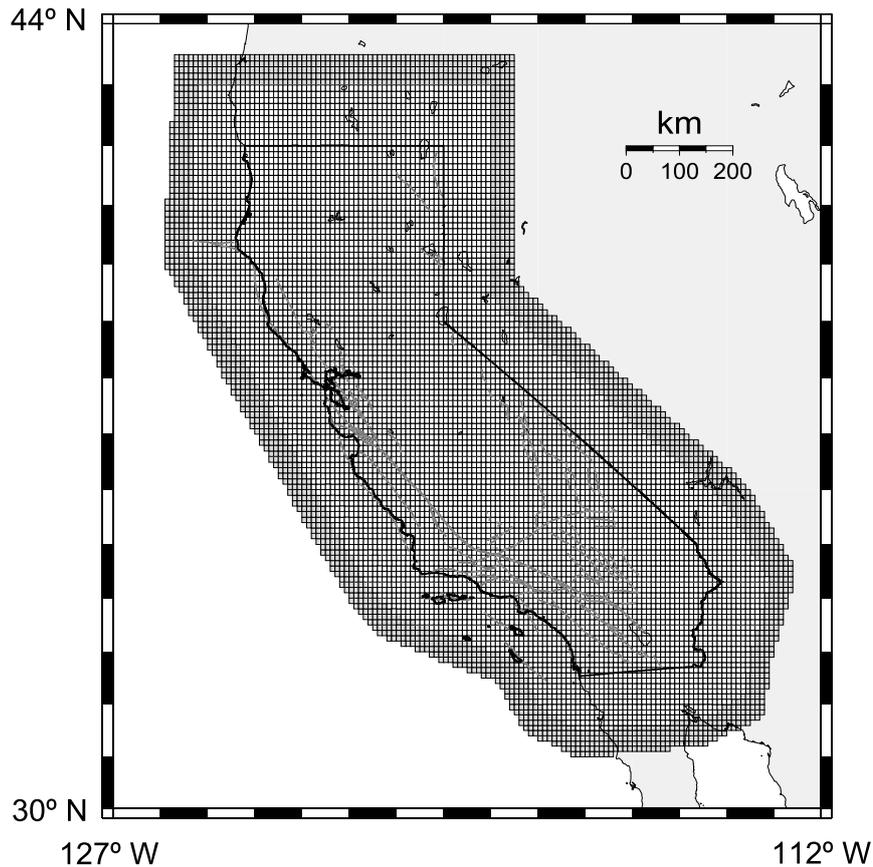


Figure 2.17: **Testing area and data collection area for real-time testing in California** (redrawn after Schorlemmer & Gerstenberger, 2007). The testing area is divided into the *white latitude-longitude cells*. The data collection area is conformed by the latter plus a buffer (*grey cells*) to reduce border effects. *Gray lines* are the traces of active faults within California (United States Geological Survey, 2015).

2.5.4.1 California

For California the experiment setup is the same as for the Regional Earthquake Likelihood Models experiment (Field, 2007; Schorlemmer & Gerstenberger, 2007), but with a lower magnitude threshold and over a different time period. The ANSS Comprehensive Earthquake Catalog is used (Advanced National Seismic System, 2015). Tests are done with earthquakes occurred in a specified testing area, which is surrounded by a buffer from which data is also used, to reduce border effects (Fig. 2.17, Schorlemmer & Gerstenberger, 2007). The start date for the data considered and the magnitude of completeness chosen (Table 2.4) are based on the the analysis by Wiemer & Schorlemmer (2007). This is the CSEP region where more earthquakes are used to calculate the maps with *Nearest* ($\sim 4,800$ earthquakes for the first one).

The Nearest model was submitted to CSEP for testing in California earlier than for other regions. It began formal testing for earthquakes that occurred since

October 1st, 2009 (Gonzalez, 2010b), and became the first alarm-based earthquake forecast ever tested on a daily basis.

During the six first years of experiment, 385 target earthquakes have taken place within the California region, according to the catalogue versions used during routine daily testing (Fig. 2.18). These are dominated by the Sierra El Mayor–Cucapah earthquake (Baja California, Mexico, April 4th, 2010, $M_w = 7.2$) and its prolific aftershock sequence (e.g. Hauksson *et al.*, 2011). This has been the largest earthquake in the region since Landers in 1992, making this a relatively anomalous period with high seismicity. Figure 2.19 shows the corresponding daily forecast map calculated by CSEP with the nearest method, with the location of this earthquake, and of the target aftershocks already occurred that day. The earthquake was preceded by foreshocks (Hauksson *et al.*, 2011), which contributed to enhance the calculated alarm function in and around the cells where eventually the mainshock took place.

This test has used so far over twelve times more target earthquakes than the RELM experiment, which was based on just 31 earthquakes with $M \geq 4.95$ occurred from January 1st, 2006, to December 31st, 2010 (Lee *et al.*, 2011; Zechar *et al.*, 2013).

2.5.4.2 Global (whole Earth)

For the global testing region, CSEP uses the Global Centroid Moment Tensor Catalogue (GCMT, formerly known as Harvard Moment Tensor Catalogue, Ekström & Nettles, 2015). The basic method used for compiling this catalogue was set out by Dziewonski *et al.* (1981) and its most recent implementation was described by Ekström *et al.* (2012).

For each earthquake, this catalogue provides, among other data, the scalar seismic moment and the coordinates of the centroid (not the hypocentre). The scalar seismic moment, M_0 , is defined as (e.g. Scholz, 2002):

$$M_0 = \mu A \bar{D}, \quad (2.8)$$

where μ is the shear modulus of the rocks involved in the earthquake rupture, A is the area of the rupture along the geologic fault where the earthquake occurred, and \bar{D} is the average displacement on A . M_0 has dimensions of energy. In the GCMT, M_0 is determined from the analysis of long-period and intermediate-period seismograms (Ekström *et al.*, 2012).

The centroid is the centre (best point-source representation) of seismic moment release. Centroids are typically not coincident with hypocentres in earthquake catalogues (e.g. Husen & Hardebeck, 2010). While the hypocentre indicates where an earthquake nucleated (is the point of initial radiation of seismic waves, and thus

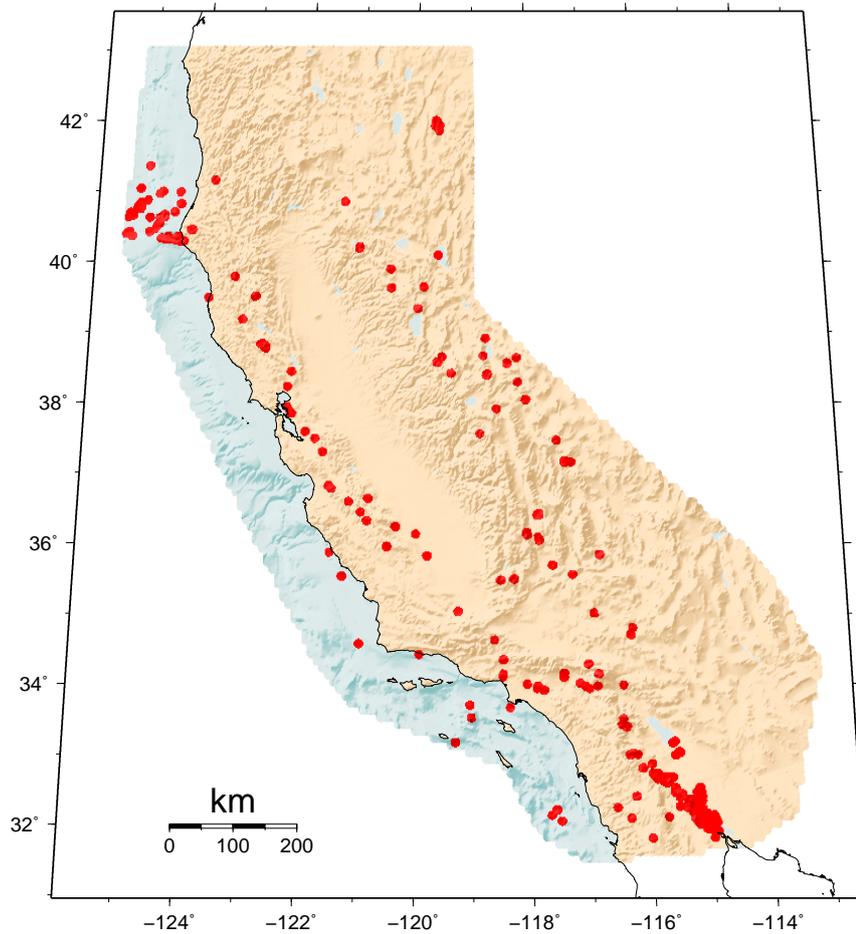


Figure 2.18: **Target earthquakes occurred during daily testing in California.** These are 385 earthquakes with $M \geq 3.95$ which took place during six years (October 1st, 2009 to September 30th, 2015). The topography of this and other maps in this chapter is derived from Amante & Eakins (2009). The $M_w = 7.2$ Sierra El Mayor–Cucapah earthquake and its foreshocks and aftershocks form a linear streak (marking the trends of mayor faults) near the southern boundary of the testing area.

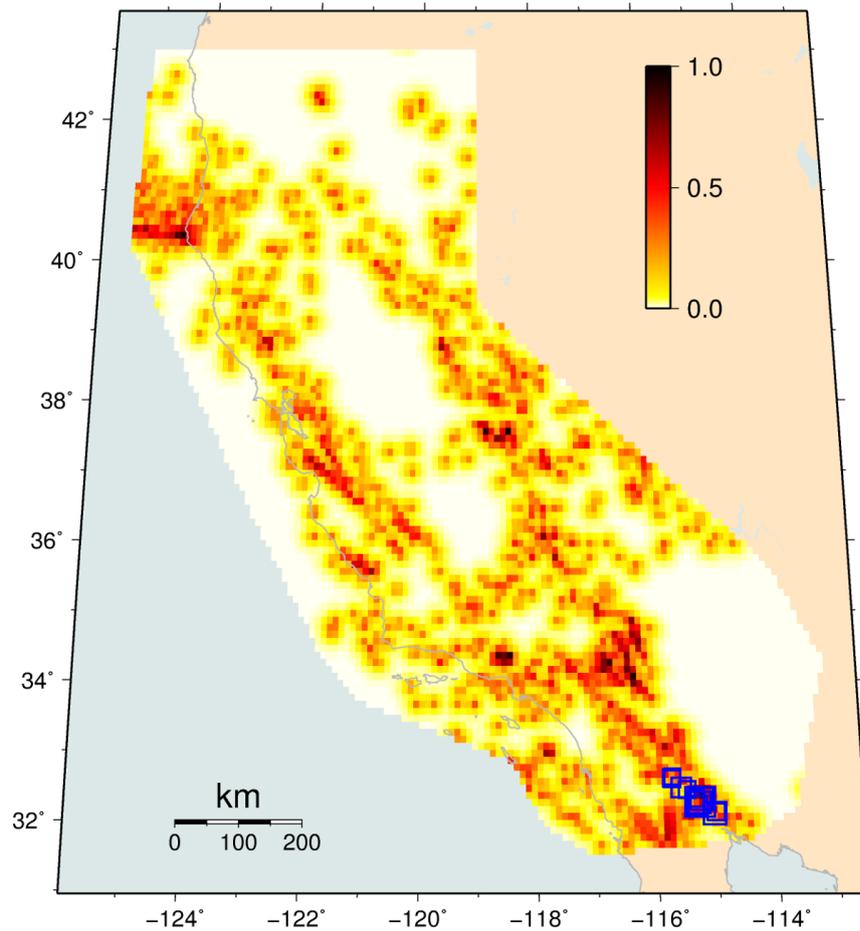


Figure 2.19: **Daily forecast map for California, April 4th, 2010, calculated automatically by CSEP.** *Colour scale:* Scaled, integrated probability of distance exceedance, (\tilde{P}') in each cell. It is higher at the latitude-longitude cells where earthquakes are considered more likely to take place. *Blue squares:* Epicentres of the target earthquakes occurred that day (symbol size proportional to magnitude). These are the $M_w = 7.2$ Sierra El Mayor–Cucapah earthquake and its first aftershocks. Only 50 of cells out of 7,682 of the testing region had equal or higher \tilde{P}' than the one ($\tilde{P}' = 0.604$) where the mainshock occurred.

the initiation point of dynamic rupture, Scholz, 2002), the centroid informs about which was the center of the final slip distribution (analogous to a centre of mass).

From the M_0 provided in the GCMT catalogue, for automatic selection of the data used for calculating the map (Table 2.4), moment magnitude (M_w) was calculated following its preferred definition (Hanks & Kanamori, 1979; USGS Earthquake Magnitude Working Group, 2002):

$$M_w = \frac{2}{3} \log M_0 - 10.7, \quad (2.9)$$

where M_0 is in dyne-cm (1 dyne-cm=10⁻⁷ Newton-metre).

An advantage of the GCMT catalogue (Kagan, 2003) is that all its earthquakes can be assigned a single M_w , systematically calculated and which does not saturate. Moreover, there is not a long latency period, since for each earthquake the catalogue provides a quick determination as soon as possible, and a final determination in no more than about four months (Ekström *et al.*, 2012). A disadvantage is that it is less complete (globally $M_c = 6.0$ Woessner & Wiemer, 2005) than the ISC Bulletin, and thus provides fewer usable earthquakes for meaningful testing.

During the testing period, 563 target earthquakes took place (Fig. 2.20), including several great ones, what makes this period more active than the long-term average. The largest ones are Tōhoku (Japan, March 11th, 2011, $M_w = 9.1$) and Maule (Chile, February 27th, 2010, $M_w = 8.8$), and will be dealt with in more detail below. Aftershocks of Tōhoku conform a significant fraction of the catalogue during these years. The most damaging event in terms of human losses, despite its lower magnitude, was the Haiti earthquake (January 12th, 2010, $M_w = 7.1$, Fig. 2.20).

2.5.4.3 Western Pacific

The western Pacific regions for CSEP testing are two partially overlapping geographic areas (NW and SW Pacific), specified by Eberhard *et al.* (2012) following Kagan & Jackson (1994), which account for more than half of Earth's recorded seismicity. The catalogue used is the same as for the global region, but the analysis is more detail (with lower target magnitude, finer gridding and greater maximum depth). The data period chosen and the M_c are based on the analysis of Woessner & Wiemer (2005). NW Pacific is the CSEP region where fewer earthquakes are used to calculate the maps with *Nearest* (~1,100 earthquakes for the first one).

NW Pacific

In the northwestern Pacific, 233 target earthquakes have taken place during the test, and the catalogue is largely dominated by the Tōhoku earthquake and its aftershocks (Fig. 2.21). The mainshock and most of its aftershocks took place

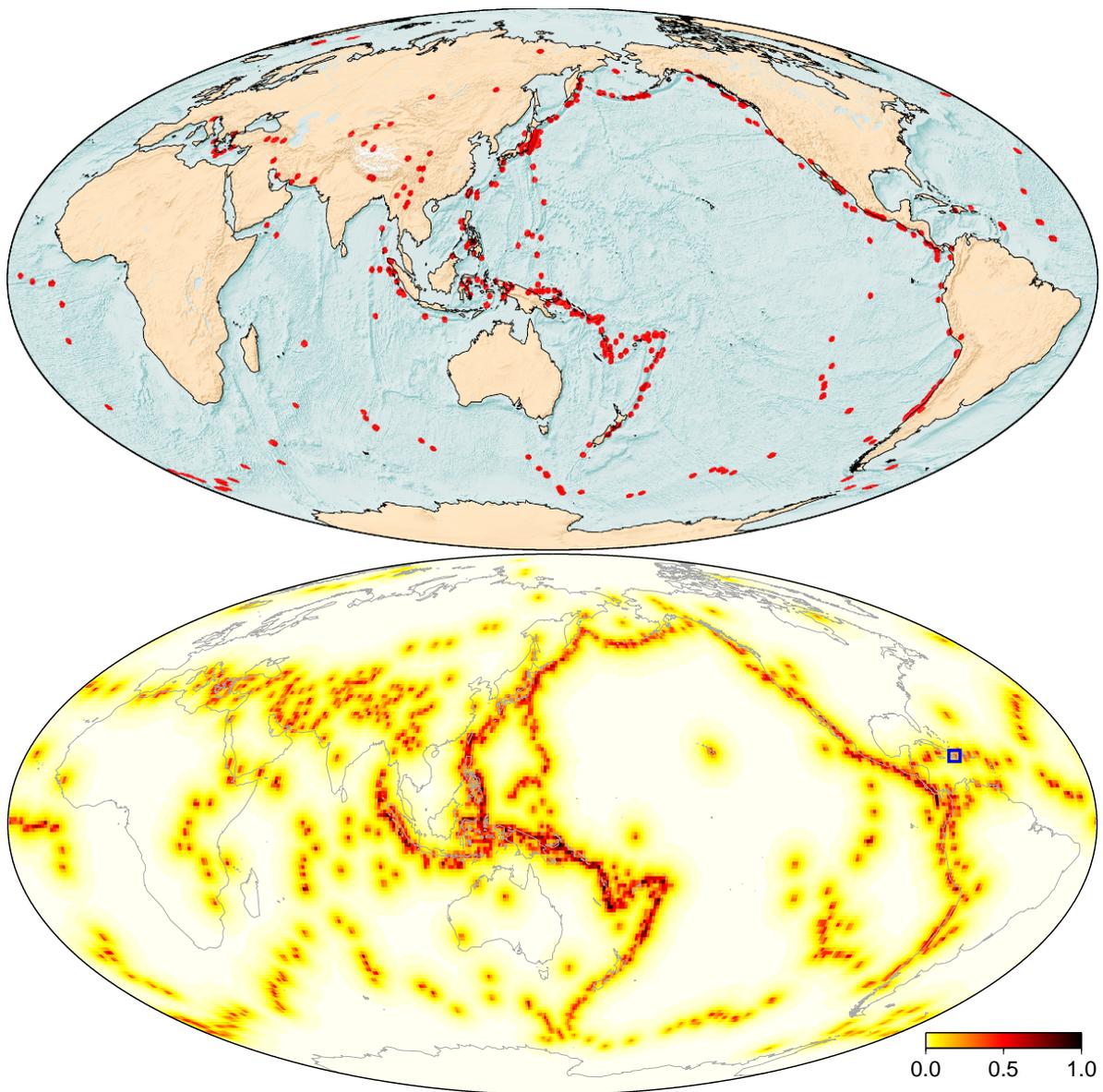


Figure 2.20: **Global daily forecast testing.** *Top:* Location of the 563 target earthquakes. They had $M_w \geq 5.95$ and took place during almost six years of daily testing (January 1st, 2010 to September 30th, 2015). *Bottom:* Daily forecast map for January 12th, 2010, with the location of the $M_w = 7.1$ Haiti earthquake marked. *Colour scale:* Scaled, integrated probability of distance exceedance (\bar{P}') in each cell. It is higher at the latitude-longitude cells where earthquakes are considered more likely to take place, in the vicinity of previous earthquakes.

in cells with moderate to high alarm function, among the largest in Japan and surroundings (e.g. Fig 2.21). Some cells towards the equator have greater alarm function, because their geographic area is much larger. Tōhoku was preceded by foreshocks (e.g. Marsan & Enescu, 2012), which contributed to enlarge this function locally. And aftershocks in the sequence rose the alarm function for those of later days, contributing to improve their spatial forecast. The calculated alarm function was high all along the subduction zone off the east coast of Japan, where the Tōhoku earthquake took place, in contrast to the official seismic hazard map of Japan, which assigned higher hazard elsewhere (Geller, 2011; Stein *et al.*, 2012).

The best forecast was obtained on August 3rd, 2010, for a target earthquake (with $M_w = 6.3$) which took place at the cell with the highest value of the function for that day (the one centred at 126.5°E, 1.5°N, under the the Molucca Sea).

SW Pacific

In the southwestern Pacific, 349 target earthquakes have taken place during the test (Fig. 2.22). Most of them concentrated along subduction zones, where previous ones had taken place before and the alarm function was high. The deadliest earthquake in the region during these years was that of Christchurch, New Zealand (February 21st, 2011, $M_w = 6.1$, Fig. 2.22). This was an aftershock of the larger but less damaging $M_w = 7.0$ Darfield earthquake, occurred nearby on September 3rd, 2010 (dates according to Universal Time). This is an example in which the mainshock caused the alarm function to significantly rise locally, eventually contributing to improve the spatial forecast of a damaging aftershock.

The best forecasts in this region were obtained for four target earthquakes which occurred at the cell with the highest value of the function in their respective days. The largest earthquake, with $M_w = 7.5$, took place on May 5th, 2015, SSW of Kokopo, Papua New Guinea (at the cell centred at 152.0°E, 5.5°S). The three other events were foreshocks of this one, and had occurred in the immediate surroundings (exactly at the same cell) a few days earlier (April 30th, and May 1st and 3rd, with M_w equal 6.6, 6.8 and 5.9, respectively).

2.5.5 Overall results and comparison with a reference model

In all regions, the vast majority of earthquakes took place in the small fraction of the area occupied by the cells with higher alarm function. The Molchan diagram with the overall results for all regions is shown in Fig. 2.23. The best results (not significantly different from each other) are obtained for the western Pacific and global regions, and are similar to those obtained for the global retrospective tests

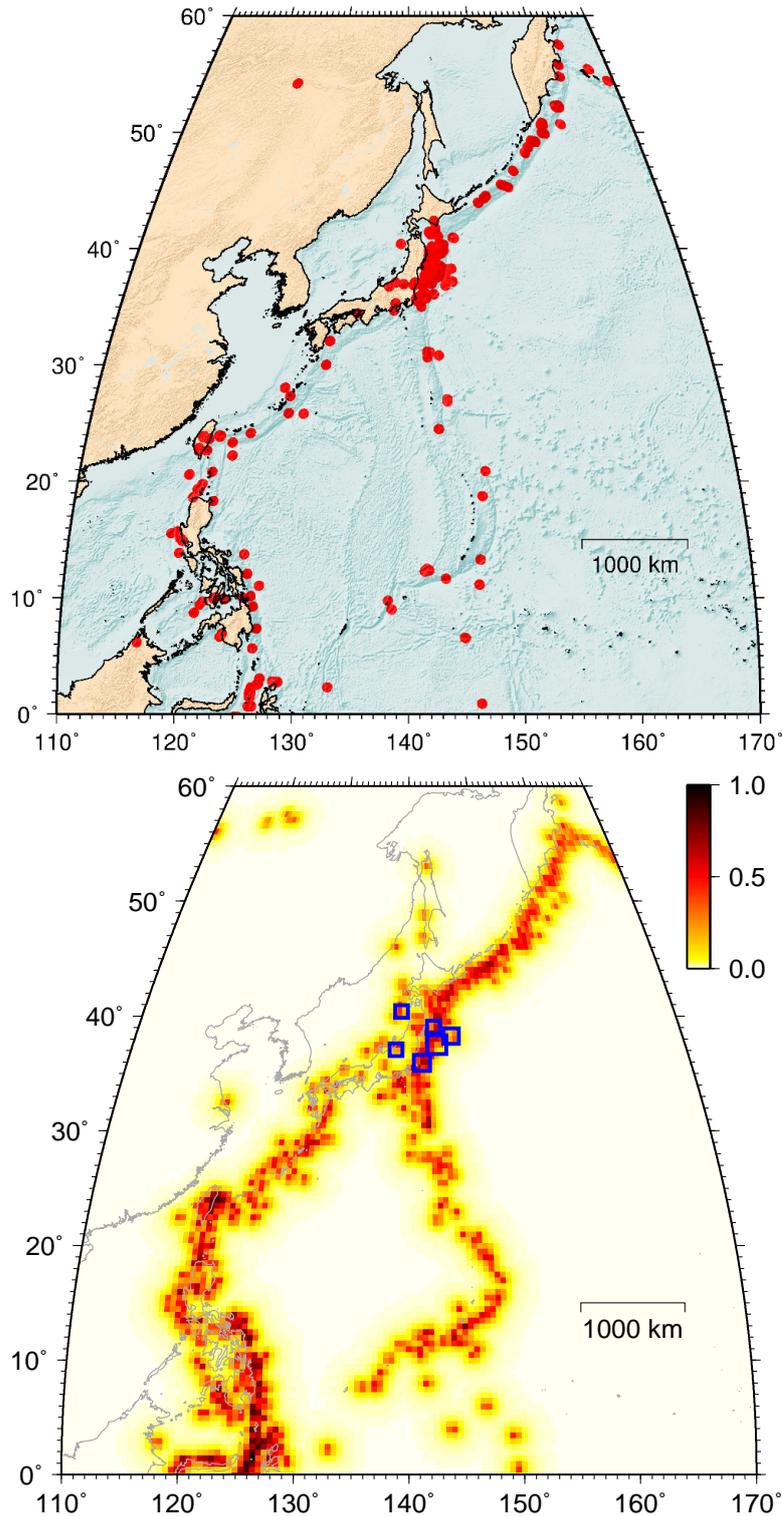


Figure 2.21: **Daily forecast testing in the NW Pacific.** *Top:* Locations of the 233 target earthquakes. They had $M \geq 5.8$ and took place during almost six years of daily testing (January 1st, 2010 to September 30th, 2015). Most of those offshore the east coast of Japan correspond to the 2011 Tōhoku earthquake sequence. *Bottom:* Daily forecast map for **March 11th, 2011.** *Blue squares:* Target earthquakes (symbol size proportional to magnitude), which are the Tōhoku earthquake ($M_w = 9.1$) and its largest aftershocks of that day. *Colour scale:* Scaled, integrated probability of distance exceedance (\bar{P}') in each cell.

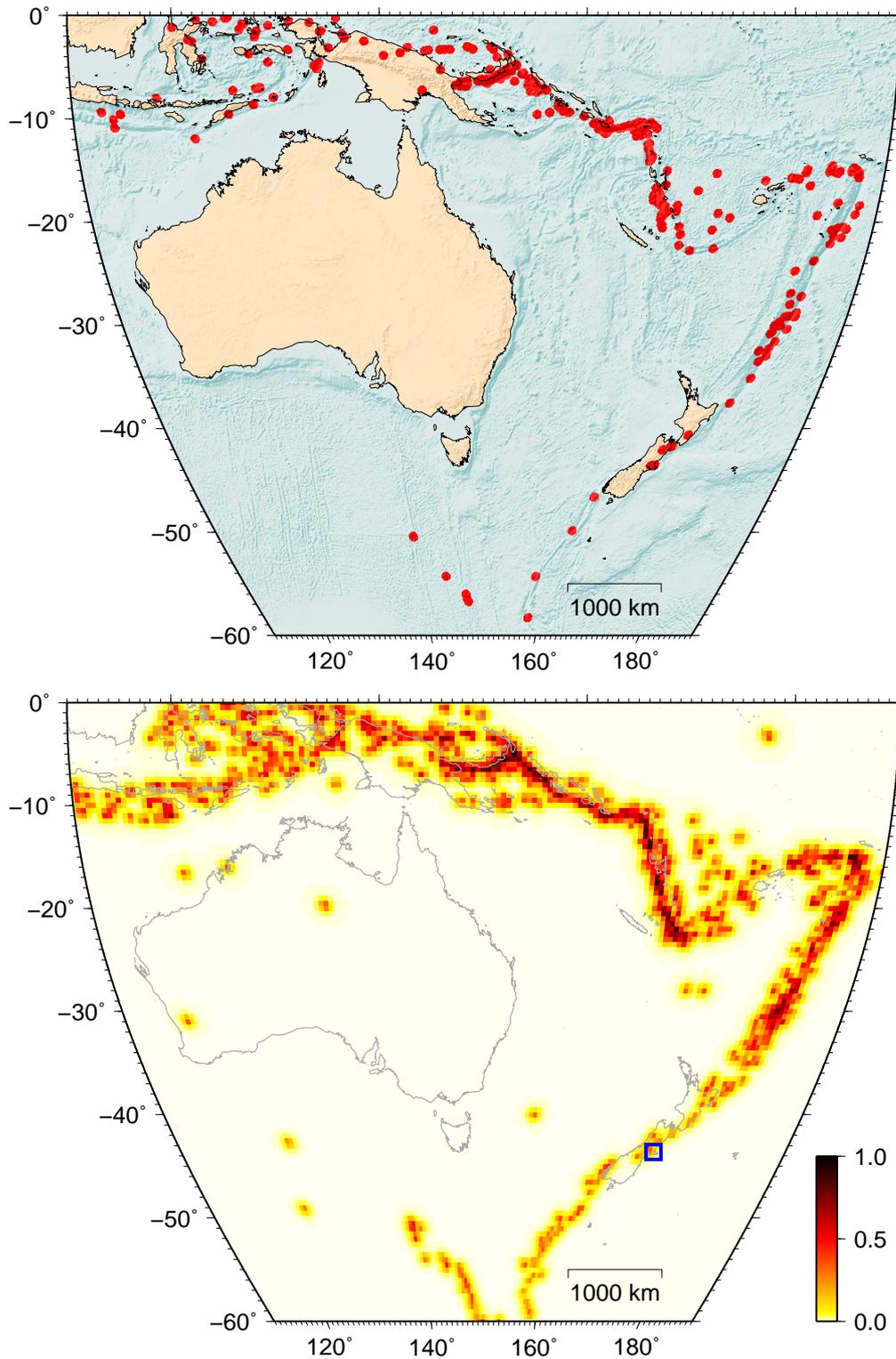


Figure 2.22: **Daily forecast testing in the SW Pacific.** *Top:* Locations of the 349 target earthquakes. They had $M \geq 5.8$ and took place during almost six years of daily testing (January 1st, 2010 to September 30th, 2015). *Bottom:* Daily forecast map for February 21st, 2011. Blue square: Location of the Christchurch, New Zealand earthquake ($M_w = 6.1$), the only target earthquake occurred that day. Colour scale: Scaled, integrated probability of distance exceedance (\bar{P}^i) in each cell.

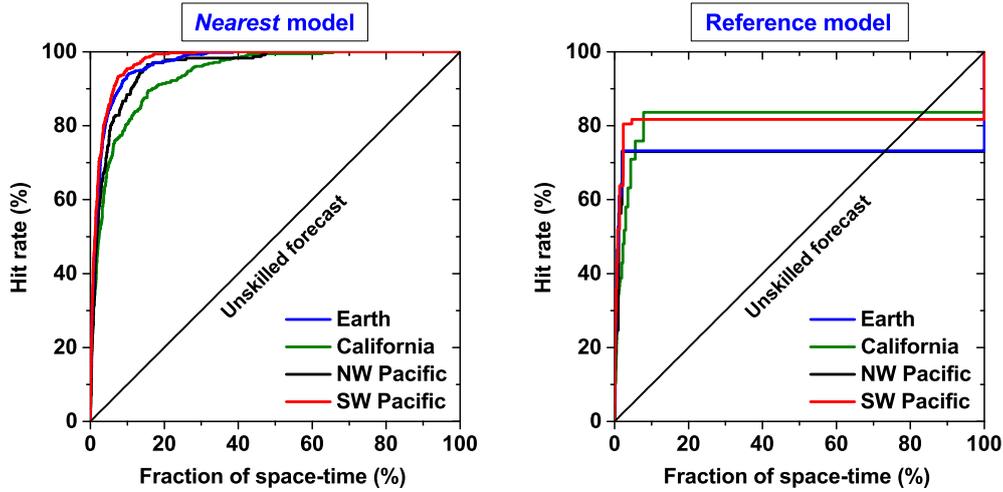


Figure 2.23: Overall results of the CSEP daily tests and a retrospective reference forecast. The results of the reference model for the whole Earth and NW Pacific mostly overlap each other. This model performs very well for the majority of earthquakes in each region, but fails completely for the remaining ones, which occurred at locations where the alarm function was null. *Nearest* outperforms this reference model in all regions.

for $M \geq 5.0$ in the previous section. The results for California are slightly worse, and similar to the retrospective ones for Southern California with $M \geq 3.0$. In all regions, a few earthquakes took place relatively far away from previous ones, so to achieve a hit rate of 100% it is necessary to cover a large fraction of the total area (total space-time, given that area is averaged for all the days of the test). No earthquake during the tests took place in a completely unexpected site (in the cells with $\tilde{P} = 0$, very far away from previous ones).

These overall results are almost insensitive to the area weighting used for each cell. That is, if instead of Eq. 2.6, the alarm function is (retrospectively) recalculated as:

$$\tilde{P}_i^* = \langle \tilde{P} \rangle_i, \quad (2.10)$$

the trajectories of the Molchan diagram are almost identical to those of Fig. 2.23 in all regions, at least for the target earthquakes tested so far.

2.5.5.1 Reference model

Dividing the map into standardized spatial bins allows for formal testing against other forecasts, using (retrospectively) the same earthquake catalogue for each region as in the CSEP test. In particular, the simplest reference forecast model is the so-called *Relative Intensity* (not to be confused with shaking intensity), proposed by Holliday *et al.* (2005) and later frequently used (e.g. Zechar & Jordan, 2008; Nanjo, 2010, 2011; Tiampo & Shcherbakov, 2012, and references therein). It just assumes

that future earthquakes will be more likely in the cells where more earthquakes (with magnitude above the completeness threshold) took place in the past. Despite its simplicity, it is able to outperform the forecasts of other models (Zechar & Jordan, 2008).

This reference method has no parameter, aside from the size of the spatial bins. Its alarm function RI , is simply defined as (Holliday *et al.*, 2005):

$$RI_i = \frac{n_i}{n_{\max}}, \quad (2.11)$$

where n_i is the number of past earthquakes within the considered ranges of date, magnitude and depth which have originated inside the cell i , and n_{\max} is the largest number of such earthquakes occurred in any cell of the map. The division by this denominator yields a maximum $RI^{\max} = 1$, and is analogous to the re-scaling used with *Nearest*. It allows taking into account that as time progresses, there will be more past earthquakes to build the map. The geographic area of the cell is not explicitly taken into account, but larger cells obviously tend to have more earthquakes.

For many cells, $RI = 0$, even if they are adjacent or close to others with high RI . This depends on the lattice spacing used. If the cell size were reduced, each cell would tend to contain either only one past earthquake or none. Smoothed versions of this method have been proposed to avoid these strictly null values (Nanjo, 2010, 2011; Tiampo & Shcherbakov, 2012, and references therein). But, as discussed at the beginning of the chapter, it is unclear which smoothing procedure is more adequate (e.g. Nanjo, 2010, 2011) and such a practice inevitably introduces at least one parameter, with is absent in the original version (Holliday *et al.*, 2005). It should be noted that the idea of counting earthquakes in a latitude-longitude grid as a representation of the geographic distribution of seismicity is actually much older (e.g. Rey Pastor, 1934).

Figure 2.24 compares the maps produced by the non-parametric *Nearest* and *Relative Intensity* methods, for the day when the second largest earthquake on Earth during the test period took place: $M_w = 8.8$ Maule (Chile). The last comparable earthquake at this location was the 1835 Concepción earthquake, observed and documented by FitzRoy (1839) and Darwin (1851) among others, and it had been measured that strain was accumulating in the area (e.g. Ruegg *et al.*, 2009). This earthquake, thus, filled a long-recognized seismic gap (e.g. Melnick *et al.*, 2012).

The Maule earthquake serves as an example of the performance of the methods when an earthquake does not is not strictly close to the previous ones from the instrumental period considered (the last decades). It took place inside a cell where no earthquake with $M \geq 6$ had taken place since 1977, so it is a clear miss for the *Relative Intensity* method. *Nearest*, instead, calculated a moderate alarm function for

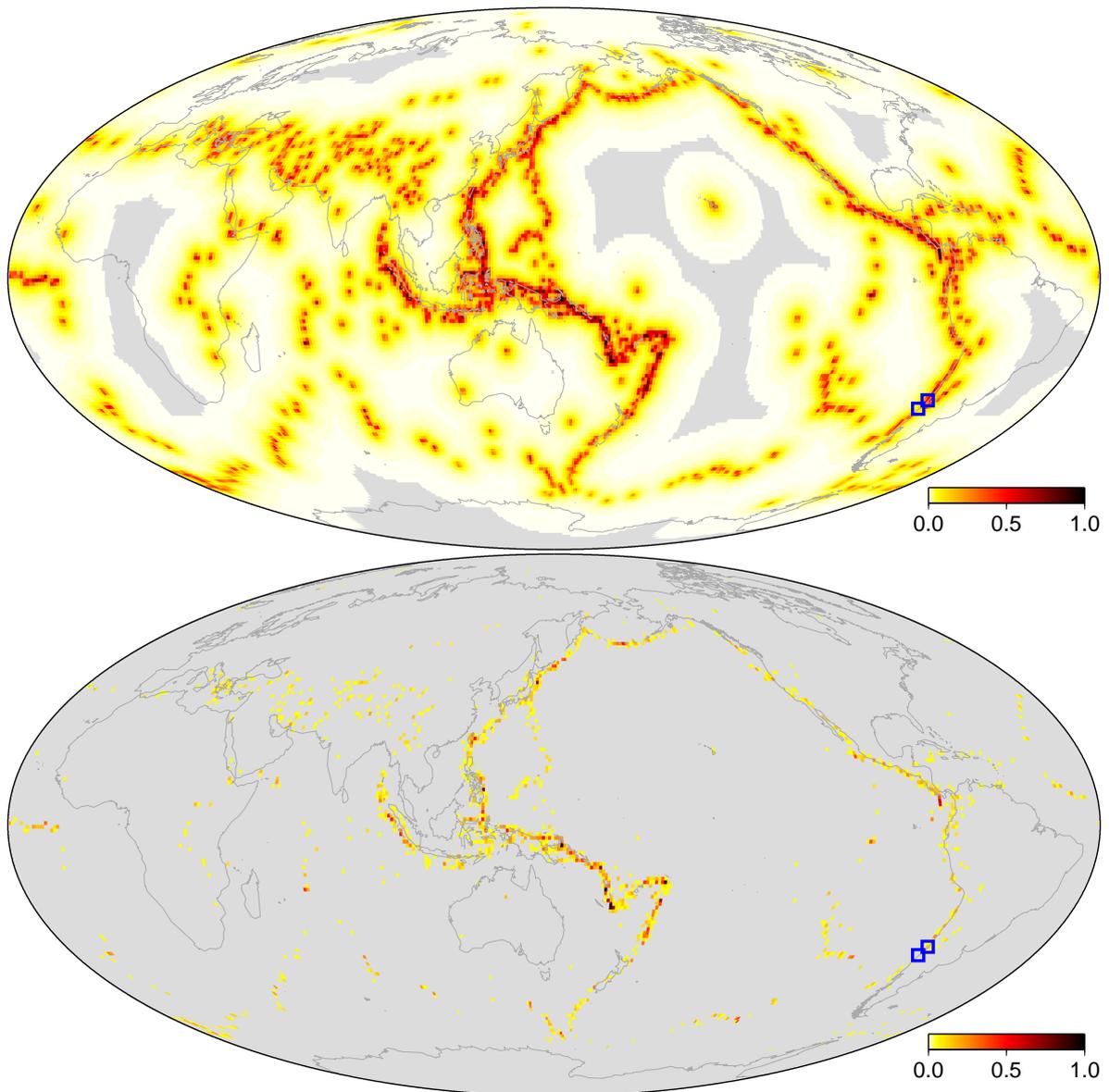


Figure 2.24: **Comparison of global forecasts for the day of the M_w 8.8 Maule earthquake (February 27th, 2010).** *Top:* Daily map calculated by CSEP with the Nearest method. *Bottom:* Retrospective map calculated with a reference method whose alarm function is proportional to the number of past earthquakes in the cell. *Colour scale:* Rescaled alarm functions (light grey where strictly zero). *Blue squares:* Location of the mainshock (the southernmost event) and the only other earthquake with $M \geq 5.95$ of that day (an aftershock). The aftershock location is well forecast by both methods, because of its proximity to earlier earthquakes. The mainshock filled a seismic gap, where Nearest assigned a moderate alarm function, while the reference model fails because it assigned strictly zero.

earthquakes with $M \geq 6$ at this cell, because previous earthquakes had occurred nearby.

Actually, in all regions, *Nearest* outperforms the reference model of *Relative Intensity*, because a large fraction of earthquakes take place in cells where no previous earthquakes had taken place (in the time, depth and magnitude ranges considered). The diagram of Fig. 2.23 shows that most earthquakes took place in the few cells where previous ones had occurred. But the remaining earthquakes ($\sim 27\%$ in the global and NW Pacific regions, $\sim 18\%$ in the SW Pacific and $\sim 16\%$ in California) are completely missed: they took place in the remaining cells, where $RI = 0$, requiring a 100% of space-time to be covered for forecasting their locations. So the main achievement of *Nearest* is to naturally assign a moderate to low relative probability to those cells, without using any parameter.

2.6 Final remarks

This chapter has presented a method to forecast the location of future earthquakes using only the location of past ones. It lacks any parameter, and it is only based on the empirical distribution of distances between past earthquakes. It is shown that this distribution can be extrapolated to calculate, in a probabilistic way, at which distance from previous earthquakes the next one will occur. Earthquakes tend to occur near past ones, and this method offers an empirical, simple way to express this fact quantitatively.

The skill of the method is shown by extensive testing, both retrospectively and prospectively, with different earthquake catalogues, for regions with different tectonic environments. In the retrospective tests, the performance of the method is tested with every new earthquake. This contrasts with the traditional tests for earthquake forecast, which typically separate the available data in two periods: an initial *training, period* whose data are used to calculate the forecast, and a later *testing period*, whose data is used to evaluate the performance of the forecast (e.g. Stock & Smith, 2002; Zechar & Jordan, 2008; Kafka & Ebel, 2011). The skill of the method tends to improve with time, as new earthquakes occur and the probability maps sharpen.

Similar results are obtained by daily, automatic, independent testing by the Collaboratory for the Study of Earthquake Predictability during six years. Testing with every new earthquake is not currently technically feasible, but it is shown that the daily updates of the maps allow improving them progressively, for example, in aftershock sequences and when foreshocks precede a larger earthquake.

The fact that the observed hit rate is so similar to the expected one implies that the method is near optimal if only the spatial distribution of earthquakes is used as input. So the method automatically delimits the smallest distances (and the smallest areas) around past earthquakes which yield a given hit rate in the long term.

Further reducing the area where the next earthquake is forecast to occur would require additional information, not used here. Improvements of this empirical method would require, for example, taking into account the temporal dependencies between earthquakes. Examples are perhaps more evident in the maps for Southern California, where aftershock sequences leave a permanent footprint, despite the probability of a next earthquakes is nowadays probably low. Refinements could be sought by disregarding earthquakes occurred too long ago, but this would require optimizing additional parameters, such as the maximum number of previous earthquakes considered to elaborate the map.

The method shows clearly a trade-off in which, to forecast the next earthquakes with larger probability, larger areas around previous earthquakes have to be considered. This trade-off is typical to any forecasting approach, and it has to be taken into account when using the forecasts for decision-making (e.g. Molchan, 1997; Mason, 2003; Stein & Stein, 2014).

Relevant consequences of these results are summarized as follows:

- This method without any parameter outperforms the one commonly used as reference (e.g. Holliday *et al.*, 2005; Zechar & Jordan, 2008), which also lacks any parameter. Thus the new method is proposed as an improved reference model to which compare the merits of more complex forecasts.
- Improved maps of spatial earthquake probability could be based on the empirical distribution of distances between earthquakes. For calculating a continuous estimate of probability density, the function used as smoothing kernel should be consistent with such empirical distribution.
- In order to forecast reliably future earthquake locations, a minimum sample of past earthquakes is necessary, on the order of 1000. The locations of only the few largest earthquakes (which frequently influence hazard maps heavily, Stein *et al.*, 2012) are not sufficient to calculate a reliable spatial probability for future earthquakes.
- The spatial component of earthquake forecast (and hazard) maps will likely benefit from using the locations of all past earthquakes for which the catalogue is complete, without declustering, in order to obtain a more complete and reliable spatial sample.

- In regions with few earthquakes in the available instrumental record (such as in stable continental interiors), retrospective tests with this method could help elucidate if the earthquake sample is large enough to justify using spatial kernels, for example for probabilistic seismic hazard assessment.

As a final note, this method can be applied to forecasting locations of any set of points which distribute in space in a persistent way. It thus could be tested with other natural events whose locations can be considered point-like, such as the opening of new sinkholes (e.g. Galve *et al.*, 2011) or volcanic vents (e.g. Becerril *et al.*, 2013). The application to earthquake catalogues constitutes a particularly complete case study because of the abundance of events available for testing.

The Spanish National Earthquake Catalogue: Evolution, precision and completeness

La Géographie sismologique nécessite comme fondement principal, la connaissance des données proportionnés par la Sismologie statis[tique] [...].

Alfonso Rey Pastor (1927, p. 1)

Chapter adapted and expanded from:

Álvaro González (2016)

The Spanish National Earthquake Catalogue:

Evolution, precision and completeness.

Journal of Seismology, in press.

3.1 Introduction

The capabilities for detecting and characterizing earthquakes in any region are limited and evolve throughout time. Consequently, the resulting earthquake catalogue is far from being obtained under controlled, laboratory-like conditions, and is prone to large temporal and spatial heterogeneities. Modifications of the seismological instrumentation, of the earthquake location routines, or of the procedures for calculating magnitudes will leave their mark on the catalogue. It is of utmost importance to decipher and unveil these heterogeneities and limitations, because, if unnoticed or disregarded, they may invalidate the analyses

of earthquake data and hamper reaching meaningful conclusions (e.g. Muñoz & Udías, 1982; Vere-Jones, 1992; Kagan, 2003).

This chapter provides an extensive review and statistical characterization of the Spanish National Earthquake Catalogue during the instrumental era, pointing out its heterogeneities and progressive improvements. The catalogue is arguably the most complete for Spain and has been used as a basis for the official seismic hazard assessments of the country (e.g. Ministerio de Fomento, 2003; García-Mayordomo *et al.*, 2004; Martínez Solares *et al.*, 2013). The analysis presented here enabled a proper statistical use of the catalogue in Chapter 2.

The Spanish National Earthquake Catalogue is compiled by the *Instituto Geográfico Nacional* (IGN, the agency responsible for earthquake monitoring and warning at a national level in Spain). It can be freely downloaded online (Instituto Geográfico Nacional, 2015). Figure 3.1 shows the epicentres and the spatial extent of the catalogue (26°N to 45°N, 20°W to 6°E). This region embraces mainland Spain, Portugal, Morocco, Andorra, parts of France and Algeria, and several Portuguese and Spanish archipelagos (Balearic, Canary, Madeira). The catalogue contains data of over 100,000 events, originated since 880 B.C. until present, most of them in recent decades, with a yearly record of $\sim 17,000$ in 2011.

At one extreme, this region has experienced the greatest earthquakes known in western Europe. The largest events are the Lisbon Earthquake on 1 November 1755 [moment magnitude $M_w \simeq 8.5$ (Martínez Solares & Mezcua, 2002; Martínez-Solares & López Arroyo, 2004)], the 25 November 1941 North Atlantic Earthquake [$M_w \simeq 8.0$ (Storchak *et al.*, 2013)], the 28 February 1969 offshore Portugal earthquake [$M_w = 7.8$ (Storchak *et al.*, 2013)] and the enigmatic, deep earthquake of Granada (south Spain) on 31 March 1954 [$M_w = 7.8$ (Storchak *et al.*, 2013)]. At the other extreme, earthquakes with magnitudes ≤ 1 are currently reported in the IGN catalogue in the geographic areas with the densest distribution of stations.

The Spanish National Seismic Network (the largest permanent network in the country, managed by IGN) provides the primary data source for the real-time compilation of the catalogue. This is expanded by IGN with data shared by seismic stations of other institutions in Spain and abroad, either in real time, or for the final catalogue review. These additional stations belong, for example, to the regional, permanent Spanish networks in Andalusia (IAG, 2015) and Catalonia (ICGC, 2015), those of the San Fernando and Ebro observatories, and the national seismic networks of France (RéNaSS, 2015), Portugal (Custódio *et al.*, 2015), Morocco and Algeria (Yelles-Chaouche *et al.*, 2013). More information on permanent networks on the Iberian region can be found in Alguacil & Dávila (2003).

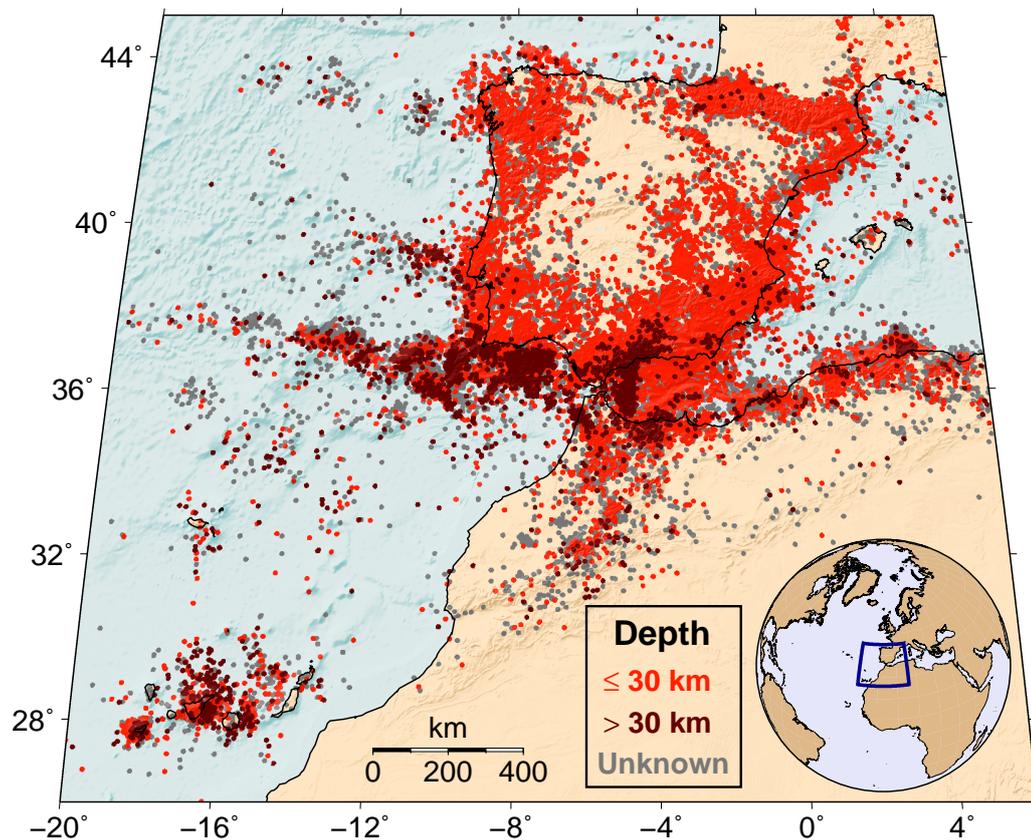


Figure 3.1: Epicentres in the Spanish National Earthquake Catalogue since the historical era to the end of 2013, separated according to hypocentral depth. Map in a sinusoidal (equal area) projection, over the global relief model of Amante & Eakins (2009), replaced by the bathymetry of Zitellini *et al.* (2009) where available. The infrequent earthquakes with depth > 30 km are highlighted by being plotted on top of the others. Depth is considered unknown if absent (e.g. for events located from intensity data only) or fixed (not calculated as an independent parameter).

Conversely, this catalogue has been used as a source for others. For example for Portugal (Carrilho *et al.*, 2004; Pena *et al.*, 2014), Pyrenees (Souriau & Pauchet, 1998; Secanell *et al.*, 2008), Catalonia (Susagna & Goula, 1999), Algeria and Morocco (e.g. Benouar, 1994; Peláez *et al.*, 2007; Hamdache *et al.*, 2010; Ayadi & Bezzeghoud, 2015; Harbi *et al.*, 2015), Euro-Mediterranean region (Grünthal & Wahlström, 2012), and the Bulletin of the International Seismological Centre (International Seismological Centre, 2015b). IGN is a nodal centre of the Euro-Mediterranean Seismological Centre and contributes data for its bulletin (Godey *et al.*, 2006). So analysing the IGN catalogue may eventually contribute to characterizing these separate, but intertwined, databases.

The chapter is structured as follows. Section 3.2 reviews the evolution of the catalogue and the seismic network, pointing at milestones in the catalogue development. The rest of the chapter focuses mainly on the instrumental era, the last decades on which the catalogue relies mostly on instrumental seismic data. Section 3.3 reviews the procedures used for locating earthquakes and assesses the typical precision of hypocentral locations. The different magnitude scales currently reported in the catalogue are described in section 3.4. Section 3.5 analyses and maps in detail the magnitude of completeness of the catalogue within different periods, separated by milestones in the instrumentation development, changes of location routines or of magnitude scales used. Closely related to the magnitude of completeness of the catalogue are the apparent daily and weekly variations, related to the variable amplitude of the artificial seismic noise, which are highlighted in section 3.6. The contamination of the catalogue by quarry and mine blasts is reviewed, and exemplified by several case studies, in section 3.7. Concluding remarks are discussed in the last section. I will use the catalogue data until the end of 2013 (~ 98,000 events in total), the last year for which it has been completely revised.

3.2 Evolution of the catalogue

Focusing on the main data source, the current catalogue can be divided into three blocks :

- Before 1901, it is mostly based (with later updates) on the catalogue by Martínez Solares & Mezcua (2002) for the Iberian Peninsula and surroundings, and on the earlier one by Mezcua & Martínez-Solares (1983) for the Canary Islands.

- From 1901 to 1980, it is taken from Mezcua & Martínez-Solares (1983), with later revisions.
- Since 1981, it has been directly compiled by IGN using data provided by the Spanish National Seismic Network (*Red Sísmica Nacional*), and seismological stations of other institutions.

The cumulative number of earthquakes in the catalogue (Fig. 3.2) shows an apparent increase of the earthquake rate with time (especially for magnitudes ≤ 3), which evidences a progressive improvement of detection capabilities. In the Canary Islands since mid-2011 the increase is mostly natural, due to the surge of volcanic swarms in El Hierro island.

The following subsections review the development of the catalogue and of its data sources, including network instrumentation.

3.2.1 Historical era (until 1900)

The most damaging earthquakes in Spain occurred in the historical era, until the late XIX century, for which the catalogue is based on chronicles that describe the effects of the earthquakes. This historical catalogue relies on primary documental sources and on earlier compilations, as discussed by Muñoz & Udías (1982), Roca *et al.* (2004), Rodríguez de la Torre (2004) and Udías (2015). Major contributions were the databases by Sánchez Navarro-Neumann (1917, 1921); Galbis (1932, 1940); Munuera Quiñonero (1963); Fontserè & Iglésies (1971); Mezcua (1982) and Rodríguez de la Torre (1990, 1993, 1997, 2004). The first catalogue entry dates from 880 B.C., although the first Iberian earthquake for which the exact date is known took place on 21 May 881 (Udías, 2015). The first event for which the catalogue details both the location and maximum intensity dates from 1048.

A digital database of macroseismic information for earthquakes until 1900 (Roca *et al.*, 2011) is available on the IGN webpage (Instituto Geográfico Nacional, 2015). Historical and early instrumental earthquakes in Spain and surroundings are being continually subject to revisions (e.g. Samardjieva *et al.*, 1999; Mezcua *et al.*, 2004, 2013; Udías, 2015, and references therein). The first systematic catalogue compiling the geological and archeological effects of earthquakes in Spain has been published by Silva *et al.* (2014).

The 1755 Lisbon earthquake, and the 1884 Andalusian earthquake prompted developments of the study of earthquakes in Spain (Udías, 2013). The latter eventually motivated the creation of the Spanish Seismological Service years after (Anduaga, 2004).

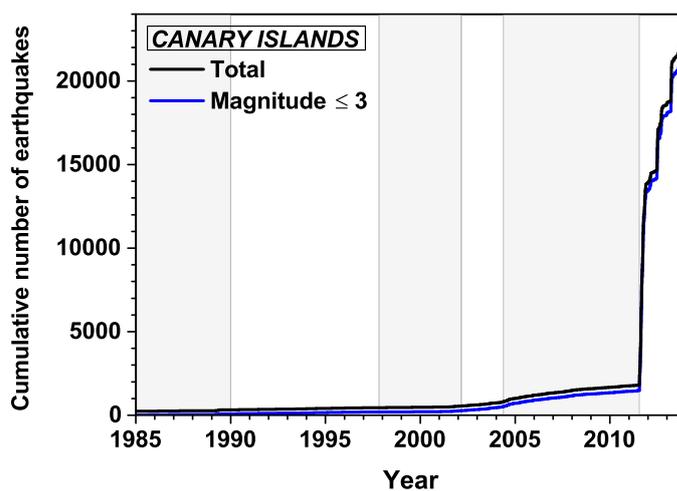
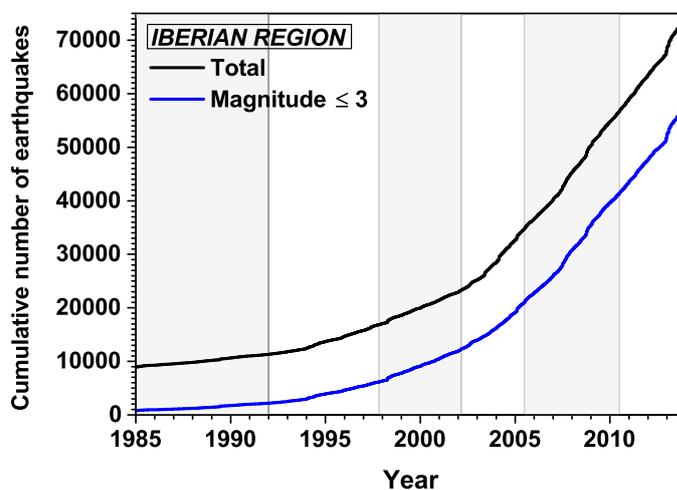
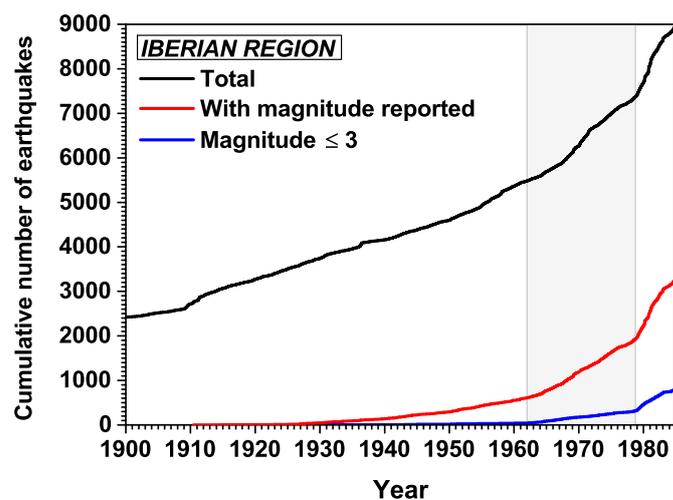


Figure 3.2: **Cumulative number of earthquakes versus time**, in the geographic regions analysed (maps of Figs. 3.12 and 3.14). Magnitude is not reported for earthquakes occurred before 1910, while it is reported for the vast majority since 1985 (not shown separately). *Vertical fringes* indicate the periods considered for completeness assessment.

3.2.2 First observatories (1897–1924)

Instrumental recording in Spain and Portugal began in the late 19th century, although macroseismic observations would remain the chief data source for the catalogue until the beginning of the strictly instrumental era (Sec. 3.2.6). A catalogue of the early Spanish seismographs was compiled by Batlló & Borrmann (2000) and Batlló (2004). The first instrumental observations in peninsular Spain were made in 1885 (Martínez Solares, 1981; Batlló, 2004; Udías, 2013), and in the Canary Islands in 1890 (von Rebeur-Paschwitz, 1892a,b; Milne, 1895; Batlló, 2004; Fréchet & Rivera, 2012). In Portugal, the first station was installed in Coimbra; it began experimental recording in 1891 and continuous recording in 1903 (Custódio *et al.*, 2012). The first continuous recordings in Spain were made with two Milne seismographs, one installed in 1897 in the Observatory of San Fernando (of the Spanish Royal Navy; Martín Dávila *et al.*, 2006), and the next one in 1898 in the mines of Río Tinto (Udías, 2013).

The next two Spanish observatories were founded by the Jesuits (Cartuja, in 1902 and Ebro, in 1904; Udías & Stauder, 1996; Anduaga, 2004; Ammon *et al.*, 2010), and were followed in 1906 by the Fabra Observatory of the *Academia de Ciencias y Artes de Barcelona*.

The first seismic observatory of the *Instituto Geográfico* (former name of the IGN) was set up in Toledo, near Madrid, in 1909 (Anduaga, 2004; Udías, 2013). It was crucial for recording early instrumental earthquakes in the Iberian region, because of its relatively advanced and well maintained instrumentation (e.g Rey Pastor, 1929; de Miguel y González Miranda, 1961; Samardjieva *et al.*, 1997, 1998; Payo Subiza & Gómez-Menor, 1998; Badal *et al.*, 2000; Batlló, 2004). Next, this Institute opened new observatories in Almería (1911), Málaga (1913) and Alicante (1914), and undertook major upgrades of the seismographic instrumentation used (circa 1924, Anduaga, 2004, 2009). It has also been gathering macroseismic data since 1909 (Roca *et al.*, 2004), with detailed questionnaires at least since the 1920's (Servicio Sismológico, 1923; Anduaga, 2009).

3.2.3 Printed bulletins (1924–1995)

Parallel to the instrumentation updates, phase picks of instrumentally recorded earthquakes started being reported by the *Instituto Geográfico* in printed bulletins in August 1924. These bulletins have been a primary source of information for the catalogue. Most of them are currently scanned and available online, with gaps for the period of the Spanish Civil War (1936–1939) and 1940's (Instituto Geográfico Nacional, 2015; Michelini *et al.*, 2005; Euroseismos Project, 2006). The last bulletin

reported the seismicity of 1995. Earthquakes occurred since 1996 are only reported in the digital catalogue (see Sec. 3.2.14). The bulletin for 1987 was the first to include the seismicity of the Canary Islands.

3.2.4 Civil War (1936-1939) and postwar

Science in general, and seismology in particular, suffered from the havoc of the Spanish Civil War (1936–1939). During war time, several observatories had to be either abandoned or operated under precarious conditions (Anduaga, 2009) resulting in a fragmentary record, evidenced by a decrease of the apparent seismicity rate during that period (Rodríguez de la Torre, 1989). The effects of the war on observational seismology were long-lasting (Anduaga, 2009). For example, the central geophysical observatory of Toledo was modern, internationally recognized and well maintained at the beginning of 1936, while by 1946 “more than half of the scientific personnel of 1936 were dead and/or exiled, and the international prestige had been lost” (Anduaga, 2009, p. 234). It was not until 1952 (Batlló, 2004; International Seismological Centre, 2015a) that the next national seismological station was opened (the first permanent station in the Canary Islands, in Tenerife).

A deep, strong, earthquake originated in 1954 under southern Spain (Frohlich, 2006) renewed the interest in Seismology in the country and eventually lead to improved stations and international cooperation (Udías, 2013).

3.2.5 Systematic earthquake location (1955–present)

Instrumental locations for earthquakes of the first half of the XX century were compiled from different databases and not recalculated using a common procedure (Mezcua & Martínez-Solares, 1983). The locations of later earthquakes have been systematically computed using software (Sec 3.3.2): retrospectively by Mezcua & Martínez-Solares (1983) for events between 1955 and 1980, and routinely for earthquakes since 1981 (Sec. 3.2.14).

3.2.6 Instrumental era (1962–present)

During the current, instrumental era, seismograms are the main source of information of the IGN catalogue, macroseismic information being complementary.

In the Iberian Peninsula, the onset of this era was marked by the set up of three stations of the World-Wide Standardized Seismographic Network (WWSSN): Toledo and Málaga (Spain) in 1962 and Porto (Portugal) in 1963 (Alguacil & Dávila, 2003; López & Muñoz, 2003; Martínez-Solares, 2003; Peterson & Hutt, 2014). And

a new Spanish national observatory (Logroño) started experimental recording ca. 1961 and full operation in 1963. The last national observatory opened in Santiago, northwestern Iberian Peninsula, in 1972. Since 1962, magnitude is systematically reported for most earthquakes, and an increase in the apparent rate of earthquakes is noticeable (Fig. 3.2, top).

In the Canary Islands, the instrumental era is considered to have begun in 1975, when earthquakes in the catalogue became typically characterized by magnitude, instead of by maximum felt intensity (Martínez Solares *et al.*, 2013). This improvement was possible by the opening of two new stations, in the islands of El Hierro and La Palma (in September 1973 and November 1974, respectively, International Seismological Centre, 2015a), together with the station already existing in Tenerife.

3.2.7 First telemetred network (1978–1992)

The first telemetred seismograph of IGN (located in Sierra de Guadarrama, near Madrid) provided the first phase picks to the catalogue in October 1978. By 1985, up to eight homogenous, short-period, vertical-component IGN stations with common time synchronization and real-time data transmission via telephone lines had been opened in peninsular Spain (e.g. García Fernández *et al.*, 1987), marking the beginning of the *Red Sísmica Nacional Española* (Spanish National Seismic Network, Instituto Geográfico Nacional, 1991; Carreño *et al.*, 2003; Carreño Herrero & Valero Zornoza, 2011). Since 1985, thanks to these enhanced detection capabilities, most earthquakes in the catalogue have magnitude ≤ 3 (Fig. 3.2).

Two subnetworks have been traditionally considered (e.g. Mezcua, 1995): one in the Iberian Peninsula, Balearic Islands and Spanish enclaves in northernmost Africa, and another in the Canary Islands. The stations in the Iberian Peninsula were not regularly spaced, but preferentially located in the onshore areas with the highest seismicity (Mezcua, 1995). Most of the currently known active faults (García-Mayordomo *et al.*, 2012) are also located there, although there are also faults in regions with low seismicity (e.g. Masana *et al.*, 2001; Gutiérrez *et al.*, 2009).

Figure 3.3 shows the number of permanent stations of the Spanish National Seismic Network in the Iberian Peninsula and surroundings (mapped in Fig. 3.4) and the Canary Islands (mapped in Fig. 3.14). To find out the actual operating period of each station, I took into account the dates at which each station reported its first and last phase pick included in the catalogue. Also, if the station was upgraded from short-period to broadband instrumentation (as reported in the station details, Instituto Geográfico Nacional, 2015), I considered the first and last picks for each kind of instrumentation. For simplicity, other temporary interruptions were not

included in this plot. For most stations, these dates are consistent with the official dates of station openings and closings (Instituto Geográfico Nacional, 2015), though there are a few exceptions. Some stations started reporting phase picks before its official opening (probably because they were already installed as temporary stations). Also, a few stations reported phase picks, occasionally, after their official closure.

In the Canary Islands, the network was very sparse in the late 1980's (Figs. 3.3 and 3.14), and recorded only minor seismicity in these initial years (Mezcua *et al.*, 1992). The need for improved monitoring was already recognized (García Fernández, 1988), and highlighted by the occurrence earthquake with magnitude 5.2 on May 1989 between the islands of Tenerife and Gran Canaria (Mezcua *et al.*, 1992; Jiménez & García-Fernández, 1996; Vinciguerra & Day, 2013). This seismic crises prompted the installation of two new permanent stations which started operating in 1990.

The telemetred short-period network was a major improvement with respect to the early observatories, but also had several limitations. The selection of sites for its stations was constrained by the availability of communication facilities (García Fernández *et al.*, 1987). Eventually, as urban areas expanded, this implied that most stations suffered from high seismic noise (Canas Torres & Vidal Sánchez, 2001). Also the records of moderate and major earthquakes were frequently saturated and clipped (e.g. Rueda, 1995), because of the narrow dynamic range of the telephone lines used for data transmission (Mezcua, 1995). This kind of problem (which also affected other networks with similar telemetric systems at the time, e.g. Amato & Mele, 2008) is particularly annoying for calculating reliable magnitudes, since it especially affects the stations closest to the earthquake, which are expected to record it with the highest signal-to-noise ratios (e.g. Yang & Ben-Zion, 2010). These issues were not solved until the new digital network was set up (Sec. 3.2.11).

3.2.8 First automatic locations (1991–1997)

In the early 1990's, when the short-period telemetred network reached its maximum development (Figs. 3.3 and 3.4), the first automatic procedure for near real-time hypocentral determination was used, which potentially enabled a more systematic and complete detection. It began experimental operation in the spring of 1991 (according to the seismic bulletin for that year) and was detailed by Mezcua (1995) and Rueda (1995). This system was replaced by a more advanced one in late 1997 (Sec. 3.2.10).

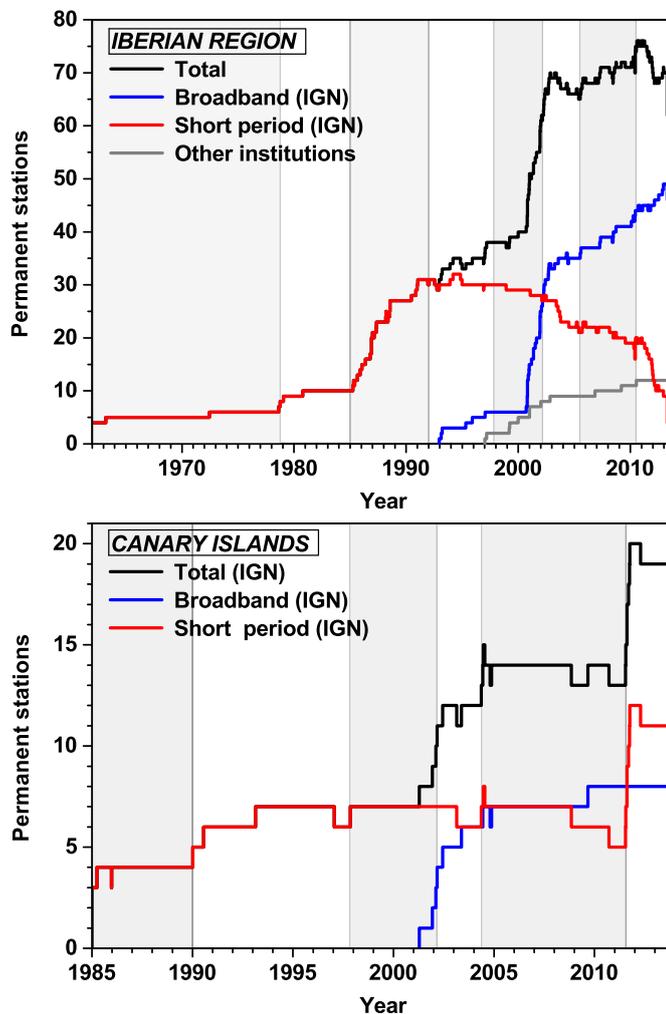


Figure 3.3: Number of permanent seismic stations of the National Seismic Network on the Iberian Peninsula and surroundings (*Iberian Region*, for short) (*top*) and on the Canary Islands (*bottom*) reporting phase picks to the catalogue. It is indicated whether the stations are managed by IGN. Stations from other institutions are only included in this plot if they were used for the real-time automatic location procedures. Many other stations from other institutions have been used for elaborating the final catalogue. Station maps are plotted in Figs. 3.4 and 3.14.

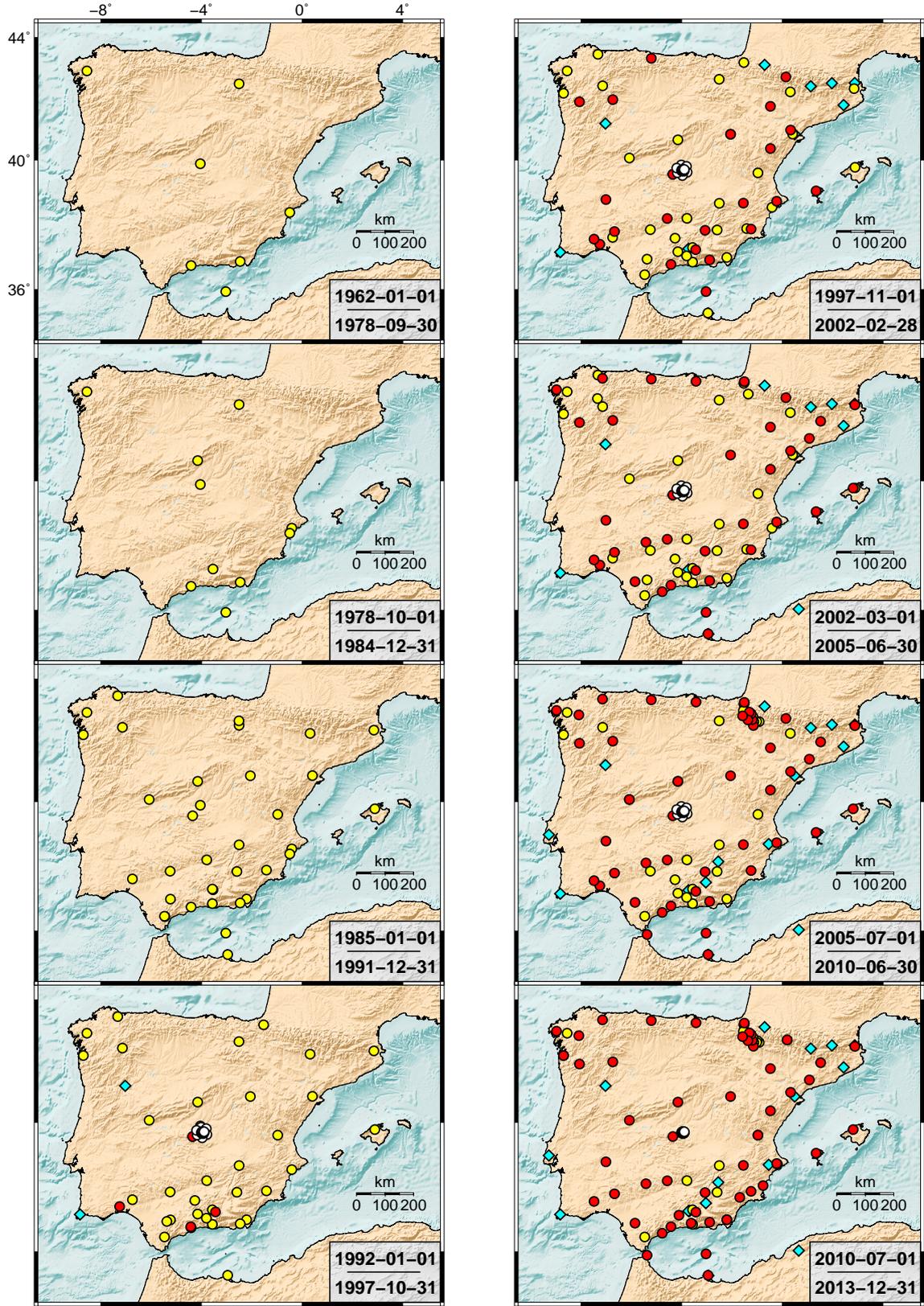
3.2.9 The Sonseca Array (1992–present)

The Sonseca Array is a small-aperture array of short-period, long-period and broadband stations, deployed inside boreholes around the town of Sonseca, at the centre of the Iberian Peninsula (Instituto Geográfico Nacional, 1991; Martínez Solares, 1992; Mezcua & Rueda, 1994; Mezcua, 1995; Dowla, 1996; Carreño *et al.*, 2001; Bahavar & North, 2002; Hasting & Barrientos, 2002; Rueda Núñez, 2006; Kværna & Ringdal, 2013). Its main purpose has been nuclear explosion monitoring, and is currently a primary station of the International Monitoring System of the Comprehensive Nuclear-Test-Ban Treaty Organization. Opened in February 1957, it was initially owned by the United States of America, until 1996, when it was formally transferred to the IGN. Data from this array had already started being received by IGN in 1988 (Mezcua, 1995), and in real time from mid-1989 (Instituto Geográfico Nacional, 2015). It became fully integrated into the permanent IGN network in 1992 (Rueda Núñez, 2006), and provided the first phase pick to the catalogue in December of that year. The Sonseca Array has been upgraded several times (Hasting & Barrientos, 2002; Instituto Geográfico Nacional, 2002a), and its long-period instruments, located on the outer part of the array, were removed in June 2010 (Fig. 3.4).

The Sonseca Array usually contributes to the hypocentral determination as a single station with superior signal detection capability (Sonseca Array beam reference point, station code ESDC, International Seismological Centre, 2015a). To do so, the records of the array elements are combined into a single one, using beam-forming techniques which enhance the signal-to-noise ratio and provide the velocity vector of an incoming wavefront (its back azimuth and slowness, e.g. Rost & Thomas, 2002; Rueda Núñez, 2006; Schweitzer *et al.*, 2009). This makes such array particularly valuable for locating events (e.g. Martínez Solares, 1992, 1995; Rueda, 1995). The array instruments typically do not supply phase picks individually to the catalogue, since this would imply an excessive number of observations recorded at almost exactly the same azimuth, that could cause a hypocentral location bias (Rueda, 1995). The main exception is the broadband station with code

Figure 3.4: **Seismic stations in the Iberian Peninsula and surroundings**, for the periods considered for completeness assessment (Fig. 3.12). They are either owned by the Instituto Geográfico Nacional (*circles*), or by other institutions (*squares*) but whose data have been used for real-time hypocentral determinations by IGN. Data from other stations (not shown) have been used to elaborate the final, revised catalogue. Each map includes the stations for which the catalogue reports phase picks during at least part of the corresponding period.

IGN stations: ● Short-period ● Broadband ○ Sonseca Array ◆ Stations of other institutions



ESLA (International Seismological Centre, 2015a), included in the array, which did contribute numerous picks individually, and is included separately in Fig. 3.3.

Because of its sensitivity, and despite ESDC being located somewhat far from seismically active areas, it provides more data to the catalogue than other stations; in particular, it is the one which until 2013 contributed more phase picks for earthquakes with $M \geq 2.0$ in the Iberian Peninsula and its close surroundings (36°N – 44°N , 10°W – 4°E).

3.2.10 New automatic earthquake location, and stations from other institutions (1997–present)

In November 1997, complex automatic location procedures came into use (Carreño *et al.*, 2003), using the EvLoc location software (Nagy, 1996). Since then, this software has been used to calculate the final locations in the catalogue (Carreño Herrero & Valero Zornoza, 2011) and will be discussed in more detail in Sec. 3.3.2.

From 1997 onwards, stations from other institutions are used for the automatic real-time hypocentral determinations, so they are in practice an extension of the Spanish National Seismic Network (Figs. 3.3 and 3.4). The whole set of permanent stations involved in real-time earthquake location will be referred to here as the *backbone network*. For elaborating the final catalogue, IGN uses data from additional stations, eventually provided (not in real time) by other institutions (Sec. 3.2.14).

3.2.11 Broadband network (1999–present)

The first few broadband stations of the Spanish National Seismic Network were installed in the 1990's (Fig. 3.3 and Mezcuca, 1995). In late 1999, installation commenced on the new network of three-component, digital broadband stations (*Red Sísmica Digital Española*, Spanish Digital Seismic Network). Data are transmitted via satellite for most of them, so they could be sited even in remote locations with low seismic noise.

The fast initial deployment of the first set of new stations was completed in 2002; since then, the number of broadband IGN stations has increased more slowly (Fig. 3.3). Thanks to the broadband network, the number of located events per unit time increased significantly from 2002 onwards (Fig. 3.2 and Carreño *et al.*, 2003), especially for those with $M \leq 3$ (but see Sec. 3.4.6 for the effect of magnitude changes). Concomitantly to the deployment of broadband stations, the analog short-period stations have been progressively replaced or closed down, especially in the Iberian Peninsula (Fig. 3.3). Lahr & Van Eck (2003) reviewed the status of

the IGN network around 2000. Rueda & Mezcua (2015) have presented a recent overview of the IGN broadband network and analysed the instrument orientations.

3.2.12 Monitoring around the Itoiz and Yesa reservoirs (2005–2013)

Apart from the Sonseca Array, the Aragón River Valley, in the Western Pyrenees, eventually became the geographic region where the IGN network was densest in the Iberian Peninsula (Fig. 3.4). Fourteen three-component stations (3 broadband, 11 short-period) and an accelerometric network were progressively deployed there between mid-2005 and mid-2010 (apart from two broadband stations already existing). Their purpose was to monitor in great detail the seismicity in the surroundings of two large water reservoirs, Itoiz and Yesa. This was motivated by a $M_w = 4.5$ earthquake on 18 September 2004, possibly triggered by the first impoundment of the recently-built Itoiz reservoir (Ruiz *et al.*, 2006; Durá-Gómez & Talwani, 2010; Luzón *et al.*, 2010; Santoyo *et al.*, 2010). Most of these stations were closed down between mid-2011 and late-2013, leaving the Spanish National Seismic Network in the Iberian Peninsula almost devoid of short-period stations by the end of 2013 (Fig. 3.3).

3.2.13 Monitoring of seismo-volcanic activity in the Canary Islands (2004 and 2011–present)

In the Canary Islands, two network improvements were spurred by seismic crises which started in 2004 and 2011.

In Tenerife, a seismic swarm took place in 2004, starting in April (e.g. Almendros *et al.*, 2007; Martí *et al.*, 2009; Domínguez Cerdeña *et al.*, 2011, and references therein). Three additional permanent seismic stations (two short-period and one broadband) were installed on the island between May–June 2004 (the first one in 17 May), plus two temporary ones which operated intermittently from June to November 2004 (Domínguez Cerdeña *et al.*, 2011, not considered in Fig. 3.3). This episode noticeably increased the recorded earthquake rate (Fig. 3.2). The short-period stations stopped contributing to the catalogue in 2008 and 2010, respectively.

Increased activity on 19 July 2011 on El Hierro (the southwesternmost island) led to dedicated geophysical monitoring, including up to nine new short-period seismic stations, most of them three-component (López *et al.*, 2012; Domínguez Cerdeña *et al.*, 2014). Seven of these stations have contributed data to the IGN catalogue (Fig. 3.3), the first one starting on 21 July. The seismic activity continued for months (Ibáñez *et al.*, 2012; Domínguez Cerdeña *et al.*, 2014), preceding a submarine eruption which started on 10 October 2011 off the southern shore of the island, and which was

declared finished as of 5 March 2012. Several seismic swarms ensued, related to magma intrusions (the first ones described by García *et al.*, 2014, and later ones until 2014). Until the end of 2013, this activity on El Hierro added $\sim 20,000$ earthquakes to the catalogue (Fig. 3.2).

3.2.14 Current practice on catalogue compilation

For each detected event, a preliminary, automatic determination of hypocentral location and magnitude is calculated automatically in real time and posted online. These initial determinations are then updated after a manual review by the around-the-clock Seismic Information Service of IGN (Instituto Geográfico Nacional, 2015).

Inspection of the catalogue indicates that a minimum of three stations are used for earthquake location. Triggering of real-time automatic hypocentral determinations requires three stations in the subnetwork of the Iberian region and four in the subnetwork of the Canary Islands (Resurrección Antón –IGN– pers. comm., 2013).

The catalogue data may be eventually subjected to further reviews. It can be regarded as fully revised only after being updated in yearly blocks, with an irregular delay of about 18–24 months (similarly as the former annual printed bulletins). As with other catalogues (e.g. Willemann & Storchak, 2001), this long latency period allows for a careful data review and incorporation of phase pick information eventually delivered by other collaborating institutions (e.g. in Spain, Portugal, France, Morocco, and Algeria). This additional data improves the catalogue especially for large earthquakes (which can potentially be detected at larger distances by other stations), or in general for earthquakes occurring where networks from other institutions have better coverage. That is, the revised catalogue is more complete than the detection threshold of the Spanish National Seismic Network alone would allow (e.g. García Fernández, 1986). Occasionally, specific earthquakes are subject to later reviews if new information or improved analyses are available. At the time of this writing, the catalogue is reviewed up to the end of 2013.

From the IGN webpage, the catalogue may be downloaded in two versions. The abridged version (option labelled as “without phase readings”) provides a list of data about origin time, location, preferred magnitude, and maximum intensity, for earthquakes from the historical period to present. The full version (option “with phase readings”) is available online only for earthquakes since November 1997 until the present, and provides a catalogue in the IMS1.0 standard format (Biegalski *et al.*, 1999), which includes the uncertainties of origin time, location and magnitude, and

the list of phase picks from all the stations used to locate the event and calculate its magnitude. It also specifies the date at which the data for each earthquake were generated (that is, the date of the last review), and the “author” field is reported as “IGN” for the preliminary solutions and as “bull_2” for the final ones. The full version of the digital catalogue for earthquakes between January 1983 and October 1997 (used in Sec. 3.3) has other formats and is available by request to IGN (see contact information for seismology on their web page). Moment tensor and moment magnitude calculations (Rueda & Mezcua, 2005) are provided in a separate database online (Instituto Geográfico Nacional, 2015).

The Spanish Accelerograph Network (managed by IGN) started being deployed in 1974 and was described by Carreño *et al.* (1991b), Mezcua (1995) and Mezcua *et al.* (2008). Accelerograms with timing calibrated by GPS are occasionally used to improve the hypocentral determinations in the IGN catalogue, e.g. for the earthquakes of Lorca (southeastern Iberian Peninsula) in 2011 (Cabañas *et al.*, 2014; Morales *et al.*, 2014), the latest damaging ones in Spain (Martínez-Díaz *et al.*, 2012). At least since the 1970s IGN also deploys temporary networks for microseismicity surveys or aftershock monitoring (e.g. seismic bulletin for 1977, Carreño *et al.*, 1991a; Herraiz & Lázaro, 1991). Since 2006, the IGN can even deploy a local temporary network with real-time data transmission via satellite (Instituto Geográfico Nacional, 2006).

Macroseismic information is compiled by IGN using questionnaires, and in real time by means of felt reports contributed online by witnesses (Instituto Geográfico Nacional, 2015).

3.3 Precision of earthquake locations

For most earthquakes, the catalogue specifies the precision of the hypocentral location (the formal uncertainty due to random errors, e.g. Husen & Hardebeck, 2010). The actual accuracy of a location may be worse, because of additional, systematic errors (e.g. in the velocity model used, Havskov & Ottemöller, 2010; Husen & Hardebeck, 2010); it can be assessed only if the exact location is known independently (such as in a controlled explosion source, e.g. Yang *et al.*, 2004; Havskov & Ottemöller, 2010). I should note that epicentral location is a well-posed problem which may be solved with increasing accuracy as the number of observations increases, while exact depth determination is an ill-posed problem (Lomnitz, 1982, 2006).

The typical uncertainty of hypocentral locations varies in time (due to changes in the network and in location procedures) and in space. Events occurring far from

the network are expected to have the largest uncertainties (e.g. García Fernández, 1986; García Fernández *et al.*, 1987; Rueda Núñez, 2006; D'Alessandro *et al.*, 2013; Cesca *et al.*, 2014; Custódio *et al.*, 2015), because they are typically recorded by few stations and with a large azimuthal gap (the largest angular gap in azimuth from the epicentre to azimuthally adjacent stations, e.g. Havskov & Ottemöller, 2010).

3.3.1 Location procedures

Instrumental locations in the IGN catalogue are single-event (that is, calculated for each earthquake individually, based on the picks of wave phase arrivals at each station), using layered, 1-D velocity models. In the Iberian Peninsula and surroundings, the model used by Mezcua & Martínez-Solares (1983) for earthquakes occurred since 1955 has been adopted for routine locations by IGN until today (IGN bulletins, Carreño *et al.*, 2003; Carreño Herrero & Valero Zornoza, 2011). For moment-tensor determinations by IGN, the model for the Iberian Peninsula and surroundings has an additional, shallow layer with lower seismic velocities (Rueda & Mezcua, 2005). In the Canary Islands and surroundings (19°W–13°W, 27°N–30°N) a different model (labelled as “Canary”) is being used; it was specified for the first time in the Bulletin for 1987 (the first one reporting earthquakes from this archipelago) and, according to Rueda Núñez (2006), is an average of the models proposed by Dañobeitia (1980) for each island. From January 2004 to December 2014, a different model (labelled as “Teide”) has been used for locating earthquakes under the Teide stratovolcano (Tenerife Island). It is actually a model for the Etna volcano (Scarpa *et al.*, 1983), which had been shown able to improve earthquake locations also in the Teide (Almendros *et al.*, 2000), with the Moho discontinuity depth adapted to match that of the Canary model. Both P and S waves are used for calculating the revised hypocentral locations since 1955 (e.g. Rueda, 1995), with a fixed velocity ratio $V_P/V_S = 1.75$ (IGN bulletin for 1987, Rueda, 1995; Carreño *et al.*, 2003; Rueda & Mezcua, 2006). D'Alessandro *et al.* (2013) discussed the limitations of the velocity model used by IGN for the Iberian region, although they assumed a different ratio ($V_P/V_S = 1.73$) and that the same model was also used for the Canary Islands. The full-format catalogue specifies the model used for locating each earthquake (if different from the default one of the Iberian region).

Double-difference relocations (Waldhauser & Ellsworth, 2000) have been calculated by IGN and collaborators in specific case studies, such as for the 2011 Lorca earthquake series (Morales *et al.*, 2014), and for the seismo-volcanic swarms of 2004-2005 in Tenerife (Domínguez Cerdeña *et al.*, 2011) and of 2011 in El Hierro (Domínguez Cerdeña *et al.*, 2014). The relocated catalogues for the Canary

Islands can be downloaded separately from the IGN webpage (Instituto Geográfico Nacional, 2015).

Before the strictly instrumental era, location uncertainties may be substantial, given that the early mechanical seismometers (Batlló, 2004) lacked common timing. For example, the catalogue locations of an earthquake sequence in southern Spain in 1951 are now taken from Batlló *et al.* (2008, see also references therein), and corrected the earlier ones (taken from Mezcua & Martínez-Solares, 1983) by tens of km. No location uncertainties are reported in the catalogue for the first half of the XX century.

For earthquakes since 1955 until October 1997, Hypo71 (Lee & Lahr, 1975) was used to calculate the final, revised hypocentral locations (IGN Bulletins, Mezcua & Martínez-Solares, 1983; Carreño *et al.*, 2003). This software implements the iterative, nonlinear, least-squares method of Geiger (1910, 1912). The only source of uncertainty that this software accounts for is the measurement error inherent to each arrival time, assumed random and normally (Gaussian) distributed (e.g. Boyd & Snoke, 1984). The location uncertainties calculated with Hypo71 were reported by Mezcua & Martínez-Solares (1983) for a few events in the 1950s and for the majority of earthquakes from 1960 to 1980. From 1981 to 1995 they are reported in most IGN bulletins.

The digital full-format catalogue from January 1983 to October 1997 (Resurrección Antón, IGN, pers. comm., 2012; not available online) will be used here for analysing location precision in this first period. In 1983–1984, an average of ~ 11 phases were used for each earthquake location, which rose to ~ 20 from January 1985 to October 1997. Hypo71 handles a default maximum of 100 phases for each earthquake (Lee & Lahr, 1975), and indeed before November 1997, only three earthquakes were located with a higher figure.

Since November 1997, hypocentral locations and their uncertainties are reported in the full-format catalogue online (Instituto Geográfico Nacional, 2015), and have been calculated with Evloc (Carreño Herrero & Valero Zornoza, 2011, and Resurrección Antón, IGN, pers. comm., 2015). This software (Nagy, 1996) implements the method by Bratt & Bache (1988) and allows taking into account random errors not only in measurements (phase pickings) but also those resulting from modelling (e.g. Yang *et al.*, 2004). In any case, the precision assessment cannot be perfect; for example, errors in Evloc are assumed independent, but correlated errors (and underestimated uncertainties) may result from using many wave arrivals which sample similar ray paths (Yang *et al.*, 2004). The number of phases used for locating each earthquake with Evloc has been highly variable, with

an average of ~ 24 , with ≥ 30 or more phases in 22% of events, and ≥ 100 phases for 2% of events.

In the following, I analyse separately only these two broad periods with different final location routines (from January 1983 to October 1997 and from November 1997 onwards), although location precision is obviously expected to vary as the recording network evolved. The reason for this choice is that the specific uncertainties for each earthquake can be checked in the catalogue, unlike the catalogue completeness, which I will analyse in greater temporal detail later on.

3.3.2 Epicentral location

Before November 1997, epicentral location uncertainties were calculated with Hypo71 and specified as the so-called ERH (standard horizontal error, in km). Unfortunately, ERH values calculated by Hypo71 (as the square root of the sum of the estimated latitude and longitude variances) indeed lack an inherent statistical interpretation (Boyd & Snoke, 1984; Husen & Hardebeck, 2010) and are generally smaller than the actual ERH.

ERH values reported by Mezcua & Martínez-Solares (1983) are in the order of several km to tens of km, and tend to decrease along time. In the early digital catalogue, the mean reported ERH is ~ 7 km during 1983–1984, ~ 4 km in 1985, and ~ 3 km from January 1986 to October 1997. The marked reduction of ERH in the mid-1980's is explained by the deployment of the telemetred network and the increasing number of phases used for locating each event.

ERH values tend to be smaller onshore the Iberian Peninsula than elsewhere, evidencing that location uncertainty increased with the epicentral distance from the network (Fig. 3.5; for this kind of map see e.g. Zaliapin & Ben-Zion, 2015).

Since November 1997, epicentral location uncertainties calculated by Evloc are reported as error ellipses at 90% confidence level in the full-format catalogue. In the absence of systematic location biases, the true epicentres would be expected to be inside such ellipses in 90% of instances.

Reported uncertainties reach the lowest values within the Iberian Peninsula, and worsen rapidly offshore as the distance from the network increases (Fig. 3.6). Rueda Núñez (2006, pp. 150–151) calculated theoretical values of the semi-axis of the error ellipse, in the Iberian region, for $M = 3.0$, and considering the stations of the real-time IGN network then in operation. In general, those values and their spatial pattern are very similar to the results of Fig. 3.6. But in regions like the Pyrenees and Portugal, location uncertainties are typically lower than calculated by Rueda Núñez (2006) thanks to the data contributed to the catalogue by other networks (see also Souriau *et al.*, 2014, for the Pyrenees).

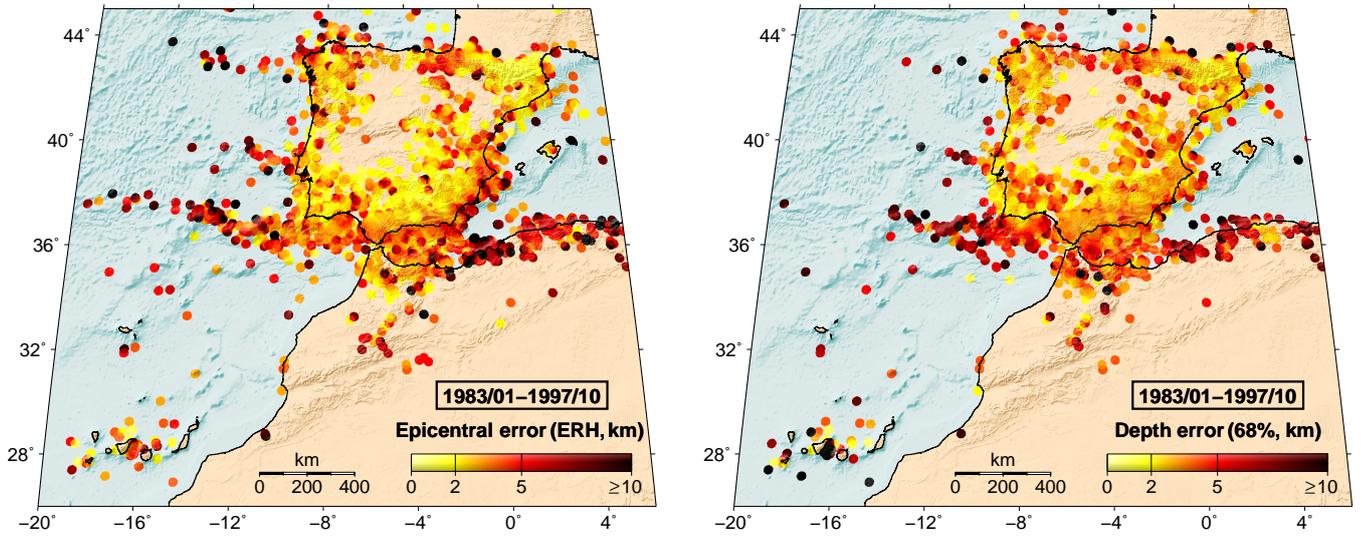


Figure 3.5: Precision of hypocentral locations from January 1983 to October 1997. *Left:* Epicentral location errors (ERH values). *Right:* Hypocentral depth errors (68% confidence ranges). Each location in the map shows the average value for the events with epicentres within a radius of 20 km.

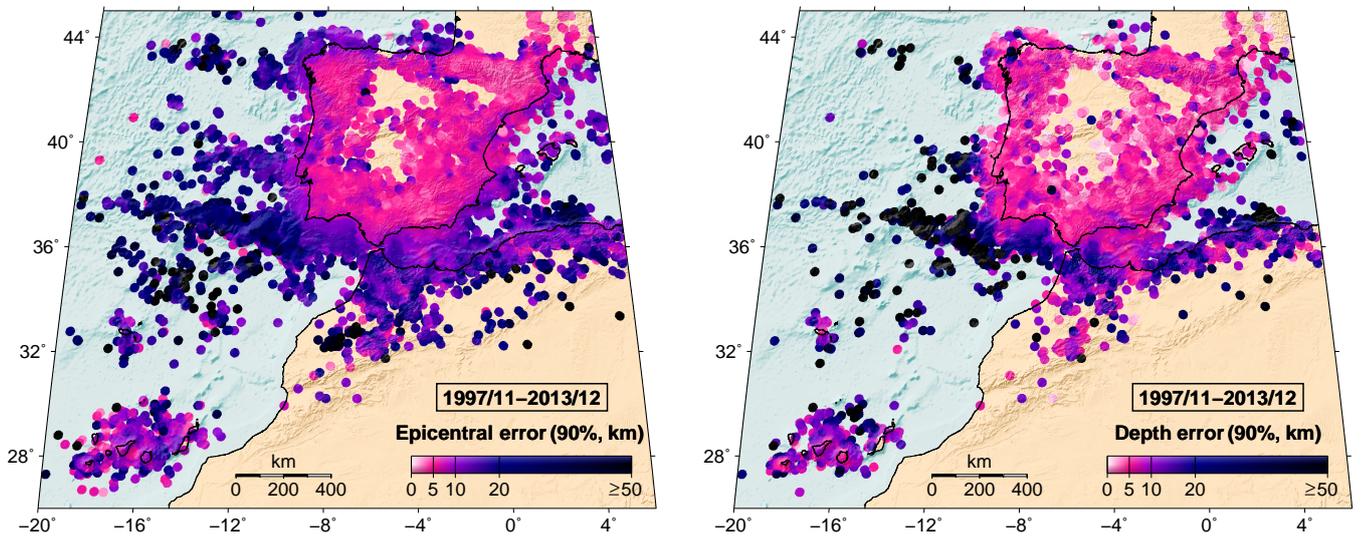


Figure 3.6: Precision of hypocentral locations from November 1997 to December 2013. *Left:* Epicentral location errors (lengths of semi-major axes of the 90% confidence ellipses). *Right:* Hypocentral depth errors (90% confidence ranges). Spatial averages are calculated as in Fig. 3.5, but the errors were reported differently in the catalogue, hence the different colour scale used.

Larger earthquakes are potentially detected by more stations, so they are expected to be typically better located than nearby smaller earthquakes. Nevertheless, location accuracy might actually worsen with magnitude if the deep part of the velocity model is inaccurate, since seismic ray traces from larger earthquakes sample its deeper layers (Husen & Hardebeck, 2010). I found that earthquakes with $M \geq 3$ within the Iberian Peninsula typically had lower nominal location uncertainties than those shown in the same area in Fig. 3.6, where the results are dominated by the more frequent (and less precisely located), smaller earthquakes. Further away, most earthquakes have $M \geq 3$ anyway, so no substantial differences are observed when considering them alone.

In the Canary Islands, location uncertainties are generally higher than in the Iberian Peninsula. Azimuthal gaps are frequently large because the network geometry is constrained by the locations of the islands themselves. Moreover, noise level on the stations is relatively high (D'Alessandro *et al.*, 2013), so the small earthquakes are typically detected by few of them. Further insight on location precision for specific earthquake series in the Canary Islands can be found in the relocation studies by Domínguez Cerdeña *et al.* (2011, 2014).

3.3.3 Focal depth

Most of the seismicity in the catalogue is shallow (with focal depth ≤ 30 km for $\sim 93\%$ of events whenever calculated), but a fringe of intermediate-depth events (depth between 30 and 170 km) extends along the Iberian-African plate boundary (e.g. Buforn *et al.*, 1997; Carreño *et al.*, 2003, Fig. 3.1). Also, since 1954, six deep earthquakes (depth > 600 km) have been recorded below the Granada region, in the South of the Iberian Peninsula (Frohlich, 2006; Buforn *et al.*, 2011; Bezada & Humphreys, 2012; Mancilla *et al.*, 2012, and references therein).

No focal depths are reported in the IGN catalogue for events whose epicentre is determined by macroseismic means. Depth begins to be reported for a minority of earthquakes of the 1950's (Mezcua & Martínez-Solares, 1983). In the abridged catalogue, it is rounded off to the nearest km, while in the full-format catalogue since November 1997 it is rounded off to the nearest tenth of km. The evolution of reported depth versus time can be observed in Fig. 3.7.

For a large proportion of earthquakes (e.g. $\sim 37\%$ since November 1997) the independent calculation of depth does not converge, and fixed depth values are used (highlighted in Fig. 3.7), which restricts the location procedure to calculating the epicentral coordinates. Similar procedures have been used in other catalogues (e.g. Gerstenberger & Rhoades, 2010; Bondár & Storchack, 2011). Fixed depths, despite being the least constrained ones, have no reported uncertainty. From 1983

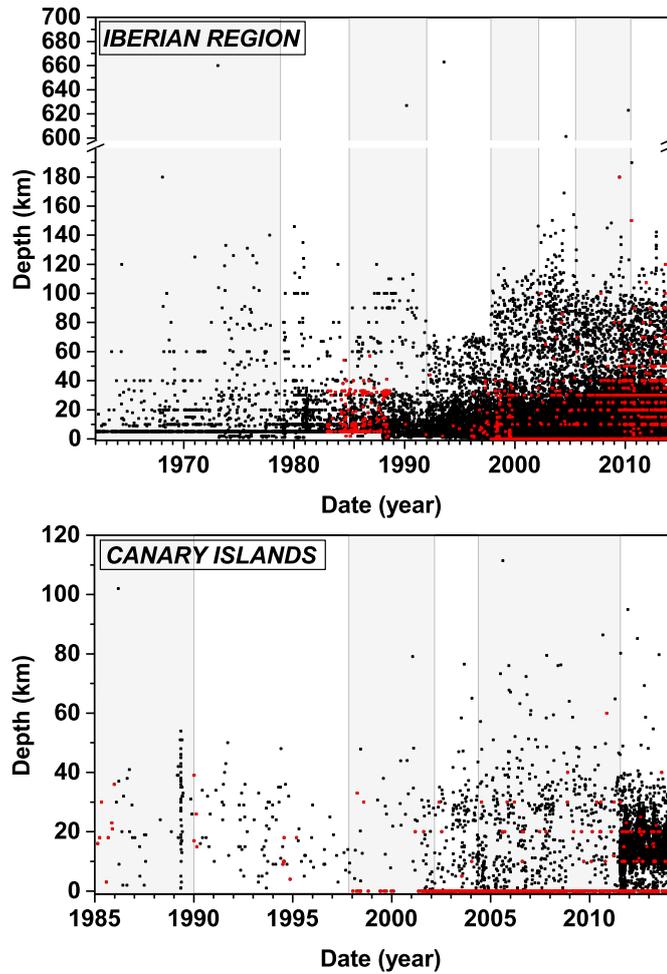


Figure 3.7: **Reported hypocentral depth versus time (dots).** *Red dots:* Depths reported as fixed, that is, not calculated as independent parameters (the distinction is only available from 1983 onwards). *Vertical fringes:* Periods considered for completeness assessment. Geographic areas are those shown on maps of Figs. 3.12 and 3.14. In the Iberian Region, note the scale break between 200 and 600 km, where no value is reported in the catalogue.

until October 1997, in the digital catalogue used here, fixed depths are assumed to be those without error reported. Since November 1997, fixed depths are flagged as such in the full-format catalogue, and omitted in the abridged version. Fixing depths has been proportionally more common in the earlier years (clearly until 1989 in the Iberian Region) and for small earthquakes (typically located with few stations) in more recent times. From 1983 until October 1997, the most usual value of fixed depths was 5 km, while since November 1997, it is 0 km, followed by 5 km and its multiples.

In the Iberian Peninsula, from January 1992 to October 1997, no depths $\gtrsim 70$ km are reported, except for the 1993 deep earthquake (Fig. 3.7). Considering the depth distribution in the rest of the catalogue, from 1962 to 2013, this gap seems unlikely to be natural. Rather, it may be related to the location procedures used only during this period (Sec. 3.2.8). Depths $\gtrsim 70$ km were not reported again until November 1997 (Fig. 3.7), when improved automatic location routines became implemented.

Fig 3.5 shows the spatial distribution of reported hypocentral uncertainty from 1983 to October 1997. This uncertainty, calculated by Hypo71, can be interpreted as 68% confidence interval under some assumptions (Boyd & Snoke, 1984; Husen & Hardebeck, 2010). Its proper statistical interpretation would require taking into account the number of phase picks used for calculating depth (Boyd & Snoke, 1984). For this period, depth determinations tend to be better on the Iberian Peninsula than offshore and in distant areas, and are generally poor in the Canary Islands, such as for the earthquake series between Tenerife and Gran Canaria in 1989.

The geographic pattern of depth uncertainties is very clear since November 1997 (Fig. 3.6). These are calculated with Evloc and reported as a 90% confidence interval in the full-format catalogue. Onshore the Iberian Peninsula, and close to its coast, they are typically ~ 5 km. Further away, they rise rapidly in proportion to the distance from the network. Beyond ~ 150 km from the coast in the Atlantic (and offshore the Algerian coast), depths are mostly unconstrained (typical 90% uncertainties are > 50 km). In the Canary Islands, reported depth uncertainties are heterogeneous, and usually larger than for the Iberian Peninsula. Uncertainties reported for the El Hierro earthquake series are typically low, thanks to the detailed monitoring (Sec. 3.2.13).

Focal depth is the most difficult spatial parameter to constrain (e.g. Vere-Jones, 1992; Carreño *et al.*, 2003; Husen & Hardebeck, 2010), but in some regions the typical 90% depth uncertainties seem lower than the epicentral ones (Fig. 3.6). The reason is that the highest depth uncertainties (those for fixed depth values) are not reported in the catalogue and cannot be considered in the corresponding maps (Figs. 3.5 and 3.6), so these have an optimistic bias.

3.4 Magnitude determination and precision

The magnitudes reported in the current version of the IGN catalogue have been calculated by using a variety of formulae, which are reviewed in this section, alongside the resulting uncertainties. These formulae have changed throughout time, and different ones were used in the original IGN bulletins before 1981 (López & Muñoz, 2003).

Magnitude values are reported in the catalogue only for earthquakes with sufficiently reliable instrumental records. No magnitudes based on macroseismic information are provided, although they have been determined elsewhere for several historical or early instrumental earthquakes (e.g. Samardjieva *et al.*, 1999; Martínez Solares & Mezcua, 2002; Mezcua *et al.*, 2004; Storchak *et al.*, 2013). The first magnitude value (=6.4) is listed for the 26 June 1910 Masqueray (Algeria) earthquake. As the records improved during the XX century due to instrumental updates, magnitudes could be calculated and reported for a progressively larger proportion of earthquakes (Fig. 3.2), and all events occurred after 1989 have a magnitude assigned.

The reported magnitudes are, where possible, averages of the values obtained from a number of stations in a diversity of azimuths around the epicenter (Rueda, 1995). Such averaging tries to counteract the variability of magnitudes calculated at different sites, which is due to the original radiation pattern, the attenuation and scattering along each particular ray path, and the diverse local site conditions which may cause local attenuation or amplification (e.g. Rueda, 1995; Vila *et al.*, 2005).

The full-format catalogue (Sec. 3.2.14) reports the standard error of the mean magnitude for each event. Note that when only one station is used this error cannot be calculated, but is reported as zero. From November 1997 onwards, when the magnitude was calculated from the phase picks of two or more stations, the mean standard error is ~ 0.2 magnitude units. Magnitudes in the catalogue are rounded to one decimal place, which sets a precision limit of 0.1 units.

Different magnitude types can be reported for the same event, the preferred type being the only one reported in the short catalogue version (Sec. 3.2.14). The revision of the different magnitude types is detailed below (Secs. 3.4.1 to 3.4.5), followed by a discussion of the overall evolution of magnitude values in the catalogue (Sec. 3.4.6).

3.4.1 M_D , Iberian Region (1910–1961)

A duration magnitude (M_D) is reported for earthquakes in the Iberian Peninsula and surroundings, from 1910 to 1961 (before the installation of the first WWSSN station in 1962), taken from Mezcua & Martínez-Solares (1983). It is based on the duration

of the earthquake as recorded on at least one of three reference observatories (Toledo, Almería and Alicante), where the original mechanical seismographs operated for decades. A specific formula was used for each observatory (Mezcua & Martínez-Solares, 1983; Samardjieva *et al.*, 1999):

$$M_D = 1.67 \log D + 0.001\Delta - 0.2 \quad (\text{Toledo}), \quad (3.1)$$

$$M_D = 1.22 \log D + 0.001\Delta + 1.20 \quad (\text{Almería}), \text{ and} \quad (3.2)$$

$$M_D = 1.44 \log D + 0.001\Delta + 0.95 \quad (\text{Alicante}), \quad (3.3)$$

where D is the duration of the earthquake in seconds and Δ the epicentral distance to the station in km. The duration was measured from the arrival of the P wave until the moment at which the earthquake signal was indistinguishable from the background noise (Mezcua & Martínez-Solares, 1983). These formulae were calibrated by regression to body-wave magnitude values (m_b) provided by the U.S. National Earthquake Information Service for a set of reference earthquakes. Mezcua & Martínez-Solares (1983) state that the average standard error of these magnitude values is < 0.4 .

3.4.2 M_D , Canary Islands (1964–1997)

In the Canary Islands, a duration magnitude was used until October 1997, using the formula (Rueda, 1995):

$$M_D = 2 \log D + 0.0035\Delta - 0.87. \quad (3.4)$$

where D is the duration of the earthquake in seconds and Δ the epicentral distance in km.

At the time of this writing, the magnitude type for these earthquakes in the catalogue is incorrectly labelled as if it were m_{bLg} (*sensu* Mezcua & Martínez-Solares, 1983, see below). No measurements of precision of magnitude values calculated with this scale are available.

3.4.3 m_{bLg} (1962–present)

Most magnitude values in the catalogue are provided in a m_{bLg} scale, a body wave magnitude estimated from the peak vertical motion of the S -wave, which at regional distances in continental crust typically occurs in the Lg phase (for the phase definition, see Storchak *et al.*, 2003).

The Lg wave propagation may be blocked for several reasons, such as crustal scattering or a wave path through a thin, oceanic, crust (e.g. Payo, 1960; Chazalon *et al.*, 1993; Zhang & Lay, 1995; Calvert *et al.*, 2000; McNamara & Walter, 2001;

Sens-Schönfelder *et al.*, 2009; Noriega *et al.*, 2015, and references therein). This may lead to m_{bLg} being underestimated with respect to M_w for offshore earthquakes (Cabañas *et al.*, 2015). A clear example is the 26 May 1975 North Atlantic earthquake, with $M_w = 7.8$ (Storchak *et al.*, 2013), but a much smaller $m_{bLg}(\text{IGN}) = 6.7$.

Two different scales of m_{bLg} have been used in the catalogue, for different periods, as described below.

3.4.3.1 m_{bLg} (1962–2002)

In the printed IGN bulletins this m_{bLg} was usually denoted simply as m_b . These formulae were used (Mezcua & Martínez-Solares, 1983; Rueda, 1995):

$$m_{bLg} = \log\left(\frac{A}{T}\right) + 1.05 \log \Delta^\circ + 3.90 \text{ for } \Delta^\circ \leq 3^\circ, \text{ and} \quad (3.5)$$

$$m_{bLg} = \log\left(\frac{A}{T}\right) + 1.66 \log \Delta^\circ + 3.30 \text{ for } \Delta^\circ > 3^\circ, \quad (3.6)$$

where Δ° is the epicentral distance in geocentric degrees, and A and T are, respectively, the peak ground amplitude in microns, and the period of peak motion in seconds, of the Lg wave.

Mezcua & Martínez-Solares (1983) calibrated Eq. 3.5 using the same reference earthquakes as for the M_D scale described before for the Iberian Region. Eq. 3.6 is the formula adopted by the International Association for Seismology and the Physics of the Earth's Interior (Båth, 1969, 1981), discussed by Nuttli (1973), Payo (1996), Samardjieva *et al.* (1999), and references therein.

López & Muñoz (2003) pointed out that Eqs. 3.5 and 3.6 are not continuous to each other; there is a -0.34 unit fall at the epicentral distance of 3° (about 334 km), implying a relative underestimation of m_{bLg} for events occurred at larger distances from the station (Fig. 3.8).

Several regressions correlating this m_{bLg} with other magnitude scales were calculated by López Casado *et al.* (2000), Rueda & Mezcua (2002), Mouayn *et al.* (2004) and Cabañas *et al.* (2015).

From November 1997 to February 2002 the mean standard error reported for this magnitude scale is 0.1. Uncertainties for earlier events are not available; Cabañas *et al.* (2015) assigned by expert judgement values of 0.3 magnitude units before 1985, and 0.2 between 1985 and 2002.

3.4.3.2 m_{bLg} (2002–present)

A new, improved m_{bLg} formula (López, 2008) replaced the previous ones in 2 June 2003 for real-time magnitude determinations, and was later applied retrospectively

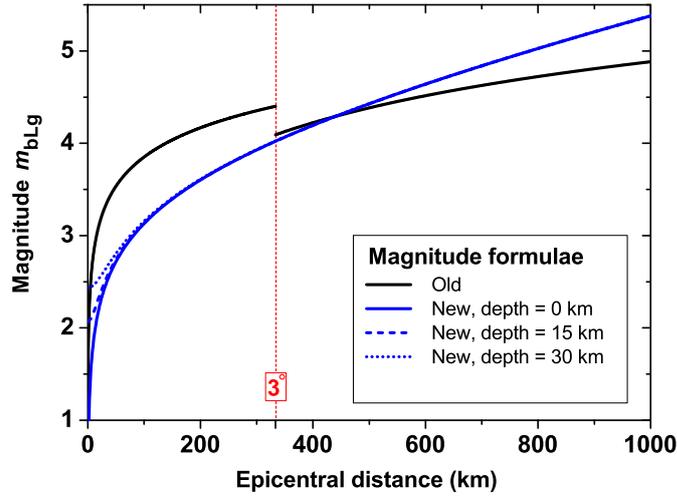


Figure 3.8: **Formulae used in the IGN catalogue to calculate the magnitude in m_{bLg} scale.** The older ones (Eqs. 3.5 and 3.6, Mezcuca & Martínez-Solares, 1983) had a break at an epicentral distance of 3° ($\simeq 334$ km). The new one (Eq. 3.7, López, 2008) is continuous and takes into account the hypocentral depth, though this effect is noticeable only for small epicentral distances. In the example shown, it is assumed that the peak ground amplitude (in microns) equals the corresponding wave period (in seconds). Other amplitude/period ratios would shift the curves vertically, but would maintain their relative differences.

for all m_{bLg} calculations in the catalogue since 1 March 2002:

$$m_{bLg} = \log\left(\frac{A}{T}\right) + 1.17 \log R + 0.0012R + 0.67, \quad (3.7)$$

where A and T are the same as in Eqs. 3.5 and 3.6. R is the hypocentral distance in km:

$$R = \sqrt{\Delta^2 + h^2}, \quad (3.8)$$

where Δ is the epicentral distance and h the hypocentral depth, both in km.

Eq. 3.7 is substantially different from the older m_{bLg} formulae, and implies a stretch of the magnitude range (Fig. 3.8). If an earthquake originates near most stations (at $\Delta < 3^\circ$) its magnitude will tend to be lower than the one calculated by the old formulae. The opposite effect will result if the earthquake source is far from most stations (at $D \gtrsim 440$ km of epicentral distance).

Cabañas *et al.* (2015) calculated a regression between this m_{bLg} and M_w . The mean standard error reported for this magnitude scale in the catalogue is $\simeq 0.2$.

3.4.4 m_b (1997–present)

Since November 1997 in the Iberian region, and since November 2003 in the Canary Islands, for those earthquakes in which the Lg wave train is not apparent (in principle for those whose rays have oceanic paths $\geq 2^\circ$, Carreño, 1999), and for those with hypocentral depth > 30 km (Cabañas *et al.*, 2015), m_b , a body-wave magnitude calculated according to Veith & Clawson (1972) is used:

$$m_b = \log \left(\frac{A}{T} \right) + P(\Delta^\circ, h) \quad (3.9)$$

where A is the maximum amplitude of a P type phase (Pn or P) in nm (Carreño, 1999); T is the corresponding period in seconds; and $P(\Delta^\circ, h)$ is a correction factor (Veith & Clawson, 1972) which depends on the epicentral distance Δ° in degrees and the hypocentral depth h in km. This magnitude was calibrated for short-period vertical-component seismograms (Veith & Clawson, 1972) although it is also being used with the seismograms from broadband stations (Martínez Solares *et al.*, 2013). Cabañas *et al.* (2015) calculated a regression between this scale and M_w , and noted that this m_b is systematically higher than M_w . Regressions to other regional magnitude scales were proposed by Mouayn (2007) and Peláez *et al.* (2012).

Between November 1997 and February 1998, m_b was used for few earthquakes only, while m_{bLg} was used for most. Almost all magnitudes during this period are labelled indistinctively as m_b , and the actual type of magnitude has to be deduced by checking, in the full-format catalogue, the type of phase used for magnitude determination. The mean standard error reported for m_b from March 1998 (when correct magnitude labelling starts) to December 2013 is $\simeq 0.2$.

Fig. 3.9 shows the geographical distribution of epicenters from March 2002 to the end of 2013 where m_b is the preferred magnitude estimate. Indeed most of them are located far from the Iberian Peninsula, Canary Islands or Balearic Islands (so the Lg waves had a relatively long path through thin crust before reaching the Spanish stations) and/or had deeper hypocenters (compare with the distribution of seismicity in Fig. 3.1).

3.4.5 M_w (2002–present)

For earthquakes with $m_{bLg} \gtrsim 3.5$ occurred since 2002, the moment magnitude, M_w (Hanks & Kanamori, 1979), and moment tensor are calculated (Rueda & Mezcua, 2005) with Dreger & Helmberger's method (1993) and reported in the abridged IGN catalogue (not in the full-format one). The detailed moment tensor results are reported in a separate catalogue online (Instituto Geográfico Nacional, 2015). The IGN catalogue also reports M_w for a few earlier earthquakes, taken from the

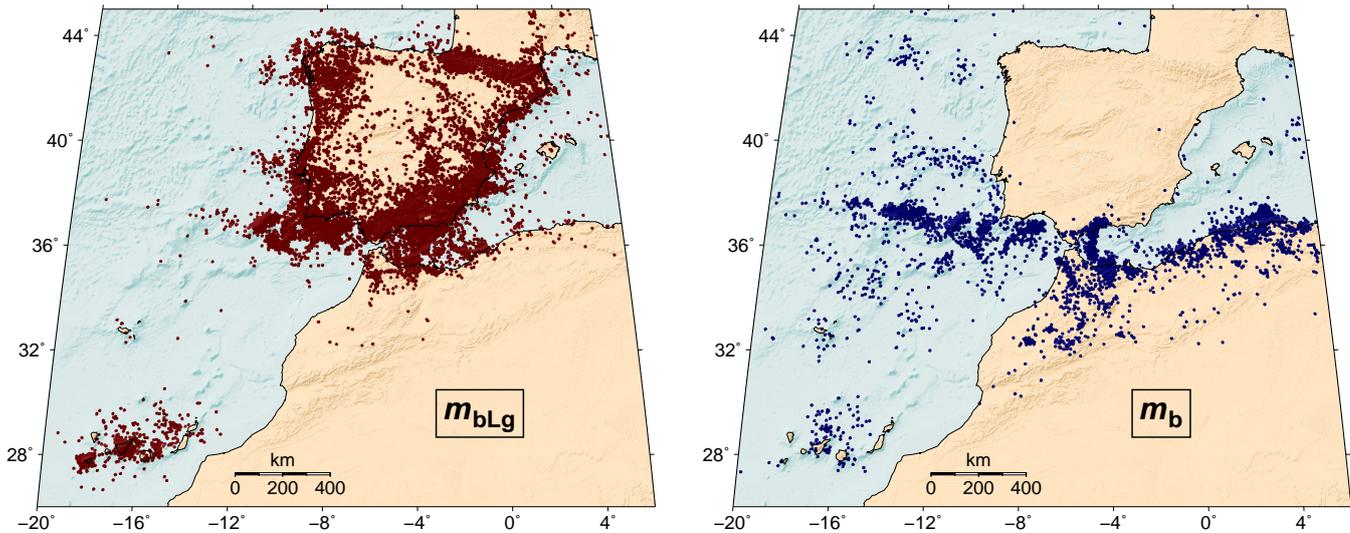


Figure 3.9: **Geographic distribution of epicentres from March 2002 to the end of 2013, separated by their preferred magnitude type** (the one reported in the abridged catalogue online). *Left: m_{bLg} . Right: m_b .* Only few earthquakes had M_w as the preferred magnitude during this period (not shown; see maps in Cabañas *et al.*, 2015).

literature (e.g. Batlló *et al.*, 2008). Complementary M_w and moment tensor catalogs in the region are those of Stich *et al.* (2010), Martín *et al.* (2015) and references therein. The M_w values determined by IGN are usually similar to those by other authors, when available (e.g. Chevrot *et al.*, 2011; Cabañas *et al.*, 2015).

3.4.6 Evolution of magnitude values in the catalogue

The overall pattern of the magnitudes throughout time is shown in Fig. 3.10. A running median magnitude (see e.g. Gentili *et al.*, 2011, for a similar graph) is calculated for each of the two regions analysed. For both there is a decreasing overall trend, correlative with the general improvement of the network, which allowed detecting and recording smaller earthquakes. Specific changes in this trend are commented and interpreted below.

In the Iberian region, at the beginning of November 1997, the median magnitude dropped by ~ 0.5 units. This effect cannot be due to the introduction of m_b since initially almost all earthquakes still used m_{bLg} . Moreover, m_b is typically higher than m_{bLg} (see Cabañas *et al.*, 2015), so the opposite effect in the median magnitude would be expected. This sudden decrease must be due to the increased reporting of smaller earthquakes thanks to the new automatic procedures for detection and location then implemented (Sec. 3.2.10), and the resulting increase in the number of

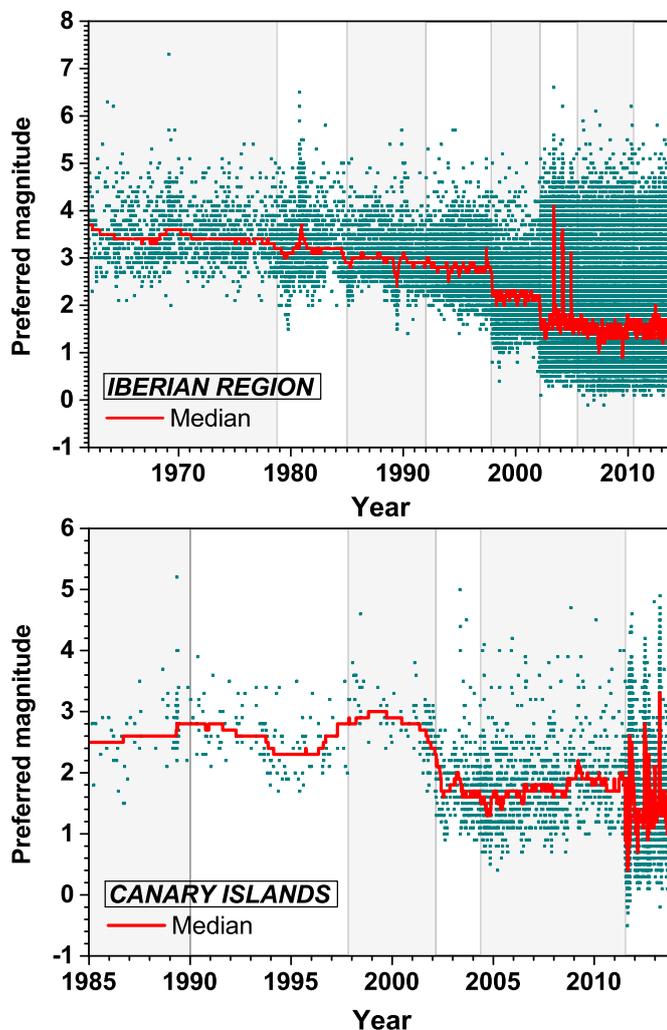


Figure 3.10: Preferred earthquake magnitudes (*i.e.* those from the abridged catalogue version, Sec. 3.2.14) and their running median, for each region analysed (Figs. 3.12 and 3.14). In the Iberian region, the running value of the median magnitude is calculated with 200 earthquakes. In the Canary Islands, due to the scarcity of earthquakes before 2011, it is calculated using 50 earthquakes. *Vertical fringes* indicate the periods considered for completeness assessment.

phases used for location (Sec. 3.3.1). This would also explain the slight increase in the rate of earthquakes detected since that date (Fig. 3.2).

In March 2002, the introduction of the new m_{bLg} formula caused two effects (see Gulia, 2012, for similar issues in other catalogues). First, as expected (Sec. 3.4.3), it stretched the magnitude range reported. Second, it caused a drop in the median magnitude, because most earthquakes occurred close to the stations and were assigned lower magnitude values than with the old formula (Fig. 3.8). Both effects may have been intensified by the increasing ability of the broadband network to detect small earthquakes nearby and moderate earthquakes far away.

Brief, transient decreases of the running median are due to the recording of earthquakes with particularly low magnitudes close to seismic stations. For example, the swarms of low-magnitude earthquakes in Loja (1985) and Villanueva de Algaidas - Antequera (1989) took place in locations with relatively dense monitoring (Carreño *et al.*, 1991a; Posadas *et al.*, 1993), leading to a global decrease of the running median magnitude.

Conversely, distant sequences where most small earthquakes are undetected tend to cause sharp increases of this running median. The clearest examples are due to large earthquakes in northern Africa and their aftershock sequences, such as: 10 October 1980 El Asnam (now Chlef, Algeria), 21 May 2003 Zemmouri (Boumerdes, Algeria), 24 February 2004 Al Hoceima (Morocco), as well as abundant moderate earthquakes south of Nador (Morocco) starting in late 2004.

In the Canary Islands, very few magnitude values are reported before 1985 (not shown in Fig. 3.10). In November 1997 there was an increase in the reported magnitude values, correlative to the replacement of the M_D scale by the old m_{bLg} . A clear decrease of about a magnitude unit in the early 2000's must be due to the deployment of the first broadband stations and the use of the new m_{bLg} scale since March 2002. From mid-2011 onwards, successive swarms at the island of El Hierro, monitored by a dense local network, have yielded recordings of numerous small earthquakes and a complicated pattern of the running magnitude median.

3.5 Magnitude of completeness

This section analyses the magnitude of completeness of the catalogue. It first reviews the concept, then describes the method used to calculate and map it in detail during the instrumental era, and finally describes the results separately for both the Iberian region and the Canary Islands.

3.5.1 Definition and influencing factors

The magnitude of completeness (M_c) is the lowest magnitude at which all earthquakes are detected and included in the catalogue (e.g. Rydelek & Sacks, 1989; Wiemer & Wyss, 2000; Woessner & Wiemer, 2005; Mignan & Woessner, 2012). Earthquake detection depends on natural and artificial factors, and requires identifying the signal above the noise level at a minimum number of stations.

Most importantly, because of seismic wave attenuation, M_c at a given point in the Earth increases with the epicentral distance to the network stations and with depth (e.g. Wiemer & Wyss, 2000; Schorlemmer & Woessner, 2008; Schorlemmer *et al.*, 2010b). In turn, higher instrumentation sensitivity and/or signal-to-noise ratios tend to favour earthquake detection and thus lower M_c . Dense, non-permanent local networks reduce M_c locally and temporarily (e.g. Schorlemmer & Woessner, 2008; Häge & Joswig, 2009a,b). Reduction of M_c may also be achieved by improving signal processing and earthquake location techniques, such as using waveform cross-correlation (e.g. Schaff & Waldhauser, 2010; Domínguez Cerdeña *et al.*, 2011).

More earthquakes are typically recorded during nighttime and weekends; this effect has been attributed to a lower M_c due to the lower artificial noise during such periods (e.g. Rydelek & Sacks, 1989; Zotov, 2007; Atef, 2009; Iwata, 2013). Seasonal variations of the noise level (e.g. Burtin *et al.*, 2008; Hillers & Ben-Zion, 2011; Custódio *et al.*, 2014, and references therein) may also cause M_c to vary systematically between different seasons (Ringdal & Bungum, 1977; Custódio *et al.*, 2015).

In swarms and aftershock sequences, M_c rises; numerous events are missed because their waveforms are masked by the coda and the overlapping wave arrivals of previous ones (e.g. Wiemer & Wyss, 2000; Kagan, 2004; Kilb *et al.*, 2007; Nanjo *et al.*, 2007). This effect even obstructs global earthquake detection after a large earthquake (e.g. Iwata, 2008); for a particular example see Heimann *et al.* (2013) and Tausin *et al.* (2013). Special techniques of waveform analysis are required to identify some of the missing aftershocks (e.g. Peng *et al.*, 2007; Peng & Zhao, 2009).

Changes in magnitude reporting will obviously also change M_c (e.g. Tormann *et al.*, 2010). And, finally, an artificially high M_c may result if earthquakes below a magnitude threshold are detected but purposely excluded from the catalogue.

All these factors imply that catalogue completeness is heterogenous in space and time, and that M_c may be only stable (and meaningful to calculate) in small enough regions and short enough periods, with a sufficiently stable network and homogenous magnitude reporting (e.g. Hutton *et al.*, 2010).

For a given network, a theoretical detection threshold (minimum magnitude which can be reliably detected) can be calculated and mapped for each location, considering the network geometry, the noise level at each station, the minimum number of stations required to detect and locate an earthquake, and the attenuation of signal amplitude with distance. For the Spanish National Seismic Network, Rueda & Mezcua (2006) provided a software to calculate such a threshold based on the detection of the Lg wave, and D’Alessandro *et al.* (2013) applied a more complex simulation-based method. But the IGN catalogue is not only compiled using the real-time backbone network (Figs. 3.4 and 3.14). Stations from other networks are also eventually considered for calculating final earthquake locations (Sec. 3.2.14). So theoretical detection thresholds are relevant to real-time network capabilities, but may be higher than the magnitude of completeness of the expanded and revised catalogue.

3.5.2 Method used to calculate M_c

Here, M_c was calculated and mapped in successive periods using only the earthquake catalogue itself, as detailed below. The Iberian Region and the Canary Islands were analysed separately, considering the catalogue differences between them.

The seismicity generated by the fault network in a rock volume commonly follows an exponential magnitude-frequency distribution (Ishimoto & Iida, 1939; Gutenberg & Richter, 1944):

$$\log_{10}[N(\geq M)] = a - bM, \text{ for } M \geq M_c. \quad (3.10)$$

where $N(\geq M)$ is the number of earthquakes with magnitude $\geq M$, and a and b are parameters. This relation has been observed to extend down at least to $M_c = -4.4$ (Boettcher *et al.*, 2009; Kwiatek *et al.*, 2010), and even to lower values in laboratory experiments of rock deformation (e.g. Yoshimitsu *et al.*, 2014).

For $M < M_c$, $N(M)$ is lower than expected from the extrapolation of Eq. 3.10. This departure can be blamed on catalogue incompleteness (e.g. Båth, 1981; Kwiatek *et al.*, 2010), since small-magnitude earthquakes are less likely to be detected (e.g. Schorlemmer & Woessner, 2008) and thus are preferentially missing. Several objective methods have been proposed to calculate the precise magnitude at which such a departure begins, albeit they do not always agree with each other (e.g. Mignan & Woessner, 2012; Michael, 2014; Roberts *et al.*, 2015).

Here the Entire Magnitude Range method (EMR) was chosen, because of its proven reliability with small samples (as low as 60 earthquakes, Woessner & Wiemer, 2005). It is available in the free software ZMAP (Wiemer, 2001) and

R (Mignan & Woessner, 2012). This method is computationally intensive, and calculates M_c by fitting the entire range of the observed magnitude-frequency distribution (both the complete and the incomplete part). For $M < M_c$, a cumulative normal (Gaussian) distribution function is used, with mean μ and standard deviation σ (as also proposed by Ringdal, 1975; Ogata & Katsura, 1993). For $M \geq M_c$, Eq. 3.10 is fitted using a robust maximum-likelihood technique (Aki, 1965; Utsu, 1965). The latter, unlike the least-squares method (Sandri & Marzocchi, 2007), is barely sensitive to possible departures from the exponential distribution of the few largest earthquakes in the range observed (e.g. Fig. 3.11). The most likely value of M_c is provided by the best fitting model, which maximizes the log-likelihood function for the four parameters: μ , σ , a and b . The EMR method is somewhat conservative, as it may overestimate M_c in some cases (Mignan, 2012). For caution, I should note that for determining b reliably, a larger sample of earthquakes would be required (preferably at least 200 with $M \geq M_c$, Schorlemmer *et al.*, 2005; Roberts *et al.*, 2015); additional insight on fitting Eq. 3.10 and related distributions was provided by Deluca & Corral (2013).

To calculate M_c and its uncertainty in a robust way, bootstrap resampling (Efron, 1979) was used, following Woessner & Wiemer (2005). Each sample of earthquakes was resampled with replacement 200 times, and M_c was calculated for each resulting bootstrap sample. This figure is high enough to obtain a stable average M_c and its standard deviation δM_c (Woessner & Wiemer, 2005), exemplified in Fig. 3.11. The presence of quarry blasts in the sample (Cattaneo *et al.*, 2014), changes in the b -value (e.g. Cesca *et al.*, 2014), or the mixture of magnitude scales may worsen the fit quality and increase δM_c .

3.5.3 Spatial sampling

To account for spatial heterogeneity, here M_c was mapped locally, using a variable resolution, so that the resulting map is more detailed in the areas where more earthquakes were recorded. A latitude-longitude grid of nodes was used, following Wyss *et al.* (1999) and e.g. Wiemer & Schorlemmer (2007) and Nanjo *et al.* (2010). The grid spacing was chosen as 0.1° both in latitude and longitude. The resulting distance between nodes is constant along the meridians ($\simeq 11.1$ km), and variable with latitude along the E-W direction as the meridians converge towards the poles (from $\simeq 10.0$ to $\simeq 7.9$ km for the whole latitudinal range of the catalogue). In this way the grid spacing is on the order of the epicentral location uncertainties (Sec. 3.3.2), so a tighter spacing would not be meaningful.

For each node, M_c and δM_c were calculated using its closest earthquakes, at least 60. Distances were measured along the great-circles between the nodes and

the epicentres, using a spherical model for the Earth, with radius 6371.007 km, the sphere of equal area of the most frequently used Earth reference ellipsoids (Moritz, 2000; NIMA, 2000). Initially, a search radius of 5 km around each node was considered (so that the search area is similar in size to the latitude-longitude cells of the grid). If there were < 60 epicentres within this radius, it was increased until it included the 60 closest earthquakes, but up to a maximum of 100 km. This limit attempts to avoid mixing earthquakes from regions too distant to each other, with different M_c (e.g. Nanjo *et al.*, 2010) and to minimize the artifact of calculating it in areas devoid of earthquakes. For most neighbouring nodes, the areas used for determining M_c will partially overlap, leading to spatial smoothing.

In the resulting map, the blank regions have too few earthquakes recorded per unit area to calculate a meaningful M_c (< 60 within 100 km from a node). Compared with the other regions, they have either an intrinsically lower seismicity rate, or a higher M_c , or both. For some regions, M_c could be calculated in periods with high seismicity rate (for example, with aftershock sequences) but not in others. It might be possible to extrapolate M_c to these blank areas using different procedures, which consider how it increases with distance from the stations (Schorlemmer & Woessner, 2008; Mignan, 2012). To properly use such methods it would be necessary to know (or to deduce, e.g. Gentili *et al.*, 2011) when each station was in operation and contributing data to the catalogue. In the IGN catalogue this would be possible for the stations of the backbone network, but becomes exceedingly complicated for the additional stations from other institutions.

Since depths are often poorly determined, or not determined independently at all (Sec. 3.3.3), I chose not to set a fixed maximum depth threshold. Only the five earthquakes with depth > 600 km (Sec. 3.3.3 and Fig. 3.7) were excluded from the analysis. This is a conservative approach: deep earthquakes are less likely to be detected than shallow ones, so the M_c calculated from a sample with mixed depths will be higher than if only the shallow earthquakes were considered. In most areas the catalogue is dominated by shallow seismicity anyway (Fig. 3.1 and Sec. 3.3.3), so except for specific locations discussed later, depth is probably a minor factor compared to the epicentral distance from the stations.

3.5.4 Choice of temporal periods and magnitude scales

To account for temporal heterogeneities in M_c , the catalogue was divided into periods. Ideally, these should be long enough to include many earthquakes, for calculating M_c in as many locations as possible, and reducing the extent of blank areas. But they should be short enough to ensure that M_c was stable, with a relatively stable network (e.g. Hutton *et al.*, 2010) and without changes in magnitude

reporting (e.g. Tormann *et al.*, 2010). Using too long periods with unstable completeness would yield underestimated M_c values, because the departure from Eq. 3.10 is controlled by the statistics of the most complete parts of the catalogue (e.g. Michael, 2014).

Periods were separated by milestones in the development of the backbone network, improvements in the earthquake detection routines, and changes of the magnitude formulae used. The evolution of other networks which contributed to the catalogue (either temporary ones or from other institutions) had to be disregarded to avoid exceeding complexity.

The preferred magnitude listed by IGN for each earthquake in the abridged catalogue was used. M_c at each node depends on the magnitude scale(s) used locally (e.g. Fig. 3.9). In the full-format catalogue, only one magnitude type is typically reported, so separating earthquakes by magnitude type would indeed create subsets less complete than the original catalogue. Thus, it is not statistically meaningful (although it would be more physically meaningful) to calculate M_c separately for each magnitude type. Reliable magnitude conversion formulae for this catalogue (Cabañas *et al.*, 2015, and references therein) are not available for $m \lesssim 3$, while it was expected that $M_c < 3$ in most of the region, so I used the catalogue as it is originally provided, without attempting to convert magnitudes to a common scale.

Assessing the reliability (e.g. Martínez Solares & Mezcua, 2002; Udías, 2015) and completeness (e.g. Albarello *et al.*, 2001) of historical earthquake catalogues is particularly challenging, so it will not be attempted here. Occasionally, the review of even relatively recent historical earthquakes in the Iberian region has substantially modified the location and intensity estimates (e.g. Ribeiro *et al.*, 2015). Determining earthquake location and size of historical events is especially difficult when originated offshore or in coastal areas (even great ones, e.g. Martínez-Solares & López Arroyo, 2004; Hough, 2013) and in scarcely populated regions. Also, aftershocks are very irregularly accounted for in the historical chronicles (e.g. Vere-Jones, 1992).

Given the catalogue differences between the Iberian Region and the Canary Islands, I will analyse them separately in the subsections below, separating periods with similar instrumentation, magnitude reporting and detection and location procedures during the instrumental era. The evolution of other networks which contributed to the catalogue (either temporary ones or from other institutions) had to be disregarded to avoid exceeding complexity.

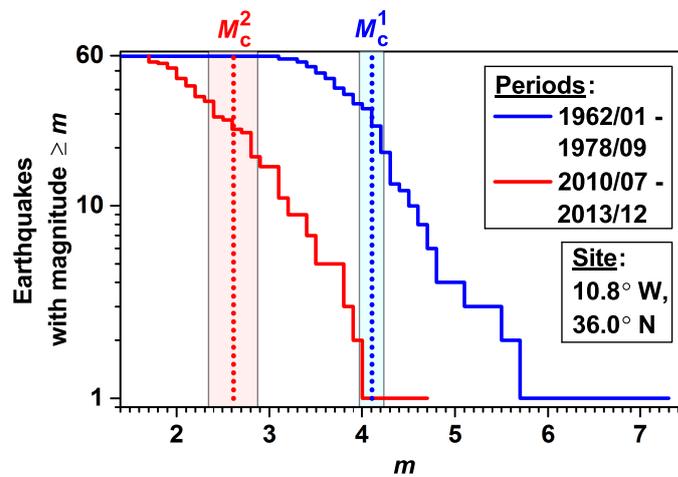


Figure 3.11: **Example of magnitude of completeness (M_c) and its standard deviation (δM_c) for a site in two different periods.** This site is the closest node in the map to the largest earthquake of the catalogue in the instrumental era (28 February 1969, offshore Portugal, $m_{bLg} = 7.3$). *Solid lines*: Magnitude-frequency distributions of the closest 60 earthquakes to the node in the periods indicated. *Dotted lines*: M_c calculated automatically with the EMR method (Woessner & Wiemer, 2005). *Colour bands*: Range of M_c plus/minus the δM_c calculated by bootstrap. The distribution is more gradually curved in the second period, leading to a larger δM_c . These values of M_c and δM_c are mapped at this node in the corresponding periods in Figs. 3.12 and 3.13 respectively.

3.5.5 M_c in the Iberian Peninsula and surroundings

According to Mezcua *et al.* (2011); Martínez Solares *et al.* (2013), the catalogue may be complete in the Iberian Peninsula for $M \gtrsim 5$ since the early XIX century, and for $M \gtrsim 4$ since the mid-XX century, being less complete offshore, in northern Africa, and in earlier times. The reader is referred to these works and references therein for further details. These assessments estimated the missing magnitudes from macroseismic intensity, and assumed that, for a given magnitude threshold, the catalogue is complete from the moment at which the earthquake rate from that threshold upwards becomes stationary (following Stepp, 1972).

The analysis made here covers the years 1962 onwards, for which the catalogue reports magnitude systematically for most events. The increasing slope of the cumulative number of earthquakes (Fig. 3.2) and the overall decrease of the median magnitude (Fig. 3.10) already evidence that M_c in the catalogue has been reduced throughout the decades. Figure 3.4 shows maps with the backbone stations in each period, while M_c and δM_c are mapped in Figs. 3.12 and 3.13 respectively.

During the first four periods considered (January 1962 to October 1997) a single m_{bLg} scale was used; relatively few earthquakes were recorded (Fig. 3.2) so M_c can be reliably determined only in the most seismically active areas.

The first period analysed is chosen to include from January 1962 (for which the m_{bLg} scale started to be used, Mezcua & Martínez-Solares, 1983) to September 1978, and during this time only a few IGN observatories operated, including the WWSSN stations of Toledo and Málaga (these from late April 1962 onwards, International Seismological Centre, 2015a). M_c was 3.0–3.7 in the Alborán Sea (the westernmost corner of the Mediterranean) and its surrounding mountain ranges (Betics to the north, Rif to the south), while it was 3.2–3.5 in the Pyrenees. In the Atlantic Ocean, the location of the largest earthquake in the instrumental era, occurred in 1969, has the highest M_c determined in this period, $\simeq 4.1$ (see also Fig. 3.11).

During the initial deployment of the short-period, telemetred, Spanish National Seismic Network (October 1978 to December 1984) M_c improved slightly, but was still > 3 in the Iberian Peninsula. In NW Algeria, the catalogue is dominated by the 1980 El Asnam (now called Chlef) earthquake and its aftershock sequence, which were recorded with $M_c \simeq 4.5$.

During the first years of the new network (January 1985 to December 1991), and with an increasing number of stations, M_c decreased to < 3 in all areas where calculable in the Iberian Peninsula and immediate surroundings, reaching down to 2.4–2.6 at and around the location of the 1985 Loja swarm in the Betics, which was monitored by IGN with temporary stations (Carreño *et al.*, 1991a). This swarm

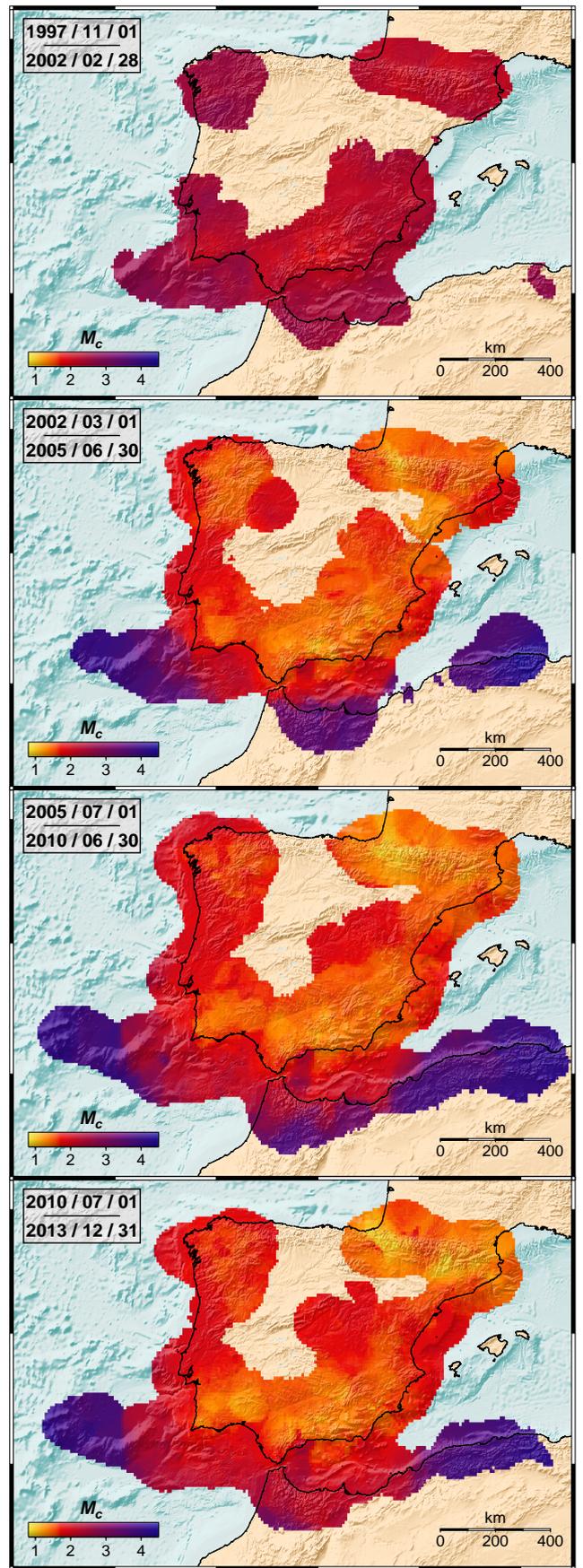
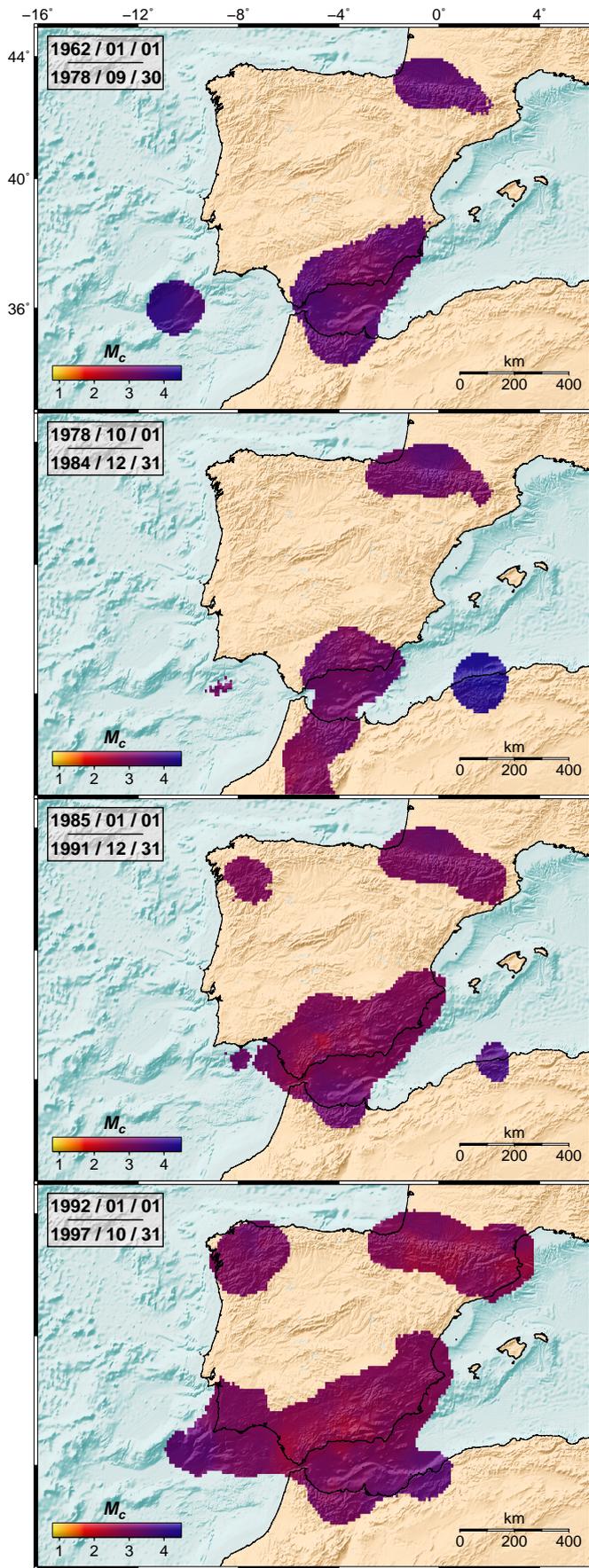
and others during this period caused a decrease of the median magnitude in the catalogue (Fig. 3.10 and Sec. 3.4.6). In the periphery of the network, such as on the northern side of the Pyrenees and in the Rif, M_c was still 3.0–3.3. Meanwhile, M_c dropped to ~ 3.7 at the same spot in NW Algeria mentioned in the previous period.

The catalogue is complete in the most seismic areas of the Peninsula down to $M_c \simeq 3$ since 1992. At that time, the first automatic location procedure was fully in operation (which also corresponds to a different pattern in depth reporting, Fig. 3.7 and Sec. 3.3.3), and since late 1992 the Sonseca Array started providing data to the catalogue. Between January 1992 and October 1997, the best monitored areas continued to be in the SE of the peninsula, M_c reaching 2.3 locally, while M_c increased according to distance from the network, reaching 3.5 offshore to the SW of the peninsula, and in sites of northern Africa.

A major improvement of the catalogue completeness is observed from November 1997, when the new automatic location procedures were implemented. This was also evidenced by a sudden drop in the median magnitude (Fig. 3.10 and Sec. 3.4.6). Anywhere in the peninsula (where measurable) $M_c \leq 2.8$, and reached down to 1.7 precisely at the site with the densest monitoring, in the Granada region (Betics). Even the values calculated offshore and in northern Africa were $\lesssim 2.8$, which is surprisingly low. This may be attributed to the underestimation of m_{bLg} for distant offshore earthquakes (Sec. 3.4.3); m_b had started to come into use, but during this period only for $< 1\%$ of earthquakes, so with very slight influence on M_c .

The most dramatic drop in M_c took place since March 2002, which also corresponds to a drop in median magnitude (Fig. 3.10 and Sec. 3.4.6), and to the most remarkable increase in the apparent frequency of earthquakes with $M \leq 3$ (Fig. 3.2). This must be due both to the deployment of the new broadband network and to the use of the new, improved m_{bLg} formula, which provides lower magnitude values for earthquakes at epicentral distances $\lesssim 440$ km from the station (Secs. 3.4.3 and 3.4.6). Between March 2002 and June 2005, $M_c \lesssim 2.0$ over the whole Iberian Peninsula, and typically $\lesssim 1.5$ in the most densely monitored areas, a pattern which continues along the broadband era. Despite the network improvement, M_c rose with respect to the previous period in locations far offshore and in regions in northern Africa, which is explained by the systematic use of m_b for distant

Figure 3.12: **Evolution of the magnitude of completeness (M_c) of the catalogue in the Iberian Peninsula and surroundings**, mapped in cylindrical equal-area projection, for the periods indicated. The corresponding maps of uncertainties are shown in Fig. 3.13.



earthquakes, that tends to provide higher magnitudes (this also occurs with the new m_{bLg} scale compared with the older one at distances $\gtrsim 440$ km).

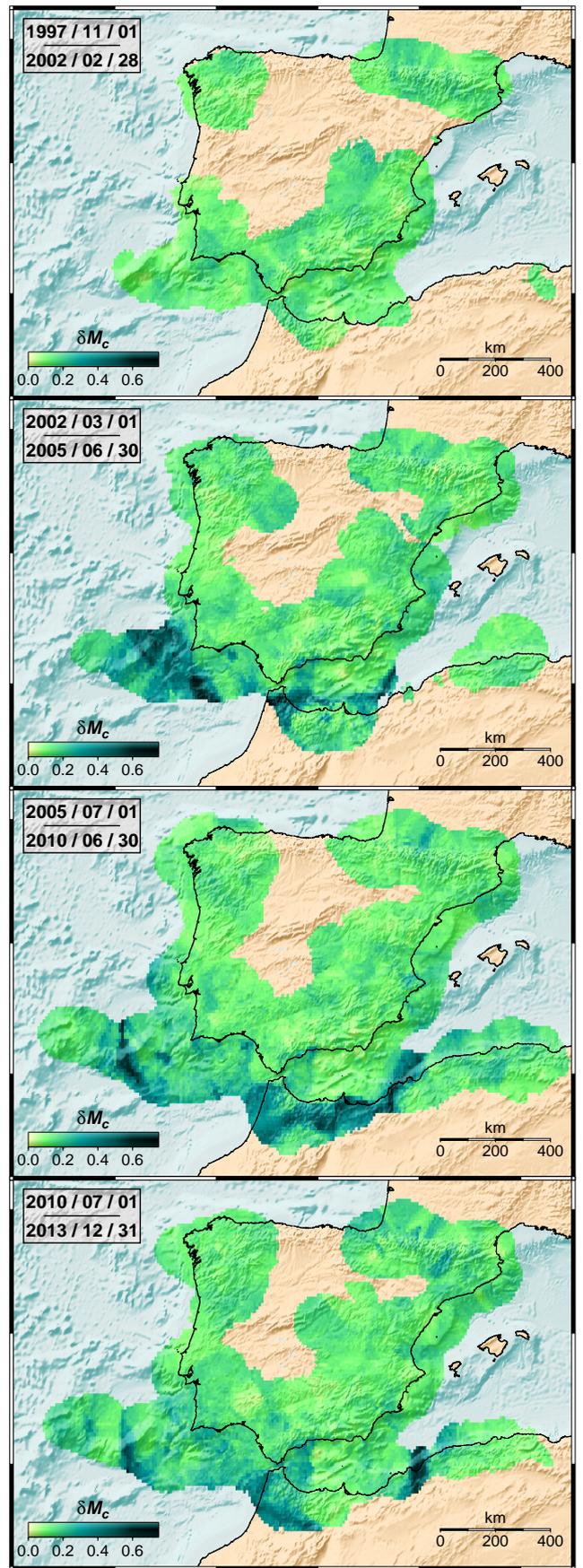
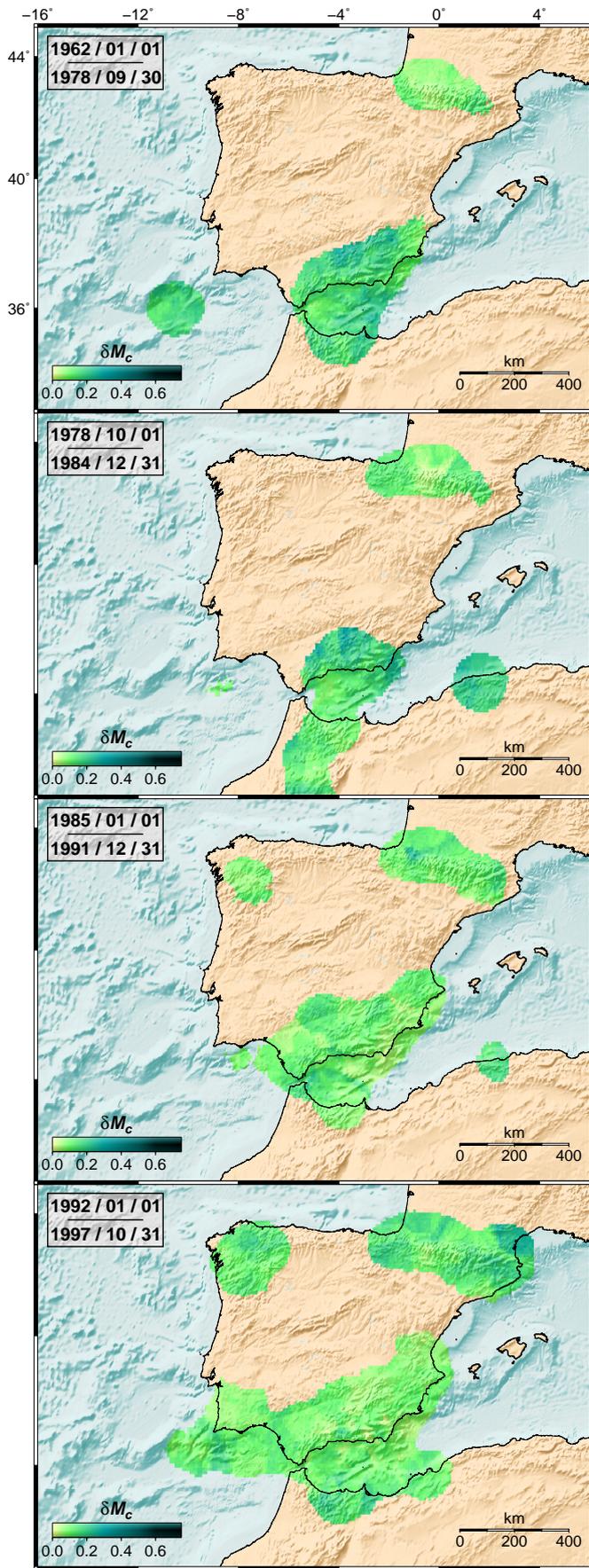
Between July 2005 and June 2010 the broadband network continued to grow. Completeness improved noticeably in offshore areas. The greatest drop of M_c , down to $\simeq 1.0$ in the western Pyrenees, was due to the deployment of the dense network around the Itoiz and Yesa reservoirs (Sec. 3.2.12). Regions exist however, where stations were closed down (Fig. 3.4) and M_c worsened nearby, especially in Galicia (NW Iberian Peninsula) and other areas in the interior of the Peninsula.

During the last period analysed (July 2010 to December 2013) the completeness is similar to the previous one, although the progressive closing of almost all remaining short-period stations of the backbone network explain local deteriorations of completeness (such as around 3°W , 39°N).

The uncertainty of the results, δM_c (Fig. 3.13), is remarkably low (typically ≤ 0.2 , that is, in the order of the magnitude uncertainties themselves, Sec. 3.4). But a few exceptions during the broadband era require additional explanation. Offshore and in Northern Africa, a sharp transition from $M_c \simeq 2.5$ to $\gtrsim 3.5$ is measured, associated to the highest δM_c values (≥ 0.4). This must be due to the use of different magnitude scales (Fig. 3.9 and Sec. 3.4), with m_b values for the more distant earthquakes being systematically higher than the m_{bLg} values of similar, closer earthquakes (because of overestimation of m_b , Cabañas *et al.*, 2015). So in the spatial fringes where the mixture of magnitudes takes place, the measured M_c changes abruptly, the theoretical fit to the magnitude-frequency distribution is poorer and with higher δM_c . In the Alborán Sea, a N-S area has higher M_c than the surroundings, and with relatively high $\delta M_c \sim 0.3$. Relatively abundant intermediate-depth earthquakes (plus shallow ones) take place here (Buforn *et al.*, 1997). So, consequently, a local mixture of M_c ensues: the deeper earthquakes are being less completely recorded (and measured in m_b scale, both factors raising the overall M_c), while the shallower ones are more completely recorded (and in m_{bLg} scale). In other offshore areas, in the Atlantic, intermediate-depth earthquakes also take place, and a similar situation probably occurs, though less spatially localized.

In Peninsular Spain, within each period, M_c generally correlates with the distance to the stations shown in Fig. 3.4, and largely agrees very well with the theoretical detection thresholds of the Spanish National Seismic Network for specific dates. Maps of these thresholds, elaborated by IGN personnel and apparently based on the method implemented by Rueda & Mezcua (2006), are

Figure 3.13: Standard deviation (δM_c) of the magnitude of completeness M_c of the catalogue mapped in Fig. 3.12.



available for the network configuration ca. 1990 (Fig. 16 of Mezcua, 1995), 1993 (Fig. 17 of Mezcua, 1995), 2002 (Fig. 3 of Carreño *et al.*, 2003) and 2010 (Rueda, 2010, which does not include the dense network in the Aragón Valley). The most recent theoretical analysis of the detection threshold was made by D'Alessandro *et al.* (2013), assuming that all the stations listed in their Table 1 were functioning simultaneously (Antonino D'Alessandro, pers. comm., 2013). That list, however, includes all stations ever operating in the backbone network, mixing stations that were opened up to mid-2010 with those already permanently closed. This prevents a meaningful comparison with such a detection threshold in most of the region. For the dense IGN network around the Itoiz and Yesa reservoirs, for which indeed the stations were simultaneously operating by mid-2010, the detection threshold (local magnitude $M_L \sim 1.0$) calculated by D'Alessandro *et al.* (2013) is like the M_c calculated there during the last two periods.

During the broadband era, M_c is lower than the expected detection thresholds (from Carreño *et al.*, 2003; Rueda, 2010) in the eastern Pyrenees (most notably since 2010), southern Portugal, several locations in Andalusia (southern Spain) and in NW Africa (with notable progressive improvements since 2005). These are areas where the catalogue truly benefits from the data from additional stations of other institutions within Spain and neighbouring countries, whose networks were also improving at the time. Conversely, M_c is higher than expected in the areas where m_b was predominantly used, because such theoretical detection thresholds were based on m_{bLg} .

3.5.6 M_c in the Canary Islands

In the Canary Islands the catalogue seems quite incomplete before the XIX century (e.g. Vinciguerra & Day, 2013), and even until the mid-1980's (only 67 earthquakes are listed from 1962 to 1984). An overview of the historical earthquakes was given by González de Vallejo (2006). The Canary catalogue until 1980 is taken from Mezcua & Martínez-Solares (1983), who warned that it is probably not valid for statistical analyses, since it relies on diverse partial sources (Galbis, 1932, 1940; Monge Montuno, 1981), and they had to complement the instrumental observations during 1960-1980 with the bulletin of the International Seismological Centre. Given those limitations, I have chosen to calculate M_c only from 1985 onwards.

For each period considered for completeness analysis, the stations of the Spanish National Seismic Network, M_c and δM_c are mapped in Fig. 3.14. Several of the milestones used for separating periods are different than for the Iberian Region, given that the Canary subnetwork had a separate temporal evolution and that (as it comprises few stations) M_c may be notably influenced by each station opening or

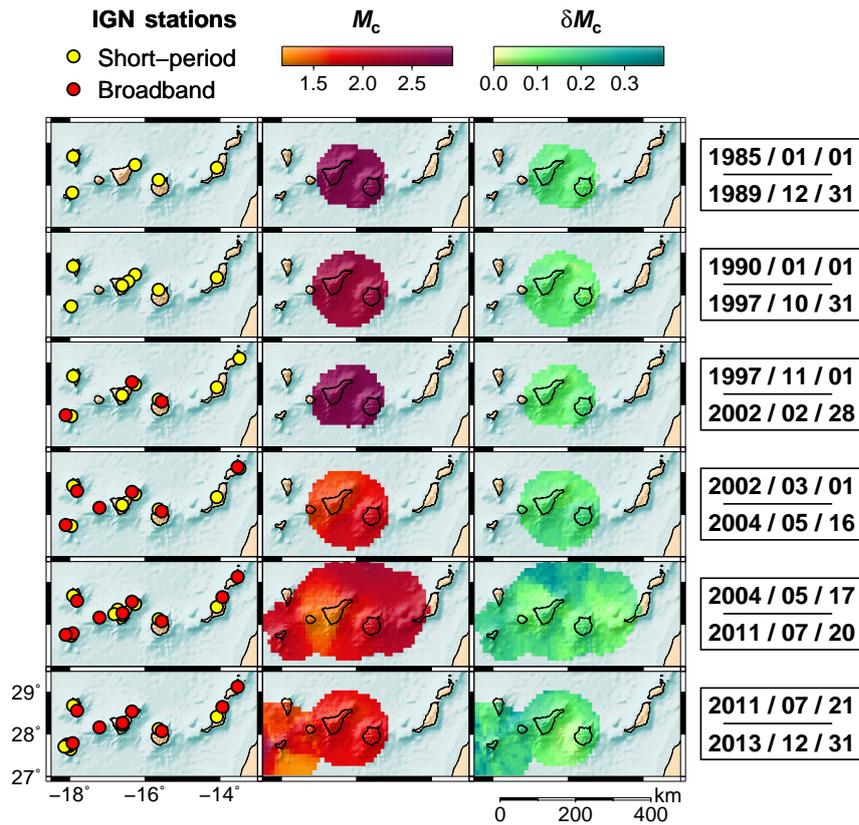


Figure 3.14: **Evolution of the network and catalogue completeness in the Canary Islands.** *Left:* Permanent IGN stations contributing to the catalogue, during the periods indicated. *Centre:* Magnitude of completeness (M_c). *Right:* Uncertainty of the magnitude of completeness δM_c . The colour scales are identical to those in Figs. 3.12 and 3.13.

closing. Before the seismic swarms below El Hierro island, most of the seismicity had taken place near the center of the archipelago, especially between the two largest islands, Tenerife and Gran Canaria. Until May 2004, M_c can be calculated with sufficient spatial detail only in this central area, while later on there are sufficient earthquakes for measuring it over a wider region.

During the early years of the short-period telemetred network (1985 to 1989), M_c (in M_D scale, Sec. 3.4.2) was 2.7–2.8. The new stations opened in Tenerife in 1990 (Sec. 3.2.7) lowered M_c there to 2.4–2.6 from January 1990 until October 1997. From November 1997 until February 2002, the median magnitude rose (Fig. 3.10) and M_c also rose to 2.7–2.9, despite the new automatic location procedures. This worsening is partially due to the change in magnitude scale (to the older m_{bLG}), and probably also due to the closing of one of the three short-period stations in Tenerife in January 1997, which left the island with only two operating stations for more than four years. Data from the first broadband stations during April 2001 to February 2002 reduced

the median magnitude (Fig. 3.10) and probably reduced M_c , but this could not be measured as too few events were recorded during this brief period.

As in the Iberian region, M_c has been greatly reduced during the broadband era. The introduction of the new m_{bLg} scale in March 2002 (Sec. 3.4.6) and the deployment of the initial broadband network reduced M_c down to 1.5–2.0 until mid-May 2004. The new stations deployed since this date in Tenerife (Sec. 3.2.13) further lowered M_c to ~ 1.3 just below this island, while it rose outwards from there up to 2.4 at the farthest locations from the network.

In the last period considered, the catalogue is dominated by the seismicity in El Hierro, thanks partly to the dense seismic network there deployed since 21 July 2011 (Sec. 3.2.13). M_c below and around El Hierro is spatially heterogenous, with the lowest values (~ 1.2) to the south of the island, probably because that was the location of the shallowest hypocentres, close to the submarine eruption site (e.g. Domínguez Cerdeña *et al.*, 2014). The apparent extension of these low M_c up to 100 km offshore SE of El Hierro, where actually no earthquakes were recorded, is an artifact of the spatial windowing procedure. West of El Hierro, where several swarms of more distant, deeper earthquakes took place after the eruption, is where the highest values (up to 2.9) are measured. During the swarms preceding the eruption, the earthquake foci migrated significantly with respect to the network, and new stations were opened (Domínguez Cerdeña *et al.*, 2014). These factors contribute to explain the temporal changes of M_c observed by Ibáñez *et al.* (2012) and Telesca *et al.* (2016). During the eruption, reported magnitudes were typically higher than before (e.g. Ibáñez *et al.*, 2012, their Fig. 3B). This may be, at least partially, a consequence of the volcanic tremor of variable amplitude which accompanied the eruption and which hampered the detection of the smallest earthquakes. All the mentioned factors bring about a sense of the complexity of determining a meaningful M_c in the El Hierro earthquake series (see also Roberts *et al.*, 2015). Meanwhile, two of the Tenerife stations were no longer in operation in this last period, what explains the increase of M_c there up to 1.6–1.8.

3.6 Daily and weekly variations

The seismic noise produced by human activities (vehicle traffic, machinery from industries, etc.) typically has daily and weekly cycles. It is higher during daytime than nighttime, and also higher on weekdays than at weekends (e.g. Bonnefoy-Claudet *et al.*, 2006; Díaz *et al.*, 2010; Custódio *et al.*, 2014). Even a relative decrease of seismic noise is sometimes observed during lunch breaks (e.g. Bonnefoy-Claudet *et al.*, 2006; Díaz *et al.*, 2010). As a result, since the ability to

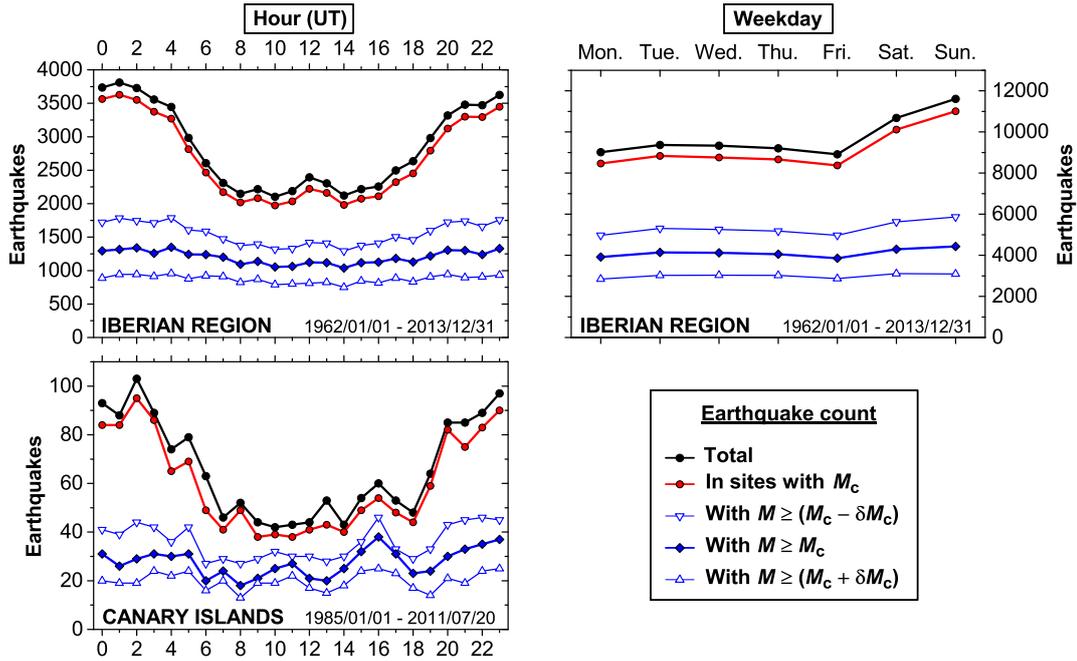


Figure 3.15: **Hourly and weekday earthquake frequencies in the Iberian region and Canary Islands** (areas of Figs. 3.12 and 3.14) for the periods indicated. The magnitude of completeness M_c and its uncertainty δM_c are those at the specific time and location of each earthquake. Solar noon is around 12 h Universal Time (UT) in the Iberian Peninsula and 11 h UT in the Canary Islands. Monday is shown as the first day of the week following the international standard (ISO, 2004).

detect and locate earthquakes decreases with noise amplitude, commonly more earthquakes are recorded in the catalogues during nighttime (e.g. Rydelek & Sacks, 1989; Iwata, 2013), lunch breaks (e.g. Custódio *et al.*, 2015) and weekends (Zotov, 2007; Atef, 2009).

Daily variations, with fewer detected events during the daytime hours attributable to higher artificial noise, are apparent in the catalogue (Fig. 3.15). This effect had already been reported by Carniel *et al.* (2008), Del Pin *et al.* (2008) and Domínguez Cerdeña *et al.* (2011) in the Canary Islands and by Custódio *et al.* (2015) in the Portuguese catalogue. The daily variation is less apparent during the period with volcano-tectonic episodes in El Hierro (not plotted in Fig. 3.15), indicating that recorded earthquake frequencies may then have been mostly dominated by natural, temporal variations of the seismicity rate. I made attempts at mapping M_c separating daytime and nighttime events, which showed a tendency for smaller M_c during nighttime, but splitting the catalogue into two such parts for each period severely limited the spatial resolution of the analysis.

A slightly higher frequency of earthquakes is recorded around noon in the Iberian region. This could be attributed either to the effect of catalogue contamination by blasts detonated precisely during the lunch break (Sec. 3.7, see also Gulia, 2012), and/or to a lower M_c thanks to the lower artificial noise at lunchtime, as proposed for the Portuguese catalogue (Custódio *et al.*, 2015).

A weekend effect is clearly observed in the Iberian region, with significantly more earthquakes being recorded on Saturdays, and especially on Sundays (Fig. 3.15). In the Canary Islands it is not observed (not plotted in Fig. 3.15 for simplicity), which might mean that artificial noise has a similar amplitude throughout the whole week.

Such cyclic effects should be smoothed out if only the complete part of the catalogue is considered, above a completeness threshold (e.g. Rydelek & Sacks, 1989). To test this (and thus indirectly the reliability of the completeness analysis), I considered for each earthquake the local M_c and δM_c at the particular latitude-longitude cell and specific period at which each earthquake occurred (Fig. 3.15). Most earthquakes in the instrumental era took place in sites where completeness had been determined. If only earthquakes with $M \geq M_c - \delta M_c$ are considered (that is, a low, optimistic, threshold), the daily and weekly variations are greatly reduced, but do not disappear completely, so the catalogue is not complete for that threshold. For $M \geq M_c$, these variations seem to vanish completely, and the remaining variability in earthquake frequency is probably real. The pattern is very similar if a higher, more conservative completeness threshold, $M \geq M_c + \delta M_c$, is used. These results support the completeness analysis, evidencing that the calculated M_c is the minimum threshold of completeness of the catalogue, and that its uncertainty, δM_c , is partially due to the daily and weekly variations in the noise level.

3.7 Catalogue contamination by blasts

Earthquake catalogues are often contaminated by artificial events, most frequently quarry or mine blasts (e.g. Wiemer & Baer, 2000; Gulia, 2010, 2012; Cattaneo *et al.*, 2014). Blasts misclassified as natural earthquakes alter the statistics made with the catalogue and may artificially rise the seismicity rate and the calculated seismic hazard. They are especially difficult to identify when they take place in seismically active regions.

The online IGN catalogue aims to only include natural earthquakes and to exclude artificial seismicity, a challenging task since the Iberian Peninsula hosts extensive mining activities. In the printed bulletins up to 1986, earthquakes

identified as artificial explosions were still listed, and specifically labelled. Artificial events are typically of low magnitude, so they started to become more frequently recorded since the 1980s. The reduction of M_c since the deployment of the broadband network in the 2000's (Section 3.5) enabled recording tens of thousands of blasts in the Iberian Peninsula, outnumbering by far the actual earthquakes (Rueda Núñez, 2006; García Vargas *et al.*, 2016). Routine screening relies on analyst judgment, based on waveform analysis and on the time of occurrence, being blasts expected preferentially during daylight hours (García Vargas *et al.*, 2016).

Despite such efforts, artificial events occasionally pass this screening and are incorporated in the catalogue as if they were natural seismicity (García Vargas *et al.*, 2016). For example, thanks to the use of a dense local network, Torcal *et al.* (2005) identified several explosions or quarry blasts which had been considered natural microearthquakes in the IGN catalogue. In a more generic analysis, Gulia (2010) identified sites with suspect contamination of quarry blasts in this and other European earthquake catalogues. Her analysis is based on the ratio of daylight to nighttime events, which takes into account that blasts at quarries or construction sites typically occur during daytime (e.g. Benson *et al.*, 1992; Wiemer & Baer, 2000; Gulia, 2012; Panzera *et al.*, 2016), while natural earthquakes are typically more abundantly recorded in the catalogues during nighttime, because of the lower artificial noise (Rydelek & Sacks 1989, and Section 3.6).

An illustrative example of blast contamination highlighted by a high daytime-to-nighttime ratio is the Trimouns talc quarry (de Parseval *et al.*, 2004), the largest open-pit talc mine in the world. It is located near Luzenac, in the French Pyrenees, where natural earthquakes also take place (42.805°N, 1.804°E, Fig. 3.16). Mining is carried out there from May to November, and blasting is performed at a very narrow time interval within the evening breaks, shortly before 16 h UT, that is 18 h summer official time (Julien Conté, former Trimouns mining project manager, pers. comm., 2015; Godey *et al.*, 2013). At this location, the IGN catalogue includes tens of highly likely quarry blasts, recorded in different years since 2004 with origin time between 15:50-15:56 UT, and $m_{bLg} \leq 2.0$ (Fig. 3.16).

The catalogue however, offers examples of blast contamination with precisely the opposite pattern: a particularly high frequency of nighttime blasts. Figure 3.17 shows the seismicity in part of the Iberian Pyrite Belt, a geologic region rich in massive sulphide deposits (Leistel *et al.*, 1998; Rosa *et al.*, 2010), which is also seismically active. Two tight clusters of small-magnitude earthquakes ($m_{bLg} \leq 2.3$, and ~ 1 in most cases) pinpoint the Portuguese mine of Neves-Corvo and the Spanish mine of Aguas Teñidas, where sulphide ores (mostly of copper and zinc) are extracted after large-scale underground blasting (Chadwick, 2007; García

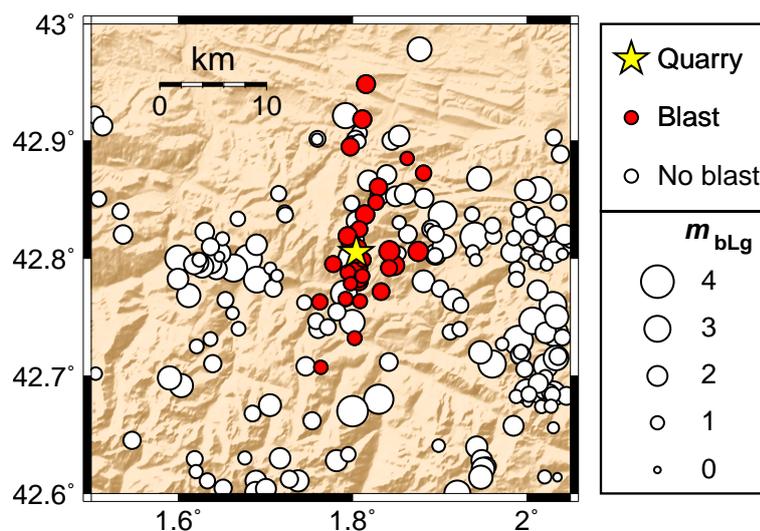


Figure 3.16: **Example of catalogue contamination by quarry blasts.** *Star*: Location of the Trimouns talc quarry in the French Pyrenees. *Circles*: Earthquakes recorded until December 2013, with size proportional to magnitude (m_{bLg} scale). *Red circles*: Highly likely blasts, that is, recorded at the typical daily blasting schedule at this quarry (a narrow range of just seven minutes, 15:50-15:56 UT) during different months and years. *White circles*: Earthquakes recorded at other times, which are most likely natural seismicity. Digital elevation model from the Shuttle Radar Topography Mission (Farr *et al.*, 2007).

Maneiro *et al.*, 2012). Neves-Corvo (currently the largest copper and zinc mine in the European Union) started operating in 1988, but seismicity is recorded in the catalogue only since 1999 (perhaps because this is about the time when the magnitude of completeness began to be low enough to record these small events, Fig. 3.12). At Aguas Teñidas, recorded seismicity began in 2009, right when commercial extraction started. Other earthquakes in the area may be tectonic, which underlines the need of distinguishing blasting from actual seismicity. In both mines, most recorded events occur in agreement with blasting schedules (Fig. 3.17). For the sake of safety, underground blasts are performed when workers are out of the mine (during meal breaks or shift changes). So they typically occur every 8 hours (at early morning, lunch time and dinner time), and also around official midnight in Aguas Teñidas (Sheila Maqueda –former drill and blast engineer at this mine–, pers. comm. 2014). Finally, these mines are exploited (and earthquakes have also been recorded at them) all days of the week. These observations cast doubts on the reliability of identifying and removing blasts in seismic catalogues based only on day/night activity ratios, or by their occurrence on working days (instead of weekends).

Identifying blasts from their location alone is hampered by location errors. This is evidenced by the wide dispersion of blast epicentres in the examples shown,

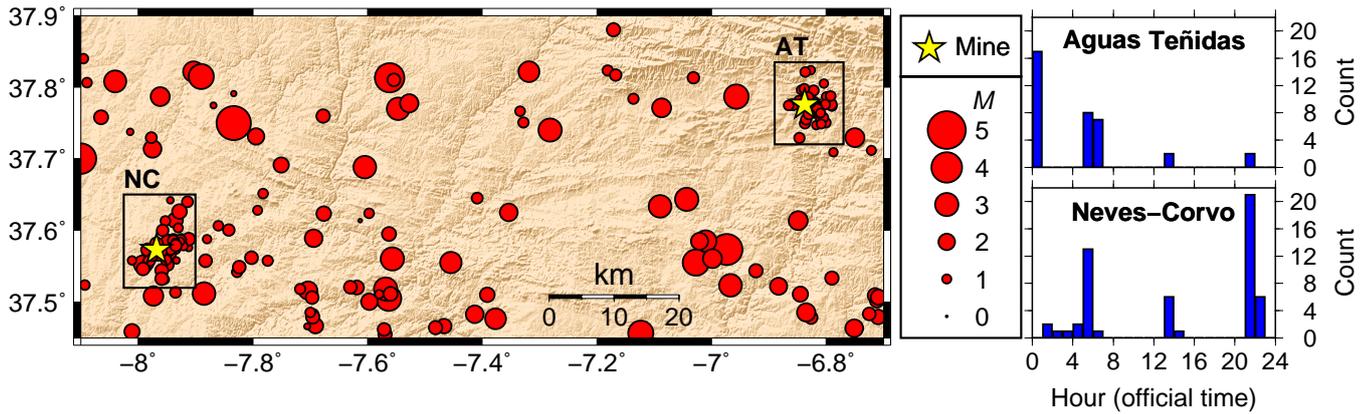


Figure 3.17: **Examples of catalogue contamination by underground mine blasts.** Stars: Location of the mines of Aguas Teñidas (AT) and Neves-Corvo (NC), in SW Iberian Peninsula. Circles: Earthquakes, with circle diameter proportional to magnitude ($m_D = 4.5$ for the largest event, and m_{bLg} scale for all the others). Histograms: Hourly frequencies of the earthquakes within the polygons surrounding each mine (which agree with the planned schedule of underground blasts). Official time is 1–2 hours ahead of Universal Time in peninsular Spain, and 0–1 hours in peninsular Portugal, depending on daylight saving time. Digital elevation model from Farr *et al.* (2007).

especially in Trimouns (Figs. 3.16 and 3.17). Quarry or mine blasts have very shallow origin depths (~ 0 km), what might help in distinguishing them from natural (deeper) earthquakes. But these small events, typically recorded by few stations, have poor depth determinations. In the three examples detailed here, most blasts have no depth determined independently (depth is fixed to 0 km in the analysis), and those for which depth is calculated have diverse reported depths, even ≥ 10 km in several cases, evidencing poor depth accuracy.

For statistical analysis of the catalogue, it is useful to set a magnitude threshold above which blasts are absent or highly unlikely (e.g. Habermann, 1987). In Europe, quarry blasts do not usually reach $M \sim 2.5$ (Giardini *et al.*, 2004; Gulia, 2010). But in Spain, blasts recorded and identified by IGN occasionally reach $m_{bLg} > 3.0$ (Rueda Núñez, 2006). An exceptional case of mining-related seismicity in Spain was the ground shaking caused by the collapse of the Reocín mine (7 January 1965), which reached $m_{bLg} = 4.1$ and whose entry was recently removed from the IGN catalogue. This early exception aside, it seems likely that only small collapse events may be confused with natural seismicity and included in the catalogue, such as a probable case with $m_{bLg} = 2.2$ analysed by Alvarez-Garcia *et al.* (2013).

To estimate the upper magnitude threshold of non-tectonic events contaminating the IGN catalogue, I compared it with the list of non-tectonic events in the same region, supplied by the EMSC for January 1998 - July 2012

(Gilles Mazet-Roux, pers. comm., 2015). The list contains over 34,000 events, mostly mine explosions, and is built with data contributions from IGN and other agencies, plus further screening by EMSC operators, which considers the location of known blasting sites and the times at which blasts are typically detonated (Godey *et al.*, 2013). I considered that an event in the IGN catalogue was also on the EMSC list if there was a difference in origin time ≤ 1.5 s and epicentral distance ≤ 10 km between them. This led to identifying 210 non-tectonic events in the IGN catalogue, all located in the Iberian Peninsula and southern France (none in Northern Africa or the Canary Islands). This figure is optimistic, considering the number of potential blasts already identified in the cases analysed here. However, it is expected to include at least the largest blasts, which should be the easiest to identify and the most consistently located in both catalogues. The IGN magnitudes of these events was $m_{bLg} \leq 2.5$ and in most cases ≤ 2.0 . It is thus unlikely that the IGN catalogue is contaminated by non-tectonic events with $m_{bLg} \gtrsim 2.5$, an observation that also holds for the blasts identified in the examples above and for the collapse event discussed by Alvarez-Garcia *et al.* (2013).

3.8 Discussion

This chapter has tried to offer a comprehensive review of the Spanish National Earthquake Catalogue and its main overall properties, focusing on the instrumental era. The evolution of the catalogue is detailed, highlighting the upgrades of the monitoring network, of the techniques used for locating earthquakes and of the determination of magnitudes. These developments have resulted in a heterogeneous catalogue with a marked overall improvement over time.

The hypocentral location precision has improved over the decades and shows a clear spatial pattern, being typically much better in the Iberian Peninsula than elsewhere, especially for the hypocentral depth. The most notable improvement began with the introduction of the new automatic location procedures in November 1997. As in other instances (e.g. Amato & Mele, 2008; Custódio *et al.*, 2015), earthquakes in distant offshore areas are the least precisely located, because they are far from the stations and recorded with a large azimuthal gap. The determination of focal mechanisms off the Iberian Peninsula shore also suffers from these factors (Domingues *et al.*, 2013). Such limitations are particularly relevant because tsunamis and the largest earthquakes in the region have indeed originated in such areas. Ocean-bottom seismometers could greatly improve offshore monitoring, as highlighted by results from temporary deployments (e.g. Geissler *et al.*, 2010) and theoretical modelling (Zahradník & Custódio, 2012).

The catalogue includes a temporal and spatial mixture of non-equivalent magnitude scales. In other catalogues such diversity is known to affect, e.g. the calculated seismicity rates and b -values, and should be carefully considered before any statistical analysis (e.g. Habermann, 1987, 1991; Zúñiga & Wyss, 1995; Zúñiga & Wiemer, 1999; Tormann *et al.*, 2010, and references therein). Homogenizing the magnitudes from this catalogue has been carried out in several instances by converting them all into moment magnitude by using empirical relations (Mezcua *et al.*, 2011; Martínez Solares *et al.*, 2013; Gaspar-Escribano *et al.*, 2015, and references therein). But reliable conversion relationships (Cabañas *et al.*, 2015) are only available for magnitudes $\gtrsim 3$ (that is, for a minority of events), since original moment magnitudes for small earthquakes in the area can rarely be calculated. Backward review of catalogue magnitudes into a single scale when possible (e.g. Tormann *et al.*, 2010) would alleviate part of this heterogeneity.

Improvements in earthquake detection have enabled recording increasing numbers of earthquakes in the region per unit time, evidencing a progressive lowering of the magnitude of completeness, M_c . For the first time for this catalogue, detailed maps of M_c are here calculated, during several periods in each region analysed. These periods, as short as data availability allows, are chosen considering the reviewed evolution of the catalogue, taking care that during each of them the network was relatively stable, and the same routines for location and magnitude determination were used. This kind of detailed temporal analysis had so far only been carried out for catalogues with many more events (e.g. Hutton *et al.*, 2010). The appropriateness of the temporal and spatial choices of the analysis is evidenced by the consistency of the results (the lowest M_c is systematically found in the areas with densest monitoring) and their typically low uncertainties (except for areas where earthquakes with different magnitude scales overlap). M_c first improved slowly from the 1960's to the 1990's. A sudden improvement was due to the introduction of new automatic location routines in late 1997, and a dramatic one was achieved by the new digital, broadband network from 2002 onwards. The catalogue is more complete in the most seismically active (better monitored) areas within the Iberian Peninsula and the Canary Islands than elsewhere, although a progressive improvement of completeness over the years is observed even in offshore areas.

A weekend effect (more earthquakes recorded during weekends) is found in the catalogue for the first time, as well as the better known daily variations (more earthquakes recorded during nighttime). Both effects are interpreted as related to the variable amplitudes of artificial noise. They disappear when only the complete part of the catalogue is regarded, which supports the values of magnitude of completeness calculated here.

Clear examples of contamination by quarry and mine blasts (with magnitudes $\lesssim 2.5$) are pinpointed, despite the monitoring agency has successfully removed the vast majority of artificial explosions from the catalogue. The cases studied here indicate that useful complementary screening would be provided by checking known mining sites and their typical blasting schedule. Examples of large-scale underground mines are found where blasts are set off during nights and weekends, indicating that it is not always correct to assume that blasts are only made during daylight times (e.g. Wiemer & Baer, 2000) and in weekdays.

Hopefully, this review will provide useful guidelines for users of this valuable database, and also insight for planning where and how this catalogue would most benefit from future upgrades in instrumentation or in routine analysis of the seismic data.

Measurement of areas on a sphere using Fibonacci and latitude–longitude lattices

Measure what is measurable, and make measurable what is not so.

Galileo Galilei.

Chapter adapted and expanded from:

Álvaro González (2010)

Measurement of areas on a sphere using Fibonacci and latitude–longitude lattices.
Mathematical Geosciences, 42 (1), 49–64.

4.1 Introduction

The area of a region is easy to measure, without explicitly considering its boundaries, by determining which points of a lattice are inside or outside the region. This point-counting method is commonly applied to estimating areas on a plane (Bardsley, 1983; Jarai *et al.*, 1997; Howarth, 1998; Gundersen *et al.*, 1999; Baddeley & Jensen, 2004). A related issue is how to approximate the region boundaries from this sampling (Barclay & Galton, 2008).

Point-counting on the sphere is commonly used for estimating the Earth coverage of satellite constellations (Kantsiper & Weiss, 1998; Feng *et al.*, 2006), the fraction of the Earth's surface efficiently seen by one or more satellites (Fig.4.1). In the simplest case, each satellite covers a circular region (cap), so the constellation covers a complex set of isolated and/or overlapping caps (Kantsiper & Weiss, 1998).

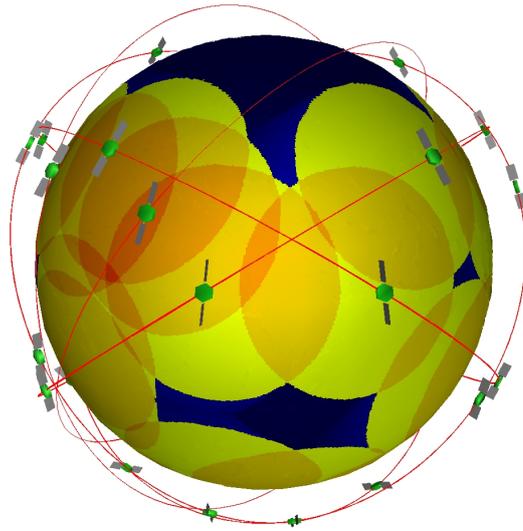


Figure 4.1: **Coverage of an example satellite constellation.** The Orbcomm constellation, consisting of thirty-six communication satellites orbiting the Earth at a height of 825 km. Image elaborated with the SaVi satellite constellation visualization software (Wood & Worfolk, 2011).

Similarly, some maps designed for earthquake forecasting depict earthquake-prone regions as complex sets of up to tens of thousands of caps (Kossobokov & Shebalin, 2003; Kafka, 2007; Kafka & Ebel, 2011, and Chapter 2). To assess these forecasts it is necessary to measure the fraction of the Earth's area covered by these regions.

Analytical solutions exist if the spherical region has a known, regular or polygonal shape (Kimerling, 1984; Bevis & Cambareri, 1987; Sjöberg, 2006). The area of a set of caps on the sphere has a complex analytical solution (Kantsiper & Weiss, 1998), which unfortunately does not indicate whether any particular location on the surface of the sphere is covered.

The numerical error of point counting should ideally decrease rapidly as the lattice density increases. Numerous works deal with errors on the plane (Bardsley, 1983; Jarai *et al.*, 1997; Howarth, 1998; Gundersen *et al.*, 1999; Baddeley & Jensen, 2004). Some particular cases using latitude–longitude lattices were analyzed elsewhere (Kantsiper & Weiss, 1998). In automated counting, the computation time is directly proportional to the number of lattice points. Satellite constellation coverage (Ochieng *et al.*, 2002) or the areas marked on rapidly-updated earthquake forecasting maps (Chapter 2) need continuous monitoring, implying a trade-off between accuracy and computational load. Thus it is important to find lattices able to measure areas as efficiently as possible.

On the plane, the regular hexagonal lattice provides optimal sampling (Conway & Sloane, 1998). On the sphere, it is impossible to arrange regularly more than 20

points (the vertices of a dodecahedron), and the optimum configuration of a large number of points is problem-specific (e.g. Saff & Kuijlaars, 1997; Conway & Sloane, 1998; Williamson, 2007; Gregory *et al.*, 2008). For optimal point-counting, the area represented by every point should be almost the same.

Traditionally, the latitude–longitude lattice is used for measuring Earth coverage (Kantsiper & Weiss, 1998; Ochieng *et al.*, 2002; Feng *et al.*, 2006). However, it is very inhomogeneous (Fig. 4.2), requiring non-uniform weighting of the point contributions. Also, its number of points is restricted by geometrical constraints.

The Fibonacci lattice is a particularly appealing alternative (Dixon, 1987, 1989, 1992; Fowler *et al.*, 1992; Svergun, 1994; Kozin *et al.*, 1997; Swinbank & Purser, 1999, 2006a,b; Winfield & Harris, 2001; Nye, 2003; Hannay & Nye, 2004; Purser & Swinbank, 2006; Purser, 2008). Being easy to construct, it can have any odd number of points (Swinbank & Purser, 2006b), and these are evenly distributed (Fig. 4.2) with each point representing almost the same area. For the numerical integration of continuous functions on a sphere, it has distinct advantages over other lattices (Hannay & Nye, 2004; Purser & Swinbank, 2006).

This chapter demonstrates that for measuring the areas of spherical caps, the Fibonacci lattice is much more efficient than its latitude–longitude counterpart. For this reason, the former was used for measuring the areas covered within each distance threshold in the retrospective forecast maps of Ch. 2. A description of both lattices (Sects. 4.2 and 4.3) is followed by an explanation of how to use them for measuring cap areas (Sect. 4.4) and how the error of this measurement is assessed (Sect. 4.5). The error results obtained from an extensive Monte Carlo simulation are described, and to some extent explained analytically (Sect. 4.6). The conclusions are set out in the final section.

4.2 Latitude–longitude lattice

The latitude–longitude lattice is the set of points located at the intersections of a grid of meridians and parallels, separated by equal angles of latitude and longitude (Fig. 4.2). This is the “latitude–longitude grid” (Swinbank & Purser, 2006b; Williamson, 2007) or “equal-angle grid” (Gregory *et al.*, 2008). The points concentrate towards the poles, due to the converging meridians, resulting in high anisotropy.

The number of points, P , depends on the angular spacing, δ , between grid lines. Since $\delta = 180^\circ / k$ with $k = 1, 2, \dots$,

$$P = 2k(k - 1) + 2. \quad (4.1)$$

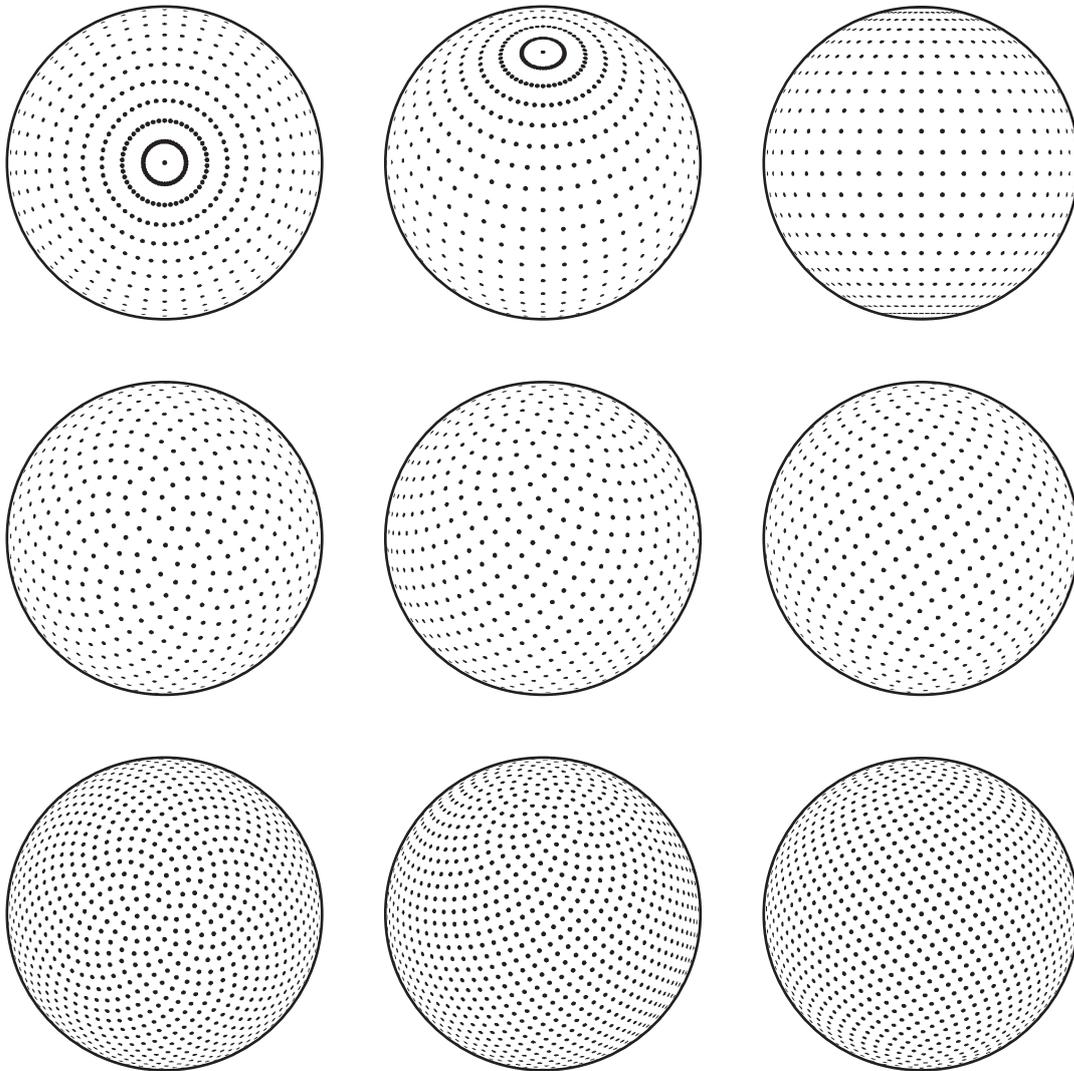


Figure 4.2: **Latitude–longitude lattice** (*top*) and **Fibonacci lattice** (*middle and bottom*), with 1014, 1001 and 2001 points, respectively. Orthographic projections, centred at the pole (*left*), latitude 45° (*middle*) and equator (*right*). In the Fibonacci lattice, the points are much more evenly spaced, and the axial anisotropy is much smaller.

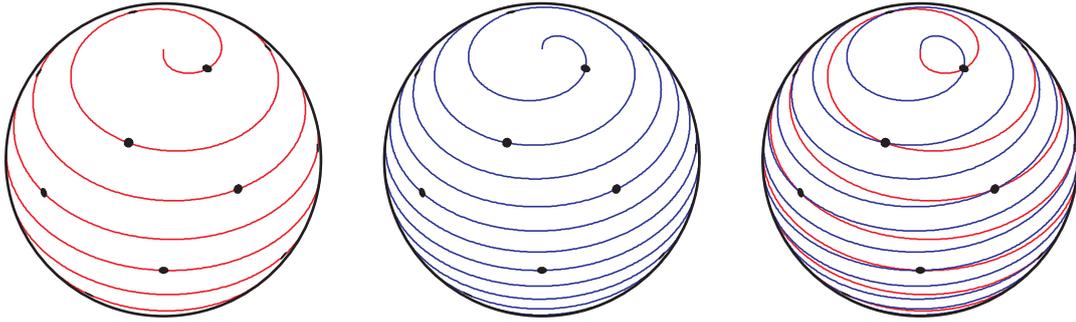


Figure 4.3: **Generative spirals of a Fibonacci lattice with 21 points.** The angle turn between consecutive points along a spiral is based on the golden ratio (ϕ): either the golden angle, $360^\circ\phi^{-2} \simeq 137.5^\circ$ (*first spiral, red*), or its complementary, $360^\circ\phi^{-1} \simeq 222.5^\circ$ (*second spiral, blue*). No point is placed at the poles in this lattice version. Orthographic projection, centred at longitude 0° , latitude 45° .

That is the number of meridians ($2k$) times the number of parallels ($k - 1$), plus the two poles. Frequently, to evaluate satellite coverage, (Feng *et al.*, 2006) $\delta = 0.25^\circ$, so more than a million points are used.

4.3 Fibonacci lattice

The points of the Fibonacci lattice are arranged along a tightly wound generative spiral, with each point fitted into the largest gap between the previous points (Fig. 4.3). This spiral is not apparent (Fig. 4.2) because consecutive points are spaced well apart. The most apparent spirals join the points closest to each other, and form crisscrossing sets (Swinbank & Purser, 2006b). The points are evenly spaced in a very isotropous way.

The following subsections describe how to construct the lattice used in this chapter, and the history of the Fibonacci lattice in various research fields.

4.3.1 Lattice construction

The Fibonacci lattice is named after the Fibonacci ratio. The Fibonacci sequence was discovered in ancient India (Singh, 1985; Knuth, 1997) and rediscovered in the middle ages by Leonardo Pisano, better known by his nickname Fibonacci (Sigler, 2002). Each term of the sequence, from the third onwards, is the sum of the previous two: 0, 1, 1, 2, 3, 5, 8, 13, 21, ... Given two consecutive terms, F_i and F_{i+1} , a Fibonacci ratio is F_{i+1}/F_i . As first proved by Robert Simson in 1753 (Weisstein, 2002), this quotient, as $i \rightarrow \infty$, quickly approaches the golden ratio, defined as $\phi = 1 + \phi^{-1} = (1 + \sqrt{5})/2 \simeq 1.618$.

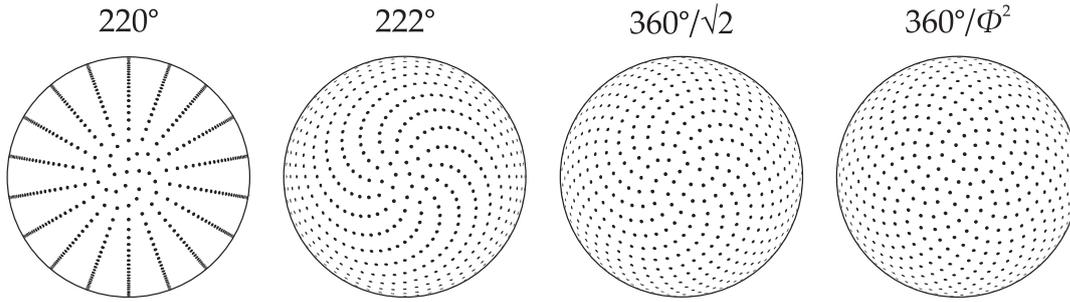


Figure 4.4: **Dependence of the lattice geometry on the longitudinal angle turn.** These four spiral lattices on the sphere (shown as orthographic projections centred at the pole) have 1001 points each, but different angle turn between consecutive points (labelled above). If this is rational, the resulting periodicities cause a heterogenous arrangement (*first two examples*). If it is irrational, strict periodicities disappear (*third and fourth examples*). Near-periodicities are avoided only if the golden angle (or its complementary) is used (Fibonacci lattice, *rightmost example*).

The Fibonacci lattice differs from other spiral lattices on the sphere (Weiller, 1966; Klíma *et al.*, 1981; Rakhmanov *et al.*, 1994; Saff & Kuijlaars, 1997; Chukkapalli *et al.*, 1999; Bauer, 2000; Hüttig & Stemmer, 2008) in that the longitudinal turn between consecutive points along the generative spiral is the golden angle, $360^\circ(1 - \phi^{-1}) = 360^\circ\phi^{-2} \simeq 137.5^\circ$, or its complementary, $360^\circ\phi^{-1} \simeq 222.5^\circ$. Some lattice versions replace ϕ by its rational approximant, a Fibonacci ratio.

The golden angle optimizes the packing efficiency of elements in spiral lattices (Ridley, 1982, 1986). This is because the golden ratio is the “most irrational” number (Weisstein, 2002), so periodicities or near-periodicities in the spiral arrangement are avoided, and clumping of the lattice points never occurs (Ridley, 1982; Dixon, 1987; Prusinkiewicz & Lindenmayer, 1990; Jean, 1994; Hannay & Nye, 2004). Such optimization is illustrated in Fig. 4.4.

The lattice version used here (Swinbank & Purser, 2006b) is probably the most homogeneous one. It is generated with a Fermat spiral (also known as the cyclotron spiral) (Vogel, 1979; Dixon, 1987, 1992), which embraces an equal area per equal angle turn, so the area between consecutive sampling points, measured along the spiral, is always the same (Swinbank & Purser, 2006b). Also, its first and last points are offset from the poles, leading to a more homogeneous polar arrangement (Purser & Swinbank, 2006; Swinbank & Purser, 2006b) than in other versions (Svergun, 1994; Kozin *et al.*, 1997; Nye, 2003; Hannay & Nye, 2004; Purser & Swinbank, 2006). When a Fibonacci ratio is used (Svergun, 1994; Kozin *et al.*, 1997; Nye, 2003; Hannay & Nye, 2004), the number of lattice points is $F + 1$, where $F > 1$ is a term of the Fibonacci sequence. The lattice used here is instead based on the golden ratio and can have any odd number of points.

To elaborate the lattice (Swinbank & Purser, 2006b), let N be any natural number. Let the integer i range from $-N$ to $+N$. The number of points is

$$P = 2N + 1, \quad (4.2)$$

and the spherical coordinates, in radians, of the i th point are:

$$\text{lat}_i = \arcsin\left(\frac{2i}{2N+1}\right), \quad (4.3)$$

$$\text{lon}_i = 2\pi i\phi^{-1}. \quad (4.4)$$

This pseudocode provides the geographical coordinates in degrees:

For $i = -N, (-N + 1), \dots, (N - 1), N$, Do {

$$\text{lat}_i = \arcsin\left(\frac{2i}{2N+1}\right) \times 180^\circ / \pi$$

$$\text{lon}_i = \text{mod}(i, \phi) \times 360^\circ / \phi$$

$$\text{If } \text{lon}_i < -180^\circ, \text{ then } \text{lon}_i = 360^\circ + \text{lon}_i$$

$$\text{If } \text{lon}_i > 180^\circ, \text{ then } \text{lon}_i = \text{lon}_i - 360^\circ$$

} End Do

Here, \arcsin returns a value in radians, while $\text{mod}(x, y)$ returns the remainder when x is divided by y , removing the extra turns of the generative spiral. For example, $\text{mod}(6, \phi) = 6 - 3 \times \phi$. The last two lines keep the longitude range from -180° to $+180^\circ$.

Every point of this lattice is located at a different latitude, providing a more efficient sampling than the latitude–longitude lattice. The middle point, $i = 0$, is placed at the equator ($\text{lat}_0 = 0$ and $\text{lon}_0 = 0$). Each of the other points $(\text{lat}_i, \text{lon}_i)$ with $i \neq 0$, has a symmetric one with $(-\text{lat}_i, -\text{lon}_i)$. The lattice as a whole is not symmetric. Thanks to the use of a Fermat spiral, each point of the lattice is inside a spherical collar of equal area (Fig. 4.5).

The number of points in this lattice version is odd because of the convention of placing a point at the equator, used by Swinbank & Purser (2006b). Removing this requirement would enable an even number of points.

The longitudinal turn between consecutive lattice points along the spiral (Eq. 4.4) is the complementary of the golden angle (Swinbank & Purser, 1999, 2006b). To use the golden angle instead, we can substitute Eq. 4.4 by

$$\text{lon}_i = -2\pi i\phi^{-2}. \quad (4.5)$$

With Eq. 4.4, the spiral progresses eastwards, while the minus sign of Eq. 4.5 indicates a westward progression.

Not yet mentioned is the fact that the lattice points are placed at the intersections between these Fermat spirals of opposite chirality, except at the poles (Fig. 4.3). To

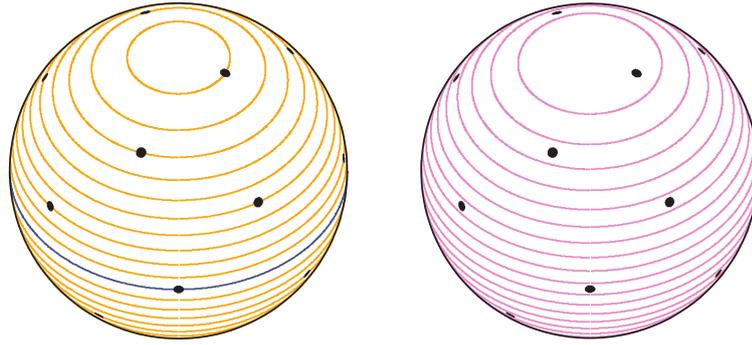


Figure 4.5: **Latitudinal distribution of points in a Fibonacci lattice.** Each of these 21 points is located at a different parallel (*left*, with the equator pictured blue). Also, each one is inside a spherical collar of equal area (*right*), i.e. a latitudinal band between consecutive violet parallels. The area of each collar is thus the total area of the sphere divided by the number of lattice points.

draw the spirals, i is made continuous in Eqs. 4.3, 4.4 and 4.5, and ranges from $(-N - 1/2)$ to $(N + 1/2)$. The $1/2$ term accounts for the polar offset (Purser & Swinbank, 2006; Swinbank & Purser, 2006b).

4.3.2 Lattice history

The Fibonacci lattice is a mathematical idealization of patterns of repeated plant elements, such as rose petals, pineapple scales, or sunflower seeds (Fig. 4.6). The study of these arrangements is known as Phyllotaxis (Prusinkiewicz & Lindenmayer, 1990; Jean, 1994; Adler *et al.*, 1997; Kuhlemeier, 2007). The Bravais brothers (Bravais & Bravais, 1837) were the first to describe them using a spiral lattice on a cylinder. They argued that the most common angle turn between consecutive elements along this spiral in plants is the golden angle. The latter provides optimum packing (Ridley, 1982, 1986), maximizing the exposure to light, rain and insects for pollination (Maciá, 2006). Structures in cells and viruses also follow this pattern (Erickson, 1973). In some experiments, elements are spontaneously ordered on roughly hemispherical Fibonacci lattices, because the system tends to minimize the strain energy (Li *et al.*, 2005) or to avoid periodic organization (Douady & Couder, 1992).

Unwrapping the cylindrical Fibonacci lattice (Fig. 4.7) produces a flat one (Bravais & Bravais, 1837; Hannay & Nye, 2004; Swinbank & Purser, 2006b), frequently used for numerical integration (Zaremba, 1966; Niederreiter, 1992, 1994; Sloan & Joe, 1994).

Projecting the cylindrical Fibonacci lattice to the sphere generates the spherical version (Hannay & Nye, 2004; Swinbank & Purser, 2006b). This can be generalized



Figure 4.6: **Spherical Fibonacci lattices in Nature** – an oblique view of two *Mammillaria solisoides*. The elements of many plants form Fibonacci lattices, for example, the areoles (spine-bearing nodes) of *Mammillaria* cacti (Fowler *et al.*, 1992; Kuhlemeier, 2007).

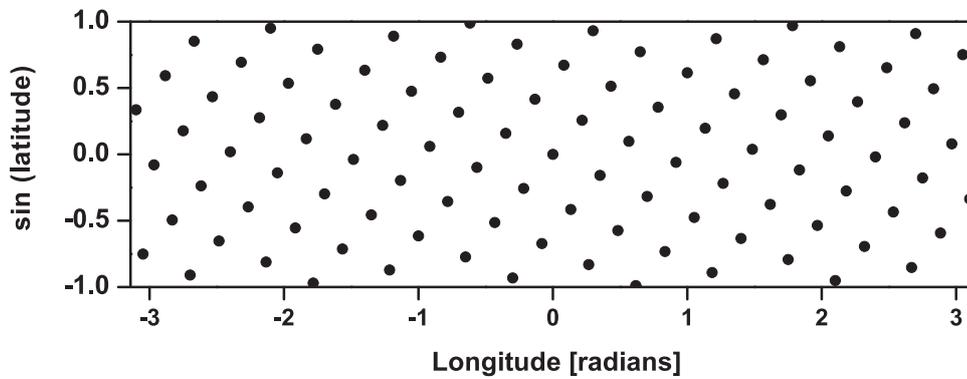


Figure 4.7: **A flat Fibonacci lattice with 101 points**, that is, a cylindrical equal-area Lambert projection of its spherical counterpart. Below, the same projection applied to the Earth's surface, using Blue Marble imagery (NASA, 2002).

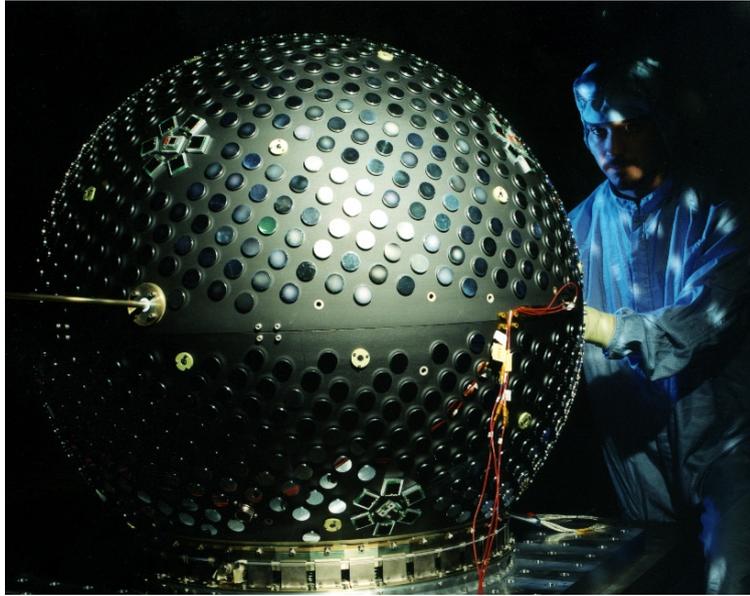


Figure 4.8: **The Starshine-3 satellite** (Maley *et al.*, 2002; Lean *et al.*, 2006) had 1500 mirrors arranged on a spherical Fibonacci lattice. Picture by Michael A. Savell and Gayle R. Fullerton, taken while the satellite was being inspected by John Vasquez. Reproduced by courtesy of the U.S. Naval Research Laboratory.

to arbitrary surfaces of revolution (Ridley, 1986; Dixon, 1992). The first graphs of spherical Fibonacci lattices used, as here, the golden ratio and a Fermat spiral (Dixon, 1987, 1989, 1992). A version based on the Fibonacci ratio is used in the modelling of complex molecules (Vriend, 1990; Svergun, 1994; Kozin *et al.*, 1997). In this case (Svergun, 1994; Kozin *et al.*, 1997), a “+” sign in the formula for the longitude should be substituted by “ \times ” (D. Svergun, personal communication, 2009). Versions with the golden ratio serve to simulate plants realistically (Fowler *et al.*, 1992) and to design golf balls (Winfield & Harris, 2001). The latter method was used by Douglas C. Winfield (B. Moore and D. C. Winfield, personal communication, 2009) in the Starshine-3 satellite (Fig. 4.8).

The spherical Fibonacci lattice is a highly efficient sampling scheme for integrating continuous functions (Nye, 2003; Hannay & Nye, 2004; Purser & Swinbank, 2006), as was observed in magnetic resonance imaging (Ahmad *et al.*, 2007). It is also advantageous for providing grid nodes in global meteorological models (Michalakes *et al.*, 1999; Swinbank & Purser, 1999, 2006a,b; Purser & Swinbank, 2006; Purser, 2008).

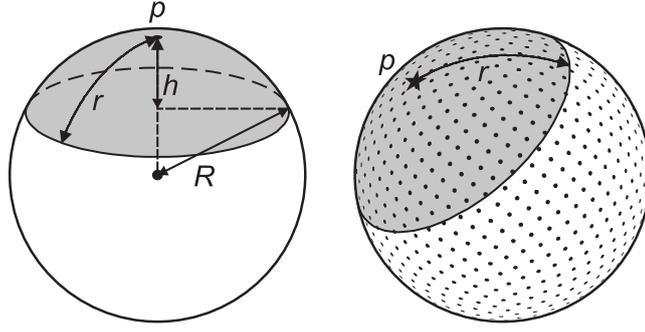


Figure 4.9: **Measurement of cap area.** The cap (*left*), centred at p , has height h and great-circle radius r , and is a region of a sphere with radius R . Placed at random (*right*), its area can be estimated by considering the lattice points it covers.

4.4 Area measurement

The area measurement starts by finding which points of the lattice are inside the considered region. This is expressed by a Boolean function f_i , such that $f_i = 1$ if the i th point is inside, and $f_i = 0$ otherwise (Kantsiper & Weiss, 1998). If the region is a cap with radius r , it suffices to measure d_i , the shortest distance (great-circle distance) between the sampling point and the cap centre (Fig. 4.9):

$$f_i = \begin{cases} 1 & \text{if } d_i \leq r, \\ 0 & \text{if } d_i > r. \end{cases} \quad (4.6)$$

Each lattice point must be assigned a weight, w_i , proportional to the area it represents. Then the estimate, \tilde{A} , of the region area A , is measured considering the sphere area, A_S , and summing the contribution of all P points of the lattice:

$$A \simeq \tilde{A} = \frac{A_S \sum_{i=1}^P f_i w_i}{\sum_{i=1}^P w_i}. \quad (4.7)$$

The weights depend on the lattice type, as described below.

4.4.1 Weights in the latitude–longitude lattice

The weights should be inversely proportional to the point density, which here increases towards the poles (Fig. 4.2). The linear spacing between parallels is constant. The length of a parallel is $2\pi R \cos(\text{lat})$, where R is the sphere radius. In any parallel there is the same number of lattice points ($2k$), so their density is inversely proportional to (Kantsiper & Weiss, 1998; Van den Dool, 2007):

$$w_i = \cos(\text{lat}_i) \quad (4.8)$$

4.4.2 Weights in the Fibonacci lattice

Thanks to the even distribution of points in this lattice, the same weight can be assumed for all of them (Purser & Swinbank, 2006; Swinbank & Purser, 2006b), namely

$$w_i = 1. \quad (4.9)$$

Each point represents the area corresponding to its Voronoi cell (Thiessen polygon). This is the region of positions closer to the corresponding lattice point than to any other (Evans & Jones, 1987; Na *et al.*, 2002). Using the exact area of each Voronoi cell as weight for its lattice point (Ahmad *et al.*, 2007) would improve the area measurement only slightly. The average cell area equals A_S/P . Here, regardless of P , only the areas of less than ten cells, located at the polar regions, differ by more than $\sim 2\%$ from this value (Swinbank & Purser, 2006b). As P increases, proportionally fewer cells depart significantly from the average area. Unlike the latitude–longitude lattice, the homogeneity of the Fibonacci lattice improves with the number of points.

4.5 Error assessment

This section details how to assess the error involved in measuring the area of spherical caps placed at random on the sphere. A good way to measure the homogeneity of a spherical lattice is to compare the proportion of lattice points in spherical regions with the normalized areas of the regions (Cui & Freeden, 1997). For this task, it is common to use spherical caps (Saff & Kuijlaars, 1997; Damelin & Grabner, 2002; Brauchart, 2004), which also appear in the applications mentioned in the introduction of this chapter.

The area of a spherical cap (see Fig. 4.9) is:

$$A_C = 2\pi R h = 2\pi R^2 \left(1 - \cos \frac{r}{R}\right). \quad (4.10)$$

The normalized cap area is:

$$F = \frac{A_C}{A_S} = \frac{1 - \cos(r/R)}{2}, \quad (4.11)$$

where $A_S = 4\pi R^2$ is the sphere area.

The absolute error of measuring a single cap is the absolute difference between the estimated fraction and the actual one:

$$E = \left| \tilde{A}_C / A_S - F \right|. \quad (4.12)$$

This depends on the lattice type, the number of points, and the size and location of the cap. If $E = 0$, the cap gets its fair share of weighted lattice points. If $A_C = 0$, $E = 0$. A plane cuts the sphere into two complementary caps. For any cap, E is the same as for the complementary cap with area $A'_C = 1 - A_C$, so it suffices to consider caps not larger than a hemisphere ($A_C = A_S/2$).

The error is characterized here numerically using a Monte Carlo method. In each particular realization, a cap is randomly placed. Every point of the sphere has the same probability of becoming the cap centre, p , with coordinates:

$$\text{lat}_p = \frac{180^\circ}{\pi} \arcsin(2X - 1) \quad (4.13)$$

$$\text{lon}_p = 360^\circ X - 180^\circ. \quad (4.14)$$

Here, X is a random number, chosen with uniform probability in the range $[0, 1]$, independently for each equation. The area of this j th cap is estimated, and its corresponding error is E_j .

The process is repeated for a total of n independently located caps of the same size, providing a sample of n values of E . The root mean squared error is (Weisstein, 2002):

$$\text{rmse} = \sqrt{\frac{1}{n} \sum_{j=1}^n E_j^2}. \quad (4.15)$$

The supremum error in Eq. 4.12 for caps of any size, for a given lattice type and P , using $w_i = 1$, is the “spherical cap discrepancy” (Saff & Kuijlaars, 1997; Damelin & Grabner, 2002; Brauchart, 2004). It cannot be determined exactly with a Monte Carlo simulation because it might result from a cap size not used, or a location not sampled. However, it is unfortunately difficult to compute explicitly (Damelin & Grabner, 2002). The maximum E measured for the Fibonacci lattice is a lower bound to its spherical cap discrepancy.

4.6 Results

This section details the maximum errors and root mean squared errors measured with the Monte Carlo simulation detailed in the previous section.

Thirteen lattice configurations of each type, from $P \simeq 10^2$ to $P \simeq 10^6$, were analyzed. The chosen values of P increase in logarithmic steps, as accurately as possible ($P \simeq 10^2, 10^{7/3}, 10^{8/3}, 10^3, \dots$). It is impossible to use identical P for both lattices because P is odd in the Fibonacci lattice but even in the latitude–longitude lattice. Moreover, there are considerably fewer possible values of P in the latter. For each configuration, 200 different cap sizes were used: from $A_C = 0.0025A_S$ to

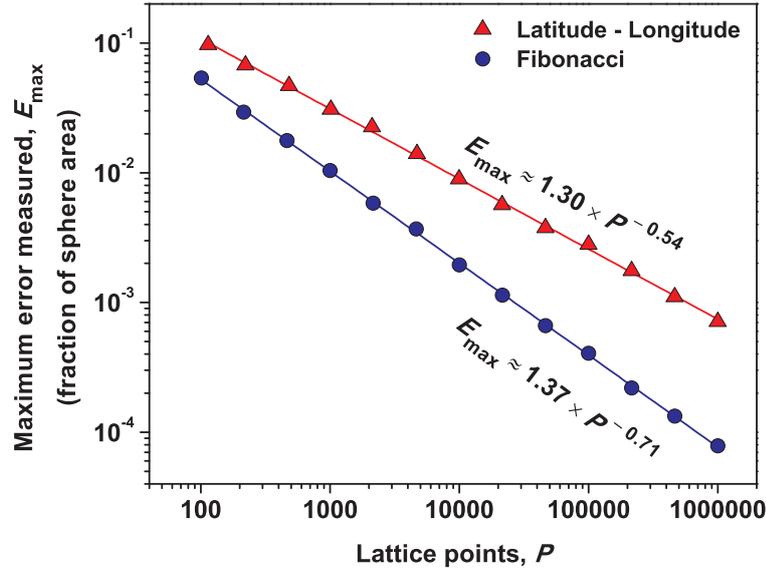


Figure 4.10: **Maximum error measured for randomly located spherical caps of any size.** For area measurements on the Earth, about a million points are frequently used (Feng *et al.*, 2006), for which the maximum error would be one order of magnitude smaller in the Fibonacci lattice.

$A_C = 0.5A_S$ in steps of $0.0025A_S$. To obtain smooth results, $n = 60,000$ was chosen for each cap size and lattice configuration.

In total, 312 million caps were measured (2 lattice types \times 13 values of $P \times$ 200 cap sizes \times 60,000 caps). After optimization, the calculation took 43 days of CPU time using 2.8 GHz, 64-bit processors.

The maximum E measured, for caps of any size and location, is represented in Fig. 4.10. For the Fibonacci lattice, it is much lower and decays faster than for the latitude–longitude lattice, despite the non-uniform weighting of points of the latter.

The rmse depends on the lattice type, number of points and cap size, as shown in Fig. 4.11. Because of the symmetry of the latitude–longitude lattice, any hemispherical cap covers one half of the points of the lattice, and (thanks to the point weights) the estimation is perfect ($E = 0$ and $\text{rmse} = 0$). This exception aside, the rmse tends to increase with the cap area, in a non-trivial way, which is more complex for the latitude–longitude lattice than for the Fibonacci lattice.

Figure 4.12 (top) shows the maximum values of rmse of each curve. They follow parallel power laws:

$$\text{rmse}_{\max} \simeq kP^{-3/4}, \quad (4.16)$$

with $k \simeq 0.505$ for the latitude–longitude lattice, and $k \simeq 0.362$ for the Fibonacci lattice. Interpolating for the same P , the rmse_{\max} would be $0.505/0.362 \simeq 40\%$ larger in the latitude–longitude lattice.

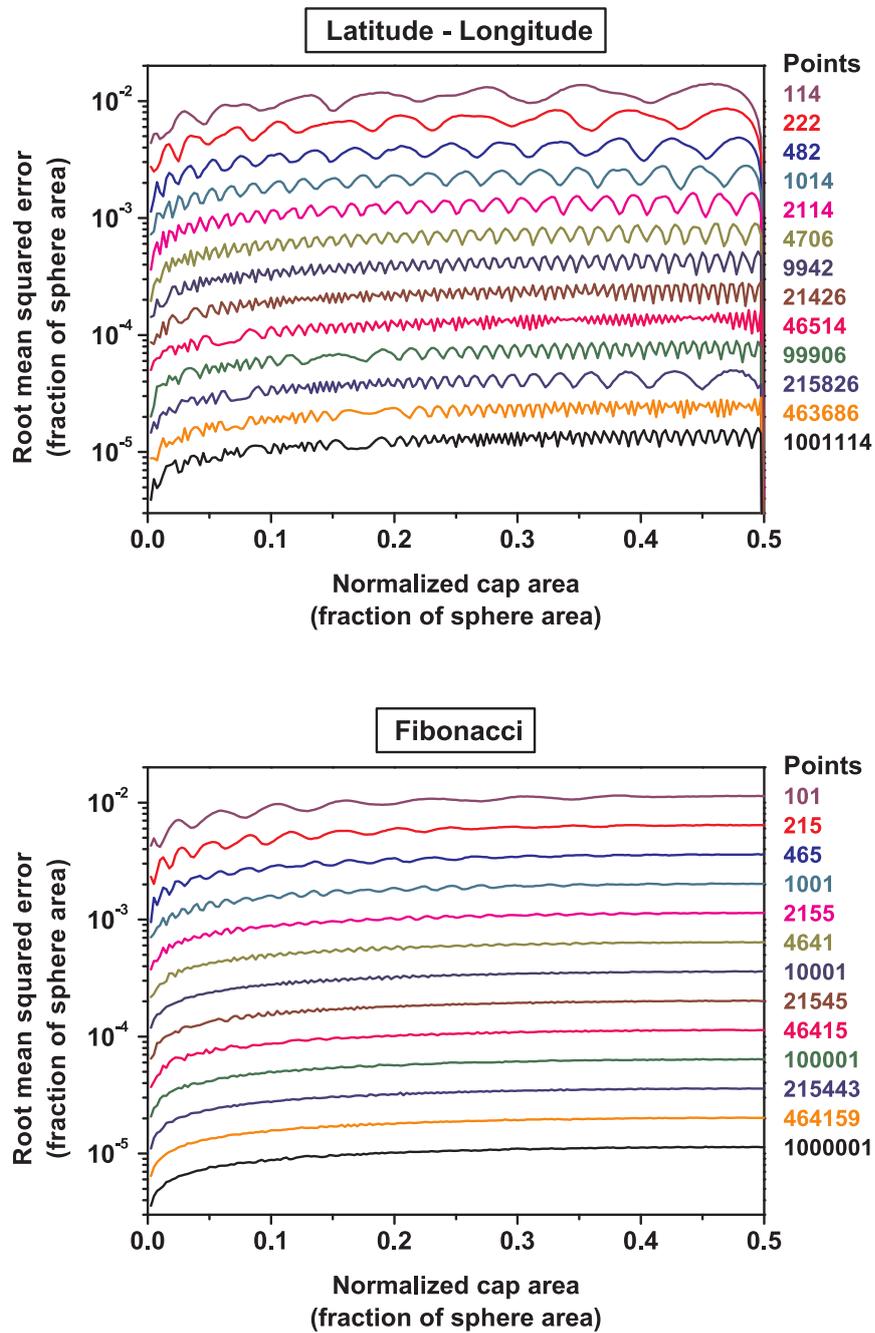


Figure 4.11: **Root mean squared error.** For randomly placed caps which occupy an area fraction given by the abscissas, the curve indicates the root mean squared error of the area measurement. Each curve corresponds to a different number of lattice points, labelled to its right. For similar lattice densities, the Fibonacci lattice provides smaller and more homogeneous errors.

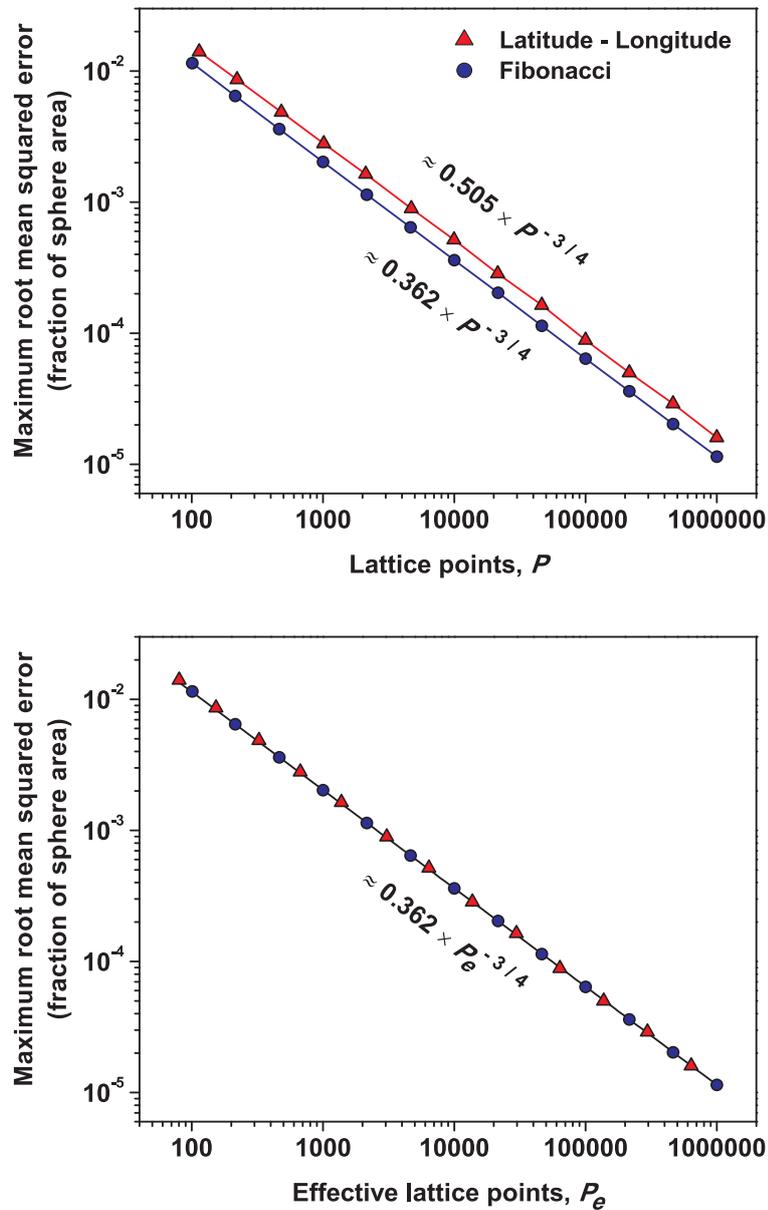


Figure 4.12: **Scaling of the maximum root mean squared error.** Each dot corresponds to the maximum value of a curve in Fig. 4.11. *Top:* For the same number of points, the values would be about 40% smaller for the Fibonacci lattice. *Bottom:* The latitude–longitude lattice is inefficient because the effective (weighted) number of points is smaller than the real one. Considering this fact, the results collapse to a single function. The power-law decay and its exponent agree with analytical scaling arguments (see text).

4.6.1 Analytical approach to the root mean squared error

The scaling shown in Fig. 4.12 can be explained using arguments from similar problems in the plane (Kendall, 1948; Huxley, 1987, 2003). The number of points of a regular square lattice enclosed by a sufficiently smooth, closed curve placed at random can be expressed as (Huxley, 2003):

$$P_{\text{in}} = AM^2 + D, \quad (4.17)$$

where A is the area enclosed by the curve, M is the inverse to the lattice spacing, and D is the discrepancy. The nominal spacing of a spherical lattice is $\sqrt{A_S/P}$ (Swinbank & Purser, 2006b). Assuming that the Fibonacci lattice is regular enough,

$$M \approx \sqrt{\frac{P}{A_S}}. \quad (4.18)$$

Substituting this into Eq. 4.17,

$$P_{\text{in}} \approx P \frac{A}{A_S} + D. \quad (4.19)$$

Dividing all the terms of this equation by P ,

$$E \approx \frac{|D|}{P} ; \text{rmse} \approx \frac{\text{rms}(D)}{P}. \quad (4.20)$$

In the planar case, the root mean squared of D , $\text{rms}(D)$, is proportional to \sqrt{M} (Kendall, 1948; Huxley, 1987, 2003). Extrapolating this fact, we obtain the scaling observed in the Monte Carlo simulation:

$$\text{rmse} \propto \frac{\sqrt{M}}{P} \approx \frac{(P/A_S)^{1/4}}{P} \propto P^{-3/4}. \quad (4.21)$$

In the latitude–longitude lattice, the same scaling holds if we consider its smaller sampling efficiency. The latter may be measured with the denominator of Eq. 4.7. This is the number of points of a homogeneous lattice that would do the same work, and is up to $\sim 36\%$ smaller than P in the range of P considered here. If Fig. 4.12 is plotted using these effective points in the abscissas, the same fit suffices for both lattice types (Fig. 4.12, bottom).

If the sampling points were placed at random, the rmse would decrease more slowly, as $\propto P^{-1/2}$ (Bevington & Robinson, 1992).

4.7 Discussion

This chapter analyses the errors involved in measuring the areas of spherical caps using lattices of sampling points: the latitude–longitude lattice (classically used for

this task), and a Fibonacci lattice (Swinbank & Purser, 2006b). The latter has low anisotropy, is easy to construct, and is shown to result from the intersection of two generative spirals (Fig. 4.3). A review of the literature (Sect. 4.3.2) reveals the successful applications of this spherical lattice dated since the 1980s.

If the Fibonacci lattice were used instead of its latitude–longitude counterpart (as done in Chapter 2), the area measurement would be more efficient, allowing a significant reduction of the computation time. For approximately the same number of lattice points, the maximum root mean squared error would be about 40% smaller (Fig. 4.12). The maximum errors would also be smaller, and would decay faster with the number of points (Fig. 4.10). If about a million points were used, as is commonly the case (Feng *et al.*, 2006), the maximum error would be one order of magnitude smaller (Fig. 4.10).

It is also found that the maximum root mean squared error obeys a single scaling relation when the sampling efficiency is taken into account (Fig. 4.12). This is partially explained using arguments from similar problems on the plane (Kendall, 1948; Huxley, 1987, 2003).

The area estimate depends also on the orientation of the sampling lattice, especially if the latter has high anisotropy. The difference may be assessed (Hannay & Nye, 2004) by rotating the lattice (Greiner, 1999). Such an issue is not relevant in the case analysed here because the caps were placed at random with uniform probability over the spherical surface.

Here the Earth's shape has been approximated by a sphere (Kantsiper & Weiss, 1998), adding a slightly higher error than other shape models (Kimerling, 1984; Earle, 2006; Sjöberg, 2006). Assessing this difference may be a topic of future research.

Finally, the article presenting this work (González, 2010a), currently cited in ~ 20 publications, has served as inspiration for further applying the spherical Fibonacci lattice to fields as diverse as biochemistry (e.g. Illes-Toth *et al.*, 2015), computer science (e.g. Keinert *et al.*, 2015), crystallography (e.g. Watson & Curtis, 2013), geodesy (e.g. Slobbe *et al.*, 2012), mathematics (e.g. Fuselier *et al.*, 2014) and industrial design (e.g. Brockmeyer *et al.*, 2013).

Part II

Temporal Earthquake Forecasting

If the stresses of an earthquake district affected only homogeneous rock and were always relieved by slipping on the same fault plane, the cycle of events would be regular; but with complexity of structure and multiplicity of alternative points of collapse, all superficial indication of rhythm is lost. If rhythmic order shall ever be found in the apparent confusion, it will be through an analysis which takes account of the points of origin of all important shocks.

Grove Karl Gilbert (1909, p. 128).

Probabilistic temporal forecasting of large earthquakes in the Parkfield section, San Andreas Fault.

Since all models are wrong the scientist cannot obtain a “correct” one by excessive elaboration. On the contrary following William of Occam he should seek an economical description of natural phenomena.

George E.P. Box (1976, p. 792)

Essentially, all models are wrong, but some are useful.

George E.P. Box & Norman R. Draper (1987, p. 424).

Chapter adapted and expanded from:

Álvaro González, Javier B. Gómez & Amalio F. Pacheco (2005)

The occupation of a box as a toy model for the seismic cycle of a fault.

American Journal of Physics, 73 (10), 946–952. [Erratum: (2007) 75 (3), 286.]

Álvaro González, Javier B. Gómez & Amalio F. Pacheco (2006)

Updating seismic hazard at Parkfield.

Journal of Seismology, 10 (2), 131–135.

5.1 Introduction

This chapter deals with the temporal forecasting of large earthquakes in individual faults. It is devoted to calculating time-varying earthquake probabilities and their intertwined binary forecasts, in which earthquakes are expected to occur in specific time intervals. It introduces a new model as an idealization of the seismic cycle, and applies it and existing ones to the series of earthquakes with magnitude ~ 6 generated by the San Andreas Fault at Parkfield (California). The reader should bear in mind that the original probability estimates included here were submitted for publication (González *et al.*, 2005, 2006a) just after the last mainshock at Parkfield, in 2004, so they can be considered as prospective, real time, forecasts.

Section 5.2 reviews the current knowledge of how regularly earthquakes occur in individual faults and in regional scales. The relative temporal regularity of large earthquakes in individual faults is shown to result from the classical Reid's classical elastic rebound theory, supported by the results of laboratory experiments and by the records of regular series of earthquakes both in geological timescales and in the recent instrumental era. The short-term temporal clustering of regional seismicity will be considered as an emergent property due to the effects of short-term earthquake triggering between different faults.

Section 5.3 reviews how the so-called renewal models are being applied to calculate time-varying probabilities of specific faults, and the importance of considering these models, and not just time-independent hazard estimates. The background and data about the Parkfield case study is set out in Section 5.4.

In section 5.5, a new model resulting from statistical physics (González *et al.*, 2005) is presented as a simple description of the seismic cycle. First the rules of the model are described, and its properties are deduced analytically. It is then used as a renewal model for calculating time-varying probabilities of the next earthquake at Parkfield.

The framework of binary forecasting and a simple strategy for the long-term forecasting of earthquakes at Parkfield using the new model is set out in Section 5.6.

Finally section 5.7 proceeds analogously as above, but considering several candidate models, instead of just one. After fitting other renewal models to the Parkfield series, the resulting probabilities for the next earthquake will be calculated, emphasizing the similarities and differences between the models, and comparing them to a time-independent estimate. The section ends with the discussion on how to use the models for binary forecasting and by mentioning more recent, relevant developments.

5.2 The seismic cycle

Forecasting the moment of occurrence of a large earthquake depends crucially on how regularly such events occur. To characterize a series of events it is natural to consider the period between consecutive ones (called *waiting time*, *inter-event time* or *recurrence interval*). A measure of the regularity of the series is the *coefficient of variation*, sometimes called *aperiodicity*, which here will be denoted α . This non-dimensional coefficient equals the standard deviation of the inter-event times (σ) divided by their mean. In a perfectly periodic sequence, $\alpha = 0$. In a Poisson process, in which the events occur randomly at a mean rate independently of when previous ones took place, $\alpha = 1$ (albeit the converse is not necessarily true). Series where $0 < \alpha < 1$ are called *quasiperiodic*, while *clustered* series are those with $\alpha > 1$ (e.g. Kagan & Jackson, 1991).

Clustering dominates earthquake occurrence at regional scales in the time periods covered by instrumental, high-quality earthquake catalogues (usually the last few decades, e.g. Kagan & Jackson, 1991). Aftershock sequences and earthquake swarms are obvious examples of temporal clustering. But earthquakes are clustered in time even when the earthquake rate (frequency) appears stationary (Corral, 2004). A very similar temporal clustering is observed for regional seismicity and for acoustic emission events produced during the formation of micro-cracks during the compression of rocks and porous materials in laboratory experiments (e.g. Davidsen *et al.*, 2007; Baró *et al.*, 2013). Usually, the pattern of temporal occurrence of earthquakes in a region is reasonably well described statistically (e.g. Saichev & Sornette, 2007; Zhuang *et al.*, 2011; Baró *et al.*, 2013; Main, 2013) by the model of epidemic-type aftershock sequences (ETAS, first introduced by Kagan & Knopoff, 1981; Ogata, 1988). In this model, of which many variants have been proposed, some earthquakes occur in a Poissonian way at a specified rate, triggering aftershocks which in turn can trigger further aftershocks (as in an epidemic outbreak, hence the name).

However, the large earthquakes (which release most of the elastic deformation) along any given fault are typically separated by recurrence intervals much longer than than the periods covered by instrumental earthquake catalogues. A recent example is the earthquake originated in 2011 offshore Tōhoku, Japan (magnitude $M_w=9.0$), which was caused by sudden slip of up to tens of meters (e.g. Romano *et al.*, 2014), worth of several hundred years of plate motion, and whose last similar predecessor at the same location took place in the year 869 CE (Sawai *et al.*, 2012). This great difference in timescales prevent extrapolating directly the properties of instrumental earthquake catalogues to the long-term seismic behaviour of faults.

In reporting the mechanism of the great California earthquake of 1906, Reid (1910) presented the *elastic rebound theory*. It hypothesizes that an earthquake is the result of a sudden relaxation of elastic strain by sudden slip or rupture along a fault. This theory extended earlier insights into the relation between earthquakes and faults by other geologists (Yeats *et al.*, 1997), especially Gilbert (1884, 1909), McKay (1890), Griesbach (1893) and Koto (1893). Since its formulation, it has been the basis for interpreting the earthquakes that occur in faults in the Earth's upper, fragile crust (e.g. Scholz, 2002; Kanamori & Brodsky, 2004).

According to Reid's theory, potential elastic energy slowly accumulates on a fault over a long time after the occurrence of an earthquake, as the rock blocks on both sides of the fault are strained elastically by tectonic forces. When the strain is large enough, the system relaxes by fast rupture and/or frictional sliding along the fault during the next earthquake. The elastic waves generated by this sudden event are the seismic waves that seismometers detect. As the tectonic forces keep operating in the long term, Reid theorized that the tectonic loading and relaxation process of a fault is cyclic. The *seismic cycle* is the time interval between two successive large earthquakes generated by slip on the same fault (e.g. Scholz & Gupta, 2000).

Reid's theory was supported by the co-seismic displacements measured on the San Andreas Fault, but it took sixty years until it was supported by laboratory experiments of rock friction. The earthquake cycle was shown to be explained by unstable, stick-slip frictional behaviour (Brace & Byerlee, 1966), so even if the long-term velocity between the rock blocks is constant, their movement can be episodic, with long periods during which the blocks are stick to each other ("locked"), while shear stress increases, until the friction threshold is reached and then sudden slip ensues, causing a *stress drop*. As frequently observed for earthquakes, this stress drop is much smaller than that associated with the fracture of intact rock (e.g. Brace & Byerlee, 1966; Scholz, 1998, 2002). Rupture and frictional slip are thus two complementary processes during earthquake generation. Once a fault is formed, it can grow and accumulate slip during thousands of earthquakes (e.g. Nicol *et al.*, 2005).

Laboratory experiments of friction of stony materials at constant slip rate typically generate either quasi-periodic cycles (of stick-slip behaviour) or stable sliding (*fault creep*), depending, for example, on the properties of the fragmented rock between the blocks (*fault gouge*) (e.g. Anthony & Marone, 2005, see Fig. 5.1). In each of these stick-slip cycles, the shear stress rises initially in a linear, elastic, way. Then increases progressively more slowly due to aseismic slip, approaching a plateau. Finally, the sudden slippage of the blocks (analogous to an earthquake)

causes an abrupt stress drop. Complex numerical models of the seismic cycle of a fault on an elastic medium display very similar patterns (Ben-Zion, 2008). When the models take into account that the lower crust and mantle deform in a viscoelastic (instead of elastic) way, the faster stress accumulation during the beginning of the cycle is more evident (Michael, 2005). These cycles with fast *post-seismic reloading* are, for example, the standard model for interpreting the cyclic deformation in subduction zones (e.g. Bourgeois, 2006). Both frictional behaviours (stick-slip or creep) are the end-members displayed by natural seismic faults (Scholz, 1998, 2002).

Elastic rebound theory implies that the seismic cycle in individual faults should be quasi-periodic, given that the fault would require some time until enough elastic strain accumulates before generating the next large earthquake. If the seismic cycle were periodic, earthquake forecasting would be easy. There is increasing information about earthquake occurrences in the seismic record, compiled with historical data and recognition of ancient large earthquakes on faults (Yeats *et al.*, 1997; McCalpin, 2009). These data show that the seismic cycle of any given fault is not strictly periodic.

Large earthquakes are indeed quasiperiodic in the few instances where their geologic record is sufficiently long and precisely dated as to be well constrained. The most complete records of major earthquakes in individual faults are probably those of the southern San Andreas Fault, in California (with 29 events in two temporal periods, with a combined $\alpha = 0.7$, Scharer *et al.*, 2010) and the central part of the Alpine Fault in New Zealand (24 events in a single period, with $\alpha = 0.33$ Berryman *et al.*, 2012). The more regular sequences are shown by the faults with simpler geometry and which are relatively isolated from others (Berryman *et al.*, 2012). Sequences are more irregular for complex fault systems (e.g. Berryman *et al.*, 2012) or for long-term local records of ground shaking caused by earthquakes from different faults (Gomez *et al.*, 2014).

Relevant examples of earthquakes for which the seismic cycle follows a quasi-periodic pattern are the so called *repeating earthquakes*, which have very similar magnitudes, hypocentral locations and are recorded with nearly identical waveforms. These earthquakes, typically of low to moderate magnitude, are usually interpreted as produced by the sudden slip of stronger patches of a fault (*asperities*), which experience stick-slip and which are surrounded by others where the fault moves by aseismic creep (e.g. Nadeau *et al.*, 1995; Okada *et al.*, 2003; Uchida *et al.*, 2012). Their quasi-periodicity makes them indeed time-predictable to some extent (Zechar & Nadeau, 2012).

The general reason for the irregularity of the earthquake cycle is that the tectonic loading and relaxation of a fault are complex nonlinear processes (e.g. Rundle *et*

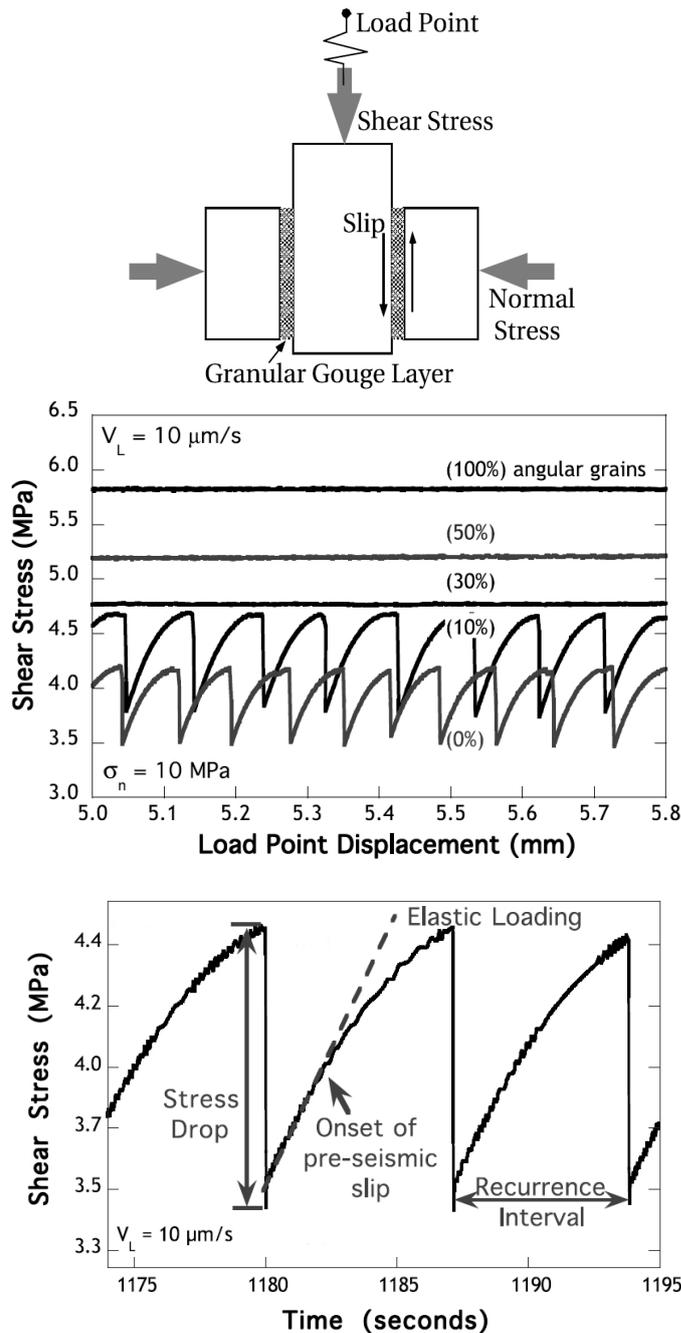


Figure 5.1: **Seismic cycles in laboratory experiments of rock friction**, from Anthony & Marone (2005) (©American Geophysical Union, reproduced with permission). *Top:* The experimental apparatus is a double-direct shear device in non-fracture loading regime, in which a block is displaced at a constant loading velocity (V_L , slip rate) between other two fixed ones, under a constant normal stress σ_n . *Middle:* The frictional regime was either quasi-periodic stick-slip or stable friction (creep), depending on the simulated fault gouge (with a variable percentage of angular sand grains mixed with spherical glass beads). *Middle and bottom:* In each seismic cycle, shear stress first increases linearly with displacement or time (in a elastic fashion), and later progressively more slowly because of stress dissipation (by aseismic slip in this case).

al., 2000). Moreover, faults occur in topologically complex networks (e.g. Bonnet *et al.*, 2001) and an earthquake occurring in a fault influences what occurs in other faults by static and dynamic stress transfer (e.g. Harris, 2000; Freed, 2005), which can operate up to great distances (e.g. Pollitz *et al.*, 2012; Parsons *et al.*, 2014). Even simple deterministic models of interacting faults may exhibit complex and chaotic behaviour (e.g. Huang & Turcotte, 1990). Irregular recurrence of large earthquakes in the same fault might also result if these only release a small fraction of the total stress accumulated (Tormann *et al.*, 2012), so that the fault could produce them (occasionally) within intervals much shorter than the mean (e.g. Kagan & Jackson, 1999; Valée & Satriano, 2014). Short-term earthquake triggering synchronizes the seismic cycles of different faults (e.g. Scholz & Gupta, 2000; Scholz, 2010; Benedetti *et al.*, 2013), what would explain the short-term temporal clustering of regional seismicity as an *emergent property* (a property of the system as a whole, which emerges from the interactions among its components that themselves do not exhibit such property).

5.3 Renewal models

The duration of the seismic cycle is thus not constant, but it follows a statistical distribution that can be empirically deduced from the earthquake time series (e.g. Savage, 1994). This distribution, if it were known, could be extrapolated to estimate the temporal probability of the next earthquake. However, it is not well known, because there are little data (typically less than ten) in the earthquake time series for any given fault or fault section. To use this probabilistic approach, it is convenient to fit the data to a theoretical statistical distribution.

Especially since the 1970s (Vere-Jones, 1970; Utsu, 1972a,b; Rikitake, 1974; Hagiwara, 1974; Rikitake, 1976b; Vere-Jones, 1978) earthquake recurrence is frequently modelled as a renewal process (Çinlar, 1975; Daley & Vere-Jones, 2003) in which the times between successive events, in this case the large earthquakes in a fault, are assumed to be independent and identically distributed random variables, disregarding correlations between them. In this interpretation, the expected time of the next event does not depend on the details of the last event, except the time it occurred. This is the simplest approach beyond a Poisson process, in which only the rate of the events is relevant and the probability for the next earthquake is constant.

In a quasi-periodic renewal process, in agreement with the elastic rebound theory, the probability of another large earthquake would be low just after the last one, and would then gradually increase, as tectonic deformation slowly stresses the

fault again. When the next large earthquake finally occurs, it resets the renewal process to its initial state.

Several well-known statistical distributions, such as the gamma (Utsu, 1984), log-normal (Nishenko & Buland, 1987) and Weibull (Hagiwara, 1974; Utsu, 1984; Kiremidjian & Anagnos, 1984; Sieh *et al.*, 1989; Abaimov *et al.*, 2008; Rundle *et al.*, 2005) have been used to describe the duration of the seismic cycle and to calculate the probabilities of future earthquakes. The gamma model fits well the (clustered) distribution of time intervals between earthquakes in a region, if time is re-scaled by the average earthquake rate, as first shown by Corral (2004) and discussed by Lippiello *et al.* (2012) and references therein. All these distributions have also been used as failure models for reliability and time-to-failure problems (Mann *et al.*, 1974).

More recently, many numerical models have been devised for simulating the tectonic processes occurring on a seismic fault (e.g. Main, 1996; Ben-Zion, 2001, 2008). These models can generate as many *synthetic earthquakes* as desired (e.g. Ward, 2000), so the statistical distribution of the time intervals between them can be fully characterized. Sometimes this distribution is alike an already known one (e.g. Weibull, Rundle *et al.*, 2005), but in general it is specific to the model (e.g. Robinson, 2004; Field, 2015) and may be affected by new features included in it (e.g. Yoder *et al.*, 2015).

Two highly idealized models are the Brownian passage time model (Ellsworth *et al.*, 1999; Matthews *et al.*, 2002) and the minimalist model (Vázquez-Prada *et al.*, 2002, 2003; Gómez & Pacheco, 2004, and next chapter). Their seismic cycle distributions have been used as renewal models, to fit actual earthquake series and estimate future earthquake probabilities (e.g. Ellsworth *et al.*, 1999; Matthews *et al.*, 2002; Gómez & Pacheco, 2004). They, as well as the gamma, log-normal and Weibull distributions, provide a reasonably good fit to the existing data (e.g. Utsu, 1984; Gómez & Pacheco, 2004).

Renewal models provide a more realistic framework than a time-independent Poisson model for earthquake recurrence in individual faults (e.g. Parsons, 2008a; Berryman *et al.*, 2012; Biasi *et al.*, 2015). They allow computing time-varying probabilities which, in the long-term, may be up to several times larger than for the Poissonian assumption (e.g. Imoto, 2004). For these reasons, they are being applied systematically, for example, in the authoritative earthquake probability assessments for individual faults in California (e.g. Working Group on California Earthquake Probabilities, 2003; Field *et al.*, 2015a). Moreover, it is possible to modify the renewal models to calculate changes in probabilities due to earthquake interactions, such as static stress triggering (e.g. Parsons *et al.*, 2000).

5.4 The Parkfield series

There are a few cases in which the recurrence interval between large earthquakes in a fault is short enough to be observed. The most studied example of this kind is the series of major earthquakes generated by the San Andreas Fault at Parkfield, California.

The San Andreas fault is a dextral, strike-slip fault which takes up most of the relative motion between the North American and Pacific plates in California. It stretches for over 1200 km, from the Salton Sea, near the Mexican border, to just north of San Francisco, where it enters the Pacific Ocean. It further extends northwards undersea, and continues into other faults southwards (Fig. 5.2).

During the last century and a half, several earthquakes with magnitude around 6 have occurred along a 35 km-long section of the San Andreas Fault that crosses the tiny town of Parkfield. This section separates a locked part of the fault to the south (whose last slip generated the great Fort Tejon earthquake in 1857, $M \sim 8$) and a creeping section to the north where steady aseismic slip takes place, accompanied by small-magnitude earthquakes. The 1989 Loma Prieta and 1906 San Francisco earthquakes were generated by slip in sections of the fault located further north (Fig. 5.2).

The apparent temporal regularity of this series lead since 1985 to attempts at forecasting when the next one might happen, and to deploy dense networks of instruments in the area, the so-called Parkfield Prediction Experiment (Bakun & Lindh, 1985; Roeloffs, 2000). Thanks to this detailed monitoring the last mainshock ($M_w=6.0$), which took place in 2004, was the best monitored earthquake in history (Bakun *et al.*, 2005).

The basic Parkfield series (Fig. 5.2) only includes the largest earthquakes generated by the San Andreas Fault in the immediate vicinity of Parkfield and disregards others originated before 1922 at uncertain locations nearby (Bakun *et al.*, 2005; Topozada & Branum, 2006). The formal definition of the next Parkfield mainshock for forecasting purposes was proposed by Michael & Jones (1998), encompassing what was known about the earlier ones. It is an event with $M_w \geq 5.7$ that produces surface rupture along the San Andreas fault, and/or the Southwest Fracture Zone (a fault parallel and adjacent to the main trace of the San Andreas Fault), between 36°N and $35^\circ 45'\text{N}$, and with its epicentre within 5 km of the mapped trace (see Michael & Jones, 1998, for a detailed map). The 2004 mainshock fulfilled this definition (Michael & Jones, 2004).

Including the most recent event, the Parkfield series consists of seven mainshocks (Bakun & Lindh, 1985; Bakun, 1988; Michael & Jones, 1998; Bakun *et al.*,

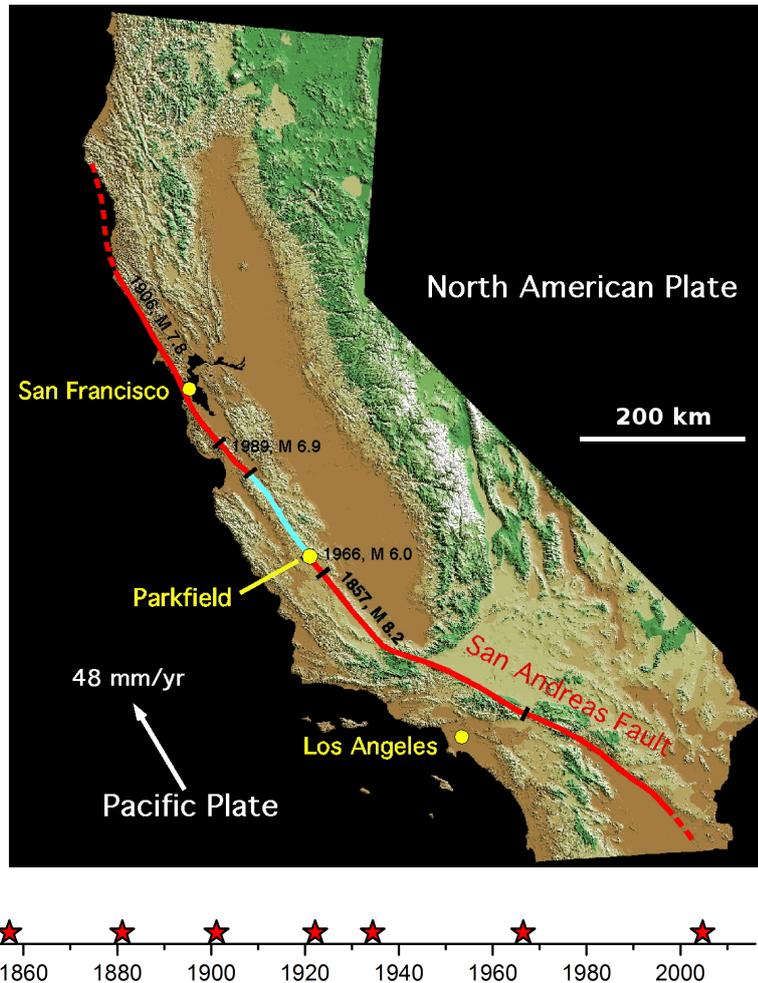


Figure 5.2: **Background on the Parkfield earthquake sequence.** *Top:* Shaded relief map of California, showing the trace of the San Andreas Fault (*red* where locked, *blue* where creeping), with segments marking the extent of the last major earthquake ruptures (years and magnitudes indicated). (Map adapted from Zoback, 2006, used with permission). *Bottom:* Sketch of the temporal sequence of mainshocks at Parkfield, including the current open interval.

2005) which occurred on 9 January 1857; 2 February 1881; 3 March 1901; 10 March 1922; 8 June 1934; 28 June 1966; and 28 September 2004. The durations, in years (calculated from the exact intervals in days) of the six observed seismic cycles are $c_1 = 24.066$, $c_2 = 20.077$, $c_3 = 21.018$, $c_4 = 12.246$, $c_5 = 32.055$, and $c_6 = 38.253$. In most other cases, the dates of large past earthquakes at a given site on a fault are not precisely known, complicating the application of recurrence models (e.g. Rhoades *et al.*, 1994; Ogata, 1999).

The last earthquake has notably increased the mean and aperiodicity of the series, which made necessary to update the existing fits of renewal models (Ellsworth *et al.*, 1999; Gómez & Pacheco, 2004; González *et al.*, 2005, the second paper being formally published just days after the last mainshock, in the October issue of the corresponding journal).

The mean of this series is

$$m = \frac{1}{6} \sum_{i=1}^6 c_i = 24.62 \text{ yr}, \quad (5.1)$$

and its sample standard deviation (square root of the bias-corrected sample variance Book, 1979) is

$$s = \left[\frac{1}{6-1} \sum_{i=1}^6 (c_i - m)^2 \right]^{1/2} = 9.25 \text{ yr}. \quad (5.2)$$

The coefficient of variation, or aperiodicity, is

$$\alpha = \frac{s}{m} = 0.3759. \quad (5.3)$$

The latter value was directly calculated from the set of intervals to avoid loss of accuracy from rounding off, since $9.25/24.62=0.3757$.

5.5 The box model

This section introduces a renewal model based on a simple *cellular automaton*. Cellular automata models are frequently used to model earthquakes and other natural hazards (e.g. Malamud & Turcotte, 2000; Jiménez, 2013). These models evolve in discrete time steps (so they are *automata*), and consist of a discrete grid of cells (so they are *cellular*), where each cell can be only in a finite number of states. Each cell's state is updated at each time step according to rules that usually depend on the state of the cell or that of its neighbours in the previous time step. Examples of more complex cellular automata applied to modelling seismicity will be mentioned in the next chapter. In the simplest models (e.g. Newman & Turcotte, 2002; Vázquez-Prada *et al.*, 2002; Tejedor *et al.*, 2008, 2015, Ch. 6), an stochastic

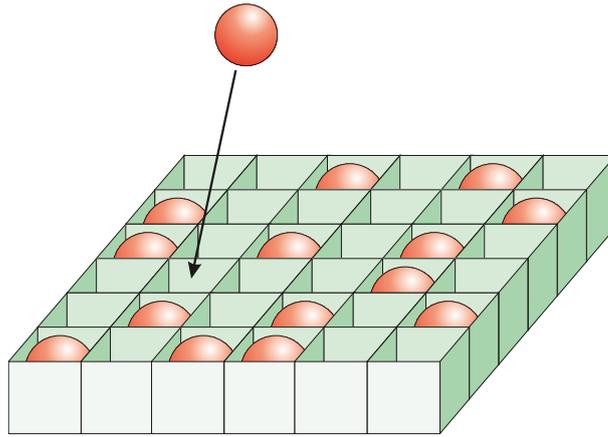


Figure 5.3: **Sketch of the box model, a new cyclic, stochastic model.** For simplicity, it is represented as a box being filled by balls, in the spirit of urn problems (classical experiments in probability theory, e.g. Johnson & Kotz, 1977). The box has a fixed number of cells (the only parameter of the model) and is loaded by balls thrown at random, one ball per time step. These are lost if they land on already occupied cells. When all cells are filled, the box is emptied and a new cycle starts.

loading is used, there are only two possible states for each cell, and the earthquakes are generated according to simplified breaking rules. The model proposed here is of this last type.

5.5.1 Rules of the model

The model will be described as a box being filled by balls, in an analogy to classical experiments in probability theory involving balls in an urn (e.g. Johnson & Kotz, 1977). Consider an array of N cells. The position of the cells is irrelevant, but we can assume that they are arranged in the shape of a box (see Fig. 5.3). At the beginning of each cycle, the box is completely empty. At each time step, one ball is thrown, at random, to one of the cells in the box. That is, each cell has equal probability, $1/N$, of receiving the ball. If the cell that is chosen is empty, it will become occupied. If it was already occupied, the thrown ball is lost. (Thus, each cell can be either occupied by a ball or empty.) When a new throw completes the occupation of the N cells of the box, it topples, becoming completely empty, and a new cycle starts. The time elapsed since the beginning of each cycle, expressed by the number of thrown balls, will be called n . The duration of the cycles is statistically distributed according to a discrete distribution function $P_N(n)$.

The box represents the area of the fault or fault segment, and the random throwing of balls represents the increase of regional stress. This randomness is a

way of dealing with the complex stress increase on actual faults. The occupation of a cell by a ball stands for the elastic strain induced by the stress in a patch or element of the fault plane. The loss of the balls that land on already occupied cells mimics stress dissipation on this plane; this dissipative rule is common to other models where the cells have only two possible states (filled or empty, e.g. Henley, 1993; Malamud *et al.*, 1998; Newman & Turcotte, 2002; Vázquez-Prada *et al.*, 2002; Tejedor *et al.*, 2009b). The total elastic strain (or conversely the total shear stress or the potential elastic energy) accumulated in the fault is represented by the number of occupied cells. This number is a state variable which completely describes the state of the system. It gradually grows up to the limit N (analogous to the failure threshold of the fault), and the toppling of the box represents the occurrence of the earthquake in the fault. It is easy to simulate the evolution of this system with a Monte Carlo approach (frequently used in stochastic models, e.g. Binder, 1997). Examples of the evolution of the cycles in the model are represented in Fig. 5.4.

This model is similar to that introduced by Newman & Turcotte (2002). The difference is that their model is a square grid of cells in which the topology is relevant: they consider that the earthquake occurs when a cluster of adjacent, occupied cells percolates the grid from one side to the opposite one. This percolating cluster (Stauffer & Aharony, 1994) happens before the grid is completely full.

5.5.2 Analytical results on the model properties

5.5.2.1 Review of the geometric distribution

To describe the box model analytically, it is convenient to recall some elements of the *geometric distribution*. Consider the probability that exactly x independent Bernoulli trials, each with a probability of success p , will be required until the first success is achieved. The probability that $(x - 1)$ failures will be followed by a success is $(1 - p)^{x-1}p$. The resulting probability function,

$$f(x; p) = (1 - p)^{x-1}p, \quad (5.4)$$

is known as the geometric distribution. Its mean and variance are

$$\langle x \rangle = \frac{1}{p}, \quad \sigma^2 = \frac{1-p}{p^2}. \quad (5.5)$$

5.5.2.2 Mean, variance and aperiodicity of the cycles

Now we deal with the box model further. In each cycle, the filling of the box proceeds sequentially and continues until the N th cell is occupied. Because each

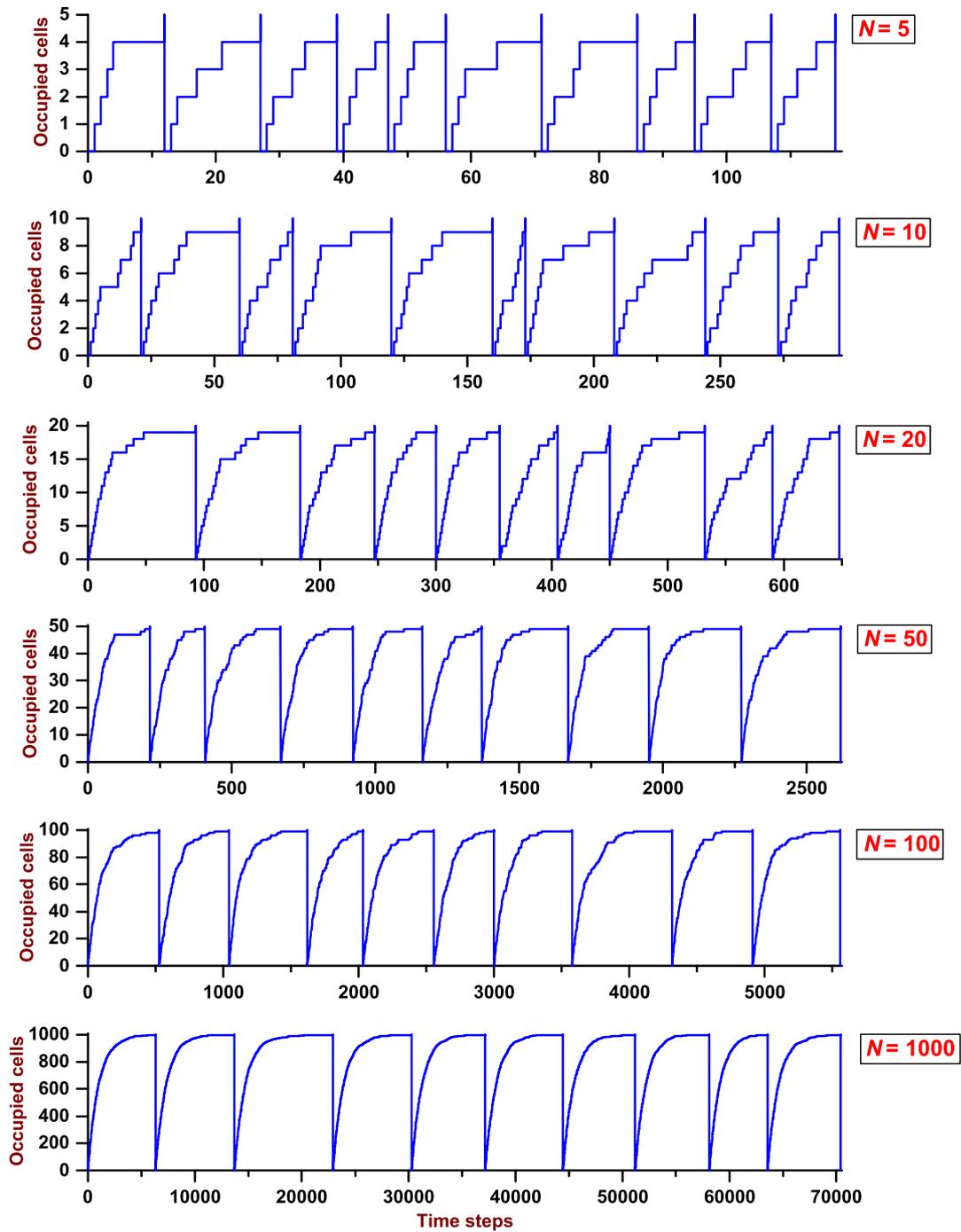


Figure 5.4: **Seismic cycles in the box model for different system sizes.** The number of occupied cells along ten cycles is represented for models with N cells. As N increases, the cycles become more regular and are longer on average. Compare with the cyclic pattern of shear stress evolution in laboratory experiments (Fig. 5.1).

of these sequential steps is an independent process, the mean number of throws to completely fill the box will be

$$\langle n \rangle_N = \langle x_1 \rangle_N + \langle x_2 \rangle_N + \cdots + \langle x_N \rangle_N, \quad (5.6)$$

where $\langle x_i \rangle_N$ is the mean number of throws it takes to fill the i th cell. Because the filling of the i th cell is geometrically distributed with $p_i = (N + 1 - i)/N$, it follows that

$$\langle x_i \rangle_N = \frac{N}{N + 1 - i}, \quad (i = 1, 2, \dots, N) \quad (5.7)$$

and therefore

$$\langle n \rangle_N = 1 + \sum_{i=2}^N \frac{N}{N + 1 - i}. \quad (5.8)$$

Relations similar to Eqs. (5.6) and (5.8) can be written for the variance of the number of thrown balls to fill the box, namely

$$\sigma_N^2 = \sigma_1^2 + \sigma_2^2 + \cdots = 0 + \sum_{i=2}^N \frac{1 - \frac{N + 1 - i}{N}}{\left(\frac{N + 1 - i}{N}\right)^2}, \quad (5.9)$$

and consequently, the standard deviation is

$$\sigma_N = \left[\sum_{i=2}^N \frac{N(i-1)}{(N+1-i)^2} \right]^{1/2}. \quad (5.10)$$

The coefficient of variation (aperiodicity) of the series, α_N , for a given N is

$$\alpha_N = \frac{\sigma_N}{\langle n \rangle_N}. \quad (5.11)$$

The mean and standard deviation of the cycle duration in the model are shown in Fig. 5.5. The aperiodicity is represented in Fig. 5.6. The latter reaches a maximum of $\simeq 0.47$ for $N = 2$ and $N = 3$, and decreases towards 0 as $N \rightarrow \infty$.

The mean and the standard deviation of the box model can be calculated by summing the $N - 1$ terms of Eqs. (5.8) and (5.10). For $N \geq 10$, these two equations can be approximated (with an absolute error < 0.01) by their asymptotic expressions (deduced in sections A.1.2 and A.1.3 of the Appendix):

$$\langle n \rangle_N \xrightarrow{N \rightarrow \infty} N(C + \ln N) + \frac{1}{2}, \quad (5.12)$$

where $C \simeq 0.5772157$ is Euler's constant, and

$$\sigma_N \xrightarrow{N \rightarrow \infty} N \left[\frac{\pi^2}{6} - \frac{1 + C + \ln N}{N} \right]^{1/2}, \quad (5.13)$$

and the aperiodicity can be estimated by using Eq. (5.11) with Eqs. (5.12) and (5.13).

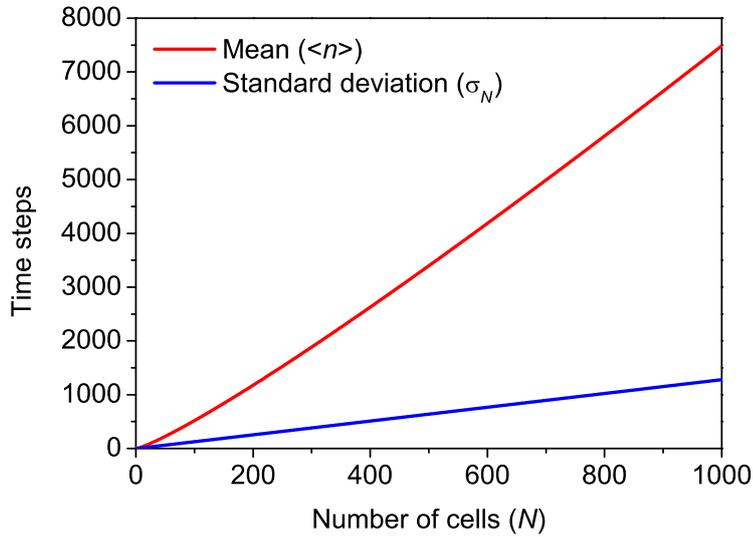


Figure 5.5: **Mean and standard deviation of the cycle duration in the box model**, as a function of the number of cells (N).

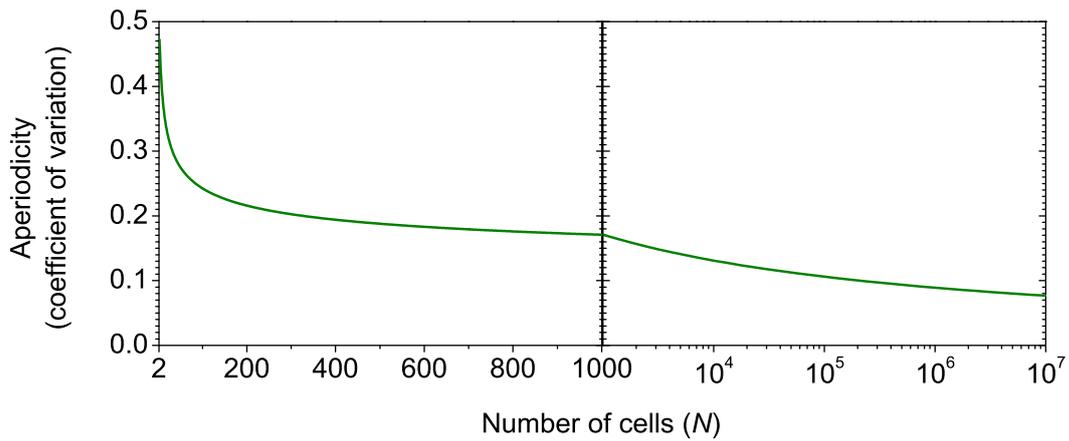


Figure 5.6: **Aperiodicity (coefficient of variation) of the sequence of cycles in the box model**, as a function of the number of cells (N). Note the logarithmic scale of the abscissas for $N > 1000$.

5.5.2.3 Probability distribution for the duration of the cycle

The function $P_N(n)$, represented in Fig. 5.7 for several instances, is not as easy to obtain as its mean and standard deviation, and is given by

$$P_N(n) = \sum_{j=1}^{N-1} (-1)^{j+1} \binom{N-1}{j-1} \left(1 - \frac{j}{N}\right)^{n-1}, \quad (5.14)$$

and the cumulative probability function, $A_N(n)$:

$$A_N(n) = \sum_{j=N}^n P_N(j) = 1 - \sum_{j=1}^{N-1} (-1)^{j+1} \binom{N-1}{j-1} \left(1 - \frac{j}{N}\right)^n \frac{N}{j}. \quad (5.15)$$

We have deduced Eq. (5.14) by means of a Markov chain approach analogous to the one we used with other model (Vázquez-Prada *et al.*, 2002, 2003). This derivation is included in the Appendix (Sec. A.1.1). In the limit $n \rightarrow \infty$, $P_N(n)$ tends to a geometric distribution (Fig. 5.7).

5.5.2.4 Conditional probability

The *conditional probability* is the probability that an earthquake will occur in the period $n + \Delta n$, given that it has not yet occurred in the step $n - 1$. That is:

$$P_{n+\Delta n}(N, n) = \frac{A_N(n + \Delta n) - A_N(n)}{1 - A_N(n - 1)}. \quad (5.16)$$

As $n \rightarrow \infty$, $P_N(n)$ decays as a geometric distribution, where the conditional probability is constant. This asymptotic conditional probability is derived in the Appendix (Sec. A.1.4) and is given by:

$$P_{n+\Delta n}(N, n) \xrightarrow{n \rightarrow \infty} \left(1 - \frac{1}{N}\right) \left[1 - \left(1 - \frac{1}{N}\right)^{\Delta n}\right]. \quad (5.17)$$

The *hazard rate* is the instantaneous conditional probability. In discrete probability functions, it equals the probability of occurrence in the next time step (e.g. Shaked *et al.*, 1995). It is given by Eq. 5.16 with $\Delta n = 1$.

5.5.2.5 Pattern of loading

In the model, the number of occupied cells increases rapidly just after a toppling, and then slows down (Fig. 5.4). This increase is due to the fact that, as a cycle progresses, there are more occupied cells, and thus it is less probable for an incoming ball to land on an empty cell. The evolution of the system in the next time step depends only on the state of the system in the current time step. If ρ_n is the fraction of occupied cells at time step n , there is a probability $1 - \rho_n$ for the next ball to be

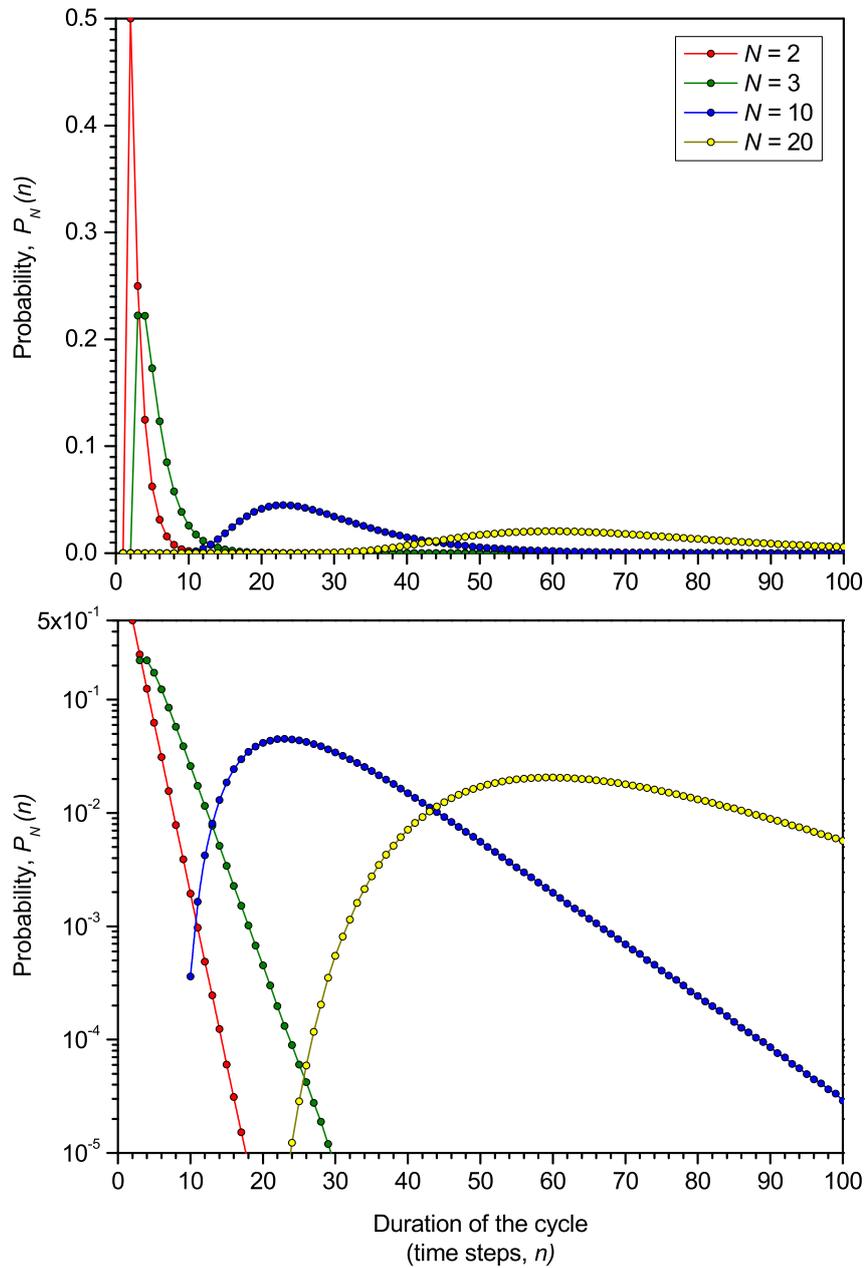


Figure 5.7: **Probability distribution function of the cycle duration (in time steps, n) in the box model, for various numbers of cells (N).** When the probability is represented in logarithmic scale (*below*), a geometric decay becomes a straight line. For $n \rightarrow \infty$, the distributions tend to geometric ones.

thrown to an empty cell. Because such a throw would increase ρ by $1/N$, the mean ρ at step $n + 1$ is

$$\langle \rho_{n+1} \rangle = \langle \rho_n \rangle + \frac{1}{N} [1 - \langle \rho_n \rangle]. \quad (5.18)$$

The box is empty at the beginning of the cycle ($\rho_0 = 0$), so from Eq. (5.18), the mean value of ρ_n is

$$\langle \rho_n \rangle = 1 - \left(1 - \frac{1}{N}\right)^n, \quad (5.19)$$

which approaches one asymptotically.

In real faults, the shear stress and elastic strain are expected to follow a similar trend, in which they increase faster just after a large earthquake, and more slowly later on (Sec. 5.2). In the box model, this behaviour emerges from the stress dissipation being proportionally larger as the cycle progresses, qualitatively in agreement with laboratory experiments (Fig. 5.1).

5.5.2.6 Stress shadow

The relaxation of a real fault by means of a large earthquake reduces the stress in the system. Thus a minimum time has to elapse before the fault accumulates enough stress to produce the next large earthquake. This effect is called *stress shadow* (e.g. Harris, 2000; Scholz, 2002). In the box model there exists a stress shadow: an earthquake cannot occur until the N th step in the cycle.

5.5.3 Fitting the box model to the Parkfield series

We will fit the cumulative probability function (Eq. 5.15) to the Parkfield series using the method of moments (Utsu, 1984). The first so-called *moments* of a statistical distribution are the mean and the variance. The method consists in making the model series to have the same mean and variance as the real series of events. Another method that could be used is that of maximum likelihood, which is more complex but typically yields similar results for quasiperiodic series (e.g. Utsu, 1984).

We have seen in Sec. 5.5.2.2 that the aperiodicity in the box model depends only on N . Thus, it is necessary to choose the N for which the aperiodicity is the closest to that of the Parkfield series, that is, $\alpha \simeq 0.3759$. The result is $N = 11$, for which, from Eq. (5.11), the aperiodicity is $\alpha = 0.3752$.

This model can be applied only to rather regular earthquake series, with coefficient of variation of the series of intervals $\alpha \leq 0.47$. To allow dealing with more irregular series, we later developed (Abadías *et al.*, 2006) a different model, based on a reduction of states (coarse-graining type) of the box model. In this new model the aperiodicity of the cycles is $0.47 < \alpha < 1$.

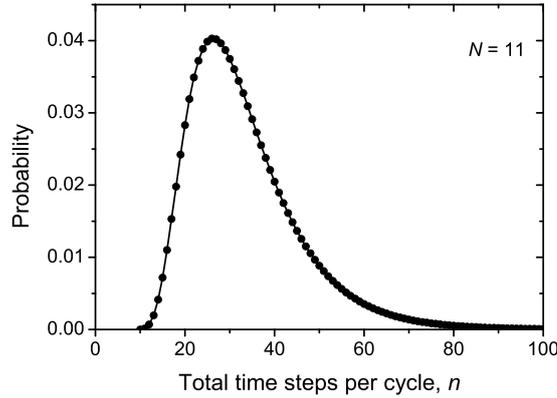


Figure 5.8: **Probability distribution function ($P_N(n)$) for the duration (in time steps, n) of the seismic cycle in the box model with $N = 11$.** This distribution has the same aperiodicity as the Parkfield earthquake series.

From Eq. (5.8) the mean value of n for $N = 11$ is $\langle n \rangle_{N=11} = 33.22$. Because the actual mean of the Parkfield series is $m = 24.62$ yr, one time step (ball throw) in the model is equivalent to $\tau = 24.62 \text{ yr} / 33.22 = 0.74 \text{ yr} \approx 9$ months. The discrete distribution function for the duration of the seismic cycle in a box model with $N = 11$, $P_{11}(n)$ is shown in Fig. 5.8.

Fig. 5.9 shows the evolution of the number of occupied cells for ten cycles with $N = 11$. Note the fluctuations in the duration of the cycles, which are consistent with the mean and the standard deviation of the series.

According to the box model, the minimum time between consecutive earthquakes (stress shadow) for the Parkfield series is $\tau N \simeq 8$ yr. All past intervals of the series are indeed larger than this value, so they cannot be used to reject the model for this case study. Moreover, at the time of this writing (November 2015), the open interval since the 2004 Parkfield mainshock is > 11 yr, also larger than the calculated period of stress shadow.

5.5.3.1 Earthquake probabilities at Parkfield with the box model

We now evaluate the quality of the box model fit for the Parkfield series and estimate the probability of the next earthquake in this fault segment.

In Fig. 5.10, the *empirical distribution function* of the Parkfield series is plotted. It is a cumulative step function ranging from 0 to 1.0, with a jump of $1/r$ at the duration of each of the r observed recurrence intervals in the sample ($1/6$ in this case). This function indicates the empirical probability (frequency) that a cycle lasts

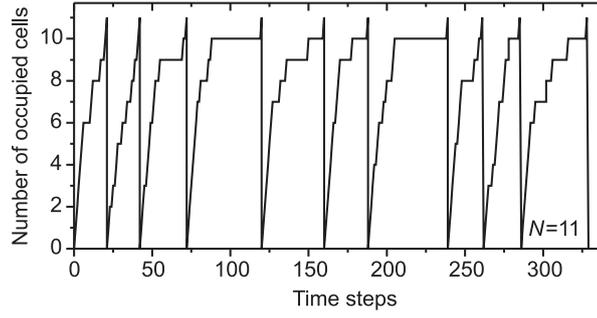


Figure 5.9: **Number of occupied cells during ten cycles of a box model with $N = 11$.** For this value, the model series has on average the same aperiodicity as the Parkfield earthquake series.

less or equal than in the time indicated in the abscissas. If the sample of cycles comes from a stable probability distribution, the empirical distribution function converges to this distribution as $r \rightarrow \infty$, according to the Glivenko–Cantelli theorem. The cumulative distribution of the box model in Eq. (5.15) for $N = 11$ with $\tau = 0.74$ yr also is drawn.

The goodness of this fit can be evaluated with a two-sided Kolmogorov-Smirnov test extended to discrete distributions (Arnold & Emerson, 2011). This nonparametric test is based on the maximum vertical distance between the empirical distribution function of the sample and the cumulative distribution function of the reference distribution. The larger the sample size (number of observed cycles in this case), the smaller that distance is allowed in order to not reject the fit with a specified level of statistical significance. According to this test, and given the small sample size (6 cycles), the fit to the Parkfield series cannot be rejected at any relevant level of statistical significance.

Now we calculate the yearly probability of the next earthquake, that is, the *conditional probability* of the next shock occurring in a certain year, given that it has not occurred previously. For the box model, from Eq. 5.16 it is

$$P_{\tau}(N, n) = \frac{A_N(n + 1/\tau) - A_N(n)}{1 - A_N(n - 1)}. \quad (5.20)$$

Note that $1/\tau$ is the number of time steps of the box model corresponding to one year. After calculating P_{τ} from Eq. (5.20), it is necessary to rescale the abscissas, n , to actual years, $n\tau + t_0$, where t_0 is the calendar year at which the last earthquake occurred ($t_0 = 2004.75$ for the Parkfield series). Fig. 5.11 shows the yearly probability for the new cycle at Parkfield according to the box model. During the first eight years after the last earthquake at Parkfield (which occurred in September 2004), the box model indicates that another big shock should not be expected. From

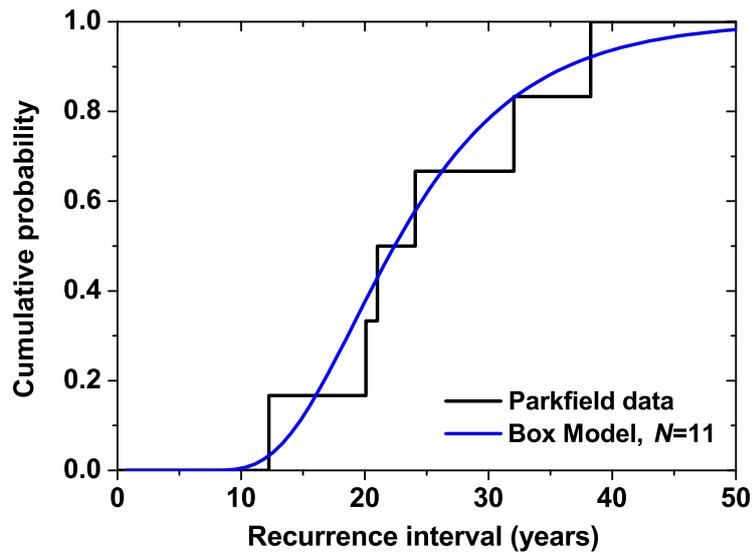


Figure 5.10: Fit of the cumulative distribution of the box model to the empirical distribution function of the Parkfield earthquake sequence. These distributions provide the probability that a cycle will be shorter than or equal to the recurrence interval indicated.

that time on, the probability of the next earthquake increases, tending to a constant equal to 11%, according to Eq. 5.17.

In the seismological literature there is a well-known question about the conditional probability of the next earthquake for a time much longer than the mean value of the series (Davis *et al.*, 1989): “The longer it has been since the last earthquake, the longer the expected time till the next?” Sornette & Knopoff (1997) discussed some statistical distributions that lead to affirmative, negative, or neutral answers to it. The result shown in Fig. 5.11 leads us to conclude that the box model produces a neutral answer. The reason is that for a long cycle duration (large n), the $P_N(n)$ of the box model decays geometrically (analog to exponentially if it were a continuous function), and asymptotically the box model behaves as a Poisson model, in which the conditional probability of occurrence of the next earthquake is constant (but different from the purely Poisson case, see Sec. 5.7).

5.6 Binary earthquake forecasting

In this section we consider the forecasting of the series of earthquakes in the framework of binary forecasts, in which the probability for the next event is not stated explicitly, but which can be evaluated in a probabilistic way (e.g. Mason, 2003; Molchan, 2003; Zechar & Jordan, 2008). In earthquake forecasting an “alarm” is sometimes turned on when it is estimated that there is a high probability for a large earthquake to occur (e.g. Keilis-Borok, 2002; Keilis-Borok & Soloviev, 2003;

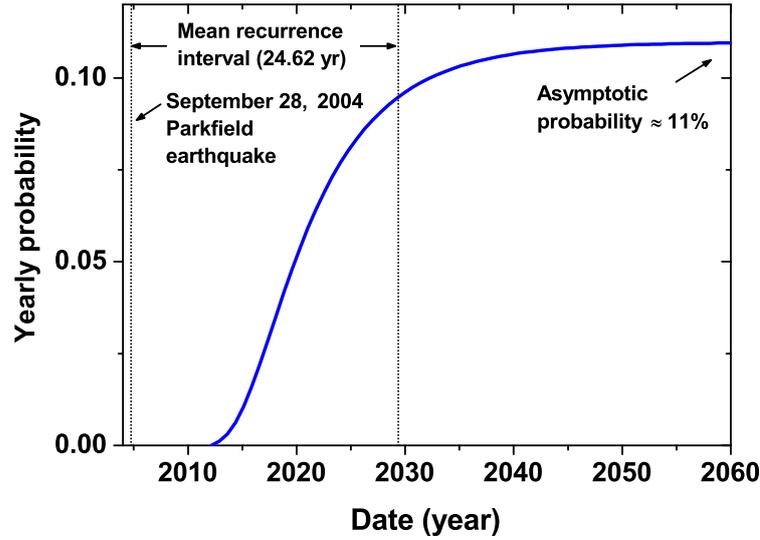


Figure 5.11: Yearly probability of the next mainshock at Parkfield, according to the box model.

Rundle *et al.*, 2003; Zechar & Jordan, 2008). If a large shock takes place when the alarm is on, the forecast is considered to be a success (a *hit*). If it takes place when the alarm is off, there has been a forecast failure (a *miss*). The *fraction of errors*, f_e , is the number of forecast failures divided by the total number of large earthquakes (equal to one minus the hit rate). The *fraction of alarm time*, f_a , is the ratio of the time during which the alarm is on to the total time of observation. A good strategy of forecasting must produce both small f_e and f_a , because both the forecast failures and the alarms are costly (e.g. van Stiphout *et al.*, 2010; Stein & Stein, 2014). As a general rule, the larger f_a , the smaller f_e . Depending on the trade-off between the costs and benefits of forecasting (Molchan, 1997), we can try to minimize a certain loss function, L . For simplicity, we will consider the simplest loss function, defined as

$$L = f_a + f_e. \quad (5.21)$$

The procedure described below to minimize L would proceed analogously if a different definition were used. For example, it would be possible to give more weight to f_e than to f_a (if failures are particularly costly) or vice versa (if it is specially impractical to declare the alarm for a long time). $L = 0$ would imply a perfect forecast, in which the instant of occurrence of every earthquake is forecast in advance (so $f_e = 0$) and the alarm time is infinitesimal ($f_a = 0$).

5.6.1 The random guessing strategy as baseline

A *random guessing strategy* consists in randomly turning the alarm on and off, and yields $L = 1$ in the long term. This result can be easily understood. The alarm will be on, randomly, during a certain fraction of time, f_a . Thus, there will be a probability $p = f_a$ for it being on when an earthquake eventually occurs, and a probability $p = 1 - f_a$ for it being off. So on average, when the number of cycles tends to infinity $f_e \rightarrow 1 - f_a$, so $L \rightarrow f_a + f_e = 1$. This strategy has two trivial cases: if the alarm is always on ($f_a = 1$), all the target earthquakes are “forecast” ($f_e = 0$). Conversely, if the alarm is always off ($f_a = 0$), all them are missed ($f_e = 1$). The random guessing strategy is considered as the simplest baseline (e.g. Molchan, 1997, 2003; Zechar & Jordan, 2008). When there is only a finite number of target earthquakes (especially when there are only a few), even a random guess may yield a $L \neq 1$, and it is necessary to evaluate statistically if the results are significantly different from those of a random guess (Zechar & Jordan, 2008).

In the long term, a forecasting strategy makes sense only if it gives $L = f_a + f_e < 1$. If $L > 1$, the strategy is performing exactly the opposite to how it should. Thus, the exact reverse strategy should be considered, and this will provide the opposite results ($f'_a = 1 - f_a$, and $f'_e = 1 - f_e$).

5.6.2 Reference strategy

The *reference strategy* is the simplest meaningful forecasting approach which can be considered for any quasi-periodic series of events (Fig. 5.12). It consists in turning on the alarm at a fixed time after the last big earthquake, and maintaining the alarm on until the next one (e.g. Newman & Turcotte, 2002; Molchan, 2003; Vázquez-Prada *et al.*, 2003). Then the alarm is turned off, and the same strategy is repeated, evaluating f_a and f_e for all cycles. The shorter this time, the larger f_a and the lesser f_e . This strategy has no false alarms; if the alarm is on, an earthquake will eventually occur, as long as the fault remains active in geological time scales.

To implement the reference strategy in the box model fitted to the Parkfield series, we consider that n is the number of time steps before sounding the alarm. If the earthquake occurs at the n th time step or earlier, the alarm has still not been sounded, and the earthquake should be considered a forecast failure. An earthquake in the box model cannot occur before the N th time step of each cycle, so $f_e = 0$ if and only if $n < N$.

In the long term, as the number of cycles tends to infinity, the analytical f_e and f_a depend only on the probability distribution of the cycle duration, and can be deduced analytically as follows. f_e is simply the fraction of cycles which last $\leq n$

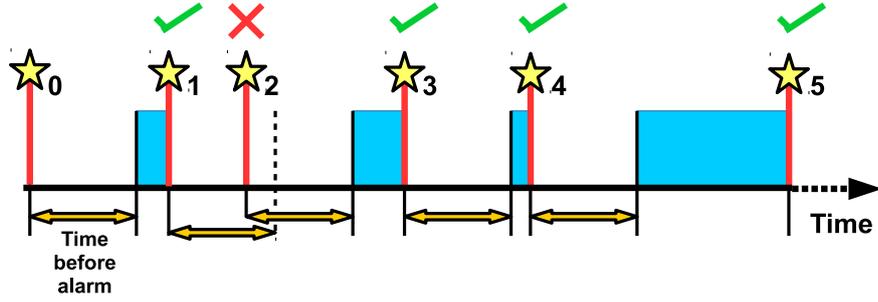


Figure 5.12: **Reference strategy for assessing the predictability of a quasi-periodic series of earthquakes.** Stars are the target earthquakes to be forecast (five in the example, excluding the initial one). In a quasiperiodic series, an alarm is turned on (*blue periods*) after some waiting time (*orange arrows*) since the last earthquake and maintained on until the next earthquake happens. An earthquake is considered as correctly forecast (*green ticks*) if it occurs when the alarm is on and as a forecast failure otherwise (*red cross*). The fraction of alarm time (f_a) will be the sum of periods during which the alarm is on divided by the total time. The fraction of errors (f_e) will be the number of forecast failures divided by the total number of target earthquakes (1/5 in the example shown).

time steps:

$$f_e(n) = \sum_{n'=1}^n P(n'). \quad (5.22)$$

While f_a is the following fraction of time:

$$f_a(n) = \frac{\sum_{n'=n+1}^{\infty} P(n')(n' - n)}{\sum_{n'=0}^{\infty} P(n')n'}. \quad (5.23)$$

To interpret the numerator of f_a , note that the alarm is turned on the cycles which last a minimum of $n' = n + 1$ time steps, but the alarm only lasts $n' - n$ time steps.

These relations are illustrated in Fig. 5.13(a), together with $L = f_a + f_e$. For each value of N , $L(n)$ has a minimum at a specific value of n , $n^*(N)$. As can be seen in Fig. 5.13(a), $n^*(11) = 19$, for which

$$f_a(n^*) = 0.432, \quad f_e(n^*) = 0.084, \quad L(n^*) = 0.516, \quad (5.24)$$

For the Parkfield sequence, n^* corresponds to

$$\tau n^* = 14.1 \text{ yr.} \quad (5.25)$$

If the distribution derived from the box model correctly describes the recurrence of large earthquakes at Parkfield in the long term, the alarm connected at this time since the beginning of the cycles and disconnected just after the occurrence of each shock would yield the results given in Eq. (5.24). Note that this time is substantially smaller than the mean duration of the cycles, $m = 24.62 \text{ yr}$.

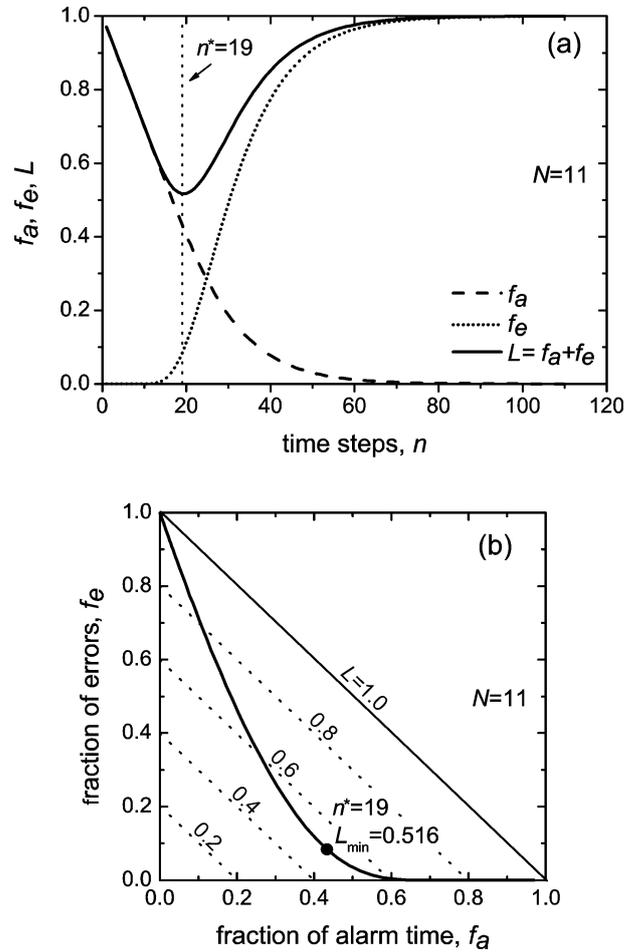


Figure 5.13: **Performance of the reference forecasting strategy in the box model applied to the Parkfield series.** *Top:* Fraction of errors (f_e), fraction of alarm time (f_a), and loss function ($L = f_a + f_e$) as a function of the time before the alarm in each cycle (time steps, n) for the forecasting strategy in a box model with $N = 11$ cells. *Bottom:* Error diagram for this strategy. Each point on the curve is the result of using a different value of n . The large dot corresponds to n^* , for which the loss function reaches a minimum. The diagonal lines are isolines of L . A random guessing strategy would render $L = 1$.

The quality of the model-earthquake forecast also can be understood visually by means of an error diagram, (Fig. 5.13b), in which f_e is plotted versus f_a (e.g. Molchan, 1997, 2003; Zechar & Jordan, 2008). This plot is analogous to the ones used in Ch. 2 (note that the hit rate equals $1 - f_e$).

5.7 Multi-model approach

This section proceeds analogously as above, but considering several candidate models, instead of just one. After fitting other renewal models to the Parkfield series, the resulting probabilities for the next earthquake will be calculated, emphasizing the similarities and differences between the models, and comparing them to a time-independent estimate. The section ends discussing how to use the models for binary forecasting and mentioning recent, relevant developments.

5.7.1 Fits with other renewal models

In this section, five previously existing renewal models are briefly described and fitted to the Parkfield series. Three of these distributions, the gamma (G), lognormal (LN), and Weibull (W) have been classically used for earthquake sequences in individual faults, and in a the broader context of reliability and time-to-failure problems. The other two distributions are derived from two simple physical models of earthquake recurrence, which have the virtue of providing an intuitive picture of the seismic cycle in a fault or fault segment. These are the Brownian Passage Time Model (BPT, Kagan & Knopoff, 1987; Ellsworth *et al.*, 1999; Matthews *et al.*, 2002; Working Group on California Earthquake Probabilities, 2003), and the Minimalist Model (MM, Vázquez-Prada *et al.*, 2002; Gómez & Pacheco, 2004, and Ch. 6). The first one represents the tectonic loading of a fault by a variable which evolves by superposition of an increasing linear trend and a Brownian noise (Parisi, 2005) term, and an earthquake occurs when this variable reaches a given threshold (Matthews *et al.*, 2002). All the earthquakes in this model are identical to each other. The Minimalist Model sketches the plane of a seismic fault, where earthquake ruptures start and propagate according to simplified breaking rules (Ch. 6). This model generates earthquakes of various sizes, and only the time between the largest ones (the *characteristic earthquakes*, that break the whole model fault, Schwartz & Coppersmith, 1984) are considered for the inter-event time distribution. The distributions derived from these two models, as well as the gamma, lognormal and Weibull, generally represent fairly well the observed distribution of large-earthquake inter-event times in individual faults (Gómez & Pacheco, 2004). However, they differ significantly in their probability estimates

for times much longer than the mean inter-event time of the data. Thus, it seems convenient to take all their different forecasts into account.

Now, we will proceed to fit the Parkfield series using the G, LN and W families of distributions (Utsu, 1984) and the BPT and MM models (Matthews *et al.*, 2002; Gómez & Pacheco, 2004). The statistical distribution of inter-event times in the BPT model is the so-called inverse Gaussian distribution (Kagan & Knopoff, 1987), which, as the three classical distributions mentioned at the beginning, is a continuum biparametric density distribution. The distribution derived from the MM is discrete, and strictly speaking has only one parameter, N , the number of cells in which the model fault plane is divided, directly related to the aperiodicity α of the series (Gómez & Pacheco, 2004), similarly as for the box model. However, for fitting the data, it is necessary to assign a definite number of years to each non-dimensional time step of the model. This second parameter will be called τ , as for the box model (Sec. 5.5.3). The distribution in the MM has an analytical solution, explicitly written for $N \leq 5$ by Gómez & Pacheco (2004). For larger values of N the distribution can be calculated numerically with Monte Carlo simulations.

Next, we will write down the explicit analytic form of the four mentioned continuum probability density distributions. Each of them has a *scale parameter* (which stretches or shrinks the distribution) and a *shape parameter* (which controls the shape of the distribution). These distributions specify the probability that the duration of a cycle equals the time t .

Gamma distribution:

$$G(t) = \frac{c}{\Gamma(r)} (ct)^{r-1} e^{-ct}, \quad c > 0, r > 0 \quad (5.26)$$

Lognormal distribution:

$$LN(t) = \frac{1}{\sqrt{2\pi}\sigma t} \exp\left[-\frac{(\ln t - n)^2}{2\sigma^2}\right], \quad n > 0, \sigma > 0 \quad (5.27)$$

Weibull distribution:

$$W(t) = a\rho t^{\rho-1} \exp(-at^\rho), \quad a > 0, \rho > 0 \quad (5.28)$$

Brownian Passage Time distribution (Matthews *et al.*, 2002):

$$BPT(t) = \left(\frac{m}{2\pi\alpha^2 t^3}\right)^{1/2} \exp\left[-\frac{(t-m)^2}{2m\alpha^2 t}\right], \quad m > 0, \alpha > 0 \quad (5.29)$$

In this last case, the parameters m and α correspond to the mean and aperiodicity defined earlier.

Gamma	$c = 0.287 \text{ yr}^{-1}$	$r = 7.078$
Lognormal	$n = 3.137$	$\sigma = 0.364$
Weibull	$a = 6.853 \times 10^{-5} \text{ yr}^{-\rho}$	$\rho = 2.889$
BPT	$m = 24.62 \text{ yr}$	$\alpha = 0.3759$
MM	$N = 495 \quad (\alpha = 0.3759)$	$\tau = 4.186 \times 10^{-3} \text{ yr}$
BM	$N = 11 \quad (\alpha = 0.3752)$	$\tau = 0.74 \text{ yr}$
Exponential	$\lambda = 1/24.62 \text{ yr}^{-1}$	

Table 5.1: **Parameter values of the fits to all the models used for the Parkfield series.** Six are renewal models, for which probabilities depend on time since the last earthquake: Gamma, lognormal, Weibull, Brownian Passage Time (*BPT*), Minimalist Model (*MM*) and Box Model (*BM*). The exponential is a time-independent (Poisson) model.

We will use the method of moments to fit the data, so within these four families of distributions, and the same for the MM, we will select that specific distribution whose mean value and aperiodicity are equal to those of the Parkfield series (Sec. 5.4). The specific values of the parameters that make the different distributions fulfill this condition are written in Table 5.1.

For reference (albeit not included in our original publications, González *et al.*, 2005, 2006a), I will also consider here explicitly the exponential distribution (which is also a special case of the G and W). This is the inter-event time distribution resulting from a Poisson process, in which the the time since the last even is irrelevant. It has only one parameter, the rate at which events occur per unit time (λ), which is the inverse of the mean of the intervals ($\lambda = 1/m$).

Exponential distribution:

$$E(t) = \lambda e^{-\lambda t}. \quad (5.30)$$

Note that in the MM model the aperiodicity of the series fixes the value of the parameter N (Gómez & Pacheco, 2004). In a minimalist system with $N = 495$, the mean recurrence interval of the characteristic earthquakes is 5881.2 non-dimensional time steps. Comparing this mean with the value $m = 24.62 \text{ yr}$ (Sec. 5.4), we deduce that one time step of the model corresponds to $\tau = 24.62 \text{ yr}/5881.2 = 4.186 \times 10^{-3} \text{ yr}$, or around 1.5 days.

5.7.2 Comparison of probability estimates at Parkfield

Fig. 5.14 shows the cumulative histogram (empirical distribution function) of the Parkfield series together with the cumulative distributions of the seven models.

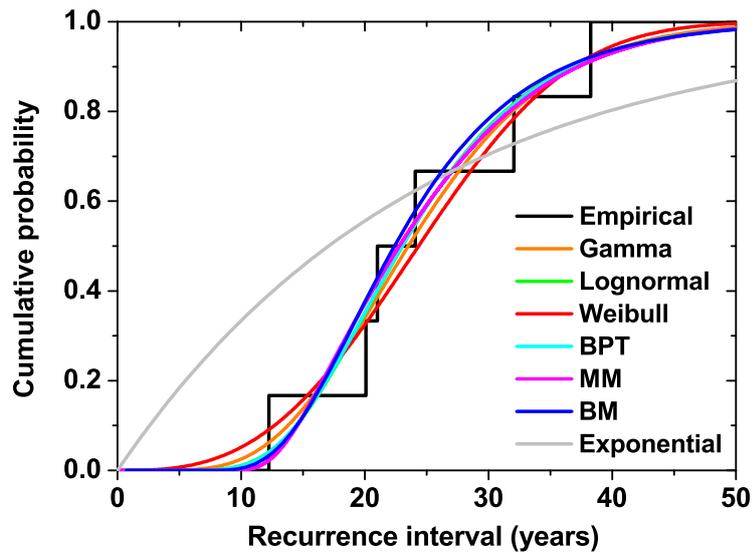


Figure 5.14: Fit to the Parkfield sequence with all the models considered.

These, for the G , LN , W , BPT and E are obtained by integrating Eqs. 5.26–5.30, and for the BM and MM by summing their discrete probability distributions. Finally, Fig. 5.15 presents the annual (conditional) probability of occurrence derived from the five models.

The results shown in Fig. 5.14 indicate that the six renewal models used in the adjustment, despite being different from each other, describe well the Parkfield data. The exponential model provides the worst fit, because it expects a higher frequency (probability) of short intervals than the observed one. In the original papers (González *et al.*, 2005, 2006a), we discussed qualitatively the residuals of the fits. Here, instead, as done with the box model in a previous section, I assessed the goodness of the fits using the two-sided Kolmogorov-Smirnov test, either the classical one for the continuous distributions or the version newly extended to discrete distributions (Arnold & Emerson, 2011) when dealing with the fits from the MM and BM . The exponential fit can be rejected only at a $\simeq 24\%$ level of significance (p value = 0.2439), so on average, approximately in one out of four sequences, with six intervals each, actually generated from a exponential distribution the fit can be as poor as in Parkfield. It is thus unlikely, but not impossible, that the series is Poissonian. Bakun *et al.* (2005), in their supplementary materials, argued that the exponential fit could be “rejected with 98% confidence ($p = 0.016$)”, but apparently used a flawed Kolmogorov-Smirnov test. For the six renewal models, the residuals are too small to be statistically significant, so no fit can be rejected at any relevant level of significance, and no one is significantly better than the others. So for the

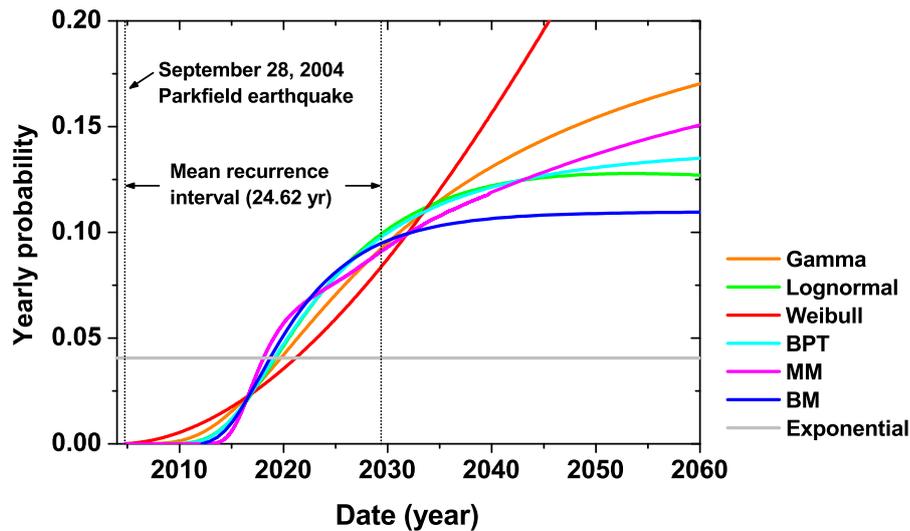


Figure 5.15: Annual probability for the new mainshock at Parkfield according to all the models.

Parkfield series no particular renewal model can be advocated as the optimal choice for forecasting purposes.

The cumulative probability is relatively similar, but not entirely alike, for the six renewal models (Fig. 5.14). The curves corresponding to the BM and MM take off later than the G, LN and W, because in the former models there is an initial stress shadow, in which no new event can occur. In the MM it has a duration of $N = 495$ time steps, i.e. $495 \times 4.186 \times 10^{-3}$ yr, that is 2.07 yr, shorter than the 8.1 yr of the BM. All the other, continuous renewal models yield low but non-null cumulative probability during the first years. The LN and BPT curves plot one upon the other in 5.14 because in the time range shown in the graph their cumulative probability distributions are similar. Note also that the Weibull model forecasts a cumulative probability considerably higher for $t < 10$ yr than the other four models, and that all six cumulative distribution functions appear to ‘converge’ in probability roughly 18 yr after the last mainshock, with a cumulative probability of around 25%. All the renewal models indicate that an immediate re-rupture of the Parkfield segment is unlikely (the cumulative probabilities are lesser than 1% for the first five years), but it will likely occur not later than 53 years after the last one (moment at which all cumulative probabilities are at least 99%). Note that these results were submitted to publication (González *et al.*, 2005, 2006a) just after the last Parkfield earthquake. The open interval at the time of this writing (over 11 years, in November 2015) is already longer than the stress shadow of the MM and BM, so the next earthquake is already possible according to all models.

The annual probability (Fig. 5.15) is relatively similar for all models until the year 2035, but there are several differences worth comment. In all these models, the probability initially rises with time since the last mainshock. In contrast, if the exponential model, the yearly probability is constant ($= \lambda$). At the beginning of the cycle the W curve is the first in the take off and the MM and BM are the last ones. This reflects what was mentioned in the previous paragraph. Later, there is an interval, roughly speaking from 2016 to 2022, in which the MM curve is on top of the others, forecasting slightly higher annual probabilities. The forecasts from the LN and the BPT are very similar until about 2040, because the BPT distribution is similar to, but not identical to, a lognormal distribution (Michael, 2005). Around 2030, when the mean recurrence interval of the series has elapsed, all the models forecast a yearly conditional probability between 8.5% and 10%.

But from 2035 onwards, the six renewal models start showing their asymptotic behaviour, or, in other words, their clear discrepancies. The behaviour of the conditional probability for long times after a large earthquake in specific faults has been long debated (e.g. Davis *et al.*, 1989; Sornette & Knopoff, 1997), and the different models used in this paper show three different possibilities for it. First, a decreasing probability, the case of the LN model, where the probability starts declining from the year 2053 onwards, approaching zero as time passes. Second, a probability which increases asymptotically to 100%, the case of the W model, according to which there is a 95% yearly probability of having a mainshock after 163 yr from the last one. And third, a probability that increases towards an asymptote smaller than 100%, the case of the curves forecast by the BPT, G, BM and MM models. The asymptotic yearly probability value is different for each of these three models: 13%, 26%, 11% and 38% respectively. This last value in the BM and MM is given by

$$\lim_{n \rightarrow \infty} P(n, \tau) = \left(1 - \frac{1}{N}\right) \left[1 - \left(1 - \frac{1}{N}\right)^{\frac{1}{\tau}}\right], \quad (5.31)$$

that is, as Eq. 5.17, where n is the number of time steps since the last characteristic earthquake in the model, and the number of time steps corresponding to a year is $\Delta n = 1/\tau$. The formula is identical for both the BM and MM since the cycle in both models finishes in the same way, with $N - 1$ cells occupied and a probability of $1/N$ for the largest earthquake to occur in the next time step (sections 5.5.2 and 6.2.2). But the asymptotic probability is different in each model, since N is also different (Table 5.1).

The discrepancies between the forecasts of these six renewal models can be considered all together to give reasonable upper and lower bounds to the annual probability of occurrence at Parkfield during the first decades since the last mainshock: about 2.5% in 2017 (after half of the recurrence interval m), between

8.5% and 10% in 2030 (after one m) and between 11% and 17% in 2042 (after 1.5 m). In contrast, the exponential model provides a constant conditional probability, which may initially overestimate the hazard and underestimate it later on.

5.7.3 Binary forecasts

Similarly as done for the box model (Sec. 5.6), we can devise a simple binary forecast, evaluating for all models the optimal moment to declare an alarm after the occurrence of the last mainshock, in order to minimize in the long term the loss function $L = f_a + f_e$ chosen as reference. All renewal, time-dependent models have some degree of predictability, but not the exponential (Poisson), which is unpredictable and would yield a result identical as the random guessing strategy ($L = 1$). Figure 5.16 shows the value of L as a function of the time waited for before connecting the alarm. The optimum times to declare the alarm (t^*), yielding the lowest L , are model-dependent. They are listed in Table 5.2, alongside their associated results. The corresponding dots in the error diagram are plotted in Fig. 5.17. Assuming that the statistical properties of the sequence can be extrapolated into the future, all models indicate that the alarm (for this definition of L) should be declared 13–17 years after the last earthquake, what would yield a low long-term f_e . If this reference strategy could had been applied retrospectively to the Parkfield sequence, all renewal models will have missed one out of six target earthquakes in the series (the one occurred in 1934, which closed the shortest cycle). The models with lowest long-term $L(t^*)$ and $f_e(t^*)$ are the MM and BM, while Weibull provides the highest values.

The reference strategy can be also applied retrospectively to the observed series of cycles without using any model. In that case, the best results so far would have been to declare the alarm at a time $t^* \sim 12.245$ yr, just slightly lower than the shortest cycle (12.246 yr). This would have resulted in no miss ($f_e = 0$), $f_a \simeq 50\%$ until the 2004 mainshock, and $f_a \simeq 46\%$ considering the current open interval until November 2015.

The optimal dates t^* turn out to be approximately coincident with the moments at which the conditional probability surpasses that of the time-independent exponential model (Fig. 5.15). This can be explained by the results of Molchan (2003), who considered a more generic alarm function (of which the one used here is a particular case). He showed that, for renewal models where the hazard rate is continuously increasing since the last earthquake (such as for the gamma, Weibull, BM and MM), the optimal moment to declare the alarm is when the hazard rate reaches that of the exponential model. For the LN and BPT distributions, the hazard rate (and conditional probability) decreases towards zero after some time since

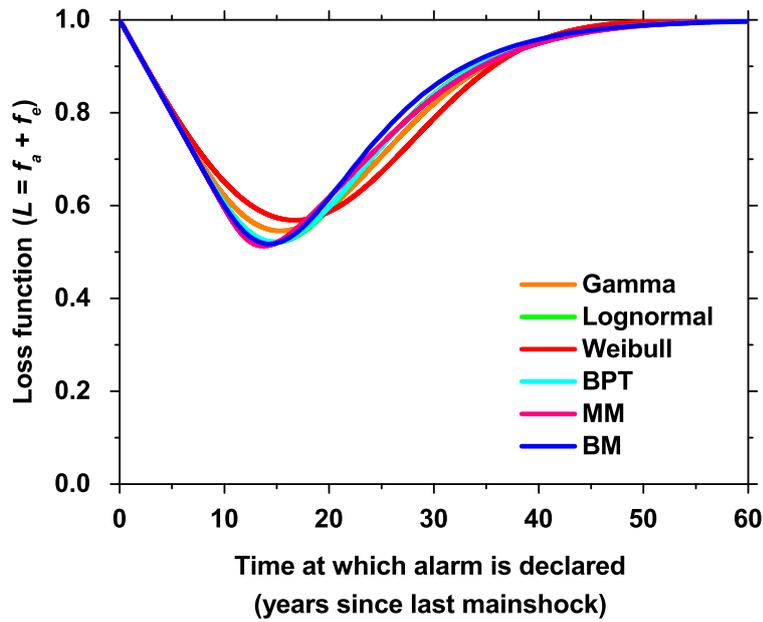


Figure 5.16: Optimization of the reference forecasting strategy for all the renewal models applied to the Parkfield sequence. The optimum moment for declaring the alarm would be that in which the loss function L reaches a minimum. In a Poisson process (exponential recurrence model), $L = 1$ in the long term regardless of when the alarm is declared.

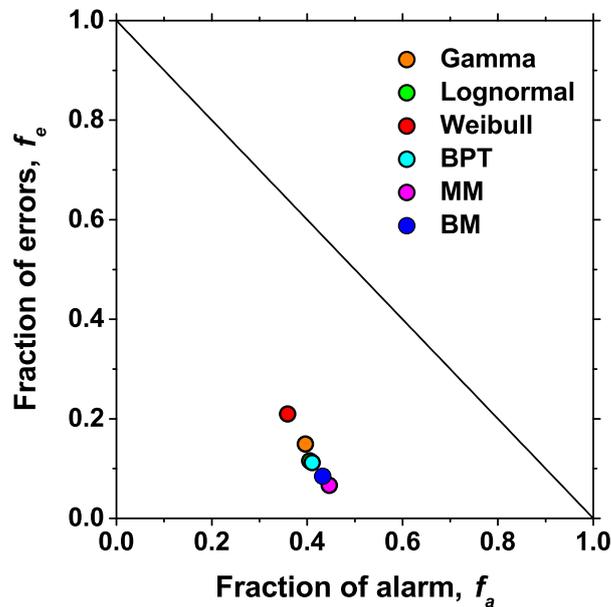


Figure 5.17: Error diagram of the optimum results of the reference forecasting strategy for all the renewal models applied to the Parkfield sequence. These optimum results correspond to forecasting most earthquakes 79–92% ($=1 - f_e$), at the expense of declaring the alarm during 35% to 45% of the total time ($=f_a$). See Table 5.2 for details. In a Poisson process (exponential recurrence model), the results would lie on the diagonal line.

Model	t^* (years)	f_a	f_e	$L = f_a + f_e$
Gamma	15.3	0.396	0.149	0.545
Lognormal	14.9	0.405	0.116	0.521
Weibull	16.7	0.359	0.210	0.569
BPT	14.8	0.410	0.112	0.522
MM	13.7	0.446	0.066	0.512
BM	14.1	0.516	0.432	0.084

Table 5.2: **Optimum results of the reference forecasting strategy applied to Parkfield.** t^* is the time after the last mainshock at which the alarm should be declared to obtain the minimum L (Fig. 5.16. Fig. 5.17 shows the corresponding error diagram.

the last mainshock, so it would give slightly better results (lower loss function) to turn the alarm on only during the period when the hazard rate is above the time-independent threshold (intersection observed in Fig. 5.15), and turning it off again when the hazard rate decays below such threshold (Molchan, 2003). These considerations stress the links between these binary, alarm-based forecasts and the purely probabilistic ones.

5.7.4 Later developments

Several later works have applied renewal models to the Parkfield sequence. Zöller *et al.* (2007) applied a more complex numerical model to fit the recurrence intervals in Parkfield. Abaimov *et al.* (2008) proposed to use the Weibull distribution, already used here (González *et al.*, 2006a). More recently, Tejedor *et al.* (2012, 2015), building upon the results presented here, proposed a negative binomial renewal model for the recurrence of large earthquakes, and applied it to the Parkfield series. In that model there is also a stress shadow (of four years), after which the yearly conditional probability rises towards a constant value. After the year 2015, this model provides lower yearly probabilities than the other renewal models considered above; after one mean recurrence interval the value is only about 6%. With the same reference forecasting strategy used here, in that model the optimal time to turn the alarm on would be at 15.0 years since the last mainshock (see Tejedor *et al.*, 2012, 2015, for details).

In a clustered series ($\alpha > 1$), the optimal binary forecast, yielding the lowest loss function, is attained by turning the alarm on immediately after each earthquake, and turning it off after some time. This is because in such situations the hazard rate (or conditional probability) for the next earthquake decreases with time since the last one (Molchan, 2003). Then, an optimal waiting time has to be found when

to turn *off* (instead of *on*) the alarm after each earthquake in order to minimize the loss function in the long term. For clustered sequences in numerical models this has been explored, for example, by Moreno *et al.* (2003) and Tejedor *et al.* (2009a). In real situations, the best examples are aftershock sequences, for which the highest probabilities, and the highest risk, are forecast to occur just after the largest earthquakes (e.g. Marzocchi *et al.*, 2015), and sometimes are high enough to ask for evacuation of the most vulnerable buildings in a cost-benefit analysis of alarms (van Stiphout *et al.*, 2010).

Insights on how to synchronize models with seismic faults for earthquake forecasting

It is far better to foresee even without certainty than not to foresee at all.

Henri Poincaré (1913, p. 129).

Chapter adapted and expanded from:

Álvaro González; Miguel Vázquez-Prada; Javier B. Gómez & Amalio F. Pacheco (2004):

Using synchronization to improve the forecasting of large relaxations in a cellular automaton model.

Europhysics Letters, 68 (5), 611–617.

Álvaro González; Miguel Vázquez-Prada; Javier B. Gómez & Amalio F. Pacheco (2006):

A way to synchronize models with seismic faults for earthquake forecasting: Insights from a simple stochastic model.

Tectonophysics, 424 (3–4), 319–334.

6.1 Introduction: Data assimilation in models of seismic faults

Numerical models are now frequently used to simulate the seismic behaviour of faults (e.g. Kato & Seno, 2003; Fitzenz and Miller, 2004; Kuroki *et al.*, 2004; Ben-Zion,

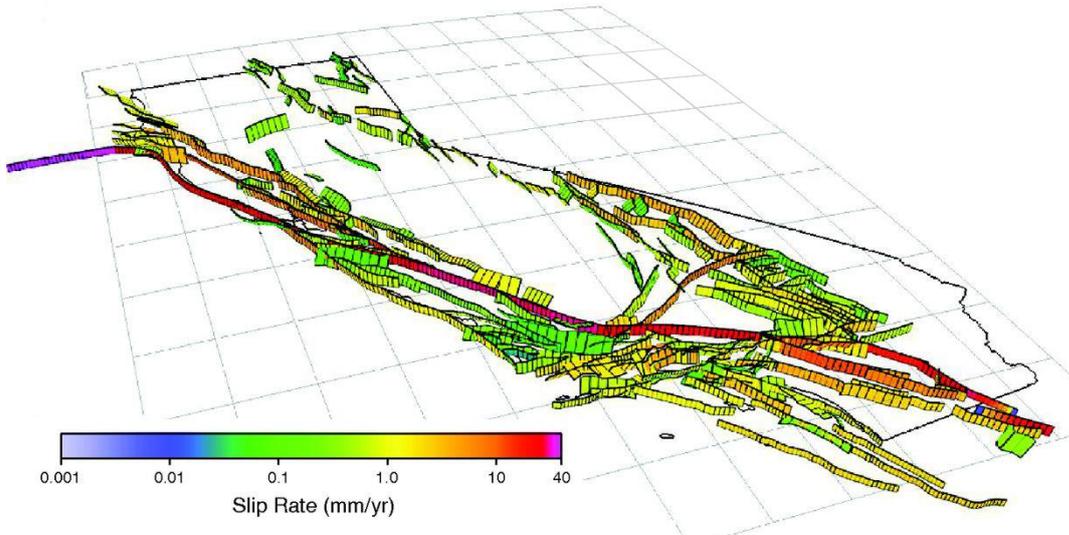


Figure 6.1: **Fault model used in the Uniform California Earthquake Rupture Forecast, version 3 (UCERF3).** Figure adapted from Field (2015), (© Seismological Society of America, used with permission). For a relief map of California for comparison, see Fig. 5.2. This is an example of a state-of-the-art, complex model used for computing earthquake ruptures and long-term earthquake probabilities in a fault network. The geometry of the faults (their traces and dips) is provided by geological mapping, and their long-term slip rate (*colour scale*) is constrained by geological and geodetic measurements. Each fault is primarily discretized as an array of subsections, down to the base of the seismogenic crust; a finer grid was used for physics-based simulations (Field, 2015). UCERF3 is the most recent authoritative hazard model for California produced by the United States Geological Survey and will be used to update “building codes, earthquake insurance products, emergency plans, and other risk-mitigation efforts” (Field *et al.*, 2015b).

2008) and fault networks (e.g. Ward, 2000; Hashimoto, 2001; Robinson & Benites, 2001; Rundle *et al.*, 2001; Soloviev and Ismail-Zadeh, 2003; Robinson, 2004; Rundle *et al.*, 2004; Field, 2015; Yoder *et al.*, 2015). In these models, fault planes separate lithospheric blocks that are strained at specific rates, and sudden slips (earthquakes) are generated by the faults according to certain friction and/or rupture laws. These simulations are becoming progressively more realistic and already play a substantial role in scientific studies of earthquake probability and hazard (e.g. Ward, 2000; Field, 2015, Fig. 6.1).

The final goals of the numerical modelling of seismicity are not different from, for example, the goals of numerical models of the atmosphere. A good model should be able to:

1. reproduce the general characteristics of the system,
2. mimic the state of the system at the present moment, and
3. forecast the future evolution of the system.

Most numerical models of seismicity have been designed to achieve the first goal, by reproducing general characteristics of earthquakes such as their size-frequency distribution (e.g. Bak and Tang, 1989; Olami *et al.*, 1992; Dahmen *et al.*, 1998; Preston *et al.*, 2000; Vázquez-Prada *et al.*, 2002), or the generation of aftershock and foreshocks (e.g. Hainzl *et al.*, 1999). When a model is designed this way, it is left to evolve freely according to its rules, and all that is checked is whether the overall results of the model are similar to the observations made in Nature or not.

The second goal requires *data assimilation*, that is, the process of absorbing and incorporating observed information into the model. By this process, the model is tuned and synchronized, at least partially, with the real system it tries to simulate. In a meteorological model, data of atmospheric pressure, temperature, humidity, cloud cover, precipitation, etc. measured in a given moment at different locations and heights can be included. With this procedure, the model becomes a reasonably good representation of the atmosphere at that moment. Then it can be used to calculate the probable future atmospheric evolution (i.e. the third goal cited above).

Seismic data assimilation poses greater problems than its meteorological equivalent. This explains (at least partially) the relative delay in developing reliable forecasts of large earthquakes. The inner workings of both the atmosphere (e.g. Houghton, 2002) and the lithosphere (e.g. Goltz, 1997; Turcotte, 1997; Keilis-Borok, 2002) are complex and chaotic, so they are inherently difficult to forecast. The atmosphere is indeed a classical example of chaotic system (Lorenz, 1963), what limits the predictability achieved by weather forecasts (e.g. Mason, 2003). However, while meteorologists can probe the atmosphere every day at different places and heights (and assimilate the obtained data in their models in near real-time), lithospheric variables of paramount importance, such as the stress and strain, can be measured only in certain places, and not at any time: earthquakes have unobservable dynamics (e.g. Rundle *et al.*, 2003). For example, the best current compendium of stress magnitudes and directions in the lithosphere is the World Stress Map (Zoback, 1992; Heidbach *et al.*, 2008), whose entries are point static time-averaged estimates of maximum and minimum principal stresses in space. And the direct measurements of stress on active fault zones at depth are still scarce (e.g. Ikeda *et al.*, 2001; Tsukahara *et al.*, 2001; Yamamoto and Yabe, 2001; Hickman and Zoback, 2004; Boness and Zoback, 2004). The dynamical models would need better spatial and temporal information of stress, both more abundant and more systematically collected than that currently available (Rundle *et al.*, 2004). It is thus necessary to seek ways to tune and synchronize the models with more abundant observable data.

A first step of data assimilation in models of earthquake faults is to introduce information regarding the geometry (shape and location) and *topology* (interrelations) of the active faults and their long-term behaviour. For example, the long-term fault slip rate, and the average recurrence interval of the largest earthquakes in the fault can be estimated from paleoseismological studies and should be included in the models (Grant and Gould, 2004). Examples of this approach are the works of Rundle *et al.* (2001); Rundle *et al.* (2004); Robinson (2004) and Field (2015, Fig. 6.1). The surface deformation measured via Global Positioning System (GPS) networks and by Synthetic Aperture Radar Interferometry (InSAR) can also constitute input data for the dynamical fault models (Rundle *et al.*, 2004).

Earthquakes themselves are indeed the most obvious observable events of lithospheric dynamics, and could provide the most detailed data available to assimilate in the models, but how? The earthquake rupture area could be an important clue.

The rupture area and slip distribution in real earthquakes can be very complex (e.g. Sieh, 1996; Langridge *et al.*, 2002; Haeussler *et al.*, 2004; Kanamori & Brodsky, 2004), but can be estimated in a variety of ways. The actual slip distribution can be obtained by inverting the observed seismic waveforms (e.g. Cohee & Beroza, 1994; Lee & Ma, 2000; Kanamori & Brodsky, 2004) or tsunami waveforms (e.g. Tanioka *et al.*, 2004; Baba & Cummins, 2005; Romano *et al.*, 2014), and/or by geodetic modelling of surface displacement (e.g. Yabuki & Matsuura, 1992). Some earthquakes produce surface ruptures, which are useful for estimating the rupture area (e.g. Stirling *et al.*, 2002). Although most surface ruptures occur in large shocks, with magnitudes larger than about 6, they have been reported for earthquakes with magnitudes down to 2.5 (see the compilation of historic earthquakes with surface rupture by Yeats *et al.*, 1997, pp. 473-485). Also, the rupture area can be estimated from the seismic moment (typically calculated from the amplitude spectra of seismic waves; Scholz, 2002; Kanamori & Brodsky, 2004), or from the moment magnitude (e.g. Wells & Coppersmith, 1994; Stirling *et al.*, 2002; Dowrick and Rhoades, 2004; Stirling *et al.*, 2013). Frequently the location of early aftershocks is used to determine the rupture area of the mainshock (e.g. Wells & Coppersmith, 1994), although the aftershock zone tends to grow with time (Kisslinger, 1996) and is not necessarily a good indicator of that area (Yagi *et al.*, 1999).

Complex models with realistic fault geometry (e.g. Fig. 6.1) are able to reproduce the rupture area and coseismic slip of historical earthquakes. It is thus possible to force the model to reproduce the rupture of a historical earthquake, and let it evolve from that moment onwards to see what could happen in the future. For example, this was already achieved by Ward (2000), who developed a model including the

network of main faults in the San Francisco Bay Area (California). He forced the model to reproduce the San Andreas Fault surface coseismic slip of the 1906 San Francisco earthquake, and let it evolve freely from that earthquake onwards, in an attempt to simulate the probable sequence of earthquake ruptures during the next 3000 years.

But considering only the data of the largest earthquake in the series is probably not sufficient to properly synchronize the model. Complex and chaotic systems are very sensitive to the initial conditions. The information regarding only one event probably does not sufficiently constrain the initial conditions, and the calculated evolution will probably be a particular case of a large range of possible outcomes. Could this panorama improve by forcing the model to reproduce all the observed earthquake ruptures, including the small ones?

To check whether this idea works, at least to forecast *synthetic seismicity*, is the purpose of this chapter. The number of large earthquakes recorded in individual faults is typically scarce (very rarely more than ten large events, Ch. 5). This hampers the ability to characterize statistically the effectiveness of any forecasting method. In contrast, synthetic earthquake catalogues, generated by numerical models of seismicity, can be as long as desired. This enables to ascertain, with robust statistics, whether a forecasting strategy could be useful, before endeavouring to apply it to real earthquake series.

In the following sections, the goal will be to forecast the largest earthquakes generated by the minimalist model, a simple numerical fault model. We will show that when all the earthquake ruptures generated by this model are imposed on other, similar models, these become partially synchronized with the former. We use them to declare alarms that efficiently mark the occurrence of the largest shocks in the first model. The results are much better than those obtained with other strategies that consider only the earthquake series. The model, albeit simple, is stochastic (it involves randomness), so its efficient forecasting is not trivial. We will describe how this stochasticity can be dealt with, by using an approach similar to the so-called ensemble forecasting used in Meteorology (e.g. Palmer *et al.*, 2005). The method could be used in other more detailed and realistic models (stochastic or not) to test our general conclusion: that they might be partially synchronized with actual faults by being forced to reproduce the series of observed earthquake ruptures.

The next section describes the minimalist model and its properties. Then, we outline the general scheme of forecasting and the strategies used as reference to assess the merits of any other forecasting method in the model. Finally, the method based on partial synchronization is explained and its possible utility discussed.

6.2 The minimalist model

The minimalist model is the numerical model whose largest earthquakes we will try to forecast. It was introduced in a previous work (Vázquez-Prada *et al.*, 2002), and has two main, apparently contradictory, advantages for the purpose of this chapter: it is simple but, at the same time, it is difficult to forecast. Because it is simple, several of its properties can be derived analytically, and it can be characterized in detail with numerical simulations which do not require an impractical amount of computer time. Because it is stochastic, it is difficult to forecast, so the results we will obtain here are not trivial. In the following paragraphs we will explain how the model works, and what are its main properties, comparing them with those of actual faults.

6.2.1 Rules

The model is a simple (hence its name) cellular automaton. Cellular automata are frequently used to model seismicity (see e.g. Malamud & Turcotte, 2000; Jiménez, 2013, for a recent review). When applied to modelling the generation of earthquakes on seismic faults, these models represent a fault plane as a grid of cells (each cell representing a fraction of the fault area), and the time evolves in discrete time steps. Each cell's state is updated at each time step according to rules that usually depend on the state of the cell or that of its neighbours in the previous time step. These rules can be designed according to certain friction laws (e.g. Ben-Zion, 2001), stress transfer (e.g. Olami *et al.*, 1992; Hainzl *et al.*, 1999; Preston *et al.*, 2000; Rundle *et al.*, 2000), and the mechanical effect of fluids (e.g. Miller *et al.*, 1999; Miller, 2002). In the minimalist model, as well as in other very simple cellular automata (e.g. Newman & Turcotte, 2002; González *et al.*, 2005; Tejedor *et al.*, 2008, 2015, Ch. 5), these details are ignored: the model is driven stochastically, there are only two possible states for each cell, and the earthquakes are generated according to simplified breaking rules.

Let us now explain the simplified view of earthquake generation that the model tries to sketch. In actual faults, the regional stress strains the rock blocks of the fault, making portions (*patches*) of the fault plane to become metastable. That is, they are static, but store enough elastic energy to propagate an earthquake rupture once triggered. Different processes (for example, fault creep –aseismic slip– and plastic deformation) dissipate stress along the fault plane, so stress is not directly converted into elastic strain (e.g. Sec. 5.2). Earthquakes rupture some of the metastable patches of the fault, that then become stable, thus relieving strain. The hypocentre of an earthquake is usually located in a particularly strong patch of the fault plane, called *asperity* (Kanamori & Stewart, 1978; Aki, 1984; Das, 2003; Lei *et al.*, 2003). Asperities

appear to be persistent features where earthquake ruptures start repetitively (e.g. Aki, 1984; Okada *et al.*, 2003, as also evidenced by the results of Ch. 2). Once the rupture starts, it propagates along the fault plane until it arrives at a patch of the fault that is not sufficiently strained. Then the rupture cannot propagate further, and is arrested. The relatively stable patch that is not sufficiently strained and that arrests the rupture is called a *barrier* (Das and Aki, 1977; Aki, 1984; Das, 2003).

The model, depicted in Fig. 6.2, sketches these features as follows. It represents the fault plane as an array of N cells. This is the only parameter of the model. Each cell is denoted by an index i . In previous papers (Vázquez-Prada *et al.*, 2002, 2003; López-Ruiz *et al.*, 2004; Gómez & Pacheco, 2004; González *et al.*, 2004), this array was drawn vertically, in order to simplify its mathematical description. Here (as in González *et al.*, 2006a) the model will be drawn horizontally, in order to sketch the fault plane in a way more similar to that of actual faults (which are usually longer along the strike than along the dip, Fig. 6.1). Typically, cellular automaton models discretize the fault plane in a similar way (e.g. Rundle *et al.*, 2004; Field, 2015).

The cells can only be in one of two states: “empty” (stable) or “occupied” (metastable). The state of the model at each time step can be described simply by stating which cells are occupied and which are not. The increase of regional stress, as in other simple models (Bak and Tang, 1989; Newman & Turcotte, 2002; Castellaro and Mulargia, 2001), and particularly in the box model (Sec. 5.5 and González *et al.*, 2005), is represented by the random addition of balls, in an analogy with classical experiments in probability theory involving balls in an urn (e.g. Johnson & Kotz, 1977). This randomness is a way of dealing with the complex stress increase in actual faults. At each time step, one cell is selected randomly, and a new ball arrives on it. That is, each cell has the same probability, $1/N$ of receiving the new ball. If the chosen cell is empty, the ball “occupies” it. This means that the regional stress has produced enough strain on that cell to make it metastable. If the cell is already occupied, that stress ball is lost; this is analogous to stress dissipation on the fault plane. The total number of occupied cells represents the total elastic strain on the fault.

In the model, we assume that there is only one, persistent, asperity: the first cell, $i = 1$, placed at one end of the array. This option is chosen because it simplifies the analytical description of the model (e.g. Vázquez-Prada *et al.*, 2002; Gómez & Pacheco, 2004). This case can be easily extended, because, if the asperity were located at a different site, the resulting model would be like two joined minimalist models which share the same asperity. Extensions of the model to two dimensions and to other spatial distributions of asperities have been explored elsewhere (Tejedor *et al.*, 2009b).

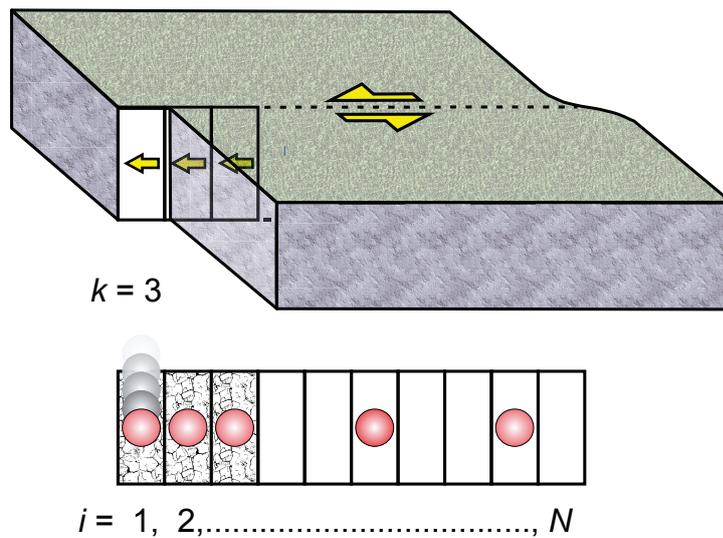


Figure 6.2: **The minimalist model as a sketch of a seismic fault.** The fault plane is divided into an array of N equal cells (the only parameter in the model), denoted with an index i , $1 \leq i \leq N$. Initially all cells are stable (empty). The increase of regional stress proceeds by randomly choosing a cell at random at every time step, which will become metastable (if already not so). As for the box model (Sec. 5.5), this is visualized by the random addition of balls to the cells, one per time step, in an analogy to classical experiments in probability theory (e.g. Johnson & Kotz, 1977). Earthquake ruptures start at an asperity, the cell $i = 1$, when randomly selected (so a ball arrives to it). The rupture propagates through all the k consecutive metastable (occupied) cells. The rupture area (a measure of earthquake size) is k (which equals 3 in the example shown).

When a ball fills cell $i = 1$, a rupture starts there, and propagates through all the consecutive metastable cells until it is arrested by a stable cell. That is, if all the successive cells $i = 1$ to $i = k$ are occupied, and cell $k + 1$ is empty, then the effect of the earthquake is to empty all the cells from $i = 1$ to $i = k$. The other cells, $i > k$ remain unaltered. The cell $k + 1$ is a barrier: it is empty (stable), so the rupture cannot propagate through it. The size (rupture area) of the earthquake is k , the number of cells broken in the synthetic earthquake. Thus, the earthquake size in the model is discrete, $1 \leq k \leq N$. Earthquakes, in practice, are instantaneous in the model (they do not last for any time step). This represents the fact that earthquake ruptures are, indeed, much faster than the slow stress loading represented by the addition of balls.

The random addition of balls is what makes the model stochastic. It also determines the rate at which earthquakes occur in the model. At each time step, independently of the previous earthquake history, there is a probability $1/N$ for the incoming ball to arrive at cell $i = 1$ and start an earthquake. Thus an earthquake, on average, occurs randomly once every N steps, that is, in a Poissonian process with rate $1/N$. In consequence, the time between any two consecutive earthquakes follows a geometric probability distribution (see Sec. 5.5.2.1 for definition).

The cellular-automaton approach of this model is similar to that of the “forest fire” models, in which clusters of interconnected occupied cells (“trees”) “burn” and are reset to empty when they are randomly struck by “lightning” (Drossel and Schwabl, 1992; Henley, 1993). The utility of this kind of models for earthquake physics has been noted by Rundle *et al.* (2003), and extensions of the minimalist model have followed such a scheme (Tejedor *et al.*, 2009b). In the minimalist model there is no random “lightning”: the clusters of interconnected metastable sites are only emptied if they are connected to the cell $i = 1$ and if this fails.

6.2.2 Main properties

The minimalist model, because of its extreme simplicity, lacks the detailed description of the seismic process that a fully dynamical model can display. For example, it does not include the effects of fault friction, elastic stress transfer, or the role of fluids that more complex models can take into account. However, it spontaneously displays several properties that are comparable to those of actual faults, outlined as follows:

1. *Earthquake size-frequency distribution.* It is of the characteristic-earthquake type (Wesnousky *et al.*, 1983; Schwartz & Coppersmith, 1984; Youngs & Coppersmith, 1985; Wesnousky, 1994), frequently observed in seismic faults

(e.g. Wyss, 2015, , and references therein), especially those with simple traces (Stirling *et al.*, 1996) and in other numerical models if the fault plane is homogeneous (e.g. Rundle and Klein, 1993; Main, 1996; Dahmen *et al.*, 1998; Steacy and McCloskey, 1999; Moreno *et al.*, 1999; Hainzl and Zöller, 2001; Heimpel, 2003; Zöller *et al.*, 2005; Ben-Zion, 2008). In this distribution (Wesnousky, 1994), the magnitude-frequency distribution follows the Gutenberg-Richter distribution (Ishimoto & Iida, 1939; Gutenberg & Richter, 1944, 1954, Eq. 3.10) for small and moderate earthquakes, but there is a relative excess of the largest events (called *characteristic earthquakes*) which involve the slip of the whole fault or most of it (Fig. 6.3).

Note that earthquake size (k) is a proxy to the area of slip (A). In natural earthquakes, magnitude scales approximately as $M \propto \log A$ (e.g. Shaw, 2009), so a Gutenberg-Richter distribution (Eq. 3.10) will be shown as a straight line (power law) decay in this log-log plot. Detailed insight on this decay for the minimalist model was presented by López-Ruiz *et al.* (2004). The Gutenberg-Richter distribution observed in regional seismicity, which includes contributions from many faults (e.g. Ch. 3) can be reproduced adding up the seismicity of an ensemble of minimalist models whose sizes (N) are distributed as in actual faults (López-Ruiz *et al.*, 2004). It can also result from changing the model by either adding more asperities where earthquakes can originate, not just one (increasing the heterogeneity of the fault plane, Tejedor *et al.*, 2009b), or from removing all barriers, so that each earthquake empties all occupied cells (Tejedor *et al.*, 2008).

2. *Duration of the earthquake cycle.* The earthquake cycle of a fault is the time interval between two consecutive large earthquakes, which involve rupture and/or slip of the whole fault or fault section (e.g. Scholz & Gupta, 2000, and Ch. 5). A seismic cycle of the minimalist model lasts the time elapsed between two consecutive earthquakes of size N . Its duration is a stochastic variable, with a discrete probability distribution function, denoted by $P_N(n)$ (Fig. 6.4).

This distribution has been proven useful as a renewal model to fit series of large earthquakes in individual faults (Gómez & Pacheco, 2004; González *et al.*, 2005, and Sec. 5.7). Note that, despite the occurrence of earthquakes of any size in the model is Poissonian, the occurrence of earthquakes of a specific size (such as N) is not, because the maximum possible size of an event depends on the size of the previous event and the time elapsed since it occurred.

3. *Stress shadow.* When a fault generates a large earthquake, the elastic strain is reduced, and typically a minimum time has to elapse until the fault, by slow

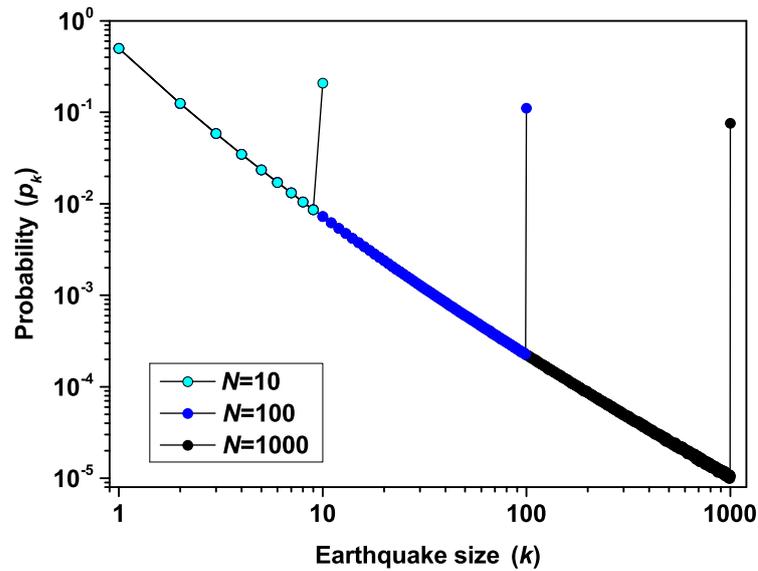


Figure 6.3: **Earthquake size-frequency distribution in the minimalist model, for different numbers of cells (N).** The distributions overlap for any N , except for the earthquakes of maximum size (characteristic earthquakes, $k = N$, which are the target ones to forecast). The probability p_k is the relative frequency of earthquakes if size k (for example, half of the earthquakes are of size $k = 1$, so $p_1 = 0.5$). For earthquakes with $k < N$ the distribution is similar to a power law (Gutenberg-Richter distribution).

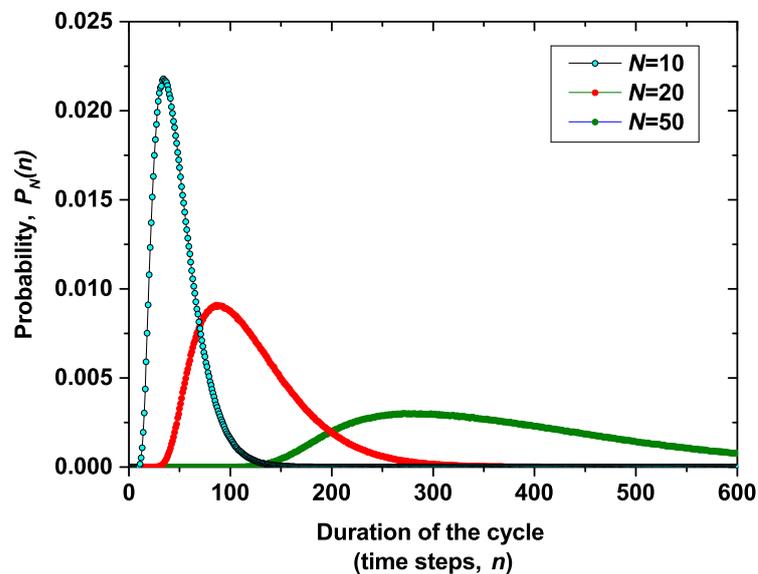


Figure 6.4: **Probability distribution function ($P_N(n)$) for the duration (in time steps, n) of the seismic cycle in the minimalist model.** The only parameter is the number of cells, N , which controls both the scale and the shape of the distribution. This distribution has been applied to model the recurrence intervals of actual earthquake series (Gómez & Pacheco, 2004; González *et al.*, 2005, and Sec. 5.7).

tectonic deformation, accumulates enough strain to generate another large earthquake. This effect is called stress shadow (e.g. Harris, 2000; Scholz, 2002, and Secs. 5.5.2.6 & 5.7.2). In the minimalist model there is a stress shadow: if an event of size k takes place, at least k time steps have to elapse until another event of that size can occur.

4. *Pattern of strain loading.* In actual faults and in laboratory experiments, the strain increases rapidly just after a large earthquake, and then more slowly, because of fast post-seismic reloading or by progressively increasing stress dissipation (e.g. Anthony & Marone, 2005; Michael, 2005; Bourgeois, 2006, Sec. 5.2). In the model, the total elastic strain (or conversely the total shear stress or the potential elastic energy) is represented by the *occupancy* (the total number of occupied cells, Fig. 6.5), which has a similar pattern. Just after a large earthquake, there are fewer occupied cells, so it is more probable for the incoming balls to land on empty cells, and the occupancy grows faster than later on (Fig. 6.5).

In this model, the system resides in a high level of occupancy during most of the time (Vázquez-Prada *et al.*, 2002). In each cycle, there are two consecutive stages. Starting from the empty state, the occupancy increases, but not in a monotonic way, because of the dissipation of balls which land in already occupied cells, and also by the occupancy drops caused by earthquakes with $k < N$. Eventually, all the $N - 1$ cells except the asperity are occupied, and the occupancy resides in this maximum level forming a plateau (Fig. 6.5) until a ball is randomly assigned to the first cell. Then, an earthquake of maximum size takes place, all the load in the system is lost, and a new cycle begins.

Because in the model there are earthquakes of different sizes, this loading pattern is richer and more complicated than in the box model (where the plateau in occupancy at the end of the cycle is analogous, but earthquakes had only one possible size, Sec. 5.5) or in simple laboratory experiments (Sec. 5.2). While in the box model the occupancy completely described the system at any time step, describing the configuration of the minimalist model requires specifying which cells are occupied (e.g. Vázquez-Prada *et al.*, 2002).

5. *Seismic quiescence.* The model displays seismic quiescence (absence of earthquakes) before the characteristic events. Once $N - 1$ cells become occupied (Fig. 6.5), only an earthquake of size N (and not smaller) can occur, and from that moment onwards this requires, on average, N time steps. Seismic quiescence has been observed preceding many large earthquakes (e.g. Wyss and Habermann, 1988; Scholz, 2002), although in other cases the

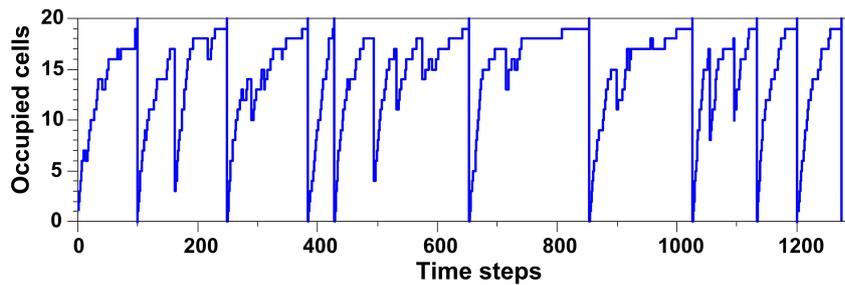


Figure 6.5: **Seismic cycles in the minimalist model.** The number of occupied cells along ten cycles is represented for a model with $N = 20$ cells. This occupancy is analogous to the total elastic strain accumulated in the fault. Sudden drops correspond to earthquakes. Each seismic cycle ends with an earthquake of size N . Compare with the simpler patterns obtained by the box model (where earthquakes had only one size, Fig. 5.4) and in simple laboratory experiments (Fig. 5.1).

opposite effect (increased activity) has been observed (e.g. Bowman *et al.*, 1998; Reasenber, 1999; Tiampo *et al.*, 2002). The minimalist model does not show this last behaviour. The statistical significance of these precursors in real seismicity has yet to be proven (e.g. Hardebeck *et al.*, 2008; Rundle *et al.*, 2011).

6.3 General scheme of binary temporal forecasting

In this section we will briefly recall the general framework for the binary temporal forecasting of earthquakes, already outlined in Sec. 5.6, and the strategies used as baselines specifically for the minimalist model. In the latter, earthquakes can have various sizes, so it is necessary to define precisely which target earthquakes we wish to forecast. In models (e.g. Pepke & Carlson, 1994; Hainzl *et al.*, 2000), as in real seismicity, they are usually the largest ones. As in previous attempts (Vázquez-Prada *et al.*, 2003), in the minimalist model it is natural to choose as target events the characteristic earthquakes (size $k = N$), as they mark a distinct peak in the size-frequency diagram, being much more frequent than other large earthquakes. Along the time series of the model, “alarms” will be turned on in periods when characteristic earthquakes are expected, and turned off when they are not.

As a first remark, we have to consider that the model is stochastic, so it is not predictable with absolute precision. Only simple deterministic systems are fully predictable. The evolution of complex systems, such as the atmosphere or the lithosphere (even if they were deterministic) is very sensitive to the initial conditions. As these complex systems cannot be fully characterized, they turn out not to be fully predictable either.

The fraction of errors, f_e , and the fraction of alarm time, f_a , will be used to quantify the forecast skill (Sec. 5.6). The fraction of false alarms, f_f , will be also computed. It is included in f_a , and defined as the ratio of the total duration of false alarms to the total duration of the time series. Of course, a good forecasting strategy should render small f_a , f_e and f_f . However, as a general rule, a strategy that renders low f_e tends to require large f_a and f_f . As in the previous chapter (Sec. 5.6) We will try to minimize the loss function $L = f_a + f_e$.

We describe here three forecasting strategies, based on the earthquake series, that we will use to asses the merits of the new strategy described later in this chapter. The first two strategies (the random guessing strategy and the so-called reference strategy) can be used in any system, and were already used in the previous chapter with the box model. The third one is specific to the minimalist model, and serves to determine its maximum theoretical predictability.

6.3.1 Random guessing strategy

In this strategy, the alarm is randomly turned on and off, during a certain fraction of alarm time, f_a . To be statistically significant, any forecasting strategy must render better results than a random guess. A natural way to measure this improvement is to consider the loss function $L = f_a + f_e$. In the long term, if $L = 1$ means that the strategy performs as a random guess, while $L = 0$ would be a perfect forecast (see Sec. 5.6.1 for details).

6.3.2 Reference strategy

The reference strategy consists simply in declaring an alarm some time after each target event, and maintaining it on until the next target event, as explained in the previous chapter (Sec. 5.6.2). For the minimalist model, we can look for the number of time steps n to use with this strategy for obtaining the smallest L . In a previous paper (Vázquez-Prada *et al.*, 2003) we observed that indeed, for each N , there is a n which minimizes L . In Fig. 6.6, the minimum L that can be obtained with this strategy is plotted for N between 2 and 20, in the curve labelled "Reference".

This method does not generate any false alarm, and only takes into account the occurrence of the largest earthquakes. The only information required is the statistical distribution (probability distribution function) of the duration of the cycles (Vázquez-Prada *et al.*, 2003, and Sec. 5.6.2). These results (Fig. 6.6) have negligible uncertainty, since they have been calculated by numerical simulations with three million seismic cycles for each N . Taking into account the effects

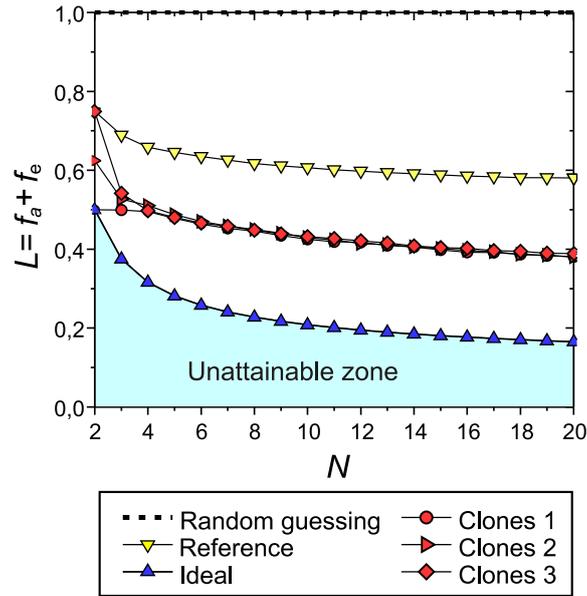


Figure 6.6: **Comparison of the best forecast results obtained for each strategy, for various system sizes (N).** The loss function (L) equals the sum of the fraction of errors (f_e) and the fraction of alarm time (f_a). A lower L indicates a better forecast of the largest earthquakes in the model. A *random guessing* strategy would render $L = 1$ for any N , while $L = 0$ would mean a perfect forecast. The model is stochastic, and thus not completely predictable: the “*ideal*” strategy marks the lowest possible L and is deduced analytically (Appendix A.2). The “*reference*” strategy is based only on the series of the largest earthquakes in the model. The three strategies labelled “*clones*” provide the best practical results, and are based on the synchronization of models with the minimalist model whose largest earthquakes we try to forecast.

of smaller earthquakes, the forecast can be modestly improved in the model (Vázquez-Prada *et al.*, 2003).

6.3.3 Ideal strategy

As the minimalist model is very simple, it is possible to explore its maximum predictability. The ideal strategy needed for getting this result, unlike the two previously described, is model-specific. It is deduced in Appendix (Sec. A.2). This ideal result could only be obtained if we could “see” inside the model to check at each time step which cells are occupied and which are not. Thus it requires a perfect knowledge of the system, and equivalent strategies cannot be used with actual faults where we cannot know the detailed state of stress and strain. Ideally the alarm should be declared at the instant in which $N - 1$ cells of the model are full (just at the beginning of the plateau with seismic quiescence commented on in Section 6.2.2). Then, it should be maintained on until the next characteristic earthquake.

This strategy has no errors and no false alarms ($f_e = 0$, and $L = f_a$). As the model is stochastic, f_a is not zero; a minimum alarm time is needed to forecast all the characteristic earthquakes. It is given by $f_a = L = N/\langle n \rangle$, where $\langle n \rangle$ is the average duration of the cycles (which depends on N). This L is also plotted in Fig. 6.6, in the curve labelled “Ideal”. This is the rigorous minimum L that can be obtained in the model. A good forecasting strategy should produce a L lower than the “Reference” curve and as close as possible to the “Ideal” curve.

6.4 Synchronization-based forecasting

In this section we will describe the novel forecasting method based on the synchronization between models, obtained by imposing the rupture area of a minimalist model onto other, similar models.

Synchronization is the process by which two systems adjust a given property of their evolution. Synchronization processes in complex systems are an expanding topic in Statistical Physics, with applications to a wide variety of natural phenomena (e.g. Arenas *et al.*, 2008). Originally (González *et al.*, 2004), the research of this chapter was partially inspired by ideas on anticipated synchronization between chaotic oscillators (e.g. Voss, 2000, 2001a,b,c; Masoller, 2001; Toral *et al.*, 2001; Ciszak *et al.*, 2003; Calvo *et al.*, 2004; Lindner *et al.*, 2004).

It has been demonstrated, both theoretically and experimentally, that it is possible to synchronize even chaotic systems under certain conditions, by coupling or forcing them (e.g. Boccaletti *et al.*, 2002). For our case it is relevant the so-called *unidirectional coupling*, in which one system (called the *master*) evolves freely and drives the evolution of the other (called a *slave* or *replica*). Synchronization between cellular automata models has been explored since long ago (e.g. Romani, 1976; Morelli & Zanette, 1998; Urías *et al.*, 1998), but, as far as we know, not with the purpose of forecasting the evolution of a system.

We will try to forecast the characteristic earthquakes generated by a minimalist model with N cells. This model will be called *master*. We will consider it as if it were an actual fault, from which we can know the rupture area of its earthquakes (equivalent to the number of cells broken, k), but not the strain or stress at depth (equivalent to the occupation state of the model cells). As in an actual fault, we cannot change the state of the master at any moment.

In this forecasting method we will use other models, which we call *clones* (González *et al.*, 2004, 2006b). These are equivalent to the models devised for forecasting the future evolution of the fault. We will modify their evolution at will, and their governing rules will be different than those of the master.

In this chapter, for simplicity, we will consider that the clones are also arrays of N cells. The average duration of the earthquake cycle in the model (average recurrence interval of the characteristic earthquakes), $\langle n \rangle$, strongly depends on N (Gómez & Pacheco, 2004). Choosing a different N for the clones will imply a different loading rate of the cells and a different average recurrence interval of the characteristic earthquakes in the clones than in the master. These effects would require further tuning of the clones, which would complicate the following discussion.

We will use a total of Q clones, that will be loaded (one ball per time step and per clone) at the same time as the master, but randomly and independently to the master and to each other. We will apply some procedures for partially synchronizing the clones with the master. Namely, if in a given time step the master does not generate any earthquake, we will oblige the clones not to generate any earthquake either. And if the master does generate an earthquake, we will force the clones to reproduce the rupture area of this earthquake, as described below in more detail. Note that, although the master and the clones are driven simultaneously, the effects of the master are dealt with first.

By using several clones, we can take into account a broad range of possible evolutions. The master and the clones are all stochastic, so each one evolves through time in a different way. By using only one clone, we could not be very sure that it is satisfactorily mimicking the evolution of the master. However, if several of these Q clones are in the same state, then it is more probable that the master is also in that state. If the clones were deterministic, only one would be required.

We have commented before (Sec. 6.3.3 and Appendix Sec. A.2) that the ideal forecasting strategy for the minimalist model will be to declare the alarm just when $N - 1$ cells of the model become occupied. Then the master enters the stage of seismic quiescence, or plateau, and the next earthquake is the characteristic one. We will try to determine this ideal instant as well as possible with the clones. For this, we will use a “democratic” procedure: we will declare an alarm when a minimum of q clones “vote” (become occupied to a certain threshold, described below). Later on we will explore the combinations of Q and q that render the best results.

Once the alarm is declared, it is maintained on until the next earthquake in the master. If it is a characteristic one, this is a successful forecast. Its rupture is imposed on the clones (so we reset all the cells of the clones to empty) and a new cycle starts. If the next earthquake is not a characteristic one, this represents a false alarm. We will disconnect the alarm, and impose the rupture on the clones as is done with any other earthquake. Of course, if a characteristic earthquake takes place when the clones have still not declared the alarm (when less than q clones have voted), this is a forecast failure. If the clones declare an alarm in the same time step in which

the master generates a characteristic earthquake, we also consider this as a forecast failure.

The exact rules for driving the clones will follow one of the three approaches commented on below. Each approach implies a different knowledge of how the master works, and a different way of imposing the rupture area on the clones. They are depicted in Fig. 6.7 and described as follows:

1. This first approach will indicate the best result that can be obtained with the synchronization-based forecasting. For this reason, the clones are indeed minimalist models identical to the master (González *et al.*, 2004). The clones are loaded only if the master does not generate an earthquake in that time step. We know that in this case the ball in the master has gone to one of the cells $i \geq 2$, so the balls in the clones will be randomly thrown to the cells $i \geq 2$. We also consider as known that, just after an earthquake with rupture area k , the first $k + 1$ cells in the master, for sure, are stable (the k just broken plus the one that acts as a barrier for the rupture). Thus, if the master generates an earthquake of size k , we will reset to empty the first $k + 1$ cells of the clones. A clone votes when $N - 1$ of its cells are full. An example of the evolution of the loading of a master and a clone obtained by using these rules is shown in Fig. 6.8.
2. In this second approach, we are more ignorant about how the master works. At every time step we will throw the balls to any of the cells in the clones. If the master generates an earthquake of size k , we only know which cells have ruptured, so we will reset to empty only the first k cells of the clones. A clone votes when its N cells are full.
3. In the third approach we know even less. At every step we will throw the stress balls to any of the cells in the clones. When an earthquake takes place in the master, we only know its size, and thus its rupture area, k , but not exactly which cells have ruptured. Thus, we will randomly empty k occupied cells of each clone. If the clone has less than k occupied cells, all are emptied. A clone votes when its N cells are full. In this approach the positions of the cells in the clone are irrelevant. Each clone is thus equivalent to the box model (González *et al.*, 2005, Sec. 5.5).

Ideally each clone should have the same number of occupied cells as the master. The fraction of time, τ , during which this happens is a measure the degree of synchronization between them. If two independent masters run simultaneously, they have the same number of occupied cells, just by chance, during a certain τ .

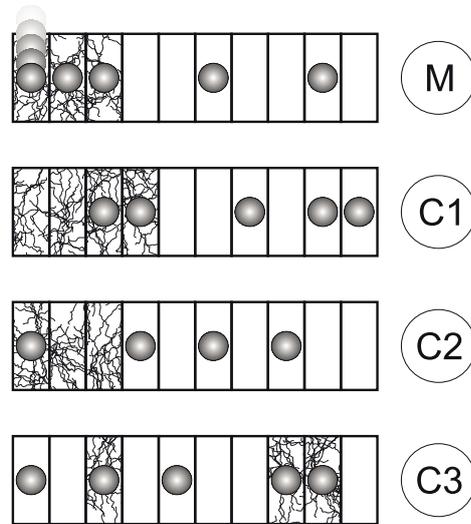


Figure 6.7: **Forcing models to reproduce the earthquake of another model.** Models called *clones* (C) are forced to reproduce the rupture area of each earthquake of a model which evolves freely (the *master*, M), so they become partially synchronized with the latter. In the example, the master generates an earthquake with rupture area $k = 3$. Three approaches are used. In the first one ($C1$), the first $k + 1$ cells rupture and will be reset to empty. In the second one ($C2$), this occurs only with the first k cells. In the third approach ($C3$), this happens to k occupied cells chosen randomly. The first cell of the clones can be occupied only in the second and third approaches.

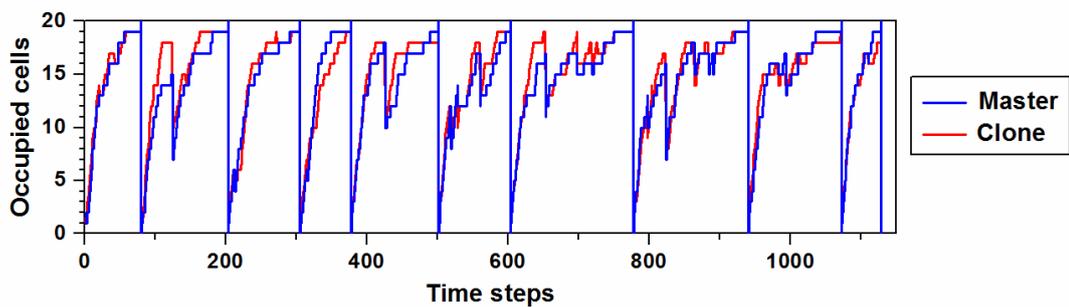


Figure 6.8: **Synchronization between a master model and its clone.** Both models are loaded independently and stochastically, but the loading state (occupied cells) of the clone becomes partially synchronized with that of the master. This is achieved by forcing the clone to reproduce the rupture area of each earthquake generated by the master. The example shows ten seismic cycles in which the first synchronization approach was used ($C1$ in Fig. 6.7).

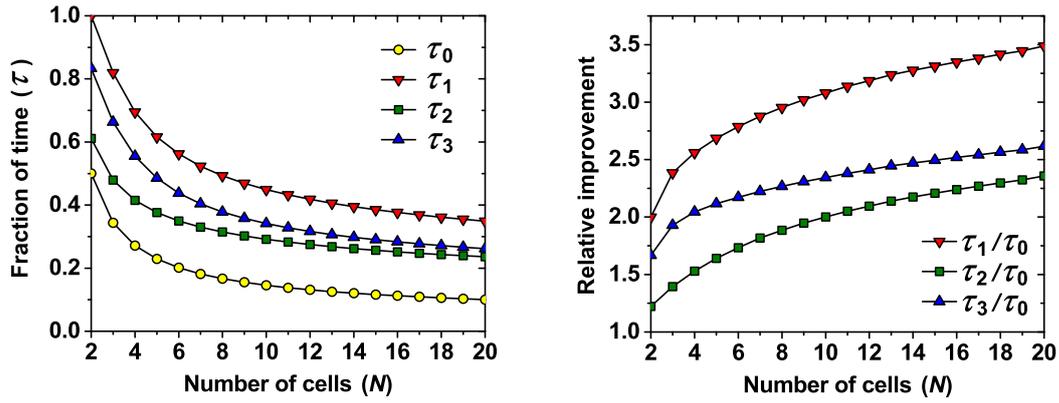


Figure 6.9: **Synchronization of occupancy in the models.** *Left:* The fraction of time, τ , during which two models have the same number of occupied cells. This depends on whether they are two independent masters (τ_0) or a master and a clone (governed by one of the three different synchronization approaches: τ_1 , τ_2 and τ_3). *Right:* The relative improvement, defined as τ_n/τ_0 .

When a clone and a master are compared, this τ greatly increases, as shown in Fig. 6.9: partial synchronization is achieved. The best results, as expected, are obtained with the first of the three approaches.

The results of f_a , f_e , f_f and $L = f_a + f_e$, for different values of Q and q can be plotted as in the diagrams of Fig. 6.10. In this figure we have plotted only results corresponding to the first of the three approaches and $N = 20$, but similar figures, with the same overall properties, can be drawn for the other two approaches and for any N (see below). There are simple trends in these graphs. In general, a forecasting strategy that produces lower f_e tends to require higher f_a and f_f . If Q is fixed (same row), the greater the q , the later the alarm is declared, so f_a and f_f are lesser and f_e is greater. If q is fixed (same column), the greater the Q , the earlier the alarm is declared, resulting in the opposite trend.

We are interested in finding the combinations of Q and q that minimize L . The interesting fact is that the sum $f_a + f_e$ shows a rectilinear “valley” for certain combinations of Q and q , marked with squares in the graph of Fig. 6.10. This valley goes down as Q and q increase. In Fig. 6.11 it can be observed that the valley goes down indefinitely, tending to a lowest asymptotic value of L . We estimate this value, as a function of Q , with a three-parameter exponential fit of the form $F = a \exp[b/(Q + c)]$, where a , b , and c are parameters. This fit describes the observations very well, and in the figure it is indistinguishable from the lines shown connecting the points of L . The value of a is the asymptotic one for $Q \rightarrow \infty$. This value is represented, for each N , in Fig. 6.6. The f_a , f_f and f_e also have asymptotic trends along this valley of L , also plotted in Fig. 6.6. They can also be fitted with

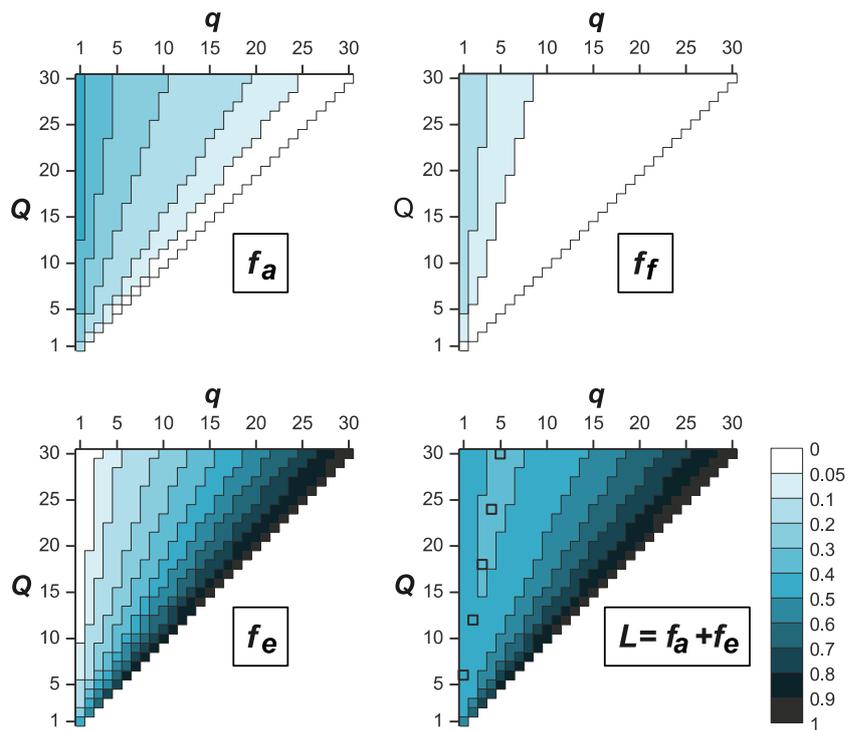


Figure 6.10: Example results of the forecasting with synchronized models. Fraction of alarm time (f_a), fraction of false alarm time (f_f), fraction of errors (f_e), and loss function (L) obtained with the the first synchronization-based forecasting approach, for $N = 20$ and different numbers of clones (Q) and votes (q). The squared cells mark a rectilinear valley in the values of L .

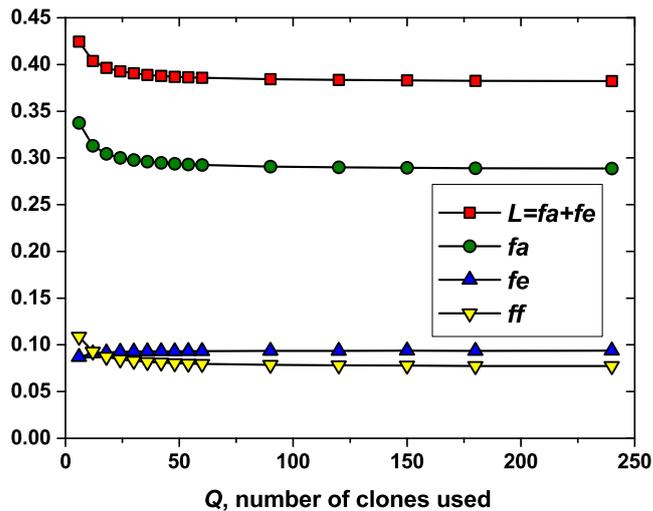


Figure 6.11: **Asymptotic results for large numbers of synchronized models.** Loss function (L), fraction of alarm time (f_a), fraction of false alarm time (f_f) and fraction of errors (f_e) obtained with the first synchronization-based forecasting approach, for $N = 20$ and different numbers of clones (Q) along the rectilinear valley observed in L in Fig. 6.10 (the first five points of each curve correspond to the cells marked in that figure). The asymptotic L is approximately reached already for several tens of clones, and for infinite ones it can be extrapolated from a precise exponential fit (which overlaps with the lines shown connecting its points). The asymptotic values of L are those shown in Fig. 6.6

the same kind of three-parameter distribution, to estimate their asymptotic values as $Q \rightarrow \infty$. A nice property is that, as shown in Fig. 6.6 for a certain case, these approaches forecast most of the characteristic earthquakes (f_e is low), and have a very small fraction of false alarms. Note also that a few tens of clones already render results close to the asymptotic ones.

As can be noted in Fig. 6.6, the synchronization-based strategies perform much better than a random guess, and also much better than the reference strategy. Their results are intermediate between the ideal forecast and the reference one. The second and third synchronization-based approaches give only slightly greater L than the first one. The differences are large only for small N . Although the first approach synchronizes more efficiently each individual clone with the master (Fig. 6.9), this effect is compensated by using many clones.

To assess the performance of the method with larger systems, we plot in Fig. 6.12 the results of L for the three approaches, for $N = 100$ and up to 60 clones. As occurred for smaller N , a rectilinear valley is observed in the graphs, and this tendency can be extrapolated to estimate the asymptotic value of L . Note that the results for the first and second approaches are almost identical (although L is slightly larger in the second approach). With the third approach, L decays to its asymptotic

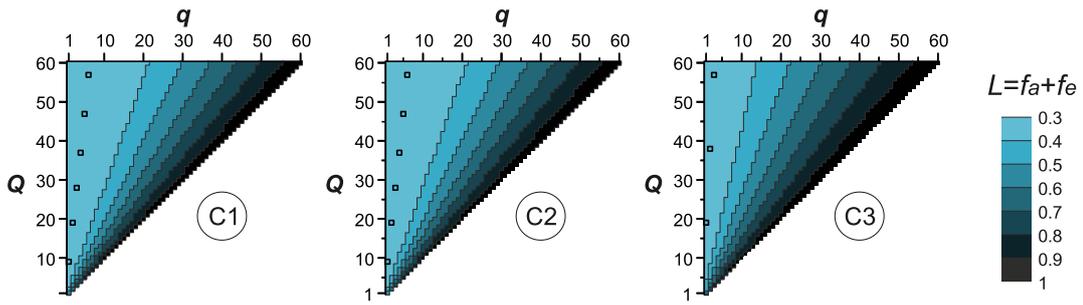


Figure 6.12: **Example forecasting results with a larger system.** Loss function (L) obtained with the three synchronization-based forecasting approaches, for $N = 100$ and different numbers of clones (Q) and votes (q). The squared cells mark rectilinear valleys in the values of L .

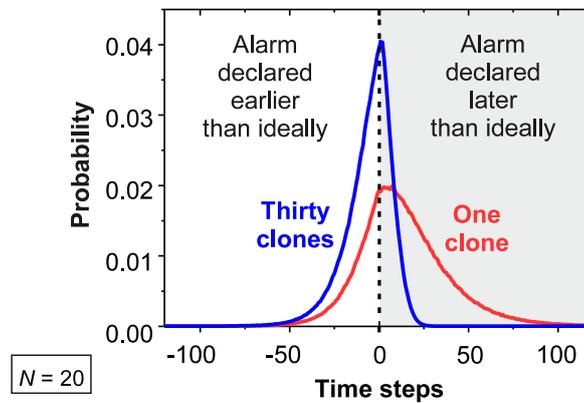


Figure 6.13: **Synchronization with the ideal moment to turn the alarm on.** Probability that a clone or thirty clones (with $q = 5$; the uppermost squared cell in L in Fig. 6.10) declare an alarm in a given time, around the instant when it should for obtaining the ideal forecast.

value more slowly (the valley floor has a smaller slope, so more clones are needed to achieve a given low L). The asymptotic values of L , however, are very similar in the three cases (0.298 for the first and second approaches; 0.306 for the third one). Note that these values are smaller than for $N = 20$, as expected from the trend observed in Fig. 6.6. The f_a , f_e and f_f show trends similar to the ones described for Figs. 6 and 7. The asymptotic f_e is very low (0.062, 0.075 and 0.071 for the first, second, and third approach, respectively).

Another way to measure the synchronization of the clones with the master is drawn in Fig. 6.13 for $N = 20$. The ideal strategy (Section 6.3.3 and Appendix A.2), would be to declare the alarm just when $N - 1$ cells of the master are full. The figure shows how a single clone declares the alarm around that moment, but a group of clones marks this moment with enhanced probability.

6.5 Discussion and conclusions

In this chapter we have tried to provide some insight into how to synchronize numerical models with seismic faults, in order to better forecast large earthquakes in them. The idea is that, although we can rarely measure the stress and strain in actual faults, we can estimate the rupture area and coseismic displacement of their earthquakes. If we force a calibrated model to reproduce every earthquake rupture of the fault it simulates, probably the model will be synchronized with the fault. Then it could be used to forecast the future evolution of the fault, including future large earthquakes. This idea is not completely new: e.g. Ward (2000) forced a model to reproduce a large-earthquake rupture and run the model forward to see what could happen in the future. The results of this chapter expand on earlier ones (González *et al.*, 2004), and are still only theoretical, but fully quantitative. We demonstrate that it is possible to partially synchronize numerical fault models between themselves, and use this to forecast synthetic earthquakes.

One of the models, called the master, evolves freely. We consider it as an actual fault, from which we can know the rupture area of its earthquakes, but not the strain or stress at depth. Our goal is to forecast the largest earthquakes it generates. In the synchronization-based forecasting, we use several other models, called clones, similar to the master (calibrated to have the same average recurrence interval of large earthquakes that the master has). These clones are equivalent to the models that can be devised to simulate a seismic fault. They are run simultaneously and independently to the master and to each other. We force them to reproduce the series of earthquake ruptures of the master, and this makes them partially synchronized with it. In simple words, if the master does not generate an earthquake, we preclude any earthquake in the clones; if the master does generate an earthquake, we impose the same rupture area on the clones. When several of the clones indicate that a large earthquake is impending in the master, we declare an alarm. This efficiently forecasts most of the largest earthquakes of the master, with a relatively low fraction of total alarm time and few false alarms. These results are robust: they are almost the same when the exact rules for imposing the earthquake ruptures vary, and this good performance is observed along the whole range of model sizes considered. This synchronization-based forecasting outperforms other procedures based only on the earthquake series of the model (Vázquez-Prada *et al.*, 2003; González *et al.*, 2004).

The master and the clones are stochastic (random), so each individual clone is only partially synchronized with the master. However, when several clones are in the same state, then it is more likely that the master is also in this state, so the group

of clones achieves much better forecasting skill than only one clone. If the clones were deterministic, as a general rule only one would be needed; more clones would have identical evolutions if run with the same initial conditions. In our approach, each clone marks a possible state of the master among a range of possible options. Several deterministic clones could also be used with different initial conditions.

The procedure developed here is a kind of *ensemble* forecasting, in which several models are run to obtain a better picture of how a system will evolve. This concept is used in atmospheric forecasting (Palmer *et al.*, 2005): several models are run simultaneously, and their average or weighted result has a larger forecasting ability than that of each constituent model individually (Houghton, 2002; Palmer *et al.*, 2005). Each model in this approach typically has slightly different initial conditions, to take into account measurement errors and then to represent one possible state of the atmosphere, among various possibilities.

Our procedure also shares some similarities with certain earthquake forecasting algorithms (Kossobokov *et al.*, 1999; Keilis-Borok & Soloviev, 2003), in which several seismicity functions are evaluated in real time. When several of these functions indicate that a large earthquake is probable, an alarm is declared. In our approach, the clones are performing a role similar to these functions, monitoring what is happening in the master.

Our proposal is that a possible way to synchronize more complex, calibrated models with real faults might be to force them to reproduce the past series of earthquakes (with the same rupture area and/or coseismic displacement). This would need to be tested in the future. Also it will be possible to test whether this procedure works in the forecasting of synthetic earthquakes in other models.

As far as we know, this is the first time that a forecasting method based on the synchronization of purely stochastic systems has been developed. It is also the first time that ensemble forecasting has been applied to numerical models of seismic faults. More recently, ensemble earthquake forecasts have been proposed for models of regional seismicity (e.g. Marzocchi *et al.*, 2012).

Forcing the models to reproduce only one large observed rupture (as in Ward, 2000) probably is not enough (this is certainly the case in our stochastic model). Complex and chaotic systems, such as the lithosphere, are very sensitive to initial conditions. Forcing the model to reproduce only one rupture probably does not constrain the initial conditions sufficiently. We propose that every observed rupture, albeit small, should be considered. Small earthquakes are much more frequent than large ones, thus providing much more data (e.g. Ch. 2). Moreover, they provide insight into the mechanical state of the crust (e.g. Seeber and Armbruster, 2000) and into the mechanics of earthquake rupture (e.g. Rubin, 2002). Their location may

indicate the patch of the fault plane which is experiencing higher stresses and is likely to rupture in the next large shock (e.g. Schorlemmer & Wiemer, 2005). Finally, they are important in the transfer of stress within the lithosphere, and in earthquake triggering (e.g. Helmstetter *et al.*, 2005).

Conclusions

We are a way for the Cosmos to know itself.

Carl Sagan (1980)

The next paragraphs summarize the main conclusions of this doctoral dissertation, following the order of the chapters in which they were presented.

Distances between past earthquakes forecast the distance to the next earthquake

- A new method without any parameter is proposed for forecasting the location of earthquakes in space, based on the locations of past earthquakes and their empirical distribution of nearest-neighbour distances.
- The method calculates the probability that the next earthquake will occur within any given distance from previous ones.
- The probabilistic maps obtained with this technique are tested and updated with every new earthquake, so that they refine automatically (self-sharpen).
- This new method is applied to forecast retrospectively the location of epicentres in three earthquake catalogues: for the whole Earth, southern California and the Iberian region. In all them, the expected fraction of earthquakes is observed to take place within each distance threshold.
- A minimum number of past earthquakes (typically ~ 1000) is required to obtain the expected results. The method could be used to assess whether the earthquakes recorded in a region are numerous enough to be representative of the longer-term spatial distribution of seismicity.
- Daily, during six years, the Southern California Earthquake Center, within the Collaboratory for the Study of Earthquake Predictability, has calculated

forecast maps with an adapted version of this method, and tested them, for the whole Earth, California and the western Pacific. This became the first contribution from Spain to this international collaboration. The results obtained are similar to the retrospective ones, and better than those of the method typically used as a reference.

- The new method could thus be used as an improved, non-parametric reference baseline to compare with the performance of more complex models.
- Kernels used for calculating spatial earthquake probabilities should mimic as closely as possible the observed distribution of nearest-neighbour distances between past earthquakes.

The National Spanish Earthquake Catalogue

- The development of the National Spanish Earthquake Catalogue (compiled by the Instituto Geográfico Nacional) is reviewed in detail.
- Catalogue heterogeneities are found to arise because of the changing seismic network, the introduction of new earthquake location procedures, and the modification of formulae used to calculate magnitudes.
- The hypocentral location precision is typically much better in the Iberian Peninsula than elsewhere, especially for the hypocentral depth. Earthquakes in distant offshore areas are the least precisely located. The precision improved with time, especially with the introduction of new procedures for automatic detection and location of earthquakes in November 1997.
- The magnitude of completeness of the catalogue during the instrumental era is analysed in detail for the first time. It is shown to have improved (lowered) along the decades, especially thanks to the new automatic detection procedures since November 1997 and to the deployment of the broadband network since the early 2000's.
- The magnitude of completeness is spatially heterogeneous, as highlighted by successive maps for different periods. It is lower in the most densely monitored areas, and increases further away and in regions where intermediate-depth earthquakes are more frequent. It is also clearly affected by the different magnitude scales used for closer and more distant earthquakes.
- A weekend effect (more earthquakes are recorded during weekends) is found in the catalogue for the first time, as well as the already known daily variations

(more earthquakes are recorded during nighttime). Both effects are interpreted as due to the less complete earthquake detection during weekdays and daytime because of the higher artificial noise.

- Clear cases of contamination by quarry and mine blasts are pinpointed, so despite the efforts of network operators, not all artificial blasts are screened out of the catalogue. In examples from large scale underground mines, blasts are set off even during nights and weekends, contrary to the usual assumption that blasting is only performed during daytime in weekdays.

Measurement of areas on a sphere using Fibonacci and latitude–longitude lattices

- The Fibonacci lattice is proposed as a homogeneous sampling scheme for measuring areas numerically on the sphere. It is applied to measure the areas of the Earth where earthquakes are expected with a given probability.
- This lattice provides a more efficient sampling (and thus less computation time) than the latitude-longitude lattice typically used for numerical area measurements.
- For approximately the same number of lattice points, the maximum root mean squared error of the area measurement is $\sim 40\%$ smaller with the Fibonacci lattice than with the latitude-longitude counterpart. This is partially explained using analytical arguments from similar mathematical problems on the plane.
- The maximum errors of the area measurement with the Fibonacci lattice are also smaller, and decay faster with the number of points used. If about a million points are used, as is commonly the case, the maximum error is one order of magnitude smaller than with the latitude-longitude lattice.

Probabilistic temporal forecasting of large earthquakes in the Parkfield section, San Andreas Fault

- The box model, a new stochastic system with only one parameter, is proposed as an idealized representation of the seismic cycle of a fault. The pattern of loading in the model is similar to the pattern of stress (or elastic strain) increase in laboratory friction experiments, and to the expected one in actual faults. Sudden unloading in the model is analogous to an earthquake, which reduces the stress and elastic strain.
- Despite the model is random, the resulting series of synthetic earthquakes is quasi-periodic (more regular in time than in a Poisson process).

- The quasi-periodic series of mainshocks with magnitude ~ 6 generated by the Parkfield section of the San Andreas Fault is not well fitted by a time-independent Poisson process. It is more adequately described as renewal process, in which the probability of the next event changes with time, depending on the period elapsed since the last one.
- Just after the occurrence of the last Parkfield mainshock (September 28th, 2004), fits of the series using renewal models were submitted for publication. These included classical probability distributions (gamma, log-normal, Weibull), as well as others based on simple descriptions of the earthquake generation (Brownian passage time, minimalist model and box model). The available data is well fitted by all these models.
- The renewal models were used to calculate the annual probability for the next mainshock at Parkfield along the next decades. All time-dependent models forecast a low or even null probability in the first years since the last mainshock, consistent with the fact that the next one has not yet occurred at the time of this writing.

Insights on a way to synchronize models with seismic faults for earthquake forecasting

- Except for a Poisson process, in which events are completely uncorrelated from each other, a stochastic system is predictable to some extent, albeit randomness imposes a limit to its predictability.
- The largest earthquakes in a stochastic, minimalist model of a seismic fault are partially predictable, and their predictability limit is found analytically.
- It is demonstrated that stochastic systems (here called *clones*) can be partially synchronized with another (the *master*), so that the former can to some extent forecast the evolution of the latter. This is the first time that a forecasting method based on the synchronization of stochastic systems is introduced.
- This synchronization is possible even without knowing the internal state of the master, but only imposing its observable dynamics to the clones.
- In particular, clones of a minimalist model become partially synchronized with it when forced to reproduce its earthquake ruptures.
- For this minimalist model, forecasting the largest earthquakes with the aid of synchronized clones clearly outperforms attempts based only on the time series of the largest earthquakes.

- The forecast results are robust even if different synchronization approaches, with decreasing knowledge about the master, are used.
- It is likely that models of seismic faults could be synchronized with actual faults if they are forced to systematically reproduce the actual earthquake ruptures (or slip distributions) that these generate. For this purpose, it would be adequate to force them to reproduce the ruptures of all earthquakes, not just of the largest ones as attempted so far.

Conclusiones

La ciencia [...], al descorrer el velo que ocultaba la explicación de un misterioso fenómeno, descubre nuevos enigmas que brindan más amplios horizontes a la investigación, cuyos dominios se ensanchan continuamente por lo mismo que son infinitos e inagotables.

José García Siñeriz (1935, p. 23).

Los siguientes párrafos resumen las principales conclusiones de esta tesis doctoral, siguiendo el orden de los capítulos en los que se han presentado.

Las distancias entre terremotos pasados pronostican la distancia al terremoto siguiente.

- Un método nuevo, que carece de parámetro alguno, se propone para pronosticar la localización de terremotos en el espacio, basándose en las localizaciones de terremotos previos y su distribución empírica de distancias a los vecinos más cercanos.
- Este método calcula la probabilidad de que el siguiente terremoto ocurra dentro de cualquier distancia dada a los previos.
- Los mapas probabilistas obtenidos con esta técnica se ponen a prueba y se actualizan con cada terremoto, de manera que se refinan automáticamente, adquiriendo mayor nitidez.
- Este nuevo método se aplica a pronosticar retrospectivamente la localización de epicentros en tres catálogos de terremotos: para toda la Tierra, para el sur de California, y para la región Ibérica. En todos ellos se observa que la fracción esperada de terremotos tiene lugar dentro de cada umbral de distancias.
- Un número mínimo de terremotos pasados (típicamente ~ 1000) se requiere para obtener los resultados esperados. Este método se podría usar para

evaluar si los terremotos registrados en una región son lo suficientemente numerosos como para ser representativos de la distribución espacial de la sismicidad a más largo plazo.

- Diariamente, durante seis años, el *Southern California Earthquake Center*, dentro del *Collaboratory for the Study of Earthquake Predictability*, ha calculado mapas de pronóstico basados en una versión adaptada de este método, y los ha puesto a prueba, para toda la Tierra, California y el Pacífico occidental. Esta ha sido la primera contribución española a esta colaboración internacional. Los resultados obtenidos son similares a los retrospectivos, y mejores que los del método típicamente usado como referencia.
- El nuevo método podría por tanto usarse como una base de referencia, mejorada y no paramétrica, con la que comparar el funcionamiento de modelos más complejos.
- Las funciones kernel usadas para calcular las probabilidades espaciales de terremotos deberían imitar tanto como sea posible la distribución observada de las distancias a vecinos más cercanos medidas entre terremotos.

El catálogo nacional español de terremotos.

- El desarrollo del catálogo nacional español de terremotos (recopilado por el Instituto Geográfico Nacional) se revisa en detalle.
- Se encuentran heterogeneidades en el catálogo, debidas a la cambiante red sísmica, a la introducción de nuevos procedimientos de localización y a la modificación de las fórmulas usadas para calcular magnitudes.
- La precisión de las localizaciones hipocentrales es típicamente mucho mejor en la Península Ibérica que en otros lugares, especialmente para la profundidad hipocentral. Los terremotos en áreas marinas distantes son los localizados con menor precisión. La precisión mejoró con el tiempo, especialmente con la introducción de nuevos procedimientos para la detección y localización automática de terremotos en noviembre de 1997.
- La magnitud de completitud del catálogo durante la era instrumental se analiza en detalle por vez primera. Se muestra que ha mejorado (disminuido) a lo largo de las décadas, especialmente gracias a los nuevos procedimientos de detección desde noviembre de 1997 y a la instalación de la red de banda ancha desde comienzos de los 2000.

- La magnitud de completitud es espacialmente heterogénea, como resaltan los mapas sucesivos para distintos periodos. Es más baja en las áreas más densamente monitorizadas, y se incrementa en lugares más alejados y en regiones donde los terremotos de profundidad intermedia son más frecuentes. También está claramente afectada por las diferentes escalas de magnitud usadas para los terremotos más próximos y más lejanos.
- Un efecto de fin de semana (más terremotos son registrados durante los fines de semana) se encuentra en el catálogo por primera vez, y también las ya conocidas variaciones diarias (más terremotos son registrados durante la noche). Ambos efectos se interpretan como debidos a la menos completa detección de terremotos durante los días laborables y las horas diurnas por el mayor ruido artificial.
- Se localizan casos claros de contaminación por explosiones en minas y canteras, de manera que, pese a los esfuerzos de los operadores de la red sísmica, no todas las explosiones artificiales se consiguen discriminar y retirar del catálogo. En ejemplos de minas subterráneas a gran escala, se detonan voladuras incluso durante las noches y fines de semana, contrariamente al supuesto habitual de que éstas se realizan sólo durante las horas de luz de los días laborables.

Medida de áreas en una esfera usando retículos de Fibonacci y latitud-longitud

- El retículo de Fibonacci se propone como un esquema homogéneo de muestreo para medir áreas numéricamente en la esfera. Esto se aplica a medir las áreas en la Tierra donde los terremotos se esperan con una probabilidad dada.
- Este retículo proporciona un muestreo más eficiente (y por tanto menor tiempo de computación) que el retículo latitud-longitud típicamente usado para medidas numéricas del área.
- Para aproximadamente el mismo número de puntos de muestreo, la máxima raíz del error cuadrático medio es $\sim 40\%$ menor con el retículo de Fibonacci que con el latitud-longitud. Esto se explica parcialmente usando argumentos analíticos a partir de problemas matemáticos similares en un plano.
- Los errores máximos de medida de áreas con el retículo de Fibonacci son también menores, y disminuyen más deprisa al aumentar el número de puntos usados. Si se usan alrededor de un millón de puntos, como es el caso

habitual, el error máximo es un orden de magnitud inferior que con el retículo latitud-longitud.

Pronóstico temporal probabilista de grandes terremotos en la sección de Parkfield, Falla de San Andrés.

- El modelo de caja, un nuevo sistema estocástico con un solo parámetro, se propone como una representación idealizada del ciclo sísmico de una falla. El patrón de carga en el modelo es similar al patrón de incremento de esfuerzo (o deformación elástica) en experimentos de fricción en el laboratorio, y al esperado en fallas reales. La repentina descarga del modelo es análoga a un terremoto, que reduce el esfuerzo y la deformación elástica.
- Pese a que el modelo es aleatorio, la serie resultante de terremotos sintéticos es cuasi-periódica (más regular en el tiempo que un proceso de Poisson).
- La serie cuasi-periódica de terremotos principales con magnitud ~ 6 generados por la sección de Parkfield de la Falla de San Andrés no se ajusta bien con un modelo de Poisson, independiente del tiempo. Se describe más adecuadamente como un proceso de renovación, en el que la probabilidad del siguiente evento cambia con el tiempo, dependiendo del periodo transcurrido desde el anterior.
- Justo tras la ocurrencia del último terremoto principal de Parkfield (septiembre de 2004), se remitieron para su publicación ajustes a la serie usando modelos de renovación. Estos incluían distribuciones clásicas de probabilidad (gamma, log-normal y Weibull), así como otras basadas en descripciones simples de la generación de terremotos (de tiempo de primer paso Browniano, modelo minimalista y modelo de caja). Los datos disponibles se ajustan bien con todos estos modelos.
- Los modelos de renovación se usaron para calcular la probabilidad anual del siguiente terremoto principal en Parkfield a lo largo de las siguientes décadas. Todos los modelos dependientes del tiempo pronostican una baja, o incluso nula, probabilidad en los primeros años desde el último terremoto principal, consistente con el hecho de que el siguiente no ha ocurrido aún en el momento de esta redacción.

Consideraciones sobre una manera de sincronizar modelos con las fallas sísmicas que éstos simulan, para el pronóstico de terremotos

- Excepto para un proceso de Poisson, en el que los eventos carecen de correlación alguna entre ellos, un sistema estocástico se puede pronosticar hasta cierto punto, aunque la aleatoriedad impone un límite a su predecibilidad.
- Los mayores terremotos en un modelo estocástico, minimalista, de una falla sísmica son parcialmente predecibles, y su límite de predecibilidad se deduce analíticamente.
- Se demuestra que sistemas estocásticos (aquí llamados *clones*) pueden ser sincronizados parcialmente con otro (el *maestro*), de manera que los primeros pueden hasta cierto punto pronosticar la evolución de éste. Esta es la primera vez en que se propone un método de pronóstico basado en la sincronización de sistemas estocásticos.
- Esta sincronización es posible incluso sin conocer el estado interno del maestro, sino sólo imponiendo su dinámica observable a los clones.
- En particular, clones de un modelo minimalista se sincronizan parcialmente con él cuando se les fuerza a reproducir las áreas de rotura de sus terremotos.
- Para este modelo minimalista, el pronóstico de los mayores terremotos con la ayuda de clones sincronizados mejora claramente los resultados de los intentos basados sólo en la serie temporal de los terremotos más grandes.
- Los resultados de pronóstico son robustos incluso si se usan diferentes métodos de sincronización, con progresivamente menor conocimiento acerca del maestro.
- Es probable que los modelos de fallas sísmicas se puedan sincronizar con fallas reales si se les fuerza a reproducir sistemáticamente la serie real de roturas (o distribuciones de desplazamiento) que éstas generan. Con este propósito, sería adecuado forzarlos a reproducir las roturas de todos los terremotos, y no sólo las de los mayores como se ha intentado hasta ahora.

Mathematical Appendix

This appendix contains the mathematical deductions which, for brevity, are omitted from the main text. First those dealing with properties of the box model (González *et al.*, 2005, and Sec. 5.5), and last the derivation of the optimal forecasting strategy in the minimalist model (González *et al.*, 2004, 2006b, and Ch. 6)

A.1 Properties of the box model

In this section the following properties of the box model are deduced:

- the discrete probability distribution for the duration of the seismic cycle;
- the asymptotic mean and standard deviation of that distribution (when the number of cells tends to infinity); and
- the asymptotic conditional probability (when the time since the last earthquake tends to infinity).

A.1.1 Discrete probability distribution for the duration of the seismic cycle, $P_N(n)$

The discrete distribution $P_N(n)$ provides the probability that, for a box model with N cells, the seismic cycle lasts n time steps.

The box model is a Markov chain, what enables to deduce $P_N(n)$ by using a technique (e.g. Durrett, 1999) that we already applied to the minimalist model (Vázquez-Prada *et al.*, 2002), which is also Markovian. A Markov chain is a stochastic process defined by a discrete random variable X with two properties: 1) it can only take a finite number of values, and 2) its value at the next time step depends only upon the value at the present time step, being independent of the way in which the present value arose from its predecessors. In other words, the evolution of a Markov system depends only on the present state of the system and

not on the history of how the state was achieved. This kind of stochastic process is named after the Russian mathematician Andrey Andreevich Markov (1856-1922), who first described it in 1906 (Basharin *et al.*, 2004).

In a box model with N cells, the state is only determined by ν , the number of occupied cells. The succession of values of this random variable defines the stochastic process of the box model. Note that exactly which cells are occupied is not relevant, but only how many of them are. The number of stable states in the model is N ; in each of them ν takes one value in the set $\{0, 1, 2, 3 \dots (N - 1)\}$. If N cells become occupied, the system instantly changes to the empty state. It does not reside any time step in the state of complete occupancy, so this is not a stable state.

The value of ν in the next time step only depends on the value of ν in the current time step, so it follows the definition of a Markov chain. For example, if the system is empty ($\nu = 0$), in the next time step, for sure (with probability equal to 1) it will move to the state of $\nu = 1$. In this second step the fraction of occupied cells is $1/N$, and the fraction of empty cells is $(N - 1)/N$. So, in the third time step, with probability $(N - 1)/N$ another cell will be occupied by a ball (ν becoming equal to 2), or the model will remain in $\nu = 1$ with probability $1/N$ (the probability of the incoming ball landing in the only occupied cell). In general, for $\nu < N - 1$, there is a probability $(N - \nu)/N$ of moving to $\nu + 1$ in the next time step. If $\nu = N - 1$, there is a probability $(N - 1)/N$ of moving to $\nu = 0$ (passing through $\nu = N$, but not residing any moment there). In each time step there is a probability ν/N of residing in the same state during the next time step.

As for any other Markov chain, for the box model a transition matrix \mathbf{M} can be written, being a table that contains all the transition probabilities of passing, in one time step, from any of the states of the system to any of the others or to itself. Each element of the matrix will be denoted in the standard way as $\mathbf{M}(i, j)$, being i the row (from top to bottom), and j the column (from left to right). Each element gives the probability of moving from the state $X = i$ in the time step n to the state $X = j$ in the step $n + 1$:

$$\mathbf{M}(i, j) = P(X_{n+1} = j \mid X_n = i). \quad (\text{A.1})$$

The transition matrix of the box model is different for each N : as shown above, the transition probabilities depend on N , and because there are N stable states, the size of the matrix is $N \times N$. Thus I will call it as \mathbf{M}_N . Given that ν ranges from 0 to $N - 1$, while i and j range from 1 to N , the element $\mathbf{M}_N(i, j)$ will be the transition probability from $\nu = i - 1$ to $\nu = j - 1$:

$$\mathbf{M}_N(i, j) = P(\nu_{n+1} = j - 1 \mid \nu_n = i - 1). \quad (\text{A.2})$$

The seismic cycle starts when $\nu = 0$, and lasts until $\nu = 0$ again. Except for $N = 1$, which is a trivial, special case of the model, there is no possible transition in one time step from $\nu = 0$ to $\nu = 0$ (remember that this impossibility causes the stress shadow in the model). Speaking more generally, in the first $n - 1$ time steps of the cycle there is no transition to the state $\nu = 0$. Because of this, to calculate $P_N(n)$ it is necessary to deduce first the probability that the system evolves from $\nu = 0$ to $\nu = N - 1$ in $n - 1$ time steps, without having passed through $\nu = 0$ in between. To calculate the probability that the system evolves from $\nu = N - 1$ to $\nu = 0$ is simpler. At the beginning of the n -th step the system has $\nu = N - 1$. Then a new particle is added to the only one empty cell, so the occupation becomes $\nu = N$, but instantly drops to $\nu = 0$ at the end of the step. The transition in the time step is thus from $\nu = N - 1$ to $\nu = 0$. The probability for this to happen is $1/N$, the chance for the incoming particle to land in the only empty cell of the array when $\nu = N - 1$.

Thus, the deduction of $P_N(n)$ proceeds according to these steps:

1. Deduce the probabilities of passing between the different states of the system in one time step. These transition probabilities will be tabulated in the transition matrix \mathbf{M}_N .
2. Remove from \mathbf{M}_N the possibility of intermediate transitions to $\nu = 0$. The resulting matrix will be called \mathbf{M}'_N .
3. Calculate the transition probabilities of passing between the different states in $n - 1$ time steps and neglecting the possibility of passing through the state with $\nu = 0$. The result is a new matrix, $\mathbf{T}_N = \mathbf{M}'_N{}^{n-1}$.
4. One of the elements of this matrix will indicate the probability of passing from $\nu = 0$ to $\nu = N - 1$ in $n - 1$ time steps without having passed through $\nu = 0$ in between. Multiplying this probability by $1/N$ we will obtain $P_N(n)$.

Regarding the first step above, the transition matrices for N equal to 2, 3 and 4 are as follows:

For $N = 2$,

$$\mathbf{M}_2 = \begin{pmatrix} 0 & 1 \\ \frac{1}{2} & \frac{1}{2} \end{pmatrix} = \frac{1}{2} \begin{pmatrix} 0 & 2 \\ 1 & 1 \end{pmatrix}; \quad (\text{A.3})$$

for $N = 3$,

$$\mathbf{M}_3 = \begin{pmatrix} 0 & 1 & 0 \\ 0 & \frac{1}{3} & \frac{2}{3} \\ \frac{1}{3} & 0 & \frac{2}{3} \end{pmatrix} = \frac{1}{3} \begin{pmatrix} 0 & 3 & 0 \\ 0 & 1 & 2 \\ 1 & 0 & 2 \end{pmatrix}; \quad (\text{A.4})$$

and for $N = 4$,

$$\mathbf{M}_4 = \begin{pmatrix} 0 & 1 & 0 & 0 \\ 0 & \frac{1}{4} & \frac{3}{4} & 0 \\ 0 & 0 & \frac{2}{4} & \frac{2}{4} \\ \frac{1}{4} & 0 & 0 & \frac{3}{4} \end{pmatrix} = \frac{1}{4} \begin{pmatrix} 0 & 4 & 0 & 0 \\ 0 & 1 & 3 & 0 \\ 0 & 0 & 2 & 2 \\ 1 & 0 & 0 & 3 \end{pmatrix}. \quad (\text{A.5})$$

Since all the elements of these matrices are probabilities, they are nonnegative and the sum of all the elements of any row is always 1. These two are necessary and sufficient properties of transition matrices of Markov chains. These matrices show evident regularities, which enable to deduce by inspection that the matrix for any N is

$$\mathbf{M}_N = \frac{1}{N} \begin{pmatrix} 0 & N & 0 & 0 & \dots & 0 \\ 0 & 1 & N-1 & 0 & \dots & 0 \\ 0 & 0 & 2 & N-2 & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & 0 & 0 & N-2 & 2 \\ 1 & 0 & 0 & 0 & 0 & N-1 \end{pmatrix}. \quad (\text{A.6})$$

Note that the matrix multiplied by $1/N$ has only three non-null diagonals, all of them trivial. The first one is the sequence $N, N-1, N-2 \dots 2$, the second one is the sequence $0, 1, 2 \dots N-1$, and the third one is only the bottom left element, which is always 1.

The second step is pruning from this matrix the transitions to $\nu = 0$. The only possible transition to $\nu = 0$ is from $\nu = N-1$, and the probability for this transition is given by the bottom left element $\mathbf{M}_N(N, 1)$. Nullifying this element, the resulting matrix, \mathbf{M}'_N , is particularly simple, because it has only two trivial, non-null diagonals:

$$\mathbf{M}'_N = \frac{1}{N} \begin{pmatrix} 0 & N & 0 & 0 & \dots & 0 \\ 0 & 1 & N-1 & 0 & \dots & 0 \\ 0 & 0 & 2 & N-2 & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & 0 & 0 & N-2 & 2 \\ 0 & 0 & 0 & 0 & 0 & N-1 \end{pmatrix}. \quad (\text{A.7})$$

The third step requires computing a new matrix, \mathbf{T}_N , which indicates all the transition probabilities, in $n-1$ time steps, between all the states, with the restriction that passing from $\nu = N-1$ to $\nu = 0$ is forbidden. In Markov chains, the m -step transition probability matrix is the m -th power of the transition matrix (Durrett, 1999). So the new matrix is

$$\mathbf{T}_N = \mathbf{M}'_N{}^{m-1}. \quad (\text{A.8})$$

This operation is done through the Jordan decomposition of \mathbf{M}'_N . The element $\mathbf{T}_N(1, N)$ of this matrix is the transition probability from $\nu = 0$ to $\nu = N - 1$ in $n - 1$ time steps and with the transition $\nu = N - 1 \rightarrow \nu = 0$ forbidden. As the probability of passing, in the next time step, from $\nu = N - 1$ to $\nu = 0$ is $1/N$, $P_N(n)$ is obtained by multiplying that element by $1/N$. The results, for $N = 2$ and $N = 3$ are as follows.

For $N = 2$,

$$\begin{aligned} \frac{1}{2^{n-1}} &= \frac{2}{2^n} = \frac{1}{2^n} \sum_{j=0}^{2-1} \left[\binom{2}{j} j^{n-1} (-1)^{1-j} (2-j) \right] = \\ &= \frac{1}{2^n} [0 + 2]; \end{aligned} \quad (\text{A.9})$$

and for $N = 3$,

$$\frac{2}{3^{n-1}} (2^{n-2} - 1) = \frac{1}{3^n} \sum_{j=0}^{3-1} \left[\binom{3}{j} j^{n-1} (-1)^{1-j} (3-j) \right]. \quad (\text{A.10})$$

By inspection, the result for a generic N is

$$\begin{aligned} P_N(n) &= \frac{1}{N^n} \sum_{j=0}^{N-1} \left[\binom{N}{j} j^{n-1} (-1)^{1-j} (N-j) \right] = \\ &= \sum_{j=1}^{N-1} (-1)^{j+1} \binom{N-1}{j-1} \left(1 - \frac{j}{N}\right)^{n-1}. \end{aligned} \quad (\text{A.11})$$

A.1.2 Asymptotic mean of $P_N(n)$

The mean duration of the cycle in the box model was indicated in Sec. 5.5.2:

$$\begin{aligned} \langle n \rangle_N &= 1 + \sum_{i=2}^N \frac{N}{N+1-i} = \\ &= 1 + \frac{N}{N-1} + \frac{N}{N-2} + \cdots + \frac{N}{2} + \frac{N}{1} = \\ &= N \left[\frac{1}{N} + \frac{1}{N-1} + \frac{1}{N-2} + \cdots + \frac{1}{2} + \frac{1}{1} \right]. \end{aligned} \quad (\text{A.12})$$

The asymptotic value of this expression can be obtained considering that (Gradshteyn & Ryzhik, 1965)

$$\sum_{i=1}^N \frac{1}{i} \xrightarrow{N \rightarrow \infty} C + \ln N + \frac{1}{2N} - o\left(\frac{1}{N}\right)^2, \quad (\text{A.13})$$

where $C \simeq 0.5772157$ is Euler's constant. Multiplying this equation by N , the asymptotic mean of $P_N(n)$ is obtained:

$$\langle n \rangle_N \xrightarrow{N \rightarrow \infty} N(C + \ln N) + \frac{1}{2}. \quad (\text{A.14})$$

The absolute error of this approximation is < 0.01 for $N \geq 9$.

A.1.3 Asymptotic standard deviation of $P_N(n)$

The variance of $P_N(n)$ (Sec. 5.5.2) is:

$$\begin{aligned}\sigma_N^2 &= \sum_{i=1}^N \frac{1 - \frac{N+1-i}{N}}{\left(\frac{N+1-i}{N}\right)^2} = \\ &= \sum_{i=1}^N \frac{1}{\left(\frac{N+1-i}{N}\right)^2} - \sum_{i=1}^N \frac{1}{\frac{N+1-i}{N}}.\end{aligned}\quad (\text{A.15})$$

To simplify the sums, the variable can be changed to $k \equiv N+1-i$. Since i ranges from 1 to N , k will range from N to 1. Then the above equation can be rewritten as

$$\begin{aligned}\sigma_N^2 &= \sum_{k=1}^N \frac{1}{\left(\frac{k}{N}\right)^2} - \sum_{k=1}^N \frac{1}{\frac{k}{N}} = \sum_{k=1}^N \frac{N^2}{k^2} - \sum_{k=1}^N \frac{N}{k} = \\ &= N^2 \sum_{k=1}^N \frac{1}{k^2} - N \sum_{k=1}^N \frac{1}{k}.\end{aligned}\quad (\text{A.16})$$

The first sum in the right-hand side of this expression can be simplified to

$$\begin{aligned}\sum_{k=1}^N \frac{1}{k^2} &= \sum_{k=1}^{\infty} \frac{1}{k^2} - \int_N^{\infty} \frac{1}{k^2} dk = \\ &= \frac{\pi^2}{6} - \left[-\frac{1}{k}\right]_N^{\infty} = \frac{\pi^2}{6} - \frac{1}{N}.\end{aligned}\quad (\text{A.17})$$

Inserting Eq. A.13 and this result, Eq. A.16 in the limit of $N \rightarrow \infty$ can be written as

$$\begin{aligned}\sigma_N^2 \xrightarrow{N \rightarrow \infty} & N^2 \left(\frac{\pi^2}{6} - \frac{1}{N} \right) - \\ & - N \left[C + \ln N + \frac{1}{2N} - 0 \left(\frac{1}{N} \right)^2 \right] = \\ & = N^2 \frac{\pi^2}{6} - N - CN - N \ln N - \frac{1}{2} = \\ & = N^2 \frac{\pi^2}{6} - N(1 + C + \ln N) - \frac{1}{2} = \\ & = N^2 \left[\frac{\pi^2}{6} - \frac{1 + C + \ln N}{N} - \frac{1}{2N^2} \right].\end{aligned}\quad (\text{A.18})$$

The asymptotic standard deviation is the square root of the above equation,

$$\sigma_N \xrightarrow{N \rightarrow \infty} N \left[\frac{\pi^2}{6} - \frac{1 + C + \ln N}{N} - \frac{1}{2N^2} \right]^{1/2}.\quad (\text{A.19})$$

Because $N \rightarrow \infty$, the term $-1/2N^2$ can be dropped, so the equation can be further simplified to

$$\sigma_N \xrightarrow{N \rightarrow \infty} N \left[\frac{\pi^2}{6} - \frac{1 + C + \ln N}{N} \right]^{1/2}.\quad (\text{A.20})$$

This approximation has an absolute error < 0.01 for $N \geq 3$.

The asymptotic aperiodicity, obtained by dividing Eq. A.20 by Eq. A.14, has an absolute error < 0.0001 for $N \geq 10$.

A.1.4 Asymptotic conditional probability

To deduce the asymptotic conditional probability in the box model I will first consider the asymptotic value of $P_N(n)$ when $n \rightarrow \infty$. This value is the first, largest term in the sum (when $j = 1$ in Eq. A.11), namely

$$P_N(n) \xrightarrow{n \rightarrow \infty} \left(1 - \frac{1}{N}\right)^{n-1} = a^{n-1}, \quad (\text{A.21})$$

where we have denoted $a \equiv 1 - 1/N$.

For calculating the asymptotic conditional probability it is necessary to deduce the value of the cumulative probability distribution, $A_N(n)$, for that large n . This is easier to do by defining the sum

$$A'_N(n) = \sum_{i=n}^{\infty} a^{i-1} = \frac{a^{n-1}}{1-a}. \quad (\text{A.22})$$

Considering that n is large enough, $P_N(n)$ can be replaced by its asymptotic value (Eq. A.21), which is the term summed in $A'_N(n)$. Then it holds that

$$A_N(n-1) = \sum_{i=1}^{n-1} P_N(n) \xrightarrow{n \rightarrow \infty} 1 - A'_N(n), \quad (\text{A.23})$$

so

$$A_N(n) \xrightarrow{n \rightarrow \infty} 1 - A'_N(n+1). \quad (\text{A.24})$$

The conditional probability for the next interval of Δn time steps is:

$$P_\tau(N, n) = \frac{A_N(n + \Delta n) - A_N(n)}{1 - A_N(n-1)}. \quad (\text{A.25})$$

Inserting Eqs. A.22 to A.24, it results that

$$\begin{aligned} P_\tau(N, n) &\xrightarrow{n \rightarrow \infty} \frac{1 - A'_N(n + \Delta n + 1) - [1 - A'_N(n + 1)]}{A'_N(n)} = \\ &= \frac{A'_N(n + 1) - A'_N(n + \Delta n + 1)}{A'_N(n)} = \\ &= \frac{a^{n+1-1} - a^{n+\Delta n+1-1}}{\frac{1-a}{a^{n-1}}} = \\ &= \frac{1-a}{1-a} = \\ &= a - a^{1+\Delta n} = a \left(1 - a^{\Delta n}\right) = \\ &= \left(1 - \frac{1}{N}\right) \left[1 - \left(1 - \frac{1}{N}\right)^{\Delta n}\right]. \end{aligned} \quad (\text{A.26})$$

A.2 Deduction of the ideal forecasting strategy in the minimalist model

In this appendix we will deduce the ideal strategy outlined in Section 6.3.3. This strategy renders the lowest (best) value of $L = f_a + f_e$ achievable in the minimalist model.

For this reasoning we would consider every cycle of the model as composed of two independent and consecutive stages. The first, that will be called the *loading* stage, starts just after the occurrence of a characteristic earthquake. During this stage the total number of occupied cells grows, but not in a monotonic way, because the particles may land in already occupied cells (and then be dissipated), and also because of the occurrence of non-characteristic earthquakes (Fig. 2). When $N - 1$ cells (all but the first one) become occupied, this first stage ends and the second stage, that will be called the *hitting* stage (or *plateau* in the occupation), starts. In this second stage, the system resides statically in the state of maximum occupancy (Fig. 2) until a particle arrives at the first cell. Then, a characteristic event occurs, all the cells are emptied, and a new cycle begins. The hitting stage can be mathematically treated as a form of Russian roulette.

Both the time spent by the system in the loading stage, x , and in the hitting stage, y , are statistically distributed. The distribution of y , denoted by $P_2(y)$, is geometric. Considering that, in each time step, the probability of hitting the first cell is $p = 1/N$, and its complementary is $q = 1 - 1/N$, it follows that

$$P_2(y) = \frac{1}{N} \left(1 - \frac{1}{N}\right)^{y-1}, \quad (\text{A.27})$$

whose mean is

$$\langle y \rangle = N, \quad (\text{A.28})$$

and whose standard deviation is

$$\sigma = N \sqrt{1 - \frac{1}{N}}. \quad (\text{A.29})$$

The time elapsed between consecutive characteristic events has been denoted by n , which is statistically distributed according to the function $P_N(n)$ (Vázquez-Prada *et al.*, 2002, 2003; Gómez & Pacheco, 2004). Because the variables x and y are independent, the mean length of the cycles $\langle n \rangle$ is the sum of the mean lengths of the two stages:

$$\langle n \rangle = \langle x \rangle + \langle y \rangle. \quad (\text{A.30})$$

It is clear that the best L would be obtained *only if* we knew the state of occupation of the system and could mark, for each cycle, the instant at which the

stage of loading concludes. In this case, $f_e = 0$, but because the hitting stage is completely stochastic, f_a (and thus L) cannot be nil.

Let us explore the result of L obtained if we turn the alarm on at a given value $y = y_0$ within the second stage of all the cycles. With this strategy, the fraction of errors is given by

$$f_e(y_0) = \sum_1^{y_0} P_2(y), \quad (\text{A.31})$$

and inserting Eq. A.27 we obtain

$$f_e(y_0) = 1 - q^{y_0}. \quad (\text{A.32})$$

With respect to to the fraction of alarm, its form is

$$f_a(y_0) = \frac{\sum_{y_0}^{\infty} (y - y_0) \cdot P_2(y)}{\langle x \rangle + \langle y \rangle}, \quad (\text{A.33})$$

and inserting Eq. A.27, we get

$$f_a(y_0) = \frac{Nq^{y_0}}{\langle x \rangle + N}. \quad (\text{A.34})$$

Note the important contribution of the first stage of the process in the denominator. Thus, the specific form of the loss function is

$$L(y_0) = 1 - q^{y_0} + \frac{Nq^{y_0}}{\langle x \rangle + N}. \quad (\text{A.35})$$

It is noteworthy that in the absence of the first stage, i.e. in the hypothesis of a pure geometric distribution, the value of L would be 1, not dependent on the value of y_0 . In this sense, the geometrical and the Poisson distributions are equivalent. The minimum value of L in Eq. A.35 as a function of y_0 is obtained for $y_0 = 0$, i.e. just after the end of the first stage, when the $N - 1$ upper cells of the system are full. And this minimum value is

$$L_{min} = \frac{N}{\langle n \rangle}. \quad (\text{A.36})$$

This result constitutes a rigorous lower bound for the expected accuracy of any forecasting strategy in the minimalist model. For this model, $\langle n \rangle$ increases rapidly as N grows (Gómez & Pacheco, 2004). This implies that the minimum L , obtained with this optimal forecasting strategy, decreases as N increases, as shown by the curve labeled as “Ideal” in Fig. 3. That is to say, minimalist models with more cells are more predictable. This is consistent with the fact that the time series of characteristic earthquakes is more periodic for larger N (Gómez & Pacheco, 2004).

Bibliography

Para un entendimiento superior [...] en vez de ciencias habría una *sola Ciencia*, cuyas ramas estuvieran representadas por las ciencias particulares. [...] El científico general [...] entrevé el tallo común a muchas ramas.

Santiago Ramón y Cajal (1916)

- Abadías, N.L.; González, Á.; Gómez, J.B. & Pacheco, A.F. (2006) Nuevo modelo de renovación para la recurrencia de terremotos en una falla. *Revista de la Real Academia de Ciencias Exactas, Físicas, Químicas y Naturales de Zaragoza*, 61, 121–134.
- Abaimov, S.G.; Turcotte, D.L.; Shcherbakov, R.; Rundle, J.B.; Yakovlev, G.; Goltz, C. & Newman W.I. (2008) Earthquakes: Recurrence and interoccurrence times. *Pure and Applied Geophysics*, 165, 777–795.
- Adler, I.; Barabé, D. & Jean, R. V. (1997) A history of the study of phyllotaxis. *Annals of Botany*, 80, 231–244.
- Advanced National Seismic System (2015) *ANSS Comprehensive Earthquake Catalog*, earthquake.usgs.gov/earthquakes/search/ [Last accessed: November 2015.]
- Ahmad, R.; Deng, Y.; Vikram, D. S.; Clymer, B.; Srinivasan, P.; Zweier, J. L. & Kuppusamy, P. (2007) Quasi Monte Carlo-based isotropic distribution of gradient directions for improved reconstruction quality of 3D EPR imaging. *Journal of Magnetic Resonance*, 184, 236–245.
- Aki, K. (1965) Maximum likelihood estimate of b in the formula $\log N = a - bM$ and its confidence limits. *Bulletin of the Earthquake Research Institute, University of Tokyo*, 43, 237–239.
- Aki, K. (1984) Asperities, barriers, characteristic earthquakes and strong motion prediction. *Journal of Geophysical Research*, 89, 5867–5872.
- Albarello, D.; Camassi, R. & Rebez, A. (2001) Detection of space and time heterogeneity in the completeness of a seismic catalog by a statistical approach: An application to the Italian area. *Bulletin of the Seismological Society of America*, 91, 1694–1703.
- Alguacil, G. & Martín Dávila, J. (2003) Stations and seismological networks in the Iberian Peninsula: State of the art. *Física de la Tierra*, 15, 29–48.
- Almendros, J.; Ibáñez, J.M.; Alguacil, G.; Morales, J.; Del Pezzo, E.; La Rocca, M.; Ortiz, R.; Araña, V. & Blanco, M.J. (2000) A double seismic antenna experiment at Teide Volcano: existence of local seismicity and lack of evidences of volcanic tremor. *Journal of Volcanology and Geothermal Research*, 103, 439–462.

- Almendros, J.; Ibáñez, J.M.; Carmona, E. & Zandomenighi, D. (2007) Array analyses of volcanic earthquakes and tremor recorded at Las Cañadas caldera (Tenerife Island, Spain) during the 2004 seismic activation of Teide volcano, *Journal of Volcanology and Geothermal Research* 160, 285–299.
- Alvarez-Garcia, I.N.; Ramos-Lopez, F.L.; Gonzalez-Nicieza, C.; Alvarez-Fernandez, M.I. & Alvarez-Vigil, A.E. (2013) The mine collapse at Lo Tacón (Murcia, Spain), possible cause of the Torre Pacheco earthquake (2nd May 1998, SE Spain), *Engineering Failure Analysis*, 28, 115–133.
- Amante, C. & Eakins, B.W. (2009) *ETOPO1 1 Arc-Minute Global Relief Model: Procedures, Data Sources and Analysis*. NOAA Technical Memorandum NESDIS NGDC-24, 19 p.
- Amato, A. & Mele, F.M. (2008) Performance of the INGV National Seismic Network from 1997 to 2007. *Annals of Geophysics*, 51, 417–431.
- Ammon, C.J.; Lay, T. & Simpson, D.W. (2010) Great earthquakes and global seismic networks. *Seismological Research Letters*, 81, 965–971.
- Anduaga, A. (2004) Earthquakes, damage, and prediction: The Spanish Seismological Service, 1898–1930. *Earth Sciences History*, 23, 175–207.
- Anduaga, A. (2009) *Geofísica, economía y sociedad en la España contemporánea*. Consejo Superior de Investigaciones Científicas, Madrid, Spain, 379 p.
- Anthony J.L. & Marone, C. (2005) Influence of particle characteristics on granular friction. *Journal of Geophysical Research*, 110, B08409.
- Arenas, A.; Díaz-Guilera, A.; Kurths, J.; Moreno, Y. & Zhou, C. (2008) Synchronization in complex networks. *Physics Reports*, 469, 93–153.
- Arnold, T.B. & Emerson, J.W. (2011) Nonparametric goodness-of-fit tests for discrete null distributions. *The R Journal*, 3 (2), 34–39.
- Atef, A.H.; Liu, K.H. & Gao, S.S. (2009) Apparent weekly and daily earthquake periodicities in the western United States. *Bulletin of the Seismological Society of America*, 99, 2273–2279.
- Ayadi, A. & Bezzeghoud, M. (2015) Seismicity of Algeria from 1365 to 2013: Maximum Observed Intensity Map (MOI₂₀₁₄). *Seismological Research Letters*, 86, 236–244.
- Baba, T. & Cummins, P. R. (2005) Contiguous rupture areas of two Nankai Trough earthquakes revealed by high-resolution tsunami waveform inversion. *Geophysical Research Letters*, 32, Art. No. L08305.
- Badal, J.; Samardjieva, E. & Payo, G. (2000) Moment magnitudes for early (1923–1961) instrumental Iberian earthquakes. *Bulletin of the Seismological Society of America*, 90, 1161–1173.
- Baddeley, A. & Jensen, E. B. V. (2004) *Stereology for Statisticians*. CRC Press, Boca Raton, Florida. 416 p.
- Badii, R. & Politi, A. (1984a) Hausdorff dimension and uniformity factor of strange attractors. *Physical Review Letters*, 52, 1661–1664.
- Badii, R. & Politi, A. (1984b) Intrinsic oscillations in measuring the fractal dimension. *Physics Letters*, 104A, 303–305.
- Badii, R. & Politi, A. (1985) Statistical description of chaotic attractors: The dimension function. *Journal of Statistical Physics*, 40, 725–750.

- Bahavar, M. & North, R. (2002) Estimation of background noise for International Monitoring System seismic stations. *Pure and Applied Geophysics*, 159, 911–944.
- Bak, P. & Tang, C. (1989) Earthquakes as a self-organized critical phenomenon. *Journal of Geophysical Research*, 94, 15635–15637.
- Bakun W. H. (1988) History of significant earthquakes in the Parkfield area. *Earthquakes and Volcanoes*, 20, 45–51.
- Bakun W. H. & Lindh, A. G. (1985) The Parkfield, California, earthquake prediction experiment. *Science*, 229, 619–624.
- Bakun, W.H. *et al.* (2005) Implications for prediction and hazard assessment from the 2004 Parkfield earthquake. *Nature*, 437, 969–974.
- Barclay, M. & Galton, A. (2008) Comparison of region approximation techniques based on Delaunay triangulations and Voronoi diagrams. *Computers, Environment and Urban Systems*, 32, 261–267.
- Bardsley, W. E. (1983) Random error in point counting. *Mathematical Geology*, 15, 469–475.
- Barenblatt, G.I. (2003) *Scaling*, Cambridge University Press, Cambridge, UK, 171 p.
- Barnsley, M.F. (2012) *Fractals Everywhere*, 3rd ed. Dover Publications, Mineola, New York, USA, 560 p.
- Baró, J.; Corral, Á.; Illa, X.; Planes, A.; Salje, E.K.H.; Schranz, W.; Soto-Parra, D.E. & Vives, E. (2013) Statistical similarity between the compression of a porous material and earthquakes. *Physical Review Letters*, 110, 088702.
- Basharin, G.P.; Langville, A.N. & Naumov, V.A. (2004) The life and work of A.A. Markov. *Linear Algebra and its Applications*, 386, 3–26.
- Båth, M. (compiler, 1969) *Handbook on Earthquake Magnitude Determinations*, 2nd ed.; VESIAC Special Report, Geophysics Laboratory, Willow Run Laboratories, University of Michigan, 158 p.
- Båth, M. (1981) Earthquake magnitude - recent research and current trends. *Earth-Science Reviews*, 17, 315–398.
- Batló, J. (2004) *Catálogo inventario de sismógrafos antiguos españoles [A Catalogue of Old Spanish Seismographs]*. Instituto Geográfico Nacional, Madrid, Spain, 413 p.
- Batló, J. & Bormann, P. (2000) A catalog of old Spanish seismographs. *Seismological Research Letters*, 71, 570–582.
- Batló, J.; Stich, D.; Palombo, B.; Macia, R. & Morales, J. (2008) The 1951 M_w 5.2 and M_w 5.3 Jaén, Southern Spain, earthquake doublet revisited. *Bulletin of the Seismological Society of America*, 98, 1535–1545.
- Bauer, R. (2000) Distribution of points on a sphere with application to star catalogs. *Journal of Guidance, Control and Dynamics*, 23, 130–137.
- Becerril, L.; Cappello, A.; Galindo, I.; Neri, M. & Del Negro, C. (2013) Spatial probability distribution of future volcanic eruptions at El Hierro Island (Canary Islands, Spain). *Journal of Volcanology and Geothermal Research*, 257, 21–30.
- Becker, T.W.; Lowry, A.R.; Faccenna, C.; Schmandt, B.; Borsa A. & Yu, C. (2015) Western US intermountain seismicity caused by changes in upper mantle flow. *Nature*, 524, 458–461.
- Benedetti, L.; Manighetti, I.; Gaudemer, Y.; Finkel, R.; Malavieille, J.; Pou, K.; Arnold, M.; Aumaitre, G.; Bourlés, D. & Keddadouche, K. (2013) Earthquake synchrony and

- clustering on Fucino faults (Central Italy) as revealed from in situ ^{36}Cl exposure dating. *Journal of Geophysical Research*, 118, 4948–4974.
- Benito, B. *et al.* (2007) An overview of the damaging and low magnitude M_w 4.8 La Poca earthquake on 29 January 2005: context, seismotectonics, and seismic risk implications for Southeast Spain. *Bulletin of the Seismological Society of America*, 97, 671–690.
- Benouar, D. (1994) Materials for the investigation of the seismicity of Algeria and adjacent regions during the twentieth century. *Annali di Geofisica*, 37, 459–860.
- Benson, R.; Lindholm, C.D.; Ludwin, R. & Qamar, A. (1992) A method for identifying explosions contaminating earthquake catalogs: Application to the Washington Regional Earthquake Catalog. *Seismological Research Letters*, 63, 533–539 .
- Ben-Zion, Y. (2001) Dynamic ruptures in recent models of earthquake faults. *Journal of the Mechanics and Physics of Solids*, 49, 2209–2244.
- Ben-Zion, Y. (2008) Collective behavior of earthquakes and faults: Continuum-discrete transitions, progressive evolutionary changes, and different dynamic regimes. *Reviews of Geophysics*, 46, RG4006.
- Berryman, K.R.; Cochran, U.A.; Clark, K.J.; Biasi, G.P.; Langridge, R.M. & Villamor, P. (2012) Major earthquakes occur regularly on an isolated plate boundary fault. *Science*, 336, 1690-1693.
- Beverington, P. R. & Robinson, D. K. (1992) *Data Reduction and Error Analysis for the Physical Sciences*, 2nd edn. McGraw Hill, Boston, 328 p.
- Bevis, M. & Cambareri, G. (1987) Computing the area of a spherical polygon of arbitrary shape. *Mathematical Geology*, 19, 335–346.
- Bezada, M.J. & Humphreys, E.D. (2012) Contrasting rupture processes during the April 11, 2010 deep-focus earthquake beneath Granada, Spain. *Earth and Planetary Science Letters*, 353–354, 38–46.
- Biasi, G.P.; Langridge, R.M.; Berryman, K.R.; Clark, K.J. & Cochran, U.A. (2015) Maximum-likelihood recurrence parameters and conditional probability of a ground-rupturing earthquake on the southern Alpine Fault, South Island, New Zealand. *Bulletin of the Seismological Society of America*, 105, 94–106.
- Biegalski, K.F.; Bohlin, J.; Carter, J.A.; Coyne, J.; Dompierre, D.; Novosel, G. & Rinehart, C. (1999) *Formats and Protocols for Messages – IMS1.0*. International Data Center, Document SAIC-99/3004, 324 p.
- Binder, K. (1997) Applications of Monte Carlo methods to statistical physics. *Reports on Progress in Physics*, 60, 487–559.
- Blackett, M.; Wooster, M.J. & Malamud, B.D. (2011) Exploring land surface temperature earthquake precursors: A focus on the Gujarat (India) earthquake of 2001. *Geophysical Research Letters*, 38, L15303.
- Boccaletti, S.; Kurths, J.; Osipov, G.; Valladares, D. L. & Zhou, C. S. (2002) The synchronization of chaotic systems. *Physics Reports*, 366, 1–101.
- Boettcher, M.S.; McGarr, A. & Johnston, M. (2009) Extension of Gutenberg- Richter distribution to M_w -1.3, no lower limit in sight. *Geophysical Research Letters*, 36, L10307.
- Bondár, I. & Storchack, D. (2011) Improved location procedures at the International Seismological Centre. *Geophysical Journal International*, 186, 1220–1244.

- Boness, N.L. & Zoback, M.D. (2004) Stress-induced seismic velocity anisotropy and physical properties in the SAFOD Pilot Hole in Parkfield, CA. *Geophysical Research Letters*, 31, L15S17.
- Bonnefoy-Claudet, S.; Cotton, F. & Bard, P.Y. (2006) The nature of noise wavefield and its applications for site effects studies. A literature review. *Earth-Science Reviews*, 79, 205–227.
- Bonnet, E.; Bour, O.; Odling, N.E.; Davy, P.; Main, I.; Cowie, P. & Berkowitz B. (2001) Scaling of fracture systems in geological media. *Reviews of Geophysics*, 39, 347–383.
- Book, S.A. (1979) Why $n - 1$ in the formula for the sample standard deviation? *The Two-Year College Mathematics Journal*, 10, 330–333.
- Bourgeois, J. (2006) Earthquakes – A movement in four parts? *Nature*, 440, 430–431.
- Bowman, D.D.; Ouillon, G.; Sammis, C.G.; Sornette, A. & Sornette, D.; 1998. An observational test of the critical earthquake concept. *Journal of Geophysical Research*, 103, 24359–24372.
- Box, G. E. P. (1976) Science and Statistics. *Journal of the American Statistical Association*, 71, 791–799.
- Box, G. E. P. & Draper, N. R. (1987) *Empirical Model Building and Response Surfaces*. John Wiley & Sons, New York, NY. 669 p.
- Boyd, T.M. & Snoke, J.A. (1984) Error estimates in some commonly used earthquake location programs, *Earthquake Notes*, 55, 3–6.
- Brace W.F. & Byerlee, J.D. (1966) Stick-Slip as a Mechanism for Earthquakes. *Science*, 153, 990–992.
- Bratt, S.R. & Bache, T.C. (1988) Locating events with a sparse network of regional arrays. *Bulletin of the Seismological Society of America*, 78, 780–798.
- Brauchart, J. S. (2004) Invariance principles for energy functionals on spheres. *Monatshefte für Mathematik*, 141, 101–117.
- Bravais, L. & Bravais, A. (1837) Essai sur la disposition des feuilles curvisériées. *Annales des Sciences Naturelles* 7, 42–110 & plates 2–3.
- Brockmeyer, E.; Poupyrev, I. & Hudson, S. (2013) PAPILLON: designing curved display surfaces with printed optics. In: *Proceedings of the 26th Annual ACM Symposium on User Interface Software and Technology*, St. Andrews, UK, p. 457–462.
- Brodsky E.E. & Lay, T. (2014) Recognizing foreshocks from the 1 April 2014 Chile earthquake. *Science*, 344, 700–702.
- Bufo, E.; Coca, P.; Udías, A. & Lasa, C. (1997) Source mechanism of intermediate and deep earthquakes in southern Spain. *Journal of Seismology*, 1, 113–130.
- Bufo, E.; Pro, C.; Cesca, S.; Udías, A. & del Fresno, C. (2011) The 2010 Granada, Spain, deep earthquake. *Bulletin of the Seismological Society of America*, 101, 2418–2430.
- Burtin, A.; Bollinger, L.; Vergne, J.; Cattin, R. & Nábělek, J.L. (2008) Spectral analysis of seismic noise induced by rivers: A new tool to monitor spatiotemporal changes in stream hydrodynamics. *Journal of Geophysical Research*, 113, B05301.
- Cabañas, L.; Alcalde, J.M.; Carreño, E. & Bravo, J.B. (2014) Characteristics of observed strong motion accelerograms from the 2011 Lorca (Spain) Earthquake. *Bulletin of Earthquake Engineering*, 12, 1909–1932.

- Cabañas, L.; Rivas-Medina, A.; Martínez-Solares, J.M.; Gaspar-Escribano, J.M.; Benito, B.; Antón, R. & Ruiz-Barajas, S. (2015) Relationships between M_w and other earthquake size parameters in the Spanish IGN seismic catalog. *Pure and Applied Geophysics*, 172, 2397–2410.
- Calvert, A.; Sandvol, E.; Seber, D.; Barazangi, M.; Vidal, F.; Alguacil, G. & Jabour, N. (2000) Propagation of regional seismic phases (Lg and S_n) and P_n velocity structure along the Africa–Iberia plate boundary zone: tectonic implications. *Geophysical Journal International*, 142, 384–408.
- Calvo, O.; Chialvo, D. R.; Eguíluz, V. M.; Mirasso, C. & Toral, R. (2004) Anticipated synchronization: A metaphorical linear view. *Chaos*, 14, 7–13.
- Canas Torres, J.A. & Vidal Sánchez, F. (2001) La nueva red sísmica nacional. In: *El Riesgo Sísmico. Prevención y Seguro*, Capote del Villar, R. & Martínez Díaz, J.J. (Eds.), pp. 147–162. Consorcio de Compensación de Seguros, Madrid (Spain).
- Carniel, R.; Tárraga, M.; Barazza, F. & García, A. (2008) Possible interaction between tectonic events and seismic noise at Las Cañadas Volcanic Caldera, Tenerife, Spain. *Bulletin of Volcanology*, 70, 1113–1121.
- Carreño, E. (1999) Instituto Geográfico Nacional (IGN), Madrid, Spain. *CSEM-EMSC Newsletter*, 15, 10–11.
- Carreño Herrero, E. & Valero Zornoza, J.F. (2011) The Iberian Peninsula seismicity for the instrumental period: 1985–2011. *Enseñanza de las Ciencias de la Tierra*, 19, 289–295.
- Carreño, E.; Galán, J. & Sánchez, M. (1991a) Microseismicity studies in Southern Spain. In: *Seismicity, Seismotectonics and Seismic Risk of the Ibero-Maghrebian Region*, Mezcua, J. & Udías, A. (Eds.), pp. 79–86. Monografía 8, Instituto Geográfico Nacional, Madrid, Spain.
- Carreño, E.; Rueda, J.; López Casado, C.; Galán, J. & Peláez, J.A. (1991b) Spanish national strong motion network. Recording of the Huelva earthquake of 20 December, 1989. *Pure and Applied Geophysics*, 136, 395–404.
- Carreño, E.; López, C.; Villamayor, A.; Gracia, F.; Alonso, A. & Ruiz, E. (2001) Estación sísmica de Sonseca. Descripción general. In: *El Riesgo Sísmico. Prevención y Seguro*, Capote del Villar, R. & Martínez Díaz, J.J. (Eds.), pp. 163–177. Consorcio de Compensación de Seguros, Madrid (Spain).
- Carreño, E.; López, C.; Bravo, B.; Expósito, P.; Gurría, E. & García, O. (2003) Seismicity of the Iberian Peninsula in the instrumental period: 1985–2002. *Física de la Tierra*, 15, 73–91.
- Carrilho, F.J.R.; Nunes, J.A.C.; Pena, J.O.A. & Senos, M.L. (2004) *Catálogo Sísmico de Portugal Continental e Região Adjacente para o Período 1970–2000*. Instituto de Meteorologia, Lisbon, Portugal, 227 p.
- Cassidy, J.F. (2013) Seismic Gap. In: *Encyclopedia of Natural Hazards*. Springer, Dordrecht, The Netherlands, p. 906.
- Castellaro, S. & Mulargia, F. (2001) A simple but effective cellular automaton for earthquakes. *Geophysical Journal International*, 144, 609–624.
- Cattaneo, M.; Caffagni, E.; Carannante, S.; D’Alema, E.; Frapiccini, M.; Ladina, C.; Marzorati, S. & Monachesi, G. (2014) A catalogue of non-tectonic earthquakes in central-eastern Italy. *Annals of Geophysics*, 57, S0328.

- Cesca, S.; Grigoli, F.; Heimann, S.; González, Á.; Buforn, E.; Maghsoudi, S.; Blanch, E. & Dahm, T. (2014) The 2013 September–October seismic sequence offshore Spain: a case of seismicity triggered by gas injection? *Geophysical Journal International*, 198, 941–953.
- Chadwick, J. (2007) The riches of Neves-Corvo and Aljustrel. *International Mining*, January 2007, 42–51.
- Chazalon, A.; Campillo, M.; Gibson, R. & Carreno, E. (1993) Crustal wave propagation anomaly across the Pyrenean Range. Comparison between observations and numerical simulations. *Geophysical Journal International*, 115, 829–838.
- Chevrot, S.; Sylvander, M. & Delouis, B. (2011) A preliminary catalog of moment tensors for the Pyrenees. *Tectonophysics*, 510, 239–251.
- Chukkapalli, G.; Karpik, S. R. & Ethier, C. R. (1999) A scheme for generating unstructured grids on spheres with application to parallel computation. *Journal of Computational Physics*, 149, 114–127.
- Çınlar, E. (1975) *Introduction to Stochastic Processes*. Prentice Hall, New York, 402 p.
- Ciszak, M.; Calvo, O.; Masoller, C.; Mirasso, C. R. & Toral, R. (2003) Anticipating the response of excitable systems driven by random forcing. *Physical Review Letters*, 90, Art. No. 204102.
- Clopper, C.J. & Pearson, E.S. (1934) The use of confidence or fiducial limits illustrated in the case of the binomial. *Biometrika*, 26, 404–413.
- Cohee, B. P. & Beroza, G. C. (1994) Slip distribution of the 1992 Landers earthquake and its implications for earthquake source mechanics. *Bulletin of the Seismological Society of America*, 84, 692–712.
- Conway, J. H. & Sloane, N. J. A. (1998) *Sphere Packings, Lattices and Groups*, 3rd edn. Springer-Verlag, New York, 703 p.
- Cornell, C.A. (1968) Engineering seismic risk analysis. *Bulletin of the Seismological Society of America*, 58, 1583–1606.
- Corral, Á. (2003) Local distributions and rate fluctuations in a unified scaling law for earthquakes. *Physical Review E*, 68, 035102(R).
- Corral, Á. (2004) Long-term clustering, scaling, and universality in the temporal occurrence of earthquakes. *Physical Review Letters*, 92, 108501.
- Crespo, M.J.; Martínez, F. & Martí, J. (2014) Seismic hazard of the Iberian Peninsula: evaluation with kernel functions. *Natural Hazards and Earth System Sciences*, 14, 1309–1323.
- Cui, J. & Freeden, W. (1997) Equidistribution on the sphere. *SIAM Journal on Scientific Computing*, 18, 595–609.
- Custódio, S.; Batlló, J.; Martins, D.; Antunes, F.; Narciso, J.; Carvalho, S.; Lima, V.; Lopes, F.C.; Ribeiro, P.; Sleeman, R.; Alves, E.I. & Gomes, C.R. (2012) Station COI: Dusting off an old seismic station. *Seismological Research Letters*, 83, 863–869.
- Custódio, S.; Dias, N.A.; Caldeira, B.; Carrilho, F.; Carvalho, S.; Corela, C.; Díaz, J.; Narciso, J.; Madureira, G.; Matias, L.; Haberland, C.; & WILAS Team (2014) Ambient noise recorded by a dense broadband seismic deployment in Western Iberia. *Bulletin of the Seismological Society of America*, 104, 2985–3007.
- Custódio, S.; Dias, N.A.; Carrilho, F.; Góngora, E.; Rio, I.; Marreiros, C.; Morais, I.; Alves, P. & Matias, L. (2015) Earthquakes in western Iberia: improving the understanding of

- lithospheric deformation in a slowly deforming region. *Geophysical Journal International*, 203, 127–145.
- Dahmen, K.; Ertaş, D. & Ben-Zion, Y. (1998) Gutenberg-Richter and characteristic earthquake behavior in simple mean-field models of heterogeneous faults. *Physical Review*, E58, 1494–1501.
- D'Alessandro, A.; Badal, J.; D'Anna, G.; Papanastassiou, D.; Baskoutas, I. & Özel, N.M. (2013) Location performance and detection threshold of the Spanish national seismic network. *Pure and Applied Geophysics*, 170, 1859–1880.
- Daley, D. J. & Vere-Jones, D. (2003) *An Introduction to the Theory of Point Processes*, 2nd ed. Springer-Verlag, New York, USA, 500 p.
- Damelin, S. B. & Grabner, P. J. (2003) Energy functionals, numerical integration and asymptotic equidistribution on the sphere. *Journal of Complexity*, 19, 231–246. [Corrigendum in 20, 883–884.]
- Dañobeitia, J.J. (1980) *Interpretación de la estructura de la corteza en el Archipiélago Canario a partir de perfiles sísmicos profundos de refracción*. Ph.D. Thesis, Universidad Complutense de Madrid, 91 p.
- Darwin, C. (1851) *Geological Observations of South America: Geological Observations on Coral Reefs, Volcanic Islands and on South America-Being the Voyage of the Beagle, Under the Command of Captain Fitzroy, R.N. During the years 1832 to 1836, Part III*. Smith, Elder & Co., London, 279 p.
- Das, S. (2003) Spontaneous complex earthquake rupture propagation. *Pure and Applied Geophysics*, 160, 579–602.
- Das, S. & Aki, K. (1977) Fault plane with barriers: a versatile earthquake model. *Journal of Geophysical Research*, 82, 5658–5670.
- Davidson, J.; Stanchits, S. & Dresen, G. (2007) Scaling and universality in rock fracture. *Physical Review Letters*, 98, 125502.
- Davis, P. M.; Jackson, D. D. and Kagan, Y. Y. (1989) The longer it has been since the last earthquake, the longer the expected time till the next?, *Bulletin of the Seismological Society of America*, 79, 1439–1456.
- Deluca, A. & Corral, Á. (2013) Fitting and goodness-of-fit test of non-truncated and truncated power-law distributions. *Acta Geophysica*, 61, 1351–1394.
- de Miguel y González-Miranda, L. (1961) *Observatorio Central Geofísico "Alfonso Rey Pastor" Toledo*. Memorias del Instituto Geográfico y Catastral, XXIX, Madrid. 19 p. + 17 slides & 2 maps.
- de Parseval, P.; Fontan, F.; Martin, F.; Freeet, J.; Jiang, S.Y. & Wang, R.C. (2004) Geology and ore genesis of the Trimouns talc-chlorite ore deposit, Pyrenees, France. *Acta Petrologica Sinica*, 20, 877–886.
- Del Pin, E.; Carniel, R. & Tárraga, M. (2008) Event recognition by detrended fluctuation analysis: An application to Teide-Pico Viejo volcanic complex, Tenerife, Spain. *Chaos, Solitons & Fractals*, 36, 1173–1180.
- Díaz, J.; Villaseñor, A.; Morales, J.; Pazos, A.; Córdoba, D.; Pulgar, J.; García-Lobón, J.L.; Harnafi, M.; Carbonell, R.; Gallart, J. & Topolberia Seismic Working Group (2010)

- Background noise characteristics at the IberArray broadband seismic network, *Bulletin of the Seismological Society of America*, 100, 618–628.
- Dixon, R. (1987) *Mathographics*. Basil Blackwell, Oxford, England, 224 p.
- Dixon, R. (1989) Spiral phyllotaxis. *Computers and Mathematics with Applications*, 17, 535–538.
- Dixon, R. (1992) Green spirals. In: *Spiral Symmetry*, Hargittai, I. & Pickover, C. A. (Eds.), pp. 353–368. World Scientific, Singapore.
- Dodge, D.A. & Walter, W.R. (2015) Initial global seismic cross-correlation results: Implications for empirical signal detectors. *Bulletin of the Seismological Society of America*, 105, 240–256.
- Domingues, A.; Custódio, S. & Cesca, S. (2013) Waveform inversion of small-to-moderate earthquakes located offshore southwest Iberia. *Geophysical Journal International*, 192, 248–259.
- Domínguez Cerdeña, I.; del Fresno, C. & Rivera, L. (2011) New insight on the increasing seismicity during Tenerife's 2004 volcanic reactivation. *Journal of Volcanology and Geothermal Research*, 206, 15–29.
- Domínguez Cerdeña, I.; del Fresno, C. & Gomis Moreno, A. (2014) Seismicity patterns prior to the 2011 El Hierro eruption. *Bulletin of the Seismological Society of America*, 104, 567–575.
- Douady, S. & Couder, Y. (1992) Phyllotaxis as a physical self-organized growth process. *Physical Review Letters* 68, 2098–2101.
- Dowla, F.U. (1996) *Calibration of the Sonseca Array with Large Magnitude Regional and Teleseismic Events*. Report UCRL-ID-125023. Lawrence Livermore National Laboratory, Livermore, California, USA, 37 p.
- Dowrick, D.J. & Rhoades, D.A. (2004) Relations between earthquake magnitude and fault rupture dimensions: How regionally variable are they? *Bulletin of the Seismological Society of America*, 94, 776–788.
- Dreger, D.S. & Helmberger, D.V. (1993) Determination of source parameters at regional distances with three-component sparse network data. *Journal of Geophysical Research*, 98, 8107–8125.
- Drossel, B. & Schwabl, F. (1992) Self-organized criticality in a forest-fire model. *Physica A*, 191, 47–50.
- Durá-Gómez, I. & Talwani, P. (2010) Reservoir-induced seismicity associated with the Itoiz Reservoir, Spain: a case study, *Geophysical Journal International*, 181, 343–356.
- Durrett, R. (1999) *Essentials of Stochastic Processes*. Springer Verlag, Berlin, Germany. 289 p.
- Dziewonski, A.M.; Chou T.-A. & Woodhouse J.H. (1981) Determination of earthquake source parameters from waveform data for studies of global and regional seismicity. *Journal of Geophysical Research*, 86, 2825–2852.
- Earle, M. A. (2006) Sphere to spheroid comparisons. *The Journal of Navigation* 59 (3), 491–496.
- Ebel, J.E. (2008) The importance of small earthquakes. *Seismological Research Letters*, 79, 491–493.
- Eberhard, D.A.J.; Zechar, J.D. & Wiemer, S. (2012) A prospective earthquake forecast experiment in the western Pacific. *Geophysical Journal International*, 190, 1579–1592.
- Efron, B. (1979) Bootstrap methods: Another look at the jackknife, *The Annals of Statistics*, 7, 1–26.

- Ekström, G. & Nettles, M. (2015) *Global CMT Web Page*, www.globalcmt.org [Last accessed: October 2015.]
- Ekström, G.; Nettles, M. & Dziewonski, A.M. (2012) The global CMT project 2004–2010: Centroid-moment tensors for 13,017 earthquakes. *Physics of the Earth and Planetary Interiors*, 200–201, 1–9.
- Ellsworth, W. L.; Matthews, M. V.; Nadeau, R. M.; Nishenko, S. P.; Reasenber, P. A. & Simpson, R. W. (1999) *A physically-based earthquake recurrence model for estimation of long-term earthquake probabilities*. United States Geological Survey Open-File Report 99–552, 22 p.
- Eneva, M. (1994) Monofractal or multifractal: a case study of spatial distribution of mining-induced seismic activity. *Nonlinear Processes in Geophysics*, 1, 182–190.
- Erickson, R. O. (1973) Tubular packing of spheres in biological fine structure. *Science*, 181, 705–716.
- Euroseismos Project (2006) *Index of bulletins of Spain*, Istituto Nazionale di Geofisica e Vulcanologia, storing.ingv.it/bulletins/Spain/index.html. [Last accessed: March 2015.]
- Evans, D. G. & Jones, S. M. (1987) Detecting Voronoi (area-of-influence) polygons. *Mathematical Geology*, 19, 523–537.
- Farr, T.G. *et al.* (2007) The Shuttle Radar Topography Mission, *Reviews of Geophysics*, 45, RG2004.
- Felzer, K.R.; Abercrombie, R.E. & Ekström, G. (2003) Secondary aftershocks and their importance for aftershock forecasting. *Bulletin of the Seismological Society of America*, 93, 1433–1448.
- Feng, S.; Ochieng, W. Y. & Mautz, R. (2006) An area computation based method for RAIM holes assessment. *Journal of Global Positioning Systems*, 5, 11–16.
- Field, E.H. (2007) Overview of the working group for the development of Regional Earthquake Likelihood Models (RELM). *Seismological Research Letters*, 78, 7–16.
- Field, E.H. (2015) Computing elastic-rebound-motivated earthquake probabilities in unsegmented fault models: A new methodology supported by physics-based simulators. *Bulletin of the Seismological Society of America*, 105, 544–559.
- Field, E.H. *et al.* (2015a) Long-term time-dependent probabilities for the third Uniform California Earthquake Rupture Forecast (UCERF3). *Bulletin of the Seismological Society of America*, 105, 511–543.
- Field, E.H. *et al.* (2015b) *UCERF3: A New Earthquake Forecast for California's Complex Fault System*. United States Geological Survey Fact Sheet 2015–3009, 6 p. [dx.doi.org/10.3133/fs20153009](https://doi.org/10.3133/fs20153009) [Last accessed: November 2015.]
- Fitzenz, D.D. & Miller, S.A. (2004) New insights on stress rotations from a forward regional model of the San Andreas fault system near its Big Bend in southern California. *Journal of Geophysical Research*, 109, B08404.
- FitzRoy, R. (1839) *Proceedings of the Second Expedition, 1831–1836, under the Command of Captain Robert FitzRoy. Volume II of Narrative of the Surveying Voyages of His Majesty's Ships Adventure and Beagle between the Years 1826 and 1836. Describing their Examination*

- of the Southern Shores of South America, and the Beagle's Circumnavigation of the Globe*, Henry Colburn, London, UK, 695 p.
- Fliegel, H.F. & van Flandern, T.C. (1968) A machine algorithm for processing calendar dates. *Communications of the ACM*, 11, 657.
- Fontserè, E. & Iglésies, J. (1971) *Recopilació de dades sísmiques de les terres catalanes entre 1100 i 1906*. Fundació Salvador Vives Casajuana, Barcelona, Spain. 546 p. [In Catalan.]
- Fowler, D. R.; Prusinkiewicz, P. & Battjes, J. (1992) A collision-based model of spiral phyllotaxis. *ACM SIGGRAPH Computer Graphics*, 26, 361–368.
- Fréchet, J. & Rivera, L. (2012) Horizontal pendulum development and the legacy of Ernst von Rebeur-Paschwitz. *Journal of Seismology*, 16, 315–343.
- Free Software Foundation (2007) *GNU General Public License Version 3, 29 June 2007*. www.gnu.org/licenses/gpl-3.0.en.html [Last accessed: November 2015.]
- Freed, A.M. (2005) Earthquake triggering by static, dynamic, and postseismic stress transfer. *Annual Review of Earth and Planetary Sciences*, 33, 335–367.
- Frohlich, C. (2006) *Deep Earthquakes*, Cambridge University Press, Cambridge, UK, 573 p.
- Fuselier, E.; Hangelbroek, T.; Narcowich, F.J.; Ward, J.D. & Wright, G.B. (2014) Kernel based quadrature on spheres and other homogeneous spaces *Numerische Mathematik*, 127, 57–92.
- Galbis, J. (1932) *Catálogo sísmico de la zona comprendida entre los meridianos 5° E . y 20° W. de Greenwich y los paralelos 45° y 25° N.; Tomo I*. Dirección General del Instituto Geográfico, Catastral y de Estadística, Madrid, Spain, 807 p.
- Galbis, J. (1940) *Catálogo sísmico de la zona comprendida entre los meridianos 5° E . y 20° W. de Greenwich y los paralelos 45° y 25° N.; Tomo II*. Instituto Geográfico y Catastral, Madrid, Spain, 277 p.
- Galve, J.P.; Remondo, J. & Gutiérrez, F. (2011) Improving sinkhole hazard models incorporating magnitude–frequency relationships and nearest neighbor analysis. *Geomorphology*, 134, 157–170.
- García, A.; Fernández-Ros, A.; Berrocoso, M.; Marrero, J.M.; Prates, G.; De la Cruz-Reyna, S. & Ortiz, R. (2014) Magma displacements under insular volcanic fields, applications to eruption forecasting: El Hierro, Canary Islands, 2011–2013. *Geophysical Journal International*, 197, 322–334.
- García Fernández, M. (1986) Relative location accuracy of the Spanish seismic network for earthquakes in the Iberian region. *Revista de Geofísica* 42, 3–8.
- García Fernández, M.; Kijko, A. & López Arroyo, A. (1987) The new Spanish seismic network: Optimization and location capability. In: *European Seismological Commission: Activity Reports 1984-1986 and Proceedings of the XX General Assembly 1986 in Kiel*, Mayer-Rosa, D.; van Gils, J.M. & Stiller, H. (Eds.), pp. 88–102. Publication series of the Swiss Seismological Service, No. 101. Federal Institute of Technology, Zürich, Switzerland.
- García Fernández, M.; Kijko, A.; Carracedo, J.C. & Soler, V. (1988) Optimum station distribution to monitor seismic activity of Teide volcano, Tenerife, Canary Islands, *Journal of Volcanology and Geothermal Research*, 35, 195–204.
- García Maneiro, C.H.; Maqueda Jiménez, S.; Pazos Pérez, L. & Carnero Ortiz, F. (2012) Voladuras de cámara de la mina de Aguas Teñidas, *Industria y Minería*, 390, 25–31.

- García-Mayordomo, J.; Faccioli, E. & Paolucci, R. (2004) Comparative study of the seismic hazard assessments in European national seismic codes. *Bulletin of Earthquake Engineering*, 2, 51–73.
- García-Mayordomo, J. *et al.* (2012) The Quaternary Active Faults Database of Iberia (QAFI v.2.0), *Journal of Iberian Geology*, 38, 285–302.
- García Siñeriz, J.; 1935, *Discurso leído en el acto de su recepción por el Sr. D. José García Siñeriz y contestación del Excmo. Señor D. Enrique Hauser y Neuburger el día 9 de enero de 1935*. Academia de Ciencias Exactas, Físicas y Naturales. C. Bermejo, Impresor, Madrid, 37 p.
- García Vargas, M.; Rueda, J.; García Blanco, R.M. & Mezcua, J. (2016) A real-time discrimination system of earthquakes and explosions for the mainland Spanish Seismic Network. *Pure and Applied Geophysics*, in press.
- Gaspar-Escribano, J.M.; Rivas-Medina, A.; Parra, H.; Cabañas, L.; Benito, B.; Ruiz Barajas, S. & Martínez Solares, J.M. (2015) Uncertainty assessment for the seismic hazard map of Spain. *Engineering Geology*, 199, 62–73.
- Geiger, L. (1910) Herdbestimmung bei Erdbeben aus den Ankunftszeiten, *Nachrichten von der Gesellschaft der Wissenschaften zu Göttingen, Mathematisch-Physikalische Klasse*, 4, 331–349.
- Geiger, L. (1912) Probability method for the determination of earthquake epicenters from the arrival time only, *Bulletin of St. Louis University*, 8, 56–71. [English translation of the 1910 article.]
- Geissler, W.H.; Matias, L.; Stich, D.; Carrilho, F.; Jokat, W.; Monna, S.; IbenBrahim, A.; Mancilla, F.; Gutscher, M.-A.; Sallarès, V. & Zitellini, N. (2010) Focal mechanisms for sub-crustal earthquakes in the Gulf of Cadiz from a dense OBS deployment. *Geophysical Research Letters*, 37, L18309.
- Geller, R.J. (2011) Shake-up time for Japanese seismology. *Nature* 472, 407–409.
- Gentili, S.; Sukan, M.; Peruzza, L. & Schorlemmer, D. (2011) Probabilistic completeness assessment of the past 30 years of seismic monitoring in northeastern Italy. *Physics of the Earth and Planetary Interiors*, 186, 81–96.
- Gerstenberger, M.C. & Rhoades D.A. (2010) New Zealand earthquake forecast testing centre. *Pure and Applied Geophysics*, 167, 877–892.
- Giardini, D.; Wiemer, S.; Fäh, D. & Deichmann, D. (2004) *Seismic hazard assessment of Switzerland, 2004*. Report, Swiss Seismological Service, ETH Zurich, 88 p.
- Gilbert, G. K. (1884) A theory of the earthquakes of the Great Basin, with a practical application. *Salt Lake City Tribune*, Sept. 30, 1883, reprinted in *American Journal of Science Series 3*, 27, 49–54.
- Gilbert, G. K. (1909) Earthquake forecasts. *Science*, 29, 121–138.
- Godey, S.; Bossu, R.; Guilbert, J. & Mazet-Roux, G. (2006) The Euro-Mediterranean Bulletin: A comprehensive seismological bulletin at regional scale. *Seismological Research Letters*, 77, 460–474.
- Godey, S.; Bossu, R. & Guilbert, J. (2013) Improving the Mediterranean seismicity picture thanks to international collaborations. *Physics of the Earth and Planetary Interiors*, 63, 3–11.
- Goltz, C. (1997) *Fractal and Chaotic Properties of Earthquakes*. Springer-Verlag, Berlin, Germany, 178 p.

- Gómez, J. B. & Pacheco, A. F. (2004) The minimalist model of characteristic earthquakes as a useful tool for description of the recurrence of large earthquakes. *Bulletin of the Seismological Society of America*, 94, 1960–1967.
- Gomez, B.; Corral, Á.; Orpin, A.R.; Page, M.J.; Pouderoux, H. & Upton, F. (2014) Lake Tutira paleoseismic record confirms random, moderate to major and/or great Hawkešs Bay (New Zealand) earthquakes. *Geology*, 43, 103–106.
- González, Á. (2009) Mapa de localizaciones probables de futuros terremotos en la Península Ibérica, Baleares y Canarias. *Seguridad y Medio Ambiente* [Fundación MAPFRE, Madrid, Spain], 114, 44–54.
- González, Á. (2010a) Measurement of areas on a sphere using Fibonacci and latitude–longitude lattices. *Mathematical Geosciences*, 42, 49–64.
- González, Á. (2010b): «Nearest»: An empirical, non-parametric, forecasting model based on nearest-neighbour distances between earthquakes. In: *Abstract Book. European Seismological Commission 32nd General Assembly*, p. 71. Montpellier, France, 6–10 September 2010.
- González, Á.; Vázquez-Prada, M.; Gómez, J. B. & Pacheco, A. F. (2004) Using synchronization to improve the forecasting of large relaxations in a cellular-automaton model. *Europhysics Letters*, 68, 611–617.
- González, Á.; Gómez, J. B. & Pacheco, A. F. (2005) The occupation of a box as a toy model for the seismic cycle of a fault, *American Journal of Physics*, 73, 946–952. [Erratum: (2007) 75, 286.]
- González, Á.; Gómez, J. B. & Pacheco, A. F. (2006a) Updating seismic hazard at Parkfield. *Journal of Seismology*, 10, 131–135.
- González, Á.; Vázquez-Prada, M.; Gómez, J. B. & Pacheco, A. F. (2006b) A way to synchronize models with seismic faults for earthquake forecasting: Insights from a simple stochastic model. *Tectonophysics*, 424, 319–334.
- González de Vallejo, L.I.; García-Mayordomo, J. & Insua, J.M. (2006) Probabilistic seismic-hazard assessment of the Canary Islands. *Bulletin of the Seismological Society of America*, 96, 2040–2049.
- Gradshteyn, I.S. & Ryzhik, I.M. (1965) *Table of integrals, series, and products*. 4th Ed. Academic Press, New York. 1086 p.
- Grant, L.B.; Gould, M.M.; 2004. Assimilation of paleoseismic data for earthquake simulation. *Pure and Applied Geophysics*, 161, 2295–2306.
- Gregory, M. J.; Kimerling, A. J.; White, D. & Sahr, K. (2008) A comparison of intercell metrics on discrete global grid systems. *Computers, Environment and Urban Systems*, 32, 188–203.
- Greiner, B. (1999) Euler rotations in plate-tectonic reconstructions. *Computers and Geosciences*, 25, 209–216.
- Griesbach, C. L. (1893) Notes on the Earthquake in Balûchistân on the 20th December 1892. *Records of the Geological Survey of India*, 26, 57–61.
- Grünthal, G. & Wahlström, R. (2012) The European-Mediterranean Earthquake Catalogue (EMEC) for the last millennium. *Journal of Seismology*, 16, 535–570.
- Gulia, L. (2010) Detection of quarry and mine blast contamination in European regional catalogues. *Natural Hazards*, 53, 229–249.

- Gulia L.; Wiemer, S. & Wyss, M. (2012) Catalog artifacts and quality control. *Community Online Resource for Statistical Seismicity Analysis*. doi:10.5078/corssa-93722864. Available at www.corssa.org. [Last accessed October 2015.]
- Gundersen, H. J. G.; Jensen, E. B. V.; Kiêu, K. & Nielsen, J. (1999) The efficiency of systematic sampling in stereology – reconsidered. *Journal of Microscopy*, 193, 199–211.
- Gutenberg, B. & Richter, C. F. (1944) Frequency of earthquakes in California. *Bulletin of the Seismological Society of America*, 34, 185–188.
- Gutenberg, B. & Richter, C. F. (1954) *Seismicity of the Earth and Associated Phenomena*, 2nd edn. Princeton University Press, Princeton, New Jersey (USA), 310 p.
- Gutiérrez, F.; Masana, E.; González, Á.; Guerrero, J., Lucha, P. y McCalpin, J.P. (2009) Late Quaternary paleoseismic evidence on the Munébrega half-graben fault (Iberian Range, Spain). *International Journal of Earth Sciences*, 98, 1691–1703.
- Habermann, R.E. (1987) Man-made changes of seismicity rates. *Bulletin of the Seismological Society of America*, 77, 141–159
- Habermann, R.E. (1991) Seismicity rate variations and systematic changes in magnitudes in teleseismic catalogs. *Tectonophysics*, 193, 277–289.
- Haeussler, P. J.; Schwartz, D. P.; Dawson, T. E.; Stenner, H. D.; Lienkaemper, J. J.; Sherrod, B.; Cinti, F. R.; Montone, P.; Craw, P. A.; Crone, A. J.; Personius, S. F. (2004) Surface rupture and slip distribution of the Denali and Totschunda faults in the 3 November 2002 M 7.9 earthquake, Alaska. *Bulletin of the Seismological Society of America*, 94, S23–S52, Part B, Suppl. S.
- Häge, M. & Joswig, M. (2009a) Microseismic feasibility study: detection of small magnitude events ($M_L < 0.0$) for mapping active faults in the Betic Cordillera (Spain). *Annals of Geophysics*, 52, 117–126.
- Häge, M. & Joswig, M. (2009b) Mapping local microseismicity using short-term tripartite small array installations – Case study: Coy region (SE Spain), *Tectonophysics*, 471, 225–231.
- Hagiwara, Y. (1974) Probability of earthquake occurrence as obtained from a Weibull distribution analysis of crustal strain. *Tectonophysics*, 23, 313–318.
- Hainzl, S. & Zöller, G. (2001) The role of disorder and stress concentration in nonconservative fault systems. *Physica A*, 294, 67–84.
- Hainzl, S.; Zöller, G.; & Kurths, J. (1999) Self-organized criticality model for earthquakes: Quiescence, foreshocks and aftershocks. *International Journal of Bifurcation and Chaos*, 9, 2249–2255.
- Hainzl, S.; Zöller, G.; Kurths, J. & Zschau, J. (2000) Seismic quiescence as an indicator for large earthquakes in a system of self-organized criticality. *Geophysical Research Letters*, 27, 597–600.
- Hamdache, M.; Peláez, J.A.; Talbi, A. & López Casado, C. (2010) A unified catalog of main earthquakes for Northern Algeria from A.D. 856 to 2008. *Seismological Research Letters*, 81, 732–739.
- Hanks, T.C. & Kanamori, H. (1979). A moment magnitude scale. *Journal of Geophysical Research*, 84, 2348–2350.
- Hannay, J. H. & Nye, J. F. (2004) Fibonacci numerical integration on a sphere. *Journal of Physics A: Mathematical and General*, 37, 11591–11601.

- Harbi, A.; Sebaï, A.; Benmedjber, M.; Ousadou, F.; Rouchiche, Y.; Grigahcene, A.; Aïni, D.; Bourouis, S.; Maouche, S. & Ayadi, A. (2015) The Algerian homogenized macroseismic database (267–1989): A deeper insight into the Algerian historical seismicity. *Seismological Research Letters*, 54, 725–747.
- Hardebeck, J.L.; Felzer, K.R. & Michael, A.J. (2008) Improved tests reveal that the accelerating moment release hypothesis is statistically insignificant. *Journal of Geophysical Research*, 113, B08310.
- Harris, R. A. (2000) Earthquake stress triggers, stress shadows, and seismic hazard. *Current Science*, 79, 1215–1225.
- Harte, D. (2001) *Multifractals: Theory and Applications*. Chapman and Hall/CRC, Boca Raton, Florida, USA, 264 p.
- Hashimoto, M. (2001) Complexity in the recurrence of large earthquakes in southwestern Japan: A simulation with an interacting fault system model. *Earth, Planets and Space*, 53, 249–259.
- Hasting, M. & Barrientos, S. (2002) The upgrade of the International Monitoring System primary station at Sonseca, Spain, (PS-40): A before and after comparison of the short-period array. In: *Proceedings of the 24th Seismic Research Review – Nuclear Explosion Monitoring: Innovation and Integration*, pp. 927–939, Ponte Vedra Beach, Florida (USA).
- Hauksson, E.; Stock, J.; Hutton, K.; Yang, W.; Vidal-Villegas, J.A. & Kanamori, H. (2011) The 2010 M_w 7.2 El Mayor-Cucapah earthquake sequence, Baja California, Mexico and southernmost California, USA: Active seismotectonics along the Mexican Pacific margin. *Pure and Applied Geophysics*, 168, 1255–1277.
- Hauksson, E.; Yang, W. & Shearer, P.M. (2012) Waveform relocated earthquake catalog for Southern California (1981 to June 2011). *Bulletin of the Seismological Society of America*, 102, 2239–2244.
- Havskov, J. & Ottemöller, L. (2010) *Routine Data Processing in Earthquake Seismology*. Springer, Dordrecht (Netherlands), 347 p.
- Heidbach, O.; Tingay, M.; Barth, A.; Reinecker, J.; Kurfeß, D. & Müller, B. (2008) The World Stress Map database release 2008. doi:10.1594/GFZ.WSM.Rel2008
- Heimann, S.; González, Á.; Wang, R.; Cesca, S. & Dahm, T. (2013) Seismic characterization of the Chelyabinsk meteor's terminal explosion. *Seismological Research Letters*, 84, 1021–1025.
- Heimpel, M.H. (2003) Characteristic scales of earthquake rupture from numerical models. *Nonlinear Processes in Geophysics*, 10, 573–584.
- Helmstetter, A. & Werner, M.J. (2014) Adaptive smoothing of seismicity in time, space, and magnitude for time-dependent earthquake forecasts for California. *Bulletin of the Seismological Society of America*, 104, 809–822.
- Helmstetter, A.; Kagan, Y.Y. & Jackson, D.D. (2005) Importance of small earthquakes for stress transfers and earthquake triggering. *Journal of Geophysical Research*, 110, B05S08.
- Henley, C.L. (1993) Statics of a “self-organized” percolation model. *Physical Review Letters*, 71, 2741–2744.
- Herraiz, M. & Lázaro, M. (1991) Microearthquake distribution in the Granada region. In: *Seismicity, Seismotectonics and Seismic Risk of the Ibero-Maghrebian Region*, Mezcuá, J. & Udías, A. (Eds.), pp. 65–77. Monografía 8, Instituto Geográfico Nacional, Madrid, Spain.

- Hickman, S. & Zoback, M. (2004) Stress orientations and magnitudes in the SAFOD pilot hole. *Geophysical Research Letters*, 31, L15S12.
- Hillers, G. & Ben-Zion, Y. (2011) Seasonal variations of observed noise amplitudes at 2–18 Hz in southern California. *Geophysical Journal International*, 184, 860–868.
- Holliday, J.R.; Nanjo, K.Z.; Tiampo, K.F.; Rundle, J.B. & Turcotte, D.L. (2005) Earthquake forecasting and its verification. *Nonlinear Processes in Geophysics*, 12, 965–977.
- Hough, S.E. (2013) Missing great earthquakes, *Journal of Geophysical Research*, 118, 1098–1108.
- Houghton, J. (2002) *The Physics of Atmospheres*. Cambridge University Press, Cambridge, UK, 320 p.
- Howarth, R. J. (1998) Improved estimators of uncertainty in proportions, point-counting, and pass-fail test results. *American Journal of Science*, 298, 594–607.
- Huang, J. & Turcotte, D.L. (1990) Evidence for chaotic fault interactions in the seismicity of the San Andreas fault and Nankai trough. *Nature*, 348, 234–236.
- Husen, S. & Hardebeck, J.L. (2010) Earthquake location accuracy. *Community Online Resource for Statistical Seismicity Analysis*, doi:10.5078/corssa-55815573. Available at www.corssa.org. [Last accessed October 2015.]
- Hüttig, C. & Stemmer, K. (2008) The spiral grid: a new approach to discretize the sphere and its application to mantle convection. *Geochemistry, Geophysics, Geosystems* 9, Q02018.
- Hutton, K.; Woessner, J. & Hauksson, E. (2010) Earthquake monitoring in Southern California for seventy-seven years (1932–2008), *Bulletin of the Seismological Society of America*, 100, 423–446.
- Huxley, M. N. (1987) The area within a curve. *Proceedings of the Indian Academy of Sciences (Mathematical Sciences)*, 97, 111–116.
- Huxley, M. N. (2003) Exponential sums and lattice points III. *Proceedings of the London Mathematical Society*, 87, 591–609.
- IAG – Instituto Andaluz de Geofísica (2015) *Red Sísmica de Andalucía*, iagpds.ugr.es/pages/red_sismica [Last checked: October 2015.]
- Ibáñez, J.M.; De Angelis, S.; Díaz-Moreno, A.; Hernández, P.; Alguacil, G.; Posadas, A. & Pérez, N. (2012) Insights into the 2011-2012 submarine eruption off the coast of El Hierro (Canary Islands, Spain) from statistical analyses of earthquake activity. *Geophysical Journal International*, 191, 659–670.
- ICGC – Institut Cartogràfic i Geològic de Catalunya (2015) *Seismic Network of Catalonia*. www.igc.cat/web/en/sismologia_xsis.html [Last accessed: October 2015]
- Ikeda, R.; Iio, Y. & Omura, K. (2001) In situ stress measurements in NIED boreholes in and around the fault zone near the 1995 Hyogo-ken Nanbu earthquake, Japan. *The Island Arc*, 10, 252–260.
- Illes-Toth, E.; Ramos, M.R.; Cappai, R.; Dalton, C. & Smith, D.P. (2015) Distinct higher-order alpha-synuclein oligomers induce intracellular aggregation. *Biochemical Journal*, 468, 485–493.
- Imoto, M. (2004) Probability gains expected for renewal process models. *Earth, Planets and Space*, 56, 563–571.

- Instituto Geográfico Nacional (1991) Spanish National Seismic Network. In: *Seismicity, Seismotectonics and Seismic Risk of the Ibero-Maghrebian Region*, Mezcuá, J. & Udías, A. (Eds.), pp. 3–15, Monografía 8, Instituto Geográfico Nacional, Madrid, Spain.
- Instituto Geográfico Nacional (2002a) Estación Sismológica de Sonseca. Homologada por el Sistema Internacional de Vigilancia Sísmica. *Boletín Informativo IGN*, 10, 1–2.
- Instituto Geográfico Nacional (2002b) Red Sísmica Digital Española-VSAT. Finalizada la nueva red sísmica vía satélite. *Boletín Informativo IGN*, 12, 1–2.
- Instituto Geográfico Nacional (2006) Desarrollo de una unidad móvil con instrumentación sísmica. *Boletín Informativo IGN*, 28, 2.
- Instituto Geográfico Nacional (2015a) *Servicio de Información Sísmica*.
www.ign.es/ign/layout/sismo.do [Last accessed October 2015].
- Instituto Geográfico Nacional (2015b) *Vigilancia Volcánica*.
www.ign.es/ign/layout/volcaVolcanologia.do [Last accessed October 2015].
- International Seismological Centre (2015a) *International Registry of Seismograph Stations* www.isc.ac.uk/registries/, Internatl. Seis. Cent.; Thatcham, United Kingdom. [Last accessed October 2015.]
- International Seismological Centre (2015b) *On-line Bulletin*, www.isc.ac.uk, Internatl. Seis. Cent.; Thatcham, United Kingdom. [Last accessed November 2015.]
- Ishimoto, M. & Iida, K. (1939) Observations sur les séismes enregistrés par le microséismographe construit dernièrement (I). *Bulletin of the Earthquake Research Institute, University of Tokyo*, 17, 443–478. [In Japanese, with French abstract.]
- ISO - International Organization for Standardization (2004) *International Standard ISO 6601. Data elements and interchange formats – Information interchange – Representation of dates and times*. 3rd ed.; International Organization for Standardization, Geneva, Switzerland, 33 p.
- Iwata, T. (2008) Low detection capability of global earthquakes after the occurrence of large earthquakes: investigation of the Harvard CMT catalogue. *Geophysical Journal International*, 174, 849–856.
- Iwata, T. (2013) Estimation of completeness magnitude considering daily variation in earthquake detection capability. *Geophysical Journal International*, 194, 1909–1919.
- Jarái, A.; Kozák, M. & Rózsa, P. (1997) Comparison of the methods of rock-microscopic grain-size determination and quantitative analysis. *Mathematical Geology*, 29, 977–991.
- Jean, R. V. (1994) *Phyllotaxis: A Systemic Study of Plant Pattern Morphogenesis*. Cambridge University Press, Cambridge, UK, 400 p.
- Jiménez, A. (2013) Cellular automata to describe seismicity: A review. *Acta Geophysica*, 61, 1325–1350.
- Jiménez, M.J. & García-Fernández, M. (1996) Aftershock sequence of the 9 May 1989 Canary Islands earthquake, *Tectonophysics*, 255, 157–162.
- Johnson, N.L. & Kotz, S. (1977) *Urn Models and Their Application: An Approach to Modern Discrete Probability Theory*, John Wiley & Sons, New York, USA, 402 p.
- Jordan, T. H.; Beroza, G.; Cornell, C. A.; Crouse, C. B.; Dieterich, J.; Frankel, A.; Jackson, D. D.; Johnston, A.; Kanamori, H.; Langer, J. S.; McNutt, M. K.; Rice, J. R.; Romanowicz, B. A.; Sieh, K. & Somerville, P. G. (2003) *Living on an Active Earth. Perspectives on Earthquake Science*. National Academies Press, Washington, DC (USA), 418 p.

- Jordan, T.H.; Chen, Y.T.; Gasparini, P.; Madariaga, R.; Main, I.; Marzocchi, W.; Papadopoulos, G.; Sobolev, G.; Yamaoka, K. & Zschau, J. (2011) Operational earthquake forecasting. State of knowledge and guidelines for utilization. *Annals of Geophysics*, 54, Article 5350, 77 p.
- Jordan, T.H.; Marzocchi, W.; Michael, A.J. & Gerstenberger, M.C. (2014) Operational earthquake forecasting can enhance earthquake preparedness. *Seismological Research Letters*, 85, 955–959.
- Kafka, A. L. (2007) Does seismicity delineate zones where future large earthquakes are likely to occur in intraplate environments? In: *Continental Intraplate Earthquakes: Science, Hazard, and Policy Issues*, Stein, S. & Mazzotti, S. (Eds.), pp. 35–48. Geological Society of America Special Paper 425, Boulder, Colorado.
- Kafka, A.L. & Ebel, J.E. (2011): Proximity to past earthquakes as a least-astonishing hypothesis for forecasting locations of future earthquakes. *Bulletin of the Seismological Society of America*, 101, 1618–1629.
- Kagan, Y.Y. (2003) Accuracy of modern global earthquake catalogs, *Physics of the Earth and Planetary Interiors*, 135, 173–209.
- Kagan, Y.Y. (2004) Short-term properties of earthquake catalogs and models of earthquake source. *Bulletin of the Seismological Society of America*, 94, 1207–1228.
- Kagan, Y.Y. (2007) Earthquake spatial distribution: the correlation dimension. *Geophysical Journal International*, 168, 1175–1194.
- Kagan, Y.Y. (2014) *Earthquakes: Models, Statistics, Testable Forecasts*. John Wiley & Sons, Oxford, UK, 283 p.
- Kagan, Y.Y. & Jackson, D.D. (1991) Long-term earthquake clustering. *Geophysical Journal International*, 104, 117–133.
- Kagan, Y.Y. & Jackson, D.D. (1994) Long-term probabilistic forecasting of earthquakes. *Journal of Geophysical Research*, 99, 13,685–13,700.
- Kagan, Y.Y. & Jackson, D.D. (1999) Worldwide doublets of large shallow earthquakes, *Bulletin of the Seismological Society of America*, 89, 1147–1155.
- Kagan, Y.Y. & Knopoff, L. (1980) Spatial distribution of earthquakes: the two-point correlation function. *Geophysical Journal of the Royal Astronomical Society*, 62, 303–320.
- Kagan, Y.Y. & Knopoff, L. (1981) Stochastic synthesis of earthquake catalogs. *Journal of Geophysical Research*, 86, 2853–2862.
- Kagan, Y.Y. & Knopoff, L. (1987) Random stress and earthquake statistics: time dependence, *Geophysical Journal of the Royal Astronomical Society*, 88, 723–731.
- Kamer, Y.; Ouillon, G.; Sornette, D. & Wössner, J. (2015) Condensation of earthquake location distributions: Optimal spatial information encoding and application to multifractal analysis of south Californian seismicity. *Physical Review E*, 92, 022808.
- Kanamori, H. & Brodsky, E. E. (2004) The physics of earthquakes. *Reports on Progress in Physics*, 67, 1429–1496.
- Kanamori, H. & Stewart, G. S. (1978) Seismological aspects of Guatemala earthquake on february 4, 1976. *Journal of Geophysical Research*, 83, 3427–3434.
- Kantsiper, B. & Weiss, S. (1998) An analytic approach to calculating Earth coverage. *Advances in the Astronautical Sciences*, 97, 313–332.

- Kato, N. & Seno, T. (2003) Hypocenter depths of large interplate earthquakes and their relation to seismic coupling. *Earth and Planetary Science Letters*, 210, 53–63.
- Keilis-Borok, V.I. (2002) Earthquake prediction: State-of-the-art and emerging possibilities. *Annual Review of Earth and Planetary Sciences*, 30, 1–33.
- Keilis-Borok, V.I. & Soloviev, A.A. (Eds.; 2003) *Nonlinear Dynamics of the Lithosphere and Earthquake Prediction*. Springer-Verlag, Berlin, Germany, 337 p.
- Keinert, B.; Innmann, M.; Sanger, M. & Stamminger, M. (2015) Spherical Fibonacci Mapping. *ACM Transactions on Graphics*, 34, 193.
- Kendall, D. G. (1948) On the number of lattice points inside a random oval. *The Quarterly Journal of Mathematics (Oxford)*, 19, 1–26.
- Kilb, D.; Martynov, V.G. & Vernon, F.L. (2007) Aftershock detection thresholds as a function of time: Results from the ANZA Seismic Network following the 31 October 2001 M_L 5.1 Anza, California, Earthquake. *Bulletin of the Seismological Society of America*, 97, 780–792.
- Kimerling, A. J. (1984) Area computation from geodetic coordinates on the spheroid. *Surveying and Mapping*, 44, 343–351.
- Kiremidjian, A. S. & Anagnos, T. (1984) Stochastic slip-predictable model for earthquake occurrences. *Bulletin of the Seismological Society of America*, 74, 739–755.
- Kisslinger, C. (1996) Aftershocks and fault-zone properties. *Advances in Geophysics*, 38, 1–36.
- Klíma, K.; Pick, M. & Pros, Z. (1981) On the problem of equal area block on a sphere. *Studia Geophysica et Geodaetica (Praha)*, 25, 24–35.
- Knuth, D. E. (1997) *Art of Computer Programming, Volume 1: Fundamental Algorithms*, 3rd edn. Addison-Wesley, Reading, Massachusetts, 672 p.
- Kossobokov, V. & Shebalin, P. (2003) Earthquake prediction. In: *Nonlinear Dynamics of the Lithosphere and Earthquake Prediction*, Keilis-Borok, V. I. & Soloviev, A. A. (Eds.), pp. 141–207 [references in pp. 311–332]. Springer, Berlin.
- Kossobokov, V.G.; Romashkova, L.L.; Keilis-Borok, V.I. & Healy, J.H. (1999) Testing earthquake prediction algorithms: Statistically significant advance prediction of the largest earthquakes in the Circum-Pacific, 1992-1997. *Physics of the Earth and Planetary Interiors*, 111, 187–196.
- Koto, B. (1893) On the cause of the great earthquake in central Japan, 1891. *Journal of the College of Science, Imperial University of Tokyo*, 5, 296–353.
- Kozin, M. B.; Volkov, V. V. & Svergun, D. I. (1997) ASSA, a program for three-dimensional rendering in solution scattering from biopolymers. *Journal of Applied Crystallography*, 30, 811–815.
- Kuhlemeier, C. (2007) Phyllotaxis. *Trends in Plant Science*, 12, 143–150.
- Kuroki, H.; Ito, H.M.; Takayama, H. & Yoshida, A. (2004) 3-D simulation of the occurrence of slow slip events in the Tokai region with a rate- and state-dependent friction law. *Bulletin of the Seismological Society of America*, 94, 2037–2050.
- Kværna T.; & Ringdal, F. (2013) Detection capability of the seismic network of the International Monitoring System for the Comprehensive Nuclear-Test-Ban Treaty. *Bulletin of the Seismological Society of America*, 103, 759–772.
- Kwiatk, G.; Plenkers, K.; Nakatani, M.; Yabe, Y.; Dresen, G. & JAGUARS-Group (2010) Frequency-magnitude characteristics down to magnitude -4.4 for induced seismicity

- recorded at Mponeng gold mine, South Africa, *Bulletin of the Seismological Society of America*, 100, 1165–1173.
- Lahr, J.C. & Van Eck, T.; 2003 (Eds.). Global inventory of seismographic networks. In: *International Handbook of Earthquake and Engineering Seismology*, Lee, W.H.K.; Kanamori, H.; Jennings, P.C. & Kisslinger, C. (Eds.), pp. 1657–1663 & supplementary CD-ROM. International Geophysics Series No. 81, Part B, Academic Press, London, UK.
- Langridge, R. M.; Stenner, H. D.; Fumal, T. E.; Christofferson, S. A.; Rockwell, T. K.; Hartleb, R. D.; Bachhuber, J. & Barka, A. A. (2002) Geometry, slip distribution, and kinematics of surface rupture on the Sakarya fault segment during the 17 August 1999 Izmit, Turkey, earthquake. *Bulletin of the Seismological Society of America*, 92, 107–125.
- Lean, J. L.; Picone, J. M.; Emmert, J. T. & Moore, G. (2006) Thermospheric densities derived from spacecraft orbits: application to the Starshine satellites. *Journal of Geophysical Research*, 111, A04301.
- Lee, W.H.K. & Lahr, J.C. (1975) *HYP071 (revised): a computer program for determining hypocenter, magnitude, and first motion pattern of local earthquakes*. U.S. Geological Survey Open File Report 75-311, 113 p.
- Lee, S. J. & Ma, K. F. (2000) Rupture process of the 1999 Chi-Chi, Taiwan, earthquake from the inversion of teleseismic data. *Terrestrial, Atmospheric and Oceanic Sciences*, 11, 591–608.
- Lee, Y.T.; Turcotte, D.L.; Holliday, J.R.; Sachs, M.K.; Rundle, J.B.; Chen, C.C. & Tiampo, K.F. (2011) Results of the Regional Earthquake Likelihood Models (RELM) test of earthquake forecasts in California. *Proceedings of the National Academy of Sciences of the United States of America*, 108, 16533–16538.
- Lei, X.; Kusunose, K.; Satoh, T. & Nishizawa, O. (2003) The hierarchical rupture process of a fault: An experimental study. *Physics of the Earth and Planetary Interiors*, 137, 213–228.
- Leistel, J.M.; Marcoux, E.; Thiéblemont, D.; Quesada, C.; Sánchez, A.; Almodóvar, G.R.; Pascual, E. & Sáez, R. (1998) The volcanic-hosted massive sulphide deposits of the Iberian Pyrite Belt. *Mineralium Deposita*, 33, 2–30.
- Li, C.; Zhang, X. & Cao, Z. (2005) Triangular and Fibonacci number patterns driven by stress on core/shell microstructures. *Science*, 309, 909–911.
- Lindner, B.; García-Ojalvo, J.; Neiman, A.; Schimansky-Geier, L. (2004) Effects of noise in excitable systems. *Physics Reports*, 392, 321–424.
- Lippiello, E.; Corral, Á.; Bottiglieri, M.; Godano, C. & de Arcangelis, L. (2012) Scaling behavior of the earthquake intertime distribution: Influence of large shocks and time scales in the Omori law. *Physical Review E*, 86, 066119.
- Liu, J.; Sieh, K. & Hauksson, E. (2003) A structural interpretation of the aftershock “cloud” of the 1992 M_w 7.3 Landers earthquake. *Bulletin of the Seismological Society of America*, 93, 1333–1344.
- Liu, M.; Stein, S. & Wang, H. (2011) 2000 years of migrating earthquakes in North China: How earthquakes in midcontinents differ from those at plate boundaries. *Lithosphere*, 3, 128–132.
- Lomnitz, C. (1982) An uncertainty principle of earthquake locations. *Bulletin of the Seismological Society of America*, 72, 669–670.

- Lomnitz, C. (2012) Three theorems of earthquake location. *Bulletin of the Seismological Society of America*, 96, 306–312.
- López, C. (2008) *Nuevas fórmulas de magnitud para la Península Ibérica y su entorno*. MsC Thesis, Departamento de Física de la Tierra, Astronomía y Astrofísica I. Universidad Complutense de Madrid (Spain).
- López, C. & Muñoz, D. (2003) Magnitude formulas in the Spanish bulletins and catalogues. *Física de la Tierra*, 15, 49–71.
- López, C. *et al.* (2012) Monitoring the volcanic unrest of El Hierro (Canary Islands) before the onset of the 2011–2012 submarine eruption. *Geophysical Research Letters*, 39, L13303.
- López Casado, C.; Molina, S.; Giner, J.J. & Delgado, J. (2000) Magnitude-intensity relationships in the Ibero-Magrebhian region. *Pure and Applied Geophysics*, 22, 271–297.
- López-Ruiz, R.; Vázquez-Prada, M.; Gómez, J. B. & Pacheco, A. F. (2004). A model of characteristic earthquakes and its implications for regional seismicity. *Terra Nova*, 16, 116–120.
- Lorenz, E.N. (1963) Deterministic non-periodic flow. *Journal of the Atmospheric Sciences* 20, 130–141.
- Luzón, F.; García-Jerez, A.; Santoyo, M.A. & Sánchez-Sesma, F.J. (2010) Numerical modelling of pore pressure variations due to time varying loads using a hybrid technique: the case of the Itoiz reservoir (Northern Spain). *Geophysical Journal International*, 180, 327–338.
- Maciá, E. (2006) The role of aperiodic order in science and technology. *Reports on Progress in Physics*, 69, 397–441.
- Main, I. (1996) Statistical physics, seismogenesis, and seismic hazard. *Reviews of Geophysics*, 34, 433–462.
- Main, I. (2013) Little earthquakes in the lab. *Physics*, 6, 20.
- Malamud, B.D. & Turcotte, D.L. (2000) Cellular-automata models applied to natural hazards. *Computing in Science and Engineering*, 2, 42–51.
- Malamud, B.D.; Morein, G. & Turcotte, D.L. (1998) Forest fires: An example of self-organized critical behavior. *Science*, 281, 1840–1842.
- Mancilla, F.; Del Pezzo, E.; Stich, D.; Morales, J.; Ibañez, J. & Bianco, F. (2012) Q_P and Q_S in the upper mantle beneath the Iberian peninsula from recordings of the very deep Granada earthquake of April 11 (2010) *Geophysical Research Letters*, 39, L09303.
- Mann, N. R.; Schafer, R. E. & Singpurwalla, N. D. (1974) *Methods for Statistical Analysis of Reliability and Life Data* John Wiley and Sons, New York (USA), 564 p.
- Marsan, D. & Enescu, B. (2012) Modeling the foreshock sequence prior to the 2011, M_w 9.0 Tohoku, Japan, earthquake. *Journal of Geophysical Research*, 117, B06316.
- Marsan, D. & Lengliné, O. (2008) Extending earthquakes' reach through cascading. *Science*, 319, 1076–1079.
- Martí, J.; Ortiz, R.; Gottsmann, J.; Garcia, A. & De La Cruz-Reyna, S. (2009) Characterising unrest during the reawakening of the central volcanic complex on Tenerife, Canary Islands, 2004–2005, and implications for assessing hazards and risk mitigation. *Journal of Volcanology and Geothermal Research*, 182, 23–33.

- Martín, R.; Stich, D.; Morales, J. & Mancilla, F. (2015) Moment tensor solutions for the Iberian-Maghreb region during the IberArray deployment (2009–2013). *Tectonophysics*, 663, 261–274.
- Martín Dávila, J.; Gárate Pasquín, J.; Pazos García, A. & Catalán Morollón, M. (2006) The Geophysics in the Royal Naval Institute and Observatory in San Fernando. *Física de la Tierra*, 18, 119–135.
- Martínez-Díaz, J.J.; Bejar-Pizarro, M.; Álvarez-Gómez, J.A.; Mancilla, F. de L.; Stich, D.; Herrera, G. & Morales, J. (2012) Tectonic and seismic implications of an intersegment rupture: The damaging May 11th 2011 Mw 5.2 Lorca, Spain, earthquake. *Tectonophysics*, 546–547, 28–37.
- Martínez Solares, J.M. (1981) *El terremoto de Andalucía del 25 de diciembre de 1884*. Instituto Geográfico Nacional, Madrid, Spain, 139 p.
- Martínez Solares, J.M. (1992) *Optimización de parámetros de localización para distintos tipos de redes sísmicas. Aplicación a la Península Ibérica*. PhD Thesis, Universidad Complutense de Madrid, 257 p.
- Martínez Solares, J.M. (1995) Métodos de localización epicentral mediante el procesamiento de señales de dispositivos sísmicos. In: *Redes sísmicas regionales*, Mezcua, J. (Ed.), pp. 113–144. Monografía 11, Instituto Geográfico Nacional, Madrid (Spain).
- Martínez-Solares, J. (2003) Historical seismicity of the Iberian Peninsula. *Física de la Tierra*, 15, 13–28.
- Martínez-Solares, J. & López Arroyo, A. (2004) The great historical 1755 earthquake. Effects and damage in Spain. *Journal of Seismology*, 8, 275–294.
- Martínez Solares, J.M. & Mezcua, J. (2002) *Catálogo Sísmico de la Península Ibérica (880 a.C. – 1900)*. Monografía 18, Instituto Geográfico Nacional, Madrid (Spain), 253 p. + map.
- Martínez Solares, J.M.; Cabañas Rodríguez, L.; Benito Oterino, M.B.; Rivas Medina, A.; Gaspar Escribano, J.M.; Ruíz Barajas, S. & Rodríguez Zaloña, Ó. (2013) *Actualización de mapas de peligrosidad sísmica de España 2012*. Instituto Geográfico Nacional, Madrid, Spain, 267 p.
- Marzocchi, W.; Zechar, J.D. & Jordan, T.H. (2012) Bayesian forecast evaluation and ensemble earthquake forecasting. *Bulletin of the Seismological Society of America*, 102, 2574–2584.
- Marzocchi, W.; Iervolino, I.; Giorgio, M. & Falcone, G. (2015) When is the probability of a large earthquake too small? *Seismological Research Letters*, 86, 1674–1678.
- Masana, E.; Villamarín, J.A.; Sánchez Cabañero, J.; Plaza, J. & Santanach, P. (2001) Seismogenic faulting in an area of low seismic activity: Paleoseismicity of the El Camp fault (Northeast Spain). *Netherlands Journal of Geosciences*, 80, 29–41.
- Masoller, C. (2001) Anticipation in the synchronization of chaotic semiconductor lasers with optical feedback. *Physical Review Letters*, 86, 2782–2785.
- Matthews, M. V.; Ellsworth, W. L. & Reasenber, P. A. (2002) A Brownian model for recurrent earthquakes. *Bulletin of the Seismological Society of America*, 92, 2233–2250.
- Maley, P. D.; Moore, R. G. & King, D. J. (2002) Starshine: a student-tracked atmospheric research satellite. *Acta Astronautica*, 51, 715–721.

- Mason, I.B. (2003) Binary forecasts. In: *Forecast Verification: A Practitioner's Guide in Atmospheric Science*, pp. 37-76 (references on pp. 215-226), Joliffe, I.T. & Stephenson, D.B. (Eds.). Wiley, West Sussex, UK.
- Matthews, M.V.; Ellsworth, W.L. & Reasenberg, P.A. (2002) A Brownian model for recurrent earthquakes. *Bulletin of the Seismological Society of America*, 92, 2233–2250.
- McCalpin, J.P. (Ed.; 2009) *Paleoseismology*. 2nd Ed.; International Geophysics Series, No. 95, Academic Press, New York. 629 p. + online companion materials.
- McKay, A. (1890) On the earthquakes of September 1888, in the Amuri District. *New Zealand Geological Survey Reports of Geological Explorations 1888–1889*, 20, 1–16.
- McNamara, D.E. & Walter, W.R. (2001) Mapping crustal heterogeneity using *Lg* propagation efficiency throughout the Middle East, Mediterranean, Southern Europe and Northern Africa. *Pure and Applied Geophysics*, 158, 1165–1188.
- Meisel, L.V. & Johnson, M.A. (1994) Multifractal analysis of imprecise data: Badii-Politi and correlation integral approaches. *Physical Review E*, 50, 4214–4219.
- Melnick, D.; Moreno, M.; Cisternas, M. & Tassara, A. (2012) Darwin seismic gap closed by the 2010 Maule earthquake. *Andean Geology*, 39, 558–563.
- Mezcua, J. (1982) *Catálogo General de Isosistas de la Península Ibérica*, Publicación 202, Instituto Geográfico Nacional, Madrid, Spain, 61 p. + 261 maps.
- Mezcua, J. (1995) Fundamentos de la red sísmica de España. In: *Redes sísmicas regionales*, Mezcua, J. (Ed.), pp. 63–86. Monografía 11, Instituto Geográfico Nacional, Madrid (Spain).
- Mezcua, J. & Martínez-Solares, J.M. (1983) *Sismicidad del Área Ibero-Mogrebí*. Publicación Técnica 203, Instituto Geográfico Nacional, Madrid (Spain), 299 p + 2 maps.
- Mezcua, J. & Rueda, J. (1994) Earthquake relative location based on waveform similarity. *Tectonophysics*, 233, 253–263.
- Mezcua, J.; Buforn, E.; Udías, A. & Rueda, J. (1992) Seismotectonics of the Canary Islands, *Tectonophysics*, 208, 447–452.
- Mezcua, J.; Rueda, J. & García Blanco, R.M. (2004) Reevaluation of historic earthquakes in Spain. *Seismological Research Letters*, 75, 75–81.
- Mezcua, J.; García Blanco, R.M. & Rueda, J. (2008) On the strong ground motion attenuation in Spain *Bulletin of the Seismological Society of America*, 98, 1343–1353.
- Mezcua, J.; Rueda, J. & García Blanco, R.M. (2011) A new probabilistic seismic hazard study of Spain, *Nat. Hazards*, 59, 1087–1108.
- Mezcua, J.; Rueda, J. & García Blanco, R.M. (2013) Iberian Peninsula historical seismicity revisited: An intensity data bank, *Seismological Research Letters*, 84, 9–18.
- Michael, A. J. (2005) Viscoelasticity, postseismic slip, fault interactions, and the recurrence of large earthquakes. *Bulletin of the Seismological Society of America*, 95, 1594–1603.
- Michael, A.J. (2012) Do aftershock probabilities decay with time? *Seismological Research Letters*, 83, 630-632.
- Michael, A.J. (2014) How Complete is the ISC-GEM Global Earthquake Catalog? *Bulletin of the Seismological Society of America*, 104, 1829–1837.
- Michael, A.J. & Jones, L.M. (1998) Seismicity alert probabilities at Parkfield, California, revisited, *Bulletin of the Seismological Society of America*, 88, 117–130.

- Michael, A.J. & Jones, L.M. (2004) Parkfield: A mainshock defined and a definition tested. *American Geophysical Union, Fall Meeting 2004, San Francisco, California*, abstract S54B-05. <http://adsabs.harvard.edu/abs/2004AGUFM.S54B..05M> [Last accessed: November 2015.]
- Michalakes, J. G.; Purser, R. J. & Swinbank, R. (1999) Data structure and parallel decomposition considerations on a Fibonacci grid. In: *Preprints of the 13th Conference on Numerical Weather Prediction, Denver, 13-17 September 1999*, pp. 129–130. American Meteorological Society.
- Michellini, A.; De Simoni, B.; Amato, A. & Boschi, E. (2005) Collecting, digitizing, and distributing historical seismological data. *Eos, Transactions, American Geophysical Union*, 86, 261–266.
- Mignan, A. (2012) Functional shape of the earthquake frequency-magnitude distribution and completeness magnitude. *Journal of Geophysical Research*, 117, B08302.
- Mignan, A. & Woessner, J. (2012) Estimating the magnitude of completeness for earthquake catalogs. *Community Online Resource for Statistical Seismicity Analysis*. doi:10.5078/corssa-00180805. Available at www.corssa.org [Last accessed October 2015.]
- Mignan, A.; Jiang, C.; Zechar, J.D.; Wiemer, S.; Wu, Z. & Huang, Z. (2013) Completeness of the mainland China earthquake catalog and implications for the setup of the China earthquake forecast testing center. *Bulletin of the Seismological Society of America*, 103, 845–859.
- Miller, S.A. (2002) Properties of large ruptures and the dynamical influence of fluids on earthquakes and faulting. *Journal of Geophysical Research*, 107, 2182.
- Miller, S. A.; Ben-Zion, Y.; Burg, J.-P. (1999) A three-dimensional fluid-controlled earthquake model: Behavior and implications. *Journal of Geophysical Research*, 104, 10621–10638.
- Milne, J. (1895) Seismic disturbances observed in Europe and at Teneriffe by Dr. E. von Rebeur-Paschwitz. *The Seismological Journal of Japan*, 4, 354–361.
- Ministerio de Fomento (2003) *Norma de Construcción Sismorresistente: Parte General y Edificación (NCSE-02)*. Ministerio de Fomento, Madrid, Spain, 96 p.
- Molchan, G.M. (1997) Earthquake prediction as a decision-making problem. *Pure and Applied Geophysics*, 149, 233–247.
- Molchan, G.M. (2003) Earthquake prediction strategies: A theoretical analysis. In: *Nonlinear Dynamics of the Lithosphere and Earthquake Prediction*, pp. 209–237 (references in pp. 311–332), Keilis-Borok, V.I. & Soloviev, A.A. (Eds.), Springer-Verlag, Berlin, Germany.
- Molchan, G.M. (2012) On the testing of seismicity models. *Acta Geophysica*, 60, 624–637.
- Molchan, G. & Kronrod, T. (2009) The fractal description of seismicity. *Geophysical Journal International*, 179, 1787–1799.
- Monge Montuno, F. (1981) *Catálogo sísmico preliminar del Archipiélago Canario*. Publicación 165, Cátedra de Geofísica, Universidad Complutense de Madrid, Madrid, Spain, 32 p.
- Morales, J.; Cantavella, J.V.; Mancilla, F. de L.; Lozano, L.; Stich, D.; Herraiz, E.; Martín, J.B.; Lopez-Comino, J.A. & Martinez-Solares, J.M. (2014) The 2011 Lorca seismic series: Temporal evolution, faulting parameters and hypocentral relocation. *Bulletin of Earthquake Engineering*, 12, 1871–1888.

- Morelli, L.G. & Zanette, D.H. (1998) Synchronization of stochastically coupled cellular automata. *Physical Review E*, 58, R8–R11.
- Moreno, Y.; Gómez, J.B. & Pacheco, A.F. (1999) Self organized criticality in a fiber-bundle-type model. *Physica A*, 274, 400–409.
- Moreno, Y.; Vázquez-Prada, M.; Gómez, J.B. & Pacheco, A.F. (2003) Error diagrams and temporal correlations in a fracture model with characteristic and power-law distributed avalanches. *The European Physical Journal B*, 34, 489–494.
- Moritz, H. (2000) Geodetic Reference System 1980. *Journal of Geodesy*, 74, 128–133.
- Mouayn, I. (2007) The M_D scale in northern Morocco: A comparative study of two empirical approaches. *Pure and Applied Geophysics*, 164, 957–974.
- Mouayn, I.; Tadili, B.A.; Brahim, L.A.; Ramdani, M.; Limouri, M. & Jabour, N. (2004) Duration magnitude scale and site residuals for northern Morocco. *Pure and Applied Geophysics*, 161, 1061–1080.
- Mulargia, F. (2001) Retrospective selection bias (or the benefit of hindsight). *Geophysical Journal International*, 146, 489–496.
- Muñoz, D. & Udías, A. (1982) Historical development of Spain's catalogs of earthquakes. *Bulletin of the Seismological Society of America*, 72, 1039–1042.
- Munuera Quiñero, J.M. (1963). Seismic Data. Datos básicos para un estudio de sismicidad en el área de la Península Ibérica. *Memorias del Instituto Geográfico y Catastral (Madrid)*, XXXII, 93 p.
- Na, H. S.; Lee, C. N. & Cheong, O. (2002) Voronoi diagrams on the sphere. *Computational Geometry – Theory and Applications*, 23, 183–194.
- Nadeau, R.M.; Foxall, W. & McEvilly, T.V. (1995) Clustering and periodic recurrence of microearthquakes on the San Andreas Fault at Parkfield, California. *Science*, 267, 503–507.
- Nagy, W. (1996) *New region-dependent travel-time handling facilities at the IDC; functionality, testing and implementation details*. Science Applications International Corporation Technical Report, SAIC-96/1179, 57 p.
- Nanjo, K.Z. (2010) Earthquake forecast models for Italy based on the RI algorithm. *Annals of Geophysics*, 53, 117–127.
- Nanjo, K.Z. (2011) Earthquake forecasts for the CSEP Japan experiment based on the RI algorithm. *Earth Planets and Space*, 63, 261–274.
- Nanjo, K.Z.; Enescu, B.; Shcherbakov, R.; Turcotte, D.L.; Iwata, T. & Ogata, Y. (2007) Decay of aftershock activity for Japanese earthquakes. *Journal of Geophysical Research*, 112, B08309.
- Nanjo, K.Z.; Ishibe, T.; Tsuruoka, H.; Schorlemmer, D.; Ishigaki, Y. & Hirata, N. (2010) Analysis of the completeness magnitude and seismic network coverage of Japan. *Bulletin of the Seismological Society of America*, 100, 3261–3268.
- NASA (2002) *The Blue Marble True-Color Global Imagery at 1km Resolution*. http://earthobservatory.nasa.gov/Features/BlueMarble/BlueMarble_2002.php [Last accessed October 2015.]
- Newman, W. I. & Turcotte, D. L. (2002) A simple model for the earthquake cycle combining self-organized complexity with critical point behavior. *Nonlinear Processes in Geophysics*, 9, 453–461.

- Nicol, A.; Walsh, J.; Berryman, K. & Nodder, S. (2005) Growth of a normal fault by the accumulation of slip over millions of years. *Journal of Structural Geology*, 27, 327–342.
- Niederreiter, H. (1992) *Random Number Generation and Quasi-Monte Carlo Methods*. Society for Industrial and Applied Mathematics, Philadelphia, 247 p.
- Niederreiter, H. & Sloan, I. H. (1994) Integration of nonperiodic functions of two variables by Fibonacci lattice rules. *Journal of Computational and Applied Mathematics*, 51, 57–70.
- NIMA – National Imaging and Mapping Agency (2000) *Department of Defense World Geodetic System 1984. Its Definition and Relationships with Local Geodetic Systems*. NIMA Technical Report 8350.2, 3rd Ed.; St. Louis (Missouri, USA), 175 p.
- Nishenko, S. P. & Buland, R. (1987) A generic recurrence interval distribution for earthquake forecasting. *Bulletin of the Seismological Society of America*, 77, 1382–1399.
- Noriega, R.; Ugalde, A.; Villaseñor, A. & Jurado, M.J. (2015) Spatial variation of *Lg*-wave attenuation in the Iberian Peninsula. *Bulletin of the Seismological Society of America*, 105, 420–430.
- Nuttli, O.W. (1973) Seismic-wave attenuation and magnitude relations for eastern North-America. *Journal of Geophysical Research*, 78, 876–885.
- Nye, J. F. (2003) A simple method of spherical near-field scanning to measure the far fields of antennas or passive scatterers. *IEEE Transactions on Antennas and Propagation*, 51, 2091–2098.
- Ochieng, W. Y.; Sheridan, K. F.; Sauer, K. & Han, X. (2002) An assessment of the RAIM performance of a combined Galileo/GPS navigation system using the marginally detectable errors (MDE) algorithm. *GPS Solutions* 5, 42–51.
- Ogata, Y. (1988) Statistical models for earthquake occurrences and residual analysis for point processes, *Journal of the American Statistical Association*, 83, 9–27.
- Ogata, Y. (1999) Estimating the hazard of rupture using uncertain occurrence times of paleoearthquakes. *Journal of Geophysical Research*, 104, 17995–18014.
- Ogata, Y. & Katsura, K. (1993) Analysis of temporal and spatial heterogeneity of magnitude frequency distribution inferred from earthquake catalogs. *Geophysical Journal International*, 113, 727–738.
- Okada, T.; Matsuzawa, T. & Hasegawa, A. (2003) Comparison of source areas of $M_{4.8} \pm 0.1$ repeating earthquakes off Kamaishi, NE Japan: are asperities persistent features? *Earth and Planetary Science Letters*, 213, 361–374.
- Olami, Z.; Feder, H. J. S. & Christensen, K. (1992) Self-organized criticality in a continuous, nonconservative cellular automaton modeling earthquakes. *Physical Review Letters*, 68, 1244–1247.
- Palmer, T.N.; Shutts, G.J.; Hagedorn, R.; Doblas-Reyes, F.J.; Jung, T. & Leutbecher, M. (2005) Representing model uncertainty in weather and climate prediction. *Annual Review of Earth and Planetary Sciences*, 33, 163–193.
- Panzer, F.; Zechar, J.D.; Vogfjörð, K.S. & Eberhard, D.A.J. (2016) A revised earthquake catalogue for south Iceland. *Pure and Applied Geophysics*, 173, 97–116.
- Parisi, G. (2015) Brownian motion, *Nature* 433, 221.
- Parsons, T. (2008a) Earthquake recurrence on the south Hayward fault is most consistent with a time dependent, renewal process. *Geophysical Research Letters*, 35, L21301.

- Parsons, T. (2008b) Persistent earthquake clusters and gaps from slip on irregular faults. *Nature Geoscience*, 1, 59–63.
- Parsons, T.; Toda, S.; Stein, R.S.; Barka, A. & Dieterich, J.H. (2000) Heightened odds of large earthquakes near Istanbul: An interaction-based probability calculation. *Science*, 288, 661–665.
- Parsons, T.; Segou, M. & Marzocchi, W. (2014) The global aftershock zone. *Tectonophysics*, 618, 1–34.
- Payo, G. (1960) Estudio sobre las ondas superficiales Lg, Rg, y Li en los registros del observatorio de Toledo. *Annali di Geofisica*, 13, 297–325.
- Payo, G. (1996) Fórmulas de magnitud a partir de registros sísmicos españoles. In: *Homenaje en honor al profesor Fernando de Miguel Martínez*, pp. 551–570, Universidad de Granada, Granada, Spain.
- Payo, G. & Gómez-Menor, R. (1998) *Historia del Observatorio Geofísico de Toledo*. Instituto Geográfico Nacional, Madrid, 222 p.
- Peláez, J.A.; Chourak, M.; Tadili, B.A.; Ait Brahim, L.; Hamdache, M.; López Casado, C. & Martínez Solares, J.M. (2007) A catalog of main Moroccan earthquakes from 1045 to 2005. *Seismological Research Letters*, 78, 614–621.
- Peláez, J.A.; Hamdache, M. & Martínez Solares, J.M. (2012) Relationship between computed mb (IGN) and ML (CRAAG) magnitudes for earthquakes located in Northern Algeria. In: *Proceedings, 7ª Asamblea Hispano Portuguesa de Geodesia y Geofísica*, San Sebastián, Spain, 189–192.
- Pena, J.O.A.; Nunes, J.A.C. & Carrilho, F.J.R. (2014) *Catálogo Sísmico de Portugal Continental e Região Adjacente para o Período 1961–1969*. Instituto Português do Mar e da Atmosfera, Lisbon, Portugal, 73 p.
- Peng, Z. & Zhao, P. (2009) Migration of early aftershocks following the 2004 Parkfield earthquake. *Nat. Geosci.*, 2, 877–881.
- Peng, Z.; Vidale, J.E.; Ishii, M. & Helmstetter, A. (2007) Seismicity rate immediately before and after main shock rupture from high-frequency waveforms in Japan, *Journal of Geophysical Research*, 112, B03306.
- Pepke, S. L. & Carlson, J. M. (1994) Predictability of self-organizing systems. *Physical Review*, E50, 236–242.
- Peterson, J. & Hutt, C.R. (2014) *World-Wide Standardized Seismograph Network: A Data Users Guide*, U.S. Geological Survey Open-File Report 2014-1218, Reston (Virginia, U.S.A.), 74 p.
- Poincaré, H. (1913) *The Foundations of Science*. The Science Press, New York, USA, 553 p.
- Pollitz, F.; Stein, R.S.; Sevilgen, V. & Bürgmann, R. (2012) The 11 April 2012 east Indian Ocean earthquake triggered large aftershocks worldwide, *Nature*, 490, 250–253.
- Posadas, A.M.; Vidal, F.; de Miguel, F.; Alguacil, G.; Peña J.; Ibañez, J. M.; Morales, J. (1993) Spatial-temporal analysis of a seismic series using the principal components method: The Antequera series, Spain, 1989. *Journal of Geophysical Research*, 98, 1923–1932.
- Preston, E.F.; Sá Martins, J.S.; Rundle, J.B.; Anghel, M. & Klein, W. (2000) Models of earthquake faults with long-range stress transfer. *Computing in Science and Engineering*, 2, 34–41.

- Prusinkiewicz, P. & Lindenmayer, A. (1990) *The Algorithmic Beauty of Plants*. Springer-Verlag, New York, 228 p.
- Purser, R. J. (2008) *Generalized Fibonacci grids: a new class of structured, smoothly adaptive multi-dimensional computational lattices*. Office Note 455, National Centers for Environmental Prediction, Camp Springs, Maryland, USA, 37 p.
- Purser, R. J. & Swinbank, R. (2006) *Generalized Euler-Maclaurin formulae and end corrections for accurate quadrature on Fibonacci grids*. Office Note 448, National Centers for Environmental Prediction, Camp Springs, Maryland, USA, 19 p.
- Rakhmanov, E. A.; Saff, E. B. & Zhou, Y. M. (1994) Minimal discrete energy on the sphere. *Mathematical Research Letters*, 1, 647–662.
- Ramón y Cajal, S. (1916) *Reglas y consejos sobre investigación científica. (Los tónicos de la voluntad)*. 5^a Ed.; Imprenta Nicolás Moya, Madrid, Spain, 318 p.
- Reasenberg, P.A. (1999) Foreshock occurrence rates before large earthquakes worldwide. *Pure and Applied Geophysics*, 155, 355–379.
- Reid, H. F. (1910) The mechanics of the earthquake. In: *The California earthquake of April 18, 1906, Report of the State Earthquake Investigation Commission*, Vol. 2, pp. 1–192. Carnegie Institution, Washington, DC (USA).
- Reiter, L. (1991) *Earthquake Hazard Analysis*, Columbia University Press, New York, USA, 254 p.
- RéNaSS (2015) *Réseau National de Surveillance Sismique*, renass.unistra.fr/ [Last accessed: September 2015.]
- Rey Pastor, A. (1927) *Traits Sismiques de la Péninsule Ibérique*. Ateliers de L'Institut Géographique et Cadastral, Madrid, Spain. 72 p. + maps.
- Rey Pastor, A. (1929) Estudio crítico de los aparatos de las Estación Sismológica de Toledo. In: *Asociación Española para el Progreso de las Ciencias. Congreso de Barcelona*, pp. 145–169. Establecimiento Tipográfico Huelves y Compañía, Madrid.
- Rey Pastor, A. (1934) *Carta de sismicidad del globo formada por A. Rey Pastor con arreglo a los datos publicados por N.H. Heck correspondientes a los años 1899–1910 y 1913–1930*. Instituto Geográfico y Catastral, Madrid (Spain), 10 p + map.
- Rhoades, D. A.; Van Dissen, R. J. & Dowrick, D. J. (1994) On the handling of uncertainties in estimating the hazard of rupture on a fault segment. *Journal of Geophysical Research*, 99, 13701–13712.
- Rhoades, D.A. & Evison, F.F. (2004) Long-range earthquake forecasting with every earthquake a precursor according to scale. *Pure and Applied Geophysics*, 161, 47–72.
- Rhodes, F. H. T. & Malamud, B. D. (2008) Preface. In: Rhodes, F. H. T.; Stone, R. O. & Malamud, B. D. (eds.): *Language of the Earth*, 2nd ed.; Blackwell Publishing, Malden (Massachusetts, USA), Oxford (United Kingdom) and Carlton (Victoria, Australia) p. ix–xi.
- Ribeiro, J.R.; Ribeiro, A.I. & Correia, A.P. (2015) Solving the mystery: The 21 October 1880 Portuguese earthquake. *Seismological Research Letters*, 86, 991–998.
- Ridley, J. N. (1982) Packing efficiency in sunflower heads. *Mathematical Biosciences*, 58, 129–139.

- Ridley, J. N. (1986) Ideal phyllotaxis on general surfaces of revolution. *Mathematical Biosciences*, 79, 1–24.
- Rikitake, T. (1974) Probability of earthquake occurrence as estimated from crustal strain. *Tectonophysics*, 23, 299–312.
- Rikitake, T. (1976b) Recurrence of great earthquakes at subduction zones. *Tectonophysics*, 35, 335–362.
- Ringdal, F. (1975) On the estimation of seismic detection thresholds. *Bulletin of the Seismological Society of America*, 65, 1631–1642.
- Ringdal, F. & Bungum, H. (1977) Noise level variation at NORSAR and its effect on detectability. *Bulletin of the Seismological Society of America*, 67, 479–492.
- Roberts, N.S. Bell, A.F. & Main, I.G. (2015) Are volcanic seismic *b*-values high, and if so when? *Journal of Volcanology and Geothermal Research*, 308, 127–141
- Robinson, R. (2004) Potential earthquake triggering in a complex fault network: the northern South Island, New Zealand. *Geophysical Journal International*, 159, 734–748.
- Robinson, R.; Benites, R. (2001) Upgrading a synthetic seismicity model for more realistic fault ruptures. *Geophysical Research Letters*, 28, 1843–1846.
- Roca, A.; Izquierdo, A.; Sousa-Oliveira, C. & Martínez-Solares J. (2004) An outline of earthquake catalogues, databases and studies of historical seismicity in the Iberian Peninsula. *Ann. Geophys.*, 47, 561–570.
- Roca, A.; Locati, M.; Susagna, T.; Fleta, J.; Pujol, J.; Batlló, J.; Valero, J.F.; Martínez-Solares, J.M.; Cruz, J.; Alves, P.; Gómez Capera, A.A. & Stucchi, M. (2011) Archivo online de datos macrosísmicos de la Península Ibérica, in *Actas del Cuarto Congreso Nacional de Ingeniería Sísmica, Granada (Spain)*, 10 p. www.earth-prints.org/handle/2122/8599 [Last accessed: April 2015.]
- Rodríguez de la Torre, F. (1989) Management and interpretation of hemerographical sources concerning research about historical Iberian seismicity, in *Proceedings of the Workshop on Historical Earthquakes in the Ibero-Maghrebian Region. Methodological Approach and Case Studies. Lisbon, Portugal, 15-17 November 1989*. IV, 23–94.
- Rodríguez de la Torre, F. (1990) *Sismicidad y sismología en la Península Ibérica durante el siglo XIX*. PhD Thesis, Universitat de Barcelona. 2615 p.
- Rodríguez de la Torre, F. (1993) *Revisión del catálogo sísmico ibérico. Años 1801 a 1850*. Internal report, Instituto Geográfico Nacional, Madrid, Spain.
- Rodríguez de la Torre, F. (1997) *Revisión del catálogo sísmico ibérico. Años 1760 a 1800*. Internal report, Instituto Geográfico Nacional, Madrid, Spain.
- Rodríguez de la Torre, F. (2004) Iberian seismic catalogue. A complete revision of all the 19th century. In: *Papers of the European Seismological Commission General Assembly, Potsdam*. www.esc-web.org/the-esc-general-assembly/111-esc-papers-september-2004.html [Last accessed March 2015.]
- Roeloffs, E. (2000) The Parkfield, California earthquake experiment: An update in 2000. *Current Science*, 79, 1226–1236.
- Romani, F. (1976) Cellular automata synchronization. *Information Sciences*, 10, 299–318.

- Romano, F.; Trasatti, E.; Lorito, S.; Piromallo, C.; Piatanesi, A.; Ito, Y.; Zhao, D.; Hirata, K.; Lanucara, P. & Cocco, M. (2014) Structural control on the Tohoku earthquake rupture process investigated by 3D FEM, tsunami and geodetic data. *Scientific Reports*, 4, 5631.
- Rosa, C.J.P.; McPhie, J. & Relvas, J.M.R.S. (2010) Type of volcanoes hosting the massive sulfide deposits of the Iberian Pyrite Belt. *J. Volcanol. Geotherm. Res.*, 194, 107-126.
- Rost, S. & Thomas, C. (2002) Array seismology: methods and applications. *Rev. Geophys.*, 40, 1008.
- Rubin, A.M. (2002) Aftershocks of microearthquakes as probes of the mechanics of rupture. *Journal of Geophysical Research*, 107, 2142.
- Rueda, J. (1995) Análisis y procesamiento de datos en la Red Sísmica Nacional. In: *Redes sísmicas regionales*, Mezcua, J. (Ed.), pp. 87–112. Monografía 11, Instituto Geográfico Nacional, Madrid (Spain).
- Rueda, J. (2010) Diseño de la alerta (sísmica) IGN, in: *Jornada técnica sobre riesgo y peligrosidad de tsunamis en España. Proyectos sobre su predicción y sistemas de alerta*. Dirección General de Protección Civil y Emergencias, Rivas de Vaciamadrid, Spain, September 10, www.proteccioncivil.org/catalogo/naturales/jornada-riesgo-tsunamis/presentaciones/index.htm [Last accessed: October 2015.]
- Rueda, J. & Mezcua, J. (2002) Estudio del terremoto del 23 de septiembre de 2001 en Pego (Alicante). Obtención de una relación $m_b L_g - M_w$ para la Península Ibérica. *Rev. Soc. Geol. Esp.*, 15, 159–173.
- Rueda, J. & Mezcua, J. (2005) Near-real-time seismic moment-tensor determination in Spain. *Seismological Research Letters*, 76, 455–465.
- Rueda, J. & Mezcua, J. (2006) *Seismic network LOcation CApability program, LOCA 1.0*, www.fundaciongarciasineriz.es/descargas/loca.zip [Last accessed: August 2015.]
- Rueda, J. & Mezcua, J. (2015) Orientation analysis of the Spanish broadband national network using Rayleigh-wave polarization. *Seismological Research Letters*, 86, 929–940.
- Rueda Núñez, J.J. (2006) Discriminación sísmica mediante el análisis de las señales generadas por explosiones y terremotos. Aplicación a la región suroeste de Europa-Norte de África. PhD Thesis, Universidad Politécnica de Madrid (Spain).
- Ruegg, J.C.; Rudloff, A.; Vigny, C.; Madariaga, R.; De Chabaliér, J.B.; Campos, J.; Kausel, E.; Barrientos, S. & Dimitrov, D. (2009) Interseismic strain accumulation measured by GPS in the seismic gap between Constitución and Concepción in Chile. *Physics of the Earth and Planetary Interiors*, 175, 78–85.
- Ruiz, M.; Gaspà, O.; Gallart, J.; Díaz, J.; Pulgar, J.A.; García-Sansegundo, J.; López-Fernández, C. & González-Cortina, J.M. (2006) Aftershocks series monitoring of the September 18, 2004 $M=4.6$ earthquake at the western Pyrenees: A case of reservoir-triggered seismicity? *Tectonophysics*, 426, 223–243.
- Rundle, J.B. & Klein, W. (1993) Scaling and critical phenomena in a cellular automaton slider-block model for earthquakes. *Journal of Statistical Physics*, 72, 405–413.
- Rundle, J.B.; Turcotte, D.L. & Klein, W. (Eds.; 2000) *GeoComplexity and the Physics of Earthquakes*. American Geophysical Union, Washington, DC, 284 p.

- Rundle, P.B.; Rundle, J.B.; Tiampo, K.F.; Sa Martins, J.S.; McGinnis, S. & Klein, W. (2001) Nonlinear network dynamics on earthquake fault systems. *Physical Review Letters*, 87, 148501.
- Rundle, J.B.; Turcotte, D.L.; Shcherbakov, R.; Klein, W. & Sammis, C. (2003) Statistical physics approach to understanding the multiscale dynamics of earthquake fault systems. *Reviews of Geophysics*, 41, 1019.
- Rundle, J. B.; Rundle, P. B.; Donnellan, A. & Fox, G. (2004) Gutenberg-Richter statistics in topologically realistic system-level earthquake stress-evolution simulations. *Earth, Planets and Space*, 56, 761–771.
- Rundle, J.B.; Rundle, P.B.; Donnellan, A.; Turcotte, D.L.; Shcherbakov, R.; Li, P.; Malamud, B.D.; Grant, L.B.; Fox, G.C.; McLeod, D.; Yakovlev, G.; Parker, J.; Klein, W. & Tiampo, K.F. (2005) A simulation-based approach to forecasting the next great San Francisco earthquake. *Proceedings of the National Academy of Sciences of the United States of America*, 102, 15363–15367.
- Rundle, J.B.; Holliday, J.R.; Yoder, M.; Sachs, M.K.; Donnellan, A.; Turcotte, D.L.; Tiampo, K.F.; Klein, W. & Kellogg, L.H. (2011) Earthquake precursors: activation or quiescence? *Geophysical Journal International*, 187, 225–236.
- Rydelek, P.A. & Sacks, I.S. (1989) Testing the completeness of earthquake catalogues and the hypothesis of self-similarity. *Nature*, 337, 251–253.
- Saff, E. B. & Kuijlaars, A. B. J. (1997) Distributing many points on a sphere. *The Mathematical Intelligencer*, 19, 5–11.
- Sagan, C. (1980) *The Shores of the Cosmic Ocean*. In: *Cosmos: A Personal Voyage*, Episode 1, Public Broadcasting Service, Arlington, Virginia, USA.
- Saichev, A. & Sornette, D. (2007) Theory of earthquake recurrence times. *Journal of Geophysical Research*, 112, B04313.
- Samardjieva, E.; Payo, G. & Badal, J. (1997) *Catalogue of Digital Historical Seismograms (1912-1962)*, *Wiechert Seismograph, Geophysical Observatory of Toledo*. Instituto Geográfico Nacional, Madrid. 315 p.
- Samardjieva, E.; Payo, G.; Badal, J. & López, C. (1998) Creation of a digital database for XXth century historical earthquakes occurred in the Iberian area. *Pure and Applied Geophysics*, 152, 139–163.
- Samardjieva, E.; Payo, G. & Badal, J. (1999) Magnitude formulae and intensity-magnitude relations for early instrumental earthquakes in the Iberian Region. *Nat. Hazards*, 19, 189–204.
- Sánchez Navarro-Neumann, M.M. (1917) Ensayo sobre la sismicidad del suelo español. *Bol. R. Soc. Hist. Nat.*, 17, 83–108.
- Sánchez Navarro-Neumann, M.M. (1921) Lista de los terremotos más notables sentidos en la Península Ibérica desde los tiempos más remotos hasta 1917, inclusive, con ensayo de agrupación en regiones y períodos sísmicos. In: *La Estación Sismológica y el Observatorio Astronómico y Meteorológico de Cartuja (Granada)*, pp. 11–65. Observatorio de Cartuja, Granada, Spain.
- Sandri, L. & Marzocchi, W. (2007) A technical note on the bias in the estimation of the b -value and its uncertainty through the least squares technique. *Ann. Geophys.*, 50, 329–339.

- Santoyo, M.A.; García-Jerez, A. & Luzón, F. (2010) A subsurface stress analysis and its possible relation with seismicity near the Itoiz Reservoir, Navarra, Northern Spain. *Tectonophysics*, 482, 205–215.
- Savage, J. C. (1994) Empirical earthquake probabilities from observed recurrence intervals. *Bulletin of the Seismological Society of America*, 84, 219–221.
- Sawai, Y.; Namegaya, Y.; Okamura, Y.; Satake, K.; & Shishikura, M. (2012) Challenges of anticipating the 2011 Tohoku earthquake and tsunami using coastal geology. *Geophysical Research Letters*, 39, L21309.
- Scarpa, R.; Patanè, G. & Lombardo, G. (1983) Space-time evolution of seismic activity at Mount Etna during 1974–1982. *Annales Geophysicae*, 1, 451–462.
- Schaff, D.P. & Richards, P.G. (2004) Repeating seismic events in China. *Science*, 303, 1176–1178.
- Schaff, D.P. & Waldhauser F. (2010) One magnitude unit reduction in detection threshold by cross correlation applied to Parkfield (California) and China seismicity. *Bulletin of the Seismological Society of America*, 100, 3224–3238.
- Scharer, K.M.; Biasi, G.P.; Weldon II, R.J. & Fumal, T.E. (2010) Quasi-periodic recurrence of large earthquakes on the southern San Andreas fault. *Geology*, 38, 555–558.
- Scholz, C.H. (1998) Earthquakes and friction laws. *Nature*, 391, 37–42.
- Scholz, C.H. (2002) *The Mechanics of Earthquakes and Faulting*, 2nd ed. Cambridge University Press, Cambridge, UK, 496 p.
- Scholz, C.H. (2010) Large earthquake triggering, clustering, and the synchronization of faults. *Bulletin of the Seismological Society of America*, 100, 901–909.
- Scholz, C.H. & Gupta, A. (2000) Fault interactions and seismic hazard. *Journal of Geodynamics*, 29, 459–467.
- Schorlemmer, D. & Gerstenberger, M.C. (2007) RELM Testing Center. *Seismological Research Letters*, 7, 30–36.
- Schorlemmer, D. & Wiemer, S. (2005) Microseismicity data forecast rupture area. *Nature*, 434, 1086.
- Schorlemmer, D. & Woessner, J. (2008) Probability of detecting an earthquake. *Bulletin of the Seismological Society of America*, 98, 2103–2117. Data available at completenessweb.gfz-potsdam.de [Last accessed November 2015.]
- Schorlemmer, D.; Wiemer, S. & Wyss, M. (2005) Variations in earthquake-size distribution across different stress regimes, *Nature*, 437, 539–542.
- Schorlemmer, D.; Christophersen, A.; Rovida, A.; Mele, F.; Stucchi, M. & Marzocchi, W. (2010a) Setting up an Earthquake Forecast Experiment in Italy. *Annals of Geophysics*, 53, 1–9.
- Schorlemmer, D.; Mele, F. & Marzocchi, W. (2010b) A completeness analysis of the national seismic network of Italy. *Journal of Geophysical Research*, 115, B04308.
- Schwartz, D.P. & Coppersmith K.J. (1984) Fault behavior and characteristic earthquakes: Examples from the Wasatch and San Andreas fault zones. *Journal of Geophysical Research*, 89, 5681–5698 .

- Schweitzer, J.; Fyen, J.; Mykkeltveit, S.; & Kværna, T. (2009) Seismic Arrays. In: P. Bormann (Ed.), *New Manual of Seismological Observatory Practice (NMSOP-1)*, IASPEI. GFZ German Research Center for Geosciences, Potsdam, Germany, 52 p.
- Secanell, R.; Bertil, D.; Martin, C.; Goula, X.; Susagna, T.; Tapia, M.; Dominique, P.; Carbon, D. & Fleta, J. (2008) Probabilistic seismic hazard assessment of the Pyrenean region. *Journal of Seismology*, 12, 323–341.
- Seeber, L. & Armbruster, J.G. (2000) Earthquakes as beacons of stress change. *Nature*, 407, 69–72.
- Sens-Schönfelder, C.; Margerin, L. & Campillo, M. (2009) Laterally heterogeneous scattering explains Lg blockage in the Pyrenees. *Journal of Geophysical Research*, 114, B07309.
- Servicio Sismológico (1923) *Las informaciones macrosísmicas. Folleto de vulgarización para que sirva de guía a los observadores de temblores de tierra e instrucciones para contestar a los cuestionarios*. Talleres del Instituto Geográfico, Madrid, 38 p.
- Shaked, M.; Shanthikumar, J.G. & Valdez-Torres, J.B. (1995) Discrete hazard rate functions. *Computers and Operations Research*, 22, 391–402.
- Shaw, B.E. (2009) Constant stress drop from small to great earthquakes in magnitude-area scaling. *Bulletin of the Seismological Society of America*, 99, 871–875.
- Shcherbakov, R.; Turcotte, D.L.; Rundle, J.B.; Tiampo, K.F. & Holliday, R. (2010) Forecasting the locations of future large earthquakes: An analysis and verification. *Pure and Applied Geophysics*, 167, 743–749.
- Sieh, K.; 1996. The repetition of large-earthquake ruptures. *Proceedings of the National Academy of Sciences of the United States of America*, 93, 3764–3771.
- Sieh, K.; Stuiiver, M. & Brillinger, D. (1989) A more precise chronology of earthquakes produced by the San Andreas fault in southern California. *Journal of Geophysical Research*, 94, 603–623.
- Sierpiński W. (1915) Sur une courbe dont tout point est un point de ramification. *Comptes rendus hebdomadaires des séances de l'Académie des Sciences (Paris)*, 160, 302–305.
- Sigler, L. E. (2002) *Fibonacci's Liber Abaci: a Translation into Modern English of Leonardo Pisano's Book of Calculation*. Springer, New York, 636 p.
- Sigmundsson, F. et al. (2015) Segmented lateral dyke growth in a rifting event at Bárðarbunga volcanic system, Iceland. *Nature*, 517, 191–195.
- Silva, P.G.; Rodríguez-Pascua, M.A.; Giner Robles, J.L.; Pérez-López, R.; Lario, J.; Perucha, M.Á.; Bardají, T.; Huerta, P.; Roquero, E. & Baustista Davila, M.B. (2014) *Catálogo de los efectos geológicos de los terremotos en España*. Instituto Geológico y Minero de España & Asociación Española para el Estudio del Cuaternario, Madrid, Spain, 358 p.
- Singh, P. (1985) The so-called Fibonacci numbers in ancient and medieval India. *Historia Mathematica*, 12, 229–244.
- Sjöberg, L. E. (2006) Determination of areas on the plane, sphere and ellipsoid. *Survey Review*, 38, 583–593.
- Sloan, I. H. & Joe, S. (1994) *Lattice Methods for Multiple Integration*. Oxford University Press, Oxford, UK, 256 p.
- Slobbe, D. C.; Simons, F. J. & Klees, R. (2012) The spherical Slepian basis as a means to obtain spectral consistency between mean sea level and the geoid. *Journal of Geodesy*, 86, 609–628.

- Soloviev, A. & Ismail-Zadeh, A. (2003) Models of dynamics of block-and-fault systems. In: *Nonlinear Dynamics of the Lithosphere and Earthquake Prediction*, Keilis-Borok, V.I. & Soloviev, A.A. (Eds.), Springer, Berlin, Germany, p. 71–139 (references in p. 311–332).
- Sornette, D. & Knopoff, L. (1997) The paradox of the expected time until the next earthquake. *Bulletin of the Seismological Society of America*, 87, 789–798.
- Souriau, A. & Pauchet, H. (1998) A new synthesis of Pyrenean seismicity and its tectonic implications. *Tectonophysics*, 290, 221–244.
- Souriau, A.; Rigo, A.; Sylvander, M.; Benahmed, S. & Grimaud, F. (2014) Seismicity in central-western Pyrenees (France): A consequence of the subsidence of dense exhumed bodies. *Tectonophysics*, 621, 123–131.
- Southern California Earthquake Data Center (2015) *Alternate catalogs*. scedc.caltech.edu/research-tools/altcatalogs.html [Last accessed: November 2015.]
- Stauffer, D. & Aharony, A. (1994) *Introduction to Percolation Theory*, 2nd ed. Taylor and Francis, London, UK, 192 p.
- Stacy, S.J. & McCloskey, J. (1999) Heterogeneity and the earthquake magnitude–frequency distribution. *Geophysical Research Letters*, 26, 899–902.
- Stein, S. & Stein, J. (2014) *Playing against Nature. Integrating Science and Economics to Mitigate Natural Hazards in an Uncertain World*. American Geophysical Union; John Wiley & Sons, West Sussex, UK, 260 p.
- Stein, S.; Geller, R.J. & Liu, M. (2012) Why earthquake hazard maps often fail and what to do about it. *Tectonophysics*, 562–563, 1–25.
- Stepp, J.C. (1972) Analysis of completeness in the earthquake sample in the Puget Sound area and its effect on statistical estimates of seismic hazard. In: *Proceedings of the International Conference on Microzonation for Safer Construction Research and Applications*, Vol. 2, pp. 1897–1910. Seattle, Washington, 30 October–3 November 1972.
- Stich, D.; Martín, R. & Morales, J. (2010) Moment tensor inversion for Iberia-Maghreb earthquakes 2005–2008. *Tectonophysics*, 483, 390–398.
- Stirling, M.W.; Wesnousky, S.G. & Shimazaki, K. (1996) Fault trace complexity, cumulative slip, and the shape of the magnitude–frequency distribution for strike-slip faults: A global survey. *Geophysical Journal International*, 124, 833–868.
- Stirling, M.; Rhoades, D. & Berryman, K. (2002) Comparison of earthquake scaling relations derived from data of the instrumental and preinstrumental era. *Bulletin of the Seismological Society of America*, 92, 812–830.
- Stirling, M.W.; Goded, T.; Berryman, K. & Litchfield, N. (2013) Selection of earthquake scaling relationships for seismic-hazard analysis. *Bulletin of the Seismological Society of America*, 103, 2993–3011.
- Stock, C. & Smith, E.G.C. (2002) Comparison of seismicity models generated by different kernel estimations. *Bulletin of the Seismological Society of America*, 92, 913–922.
- Storchak, D.A.; Bird, A.L. & Adams, R.D. (2000) Discrepancies in earthquake location between ISC and other agencies. *Journal of Seismology*, 4, 321–331.
- Storchak, D.A.; Schweitzer, J. & Bormann, P. (2003) The IASPEI standard seismic phase list. *Seismological Research Letters*, 74, 761–772.

- Storchak, D.A.; Di Giacomo, D.; Bondár, I.; Engdahl, E.R.; Harris, J.; Lee, W.H.K.; Villaseñor, A. & Bormann, P. (2013) Public release of the ISC-GEM Global Instrumental Earthquake Catalogue (1900-2009). *Seismological Research Letters*, 84, 810–815.
- Susagna, T. & Goula, X. (1999) *Atles sísmic de Catalunya. Volum 1: Catàleg de sismicitat*. Institut Cartogràfic de Catalunya, Barcelona, Spain, 436 p. www.igc.cat/web/files/atles_sismic.pdf [Last accessed: September 2015]
- Svergun, D. I. (1994) Solution scattering from biopolymers: advanced contrast-variation data analysis. *Acta Crystallographica*, A50, 391–402.
- Swinbank, R. & Purser, R. J. (1999) Fibonacci grids. In: *Preprints of the 13th Conference on Numerical Weather Prediction, Denver 13–17 September 1999*. pp. 125–128, American Meteorological Society.
- Swinbank, R. & Purser, R. J. (2006a) *Standard test results for a shallow water equation model on the Fibonacci grid*. Forecasting Research Technical Report 480, Met Office, Exeter, UK, 12 p.
- Swinbank, R. & Purser, R. J. (2006b) Fibonacci grids: a novel approach to global modelling. *Quarterly Journal of the Royal Meteorological Society*, 132, 1769–1793.
- Tanioka, Y.; Hirata, K.; Hino, R. & Kanazawa, T. (2004) Slip distribution of the 2003 Tokachi-oki earthquake estimated from tsunami waveform inversion. *Earth, Planets and Space*, 56, 373–376.
- Taroni, M.; Zechar, J.D. & Marzocchi, W. (2014) Assessing annual global $M6+$ seismicity forecasts. *Geophysical Journal International*, 196, 422–431.
- Tauzin, B.; Debayle, E.; Quantin, C. & Coltice, N. (2013) Seismoacoustic coupling induced by the breakup of the 15 February 2013 Chelyabinsk meteor. *Geophysical Research Letters*, 40, 3522–3526.
- Tejedor, A.; Ambroj, S.; Gómez, J.B. & Pacheco, A.F. (2008) Predictability of the large relaxations in a cellular automaton model. *Journal of Physics A: Mathematical and Theoretical*, 41, 375102.
- Tejedor, A.; Gómez, J.B. & Pacheco, A.F. (2009a) Prediction of stasis and crisis in the Bak-Sneppen model. *Physics Letters A*, 373, 4077–4081.
- Tejedor, A.; Gómez, J.B. & Pacheco, A.F. (2009b) Earthquake size-frequency statistics in a forest-fire model of individual faults. *Physical Review E*, 79, 046102.
- Tejedor, A.; Gómez, J.B. & Pacheco, A.F. (2012) One-Way Markov process approach to repeat times of large earthquakes in faults. *Journal of Statistical Physics*, 149, 951–963.
- Tejedor, A.; Gómez, J.B. & Pacheco, A.F. (2015) The negative binomial distribution as a renewal model for the recurrence of large earthquakes. *Pure and Applied Geophysics*, 172, 23–31.
- Tiampo, K.F. & Shcherbakov, R. (2012) Seismicity-based earthquake forecasting techniques: Ten years of progress. *Tectonophysics*, 522–523, 89–121.
- Tiampo, K.F.; Rundle, J.B.; McGinnis, S.A. & Klein, W. (2002) Pattern dynamics and forecast methods in seismically active regions. *Pure and Applied Geophysics*, 159, 2429–2467.
- Telesca, L.; Lovallo, M.; Lopez, C. & Marti Molist, J. (2016) Multiparametric statistical investigation of seismicity occurred at El Hierro (Canary Islands) from 2011 to 2014. *Tectonophysics*, 672–673, 121–128.

- Tong, X.; Smith-Konter, B. & Sandwell, D.T. (2014) Is there a discrepancy between geological and geodetic slip rates along the San Andreas Fault System? *Journal of Geophysical Research*,
- Topozada, T. & Branum, D. (2006) San Andreas $M \sim 6$ earthquakes within 40 km of the Priest Valley end zone of the 1857 faulting. *Bulletin of the Seismological Society of America*, 96, S385–S396.
- Toral, R.; Mirasso, C. R.; Hernández-García, E. & Piro, O. (2001) Analytical and numerical studies of noise-induced synchronization of chaotic systems. *Chaos*, 11, 665–673.
- Torcal, F.; Serrano, I.; Havskov, J.; Utrillas, J.L. & Valero, J. (2005) Induced seismicity around the Tous New Dam (Spain). *Geophysical Journal International* 160, 144–160.
- Tormann, T.; Wiemer, S. & Hauksson, E. (2010) Changes of reporting rates in the Southern California earthquake catalog, introduced by a new definition of M_L . *Bulletin of the Seismological Society of America*, 100, 1733–1742.
- Tormann, T.; Wiemer, S. & Hardebeck, J.L. (2012) Earthquake recurrence models fail when earthquakes fail to reset the stress field. *Geophysical Research Letters*, 39, L18310.
- Tsukahara, H.; Ikeda, R. & Yamamoto, K. (2001) In situ stress measurements in a borehole close to the Nojima Fault. *The Island Arc*, 10, 261–265.
- Tsuruoka, H.; Hirata, N.; Schorlemmer, D.; Euchner, F.; Nanjo, K.Z. & Jordan, T.H. (2012) CSEP Testing Center and the first results of the earthquake forecast testing experiment in Japan. *Earth, Planets and Space*, 64, 661–671.
- Turcotte, D.L. (1997) *Fractals and Chaos in Geology and Geophysics*. 2nd ed, Cambridge University Press, Cambridge, UK, 412 p.
- Uchida, N.; Matsuzawa, T.; Ellsworth, W.L.; Imanishi, K.; Shimamura, K. & Hasegawa, A. (2012) Source parameters of microearthquakes on an interplate asperity off Kamaishi, NE Japan over two earthquake cycles. *Geophysical Journal International*, 189, 999–1014.
- Udías, A. (2013) Development of seismology in Spain in the context of the three large earthquakes of 1755, 1884 and 1954. *Earth Sciences History*, 32, 186–203.
- Udías, A. (2015) Historical earthquakes (before 1755) of the Iberian Peninsula in early catalogs. *Seismological Research Letters*, 86, 999–1005.
- Udías, A. & Mezcua, J. (1997) *Fundamentos de geofísica*. Ed. Alianza, Madrid (Spain), 476 p.
- Udías, A. & Stauder, W. (1996) The Jesuit contribution to seismology. *Seismological Research Letters*, 67, 10–19 & journal cover.
- United States Geological Survey (2015) *Quaternary Fault and Fold Database for the United States*, earthquakes.usgs.gov/regional/qfaults/ [Last accessed: November 2015].
- Urías, J.; Salazar, G. & Ugalde, E. (1998) Synchronization of cellular automaton pairs. *Chaos*, 8, 814–818.
- USGS Earthquake Magnitude Working Group (2002) *USGS Earthquake Magnitude Policy (implemented on January 18, 2002)*. earthquake.usgs.gov/aboutus/docs/020204mag_policy.php [Last accessed October 2015.]
- Utsu, T. (1965) A method for determining the value of b in a formula $\log n = a - bM$ showing the magnitude-frequency relation for earthquakes, *Geophys. Bull. Hokkaido Univ.*, 13, 99–103 [in Japanese, with English abstract].

- Utsu, T. (1969) Aftershocks and earthquake statistics (I) : Some parameters which characterize an aftershock sequence and their interrelations. *Journal of the Faculty of Science, Hokkaido University*. Series 7, 3, 129–195.
- Utsu, T. (1972a) Large earthquakes near Hokkaido and the expectancy of the occurrence of a large earthquake of Nemuro. *Report of the Coordinating Committee for Earthquake Prediction*, 7, 7–13.
- Utsu, T. (1972b) Aftershocks and earthquake statistics (IV). *Journal of the Faculty of Science, Hokkaido University, Series VII, Geophysics*, 4, 1–42.
- Utsu, T. (1984) Estimation of parameters for recurrence models of earthquakes. *Bulletin of the Earthquake Research Institute, University of Tokyo*, 59, 53–66.
- Uyeda, S.; Nagao, T. & Kamogawa, M. (2011) Earthquake precursors and prediction. In: *Encyclopedia of Solid Earth Geophysics*, p. 168–178, Gupta, H.K. (Ed.). Springer, Dordrecht, The Netherlands.
- Valée, M. & Satriano, C. (2014) Ten year recurrence time between two major earthquakes affecting the same fault segment. *Geophysical Research Letters*, 41, 2312–2318.
- Van den Dool, H. M. (2007) *Empirical Methods in Short-Term Climate Prediction*. Oxford University Press, Oxford, UK, 240 p.
- van Stiphout, T.; Wiemer, S. & Marzocchi, W. (2010) Are short-term evacuations warranted? Case of the 2009 L'Aquila earthquake. *Geophysical Research Letters*, 37, L06306.
- van Stiphout, T.; Zhuang, J. & Marsan, D. (2012) Seismicity declustering. *Community Online Resource for Statistical Seismicity Analysis* doi:10.5078/corssa-52382934. www.corssa.org [Last accessed: November 2015.]
- Vázquez-Prada, M.; González, Á.; Gómez, J. B. & Pacheco, A. F. (2002) A minimalist model of characteristic earthquakes. *Nonlinear Processes in Geophysics*, 9, 513–519.
- Vázquez-Prada, M.; González, Á.; Gómez, J. B. & Pacheco, A. F. (2003) Forecasting characteristic earthquakes in a minimalist model. *Nonlinear Processes in Geophysics*, 10, 565–571.
- Veith, K.F. & Clawson, G.E. (1972) Magnitude from short-period *P*-wave data. *Bulletin of the Seismological Society of America*, 62, 435–452.
- Vere-Jones, D. (1970) Stochastic models for earthquake occurrence (with discussion). *Journal of the Royal Statistical Society, Series B (Statistical Methodology)*, 32, 1–62 (1970).
- Vere-Jones, D. (1978) Earthquake prediction: A statistician's view. *Journal of Physics of the Earth*, 26, 129–146.
- Vere-Jones, D. (1992) Statistical methods for the description and display of earthquake catalogs. In: *Statistics in the Environmental and Earth Sciences*, Walden, A.T. & Guttorp, P. (Eds.), pp. 220–246. Edward Arnold, London, UK.
- Vila, J.; Macià, R. & Correig, A.M. (2005) Spatial variations of local magnitude in earthquakes recorded in northeastern Spain. *Pure and Applied Geophysics*, 162, 515–530.
- Vinciguerra, S. & Day, S. (2013) Magma intrusion as a driving mechanism for the seismic clustering following the 9 May 1989 earthquake swarms at the Canary Islands. *Acta Geophys.*, 61, 1626–1641.
- Vogel, H. (1979) A better way to construct the sunflower head. *Mathematical Biosciences*, 44, 179–189.

- von Rebeur-Paschwitz, E. (1892a) Das Horizontalpendel und seine Anwendung zur Beobachtung der absoluten und relativen Richtungs-Aenderungen der Lothlinie: Ergebnisse einiger mit Unterstützung der Königlich Preussischen Akademie der Wissenschaften in den Jahren 1889-1892 auf den Observatorien zu Wilhelmshaven und Potsdam sowie in Puerto Orotava auf Teneriffa ausgeführter Beobachtungsreihen. *Nova Acta Academiae Caesareae Leopoldino-Carolinae Germanicae Naturae Curiosorum*, 60, 1–216.
- von Rebeur-Paschwitz, E. (1892b) Über Horizontalpendel-Beobachtungen in Wilhelmshaven, Potsdam und Puerto Orotava auf Teneriffa. *Astron. Nachr.*, 130, 193–216.
- Voss, H. U. (2000) Anticipating chaotic synchronization. *Physical Review*, E61, 5115–5119.
- Voss, H. U. (2001a) Erratum: Anticipating chaotic synchronization [Phys. Rev. E 61, 5115 (2000)]. *Physical Review*, E64, Art. No. 039904.
- Voss, H. U. (2001b) Dynamic long-term anticipation of chaotic states. *Physical Review Letters*, 87, 014102.
- Voss, H. U. (2001c) A backward time shift filter for nonlinear delayed-feedback systems. *Physics Letters A*, 279, 207–214.
- Vriend, G. (1990) WHAT IF: a molecular modeling and drug design program. *Journal of Molecular Graphics*, 8, 52–56.
- Waldhauser, F. & Ellsworth, W.L. (2000) A double-difference earthquake location algorithm: Method and application to the northern Hayward fault. *Bulletin of the Seismological Society of America*, 90, 1353–1368.
- Wallace, P. (2000) Dance Beneath The Modern Sun. *Seeker Magazine*, 6 (8). ellenodale.com/v0800/wallacedance.html [Last accessed November 2015.]
- Ward, S. N. (2000) San Francisco Bay area earthquake simulations: A step toward a standard physical earthquake model. *Bulletin of the Seismological Society of America*, 90, 370–386.
- Watson, M.C. & Curtis, J.E. (2013) Rapid and accurate calculation of small-angle scattering profiles using the golden ratio. *Journal of Applied Crystallography*, 46, 1171–1177.
- Weiller, A. R. (1966) Probleme de l'implantation d'une grille sur une sphere, deuxième partie. *Bulletin Géodésique*, 80, 99–111.
- Weisstein, E. W. (2002) *CRC Concise Encyclopedia of Mathematics CD-ROM*, 2nd edn. CRC Press, Boca Raton, Florida, 3252 p.
- Wells, D. L. & Coppersmith, K. J. (1994) New empirical relationships among magnitude, rupture length, rupture width, rupture area, and surface displacement. *Bulletin of the Seismological Society of America*, 84, 974–1002.
- Wesnousky, S.G. (1994) The Gutenberg-Richter or characteristic earthquake distribution, which is it? *Bulletin of the Seismological Society of America*, 84, 1940–1959.
- Wesnousky, S.G.; Scholz, C.H.; Shimazaki, K. & Matsuda, T. (1983) Earthquake frequency distribution and the mechanics of faulting. *Journal of Geophysical Research*, 88, 9331–9340.
- Wessel, P. & Smith, W. H. F. (1998) New, improved version of Generic Mapping Tools released. *Eos, Transactions, American Geophysical Union*, 79, 579.
- Wessel, P.; Smith, W.H.F.; Scharroo, R.; Luis, J. & Wobbe, F. (2013) Generic Mapping Tools: improved version released. *Eos, Transactions, American Geophysical Union*, 94, 409–410.

- Wiemer, S. (2001) A software package to analyze seismicity: ZMAP. *Seismological Research Letters*, 72, 374–383.
- Wiemer, S. & Baer, M. (2000) Mapping and removing quarry blast events from seismicity catalogs. *Bulletin of the Seismological Society of America*, 90, 525–530.
- Wiemer, S. & Schorlemmer, D. (2007) ALM: An asperity-based likelihood model for California. *Seismological Research Letters*, 78, 134–140.
- Wiemer, S. & Wyss, M. (2000) Minimum magnitude of completeness in earthquake catalogs: Examples from Alaska, the western United States, and Japan. *Bulletin of the Seismological Society of America*, 90, 859–869.
- Willemann, R. J. & Storchak, D. A. (2001) Data collection at the International Seismological Centre. *Seismological Research Letters*, 72, 440–453.
- Williamson, D. L. (2007) The evolution of dynamical cores for global atmospheric models. *Journal of the Meteorological Society of Japan*, 85B, 241–269.
- Winfield, D. C. & Harris, K. M. (2001) *Phyllotaxis-Based Dimple Patterns*. Patent number WO 01/26749 A1. World Intellectual Property Organization, Geneva, Switzerland.
- Woessner, J. & Wiemer, S. (2005) Assessing the quality of earthquake catalogues: Estimating the magnitude of completeness and its uncertainty. *Bulletin of the Seismological Society of America*, 95, 684–698.
- Wood, L. & Worfolk P. (2011) *SaVi - Satellite Constellation Visualisation Software*. [savi.sourceforge.net](https://sourceforge.net) [Last accessed October 2015.]
- Wordsworth, W. & Coleridge, S.T. (1798) *Lyrical Ballads*. J. & A. Arch, London, UK, 210 p.
- Working Group on California Earthquake Probabilities (2003) *Earthquake probabilities in the San Francisco Bay Region: 2002–2031*. United States Geological Survey Open-File Report 03-214. Menlo Park, California (USA), 234 p.
- Wyss, M. (2015) Testing the basic assumption for probabilistic seismic-hazard assessment: 11 failures. *Seismological Research Letters*, 86, 1405–1411 .
- Wyss, M. & Habermann, R.E. (1988) Precursory seismic quiescence. *Pure and Applied Geophysics*, 126, 319–332.
- Wyss, M.; Hasegawa, A.; Wiemer, S.; & Umino, N. (1999) Quantitative mapping of precursory seismic quiescence before the 1989, *M* 7.1 off-Sanriku earthquake, Japan. *Annali di Geofisica*, 42, 851–869.
- Yabuki, T. & Matsuura, M. (1992) Geodetic data inversion using a bayesian information criterion for spatial-distribution of fault slip. *Geophysical Journal International*, 109, 363–375.
- Yagi, Y.; Kikuchi, M.; Yoshida, S. & Sagiya, T. (1999) Comparison of the coseismic rupture with the aftershock distribution in the Hyuga-nada earthquakes of 1996. *Geophysical Research Letters*, 26, 3161–3164.
- Yamamoto, K. & Yabe, Y. (2001) Stresses at sites close to the Nojima Fault measured from core samples. *The Island Arc*, 10, 266–281.
- Yang, W.; & Ben-Zion, Y. (2010) An algorithm for detecting clipped waveforms and suggested correction procedures. *Seismological Research Letters*, 81, 53–62.
- Yang, X.; Bondár, I.; Bhattacharyya, J.; Ritzwoller, M.; Shapiro, N.; Antolik, M.; Ekström, G.; Israelsson, H. & McLaughlin, K. (2004) Validation of regional and teleseismic travel-time

- models by relocating ground-truth events. *Bulletin of the Seismological Society of America*, 94, 897–919.
- Yeats, R. S.; Sieh, K. & Allen, C. R. (1997) *The Geology of Earthquakes*. Oxford University Press, New York (USA), 568 p.
- Yelles-Chaouche, A.; Allili, T.; Alili, A.; Messemen, W.; Beldjoudi, H.; Semmane, F.; Kherroubi, A.; Djellit, H.; Larbes, Y.; Haned, S.; Deramchi, A.; Amrani, A.; Chouiref, A.; Chaoui, F.; Khellaf, K. & Nait Sidi Said, C. (2013) The new Algerian Digital Seismic Network (ADSN): towards an earthquake early-warning system *Adv. Geosci.*, 36, 31–38.
- Yoder, M.R.; Schultz, K.W.; Heien, E.M.; Rundle, J.B.; Turcotte, D.L.; Parker, J.W. & Donnellan, A. (2015) The Virtual Quake earthquake simulator: a simulation-based forecast of the El Mayor-Cucapah region and evidence of predictability in simulated earthquake sequences. *Geophysical Journal International*, 203, 1587–1604.
- Yoshimitsu, N.; Kawakata, H. & Takahashi, N. (2014) Magnitude -7 level earthquakes: A new lower limit of self-similarity in seismic scaling relationships. *Geophysical Research Letters*, 41, 4495–4502.
- Youngs, R.R. & Coppersmith, K.J. (1985) Implications of fault slip rates and earthquake recurrence models to probabilistic seismic hazard estimates. *Bulletin of the Seismological Society of America*, 75, 939–964.
- Zahradník, J. & Custódio, S. (2012) Moment tensor resolvability: Application to southwest Iberia. *Bulletin of the Seismological Society of America*, 102, 1235–1254.
- Zaliapin, I. & Ben-Zion, Y. (2015) Artifacts of earthquake location errors and short-term incompleteness on seismicity clusters in southern California. *Geophysical Journal International*, 202, 1949–1968.
- Zaliapin, I.; Gabrielov, A.; Keilis-Borok, V. & Wong, H. (2008) Clustering analysis of seismicity and aftershock identification. *Physical Review Letters*, 101, 018501.
- Zaremba, S. K. (1966) Good lattice points, discrepancy, and numerical integration. *Annali di Matematica Pura ed Applicata*, 73, 293–317.
- Zechar, J.D. & Jordan, T.H. (2008) Testing alarm-based earthquake predictions. *Geophysical Journal International*, 172, 715–724.
- Zechar, J.D. & Jordan, T.H. (2010a) The area skill score statistic for evaluating earthquake predictability experiments. *Pure and Applied Geophysics*, 167, 893–906.
- Zechar, J.D. & Jordan, T.H. (2010b) Simple smoothed seismicity earthquake forecasts for Italy. *Annals of Geophysics*, 53, 99–105.
- Zechar, J.D. & Nadeau, R.M. (2012) Predictability of repeating earthquakes near Parkfield, California. *Geophysical Journal International*, 190, 457–462.
- Zechar, J.D.; Schorlemmer, D.; Liukis, M.; Yu, J.; Euchner, F.; Maechling, P.J. & Jordan, T.H. (2010) The Collaboratory for the Study of Earthquake Predictability perspective on computational earthquake science. *Concurrency and Computation: Practice and Experience*, 22, 1836–1847.
- Zechar, J.D.; Schorlemmer, D.; Werner, M.J.; Gerstenberger, M.C.; Rhoades, D.A. & Jordan, T.H. (2013) Regional Earthquake Likelihood Models I: first-order results. *Bulletin of the Seismological Society of America*, 103, 787–798.

- Zhang, T.-R. & Lay, T. (1995) Why the L_g phase does not traverse oceanic crust. *Bulletin of the Seismological Society of America*, 85, 1665–1678.
- Zhuang, J.; Werner, M.J.; Hainzl, S.; Harte, D. & Zhou, S. (2011) Basic models of seismicity: spatiotemporal models. In: *Community Online Resource for Statistical Seismicity Analysis*, doi:10.5078/corssa-07487583. www.corssa.org [Last checked: October 2015.]
- Zitellini, N.; Gràcia, E.; Matias, L.; Terrinha, P.; Abreu M.A.; DeAlteriis, G.; Henriët, J.P.; Dañobeitia, J.J.; Masson D.G.; Mulder, T.; Ramella, R.; Somoza, L. & Diez, S. (2009) The quest for the Africa–Eurasia plate boundary west of the Strait of Gibraltar. *Earth Planet. Sci. Lett.* 280, 13–50.
- Zoback, M.L. (1992) 1st-order and 2nd-order patterns of stress in the lithosphere–The World Stress Map project. *Journal of Geophysical Research*, 97, 11703–11728.
- Zoback, M.D. (2006) SAFOD Penetrates the San Andreas Fault. *Scientific Drilling*, 2, 32–33.
- Zöller, G.; Holschneider, M. & Ben-Zion, Y. (2005) The role of heterogeneities as a tuning parameter of earthquakes dynamics. *Pure and Applied Geophysics*, 162, 1027–1049.
- Zöller, G.; Ben-Zion, Y.; Holschneider, M. & Hainzl, S. (2007) Estimating recurrence times and seismic hazard of large earthquakes on an individual fault. *Geophysical Journal International*, 170, 1300–1310.
- Zotov, O.D. (2007) Weekend effect in seismic activity. *Izvestiya Phys. Solid Earth*, 43, 1005–1011.
- Zúñiga, F.R. & Wiemer, S. (1999) Seismicity patterns: Are they always related to natural causes? *Pure and Applied Geophysics*, 155, 713–726.
- Zúñiga, F.R. & Wyss, M. (1995) Inadvertent changes in magnitude reported in earthquake catalogs: Their evaluation through b -value estimates. *Bulletin of the Seismological Society of America*, 85, 1858–1866.



Universidad
Zaragoza