

Magnitude frequency, spatial and temporal analysis of large seismicity catalogs: The Californian Experience

A thesis submitted to attain the degree of
DOCTOR OF SCIENCES of ETH ZURICH
(Dr. sc. ETH Zurich)

presented by
YAVOR KAMER
Master of Control and Automation Engineering,
Istanbul Technical University (Turkey)

born on February 2nd 1985
Citizen of Bulgaria and Turkey

accepted on the recommendation of

Prof. Dr. Didier Sornette	ETH Zürich	Examiner
Prof. Dr. Eduard Kissling	ETH Zürich	Co-examiner
Prof. Dr. David Jackson	UCLA	Co-examiner
Dr. Guy Ouillon	Lithophyse	Co-examiner

"Let us first try to understand what you seek in devoting yourself to science. Is it only the pleasure - doubtless immense - which we derive from the study of nature and the exercise of our intellectual faculties? In that case I ask you in what respect does the philosopher, who pursues science in order that he may pass life pleasantly to himself, differ from that drunkard there, who only seeks the immediate gratification that gin affords him? The philosopher has, past all question, chosen his enjoyment more wisely, since it affords him a pleasure far deeper and more lasting than that of the toper. But that is all! Both one and the other have the same selfish end in view, personal gratification.

But no, you have no wish to lead this selfish life. By working at science you mean to work for humanity, and that is the idea which will guide you in your investigations.

A charming illusion! Which of us has not hugged it for a moment when giving himself up for the first time to science?"

An Appeal to the Young, 1880

P. Kropotkin

Acknowledgements

I would like to thank my doktoratvater Didier Sornette for sharing his wisdom, experience and immense knowledge. For giving me the opportunity to explore diverse and controversial topics but also making sure I was not led astray. I am greatly indebted also to my advisor and “master” Guy Ouillon for all the time and patience he invested in me. His humor, humility and true passion for science will always inspire me. On several occasions during this thesis I was in a dead end and Guy was always there to dig a tunnel or blast a wall, and always make a way. I respect these two men for their willingness to leave the comfort zone of incremental science and undertake challenging endeavors.

I also thank Edi Kissling for guiding me through the arduous task of relocating the Southern Californian catalog. Whenever I left his office I always felt like a lump of forged metal; hammered, bent, scorched but slowly converging into its final form. Even though I am not a physicist, and regardless of our occasional disputes, Edi has always made me wish I was one.

I thank also Dave Jackson who was kind enough to come all the way from US to join my committee. Little as I may know him personally, I am already a fan of his eloquent style and the joy he brings to what would otherwise be boring workshop discussions.

I would also like to thank my co-author and friend Stefan Hiemer for his courage and enthusiasm during our struggle with the *b*-lievers. They told us to “be assured that it’s not over” and that they will “get to the bottom of it”. In end, while they talked the talk, together with Stefan we walked the walk.

I acknowledge Jochen Wössner for bringing up the difference between engineers and scientists during my Phd interview. At the end of my Phd I see that this division is an illusion. Science without practical implications is a mere hearsay. You are not a scientist if you can’t code your own algorithm or implement your own experiment. I would therefore like to acknowledge Mustafa Erdik from KOERI who introduced me to the field of earthquake engineering. In KOERI I felt the thrill of working somewhere where earthquakes actually matter; getting calls in the middle of the night, press conferences, scary aftershock deployments... These days here, when I drive back from Istanbul, Swiss border control asks me why my Phd is taking so long.

In this regard I want to thank Steffi, Shyam, Eszster, Max, Men, Kees, Yaming, Ylona ... for making me wish my Phd could last even longer. I want to also thank my family in Istanbul, Sofia and San Francisco for their love and support.

Lastly I want to acknowledge Hakkı, my friend from the EEG group. In my opinion he was forced to quit his Phd in an unfair conduct. This will always remain with me as a moment of helplessness in the face of injustice.

Abstract

Recent years have seen a substantial increase in the amount of available seismic catalog data. Extracting information from these data, on a case-by-case basis is gradually becoming infeasible. In order to harness the influx of data, we need new methods that are objective and data-driven. These methods need to provide the community with models that are modest in their complexity and assumptions. This thesis undertakes the task of developing such methods that can be readily applied to large seismic catalogs.

One of the main characteristics of earthquake phenomena is the power-law relation of moment release and frequency, i.e. the Gutenberg-Richter law. We developed a penalized likelihood-based method for spatial estimation of Gutenberg-Richter's b-value. We used the obtained results to generate an earthquake forecast model for California. Our forecast shows a significant information gain with respect to the competing models. These results provide evidence for small amplitude, large-scale spatial b-value variations that can, in most cases, be attributed to inconsistencies in magnitude estimations.

Another widely observed feature of seismicity is its multifractal spatial distribution. We developed an adaptive multifractal estimation method that reduces edge effects and improves precision. The method was successfully validated on synthetic fractals and used to assess the scaling properties of the widely studied diffusion-limited aggregation process. We developed also a condensation technique that allows accounting for the location uncertainty information in spatial analysis. We utilized these two approaches to investigate the spatial distribution of Southern Californian seismicity. Our analysis provided evidence for possible issues with the relocation procedure of the current catalog. This motivated us to re-locate the last three decades of Southern Californian seismicity by employing current state-of-the-art probabilistic non-linear methods. The outcome of these efforts is the new KaKiOS-15 catalog that contains the full expression of each event's location uncertainty.

Lastly, we developed a new fault network reconstruction method based on agglomerative clustering. We applied this method to the events in the condensed KaKiOS-15 catalog and obtained a fault network for Southern California. In retrospective tests, we were also able to demonstrate that the obtained fault network has significantly higher forecasting skills compared to commonly used methodologies.

Резюме

Размерът на достъпни сеизмични данни се е увеличава значително през последните години. Добиване на информация от тези данни, за всеки случай поотделно постепенно става невъзможно. За да може да се възползваме от притока на тези данни, има нужда от нови методи, които са обективни и адаптивни. Тези методи трябва да предоставят сейсмологичната общност с модели, които са скромни в техните сложности и предположения. Тази теза се отнася до разработването на такива методи, които могат лесно да бъдат приложени за големи сеизмични каталоги.

Една от основните характеристики на земетръсните явления е отношението между тяхната енергия и честота, т.е. закона на Гутенберг-Рихтер. Разработихме метод, основан на санкционирана вероятност, за пространствено изчисление на б-стойност по закона на Гутенберг-Рихтер. Използвахме получените резултати за да генерираме прогнозен модел за земетресенията в Калифорния. Нашата прогноза достига значително подобреие спрямо конкурентните модели. Тези резултати предоставят доказателство че пространствените вариации на б-стойност са широко мащабни и с малък амплитуд. Те могат, в повечето случаи да се обяснят с несъответствия в измерванията на магнитуд.

Една от другите често наблюдавани особености на сейсмичността е нейното мултифрактално пространствено разпределение. Разработихме адаптивен метод за изчисление на мултифракталния спектър, който намалява периферичните ефекти и подобрява точността. Методът бъде успешно утвърден чрез изчисляването на скалиращите свойства на различни синтетични фрактали. Също така, разработихме кондензационен метод, който дава възможност за отчитане на локационното неопределение в пространствените анализи. Използвахме тези два метода за да изследваме пространственото разпределение на сейсмичността в Южна Калифорния. Нашият анализ предоставя индикации за проблеми относно локализационната процедурата на текущия каталог. Това ни мотивира да локализираме сейсмичността в Южна Калифорния отначало, използвайки съвременните нелинейни и вероятностни методи. Резултатът на тези усилия е новият каталог KaKiOS-15, който съдържа пълния израз на локационното неопределение за всяко заметресение.

Наи последно, разработихме нов метод за разломна реконструкция основан на обединителено групиране. Приложихме този метод към кондензираните земетресенията в каталога KaKiOS-15 и получихме разломна мрежа за Южна Калифорния. В ретроспективни изследвания, можахме да демонстрираме, че разломната мрежата има значително по-високо умение за прогнозиране в сравнение с често използваните методологии.

Özet

evc ut sismi veri mi tarı son yillarda ne mli l de artmı tır. e r bir va a elinde ayrı ayrı bilgi tl eme gidere olana si hale gelme tadir. u amansı veri a inı ndan istifade edebilme i in veri g d ml ob e tif y ntemler e ihtiyac vardır. u y n temlerin, a rma ı l ı ve varsayımda m t eva ı modeller ortaya o yabilmesi gere ir. u te er evesinde, ola ylı la b y sismi ata loglara uygulanabilece bu t r y ntem ler geli tirmeye gayret etti .

e prem olayının temel elli l erinden biri ut enberg-Richter a nunu, ba a bir deyi le deprem ener isi ve fre a nsı arasında i g yasası ili i sidir. Bu tezde, Gutenberg-Richter a nununda i b de e rinin me a nsal e stirimi i in, d eltilmi olabilirli temelli bir y ntem gelidi tirdi . lde edilen sonu ları ull anara Kaliforniya i in deprem tahmin modeli olu turdu . odelim i ra ip modellere ı yasla belirgin bir tahmin a biliyeti g s terdi. u sonu lar, b- de e rende i me a nsal far lılı lar in genli li ve b y ı e li oldu unu i aret etme tadir . u far lılı lar in o unlu la fi i sel olma tan te , deprem b y l ı m lerinde i tutarsı li lar dan a yna landı ina dair bulgular elde etti .

Sismisitenin yaygın olara g lemlenen bir di er elli i multifractal me a nsal da ılimdir. una y ne li olara enar et i lerini a altan ve e stirim do ru lu unu iyile ti ren adaptif bir multifraktal kestirim y nt emi gelidi tirdi . Y ntemi senteti fra tallar erinde uygulayara ba a rıyla do ruladı . olara , me a nsal anali esnasında onum belirsi li verisinin hesaba a tilmasını sa layan bir yo unla tırma te ni i gelidi tirdi . ne y Kaliforniya sismisitesinin me a nsal da ılimini inceleme i in bu i i y n temi uyguladı . u anali sonucunda, mevcut a talogta ull anilan on um belirleme prosed r n de sorunlar oldu u y n nde bulgulara ıla tı . Bu bulgular, bi i son otu yılda meydana gelen ney Kaliforniya depremlerini yeni ba tan onuml andırmaya te vi etti. o rusal olmayan, olasılı sal onu mlandırma ya ılimlarını ull anara her deprem i in a psamlı bir onum belirsi li verisinin yer aldı ı yeni KaKiOS-15 a talo unu olu turdu .

Son olara , agglomeratif mele nmeye dayalı yeni bir fay a ı re onst r si yon y ntemi gelidi tirdi . u y ntemi yo unla tı rılmı KaKiOS-15 a talo una uygulayıp ne y Kaliforniya i in bir fay a ı elde etti . Retrospe ti f testler, elde edilen fay a inın yaygın olara ull anilan y ntemler e ı yasla daha y se tahmin becerilerine sahip oldu unu g ster mi tır.

Résumé

La quantité de données disponibles dans les catalogues de sismicité a récemment substantiellement augmenté. Extraire de l'information de ces données devient graduellement impossible. Afin d'exploiter ce flux de données, nous avons besoin de nouvelles méthodes objectives et dirigées par les données. Ces méthodes doivent procurer à la communauté des modèles aux hypothèses et à la complexité modestes. Cette thèse entreprend la tâche de développer de telles méthodes qui peuvent être aisément appliquées à de grands catalogues sismiques.

Une des caractéristiques principales des séismes est la relation en loi de puissance du moment relâché et de sa fréquence, i.e. la loi de Gutenberg-Richter. Nous avons développé une méthode basée sur une pénalisation de la vraisemblance afin d'estimer la variation spatiale de la valeur de b de la loi de Gutenberg-Richter. Nous avons utilisé les résultats pour générer un modèle de prévision de l'aléa sismique pour la Californie. Notre prévision montre un gain d'information significatif par rapport aux modèles concurrents. Ces résultats montrent des variations de b à grande échelle, mais de faibles amplitudes, qui peuvent la plupart du temps être attribuées à une incohérence dans l'estimation des magnitudes.

Une autre caractéristique fréquemment observée de la sismicité est sa distribution spatiale multifractale. Nous avons développé une méthode adaptative d'estimation du spectre multifractal qui réduit les effets de bord et améliore la précision. La méthode a été validée avec succès sur des fractals synthétiques et utilisée afin d'évaluer les propriétés d'échelle du célèbre processus d'aggrégation limité par la diffusion. Nous avons également développé une technique de condensation qui permet de tenir compte de l'incertitude des localisations lors d'analyses spatiales. Nous avons utilisé ces deux approches pour afin d'étudier la distribution spatiale de la sismicité de Californie du Sud. Cette analyse indique des problèmes possibles lors de la re-localisation du catalogue actuel. Ceci nous a conduit à ré-estimer la localisation de la sismicité Sud-Californienne des trois dernières décennies en employant des méthodes probabilistes non-linéaires sophistiquées. Le produit de ces efforts est le nouveau catalogue KaKiOS-15 qui contient une description complète de l'incertitude de localisation pour chaque événement.

Enfin, nous avons développé une nouvelle méthode de reconstruction de réseaux de failles basée sur le regroupement agglomératif. Nous avons appliqué cette méthode aux événements du catalogue KaKiOS-15 condensé, et avons obtenu un réseau de failles pour la Californie du Sud. A l'aide de tests rétrospectifs, nous avons également démontré que le réseau obtenu permet de meilleures prévisions comparé aux autres méthodes communément utilisées.

Curriculum Vitae



Yavor Kamer

born 1985, Pleven, Bulgaria

- 2003-2007 B.Sc in Control Engineering, Istanbul Technical University, Turkey
- 2007-2010 M.Sc in Control and Automation Engineering, Istanbul Technical University, Turkey
- 2010-2012 Phd in a rthqua e n gineering, o a i i Unive rsity, Tur e y (withdraw)
- 2012-2015 Phd in Earthquake Forecasting, ETH Zürich, Switzerland

Publications

Y. Kamer, G. Ouillon, D. Sornette, and J. Wössner (2015), *Condensation of earthquake location distributions: Optimal spatial information encoding and application to multifractal analysis of south Californian seismicity*, Phys. Rev. E, 92(2), 022808

Y. Kamer, S. Hiemer (2015) *Data-driven spatial b-value estimation with applications to California seismicity: To b or not to b* JGR Solid Earth, 120(7), 5191-5214.

Y. Kamer, (2014) *Comment on "Systematic survey of high-resolution b value imaging along Californian faults: Inference on asperities" by T. Tormann et al.* JGR Solid Earth, 119, 5830-5833

Y. Kamer, S. Hiemer (2013) *Comment on “Analysis of the b-values before and after the 23 October 2011 Mw 7.2 Van-Ercis, Turkey, Earthquake”* Tectonophysics, 608: 1448-1451

Y. Kamer, G. Ouillon, D. Sornette (2013) *Barycentric fixed-mass method for multifractal analysis* Physical Review E, 88: 022922

Y. Kamer, S. Ikizoglu (2013) *Effective accelerometer test beds for output enhancement of an inertial navigation system* Measurement, 46-5: 1641-1649

M. Erdik, Y. Kamer, M. Demircioglu, K. Sesetyan, (2012) *23 October 2011 Van (Turkey) earthquake* Natural Hazards, 64-1: 651-665

K. Sesetyan, C. Zulfikar, M. Demircioglu, U. Hancilar, Y. Kamer and M. Erdik. (2011) *Istanbul Earthquake Rapid Response System: Methods and practices.* Soil Dynamics and Earthquake Engineering, 31(2):170 - 180

Contents

Acknowledgements	vii
Abstract	viii
Резюме	ix
Özet	x
Résumé	xi
Curriculum Vitae.....	xii
Contents	xiv
List of Figures	xviii
Chapter 1 Introduction	27
Chapter 2 Data-driven spatial b-value estimation with applications to California seismicity: To b or not to b	31
2.1 Abstract	32
2.2 Introduction	32
2.3 Traditional spatial b-value mapping.....	33
2.4 A penalized likelihood-based method	35
2.5 Application to synthetic datasets.....	38
2.5.1 Uniform event distribution	38
2.5.2 Fractal event distribution	39
2.5.3 Convergence with increasing sample size	41
2.6 Application to Californian seismicity	42
2.6.1 Results for the time period of 01.01.1984 – 01.01.2014	43
2.6.2 Results for the time period of 01.01.1984 – 01.01.2014	46
2.7 Discussion	47
2.7.1 Magnitude inconsistencies between local networks	47
2.7.2 Bias and errors in magnitude estimations	49
2.8 Conclusions	51
2.9 Appendix A: Comparative Synthetic Test Based on Californian Seismicity	54
2.10 Appendix B: How plausible is the GR law for Californian seismicity?	56

2.11 Acknowledgements	58
Chapter 3 Improved seismicity forecast with spatially-varying magnitude distribution.....	61
3.1 Abstract	62
3.2 Introduction	62
3.3 Method	64
3.3.1 The Helmstetter et al. (2007) approach	64
3.3.2 The Wiemer and Schorlemmer (2007) approach.....	66
3.3.3 A novel approach for probabilistic seismicity forecasting	67
3.4 Results	70
3.5 Conclusions	72
3.6 Discussion	74
3.7 Acknowledgements	75
Chapter 4 The Barycentric Fixed Mass Method for Multifractal Analysis	77
4.1 Abstract	78
4.2 Introduction	78
4.3 The Barycentric Fixed Mass Method.....	79
4.3.1 Review of multifractal analysis methods.....	79
4.3.2 Barycentric Pivot Selection	80
4.3.3 Non-Overlapping Coverage	81
4.4 Multifractal spectra of synthetic datasets	83
4.4.1 First example with the synthetic multifractal Sierpinski triangle of Figure 4-4	83
4.4.2 Multifractal spectra of synthetics datasets	84
4.5 Application to the Diffusion-Limited Aggregation (DLA) process.....	88
4.6 Conclusions	92
Chapter 5 Condensation of earthquake location distributions: optimal spatial information encoding and application to multifractal analysis of South Californian seismicity	93
5.1 Abstract	94
5.2 Introduction	94
5.3 The Condensation Method	96
5.3.1 Description and illustration of the method	96
5.3.2 Synthetic Test with Fractal Distributions	98

5.4	Application to South Californian Seismicity	100
5.4.1	Condensing the catalog.....	100
5.4.2	Visualizing and quantifying weight transfer due to condensation.....	101
5.5	Multifractal properties of the original and condensed catalogs	104
5.5.1	The multifractal formalism and classical estimation methods	104
5.5.2	The Barycentric Fixed Mass estimation Method.....	106
5.5.3	Multifractal analysis of the Southern Californian Seismicity with robust estimation of the different scaling regimes.....	107
5.5.4	Multifractal analysis of the Southern Californian Seismicity: condensed versus original catalogs.....	109
5.5.5	Multifractal analysis of a multiscale synthetic dataset	111
5.6	Discussion	112
5.6.1	Consequences for the spatial distribution of earthquake loci	112
5.6.2	Consequences for earthquake triggering models.....	113
5.7	Conclusion.....	116
5.8	Acknowledgments.....	117
Chapter 6 Probabilistic, non-linear, absolute relocation of the 1981-2011 South Californian seismicity	119	
6.1	Abstract	120
6.2	Introduction	120
6.3	A Minimum 1D Velocity model for Southern California	121
6.3.1	Data selection.....	121
6.3.2	Inversion results and validation	122
6.4	Hypocenter relocation using NonLinLoc	125
6.5	Comparison of the KaKiOS-15 and HYS-12 catalogs.....	127
6.5.1	Vertical and Horizontal shifts	127
6.5.2	Multifractal analysis of the two catalogs	129
6.6	Conclusion.....	132
Chapter 7 Fault Network Reconstruction using Agglomerative Clustering: Applications to South Californian Seismicity	133	
7.1	Abstract	134
7.2	Introduction	134
7.3	The agglomerative clustering method	135
7.3.1	Recent developments in fault reconstruction.....	135

7.3.2 Method description	136
7.4 Application to seismicity.....	139
7.4.1 Small Scale application to Landers aftershocks	140
7.5 Condensation of the KaKiOS-15 catalog.....	141
7.6 Large scale application to Southern California.....	144
7.7 Validation through a spatial forecast test	146
7.8 Conclusion.....	148
Chapter 8 Conclusion and Future perspectives	149
8.1 Achieved results	149
8.2 Future perspectives.....	151
References	153

List of Figures

- Figure 2-1 a) Underlying spatial b-value distribution for the generation of synthetic magnitudes for 8000 uniformly distributed events. b) and c) Resulting b-value maps using the classical nearest neighbor mapping method for $N_{\min} = 150$ and $N_{\min} = 1000$, respectively. d) and e) Resulting b-value maps based on the penalized likelihood approach for 50,000 randomly generated models in the complexity range of 1 - 100 Voronoi nodes (showing the median of the 100 and 1000 best performing models, respectively.) Ranges of resulting b-values are given in the title of each subplot..... 38
- Figure 2-2 Fractal ($D=1.59$) distribution of synthetic events. Symbol size scales with event size. Color-coding depicts the underlying b-value: 6000 background events were drawn from a GR law using $b_{\text{bkg}}=1.0$, and 2000 events from $b_{\text{ano}}=0.75$. The boundaries of the anomaly are 0.25/0.75 in both x- and y-direction. b) Cumulative frequency-magnitude distributions and estimated b-values for both regions..... 39
- Figure 2-3 a) Ensemble median b_{med} and b) ensemble interquartile range (IQR) of the best 102 solutions. The median spatial b-value within the region of interest ROI is 0.79 (dashed line, compare Fig. 1). c) 14500 randomly generated solutions are shown in terms of their BIC (black dots) and their average negative log-likelihood (gray line, error bar corresponds to one standard deviation). Red circles highlight the best 102 solutions. d) Best ten individual models. Black dots denote the positions of the Voronoi nodes; resulting partitions are color-coded by their estimated b-values..... 41
- Figure 2-4 Convergence of the method for four different b-value anomalies (0.5, 0.75, 1.25 and 1.5) with varying sample sizes (from 100 to 5000). The three panels show results for different sample size ratios with respect to the surrounding events. Solid lines and shaded areas denote the corresponding median b-value and 0.16-0.84 quantiles, respectively. 42
- Figure 2-5 a) Entire solution space for the ANSS catalog (1984-2014, $m \geq 2.2$): Average BIC (black line) and negative log-likelihood (gray line) as a function of number of Voronoi cells. Error bars correspond to one standard deviation. Red circles denote the BIC of the best 1000 solutions. b) and c) show the corresponding ensemble median b-value (in the range of 0.78-1.28) and ensemble IQR (in the range of 0.02-0.65), respectively, computed at all 1x1km cells with more than one event..... 44
- Figure 2-6 Same as Figure 2-1 for the last decade of the ANSS catalog (2004-2014, $m \geq 2.2$). Ranges for median b-values and IQRs are [0.79- 1.15] and [0.01- 0.35], respectively..... 44

Figure 2-7 Average Voronoi areas for a) random placement of Voronoi seeds and for b) Voronoi seed positions of the best 1000 solutions, white circles highlight the Mendocino Triple Junction and Los Angeles c) The ratio of averaged areas (b/a) highlights regions covered by smaller (<1.0) and larger (>1.0) Voronoi cells with respect to random placement. 45

Figure 2-8 Median b-values from Figure 2-6b, median cut-off magnitude m_c and median number of events with $m \geq m_c$ (N_{comp}) as a function of longitude. Dark and light gray bars correspond to percentiles of [0.25 - 0.75] and [0.16 - 0.84], respectively. Top right inset shows Californian seismic stations, colored according to their network affiliation. Parts of Nevada and Southern California networks cover the same longitude range, thus we show values from the Southern California network separately (right panels). 47

Figure 2-9 Magnitude and b-value comparison using matching events from a) NCEDEC-SCEDC, b) NCEDEC-Nevada, and c) Nevada-SCEDC. Top panels show magnitudes and magnitude probability density functions. Black line is a reference for equal magnitudes. Bottom panel shows b-value with interquartile range as a function of cut-off magnitude. Dashed lines denote the mode of each magnitude distribution. 48

Figure 2-10 Distribution of magnitude error (estimated-true) for different network configurations assuming a P pick error of 0.1s and $V_p = 6$ km/s. Stations are shown with red triangles, true epicenters with a red cross and samples from the location PDF with white dots. Low azimuthal coverage results in large magnitude errors and bias: overestimation in a) and underestimation in c). Moving a single station to increase the azimuthal coverage (shown in b) and d)) reduces the error and bias significantly in both configurations. 50

Figure 2-11 a) Box plot of median absolute deviation versus magnitude for the Northern California NCEDEC catalog, 1984-2014, binned at M0.5 intervals. The corresponding number of events with reported MAD values is given in the top. Box height corresponds to the IQR and whiskers extend to the most extreme values within 1.5 IQR in both directions. b) Schematic plot of the bias in the estimated GR relation in the presence of magnitude uncertainties (adapted from [Rhoades and Dowrick, 2000]). 51

Figure 2-12 Ensemble median b_{med} (a) and IQR(b) for the synthetic dataset based on the [2004-2014] catalog with magnitudes generated from the annotated b-values in each zone. The remaining magnitudes are generated from $b=1$. c) The entire solution space for the synthetic catalog: BIC scores (black dots) and negative log-likelihood (gray line) as a function of number of Voronoi cells. Error bars correspond to one standard deviation. Red circles denote the BIC of the best 80 solutions. 55

Figure 2-13 Same synthetic dataset as Figure 2-12 analyzed via constant sized non-overlapping grids with increasing sizes from 10 to 300 km with $N_{min}=50$. The resulting BIC scores and b-value ranges are given in the title of each grid plot. 56

Figure 2-14 Median p -values obtained by 1000 Monte-Carlo simulations with the top ranking partitions for the 2004-2014 California catalog (Fig. 6). The p -value represents the probability of observing a worse GR law fit (in terms of Kolmogorov-Smirnov distance) estimated from the Monte-Carlo simulations. Values lower than $p_{min}=0.1$ indicate implausibility of the GR law.....	57
Figure 2-15 Median p -values from Fig. B1 as a function of longitude. Dark and light gray bars correspond to percentiles of [0.25-0.75] and [0.16-0.84], respectively.....	58
Figure 3-1 Spatial distribution within R L 's testing area of the expected $m \geq 5$ earth- quake rate λ for the 5-years target period of 2006 - 2011 for the Wiemer and Schorlemmer (2007) and Helmstetter et al. (2007) forecast. Small dots label cells whose rates were defined by the “water-line” criteria.....	64
Figure 3-2 Entire solution space for the declustered 1984 - 2005.5 ANSS catalog (as used by Wiemer and Schorlemmer, 2007). Average BIC (black line) and negative log-likelihood (gray line) are shown as a function of number of Voronoi cells. Error bars correspond to one standard deviation. Crosses denote the BIC of the best 400 solutions and black circles highlight values for the two models shown in Figure 3-3.....	68
Figure 3-3 Map view of two individual tessellations. Black dots denote location of Voronoi seeds. The two models represent solutions of relatively low and high complexity, respectively. Corresponding BIC scores are given in the title and in Figure 3-2.....	69
Figure 3-4 Top panels show a close-up of a single Voronoi region (see top left corner in Figure 3-3) and the corresponding frequency-magnitude distribution. Dashed line denotes cut-off magnitude. Bottom panels visualize results of different techniques for modelling the spatial distribution of seismicity. λ is the estimated number of expected $m \geq 5$ events as derived from the R parameters. Note that empty gray squares indicate $\lambda = 0$. The total sum $\Sigma\lambda$ over all spatial cells in a single Voronoi region is identical for the different techniques.....	69
Figure 3-5 Same as Figure 1 for the ensemble forecasts using the top 400 individual models (crosses in Figure 3-2). Notice that the relative intensity based forecast involves cells with zero expectation (compare Figure 3-4B).....	70
Figure 3-6 Rate-corrected average information gain per RELM target mainshock (20 $m \geq 4.95$ events in 2006 - 2010) with respect to the Wiemer & Schorlemmer model. Confidence intervals are derived from T-test statistics, and letter T highlights significance for positive information gains. Note that the number of target events varies for different RELM member models due to differences in spatial coverage (for corresponding numbers see Figure 5 in Zechar et al., 2013).	71
Figure 3-7 Rate-corrected average information gain per RELM target mainshock for the individual models that were used to construct the ensemble forecasts (Figure 3-2 and Figure 3-5). Top panels show results with respect to the Wiemer & Schorlemmer model (for ensemble T-scores see Figure 3-6) and bottom panels	

show results with respect to the Helmstetter model (the T-scores for the ensemble models are: A=-1.22±0.50, B=0.25±0.24, C=0.10±0.29, and D=0.17±0.23). Inset text boxes quantify the fractions of models that feature positive information gains/ significantly positive information gains/ zero-rate cells with target events. Notice that 8% of the forecasts in B failed the spatial consistency test due to target events occurring in such zero expectation cells.....	72
Figure 4-1 Illustration of the barycentric pivot point selection criterion.....	80
Figure 4-2 Minimum overlap of circles covering an area	82
Figure 4-3 Generation of the synthetic multifractal measure shown in Figure 4-4 from the density matrix given by the 2x2 table on the left.....	82
Figure 4-4 Coverage of a synthetic multifractal Sierpinski triangle with fixed-mass circles	83
Figure 4-5 a) Averaged radii versus fixed-mass for increasing τ ; b) Analytical and estimated D_q -q curves for the multifractal Sierpinski triangle obtained recursively with the density generator [1 0; 1 2] as explained in the text and with Figure 4-3.....	84
Figure 4-6 Benchmark results for regular and random synthetic distributions	86
Figure 4-7 D_q -q curves estimated for 100 DLA clusters with increasing masses given in the inset with different colors	89
Figure 4-8 Convergence of D_q versus the number of samples in DLAs.....	90
Figure 4-9 D_q -q curves obtained for mono and multifractal toy DLA synthetics	91
Figure 4-10 Comparison with D_q -q curves reported by Hanan et. al. 2012 Vicsek et. al. 1990.....	91
Figure 5-1 Illustration of the condensation procedure for a set of 1D distributions. These are labeled with letters A to E according to their standard deviations [0.5, 1, 1.5, 2, and 2.5]. At each step, source events loosing mass are represented by downward arrows while target events gaining mass are labeled with upward arrows. The portion of the probability weight assigned to each event is depicted with its respective color	97
Figure 5-2 Synthetic distributions of 3360 points with different fractal dimension (D). Each distribution is generated by iteratively replicating and permuting the 3D template matrices given in the upper left corners. L and M denote the factors of length reduction and mass increment per iteration, where $\log(M)/\log(L)=D$. See Kamer et al. (2013) for details.....	98
Figure 5-3 Log likelihood gain of the condensed catalog with respect to the original catalog. Each curve corresponds to a different distribution with a given fractal dimension D, calculated for an increasing number of events. b) All curves, except D=0, plotted against the minimum spacing calculated from Equation (5-3).	100

Figure 5-4 Probability density distributions of vertical, horizontal and isotropic errors of the original and condensed catalogs: first row of panels for the whole catalog; second row of panels for events with ≥ 2	101
Figure 5-5 Areas of weight enrichment (red or dark gray) and depletion (green or light gray) comparing a) the original and condensed catalogs of Southern California, b) two uniform random spatial PDFs with similar extends for Gaussian filters with bandwidths of $\sigma=1$ and $\sigma=5$ m. c) Percent of displaced weight as a function of filter bandwidth σ	102
Figure 5-6 Areas of weight enrichment (red or dark gray) and depletion (green or light gray) resulting from condensation at bandwidth of 3km, superimposed with the fault traces obtained from the Community Fault Model. Individual faults are labeled with the following abbreviations: San Andreas (SAn), Santa Cruz (SCr), San Clemente (SCI), San Jacinto (SJc), Brawley (Brw) and Laguna Salada (LgS).	104
Figure 5-7 D_q vs q curves for two synthetic multifractal point distributions given in the upper right insets: A multifractal Sierpinski triangle (left) and Sierpinski carpet (right). Analytically derived true D_q (solid line) is shown together with the BFM method (diamonds), fixed-size (squares) and fixed-mass (circles) methods. Reproduced from Figure 6 of Kamer <i>et al.</i> [2013]	106
Figure 5-8 Averaged radii versus fixed-mass for increasing τ obtained from a single measurement (left) and averaged over 100 measurements (right) of the ≥ 2 Southern Californian seismicity. The two arrows mark the transitions from small to medium scales (R_{S-M}) and medium to large scales (R_{M-L}).	107
Figure 5-9 Best fitting piecewise linear models with a) 3 and b) 6 segments fitted to the moment curves shown in Figure 5-8. Different segments are represented by different colors (shading). The staircase slope functions obtained for the best fitting models shown in a,b) are shown in c,d).....	109
Figure 5-10 Continuous multifractal spectra of the a) original and b) condensed catalogs. The horizontal axis indicates the increasing radius (bottom) and mass (top); the vertical axis represents the q value and individual colors represent D_q within the range [0-3]. The two arrows mark the transitions from small to medium scales (R_{S-M}) and medium to large scales (R_{M-L}).	110
Figure 5-11 a) Spatial distribution of the synthetic distribution, inset shows a close up with the small scale linear features. b) Radii vs mass curves for different increasing τ exponents c) Continuous multifractal spectra of the distribution (similar to Figure 10); insets show the replicating density matrices of each scaling regime together with the analytically calculated D_0 , D_1 and D_2 values....	112
Figure 6-1(Left) Routine epicenter locations colored according to the number of P observations within 100 km of their epicenters. (Right) Colored according to the azimuthal gap.....	122
Figure 6-2 Training set of events selected as the most well-locatable events within a 6.3km cell grids.....	122

Figure 6-3 Top: Initial (squares) and final (crosses) values of Vp velocities for different models. Bottom: Progression of the RMS with increasing iteration numbers. Final RMS values are shown in the legend.....	123
Figure 6-4 Same as Figure 6-3 obtained for the YKEK05 initial model using the four validation sets.....	124
Figure 6-5 Left: Station corrections shown as O (negative) and + (positive) signs scaled according to their amplitude. Right: The same station corrections superimposed over the geological map of California.....	125
Figure 6-6 Same as Figure 6-3. Red: No pick quality information; blue: randomly assigned pick quality; green: qualities assigned according to onset type (impulsive=0, emergent=1).....	126
Figure 6-7 A NonLinLoc location PDF sampled with 1000 points shown as crosses superimposed on the optimal BIC Gaussian mixture model with components weights 0.59, 0.31 and 0.09 (darker shading for larger weights). Gaussian ellipsoids are plotted at $\pm 1\sigma$ level.....	127
Figure 6-8 Horizontal and vertical shifts PDFs between the HYS-12 and KaKiOS-15 catalogs, calculated with respect to the later.....	128
Figure 6-9 Hypocentral depth distribution of the HYS-12 and KaKiOS-15 catalogs. Gray dashed lines indicate the velocity model layer boundaries.	128
Figure 6-10 Averaged radii versus fixed-mass curves for $\tau = [-3, 6]$ averaged over 100 measurements. Left panel: the HYS-12 catalog; right panel: the KaKiOS-15 catalog.....	129
Figure 6-11 Continuous multifractal spectra of the HYS-12 (top) and KaKiOS-15 (bottom) catalogs. The horizontal axis indicates the increasing radius; the vertical axis represents the q value and individual colors represent D_q within the range [0-3]	130
Figure 6-12 (Left) Multifractal spectra of three different uniform synthetic distributions with dimensions 400x400x16km. Insets show the point densities with color correspondence in the right panel. (Right) The radius versus mass curves plotted with their respective colors. Inset shows a 20x20km sample region of densest catalog.....	131
Figure 6-13 Continuous multifractal spectra of the KaKiOS-15 catalog, using all the individual Gaussian components of each earthquake.....	132
Figure 7-1 Iterative splits on the 1992 Landers aftershock data. Points with different colors represent seismicity associated with each plane. Black dots show the center points of the planes resulting from the next split. Notice how in the 2nd step the planes fail to converge to the local branches (shown with arrows).	135
Figure 7-2 Top: Synthetic fault network created by uniform sampling of 5 faults, each shown with a different shaded according to its total number of points. Empty dots represent the uniform random background. Bottom: The number of initial and final (undiscarded) cluster for the case with and without background noise.	137

Figure 7-3 Left: The initial <i>protoclusters</i> for the synthetic dataset given in Figure 7-2. Notice that the number of clusters includes the uniform background kernel as well.	138
Figure 7-4 The final models obtained using the local (left) and global (right) criteria. Notice that the number of clusters includes the uniform background kernel as well	139
Figure 7-5 Left: Top view of the 1992 Landers aftershocks. Fault networks obtained from these event using the local (center) and global (right) merging criterion.....	140
Figure 7-6 Two uniform distributions (dotted gray lines), their Gaussian approximations (solid gray lines) and the Gaussian resulting from their merger (solid black line). Notice that the joint Gaussian has higher densities at the tails compared to its constituents.....	141
Figure 7-7 Idealized schematic representations of a) Condensation: each event is represented by a different shade, weight is transfer is represented by the arrows, notice that there are no intra-event weight transfers b) Final condensed catalog: total weight sum is preserved, one component is discarded. c) Sampling of the event PDFs: this step is independent of the condensed catalog d) Maximum likelihood assignment of the three events onto two of the condensed.....	142
Figure 7-8 Cumulative weights of the 600,463 condensed KaKiOS-15 components representing a total of 475,371 events. The components are ranked according to increasing weights.....	143
Figure 7-9 Left: Mean locations of condensed 600,463 Gaussian components shaded according to their weights. Right: The same components shaded according to the total number of events assigned to them after the maximum likelihood assignment	143
Figure 7-10 Fault network reconstructions for the KaKiOS-15 catalog. Top row shows results for the case of 5 initial subsets with local (left) and global (right) merging criterion. Bottom row shows the same for 30 initial subsets. The number of cluster, background weight and BIC per data point is given in the titles.145	
Figure 7-11Average Negative Log Likelihood for the target dataset limited to events above M2.5 (light gray), M3.0 (dark gray) and M3.5 (black). Performance of the TripeS models is evaluated as function of the isotropic kernel bandwidth (dotted lines). The fault network performance is shown with constant level solid lines.....	147

Chapter 1 Introduction

Science, being a mainly public funded enterprise, strives to yield a benefit to its funders, i.e. the public. For some fields the immediate goals of this strive may be more obvious than others; compare medical research (e.g. vaccine development) with particle physics (e.g. hadron collider experiments). In geophysics, one can even find fields that seem to work against one another; geophysicists toiling away to advance hydrocarbon exploration are often under the same roof with their colleagues who try to convince the public of the immediate effects of green house gases and global warming. The field of natural hazards provides a safe haven from such moral dilemmas. The goals are clear; understanding, modeling and ultimately forecasting the deadly wrath of Nature.

These goals have proven relatively easier to achieve for some natural disasters than others. For instance, meteorological models benefit from extensive observational techniques together with well established precursory signals (low barometric readings, seasonal correlation, etc). Volcanoes undergo observable changes (uplift, increased seismic activity) while landslides can be correlated with regions with heavy rain falls and unstable slopes. Once such causal relationships are posed and established, public funding provides leverage by focusing wider research effort, more accurate instrumentation and better data coverage. In most cases these investments yield to better models and ultimately better forecasts of the studied natural disaster.

Unfortunately, this has hardly been the case in the study of earthquakes. Today we have the highest number of geodesic/seismic stations, the widest station coverage, the largest seismic catalogs and the highest number of researchers working in the field than we ever had before; but our understanding and hence our ability to forecast earthquakes has seen little improvement over the last century. Despite extensive research efforts there is still an absence of a distinct characteristic or variable that can be associated with the magnitude of an impending earthquake. Our best forecasting models still rely on empirical laws posed more than half a century ago, namely the aftershock rate decay (Omori's law) and the magnitude frequency distribution (Utsu's law). The resulting information gains are minute and hence they cannot warrant any short-term precautionary measure or be used for any decision making process.

There are several possible reasons why meaningful earthquake forecasting has remained elusive. Some of these reasons are due to the distinct nature of the earthquake phenomena. The frequency of earthquakes versus their released energy follows a well defined power law meaning that large and devastating events occur rarely leading to scarce observations. When observations are available, the information regarding the actual process is usually limited to a triplet of location, time and magnitude, which are often of unassessed or poor quality. Nevertheless, this single frequency energy distribution observed through multiple scales implies that the relationship between earthquakes is self-similar, lacking distinct grouping with characteristic interaction scales or magnitudes. This property of earthquakes leaves us with the prospect of studying small earthquakes and inferring statistical and physical properties of the larger ones. Due to the same self-similar

distribution, we can expect to expand our catalogs by tenfold for a unit of magnitude improvement in our detection threshold. While this aspect should in principle empower the research community, in practice it has led to two distinct approaches to data analysis that in turn lead to irreconcilable results.

The simplistic approach maintains that as data accumulates the empirical laws observed globally become more robust, supporting a universal behavior and deeming deviations as insignificant results of statistical fluctuations. Whereas the pluralists, who view nature as infinitely complex, maintain that more data reveals more complexity. Hence this latter approach tends to focus on exceptions of the empirical law and tries to collect evidences based on a case studies rather than large scale analysis. Ultimately, the main distinction between these two approaches and the reasons for their conflicting results stem from the different ways chosen to treat the same data. As illustrative examples, we can consider the studies by Kagan who reports a stable universal Gutenberg-Richter law studying large global datasets [Kagan, 1999b, 2002]. Pisarenko and Sornette (2003) confirmed these results through an extensive investigation of the distribution tails, with the exception of mid-ocean ridges where they found significant differences in the distribution. Similar results of universal scaling have been reported using a local but large dataset of seismicity in a South African mine [Boettcher *et al.*, 2009]. These results contrast with several reports of spatial and temporal variation observed by studying smaller sample sizes [Wiemer and Wyss, 1997; Wyss *et al.*, 2000; Tormann *et al.*, 2014]. Similar dissidence is found while studying the fractal dimension of earthquake locations: while some report stable scaling regimes for Californian and Japanese earthquakes [Hirata, 1989], other researchers find fluctuations which seemingly correlate with deviations from the Gutenberg-Richter law [Enescu and Ito, 2001; Wyss *et al.*, 2004].

Neither approach can be viewed as correct since their final conclusions seem to be dependent on their initial assumptions about the origin of the data while, ideally, it should be the opposite way, data should be processed with as few assumptions as possible and the results should provide evidence in favor or against the competing hypothesis. Faced with this dichotomy, in this Phd thesis we tried to advance alternative approaches that rely on data-driven techniques, information criteria for penalizing complexity, synthetic tests for verification, and forecasting experiments for final validations. This holistic and objective approach has allowed us to expose distinct properties of seismicity that seem to indicate robust empirical laws and also other features than can be understood as artifacts in terms of measurement errors and inconsistencies in data quality.

The methods introduced in this thesis deal with all the dimensions of a typical seismicity catalog, namely magnitude, location and time. It is important to notice that all these parameters are jointly inferred from a limited set of seismic observations under the assumption of a previously modeled earth structure. Thus, the observational and modeling errors affect all of them through different non-linear couplings. For instance, a misidentified phase arrival can bias the hypocentral location that in turn might lead to a bias in both the time and magnitude of the earthquake. Similarly, a velocity model biased towards faster seismic wave velocities will tend to cause overestimated depths. The information potential of these catalog parameters is directly proportional to their range and inversely proportional to their uncertainty. Our modeling approaches should be tailored according to these primary differences and our consequent interpretations should not exceed the upper bounds set by our limited instrumental capacity. We can only then build up on our acquired understandings and extend our investigations.

We thus begin our analysis with the investigation of the frequency-magnitude relationship distribution of Californian seismicity. Our motivation is that if earthquakes exhibit spatially varying magnitude probabilities, then through identifying these regions we can create better models to forecast seismicity. If such spatial variations do exist, further investigations regarding their origin can be beneficial. For this purpose, we develop in Chapter 2 a data-driven spatial analysis method that is based on Voronoi tessellations and penalized likelihood. In Chapter 3, we put our modeling results through retrospective forecast tests and validate that the introduced method provides improvements over the state-of-the-art methods. Such studies lead us to the conclusion that the spatial variation in frequency-magnitude distribution, while significant in terms of forecasting capability, is mainly driven by network discrepancies and magnitude errors. Thus, we decide to turn our attention on a seemingly better constrained parameter, namely the hypocenter locations in the relatively well instrumented Southern part of California. Similarly to the frequency-magnitude distribution, the spatial distribution of seismicity also exhibits a power-law scaling, namely a multifractal spectrum. However, at this point we are hampered by a) the poor performance of the classical multifractal estimation methods and b) the vast computational resources needed for the analysis of such large datasets. To address the first issue, we developed a new approach for point distributions in Chapter 4. This new method is tested rigorously on deterministic and stochastic multifractal sets to establish its superior performance. To tackle the problem of large datasets, we introduce in Chapter 5 a new compression technique that reduces the size of seismicity catalogs while preserving the spatial information content. In the same study, we perform a multifractal analysis of the Southern Californian seismicity. Our results suggest the presence of artifacts that could be associated with the relocation procedure of the catalog. Motivated by these findings, we then undertake in Chapter 6 the comprehensive task of developing a consistent velocity model for Southern California, and relocating the seismicity of the last 30 years with a probabilistic non-linear method. Finally, in Chapter 7, we develop a new pattern recognition method for seismicity-based fault reconstruction. We apply our new method to the new catalog, and obtain a detailed fault network that is validated through a retrospective forecasting test. In the final Chapter 8 of this thesis, we discuss our results and give an outlook for future studies.

Chapter 2

Data-driven spatial b-value estimation with applications to California seismicity: To b or not to b

by

Yavor Kamer, ETH Zürich
Stefan Hiemer, ETH Zürich

published in
Journal of Geophysical Research: Solid Earth, July 2015

2.1 Abstract

In this paper we present a penalized likelihood-based method for spatial estimation of Gutenberg-Richter's b-value. Our method incorporates a non-arbitrary partitioning scheme based on Voronoi tessellation, which allows for the optimal partitioning of space using a minimum number of free parameters. By random placement of an increasing number of Voronoi nodes we are able to explore the whole solution space in terms of model complexity. We obtain an overall likelihood for each model by estimating the b-values in all Voronoi regions and calculating its joint likelihood using A i's formula. Accounting for the number of free parameters we then calculate the Bayesian Information Criterion for all random realizations. We investigate the ensemble of the best performing models and demonstrate the robustness and validity of our method through extensive synthetic tests. We apply our method to the seismicity of California using two different time spans of the ANSS catalog (1984-2014 and 2004-2014). The results show that for the last decade the b-value variation in the well-instrumented parts of mainland California is limited to the range of $[0.94 \pm 0.04 - 1.15 \pm 0.06]$. Apart from the Geysers region, the observed variation can be explained by network related discrepancies in the magnitude estimations. Our results suggest that previously reported spatial b-value variations obtained using classical fixed radius or nearest neighbor methods are likely to have been overestimated, mainly due to subjective parameter choices. We envision that the likelihood-based model selection criteria used in this study can be a useful tool for generating improved earthquake forecasting models.

2.2 Introduction

The Gutenberg-Richter (GR) law provides a fundamental description of the relation between earthquake occurrence frequencies and magnitudes: $\log_{10}(N(m)) = a - bm$, where a and b are constants, and $N(m)$ is the cumulative number of earthquakes of magnitude m or greater [Ishimoto and Iida, 1939; Gutenberg and Richter, 1954]. The widely accepted empirical formula has prevailed through time, having been validated numerous times by earthquake catalogs, laboratory experiments (e.g. Mogi, 1962; Amitrano, 2003, 2012) and numerical models (e.g. Bak & Tang 1989; Olami et al. 1992). Thus, for earthquake forecasting, the frequency magnitude analysis has been quite popular among other more controversial and complex concepts such as the characteristic earthquake model, static stress triggering and seismic gap models. Especially the b-value, which describes the relative amount of large to smaller earthquakes, has attracted widespread research interest due to the fact that it is one of the main parameters governing seismic hazard.

Different methods have been proposed for the estimation of the b-value. Initially, seismologists used least squares fitting on a log-log plot to determine the slope that corresponds to the b-value. After the introduction of A i's maximum likelihood estimator [Aki, 1965] it has been repeatedly demonstrated that this should be the preferred method [Marzocchi and Sandri, 2003; Amorese et al., 2010]. Corrections for the magnitude binning have also further improved the performance of the estimator [Utsu, 1965; Bender, 1983; Tinti and Mulargia, 1987]. Additionally, analytical derivations for the standard estimation error with respect to sample size and b-value have been derived [Shi and Bolt, 1982].

Many studies have investigated the variability of the b-value both in the space and time domains. Despite the well established methodology for b-value estimation, there is not a consensus on the reporting and the interpretation of its results. An important question is whether the observed

spatiotemporal variations of the b-value are statistically significant. Currently there are two main conceptions: The parsimonious view maintains that the b-value is constant both in space and time and that the observed fluctuations are artifacts resulting from undersampling, magnitude errors and non-homogenous detection capabilities [*Shi and Bolt*, 1982; *Frohlich and Davis*, 1993; *Kagan*, 1999b, 2002, 2010; *Amorese et al.*, 2010]. The other side of the debate argues that the observed b-value variations are not statistical or network related artifacts. Thus they have been used for physical interpretation of various faulting regimes [*Wiemer and Wyss*, 1997, 2002]; as a proxy for shear stress and pore pressure [*Scholz*, 1968; *Schorlemmer et al.*, 2005; *Bachmann et al.*, 2012]; as an indicator of material heterogeneity [*Mogi*, 1962; *Mori and Abercrombie*, 1997]; or even as precursors of future rupture areas and sizes [*Schorlemmer and Wiemer*, 2005]. The debate has remained inconclusive; mainly due to the lack of objective methods that can produce quantitative results [*Kagan*, 1999a].

The most commonly used methods for mapping b-values in space are the fixed radius and nearest neighbor method [*Wiemer and Wyss*, 2002] that estimate b-values from close-by events within a proximity limit. The parameter sensitivity of these approaches has been previously demonstrated [*Schorlemmer et al.*, 2004; *Kamer and Hiemer*, 2013]. Imoto (1987, 1991) and Ogata and Katsura (1993) addressed the issue of optimal parameter choices by introducing likelihood-based b-value mapping approaches that take into account model complexities. In these methods, the study region is divided into segments of equal space-time volumes. B-values are estimated for each of these volumes in an optimization setting that seeks to maximize a likelihood-based information criterion [Imoto, 1987, 1991]. These approaches were the first to introduce likelihood criteria for parameter optimization, however they can be further improved by addressing the limitation of using regularly shaped, equal volume segments.

In this study we build on Ogata's principle of likelihood criteria, and we propose a new method for the investigation of spatial b-value variations, which incorporates a data-driven partitioning scheme. The paper is structured as follows: first, we give an overview of the commonly used methods for spatial mapping of b-values. Secondly, we present our method based on optimal partitioning using Voronoi tessellation, penalized likelihood and wisdom of the crowd philosophy. To assess the validity and the performance of our method we perform synthetic tests using realistic spatial distributions of events with varying b-values and sample sizes. Finally, we apply our method to the seismicity of entire California and discuss the implications of our results for previous and future studies.

2.3 Traditional spatial b-value mapping

Since its introduction in the 90s, the spatial mapping of Gutenberg-Richter's b-value has become an increasingly popular analysis. B-value maps obtained with the commonly used techniques often feature large spatial variations at small scales, which are rarely observed in well-sampled large catalogs. However, the input parameters (such as grid spacing, cylindrical mapping radius, and minimum number of events) have to be chosen somewhat arbitrarily.

The classical spatial b-value analysis [*Wiemer and Wyss*, 2002] involves the following steps: (1) the earthquakes within the study area are projected on a planar surface (e.g. the earth's surface, a fault cross section), (2) the planar surface is gridded into equally sized cells, (3) each cell is used as a center for a circle of radius R , and (4) if the number of encircled events exceeds a predefined minimum (N_{min}), the b-value is estimated and assigned to the location of the cell. A variant of the

same method involves using variable circle radii, which are expanded until a constant number of nearest earthquakes are enclosed. For both approaches the b-values of neighboring cells are inevitably correlated because R is always larger than the cell grid interval D_g which results in substantial data overlaps. In the constant number approach, the b-value at each node is estimated from equally sized samples, thus the uncertainty of the results is assumed to be spatially constant while the spatial resolution varies with the density of earthquakes. The constant radius approach keeps the spatial resolution constant, whereas the b-value uncertainty varies spatially.

Regardless of the choice of constant number or radius, the results of the method are sensitive to the values of the free parameters R , N_{min} and D_g . For qualitative and quantitative examples the reader is referred to Figure 4 of Schorlemmer et al.(2004) and Figure 3 of Kamer (2014). While significant effort is devoted to find patterns in the resulting b-value maps, further research is needed to assess the influence of these free parameters or to develop objective strategies for determining their values. For instance, the sampling radius R is sometimes chosen based on qualitative criteria such as the “crispiness” of the resulting b-value image [Wyss et al., 2000; Schorlemmer et al., 2004]; or observations that certain values reduce b-value contrast [Schorlemmer et al., 2004]. The minimum number of samples (N_{min}), which governs the standard error of the estimations, is also often set to low values. Using small sample sizes (~50) is favorable, since it can image small scale details, and can be justified if there is prior knowledge that differences in b-value are large [Wiemer and Wyss, 2002]. Other studies, on synthetic catalogs with known b-values, suggest that such small sample sizes can cause variations in the estimations [Shi and Bolt, 1982; Pickering et al., 1995; Felzer, 2006]. Nevertheless, due to the difficulty in detecting sufficient amount of events in certain field conditions, such as induced seismicity, some studies are carried out with even smaller sample sizes (~25) [Bachmann et al., 2012].

Studies that aim to assess the significance of the obtained spatial b-values usually investigate standard deviations [Efron, 1979] and conduct statistical significance tests [Utsu, 1992] based on the subset of events assigned to each grid cell [Enescu and Ito, 2002; Schorlemmer et al., 2003, 2004]. These tests can indicate which of the estimated b-values are reliable; however varying the mapping parameters can change the data assigned to the grid cells and hence the obtained test results. A recent study introduced a semi-synthetic test to facilitate parameter choices [Tormann et al., 2014], whereas the applicability of the proposed method, in the absence of prior knowledge on b-value variation, is debated [Kamer, 2014; Tormann and Wiemer, 2014].

Consequently, some issues arise in the interpretation of b-value maps obtained by this method: 1) due to the subsampling induced by low values of the parameters R and N_{min} , the b-value maps feature extreme spatial variations at small scales that are rarely observed in large catalogs [Howell, 1985; Bird and Kagan, 2004]. 2) Earthquake epicenter distributions and faults are known to be scale-invariant [Kagan and Knopoff, 1980; Mandelbrot, 1982; Ouillon and Sornette, 1996; Kagan, 2007]. In other words, the notion of characteristic scale (expressed by the R parameter) does not exist [Sornette and Werner, 2009]. The resolvable scaling range, which is bounded by location uncertainties at the lower end, expands as a function of sample size. 3) Previous studies have shown that earthquake clusters can be anisotropic [Ouillon et al., 2008; Ouillon and Sornette, 2011], thus the choice of isotropic mapping volumes may not always be the optimal choice. Cylindrical volumes become especially problematic in the presence of closely spaced faults with diffused seismicity (e.g. the Hellenic subduction zone, pers. comm. G. Chouliaras). 4) The b-values are estimated using a maximum likelihood approach, however due to the overlapping volumes a single

earthquake contributes to the likelihood of several b-value estimates. Thus, there is no straightforward way to calculate an overall likelihood for the whole b-value map with respect to the entire catalog. Due to this data overlap it is also not clear how to obtain a synthetic earthquake catalog using a b-value map such as the generating model, hence the results cannot be statistically or numerically tested as a whole. For a detailed comparative synthetic test that also highlights these difficulties, the reader is referred to Appendix A.

2.4 A penalized likelihood-based method

In order to address the limitations described in the previous section, we propose an automated, parameter-free method which is based on optimal partitioning using Voronoi tessellation, penalized likelihood and wisdom of the crowd philosophy. The problem of mapping b-values can be viewed as a two step inverse problem involving (1) the classification of a subset of earthquakes (corresponding to the delineation of a region) which is in accordance with the GR law and (2) the estimation of a b-value. Not only the number of subsets but also their shapes, sizes and b-values have to be inferred jointly. That is using the data themselves to constrain the maximum allowable complexity in the model rather than specifying this beforehand. Considering the broad solution space, one seeks an ensemble of solutions instead of a single best model and derives properties of that ensemble to infer information about the unknown spatial model [Sambridge *et al.*, 2013].

Earthquakes are clustered in space but the shapes and sizes of these clusters vary greatly, hence the need for a flexible partitioning approach facilitating the search of segmentations that might be present in the data. For this purpose we use Voronoi tessellation that uses a set of points (nodes) to partition space assigning each node with its nearest neighborhood region [Voronoi, 1908]. By random perturbation of these nodes, it is possible to obtain anisotropic regions of various shapes and sizes. Another benefit of the Voronoi tessellation is that the complexity of the partitioning model is minimal, requiring only the coordinates of the nodes to be specified. This is of particular importance because increasing the number of Voronoi nodes allows for exploration of smaller scales. If one were to use polygon vertices for defining the partitioning, the number of free parameters in the model would increase rapidly with the number of regions.

We start with a single Voronoi cell, which assigns a single b-value for the whole catalog; this solution can be regarded as the null hypothesis as it implies that there is no spatial b-value variation. We then gradually increase the number of Voronoi nodes, which generates models of increased complexity. At each increment we randomize the locations of the Voronoi nodes investigating possible segmentations. For each Voronoi region (cell) we first estimate the magnitude of completeness (m_c) of the enclosed events using the maximum curvature method [Wyss *et al.*, 1999; Mignan and Woessner, 2012], which is further complemented with statistical significance tests [Clauset *et al.*, 2009]. The probability density function of the magnitudes $m \geq m_c$ is given by:

$$P(m) = b \log(10) 10^{-b(m-m_c)} \quad 2-1$$

The likelihood that a set of N magnitudes m_i is a realization of equation (2-1) is given by:

$$L(b | m) = \prod_{i=1}^N P(m_i) \quad 2-2$$

and the corresponding log-likelihood reads:

$$\log(L(b | m)) = N \log(b) + N \log(\log(10)) - b \log(10) \sum_i^N (m_i - m_c) \quad 2-3$$

Using equation (2-3) we calculate the log-likelihood of all $m \geq m_c$ events in each Voronoi cell. The b-value is estimated with Aki's (1965) maximum likelihood method corrected for binned data using the formula provided by Tinti and Mularia (1987):

$$b = \frac{1}{\log(10)\Delta m} \log\left(1 + \frac{\Delta m}{\hat{\mu} - m_c}\right) \quad 2-4$$

where $\hat{\mu}$ is the sampling average of the magnitudes. We want to point out that we use this formula rather than the commonly used corrections [Utsu, 1966; Bender, 1983] because it gives more accurate and stable results under varying sample sizes [Marzocchi and Sandri, 2003]. We point out that Aki's maximum likelihood estimation is based solely on the mean magnitude, i.e. datasets with similar means will result in similar b-values, regardless of their parent distributions. Due to the absence of any measure on the quality of the fit, Deluca and Corral (2013) suggest that equation (2-4) should be rather called "minimum unlikelihood estimation". This issue is addressed in Appendix B where we investigate the plausibility of the GR law for the Californian seismicity. It should be noted that equation (2-1) implies infinite seismic moment/energy rate, thus for physical consistency it must be supplemented by an upper magnitude limit. In this study we use the standard GR relation because, if treated as a fitting parameter, the upper magnitude would be biased by the size of each Voronoi region. The issues arising from the assumption of an infinite upper limit will be further discussed in Appendix B.

Unlike the classical method that has significant data overlaps, the tessellation ensures that each event is used only in the estimation of the b-value of its Voronoi cell. Thus by summing the log-likelihoods over all Voronoi cells, we can calculate the overall log-likelihood for each random tessellation:

$$\log(\hat{L}) = \sum_{v=1}^{N_v} \log(L(b_v | m)) \quad 2-5$$

where N_v is the total number of Voronoi nodes and b_v is the estimated b-value in each Voronoi cell. This provides us with an objective criterion that can be used to search for the optimal number of partitions. However, as in every optimization problem, increasing the model complexity (i.e. the degrees of freedom) will increase the likelihood, thus to compare models with different complexities it is necessary to penalize the likelihood based on the number of free parameters. For this purpose, we use the Bayesian Information Criteria (BIC) [Schwarz, 1978] given by:

$$BIC = -\log(\hat{L}) + \frac{k}{2} \log(N)$$

2-6

where \hat{L} is the overall likelihood of a model (cf. eq. 5), k is the number of free parameters in the model and N is the number of data points. Each Voronoi cell is characterized by the coordinates of its node and its b-value, thus for 2D applications each Voronoi cell would require 3 free parameters (x, y, b). Such parameterization would allow the node locations to be specified with infinite precision (since $x, y \in \mathbb{R}$). However, as already mentioned in the previous section, the smallest resolvable scale is bounded by the earthquake location uncertainties. Thus we discretize the Voronoi node selection space in accordance with the location uncertainties. Intuitively, it follows that this reparametrization should reduce the model complexity since the location parameters become less potent. Indeed, when the space for possible Voronoi nodes is limited and discretized, instead of specifying two variables for each coordinate we can initially communicate an enumerated list of all possible finite locations and their coordinates. With the help of such a list it is possible to refer to Voronoi node locations only by their indices and hence use only one free parameter instead of two. Since the list has to be communicated for all models, regardless of their complexity, it can be regarded as a constant and thus be neglected in the BIC formulation. With this reformulation, the number of free parameters k in equation (2-6) is given as $k=2*N_v$, where N_v is the number of non-empty Voronoi cells. Apart from decreasing the overall model complexity, the discretization reduces the procedure of random Voronoi placement to N_v -permutations of N_p where N_p is the number of all possible discrete locations. This makes the random search more tractable and avoids repeated placement of nodes on the same coordinate.

The total number of randomly generated tessellations is given as $T_{Tot} = T(N_{v,\max} - 1)$, where $N_{v,\max}$ is the highest number of Voronoi cells considered and T is the number of random realizations for each N_v . We consider all models, which perform better (in terms of BIC) than the prior that is a single b-value for the entire dataset. After ranking all models according to their BIC per event we obtain an ensemble solution by BIC weighted averaging of the top ranking models. This is a manifestation of the wisdom of the crowd philosophy since models with different complexities but similar penalized likelihoods have equal influence on the ensemble inference. Similar concepts, often referred to as “boosting” [Schapire, 1990], are widely used in machine learning. The ensemble inference provides not only an average spatial b-value map but also a corresponding standard deviation for each location. This allows inferring the stability of the results and provides insights about the penalized likelihood surface, such as presence of competing local maxima. The method bears similarities with, and is in fact inspired by, the concept of transdimensional inference developed by Sambridge and others [Sambridge *et al.*, 2006, 2013; Bodin and Sambridge, 2009]. However, our method is considerably simpler in its formulation and implementation due to the fact that it relies on BIC and random search of the solution space. By contrast, other sampling techniques such as Markov chain Monte Carlo require prior distributions for parameters and for probabilities of changing the model complexity.

2.5 Application to synthetic datasets

2.5.1 Uniform event distribution

To demonstrate the performance of the proposed method we conduct a simple checkerboard test that features b-value anomalies [0.5-1.5] with different spatial extents and different sample sizes (Figure 2-1a). For the sake of simplicity, we assume a uniform distribution of event locations. First we applied the classical nearest neighbor mapping method using two different values for the mapping parameter N_{\min} . Setting $N_{\min} = 150$ [e.g. Tormann et al., 2014] results in the emergence of artifacts due to undersampling (strong small-scale b-value variations [0.39-1.99], Figure 2-1b). Increasing N_{\min} to a larger value (1000) prevents such artifacts but fails to reproduce the b-value anomalies [0.53-1.35] due to mixing different magnitude distributions (Figure 2-1c). By contrast to the classical b-value mapping approach, the proposed method does not feature any adjustable tuning parameters. For this synthetic example case we randomly generated 50,000 models in the complexity range of 1-100 Voronoi seeds for data partitioning. We ranked all models according to their penalized likelihood score and obtained the ensemble median of the best performing models ([0.47-1.48], Figure 2-1d,e). The ensemble is stable with respect to the number of individual models considered (compare Figure 2-1d and e).

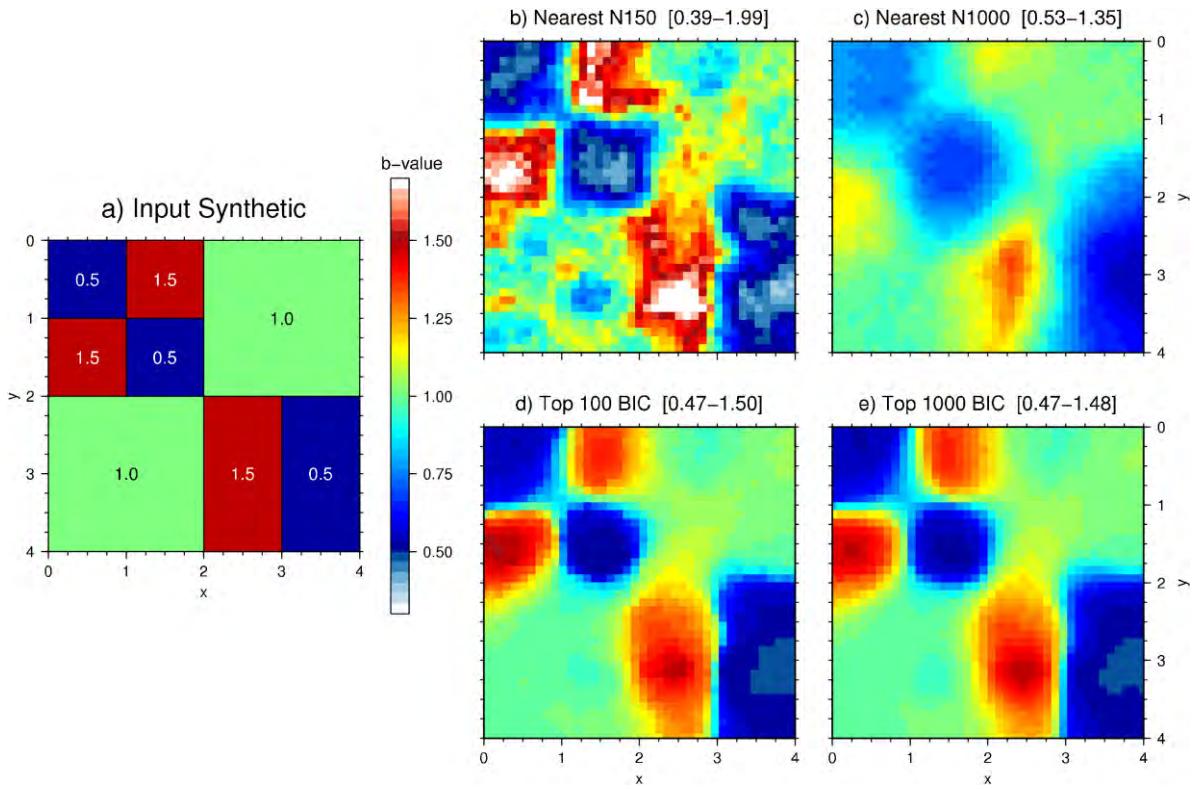


Figure 2-1 a) Underlying spatial b-value distribution for the generation of synthetic magnitudes for 8000 uniformly distributed events. b) and c) Resulting b-value maps using the classical nearest neighbor mapping method for $N_{\min} = 150$ and $N_{\min} = 1000$, respectively. d) and e) Resulting b-value maps based on the penalized likelihood approach for 50,000 randomly generated models in the complexity range of 1 - 100 Voronoi nodes (showing the median of the 100 and 1000 best performing models, respectively.) Ranges of resulting b-values are given in the title of each subplot.

2.5.2 Fractal event distribution

To further demonstrate the performance and robustness of the proposed method and to illustrate a comparison between the ensemble and individual models, we conduct a more realistic synthetic test. This involves generating synthetic earthquake catalogs using two spatially separated frequency magnitude distributions (FMDs) with: 1) a b-value for background events (b_{bkg}) and 2) an anomalous b-value (b_{ano}) for a subregion. Instead of using an unrealistic homogenous spatial event distribution we generate a fractal distribution by recursively replicating the matrix M:

$$M = \begin{bmatrix} 1 & 0 & 1 & 0 & 1 \\ 0 & 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 & 1 \\ 0 & 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 & 1 \end{bmatrix}$$

This produces a point distribution with a fractal dimension of $D=\log(13)/\log(5)=1.59$. For a more detailed description of this process, the reader is referred to Kamer et al. (2013). This value is chosen such that it is in accordance with the reported fractal dimensions of actual catalogs: ≈ 1.6 for southern California [Helmstetter et al., 2005] and similar values of $D=1.5-1.6$ for the Japan Arc [Hirata, 1989]. Unlike a homogenous event distribution, such a synthetic fractal is a more realistic spatial representation of seismicity and poses more sensible challenges to any mapping technique in general.

Using this synthetic location distribution that extends a unit square, we delineate a region of interest (ROI) which contains 25% of the events (Figure 2-2a). We simulate magnitudes for the events in the ROI using a GR law with a b-value of $b_{\text{ano}}=0.75$; the remaining (surrounding) event magnitudes are generated using a b-value of $b_{\text{bkg}}=1$. The generated catalog consists of a total of 8000 events, 2000 being located in the ROI (Figure 2-2a). Notice that the estimated b-value for the anomaly deviates from the input ($b_{\text{ano}}=0.77$) due to the relatively small sample size (Figure 2-2b).

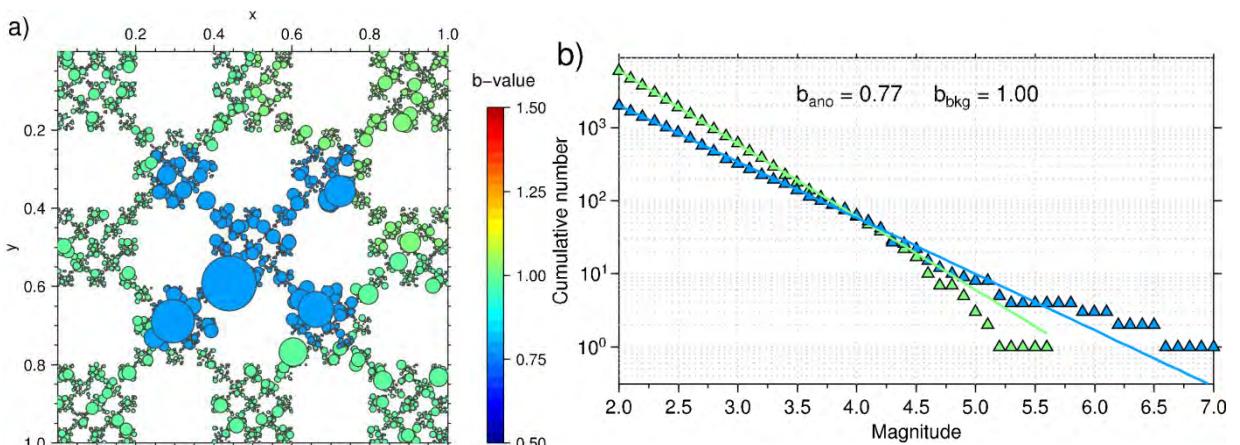


Figure 2-2 Fractal ($D=1.59$) distribution of synthetic events. Symbol size scales with event size. Color-coding depicts the underlying b-value: 6000 background events were drawn from a GR law using $b_{\text{bkg}}=1.0$, and 2000 events from $b_{\text{ano}}=0.75$. The boundaries of the anomaly are 0.25/0.75 in both x- and y-direction. b) Cumulative frequency-magnitude distributions and estimated b-values for both regions.

We discretize the spatial domain by a 20x20 grid, which results in 400 possible locations for Voronoi node placement. The number of Voronoi cells is increased from $N_v = 1$ to $N_{v,max} = 30$ at an increment of one, performing 500 random tessellations at each step. This results in $T_{Tot} = 500(30 - 1) = 14500$ models, which are ranked according to their BIC (Figure 2-3c). Only 102 of these perform better than the prior model (i.e. a single b-value for the whole catalog). Notice that $N_{v,max}$ is not a free parameter since above a certain number of Voronoi cells ($N_v=15$), none of the tessellations performs better than the prior. We use all the 102 models to obtain an ensemble model, which is expressed in terms of median and interquartile range (IQR). Such a non-parametric representation avoids the assumption of a certain distribution and symmetry. For objective evaluation of the output we report the median b-value within the predefined ROI, which is 0.79 (Figure 2-3a). High IQR depicts regions in which the ensemble signal becomes unreliable (Figure 2-3b). These can correspond to regions with limited data or transition zones between distinct b-value regions. Notice that despite the arbitrary discretization of Voronoi locations, the ensemble conveys sufficient spatial detail to distinguish the ROI. The ensemble signal is invariant with respect to the discretization of the domain due to the fact that the optimal complexity (i.e. the number of Voronoi cells) is governed by the number of data points and their likelihood. To illustrate the Voronoi tessellation obtained by individual models, we show the top ten solutions labeled by their number of Voronoi nodes and BIC (Figure 2-3d). We observe that even though most of the individual solutions did not delineate the ROI correctly, the ensemble provides a stable solution ($b_{ano} \approx 0.79$, $IQR \approx 0.05$) and most importantly spatial variance estimation.

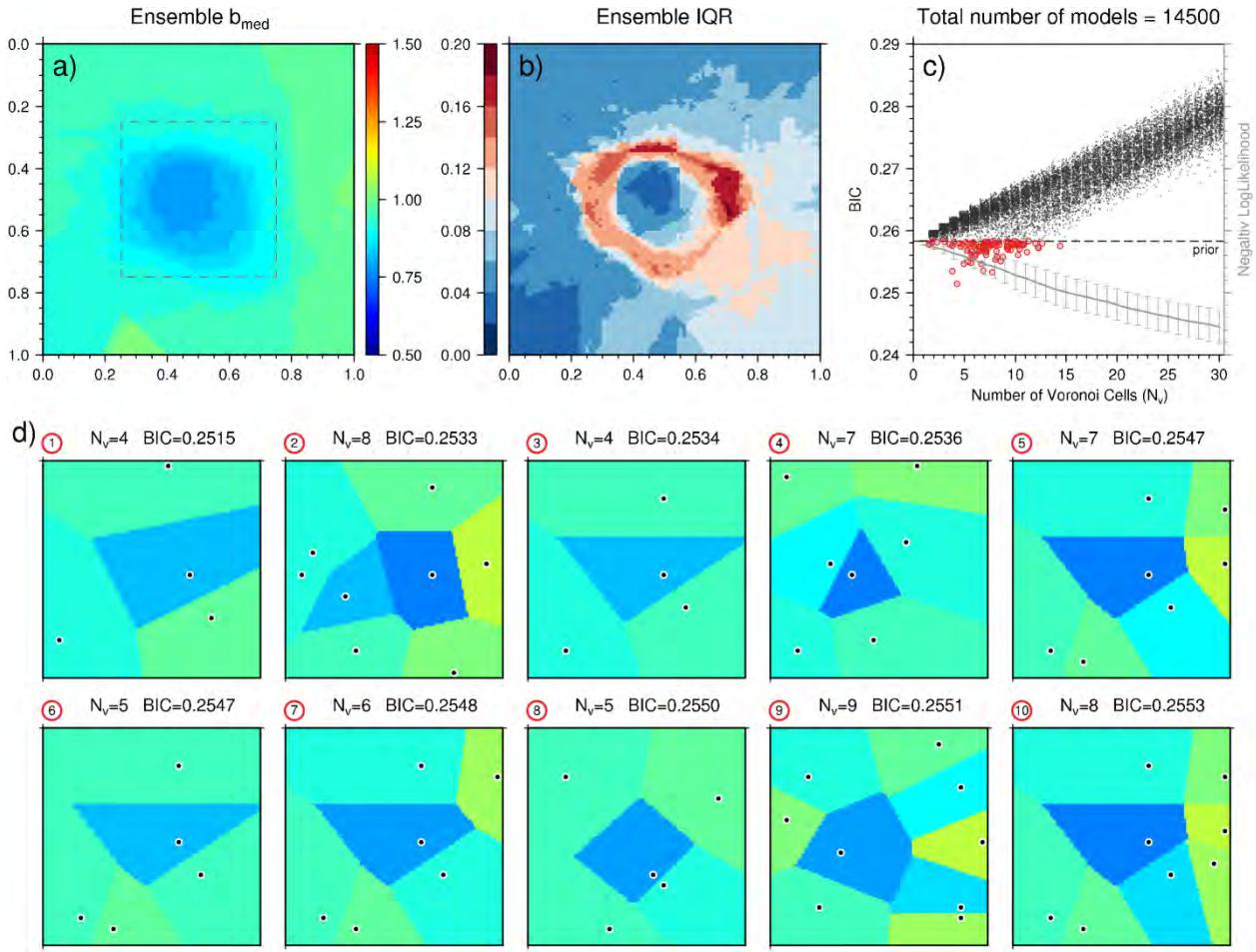


Figure 2-3 a) Ensemble median b_{med} and b) ensemble interquartile range (IQR) of the best 102 solutions. The median spatial b -value within the region of interest ROI is 0.79 (dashed line, compare Fig. 1). c) 14500 randomly generated solutions are shown in terms of their BIC (black dots) and their average negative log-likelihood (gray line, error bar corresponds to one standard deviation). Red circles highlight the best 102 solutions. d) Best ten individual models. Black dots denote the positions of the Voronoi nodes; resulting partitions are color-coded by their estimated b -values.

2.5.3 Convergence with increasing sample size

To assess the performance and convergence of our method we extend the presented synthetic test by: 1) investigating a wider range of commonly reported b -values, analyzing $b_{ano}=0.50, 0.75, 1.25$ and 1.50 ; 2) incrementing the sample size of the ROI (N_{ano}) from 100 to 5000 at steps of 100; and 3) considering different ratios between the number of events in the ROI and the background ($N_{bkg}/N_{ano}=3, 2$ and 1). This results in a total of $4 \times 50 \times 3 = 600$ test cases. The overall error of the estimated b -value during such a test is influenced by both the synthetic catalog generation and the optimal partitioning process. We generate 100 synthetic catalogs for each case; for each of these catalogs we apply our method and report the median b -value within the ROI obtained from the best model. We then use these 100 b -values to calculate confidence intervals for the result of each test case. The results of the test are grouped into three panels according to the N_{bkg}/N_{ano} ratio (Figure 2-4).

Intuitively, as the difference between $b_{bkg}=1$ and b_{ano} increases the detection of the anomaly should become easier. This is observed for negative b -value anomalies ($b_{ano} < 1$), which converge rather quickly to their input values, however, for the positive anomalies ($b_{ano} > 1$) the convergence is much slower (Figure 2-4). Not surprisingly this is due to the fact that the b -value is an expression of slope on a log-log plot: as the slope becomes steeper (higher b -values) small variations in the

abscissa (magnitude axis) will induce greater fluctuations. This is also why, as a rule of thumb, any candidate power-law is required to span at least two orders of magnitude in both x and y axes [Stumpf and Porter, 2012]. This test demonstrates the robustness of our method and also highlights the inherent difficulties associated with estimating power-law exponents. We remind the reader that for a single dataset, the reliability analysis is carried out by investigating the spatial variance distribution obtained from the ensemble averaging procedure (cf. Figure 2-3b).

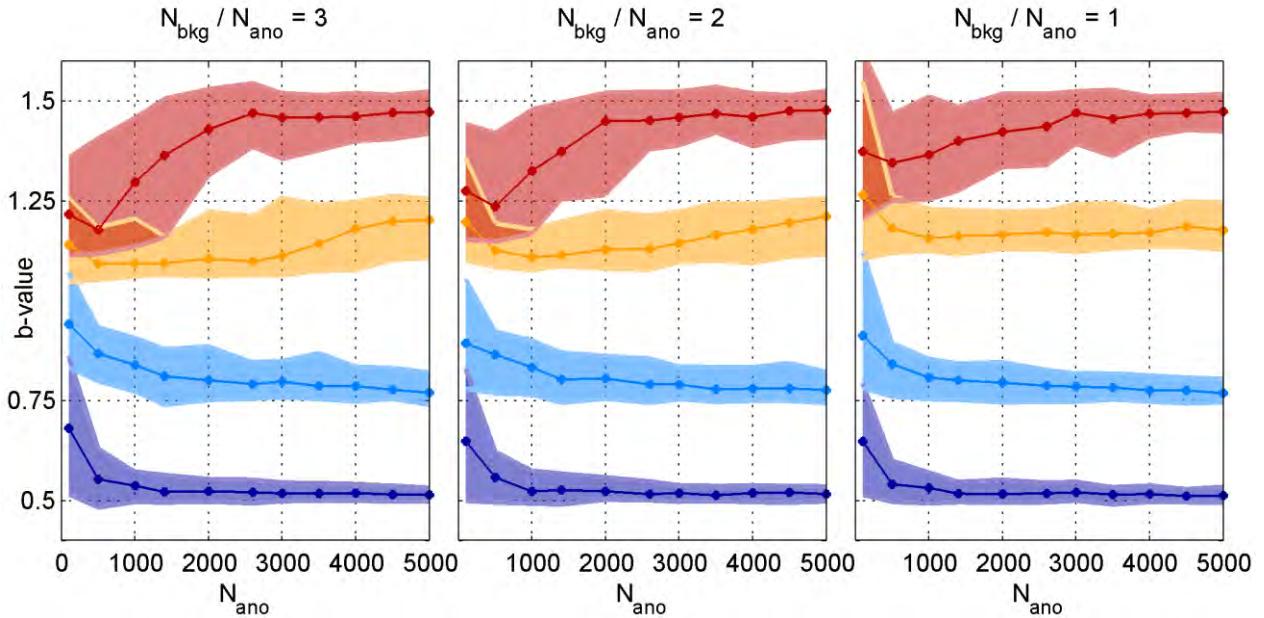


Figure 2-4 Convergence of the method for four different b-value anomalies (0.5, 0.75, 1.25 and 1.5) with varying sample sizes (from 100 to 5000). The three panels show results for different sample size ratios with respect to the surrounding events. Solid lines and shaded areas denote the corresponding median b-value and 0.16-0.84 quantiles, respectively.

2.6 Application to Californian seismicity

We applied our method to Californian seismicity using the Advanced National Seismic System composite catalog (ANSS, <http://www.ncedc.org/anss/>). Our study region is delineated by a polygon that envelopes all listed crustal events for the period of 01.01.1900 - 01.01.2014 ($31^\circ \leq \text{latitude} \leq 43^\circ$, $-127^\circ \leq \text{longitude} \leq -113^\circ$, $0 \leq \text{depth} \leq 30 \text{ m}$). The event magnitudes are binned into magnitude bins of $\Delta m = 0.1$. We then perform a temporal mc analysis using a sliding window of 5000 events with a sliding step of 1000 events. For mc estimation, we used the maximum curvature method, which consists of finding the magnitude bin with the largest number of events. This non-parametric method requires fewer events than other techniques to reach a stable result [Mignan and Woessner, 2012]. We observe that the estimated mc values show a decreasing trend with short-term spikes that are characteristic for aftershock sequences of large earthquakes. Because the Southern Californian network underwent a substantial upgrade in 1984, we use all $m \geq 2.2$ (chosen as the highest observed value during this time period) events in the period of 01.01.1984 - 01.01.2014 (125,760 events with $m \geq 2.2$). Due to spatially inhomogeneous station coverage, the catalog is not complete everywhere at this magnitude level. However, our method accounts for possible spatial completeness variability since mc is locally estimated in each Voronoi partition.

2.6.1 Results for the time period of 01.01.1984 – 01.01.2014

To allocate discrete locations for Voronoi node placements, we discretize the study area into 1x1 km cells obtaining a total of 712,493 cells. This discretization is chosen in accordance with the horizontal location uncertainty of the events and allows for delineation of small-scale features while reducing the computational demands. We have investigated the results' sensitivity with respect to the chosen grid size by testing 5x5 and 10x10 km cells and confirm that the ensembles remain stable. Instability arises when using much larger grids, which affect the shape and size of the tessellations, neglecting small scale features supported by the data. In order to further increase the efficiency of our random searching algorithm, we consider only cells containing more than one event. This reduces the number of potential Voronoi node locations from 712,493 to 16,152.

We increment the number of Voronoi nodes from 1 to $N_{v,max} = 500$, and spatially randomize each configuration 2,000 times. As a result, the solution space comprises 10^6 models, each representing a unique spatial segmentation of the study region. All models are labeled and ranked by their penalized likelihood using BIC. Preliminary results showed that some of the b-value estimates (especially in the Cascadia subduction zone) were affected by an underestimation of m_c . This is due to the tendency of the maximum curvature method to underestimate m_c in the case of gradually curved FMDs [Mignan and Woessner, 2012]. When m_c is underestimated, the maximum likelihood estimate of the b-value (which is driven by the small events) becomes unreliable for the large magnitudes. To counter this problem, we apply the method of Clauset *et al.* (2009) which optimizes the lower data cut off (m_c) by minimizing the Kolmogorov-Smirnov distance between the empirical CDF obtained from the data and the analytical CDF obtained using the estimated GR law exponent. The use of this method was also proposed by Amitrano (2012) who pointed out improper statistical methods as a potential explanation for observed b-value variations. For visual inspections of the individual GR law fits, we have generated videos featuring each Voronoi region of the best 100 tessellations [Kamer and Hiemer, 2014a, 2014b]. These videos, which are publicly accessible, demonstrate the robustness of the estimates. We find that all 10^6 random realizations performed better than the prior model (Figure 2-5a), which suggests that the use of a single b-value to describe such a heterogeneous dataset is an oversimplification (i.e. an underfitting model). We calculate the ensemble model using the best 1000 solutions. Unlike the synthetic test maps, we use only the non-empty 1x1 km cells to display both the median b-values and the interquartile ranges (Figure 2-5b,c). The results are stable with respect to the number of averaged top solutions since model contributions are weighted by their BIC. As verification, we have averaged as few as 10 and as much as 10,000 best solutions, obtaining consistent results in terms of both median and interquartile range. Normally $N_{v,max}$ should be set to a number large enough to ensure that the average BIC trend goes above the prior and thus all model complexities better than the prior have been considered. In this application we observed that the majority of the best performing models is concentrated at $Nv < 100$ thus, in order to save computational time, we set $N_{v,max} = 500$. We confirm that even when considering the 10,000 best performing models, all of their complexities were well below $Nv=100$.

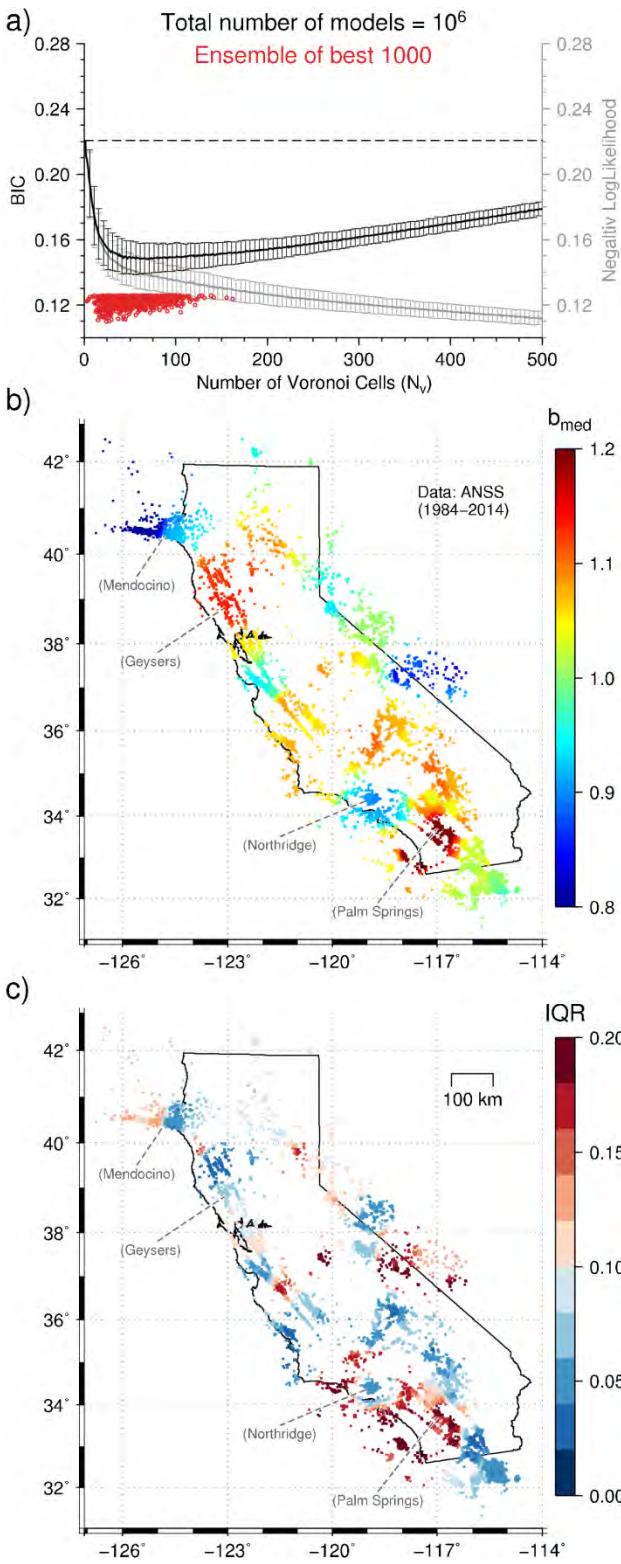


Figure 2-5 a) Entire solution space for the ANSS catalog (1984-2014, $m \geq 2.2$): Average BIC (black line) and negative log-likelihood (gray line) as a function of number of Voronoi cells. Error bars correspond to one standard deviation. Red circles denote the BIC of the best 1000 solutions. b) and c) show the corresponding ensemble median b -value (in the range of 0.78-1.28) and ensemble IQR (in the range of 0.02-0.65), respectively, computed at all 1x1km cells with more than one event.

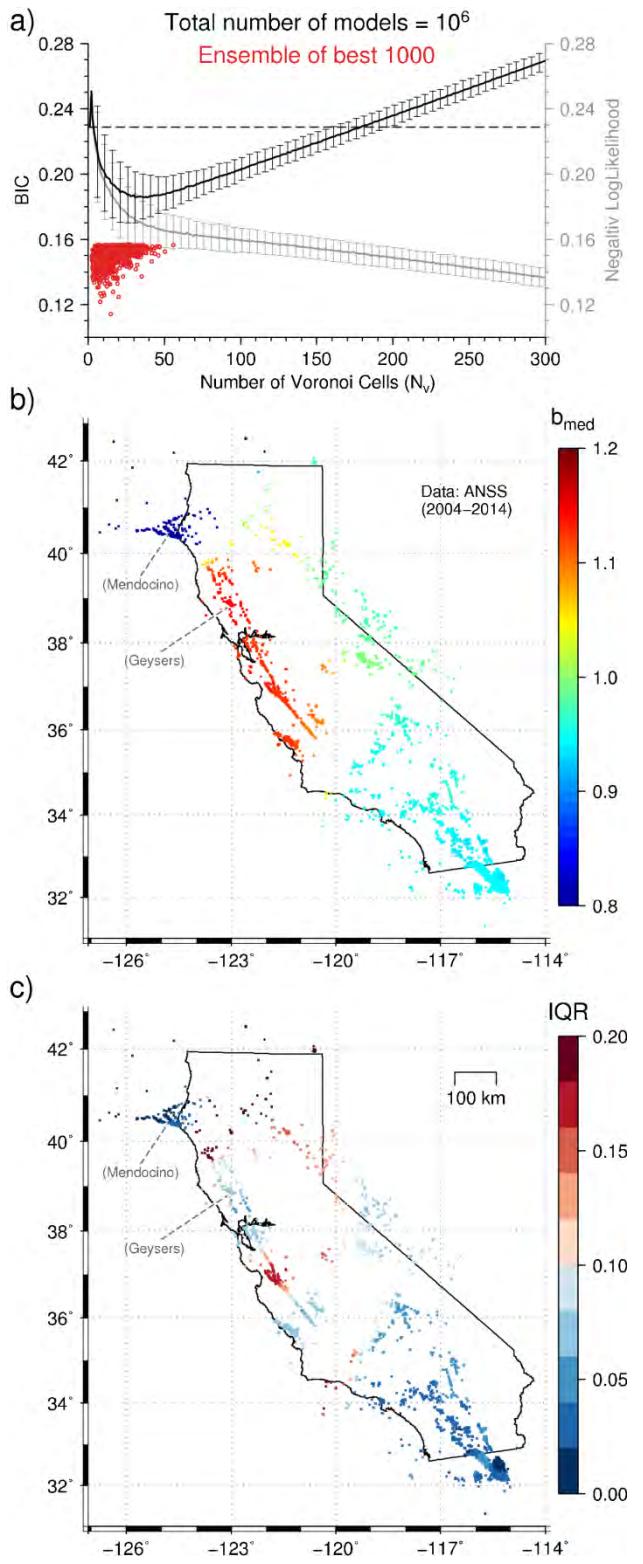


Figure 2-6 Same as Figure 2-1 for the last decade of the ANSS catalog (2004-2014, $m \geq 2.2$). Ranges for median b -values and IQRs are [0.79- 1.15] and [0.01-0.35], respectively.

Since our method is based on data driven partitions of various shapes and sizes, the notion of spatial resolution might be of particular interest. For this purpose, we investigate the average Voronoi areas of the top 1000 model segmentations used in the ensemble. These segmentations span the whole study region, which consists of 712,493 cells of 1x1km. Each cell is assigned the area of the Voronoi region it resides in. This is repeated for all the top 1,000 segmentations in order to report a mean area for each cell. The Voronoi areas are constrained by both the edges of the polygon defining the study region and the possible discrete locations for Voronoi seed placements. It is clear that these factors alone (which are independent of the model) will produce a particular pattern that should be regarded as an edge effect. To account for this we apply a correction technique inspired by Ouillon and Sornette (1996) who account for edge effects by normalizing their fractal fault measurements with measurements on randomly distributed faults within the same bounding polygon. To do this, we first compute a measure for the edge effects by randomizing the Voronoi node locations of the top 1,000 solutions and calculating their average Voronoi area distribution (Figure 2-7a). We then repeat the calculation using the actual optimal Voronoi seed locations obtaining a measure, which includes both the edge effects and the preferred model segmentations (Figure 2-7b). By normalizing the later with respect to the former we obtain an relative spatial resolution map which indicates if on average a region has been partitioned by a larger or a smaller segment with respect to random placement (Figure 2-7c).

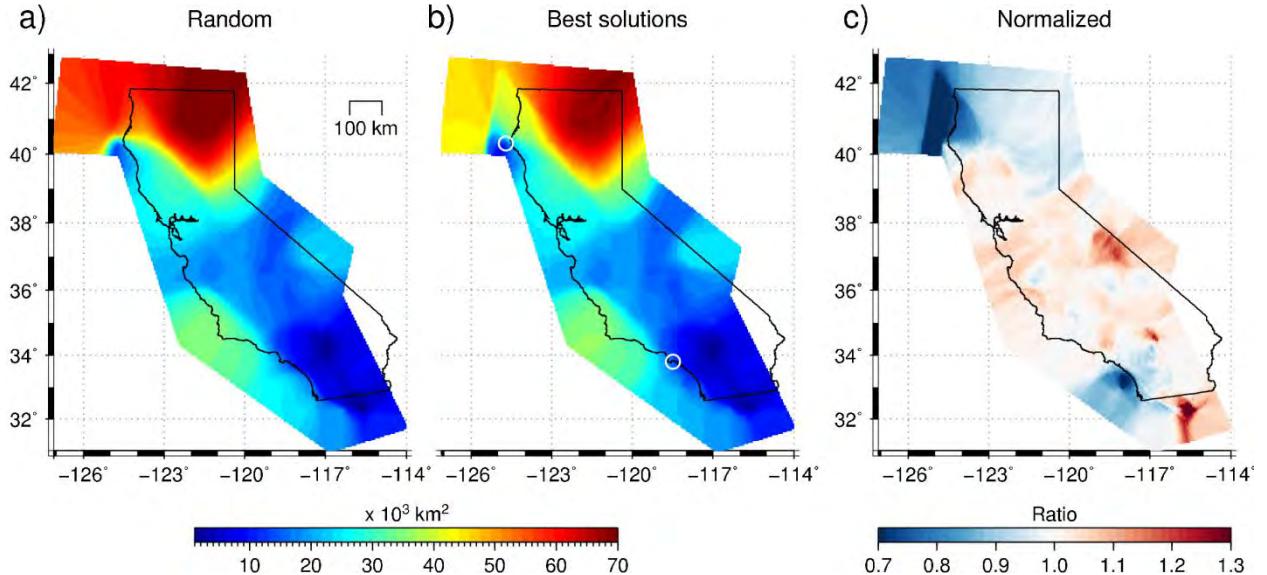


Figure 2-7 Average Voronoi areas for a) random placement of Voronoi seeds and for b) Voronoi seed positions of the best 1000 solutions, white circles highlight the Mendocino Triple Junction and Los Angeles c) The ratio of averaged areas (b/a) highlights regions covered by smaller (<1.0) and larger (>1.0) Voronoi cells with respect to random placement.

For instance Figure 2-7b shows that both the Mendocino Triple Junction and Los Angeles (highlighted with white circles) were contained by Voronoi cells having an average area of $\sim 20,000 \text{ km}^2$. However, the ratios in Figure 2-7c indicate that for Los Angeles this value is similar to the one obtained using random placement, whereas for the Mendocino Triple Junction the same value is more than 30% lower with respect to random placement. This indicates that optimizing the likelihood induced relatively smaller Voronoi partitions in this region.

When applied to the Californian seismicity of the last 30 years, our method reveals spatial b-value variations limited within the range 0.78-1.28 (Figure 2-5b). To investigate the source of this

variation, we focus on four regions with anomalous b-values: (1) Mendocino fault zone/ Cascadia subduction zone ($b=0.78\pm0.05$), (2) Geysers ($b=1.14\pm0.07$), (3) Palm Springs ($b=1.28\pm0.18$) and (4) Northridge ($b=0.90\pm0.04$).

(1) The Cascadia subduction zone, the Mendocino transform fault and the San Andreas Fault intersect at the Mendocino triple junction [Silver, 1971; Bird, 2003]. Complex interactions among the North American, Pacific, and Juan de Fuca plates are responsible for one of the highest rates of seismicity in continental North America [Furlong and Schwartz, 2004]. The low b-values in this region have relatively narrow confidence intervals, expressed by low IQR values, which point out a significantly different local frequency magnitude distribution.

(2) The geothermal region of Geysers has a high seismicity rate producing mainly small events and hence a significantly higher b-value. Due to the non-tectonic origin of these events the FMD in this region exhibits an upper magnitude limit. The characteristics of this region are further discussed in the conclusion section.

(3) The area with the relatively high b-value can be associated with the 1984 sequence of North Palm Springs. The sequence was initiated by a $m=5.9$ event followed by 1,437 aftershocks with magnitudes $m\geq1.8$. Kisslinger and Jones (1991) estimated the b-value of these events as 1.12. The ensemble map has relatively large IQR (≈0.18) values in this region, which undermine the statistical significance of the deviation with respect to $b=1$. Similarly to the synthetic tests, these high IQR values could also be indicative of potential transitions between regions with different b-values.

(4) Although the low tendency of the b-value in this region is not strong, it might be of interest due to the low ensemble variance associated with it. Seismicity in this region is dominated by the 1994 $m=6.7$ Northridge earthquake and its aftershocks. This sequence has been previously characterized by a b-value of 0.91 [Wiemer and Katsumata, 1999; Shcherbakov et al., 2005] which agrees well with the ensemble value reported here. Aftershock sequences of such large events can introduce bias in b-value estimations due to the sudden decrease in the network detection capabilities following the increased number of large aftershocks [Mignan and Woessner, 2012].

2.6.2 Results for the time period of 01.01.1984 – 01.01.2014

In this paper, we have analyzed the b-value variations in the spatial domain assuming that they are temporally constant. However, as mentioned in the section above, previous studies indicate that most observed b-value variations can be associated with specific events and thus can be regarded as time-dependent. To investigate the temporal stability of the initial results, we applied our method to a constrained dataset with only events occurring in the last decade (01.01.2004–01.01.2014), which contains 32,042 out of the 125,760 initially used events. In comparison, we observe that the spatial b-value variability is greatly reduced as small-scale anomalies such as (3) and (4) have disappeared (Figure 2-6b,c). The corresponding variations observed in Figure 2-5b,c can thus be attributed to seismic sequences prior to 2004. For entire Southern California we observe a uniform b-value of 0.95 with a low variance ($IQR\leq0.05$). This b-value also extends to the region of Long Valley and Nevada with an $IQR\leq0.1$ (Figure 2-6b,c). For Northern California, the low b-values around the Mendocino Triple Junction and the high b-values in the Geysers region are preserved when compared to the long term catalog.

As previously discussed, making statistically significant inferences of the results requires the ensemble median to be considered jointly with the ensemble variance. To facilitate a clearer

inference, we make use of a representation used by Felzer (2006), and show the median b-values, cut-off magnitude m_c and number of events with $m \geq m_c$ as a function of longitude (Figure 8). For each median value we further show both the 0.25-0.75 (IQR, Figure 2-6c) and 0.16-0.84 percentile ranges. This representation indicates that the median b-values in Southern California and Nevada are within the range of [0.92-0.98]. The Central San Andreas Fault Zone (SAFZ) has relatively higher median b-values that vary in the range of [1.09-1.13]. We observe that the IQR values in SAFZ are significantly higher compared to Southern California and Nevada. Considering their confidence intervals, only the regions of Cascadia and Geysers are distinguished significantly with low ($b=0.79 \pm 0.02$) and high ($b=1.15 \pm 0.06$) values.

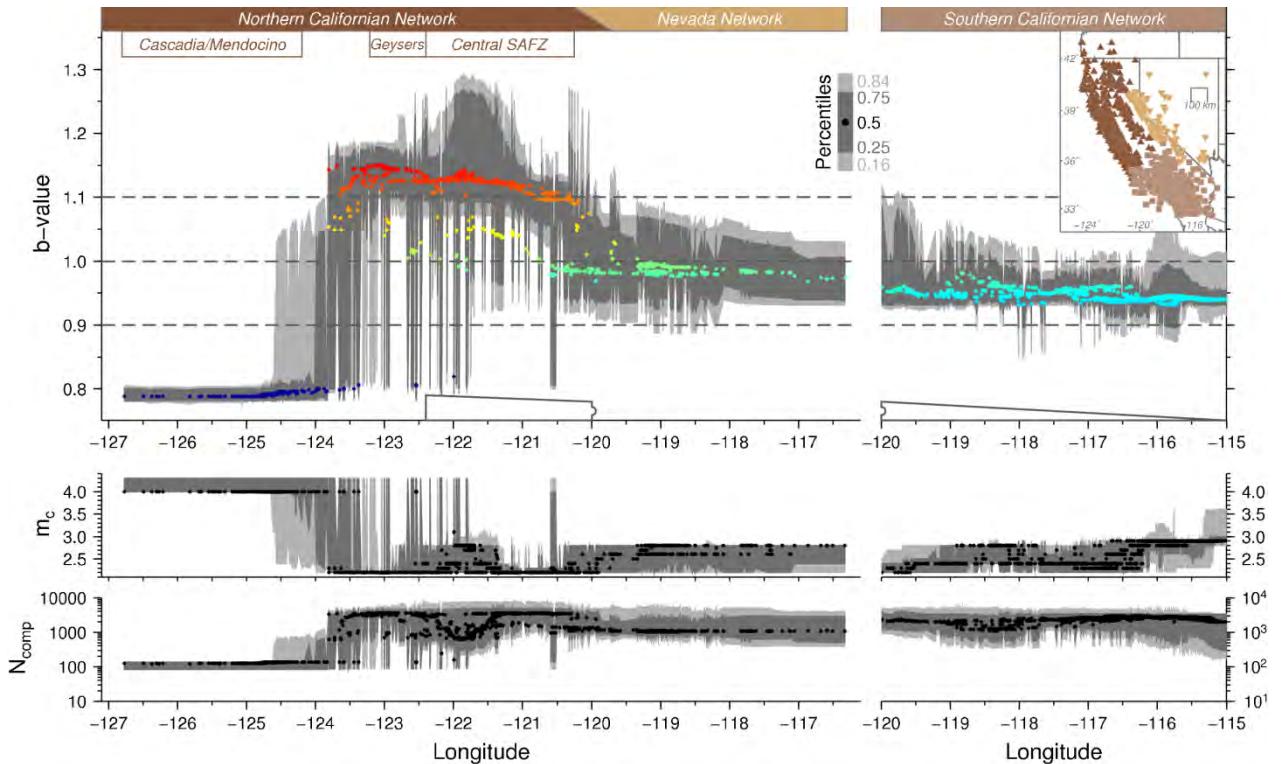


Figure 2-8 Median b-values from Figure 2-6b, median cut-off magnitude m_c and median number of events with $m \geq m_c$ (N_{comp}) as a function of longitude. Dark and light gray bars correspond to percentiles of [0.25 - 0.75] and [0.16 - 0.84], respectively. Top right inset shows Californian seismic stations, colored according to their network affiliation. Parts of Nevada and Southern California networks cover the same longitude range, thus we show values from the Southern California network separately (right panels).

2.7 Discussion

2.7.1 Magnitude inconsistencies between local networks

Although the amplitude of the variations in onshore California is small, we observe an abrupt offset in the b-value along the SAFZ occurring around longitude -120° (Figure 2-6a, Figure 8). This region coincides with the transition zone between the North and South Californian seismic networks. To investigate if the observed difference is caused by differences in the magnitude estimation procedure of the two networks, we analyzed the magnitude distribution of events that have been located by both networks. For this purpose we acquired both catalogs separately from the two earthquake data centers for the period of 01.01.2004 - 01.01.2014 (NCEDC: <http://quake.geo.berkeley.edu/ncedc/catalog-search.html>, SCECD: <http://www.data.scec.org/eq>

catalogs/date_mag_loc.php). To select identical events, we applied a joint criterion of maximum difference in detection time (3 seconds), location (10 m) and magnitude ($\Delta m < 1$), which resulted in a total of 1026 events. We find that small magnitudes ($m < 3$) are systematically overestimated in the North with respect to the South (Figure 2-9a). If the overestimation was in the form of a constant bias effective over the entire magnitude range, this would not have implications on the b-value. However, the overestimation decreases as a function of cut-off magnitude; consequently this results in a change in the slope of the probability density functions of the two networks (Figure 2-9a). We observe that for $m_c \geq 2.5$, where both datasets can be regarded as complete, the b-values calculated for the Northern network are consistently higher ($\Delta b \approx 0.1-0.2$) than those for the Southern network (Figure 2-9a). Since this analysis is based on a small subsample of the catalog, we do not expect an exact match with the b-values computed in the ensemble map. However, we maintain that such a network-related discrepancy can be used to explain the deviation between Northern and Southern California in Figure 2-6. It is important to note that similar network related issues have caused the emergence and misinterpretation of artificial seismicity features, such as the Byerly-Gutenberg discontinuity [Gawthrop, 2014].

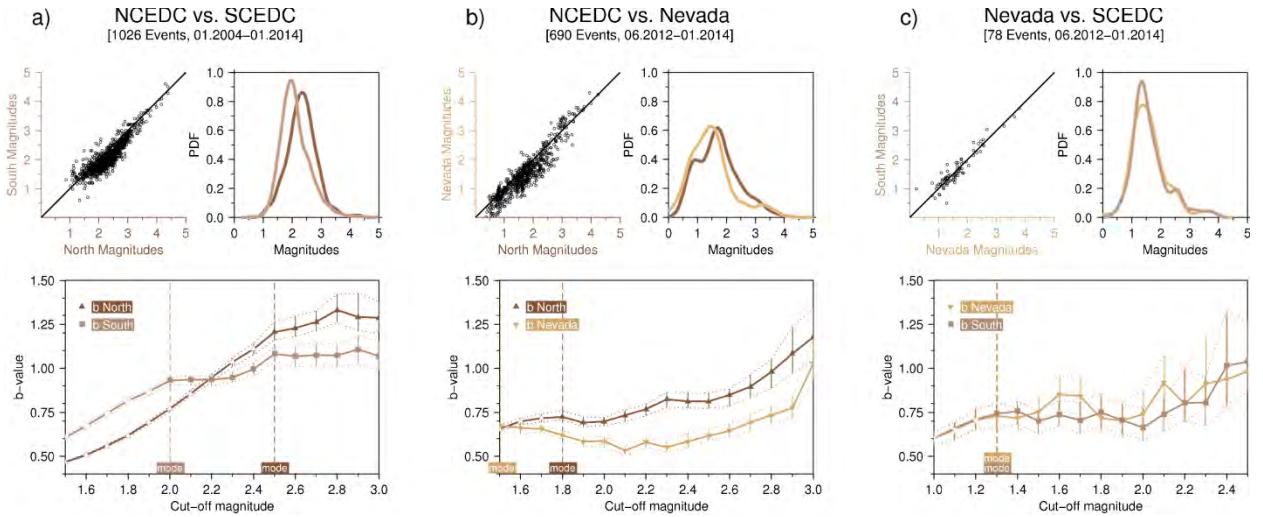


Figure 2-9 Magnitude and b-value comparison using matching events from a) NCEDC-SCEDC, b) NCEDC-Nevada, and c) Nevada-SCEDC. Top panels show magnitudes and magnitude probability density functions. Black line is a reference for equal magnitudes. Bottom panel shows b-value with interquartile range as a function of cut-off magnitude. Dashed lines denote the mode of each magnitude distribution.

Furthermore we propose that the same reasoning can be used in Northern California since the region is covered by two different seismic networks: the Northern California Seismic network operated by USGS and the Nevada network operated by the University of Nevada, Reno. We observe that the drop in the b-value between west ($b \approx 1.12$) and east ($b \approx 0.98$) Northern California coincides with the network boundary (top bar in Figure 2-8). To test this reasoning we performed the same analysis using events located by the Nevada Network. The publicly available dataset includes events only after 06.2012 (<http://www.seismo.unr.edu/Earthquake>). Thus, the comparisons with respect to the Nevada network were based on a limited time period (Figure 2-9b for Northern and Figure 2-9c for Southern Californian network). Similarly to the results of the South-North networks comparison, we observe a discrepancy between the reported magnitudes for the 690 matching events between the Nevada and North California networks. This leads to a systematic

shift in the estimated b-values. The subset of events detected by the Nevada network has a lower b-value compared to the North California network, which is similar to the spatial trend observed in Figure 2-6 and Figure 2-8. For the comparison with the South Californian network (Figure 2-9c), we observe a close correspondence in both the shapes of the magnitude PDFs and the estimated b-values. However, the results of this comparison should be regarded with care due to the limited number of matching events (78 events).

2.7.2 Bias and errors in magnitude estimations

It might be tempting to infer the consistently low b-values ($b \approx 0.8$) reported offshore the Mendocino triple junction as being indicative of a distinct tectonophysical process. However, because the region is located far from the California network stations (low azimuthal coverage) and has the smallest sample size ($\sim 130 m \geq 4$ events, Figure 2-8), we need to consider the possibility of bias due to systematic magnitude errors. To demonstrate the possible effect of station coverage and configuration on magnitude estimation, we conduct a synthetic earthquake location test in 2D. For an earthquake location we first calculate true P arrival times at all stations assuming a constant $V_p = 6$ km/s. We then perturb these arrival times with a Gaussian pick in g time error $\sigma = 0.1$ s. Using these perturbed P picks for each station pair we obtain hyperbolas that indicate possible locations of the earthquake. Discretizing the grid at a resolution of 100m, we convert the hyperbolas into discrete probability mass distributions that are then multiplied element-wise to calculate the joint probability of the event location (Figure 2-10). We sample the resulting spatial probability distribution with a total of 100 points. Finally, for each of the sampled locations we calculate a magnitude average over all stations based on the local magnitude formula [Richter, 1935]:

$$M_L = \log_{10}(A) + 2.76 \log_{10}(D) - 2.48 \quad 2-7$$

where A is the P wave amplitude at 0.8 Hz in micrometers and D is the epicentral distance in kilometers. Since we assume no error on the amplitude measurement, A is calculated from the true magnitude M_L^* and true epicentral distance D^* using equation (2-7). Thus the magnitude error $M_L - M_L^*$ simplifies to:

$$M_L - M_L^* = 2.76 \log_{10} \left(\frac{D}{D^*} \right) \quad 2-8$$

Figure 2-10 shows how different network configurations with low azimuthal coverage can result not only in large deviations from the true magnitude but also in systematic biases. Actual earthquake locations are prone to a more extensive cascade of errors including phase misidentifications, velocity model errors, station delays and asymmetric picking time errors. It is likely that for poor station coverage and limited number of events such errors can lead to biases in the estimated magnitudes and consequently in the b-value.

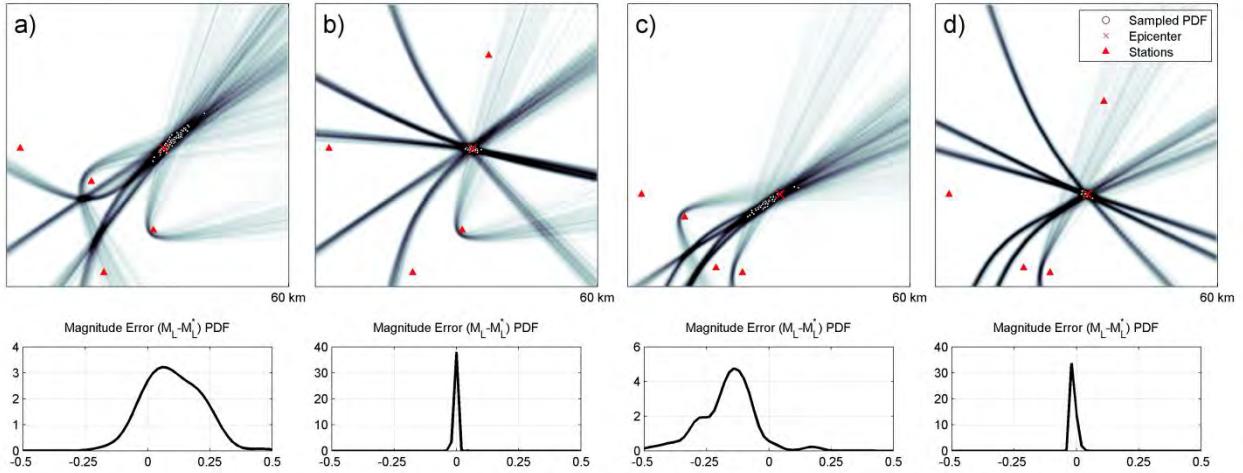


Figure 2-10 Distribution of magnitude error (estimated-true) for different network configurations assuming a P pick error of 0.1s and $V_p = 6$ km/s. Stations are shown with red triangles, true epicenters with a red cross and samples from the location PDF with white dots. Low azimuthal coverage results in large magnitude errors and bias: overestimation in a) and underestimation in c). Moving a single station to increase the azimuthal coverage (shown in b) and d)) reduces the error and bias significantly in both configurations.

In his paper, investigating the global Centroid Moment Tensor (CMT) [Ekström *et al.*, 2012] catalog, Kagan (2010) observed that moment magnitude errors decrease with increasing magnitude causing inflated b-value estimates. However, in his analysis of the GeoNET New Zealand catalog, Rhoades (1996) observed that magnitude errors increase linearly with magnitude leading to underestimation of b-values. Such overall positive or negative biases cannot explain regional differences as all b-values would be equally affected. If the magnitude error function changes with magnitude, however, regions spanning different magnitude ranges would be biased differently, leading to regional b-value differences. The ANSS catalog does not report magnitude errors, however as [Werner and Sornette, 2008] pointed out in their study, it is possible to obtain a “raw” catalog from the NCECD (the authoritative network for North California) that include median absolute deviation (MAD) of the reported magnitudes. We obtained all events with $m > 2$ for the time period of 01.01.1984-01.01.2014 and depth < 30 km resulting in a total of 80,719 events with reported MAD values. We group events at 0.5 magnitude intervals and plot box plots of their MAD values. We observe that the median MAD value remains stable for $2 < m \leq 4$ and then increases rapidly for $m > 4$ as a function of magnitude (Figure 2-11a). As demonstrated by Rhoades (1996) the MAD increase at large magnitudes would cause the b-value to be underestimated in regions with large magnitudes (Figure 2-11b). The region offshore the Mendocino triple junction contains the largest earthquakes in our study region and has the highest m_c value of $m_c = 4$ (Figure 2-8). Thus it is safe to assume that the b-value in this region is underestimated with respect to onshore regions that have lower m_c values.

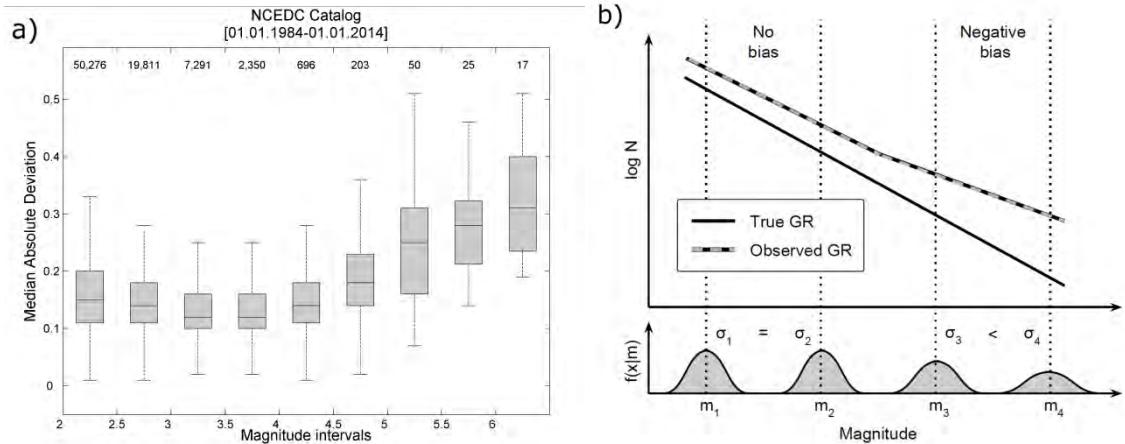


Figure 2-11 a) Box plot of median absolute deviation versus magnitude for the Northern California NCECD catalog, 1984-2014, binned at M0.5 intervals. The corresponding number of events with reported MAD values is given in the top. Box height corresponds to the IQR and whiskers extend to the most extreme values within 1.5 IQR in both directions. b) Schematic plot of the bias in the estimated GR relation in the presence of magnitude uncertainties (adapted from [Rhoades and Dowrick, 2000]).

To determine whether this reasoning can be generally applicable to explain low b-value anomalies observed in other regions, we revisit the studies of Kagan and Rhoades, repeating their error analysis with more up-to-date datasets (see Supplementing Information). We observe that for shallow events in the CMT catalog, the magnitude error decreases for $5 < m \leq 6.5$ but increases for $m > 6.5$ (Figure S1a1). Kagan also observed a similar break-point at $m=6.5$ but argued that the small number of events available at the time cannot support a more complex trend beyond this range. For deep events, the magnitude errors are generally smaller and decrease consistently throughout the whole magnitude range (Figure S1a1). For the GeoNET New Zealand catalog, we observe, similarly to the NCECD catalog, that the median magnitude error remains stable ($2.5 < m \leq 5$) and then increases again ($m > 5$) as a function of magnitude (Figure S2a1). For deep events, we observe that the median error remains fairly constant for all $m > 3$ events. These analyses suggest that the bimodal relationship between magnitude errors versus magnitude might be typical in shallow seismicity regions, and hence cause biases in the b-value estimations.

2.8 Conclusions

In this paper we introduced a fully automated, objective method for spatial b-value variations estimation, which does not involve any adjustable input parameter. The core components of the method are (1) Voronoi tessellation, (2) penalized likelihood in terms of BIC and (3) ensemble inference. The Voronoi tessellation allows for adaptive spatial resolution, putting more emphasis on areas that are highlighted by data availability. The flexible partitioning makes the delineation of any irregular shape possible. The BIC acts like a regularizer between the data and the complexity required to model it. The ensemble averaging procedure ensures that results comprise most likely models of different complexities.

We used synthetic tests to demonstrate the efficiency of the BIC for determining optimal complexity. An alternative approach might be cross validation which involves separating the data into two complementary sets and evaluate the performance of the model derived from one set (i.e. the training set) based on the likelihood of the other (i.e. the validation set). While this approach yields satisfactory results in other fields, its application to spatial b-value mapping remains

problematic due to the fact that the results depend greatly on the selection of the validation set. Following the common practice of using a random selection, one would end up with mostly small events in the validation set. Such a validation set would prefer models with high b-values, since they provide higher likelihoods for small events. Yet, if an arbitrary number of large earthquakes are set aside for validation (as they are of greatest interest) then many parts of the spatial domain will remain invalidated. In this regard, the BIC not only helps us avoid such arbitrary choices but also allows the use of the full dataset.

Applications to the Californian seismicity catalog of the last 30 years indicate that there are statistically significant b-value variations in the range of [0.78-1.28] (Figure 2-5b). As previously pointed out by Kagan [*Kagan*, 1999b, 2002, 2010], it is possible to attribute observed b-value variations to magnitude errors and temporal variations in the detection capabilities. Since we would expect these effects to subside with time, we have analyzed separately the last 10 years of the seismic catalog and found that the range drops to [0.79-1.15] (Figure 2-6b). We observe the following three main features: (1) abrupt change in b-value between Northern California and Southern California / Nevada (Figure 2-8), (2) consistently high b-values in the geothermal area of Geysers and (3) consistently low b-values offshore the Mendocino triple junction.

(1) We showed that discrepancies in magnitude estimations between different networks alone result in b-value variations (~ 0.2 , Figure 2-9) that are comparable or even greater than the reported changes (~ 0.15 , Figure 2-8). This effect should not be overlooked when dealing with composite catalogs featuring different magnitude estimation practices of the contributing agencies. A possible way to address this issue is to obtain empirical magnitude relationships (see Figure 2-9) between different networks. However, this would require large amount of matching events and that the magnitude estimation practices do not change with time [*Agnew*, 2010].

(2) The Geysers seismicity is thought to arise from increased pore fluid pressure and volumetric contraction due to water injection [*Boyle and Zoback*, 2014]. Many local studies have characterized this region by an anomalously high b-value (e.g [*Helmstetter et al.*, 2007]). Our method detects the same pattern in a fully automated fashion. However, it is expected that the frequency of large events will deviate from the GR law, which ignores the fact that rupture surfaces and stimulated volumes are finite [*Shapiro et al.*, 2013]. Accordingly we observe relatively low p-values that indicate the poor performance of the unbounded GR model in this region (Figure 2-14). Another noteworthy observation regarding Geysers is that although the geothermal region has a small spatial extent, the low b-values associated with it span larger extents. This is due to the relatively large number of complete events in this region compared to its surrounding. This large density tends to dominate its vicinity through the fully random segmentation search that is often unable to make a tight delineation of the small region. The proposed method is based on data-driven segmentation in the absence of prior knowledge; however it is possible to improve the method's performance by introducing prior knowledge of regions with distinct geophysical characteristics. For instance by ensuring that Geysers region is tightly delineated in all randomly generated segmentations.

(3) We have demonstrated that the magnitudes reported offshore the Mendocino triple junction, which has consistently low b-values ($b \approx 0.8$), are likely to have larger magnitudes errors compared to the smaller onshore events. Simple synthetic tests suggest that the low azimuthal coverage in this region is likely to contribute to larger magnitude errors. Such a magnitude error difference can give rise to underestimation of the b-value [*Rhoades*, 1996; *Rhoades and Dowrick*,

2000]. These findings preclude us from making physical interpretations based on the observed b-value anomaly.

The fact that the b-value map for the period of 1984-2014 differs from the more recent one of 2004-2014 gives rise to two possible explanations: (I) there are actual temporal b-value variations in California and (II) the results for 1984-2014 may have been affected by temporal completeness issues and changing network practices. We believe the truth lies somewhere in between. Since our method does not feature segmentation in the time domain it is difficult to give a definitive answer. Nevertheless, based on data quality, we put more confidence on the results obtained from the last 10 years of data.

It is important to point out that our results are irreconcilable with many recent studies: for instance, Wiemer and Schorlemmer (2005) presented a spatial b-value map for whole California featuring variations saturated in the range of [0.5-1.5]. It might be worth noting that their study was conducted on a declustered catalog while the results presented here are obtained from a non-declustered catalog. The same authors reported a similar range in their Parkfield study conducted over a more limited non-declustered dataset [Schorlemmer and Wiemer, 2005]. Recently, [Tormann et al., 2014] conducted a b-value analysis on major Californian faults with a constant radius of 7.5 km and minimum 50 events and estimated probabilities of m6+ events. Such b-value maps, and many published maps obtained using the classic method, feature abrupt changes of the b-value at small scales. Until now, the prevalent inference of this heterogeneity has been that b-values can be used as proxy for stress, which is also believed to be heterogeneous in the upper crust. Our results indicate that the b-value variation is smaller, observed over much larger scales and in most cases can be explained by magnitude errors and magnitude discrepancies due to network effects. Thus they, do not support any correlation with stress, be it tensile or compressive. These conclusions are supported also by previous studies: Frohlich and Davis (1993) reported similar b-value ranges after rigorous analysis of several global catalogs, Imoto (1987) and Ogata and Katsura (1993) observed such variations only at large scales in New Zealand and Japan using penalized likelihood approaches.

Throughout this paper we have highlighted drawbacks of the previously used methods and how the proposed penalized likelihood-based method addresses these. Nonetheless, there is room for further improvement.; As we have demonstrated by processing two different time periods, the catalog contains temporal features (such as the 1994 Northridge and 1984 Palm Springs sequences). The Voronoi partitioning can be readily applied in n -dimensions, and thus we can extend our method to partition not only the spatial but also the temporal domain. Such a partitioning will enable us to capture the spatiotemporal changes in m_c . This spatiotemporal flexibility of m_c will allow for the use of the recently introduced entire magnitude range angular FMD that is compatible with A i's maximum likelihood b-estimator [Mignan, 2012]. The angular FMD will facilitate the use of both complete and incomplete events, which have often been discarded previously.

To summarize our conclusions regarding the last decade of recorded seismicity:

- 1) The spatial b-value in the well instrumented parts of mainland California is limited to the range of $[0.94 \pm 0.04 - 1.15 \pm 0.06]$.
- 2) Except for the geothermal area of Geysers, discrepancies in magnitude estimations between different networks and bias due to magnitude errors alone can explain the observed b-value variations. This precludes non-testable hypotheses such as crustal stress variations.

Our results for California support the recent decision of the 2014 Working Group on California Earthquake Probabilities to use a uniform b-value for the entire state of California [Field *et al.*, 2014]. Lastly, we underline that the results of our method are objectively quantifiable in terms of likelihood; this is not the case with the classical methods, which have been wildly applied up till now. It must be noted that in the absence of an objective evaluation criterion, any method can turn into a data distortion tool or, as Coase (1995) puts it, “if you torture data enough, nature will always confess”.

2.9 Appendix A: Comparative Synthetic Test Based on Californian Seismicity

In this appendix we demonstrate the reliability of the proposed method and highlight its advantages over the previously used methods via a synthetic test. We consider the same dataset as in Section 5.2, namely the 32,042 events occurring in the period (01.01.2004-01.01.2014) with $m \geq 2.2$. We keep the original event locations and define three distinct b-value zones based on the results obtained in Figure 2-6 and Figure 2-8. The zones and their assigned b-values are: Cascadia/Mendocino ($b=0.8$), Geysers ($b=1.2$) and Southern California/Southern Nevada ($b=0.95$) (see Figure 2-12a). We assigned magnitudes to the events in each zone from a GR distribution generated with the zones' b-value. The magnitudes of the events outside these three zones were generated from $b=1$. Similarly to the analysis conducted on the real data, we perform 2,000 random tessellations at each Voronoi increment. We stop the random tessellation procedure at $N_v=40$ since all the resulting models perform worse than the single b-value model at this level. The top 80 of all 80,000 models are assembled to obtain the median b-value and IQR distributions (Figure 2-12a, b). The results are stable with respect to the assembled number of top solutions; we present the case of 80 since it is consistent with the ratio used in the application to the real data.

In order to put these synthetic results into the perspective of the currently used methods, we analyze the same dataset with a constant radius approach. As pointed out in Section 2, the data overlap in the classical methods hinders the computation of an overall likelihood that is needed to facilitate a comparison. That is why we use non-overlapping grids of different cell sizes. The b-value is calculated in each grid cell containing more than 50 events (as commonly used in classical b-value mapping). The overall likelihood and BIC terms are calculated using the formulas given in equations (2-3) and (2-6). Similarly to the case of Voronoi tessellations, the number of free parameters is $k=2*N_g$ where N_g is the number of non-empty grid cells. In Figure 2-13 we present the resulting maps with their BIC scores for grid sizes of 10, 25, 50, 100, 200 and 300 km. We point out that all of the top Voronoi models regardless of their complexity (varying from 3 to 17 cells), outperform the gridded segmentations in terms of BIC.

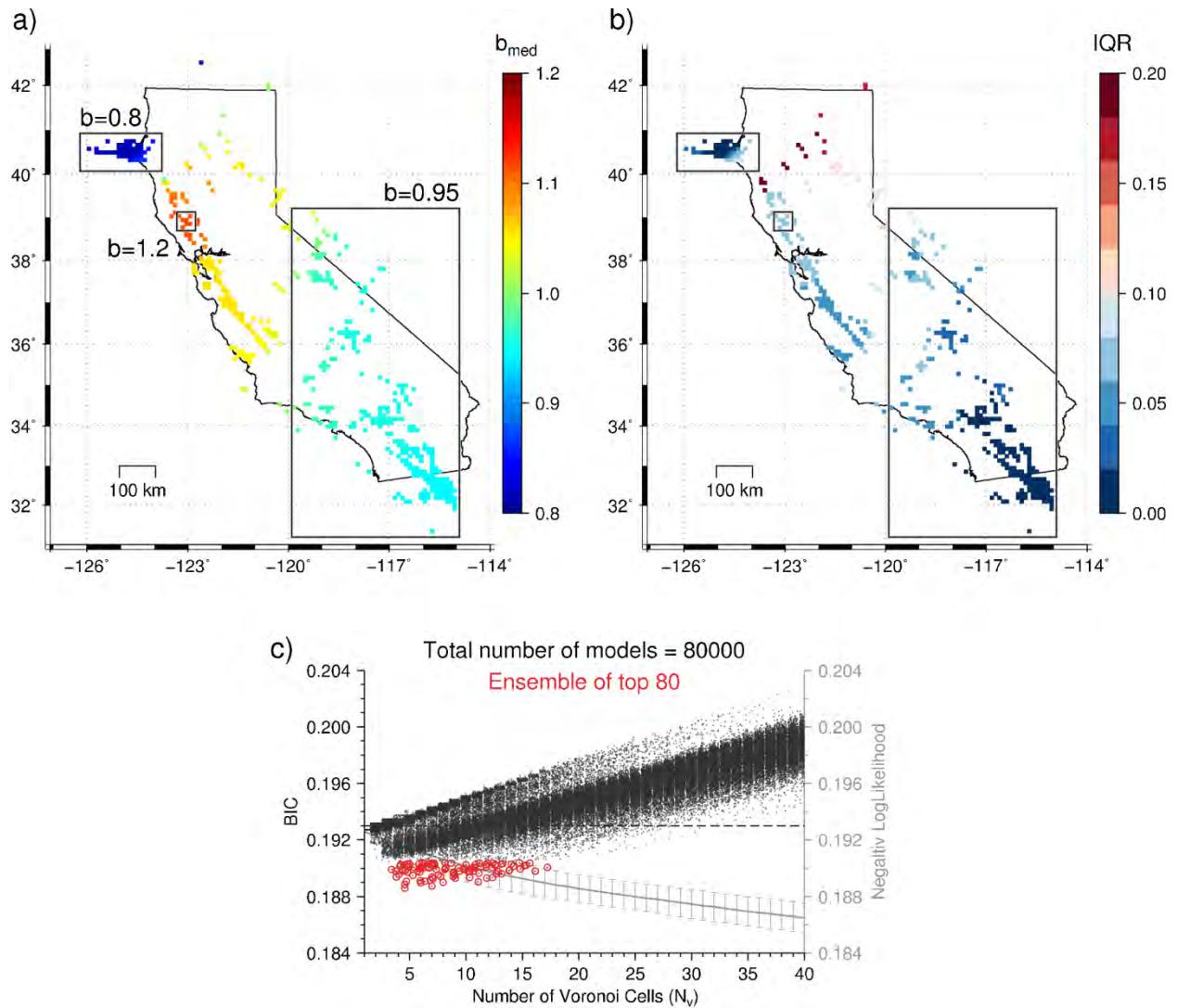


Figure 2-12 Ensemble median b_{med} (a) and IQR(b) for the synthetic dataset based on the [2004-2014] catalog with magnitudes generated from the annotated b -values in each zone. The remaining magnitudes are generated from $b=1$. c) The entire solution space for the synthetic catalog: BIC scores (black dots) and negative log-likelihood (gray line) as a function of number of Voronoi cells. Error bars correspond to one standard deviation. Red circles denote the BIC of the best 80 solutions.

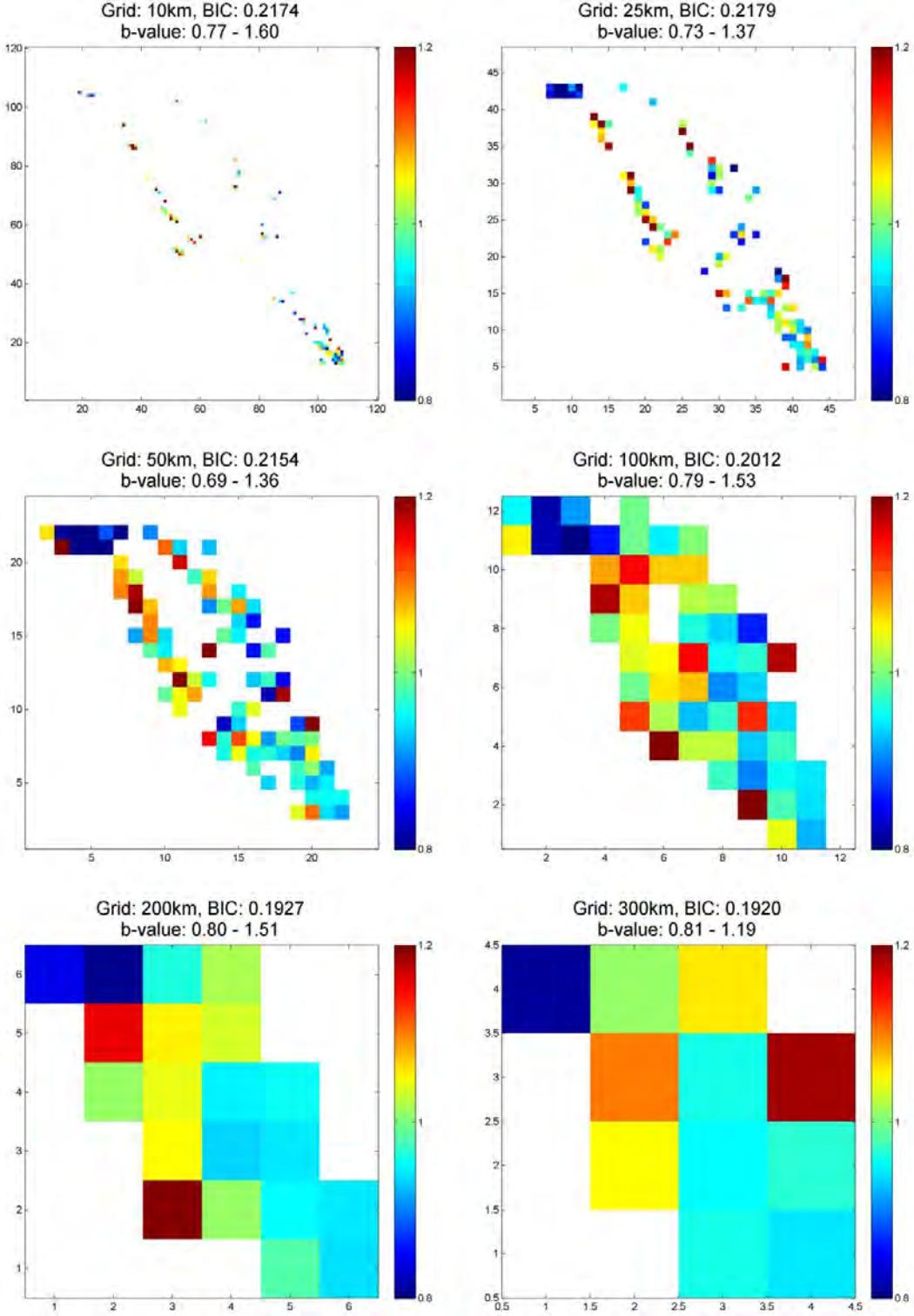


Figure 2-13 Same synthetic dataset as Figure 2-12 analyzed via constant sized non-overlapping grids with increasing sizes from 10 to 300 km with $N_{min}=50$. The resulting BIC scores and b-value ranges are given in the title of each grid plot.

2.10 Appendix B: How plausible is the GR law for Californian seismicity?

In this appendix we acknowledge the fact that any dataset can be fitted with any parametric model regardless of the model's capacity to describe the data. We investigate the plausibility of the GR law in each of the ensemble Voronoi partitions using the Kolmogorov-Smirnov (KS) distance

as a measure for the goodness of fit. To this end we follow the methodology proposed by Clauset et al. (2009): For each Voronoi region the lower cut-off (m_c) and the b-value are assigned jointly by minimizing the KS distance between the empirical CDF and the analytical CDF obtained using the GR law exponent estimated from events with $m \geq m_c$. Note that this is the same method used in the analysis of Californian seismicity presented in this study. Next, we compare the observed KS distance with KS distances obtained by applying the same procedure to several randomly generated synthetic datasets. For each dataset containing a total of $N_T = N_c + N_i$ events, where N_c and N_i are the number of events with $m \geq m_c$ and $m < m_c$ respectively, we generate N_c synthetic magnitudes from a GR distribution with the b-value estimated from the original dataset. N_i magnitudes are generated by bootstrapping the original magnitudes with $m < m_c$. This synthetic dataset is then used in the same way as the original one to estimate the optimal cut-off, b-value and KS distance. The procedure is repeated several times keeping count of the instances where the KS distance measured in the original data are smaller than the one obtained from the synthetics. The reported p-value is the ratio of the number of such instances to the total number of simulations. A high p -value indicates that the GR law is plausible since synthetic datasets rarely perform better in terms of KS, while a low p -value would indicate that the observed KS distance is large compared to what one would expect from a true GR law and thus it is not a plausible model for the data. Clauset et al. (2009) consider $p > 0.1$ to be sufficiently large enough not to rule out the candidate model.

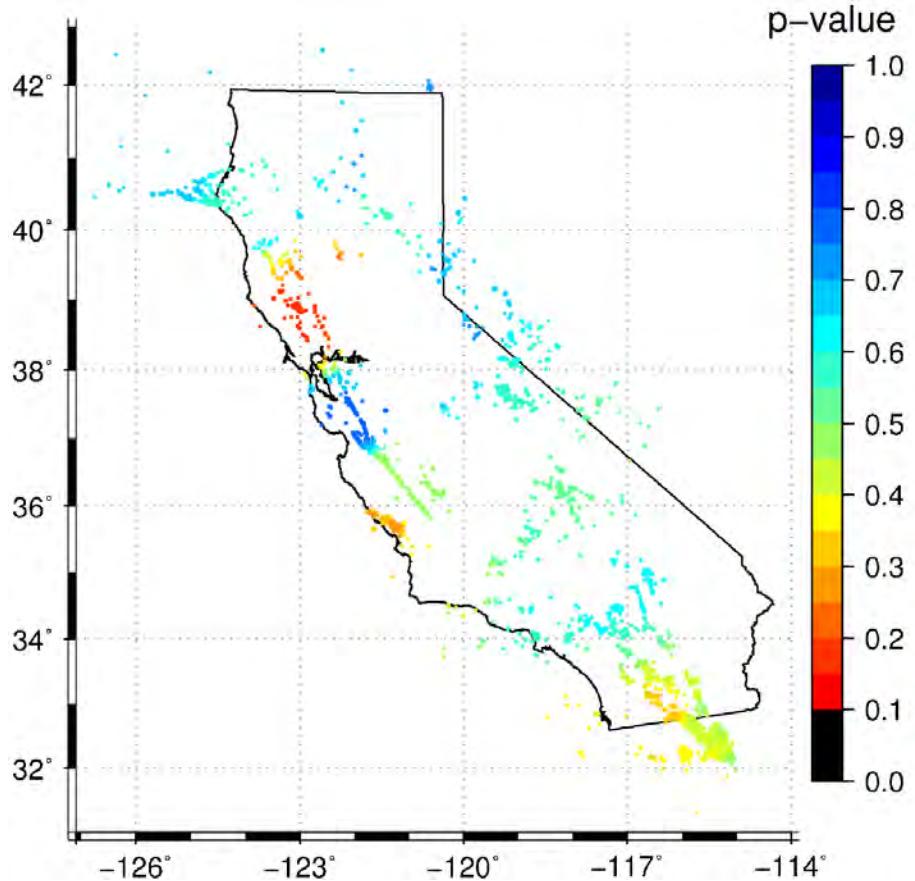


Figure 2-14 Median p -values obtained by 1000 Monte-Carlo simulations with the top ranking partitions for the 2004-2014 California catalog (Fig. 6). The p -value represents the probability of observing a worse GR law fit (in terms of Kolmogorov-Smirnov distance) estimated from the Monte-Carlo simulations. Values lower than $p_{min} = 0.1$ indicate implausibility of the GR law.

Here we present results of the p -value analysis conducted for the 2004-2014 Californian seismicity. We calculate a p -value for each Voronoi region of all the top ranking tessellation models (Figure 2-6) using 1000 synthetic runs. The median p -values are assigned to the same non-empty 1x1 km cells used in Figure 2-6 (Figure 2-14). For better visualization we also present the p -value distributions as a function of longitude (Figure 2-15). The results suggest that the GR law is plausible in all the top models that contribute to the ensemble. The spatial variability of the p -value can be attributed to several aspects related to the generation of the synthetics and the consequently calculated KS distance. For instance the low p -values near the Californian state border with Mexico are most likely due to the fact that small events falling outside the state boundaries are not included in the catalog, this would cause instabilities in the estimation of the lower cut-off, increased KS distances and thus low p -values. The e yser's region also features low p -values, however in this case the most likely cause is the actual frequency magnitude distribution, which is expected to deviate from a power law at small cut-off due to the fact that the magnitudes in this range exhibit swarm characteristics [Shapiro *et al.*, 2013].

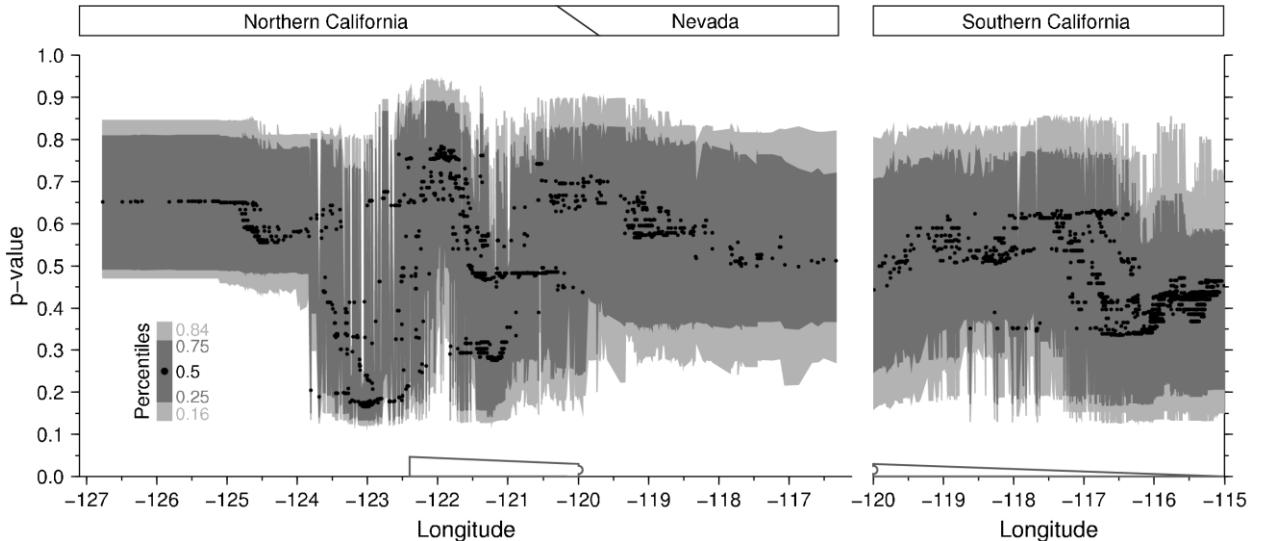


Figure 2-15 Median p -values from Fig. B1 as a function of longitude. Dark and light gray bars correspond to percentiles of [0.25-0.75] and [0.16-0.84], respectively.

2.11 Acknowledgements

We would like to acknowledge Guy Ouillon for numerous discussions, constructive criticism and his constant support throughout this research. We appreciated valuable suggestions from Edi Kissling, Didier Sornette, Malcolm Sambridge, Shyam Nandan and a thorough review by Yan Kagan. We thank Jeroen Tromp for raising the question of parameter dependence of classical b-value studies during the 2012 SED directorship talks and thus triggering the research, which has eventually led to this publication. Our thanks extend to Robert Geller for his encouragement in the final stage of this work. This study has also benefited from thought-provoking debates within the SED statistical seismology group. S.H. thanks Stefan Wiemer for his support and feedback. Figures were produced using the generic mapping tools (<http://gmt.soest.hawaii.edu/>). Earthquake data were obtained from ANSS; <http://www.ncedc.org/anss/catalog-search.html>, Northern California; <http://quake.geo.berkeley.edu/ncedc/catalog-search.html>, Southern California; http://www.data.scec.org/eq-catalogs/date_mag_loc.php, Nevada;

<http://www.seismo.unr.edu/Earthquake>, New Zealand; <http://quakesearch.geonet.org.nz/> and Global Centroid Moment Tensor database; <http://www.globalcmt.org/CMTsearch.html>.

Chapter 3

Improved seismicity forecast with spatially-varying magnitude distribution

by

Stefan Hiemer, ETH Zürich
Yavor Kamer, ETH Zürich

under review in
Seismological Research Letters, as of August 2015

3.1 Abstract

In this article we present a time-independent earthquake rate forecast for California. Our model features spatial variations of the Gutenberg-Richter b-value using the method of Kamer and Hiemer, (2015) . We account for lessons learned from the outcome of the Regional Earthquake Likelihood Models (RELM) experiment and use RELM's framework to investigate different concepts for modelling the spatial distribution of seismicity. Our resulting forecast shows a significant information gain with respect to all other first-generation RELM mainshock forecasts, including the much-praised smoothed seismicity model by Helmstetter et al., (2007) . Our findings indicate that large-scale b-value variations are a considerable feature for increasing the skill of Californian seismicity forecasts. We underline the importance of statistical rigor when implementing earthquake occurrence hypotheses. Our results have implications for seismic hazard studies, where the b-value is either chosen as a regional constant or varies spatially between local zones. Future improvements of our model may help to overcome the inherent subjectivity in choosing either of these approaches.

3.2 Introduction

Rigorous testing of earthquake occurrence hypotheses is key to enhance both our statistical models and physical understanding of seismicity. The Regional Earthquake Likelihood Models (RELM) experiment evaluated the reliability and skill of a variety of probabilistic seismicity forecasts for California over the five-year target period 2006-2011 in a truly prospective fashion (using all $m \geq 4.95$ events, [Schorlemmer et al., 2007]). RELM can be regarded as a milestone for forecasting-related research in seismology because it successfully addressed a major deficiency faced previously: the lack of a controlled environment for model performance evaluations. Modellers expressed their forecasts in a common format and submitted them to a testing center that used pre-agreed statistical criteria to test the models against future seismicity. Accordingly, such a hands-off approach ensured the prevention of retrofitting model parameters or of any other hindsight model adjustments. The spirit of RELM has been subsumed by the broader initiative of CSEP (Collaboratory for the Study of Earthquake Predictability; e.g. [Jordan, 2006; Zechar et al., 2009]) that conducts similar forecasting experiments in various regions in the world (for an overview visit www.csep-testing.org).

The first-generation RELM models were based on diverse input data and modeling approaches, including non-parametric density estimates of past earthquake locations, pattern recognition in seismicity rate changes, geodetic estimates of strain rates from GPS data, geologic fault slip rates, and physics-based numerical earthquake simulations (see [Field, 2007] and references therein). Many RELM and CSEP related studies have focused on the revision and improvement of evaluation metrics [Zechar et al., 2010; Imoto et al., 2011; Rhoades et al., 2011; Eberhard et al., 2012; Molchan, 2012; Tarom et al., 2013], on model combination and ensemble forecasting [Marzocchi et al., 2012; Rhoades et al., 2014; Shebalin et al., 2014], on the development of alternative testing approaches [Clements et al., 2011; Lee et al., 2011; Sachs et al., 2012; Smyth et al., 2012; Bray and Schoenberg, 2013; Ogata et al., 2013], and on the future design of earthquake forecasts experiments [Werner and Sornette, 2008; Lombardi and Marzocchi, 2010;

Werner et al., 2010]. However, little has been done regarding both the improvement of the models themselves and the assessment of R L's results in terms of what we have learned about the earthquake occurrence hypotheses governing the models.

The first-order results of RELM [Zechar et al., 2013] revealed the kernel smoothed-seismicity model by Helmstetter et al., (2007) as the most skilful forecast (hereafter referred to as the HKJ model). Their underlying working hypothesis is that the spatial density of past earthquakes indicates where future earthquakes are most likely to occur. The model's magnitude distribution is essentially based on a tapered Gutenberg-Richter distribution using a uniform b-value across California. By contrast, the Wiemer and Schorlemmer (2007) forecast was built on the hypothesis that spatial variations of the Gutenberg-Richter b-value hold important information for forecasting future rates of earthquakes. Their forecast was found to be inconsistent with the spatial distribution of target events; a result that was not only reported by the prospective RELM experiment in California [Zechar et al., 2013], but also by retrospective CSEP evaluations in Italy regarding a similar model [Gulia et al., 2010; Werner et al., 2010]. Despite its inconsistency, the Wiemer and Schorlemmer (2007) model finished among the runner-up forecasts when ranking all RELM mainshock models in accord to their information gains (notice that the three best-performing models are based on seismicity data alone, [Rhoades et al., 2011; Zechar et al., 2013]). This result thus might indicate that there is potential for the gainful use of spatially varying b-values in probabilistic earthquake rate forecasting.

In the present study we investigate the reasons for the deficiency of the Wiemer and Schorlemmer (2007) forecast (hereafter referred to as the WS model) and we address the question whether the rejection of the model in the RELM framework implies the rejection of its underlying earthquake occurrence hypothesis. In that regard it is important to note that the WS model was based on a b-value mapping technique that involves the somewhat arbitrary choice of several mapping parameters, such as mapping radii and minimum number of events for reliable b-value estimation (see review article by [Wiemer and Wyss, 2002]). [Tormann et al., 2014] used a similar approach to conduct a systematic survey of b-value images along Californian faults. However, [Kamer, 2014] argued in his comment article that the results of the classical b-value mapping technique heavily depend on external parameters, whose choice requires a priori knowledge of the spatial b-value distribution. Moreover, it has been underlined that many conclusions based on b-value variations should be viewed with caution mainly because these studies often fail to acknowledge sources of errors (e.g. under-sampling, magnitude errors, non-homogeneous detection capabilities, lack of statistical rigor; [Frohlich and Davis, 1993; Kagan, 1999b, 2010; Amorese et al., 2010; Kamer and Hiemer, 2013; Geist and Parsons, 2014]).

In this article we propose an alternative, objective implementation of the hypothesis that spatial b-value variations can provide information gain for probabilistic earthquake rate forecasting. The magnitude dimension of our time-independent mainshock model is based on the recently introduced method by Kamer and Hiemer (2015) that circumvents the arbitrary choice of external mapping parameters. The development of the spatial component of our forecasts builds up on the first-order results of the RELM experiment; specifically on the success of the HKJ model and the failure of the WS model (Figure 3 in [Zechar et al., 2013]).

The article is organized as follows: in the method section, we first review the fundamental differences in the construction of the HKJ and WS model. We then detail the lessons learned from R L for the development of our forecast. In the result section, we test our model's consistency in

terms of number, magnitude and space; particularly when using different concepts for the description of the spatial distribution of seismicity. We compare our forecasts against all first-generation RELM mainshock forecasts applying state of the art CSEP comparison tests. We finally discuss the strengths and weaknesses of our forecasts and indicate avenues for further improvements.

3.3 Method

The two most successful RELM mainshock forecasts (HKJ and WS, Figure 3-1) exhibit fundamental differences in their model constructions albeit being both based on earthquake catalog data as input alone. The analysis of those differences is important for the development of reliable next-generation seismicity based earthquake forecasts. The transparent policy of the RELM experiment allows for the thorough investigation of the advantages and drawbacks of individual model components and modeller's choices.

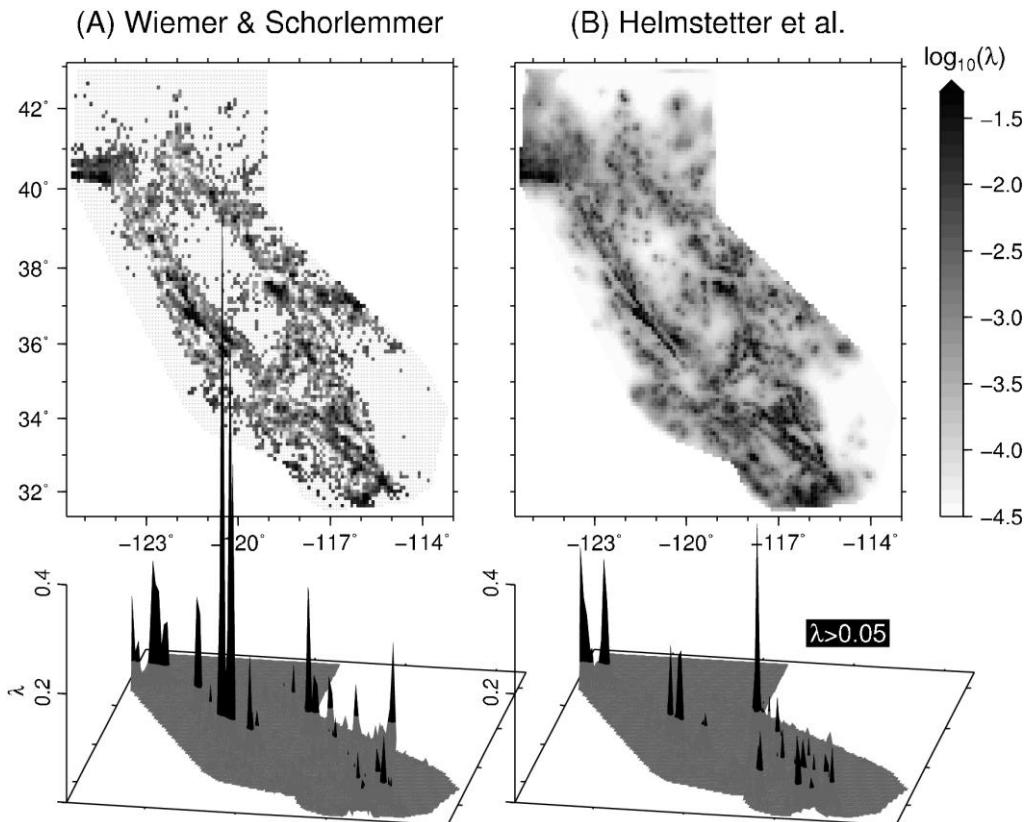


Figure 3-1 Spatial distribution within R. L.'s testing area of the expected $m \geq 5$ earthquake rate λ for the 5-years target period of 2006 - 2011 for the Wiemer and Schorlemmer (2007) and Helmstetter et al. (2007) forecast. Small dots label cells whose rates were defined by the “water-line” criteria.

3.3.1 The Helmstetter et al. (2007) approach

Helmstetter et al., (2007) used earthquakes of $m \geq 2$ in the Advanced National Seismic System (ANSS) catalog, in the time period from 1 January 1981 to 23 August 2005. The authors applied the [Reasenberg, 1985] declustering method to remove large fluctuations of seismicity rate in space and time due to aftershock sequences. Notice that they did not follow the Monte Carlo approach suggested by the RELM testing center [Schorlemmer et al., 2007]. Instead, Helmstetter et al., (2007) applied their own modification of Reasenberg's code using one specific declustering

parameter combination. The first step for the construction of the HKJ forecast is to estimate the density of seismicity in each cell by smoothing the locations of all past earthquakes with an isotropic power-law kernel of variable smoothing distance. The smoothing distance connected to each event is obtained by measuring the horizontal distance to its n^{th} nearest neighbor. In retrospective evaluations (i.e. using cross-validation) the performance of the spatial component of their forecast was found to be optimal in terms of likelihood gain for $n = 6$.

In a second step, the local densities needed to be corrected to account for spatial variations in the magnitude of completeness. In order to estimate the completeness magnitude in each cell, Helmstetter et al. (2007) extended their spatial smoothing approach by incorporating magnitude smoothing using a Gaussian kernel with a fixed bandwidth of 15 km. The magnitude of completeness m_0 is then estimated as the maximum of the local smoothed magnitude distribution. Such an approach bears similarities with the maximum curvature method applied to empirical magnitude distributions [Wiemer and Wyss, 2000]. In order to avoid non-physical small-scale fluctuations the authors smoothed the resulting spatial distribution of m_0 using a Gaussian filter with a standard deviation of 15 km. Under the assumption that the magnitudes follow the Gutenberg-Richter law with $b=1$, the density of events with $m \geq m_{\min}$ could then be corrected by the factor of $10^{m_0 - m_{\min}}$.

The magnitude distribution of the HKJ model is assumed to obey a tapered Gutenberg-Richter law with a uniform b-value and corner magnitude m_c . The corresponding cumulative density function reads

$$F(m) = 10^{-b(m-m_{\min})} \exp\left[10^{1.5(m_{\min}-m_c)} - 10^{1.5(m-m_c)}\right] \quad 3-1$$

with the minimum magnitude $m_{\min} = 4.95$ and the corner magnitude $m_c = 8.0$ (as suggested by Bird and Kagan (2004) for continental transform fault boundaries). The b-value was estimated as $b = 0.89$ on the basis of the catalog provided by the RELM testing center (using all $m \geq 4.95$ events weighted by their independence probabilities, [Schorlemmer et al., 2007]). Helmstetter et al. (2007) dedicated special attention to the region of Geysers, where they decided to use a local b-value estimate of 2.0 after having found that using $b = 0.89$ gives a very large expected rate of $m \geq 4.95$ events. They also examined magnitude distributions of other geothermal or volcanic areas within the RELM testing region (Mammoth Lakes, Coso, Salton Sea), but did not consider them as anomalous.

It is important to note that the HKJ model relies on estimating probability density functions for both the spatial and magnitude dimension. Finally these densities are scaled by the total expected number of target events (4.41 $m \geq 4.95$ events per year), which e lmstetter et al. (2007) estimated by counting target earthquakes in the RELM catalog that occurred during the time window 1984-2004. Accordingly, the final rates $\lambda(i_s, i_m)$ for each spatial cell i_s and magnitude bin i_m for the HKJ model are given by:

$$\lambda(i_s, i_m) = N \mu(i_s) P(i_m) \quad 3-2$$

where N is the expected number of target events for the next 5 years, $\mu(i_s)$ describes the spatial density within each longitude-latitude cell is, and $P(i_m)$ is the magnitude density within the magnitude bin i_m (estimated from equation 3-1).

The non-systematic approach for estimating the magnitude distribution in each cell can be regarded as the main deficiency of the HKJ approach, as underlined by Helmstetter et al. (2007) in their conclusion section. Furthermore, we noticed that their density correction scheme is questionable as it depends on the local degree of incompleteness (i.e. two spatial cells with identical m_0 and identical number of events $\geq m_0$ will not get the same spatial density if their numbers of events $\leq m_0$ are different).

3.3.2 The Wiemer and Schorlemmer (2007) approach

The WS model is based on the idea that small-scale spatial variations of the Gutenberg-Richter b-value are important for improving the skill of a probabilistic earthquake rate forecast. The basic method to construct their model was established by a number of case studies of cross-sectional b-value mapping along faults [Wiemer and Wyss, 2000].

Wiemer and Schorlemmer (2007) pointed out that due to the lack of objective criteria to estimate all involved mapping parameters; they used their intuition in the physical processes of earthquake occurrences as guidelines for model decisions. The WS model used all events in the ANSS catalog with depth ≤ 30 m in the time period from 1 January 1984 to 30 June 2005. Aftershocks were removed by applying the mainshock-window approach proposed by [Uhrhammer, 1986]. The WS forecast requires the estimation of three local parameters: the m_0 -value, the b -value, and the a -value.

In a first step, a map of the local magnitude of completeness $m_0(i_s)$ was generated by applying the entire magnitude range (EMR) method of [Woessner and Wiemer, 2005]. The EMR method performs a bootstrap estimate of m_0 that includes also events below m_0 . Note that it has been shown that the magnitude probability density function estimated with the EMR technique becomes discontinuous due to problems with its mathematical formulation (Figure 9 in [Mignan and Woessner, 2012], and references therein). Wiemer and Schorlemmer (2007) added an extra safety factor of 0.2 to the local m_0 estimates and smoothed the results using a Gaussian kernel with an unspecified width.

The regional b-value for the declustered catalog within the RELM testing area was reported to be $b = 0.823$ (using a magnitude cut-off of $m_0 = 3.5$ and A i's maximum likelihood formula corrected for magnitude binning, [Aki, 1965; Utsu, 1966; Bender, 1983]). For each cell is the local b-value $b(i_s)$ is estimated for all events with $m \geq m_0(i_s)$ that are within a cylindrical volume of radius R. The value of R was incremented in the range of $7 \leq R \leq 20$ m provided that the fixed regional b-value lead to a better penalized likelihood score than the use of the locally estimated b-value (as justified by the corrected Akaike Information Criterion; [Burnham and Anderson, 2002]). In cells with insufficient number of events (defined as $N_m \geq m_0 < 50$ for $R = 20$ km), the regional b-value estimate was used as a default value. Finally, the local a-value $a(i_s)$ was calculated using the formula $a(i_s) = \log_{10}[N_0(i_s)] + b(i_s) \cdot m_0(i_s)$, where N_0 is the total number of earthquakes with $m \geq m_0(i_s)$ located within the spatial cell is. Wiemer and Schorlemmer (2007) chose a default value of 0.01 m ≥ 0 events for all cells that did not contain observed events above $m_0(i_s)$.

The resulting forecast can be regarded as a modified version of the common relative-intensity based earthquake forecasts (e.g. Nanjo, 2010 and references therein). The cumulative

rates $\lambda_{cum}(i_s, i_m)$ for each spatial cell i_s and magnitude bin i_m for the WS model are given by:

$$\lambda_{cum}(i_s, i_m) = \frac{10^{a(i_s) - b(i_s)m(i_m)}}{T_c} T_t \quad 3-3$$

where $T_c = 21.5$ years and $T_t = 5$ years are the time spans of the learning and forecasting period, respectively. Notice that in the original code for computing the WS forecast (D. Schorlemmer, pers. comm.) all rates were estimated using $T_c = 22.5$ years, which does not correspond to the length of the input catalog. Moreover, we could not reproduce the reported regional b -value of 0.823 (using the same catalog results in $b = 0.92$ instead).

Wiemer and Schorlemmer (2007) pointed out that their method's results are not quantifiable in terms of a likelihood score, i.e. that the method does not allow for an optimization considering the entire parameter space. The WS model's sensitivity to its chosen external parameters (R , $N_m \geq m_0$, and the default a -value for empty cells) hinders the investigation for the particular reason of the forecast's skill in the RELM experiment.

3.3.3 A novel approach for probabilistic seismicity forecasting

In this subsection, we outline a novel method for the development of a forecast that incorporates spatial b -value variations. The basic concept of our forecast is based on the method of Kamer and Hiemer (2015), which uses penalized likelihood criteria to investigate the optimal b -value complexity supported by the data. Kamer and Hiemer (2015) proposed random Voronoi tessellations to explore optimal spatial partitioning using a minimum number of free parameters. For each random realization the overall likelihood is obtained by estimating the b -values in all Voronoi regions using Aki's formula. All individual models can then be ranked by their likelihood scores accounting also for the number of free parameters by means of the Bayesian Information Criterion [Schwarz, 1978]. Finally, an ensemble model is calculated by BIC-weighted averaging of the top-ranked models (for a detailed description and illustration of the method the reader is referred to Kamer and Hiemer, 2015).

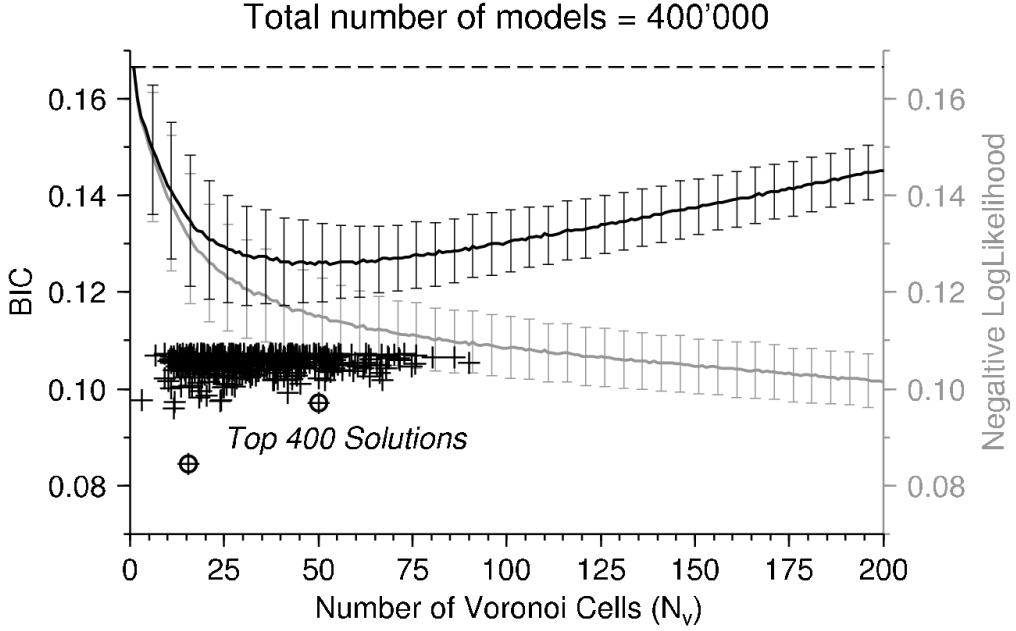


Figure 3-2 Entire solution space for the declustered 1984 - 2005.5 ANSS catalog (as used by Wiemer and Schorlemmer, 2007). Average BIC (black line) and negative log-likelihood (gray line) are shown as a function of number of Voronoi cells. Error bars correspond to one standard deviation. Crosses denote the BIC of the best 400 solutions and black circles highlight values for the two models shown in Figure 3-3.

For an unbiased comparison, we use the same catalog as the WS forecast to construct our model. We increment the number of Voronoi nodes N_v from 1 to 200, and spatially randomize each configuration 2000 times. The resulting solution space is composed of 400'000 models, each representing a unique spatial segmentation of R L's testing region. All models are ran ed by their penalized likelihood using BIC (Figure 3-2). We used the same fraction of top-ranked models as in Kamer and Hiemer (2015). Figure 3-3 illustrates two individual solutions with significantly different complexities. Each Voronoi region features a local estimate of the b-value (e.g. top panels in Figure 3-4), which can be used to correct the local spatial density estimates as a function of magnitude (as done in the HKJ model with a uniform b-value). We compare four different methods for describing the spatial component of our forecast (Figure 3-4):

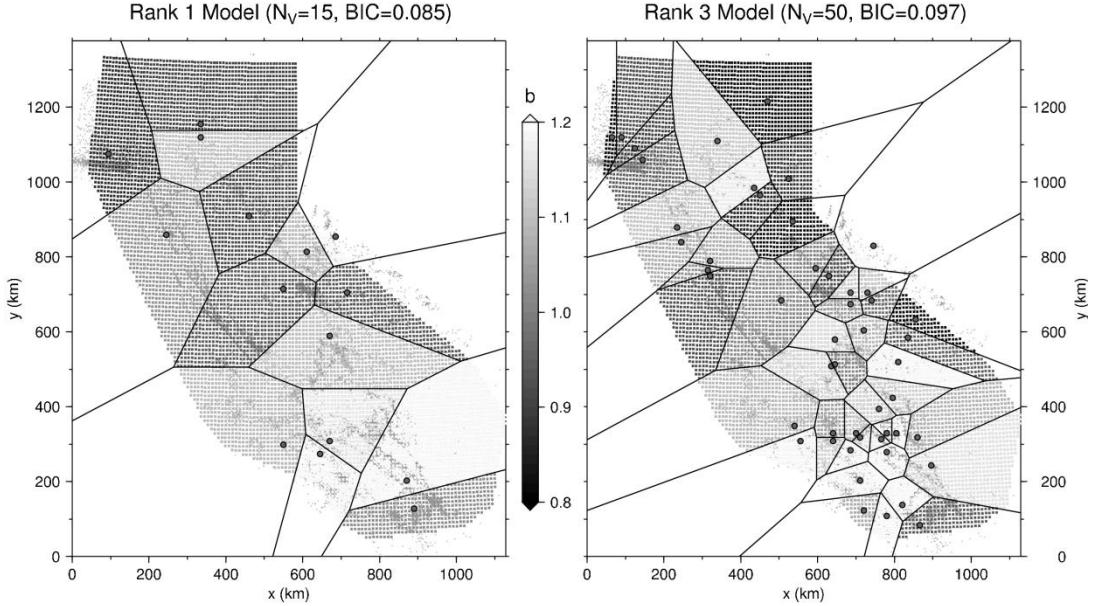


Figure 3-3 Map view of two individual tessellations. Black dots denote location of Voronoi seeds. The two models represent solutions of relatively low and high complexity, respectively. Corresponding BIC scores are given in the title and in Figure 3-2.

(A) Homogeneous: treating each Voronoi region like a classical area source zone in seismic hazard related source models), (B) relative intensity: simple counting of the number of events hosted by each cell above the cut-off magnitude (i.e. as applied in the WS forecast), (C) constant-width kernel: kernel-density estimation using a constant power-law kernel (set to 5 km as a benchmark; J. Zechar, pers. comm.), and (D) variable-width kernel: kernel-density estimation using a variable kernel with $n = 6$ (i.e. the same technique as used for estimating the spatial dimension of the HKJ forecast).

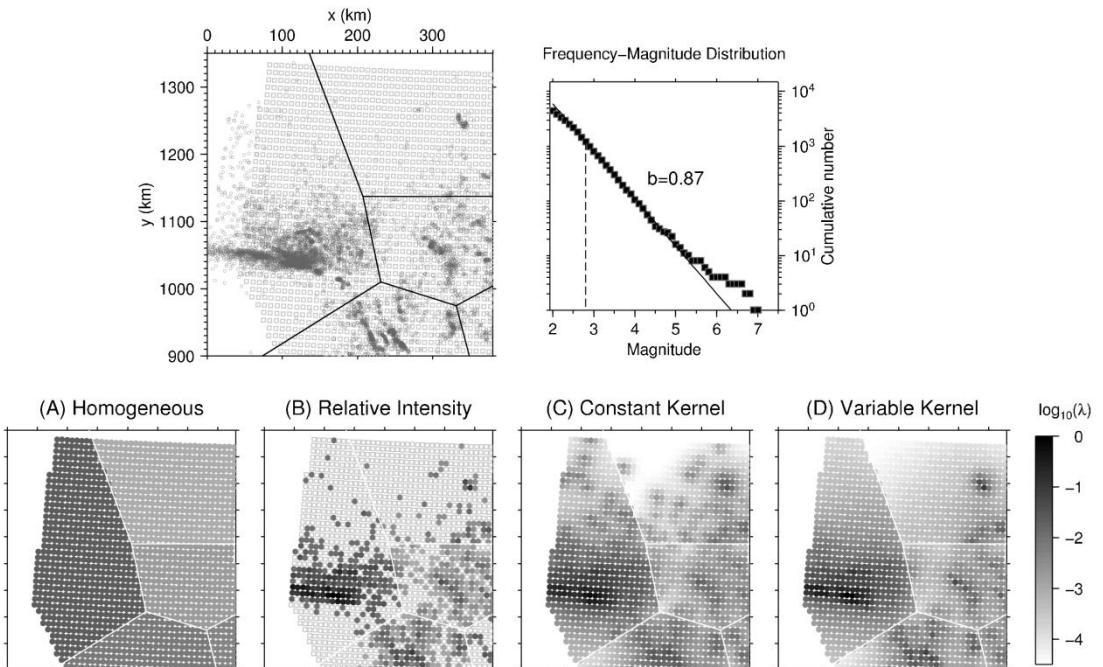


Figure 3-4 Top panels show a close-up of a single Voronoi region (see top left corner in Figure 3-3) and the corresponding frequency-magnitude distribution. Dashed line denotes cut-off magnitude. Bottom panels visualize results of different techniques for modelling the spatial distribution of seismicity. λ is the estimated number of expected

$m \geq 5$ events as derived from the R parameters. Note that empty gray squares indicate $\lambda = 0$. The total sum $\Sigma \lambda$ over all spatial cells in a single Voronoi region is identical for the different techniques.

In contrast to the HKJ model, the resulting spatial probability density $\mu(i_s, i_m)$ of our forecast differs for each magnitude bin (as a consequence of incorporating spatially varying b-values). The rates $\lambda(i_s, i_m)$ for each spatial cell is and magnitude bin im in our model are given by:

$$\lambda(i_s, i_m) = N\mu(i_s, i_m)P(i_m) \quad 3-4$$

where we followed the same formulation as Helmstetter et al. (2007) for estimating the overall magnitude probability distribution P and the total expected number of target events N. We construct four different ensemble forecasts based on the four different spatial component modelling techniques (Figure 3-4). Each forecast is ensembled by averaging all spatial probability density maps obtained from the top ranking individual tessellations.

3.4 Results

Figure 3-5 displays the cumulative spatial rates of our four ensemble forecasts using the same visual representation as for the WS and HKJ forecasts in Figure 3-1.

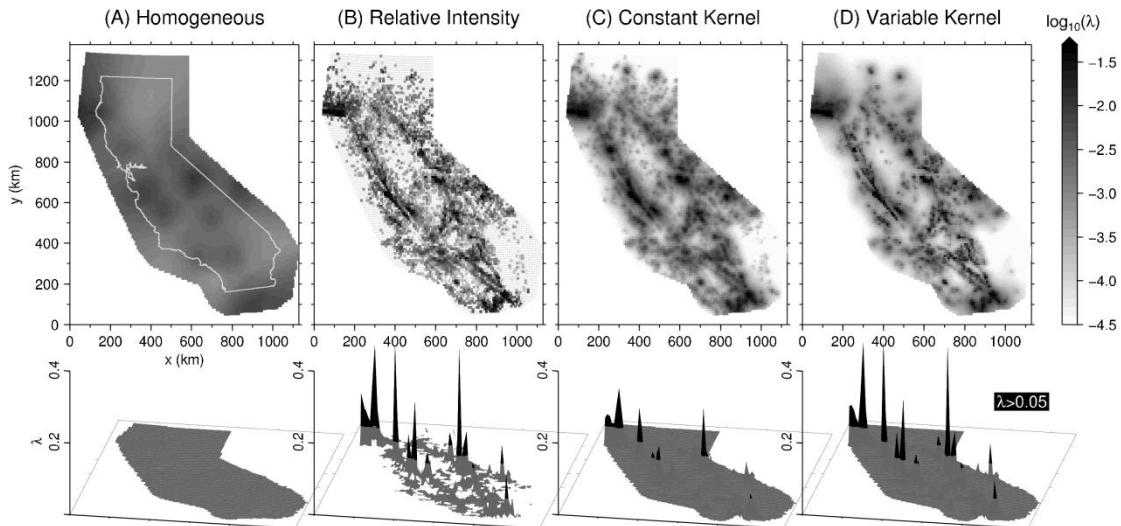


Figure 3-5 Same as Figure 1 for the ensemble forecasts using the top 400 individual models (crosses in Figure 3-2). Notice that the relative intensity based forecast involves cells with zero expectation (compare Figure 3-4B).

We subject our forecasts to a pseudo-prospective evaluation based on the metrics as outlined by the RELM experiment. The target earthquakes are those with $m \geq 4.95$ (in the ANSS catalog) occurring between 1 January 2006 and 31 December 2010. During these five years, 31 such target earthquakes occurred in the testing region and 20 of these were classified as target mainshocks (Table 1 in Zechar et al., 2013). We conducted R L's conditional likelihood and spatial consistency test to evaluate the statistical agreement of our forecasts with the observations. Note that for the number and magnitude consistency test we obtain the same score as the HKJ forecast, because the respective underlying distributions are the same. We found that all our forecasts that account for the spatial clustering of seismicity (Figure 3-5B,C,D) are consistent with the observation assuming a critical significance value of 5%. The homogeneous model (Figure 3-5A) failed both the spatial and overall conditional likelihood tests.

We focus on the evaluation of the skill of our forecasts relative to the first-generation models of Helmstetter et al. (2007) and Wiemer and Schorlemmer (2007). As pointed out by Eberhard et al. (2012), we replaced the inappropriate R-test used for pairwise model comparison by the T- and W-test (Rhoades et al., 2011). Both tests assess whether the average information gain per earthquake of one model (Λ^A) over another model (Λ^B) is significantly different from zero. The average information gain per earthquake $I_N(\Lambda^A, \Lambda^B)$ is defined as

$$I_N(\Lambda^A, \Lambda^B) = \frac{\sum_{t=1}^{N_T} [\ln \lambda_A(i_t) - \ln \lambda_B(i_t)] - (\sum \lambda_A - \sum \lambda_B)}{N_T} \quad 3-5$$

where it denotes longitude-latitude magnitude bins in which target earthquakes occurred and the term $(\sum \lambda_A - \sum \lambda_B)/N_T$ corrects the result for each forecast's overall rate. N_T is the total number of target events ($N_T = 20$ for RELM). The T-test assumes normally distributed rates and is based on Student's paired t-test, and the W-test is the corresponding non-parametric alternative that uses the Wilcoxon signed-rank test.

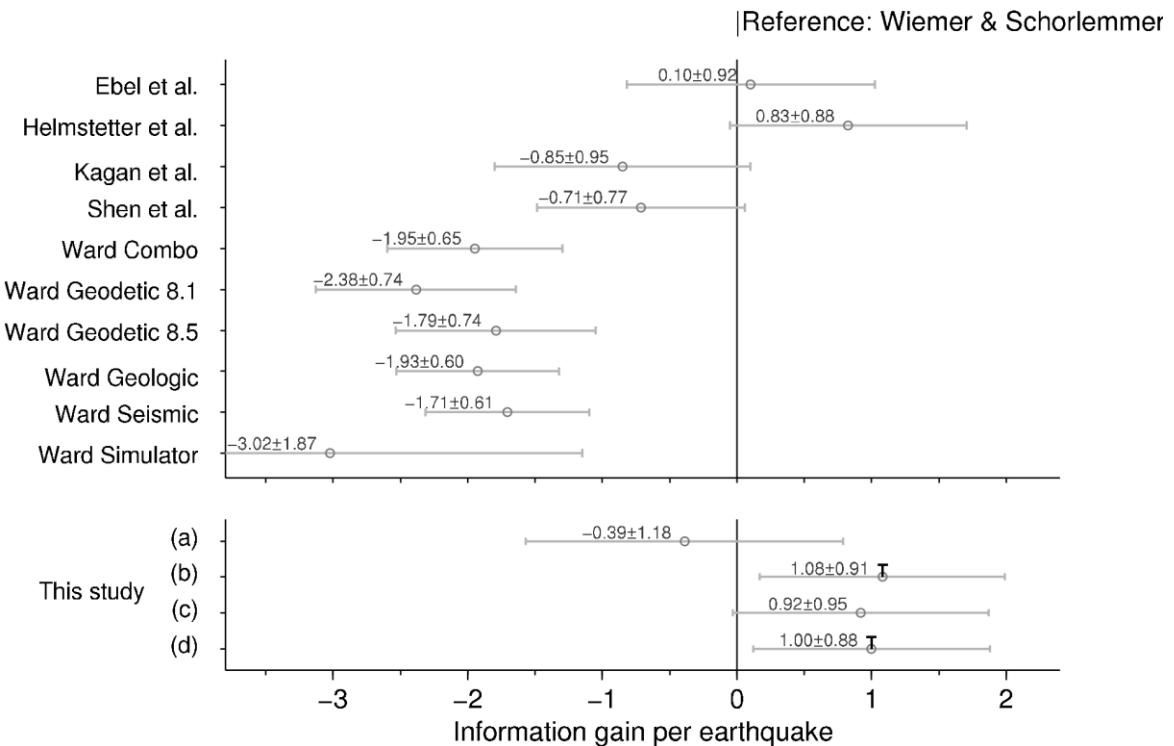


Figure 3-6 Rate-corrected average information gain per RELM target mainshock (20 $m \geq 4.95$ events in 2006 - 2010) with respect to the Wiemer & Schorlemmer model. Confidence intervals are derived from T-test statistics, and letter T highlights significance for positive information gains. Note that the number of target events varies for different RELM member models due to differences in spatial coverage (for corresponding numbers see Figure 5 in Zechar et al., 2013).

We find that the forecasts (B), (C), and (D) have a positive rate-corrected average information gain over the WS forecast as a reference, whereas the forecast (A) does not (Figure 3-6). For the forecasts (B) and (D) these results are significant given both the T- and W-test statistics. Notice that the confidence intervals given in our Figure 3-6 for the first-generation of RELM models do not agree with the ones published in the Electronic Supplement of Zechar et al. (2013). Zechar et al. (2013) reported values for I_N together with one sample standard deviation, but

did not scale the standard deviation appropriately (i.e. multiplied by the t-quantile and divided by $\sqrt{N_T}$, see Rhoades et al., 2011). However, their conclusions regarding the significance of the final ranking of RELM models are not affected by this technical error.

We repeated the comparison evaluation by using the HKJ model as a reference. The rate-corrected average information gains for our ensemble forecasts (A-D) are -1.22 ± 0.50 , $+0.25 \pm 0.24$, $+0.10 \pm 0.29$, and $+0.17 \pm 0.23$, respectively (see Figure 3-7). Although not significant in terms of T- and W-test, the positive gains obtained for the ensemble forecasts (B-D) seem promising. To further investigate the reasons for the skilfull performance of these ensemble models, we constructed individual forecasts out of each of the 400 individual partitions (including the two partitions shown in Figure 3-3). The T-test results indicate that 20% of the models in (B) and 7% of the models in (D) are significantly better than the HKJ forecast. However, it is important to notice that 8% of the models in (B) failed to enter the comparison test due to the occurrence of target events in cells with an expected rate of zero.

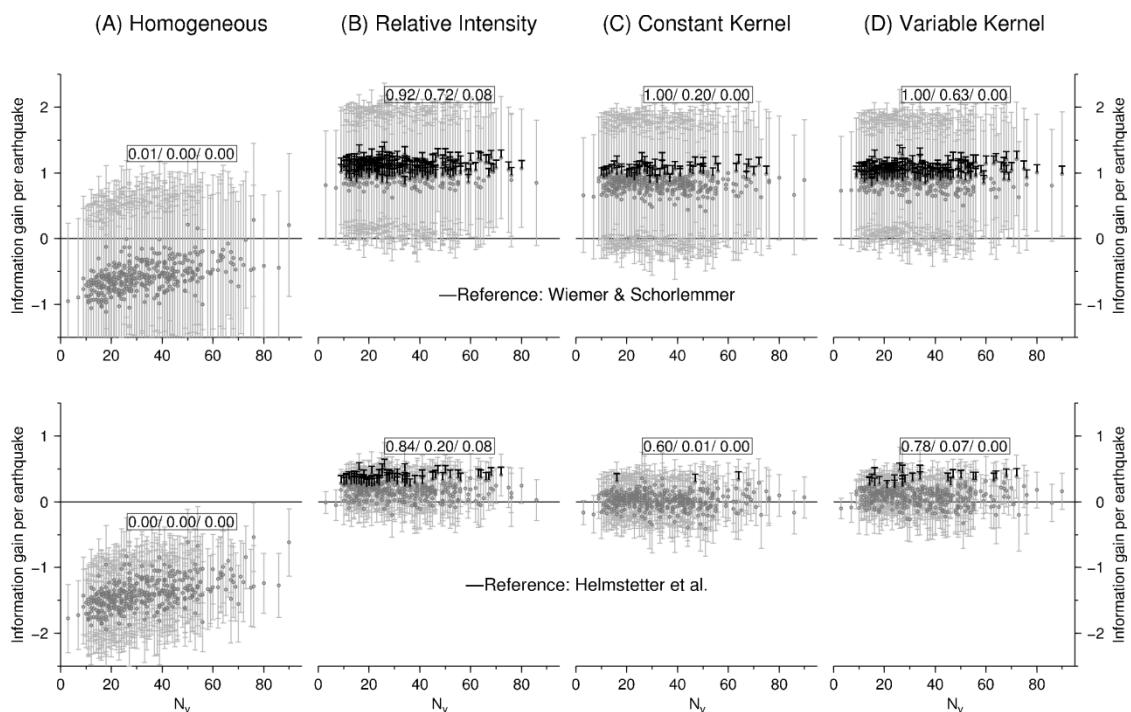


Figure 3-7 Rate-corrected average information gain per RELM target mainshock for the individual models that were used to construct the ensemble forecasts (Figure 3-2 and Figure 3-5). Top panels show results with respect to the Wiemer & Schorlemmer model (for ensemble T-scores see Figure 3-6) and bottom panels show results with respect to the Helmstetter model (the T-scores for the ensemble models are: A= -1.22 ± 0.50 , B= 0.25 ± 0.24 , C= 0.10 ± 0.29 , and D= 0.17 ± 0.23). Inset text boxes quantify the fractions of models that feature positive information gains/ significantly positive information gains/ zero-rate cells with target events. Notice that 8% of the forecasts in B failed the spatial consistency test due to target events occurring in such zero expectation cells.

3.5 Conclusions

In this article we outlined the construction of a time-independent earthquake rate forecast that uses seismicity catalog data as input. The development of the spatial component of our model was facilitated by the lessons learned from the prospective RELM experiment (regarding the performance of the HKJ and WS forecast). In the light of both the outcome of RELM and the

pseudo-prospective evaluations presented in this paper, we draw the following three main conclusions for the construction of a reliable and skilful probabilistic seismicity forecast.

(1) Kernel density estimation emerges as a suitable technique for transforming discrete earthquake locations into a non-parametric spatial probability density function (e.g. [Stock and Smith, 2002]). Both the success of the HKJ model and our own findings support the use of spatially varying kernels, however, we did not optimize the constant kernel size and we did not investigate different constant kernel shapes. The Relative- Intensity method, i.e. the simple counting of events inside spatial cells, suffers from its dependence on the initial definition of the spatial cells. Although the method performs well in terms of information gain, we find that, unlike the others models, 8% of all the individual forecasts failed the spatial consistency tests. Wiemer and Schorlemmer (2007) used a minimum “water line” level to fill up empty spatial cells, but the chosen value was too low to provide spatial consistency (as further reported by Werner et al., 2010). Note that the Relative-Intensity method can be regarded as a variant of the kernel density estimation technique, where each kernel represents a constant-width boxcar function (centered on the cell closest to the earthquake location; using the same size as the cell itself).

(2) We revised two different approaches for modelling the magnitude dimension of a seismicity forecast. The HKJ forecast used a tapered Gutenberg-Richter distribution [Kagan, 2002], and estimated its parameters (the b-value and the corner magnitude) based on regional large data sets (Bird and Kagan, 2004). By contrast, the WS model relied on estimating the GR parameters on a local scale and the overall magnitude distribution was obtained by the summation of all local ones. However, Woo (1996) showed that the partition of a large zone into subzones of non-equal b-value is algebraically inconsistent. This is supported by Figure 4 presented in Wiemer and Schorlemmer (2007): The sum of all local magnitude distributions is dominated by the contributions from small b-values, which in turn leads to deviations from the overall observed magnitude distribution. To avoid this “b-value pitfall” (Woo, 1996) we used the regional approach by Helmstetter et al. (2007) to constrain the magnitude dimension of the forecasts presented in this paper. Note that in the RELM experiment both the HKJ and WS forecasts passed the magnitude consistency test (Zechar et al., 2013).

(3) We find that the incorporation of spatial b-value variations has the potential to significantly increase the skill of a forecast (comparing our forecasts to the HKJ model as a reference). Helmstetter et al. (2007) manually investigated several selected regions in California (in a non-systematic way). They modified the magnitude distribution in the Geysers geothermal area, which has an unusually large b-value. Wiemer and Schorlemmer (2007) estimated the b-value locally in each cell using a specific set of external mapping parameters (mapping radius R , minimum number of events) and an information criterion that favors the search for local b-values. We note that their resulting b-value value map features extreme variations at small scales (in the range of $0.35 \leq b \leq 2.79$, not visible in their Figure 3A due to saturation of the color bar, Wiemer and Schorlemmer, 2007). The WS forecast does not have a positive information gain over the uniform b-value HKJ model (Figure 3-6), which might indicate that spatial b-value variations do not contribute to increasing the skill of a Californian seismicity forecast. However, we argue that such a conclusion might be unwarranted due to the poor implementation and lack of objective criteria for setting the external mapping parameters in the WS model. We proposed a seismicity forecast that remedies such deficiencies by using the method of Kamer and Hiemer (2015). The presented model features a positive information gain over the HKJ model. We conclude that the

overall success of the HKJ forecast (as a representative for simple smoothed seismicity models) can be attributed to their rigorous methodology regarding the spatial and overall magnitude dimension. However, we showed for the case of the RELM experiment that these forecasts are lacking skill with respect to models that feature large-scale b-value variations.

3.6 Discussion

Hindsight is much easier than foresight and only truly prospective testing will reveal if our conclusions based on pseudo-prospective evaluations will retain their validity. Nevertheless, we point out that our forecast does not require the subjective choice of external parameters, which facilitated both the model construction and the investigation of the forecast's performance regarding different methods for modelling the spatial distribution of seismicity. Our results confirm one of the main outcomes of RELM that the kernel density smoothing technique should be preferred over the Relative Intensity method for estimating the spatial forecast dimension. Our simple investigation supports the use of variable kernels in favour of constant kernels (Figure 3-6). However, we did not address the issue of kernel size and shape optimization in this study. For instance, Kagan and Jackson (2012) outlined drawbacks of variable kernels and promoted the use of adaptive kernels instead (as defined by Silverman, (1986)).

None of the first-generation RELM models features a positive information gain with respect to the HKJ model (Zecher et al., 2013, their Figure 5), which makes Helmstetter et al. (2007) the official winner of the RELM experiment. We showed in this paper that the HKJ model lacks skill due to the disregard of large-scale b-value variations. Notice that our findings regarding the rate-corrected average information gains are stable with respect to the number of individual models used for constructing the ensemble forecasts (Figure 3-5).

We assert that the method introduced by Kamer and Hiemer (2015) is a step in the right direction to investigate spatial b-value variations and to incorporate these in probabilistic earthquake rate forecasts. However, we find that the repeated random placement of Voronoi seeds is computational expensive considering the high number of generated models needed to explore the whole solution space. For instance we observe that despite the extensive random search conducted in our study, none of the individual optimal partitions succeeded to delineate the Geysers cluster (e.g. Figure 3-3) such that the local b-value is consistent with the one reported by Helmstetter et al. (2007). Therefore our model features a very large expected rate of $m \geq 5$ earthquakes in that region, close to the largest value in all California (Figure 3-5; compare with HKJ in Figure 3-1). Despite our model'siful performance, we underline that this setbac lead to losses in information gain with respect to the HKJ forecast.

To address the apparent weaknesses of the current random tessellation method we intent to investigate more efficient searching schemes in a related future study. Note that by contrast to classical b-value mapping techniques (Wiemer and Wyss, 2002), the framework provided by Kamer and Hiemer (2015) allows for the objective quantification of the performance of alternative tessellation methods in terms of penalized likelihood (e.g. Figure 3-2). We consider such an investigation as a much-needed benchmark before releasing any model for truly prospective testing within CSEP. We underline that only prospective testing in a well-defined controlled environment will help to advance earthquake forecasting related research.

The promotion of healthy competition among models and modellers can be regarded as one of the major aims of CSEP. Unlike many Hollywood movies, we argue that there is a strong case

for a RELM sequel in California to account for all the lessons learned from the first round. The seismicity forecast presented in this paper and many recent iterations of the HKJ approach (e.g. Werner et al., (2011); Helmstetter and Werner, (2012, 2014)) underline the benefits of RELM regarding the improvement of probabilistic seismicity forecasts.

3.7 Acknowledgements

We are thankful to J. Zechar and S. Wiemer for comments on an earlier version of this manuscript. All figures were produced using the Generic Mapping Tools version 4.2.0 (<http://www.soest.hawaii.edu/gmt>).

Chapter 4

The Barycentric Fixed Mass Method for Multifractal Analysis

by

Yavor Kamer, ETH Zürich

Guy Ouillon, Lithophyse

Didier Sornette, ETH Zürich

published in
Physical Review E, August 2013

4.1 Abstract

We present a novel method to estimate the multifractal spectrum of point distributions. The method incorporates two motivated criteria (*barycentric pivot point selection* and *non-overlapping coverage*) in order to reduce edge effects, improve precision and reduce computation time. Implementation of the method on synthetic benchmarks demonstrates the superior performance of the proposed method compared with existing alternatives routinely used in the literature. Finally, we use the method to estimate the multifractal properties of the widely studied growth process of *Diffusion Limited Aggregation* and compare our results with recent and earlier studies. Our tests support the conclusion of a genuine but weak multifractality of the central core of DLA clusters, with D_q decreasing from 1.75 ± 0.01 for $q=-10$ to 1.65 ± 0.01 for $q=+10$.

4.2 Introduction

Since their popularization by Mandlebrot [Mandelbrot, 1977], fractals and fractal geometry have been empirically observed and extensively studied in a wealth of natural and experimental physical phenomena. A common way to quantify the fractal or multifractal properties of a given set of data points is to calculate its generalized (Renyi) dimensions [Renyi, 1970], given as:

$$D_q = \lim_{\varepsilon \rightarrow \infty} \frac{\frac{1}{1-q} \log(\sum_i p_i^q)}{\log(\frac{1}{\varepsilon})} \quad 4-1$$

where ε is the scale of observation, $p_i(\varepsilon)$ is the fraction of data points (e.g, estimated measure) within box i of size ε , q is a real-valued moment order and the sum is performed over all boxes covering the data set under investigation. The most popular generalized dimensions are: D_0 the box counting dimension, D_1 the information dimension, and D_2 the correlation dimension. Varying the q parameter, D_q characterizes the scaling of the underlying measure within the distribution. Thus, $D_{-\infty}$ and D_{∞} respectively correspond to the local scaling of the lowest and highest densities, i.e. to the weakest and strongest singularities. For monofractal sets, D_q is a constant independent of q . For multifractal distributions, D_q decreases monotonically with q , and the resulting functional dependence of D_q as a function of q fully characterizes the underlying scaling properties. However, in practical implementations, strong departures from the theoretical values may occur due to edge effects related to the shape of the sampled zone, or to the finite number of data points (see for instance [Ouillon and Sornette, 1996]).

In many studies, researchers have tried to account for the edge and finite size effects using preprocessing, filtering and exclusion of data [Weitz and Oliveria, 1984; Tence et al., 1986; Torcini et al., 1991], which often suffer from some arbitrariness. The results' sensitivity to these subjective choices is often ignored or deemed incomputable, since these choices often alter not only the applied method but also the data used. Although methods for assessing and correcting such bias have been formerly introduced (see [Grassberger and Procaccia, 1983; Ouillon et al., 1995a; Ouillon and Sornette, 1996] for instance), some recent fractal analysis studies continue to use methodologies which exhibit errors up to 0.15 in D_0 for uniform 2D distributions [Márquez-Rámirez et al., 2012]. Such large error margins and the need for additional corrections to obtain unbiased measures have hindered the interpretation and comparison of the results between different

multipfractal analyses. In this study, we address the issue of edge effects by introducing a novel method that accounts for such errors/biases intrinsically during the analysis. This is done with the help of two data-driven, non-arbitrary criteria: barycentric pivot selection and non-overlapping coverage.

We test the performance of the method on synthetically generated monofractal and multipfractal distributions and compare the obtained empirical results with the analytically predicted ones. Encouraged by the results, we then proceed with the analysis of large clusters resulting from the growth process of diffusion limited aggregation. We provide new results that further inform the debate about the possible multipfractal nature of such a generic growth process.

4.3 The Barycentric Fixed Mass Method

4.3.1 Review of multipfractal analysis methods

In order to put our proposed method in perspective, we shall first give a brief overview of the commonly used multipfractal analysis methods. Generally, they are classified as either fixed-size or fixed-mass methods. Fixed-size methods (FSMs) [Jensen *et al.*, 1985; Tél *et al.*, 1989] estimate D_q via the scaling of the total mass M within a constant r -sized ball, as r is increased:

$$\log \langle M(<r)^{q-1} \rangle \approx (q-1)D_q \log(r) \quad 4-2$$

The box counting method, which consists in covering the distribution with boxes and increasing their sizes, is a classical example of FSMs. Due to its significant bias for small samples, the box counting method is regarded as impractical [Greenside *et al.*, 1982]. Inspired from the correlation dimension algorithm [Grassberger and Procaccia, 1983], the sand box method [Tél *et al.*, 1989] performs better than box-counting by centering circles at arbitrary points on the fractal and averaging the mass accumulation as the radii is increased. However, this method is not reliable for D_q values when $q < 1$, which quantify the scaling properties of the weakest singularities, i.e. the low density parts of the multipfractal.

On the other hand, fixed-mass methods (FMMs) estimate D_q via the scaling of the smallest radius r to include a fixed mass m , as m is increase. Several studies report FMMs to be superior to FSMs, especially for negative q values [Badii and Broggi, 1988; Grassberger *et al.*, 1988; Hirabayashi *et al.*, 1992]:

$$\begin{aligned} \log \langle R(<m)^{-(q-1)D_q} \rangle &\approx -(q-1) \log(m) \\ \tau(q) = (q-1)D_q \\ \log \langle R(<m)^{-\tau(q)} \rangle^{\eta \tau(q)} &\approx \log(m)^{-\eta D_q} \end{aligned} \quad 4-3$$

For a detailed review of both FSMs and FMMs, the reader is referred to [Theiler, 1990]. Other methods such as wavelet analysis have also been introduced [Arneodo *et al.*, 1988; Ouillon *et al.*, 1995b]; however they are also prone to biases due to finite size effects, and their efficient implementation generally necessitates to discretize the underlying distribution.

The barycentric fixed-mass method (BFM) introduced in the present study uses Equation (4-3) to estimate D_q . The method uses two criteria in order to reduce the finite size and boundary effects, which we now describe.

4.3.2 Barycentric Pivot Selection

In both FSM and FMM, the data points serving as centers for the fixed radius or mass circles are chosen randomly within the sample. The D_q 's measured using a small selection of such random centers is considered to be good approximation if those centers (hereafter pivot points) are chosen according to a uniform distribution on the fractal [Tél *et al.*, 1989]. This assumption reduces the computation load and allows a quick analysis of large datasets. However the measured multifractal spectrum will depend on the location of the randomly selected pivot points, as the finite size and irregular boundaries effects will vary: pivots in the inner core of the fractal will accumulate more mass compared to pivots on the outer edges. Repeating the analysis with a different set of pivot points will result in variation of the estimated D_q , which controls the precision of the analysis. Using all data points as pivots would give a single D_q estimate, increasing the precision, but this would not account for the edge effects and would require more computational resources.

The barycentric pivot selection criterion tackles these two issues and is illustrated in Figure 4-1. We consider a given data point (plotted in yellow color in Figure 4-1) as a potential pivot. As the mass m has been previously defined, we consider its m closest neighbors and compute the barycenter of those m datapoints; we also compute r , which is the distance from the pivot to the farthest of those neighbors. If the barycenter stands closer to the pivot than to any neighbor, then the corresponding couple (m, r) contributes to the averaging term in Equation (4-3). This is the case in Figure 4-1 for the circles labeled as A and B , as well as for the corresponding barycenters of the enclosed data points labeled the same way. Those circles correspond respectively to masses $m=5$ and 10 and radii r_A and r_B .

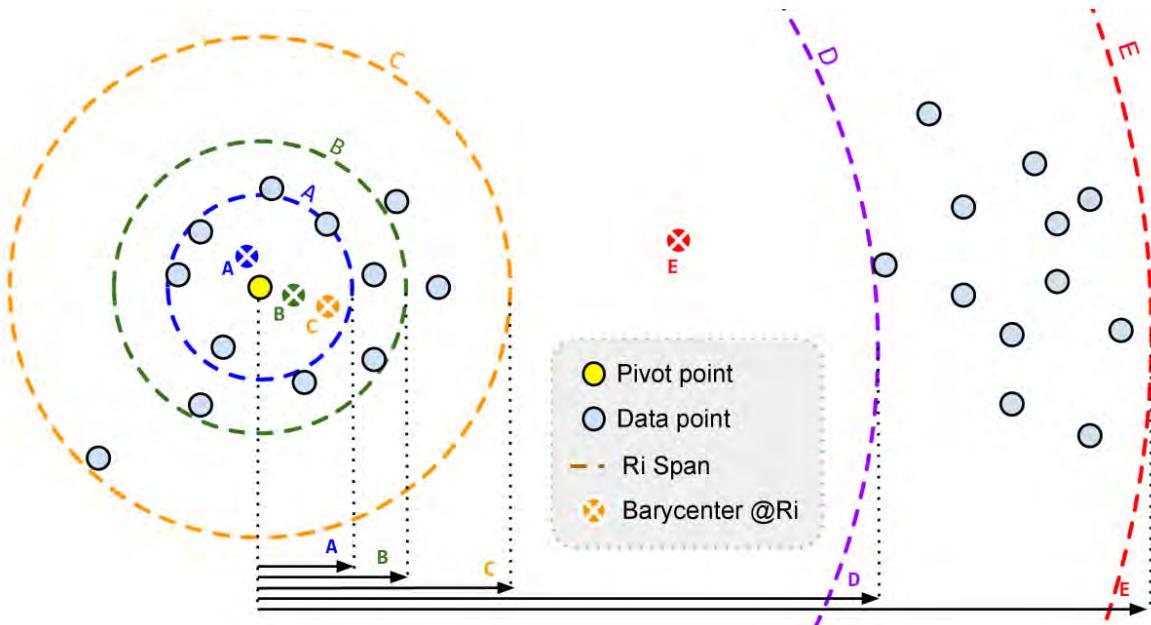


Figure 4-1 Illustration of the barycentric pivot point selection criterion

In the example shown in Figure 4-1, this criterion then ceases to be valid when one extends the mass m by 3 units or more, as the circle radius now becomes equal to r_C (corresponding to the circle labeled C on the figure). The corresponding barycenter is found to be offset to the right so that the previous pivot point ceases to be active, as it is no longer the closest point to the barycenter. Another example is shown with a circle and its center both labeled as E , corresponding to a mass $m=25$. In the usual methods, be it FSM or FMM, the circle around the yellow point is allowed to extend till it encloses all its neighbors. In contrast, using the barycentric criterion, the pivot point will be active only up to radius r_B in the example shown in Figure 4-1. Applying this criterion to each data point, we determine the set of radii for which it can be used as a pivot.

4.3.3 Non-Overlapping Coverage

Implementing the classical methods, all N pivot points spread over the whole self-similar set, so that each data point contributes N times to the averaging term in Equation (4-3). The barycentric pivot selection criterion we introduced above results in pivot points being preferably chosen within dense areas where the mass concentration is higher than in their neighborhood. Points located within these areas will be more likely to satisfy the barycentric condition over large radii, resulting in high density areas having a higher contribution to the averaging term. To account for the bias that could result from this selection, we introduce an additional non-overlapping coverage (NOC) criterion for each pivot point. For each fixed mass value, we require that the ensemble of selected pivot points and their respective fixed mass spheres define a non-overlapping configuration. In the absence of the implementation of the NOC condition, the data points located in dense regions would be multiply counted by many spheres, leading to an oversampling of the strongest singularities. By introducing the NOC criterion, we are effectively equalizing the probability of low-density areas to be correctly sampled. The NOC criterion is akin to the construction of the packing dimension, which is obtained by “packing” equal sized spheres inside a given subset[Tricot, 1982]. However, because our method is a fixed-mass method, the spheres will have different sizes, contrary to the usual definition of the packing dimension. Allowing for only a limited amount of overlap in the location of the spheres will lead to a tight covering of the dataset, a configuration similar to the definition of the Hausdorff dimension[Hausdorff, 1918]. The minimum amount of overlap can be approximated by considering a set of three circles covering an area as given in Figure 4-2. Once the radius of the circle is fixed, we compute the minimum distance between the centers of the circles as the one for which all three circles intersect at the same location (ensuring that the whole space within the dashed triangle is completely covered by the three circles). By simple geometrical reasoning applied to the case of three circles of equal radii, we estimate the radius overlap as $2R/2r = \sqrt{3}/2 = 34\%$ where r is the radius of the circles and $2R$ the distance between two distinct circle centers.

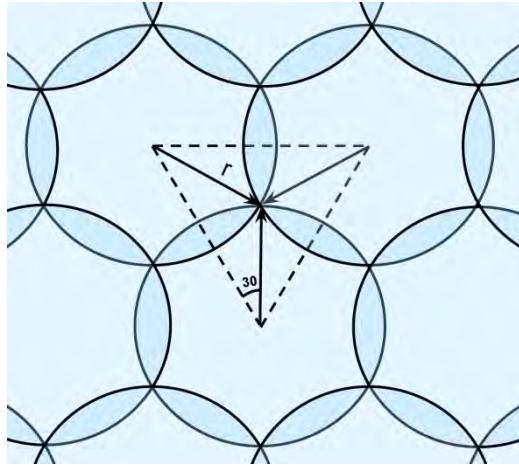


Figure 4-2 Minimum overlap of circles covering an area

The NOC criterion is implemented by placing a first random candidate pivot point with a circle with radius R_0 (satisfying the barycentric pivot selection) and then discarding all pivot points (with radii R_i) within a distance of $0.86(R_0+R_i)$. By downscaling the radii to 86.6%, we tend to induce an overlap of 13.4% necessary for full coverage. The next pivot point is again chosen randomly from within the remaining set of possible pivot points. The random selection and consequent discarding is then repeated over the remaining set of points until all candidate pivot points are placed (or discarded). For a synthetic dataset we use the multifractal Sierpinski measure which is obtained by recursive replication of the density matrix [1 0; 1 2]. Figure 4-3 illustrates the first recursion of the generation procedure. At each recursion, the output grid replicates itself multiplicatively over each element of the density matrix. This results in tripling (due to the 3 non zero elements) of the area and quadrupling ($1+0+1+2=4$) of the mass with each recursion. The reader is referred to [Lynch, 2004] for details. The distributions in Figure 4-4 are obtained by 6 recursive replications, resulting in a total mass of $(1+0+1+2)^6=4096$ points and a maximum mass concentration of $2^6=64$ points. The same figure displays two coverages of a multifractal Sierpinski triangle with fixed-mass circles with masses of respectively 84 and 136. The candidate pivot points satisfying the barycentric pivot selection are plotted as gray dots and the selected circle centers are plotted as black dots.

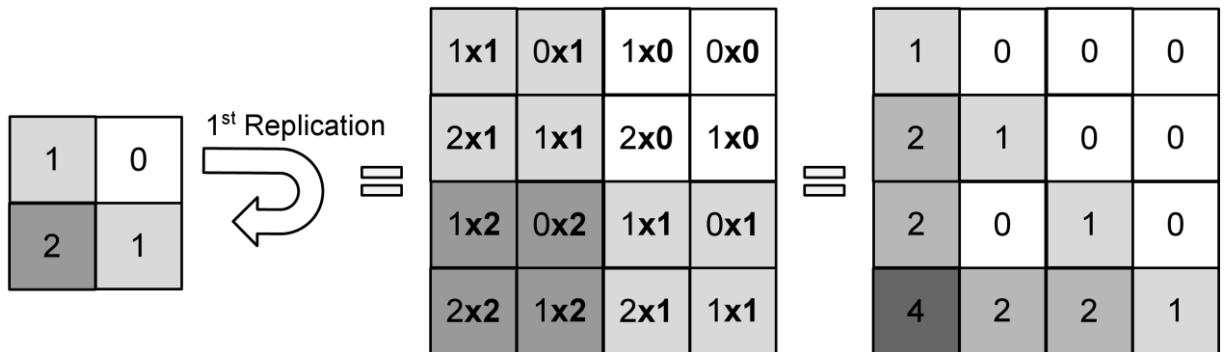


Figure 4-3 Generation of the synthetic multifractal measure shown in Figure 4-4 from the density matrix given by the 2x2 table on the left

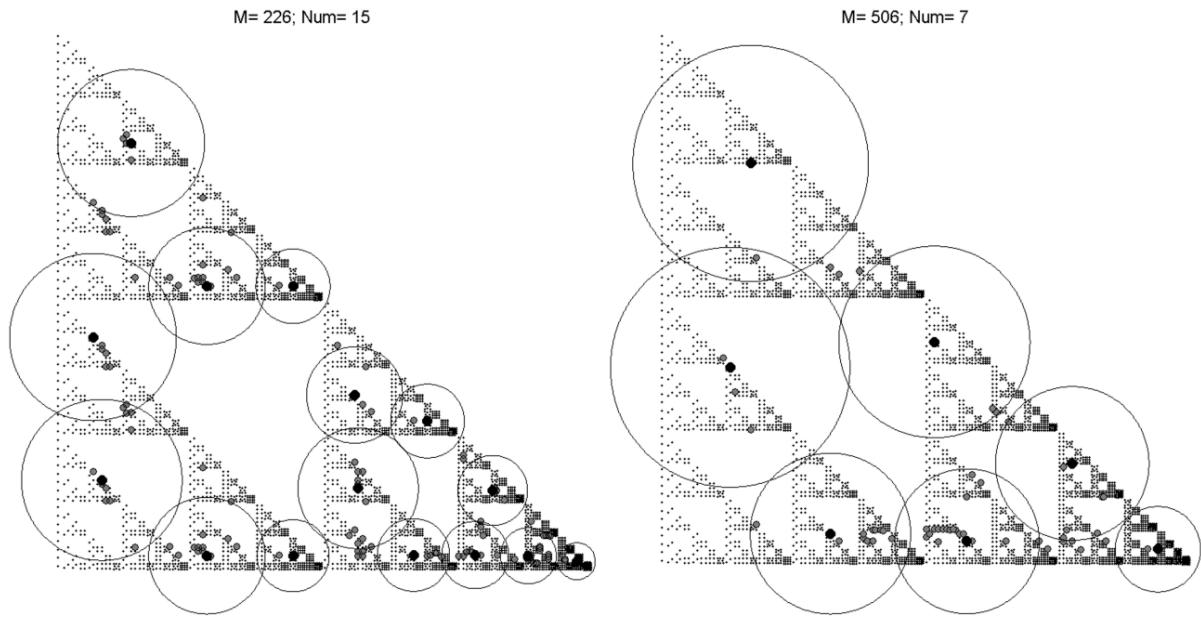


Figure 4-4 Coverage of a synthetic multifractal Sierpinski triangle with fixed-mass circles

4.4 Multifractal spectra of synthetic datasets

4.4.1 First example with the synthetic multifractal Sierpinski triangle of Figure 4-4

To estimate D_q , we increment τ and calculate the expression $\log \langle R(m)^{\tau} \rangle^{\frac{1}{\tau}}$ for sets of fixed-mass circles covering the point distribution. The mass range m is sampled at logarithmically spaced steps, giving $m_i = m \times 10^{i\alpha}$, with $\alpha=0.05$, where the role of the smallest value m is discussed in section 3.2. The curves of averaged radii versus fixed-mass are given on a log-log plot in Figure 4-5a. Calculating the slope for each τ exponent (represented in shades of gray), we estimate both D_q and q . The analytical and estimated curves of D_q as a function of q are shown in Figure 4-5b; the solid curve represents the exact analytical expression of D_q for the synthetic multifractal Sierpinski triangle of Figure 4-4, the dashed curve is the estimation and the gray band corresponds to $\pm 1\sigma$ obtained over 100 trial measurements, resulting in different configurations of circles locations. We thus check that, for this dataset, our method obtains excellent results, even for negative q values and such a small dataset.

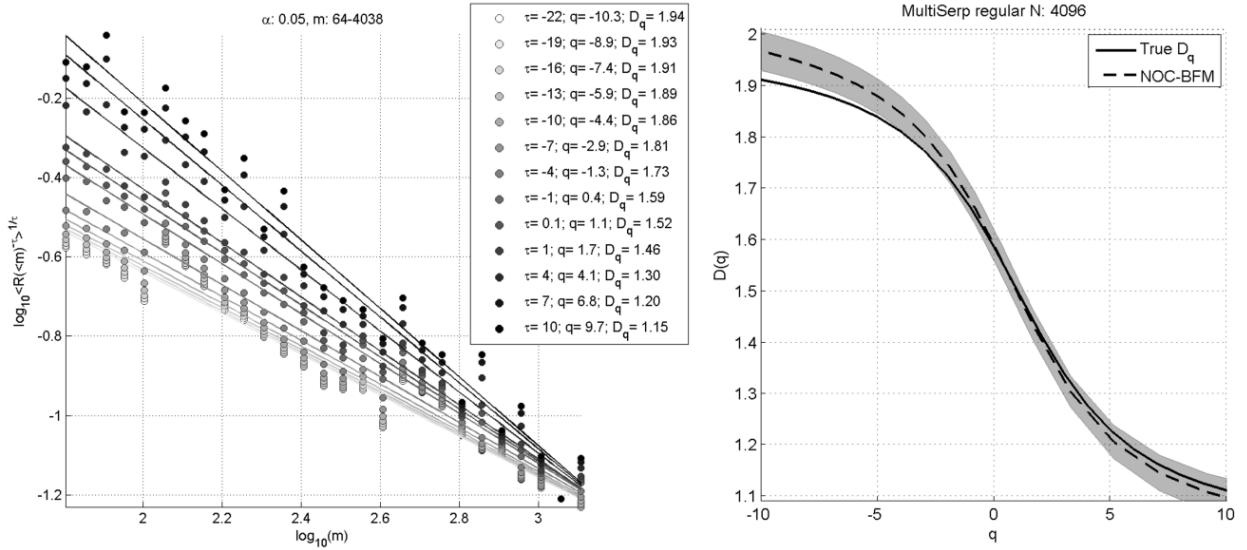


Figure 4-5 a) Averaged radii versus fixed-mass for increasing τ ; b) Analytical and estimated D_q - q curves for the multifractal Sierpinski triangle obtained recursively with the density generator [1 0; 1 2] as explained in the text and with Figure 4-3.

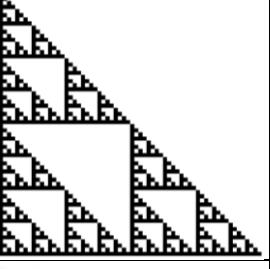
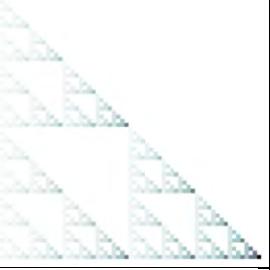
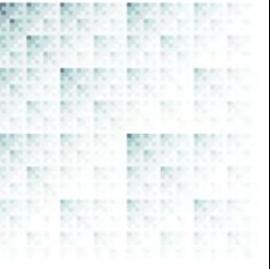
4.4.2 Multifractal spectra of synthetics datasets

This subsection extends the previous one by exploring the merits and limitations of our barycentric fixed mass method applied to different mono- and multifractal measures obtained synthetically. The synthetic datasets are constructed by recursive replication of a 2 by 2 density matrix (See Figure 4-3 for an example). The resulting density grid is normalized so that the lowest mass in a cell is 1 and each other grid cell is uniformly sampled by a number of points according to its computed mass. An optimal uniform distribution within grid cells featuring more than one point is ensured by assessing the locations according to an optimal circle packing within each square. The analytical values of D_q of the obtained distribution are given by Equation (4-4) in terms of the elements of the density matrix (p_i):

$$D_q = \frac{1}{1-q} \frac{\log(\sum_{i=1}^4 p_i^q)}{\log(2)} \quad 4-4$$

The formulation of D_q does not depend on the locations of the p_i elements, thus shuffling the elements of the density matrix would result in a different fractal with the same D_q . To test the robustness of the new method, we conduct two test cases for each density matrix: (i) a regular distribution where the density matrix is constant through all iterations, and (ii) a random distribution where the density matrix is shuffled in the beginning and rotated 90 degrees after each iteration. We choose rotation rather than shuffling in order to ensure that the replication is not done with the same matrix in consecutive generations (a possibility when shuffling). This is done to minimize the effect of discrete scale invariance, which is intrinsic in deterministic synthetic fractals [Sornette, 1998]. The 6 test-sets and their details are given in Table 4-1. The figures show the density of data points, which increases with the darkness of the grey level.

Table 4-1 Synthetic datasets used for the benchmark test

Name	Density Matrix	Replication No	Total Mass	Density Grid	
				Regular	Random
Monofractal Sierpinski Triangle	$\begin{bmatrix} \frac{1}{3} & 0 \\ \frac{1}{3} & \frac{1}{3} \end{bmatrix}$	8	6561 (3 ⁸)		
Multifractal Sierpinski Triangle	$\begin{bmatrix} \frac{1}{4} & 0 \\ \frac{1}{4} & \frac{2}{4} \end{bmatrix}$	6	4096 (4 ⁶)		
Multifractal Sierpinski Carpet	$\begin{bmatrix} \frac{2}{5} & \frac{1}{5} \\ \frac{1}{5} & \frac{1}{5} \end{bmatrix}$	5	3125 (5 ⁵)		

The proposed non-overlapping barycentric fixed-mass method (NO-BFM) was benchmarked against the usual fixed-sized (FS-SB) and fixed-mass sandbox (FM-SB) methods, both of these methods being prone to significant edge effects: the mass vs radius (or vice versa) growth for a point located at some edge of the fractal differs significantly from a point in the center of the fractal, as the corresponding circles include more and more empty space as their radius increases.

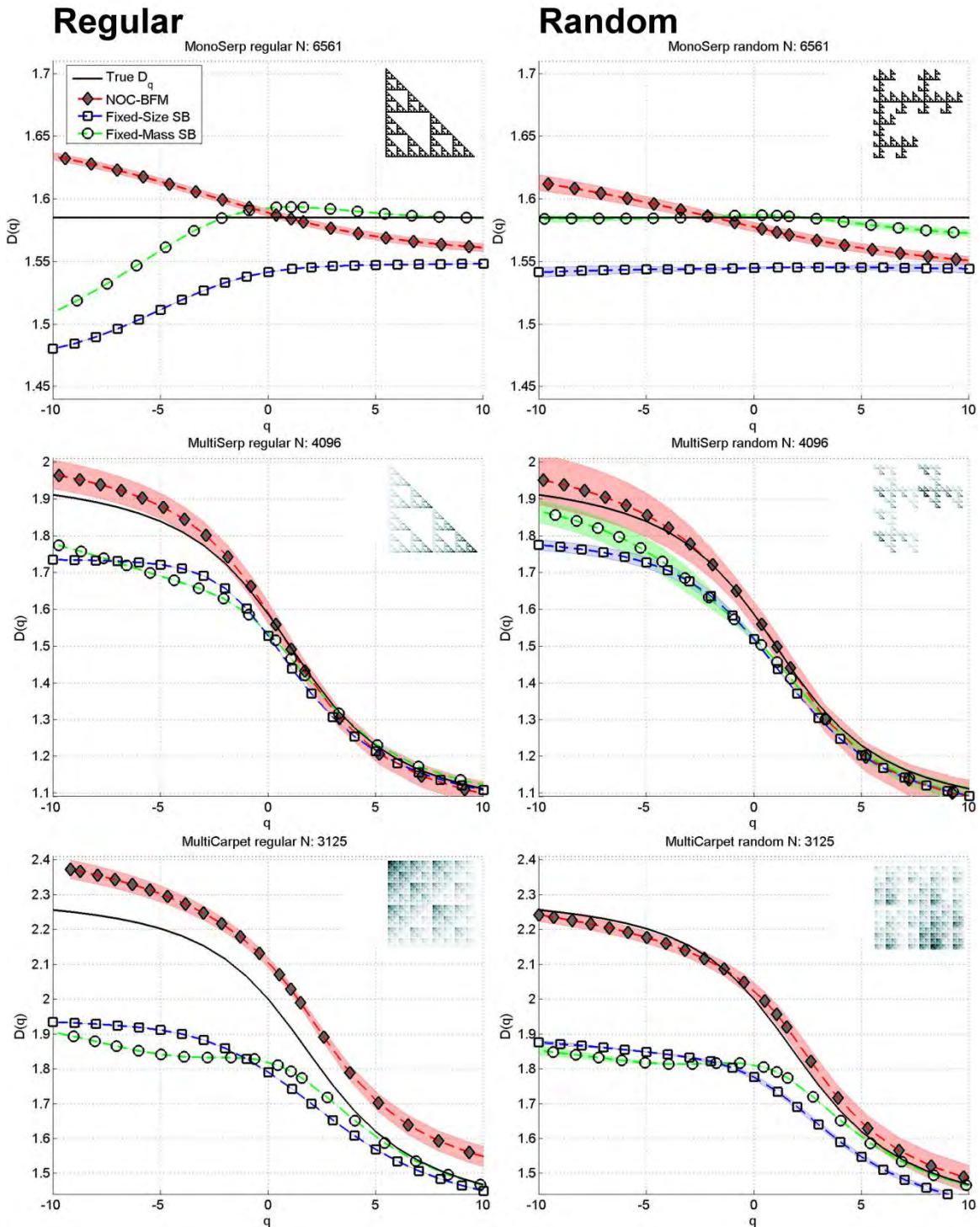


Figure 4-6 Benchmark results for regular and random synthetic distributions

In order to present an objective comparison, we tried to optimize the implementation and maximize the performance of all methods in competition. For the FS-SB method, we limited the spanning radius range starting from the maximum of the closest neighbour distance up to the minimum of the furtherest neighbour distance (corresponding to the smallest enclosing radius). This R interval was logarithmically sampled as $R_i = R_1 10^{i\alpha}$ with $\alpha=0.05$. For the FM-SB method, the mass range m was sampled logarithmically as $m_i = m_1 10^{i\alpha}$ with $\alpha=0.05$. The minimum mass was set by a condition imposed on the maximum number of points within an elementary cell, so as to avoid that the spatial distribution of points within that cell becomes uniform ($D=2$), thus breaking self-

similarity. The maximum mass was limited to one quarter of the total mass ($1/2^D$ with $D=2$). These values were also used for the NO-BFM method with the exception of the maximum mass. The latter is determined automatically as the maximum mass level at which only one non-overlapping circle can be placed on the fractal, since the averaged $\langle R \rangle$ term would require at least two circles. It is important to point out that FS methods are sensitive to the sampling of the R range: as a result, $\langle M \rangle$ vs R curves can become unstable at oversampled small scales where the radius increment fails to result in mass increment. In contrast, the FM methods are more robust since the radius increases to enclose the M th closest neighbor, thus ensuring $\langle R \rangle$ to increase with M . Another issue regarding the benchmarking of the methods is the use of the full (or partial) dataset for the generalized multifractal dimensions estimation. The general practice for both FM-SB and FS-SB methods is to select a random sample dataset (usually 10 percent of the whole) as pivot points, and perform the measurements at these points. The standard deviation of the measure is calculated by different randomizations of this subset. For comparability with the NO-BFM method, which considers the full dataset for determining the candidate pivot points and eliminating overlaps, we also use the full dataset as pivot points for both FM-SB and FS-SB methods. Thus, for the regular fractals, these methods do not allow us to compute a standard deviation, while NO-BFM provides a standard deviation related to the different possible packing configurations.

The D_q estimates of the six datasets for the three methods are displayed in Figure 4-6. For the random synthetics, estimates are plotted with confidence bounds of $\pm 1\sigma$ obtained over 100 different randomized distributions. One should be aware that all the possible randomization outcomes depend on the unique elements of the density matrices. The results clearly show that, for the multifractal sets, NO-BFM outperforms both fixed-mass and fixed-sized methods for both negative and positive values of q . We observe that, for the regular multifractals, the method tends to slightly overestimate D_q . Since we do not observe this in the randomized distributions, we conclude that this overestimation is due to the discrete scale invariance effect[Sornette, 1998], which becomes more pronounced as the high density mass is progressively being concentrated in one part of the fractal. We have also investigated individual coverage configurations of the regular multifractals, and observe sudden drops in the number of covering circles as their mass is increased, a signature of discrete scale invariance.

In order to highlight the implications of our synthetic test results, we draw the reader's attention to the randomly generated fractal sets on the right panel of Figure 4-6. We consider these sets to be more representative of typically observed distributions due to their stochastic structure. For all three cases, the estimation error of NO-BFM is bounded within the range of $\pm 5\%$. Our method performs slightly worse for the regular sets on the left side of Figure 4-6, however this is due to the pronounced discrete scaling pattern which is irrelevant for most natural systems. While the other methods perform fairly well in the monofractal case, their errors increase drastically when the data starts exhibiting multifractality. These results indicate that the sandbox-based methods might still be used for monofractal sets, but since such a conclusion implies a prior knowledge about the input distribution it should be avoided. Another practical consideration regarding the results obtained with the classic methods is that they represent the best-case scenario. As these methods are computationally demanding, many applications are conducted over a small subset of pivot points rather than the full dataset as presented here. The reason for the poor performance of the classical methods is their susceptibility to finite size and edge effects that remained overlooked

because previous benchmarks were carried out on 1D multifractals or 2D monofractals [Tél *et al.*, 1989; Vicsek *et al.*, 1990].

In terms of computation resources, although NO-BFM includes both pivot selection (BFM) and no-overlap (NOC) criteria, it is still superior to both methods. The computation of the averaging term is significantly accelerated since NOC decreases the number of averaged points as the fixed-mass increases. On the other hand, BFM limits the number of points considered in NOC, minimizing the time needed for its computation.

4.5 Application to the Diffusion-Limited Aggregation (DLA) process

Having evaluated the accuracy and precision of the proposed NO-BFM method, we now revisit the diffusion-limited aggregation (DLA) growth process, which has been the subject of many fractal analyses. Diffusion-limited aggregation occurs when particles following a random walk stick to a stationary seeding point, becoming themselves locations for the attachment of other later incoming particles, leading to the growth of a complex aggregate. The growth of such an aggregate is governed by branch (finger) formations and the consequent screening effects of channels between the fingers. Due to its simple mechanism and widespread occurrence in natural phenomena, the DLA process has been studied extensively. However analyses of its fractal properties resulted in different conclusions: some studies[Meakin, 1983; Argoul *et al.*, 1988, 1990; Li *et al.*, 1989] suggest that the DLA cluster is a monofractal set with a constant D_q , independent of q while others[Meakin and Havlin, 1987; Nittmann *et al.*, 1987; Nagatani, 1988; Vicsek, 1990; Boularot and Albinet, 1996] propose that it is a multifractal. The differing findings are likely to be influenced by the finite size and boundary effects, which affect all D_q estimation methods to different degrees. In a recently published study[Hanen and Heffernan, 2012], the authors tried to address this and several other issues by limiting the pivot point selection to points within a distance $0.5R_g < d < 1.5R_g$ from the cluster seed point, where R_g is the radius of gyration defined as:

$$R_g^2 = \frac{1}{N} \sum_{i=1}^N (r_i - r_{mean})^2 \quad 4-5$$

where r_i is the distance of the i -th particle to the barycenter of the DLA cluster.

The R interval for the slope estimation was chosen as $0.032R_g < R < 0.32R_g$. The study[Hanen and Heffernan, 2012] suggests that the multifractality of the DLA cluster is less pronounced than previously thought by some authors, proposing a constant D_q value of 1.69. It is difficult to assess the consistency of these results, as they are likely to be affected by the somewhat arbitrary criterion for pivot point selection and R interval. The authors also indicate that their estimate of D_q is likely to be underestimated due to the boundary effects inherent in the fixed-size sandbox method.

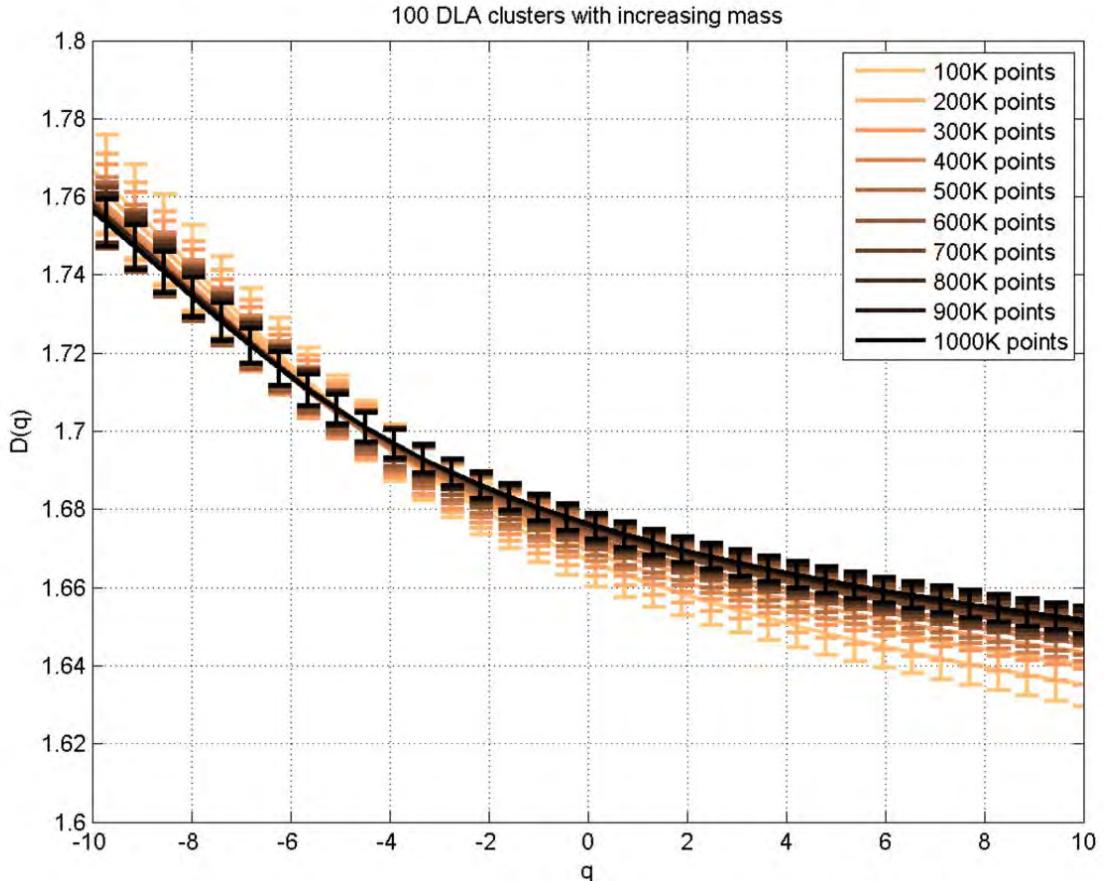


Figure 4-7 D_q - q curves estimated for 100 DLA clusters with increasing masses given in the inset with different colors

We performed a multifractal analysis on DLA clusters using our new NO-BFM method. For this purpose, we grew 100 off-lattice DLA clusters up to a total mass of 10^6 points each. To study the evolution of D_q as the clusters grow, we estimated D_q at mass increments of 10^5 points, resulting in 10 estimates for each cluster as it grows. For robustness of the analysis, as with the synthetics, we required a minimum of 2 circles to calculate the averaging term. The mass was sampled with a logarithmic step $\alpha=0.05$. The sampling is initiated at a mass of 7 points based on the fact that one circle surrounded by six other circles corresponds to the hexagonal lattice, which is the lattice arrangement of circles with the highest density, as proven by Lagrange. The D_q - q curves for the 10 different masses are shown in Figure 4-7, where the error bars indicate the standard deviation obtained over the 100 clusters. To check the convergence of our results in Figure 4-8, we plot D_q as a function of the number of data points in the cluster for $q=-10, 0$ and 10 .

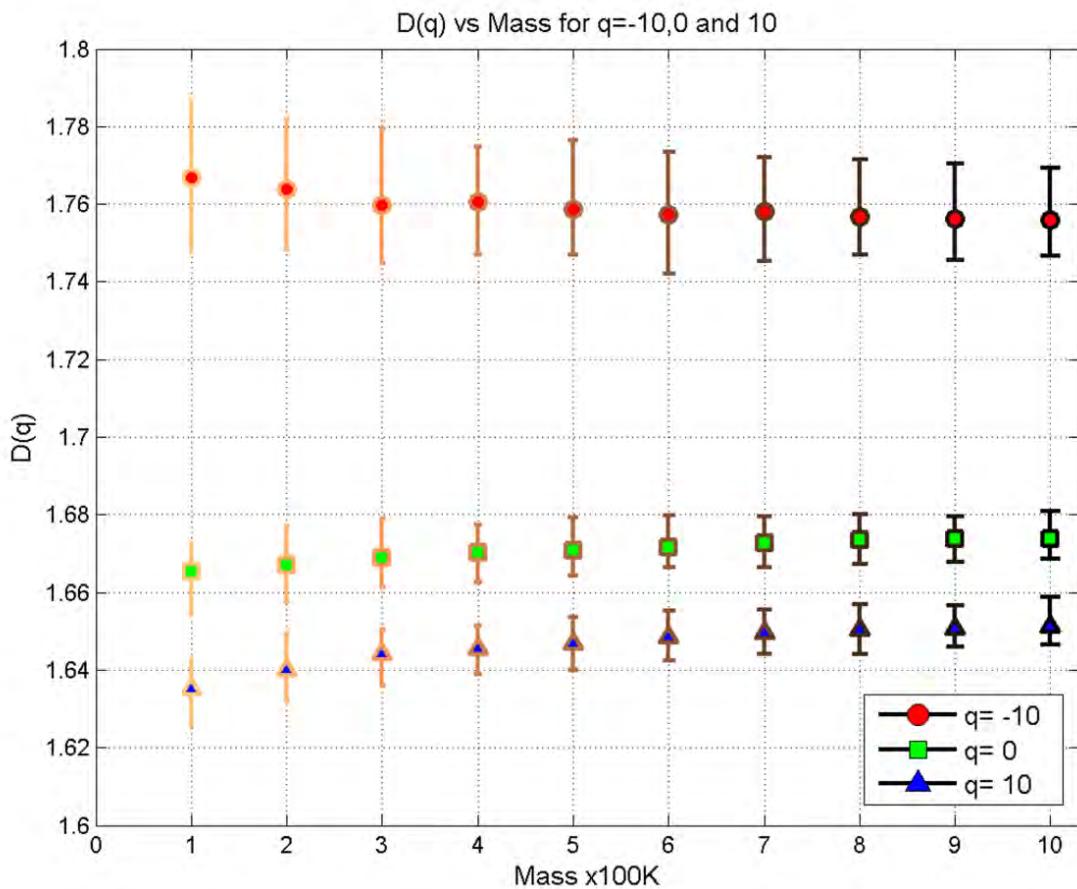


Figure 4-8 Convergence of D_q versus the number of samples in DLAs

To assess the significance of the obtained measure, we conducted the following test: we created a synthetic monofractal set with $D_q=1.63$ ($\log(6)/\log(3)$) and confirmed that the method retrieves the correct dimension with a maximum error of 0.03 with sample sizes as low as $6^6=46656$ points. Secondly, we created a synthetic multifractal having a D_q curve similar to what we observe in the DLA datasets. The density matrix was chosen as $[2 \ 0 \ 2; 0 \ 2 \ 0; 2 \ 1 \ 2]$. We confirmed that, for a sample size of $11^5=161051$ points, the method retrieves the correct dimensions with a maximum error of 0.02. The results are presented in Figure 4-9. We chose regular constructions because they can be associated with the mean DLA branching number 5, resulting in synthetic clusters that look somewhat similar to DLA clusters.

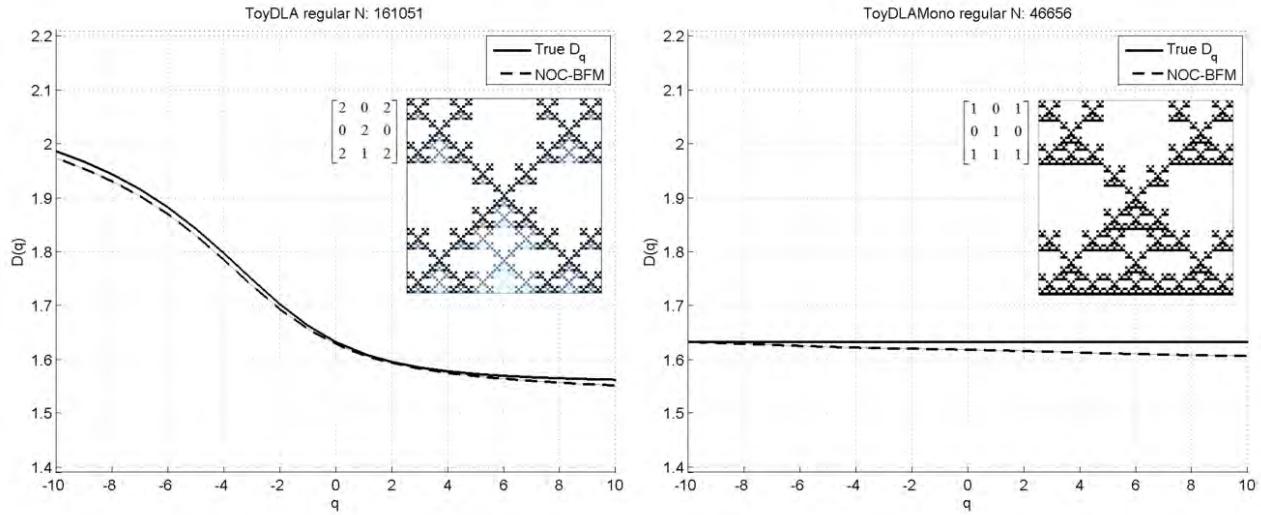


Figure 4-9 D_q - q curves obtained for mono and multifractal toy DLA synthetics

In order to put our results in perspective, Figure 4-10 shows our obtained multifractal D_q curves together with those obtained in [Hanen and Heffernan, 2012] and [Vicsek et al., 1990] for DLA clusters with similar mass. All three results are for DLA clusters of 1 million particles. In contrast to [Hanen and Heffernan, 2012] that uses only 10^4 random pivot points, [Vicsek et al., 1990] uses 5×10^4 random pivot points and square boxes for the fixed-size sandbox method. Also, the R interval for this study was wider: $0.037R_{max} < R < 0.5R_{max}$.

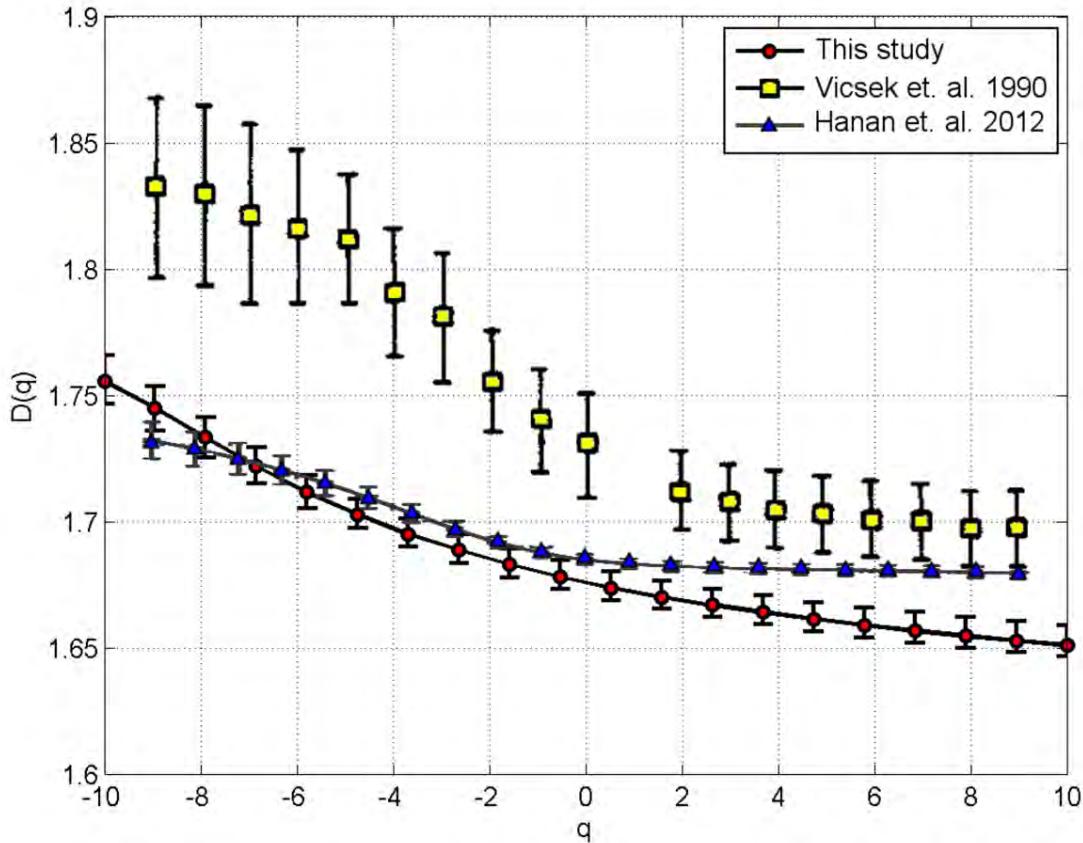


Figure 4-10 Comparison with D_q - q curves reported by Hanan et. al. 2012 Vicsek et. al. 1990

4.6 Conclusions

We have introduced a novel method, coined the Barycentric Fixed Mass Method, to address the ubiquitous edge and finite size effects that plague current determination of multifractal spectra. The method incorporates two motivated criteria (the barycentric pivot point selection and the non-overlapping coverage) in order to reduce edge effects, improve precision and reduce computation time. We have presented extensive tests on synthetically generated mono- and multifractals, both deterministic and random, which demonstrate the superior performance of our proposed method. We have then applied it to the still open question of whether large clusters generated by diffusion-limited aggregation (DLA) exhibit genuine multifractality. Our tests support the conclusion of a genuine by weak multifractality of the central core of DLA clusters, with D_q decreasing from 1.75 ± 0.01 for $q=-10$ to 1.65 ± 0.01 for $q=+10$.

Chapter 5

Condensation of earthquake location distributions: optimal spatial information encoding and application to multifractal analysis of South Californian seismicity

by

Yavor Kamer, ETH Zürich

Guy Ouillon, Lithophyse

Didier Sornette, ETH Zürich

Jochen Wössner, ETH Zürich

published in

Physical Review E, August 2015

5.1 Abstract

We present the “condensation” method that exploits the heterogeneity of the probability distribution functions (PDF) of event locations to improve the spatial information content of seismic catalogs. As its name indicates, the condensation method reduces the size of seismic catalogs while improving the access to the spatial information content of seismic catalogs. The PDFs of events are first ranked by decreasing location errors and then successively condensed onto better located and lower variance event PDFs. The obtained condensed catalog differs from the initial catalog by attributing different weights to each event, the set of weights providing an optimal spatial representation with respect to the spatially varying location capability of the seismic network. Synthetic tests on fractal distributions perturbed with realistic location errors show that condensation improves spatial information content of the original catalog, which is quantified by the likelihood gain per event. Applied to Southern California seismicity, the new condensed catalog highlights major mapped fault traces and reveals possible additional structures while reducing the catalog length by $\sim 25\%$. The condensation method allows us to account for location error information within a point based spatial analysis. We demonstrate this by comparing the multifractal properties of the condensed catalog locations with those of the original catalog. We evidence different spatial scaling regimes characterized by distinct multifractal spectra and separated by transition scales. We interpret the upper scale as to agree with the thickness of the brittle crust, while the lower scale (2.5km) might depend on the relocation procedure. Accounting for these new results, the Epidemic Type Aftershock Model formulation suggests that, contrary to previous studies, large earthquakes dominate the earthquake triggering process. This implies that the limited capability of detecting small magnitude events cannot be used to argue that earthquakes are unpredictable in general.

5.2 Introduction

The latest advances in the instrumentation field have increased the station coverage and lowered seismic event detection thresholds. This has resulted in a vast increase in the yearly number of located events. The abundance of data comes as a double-edged sword: while it facilitates more robust statistics, this comes at the cost of larger computations, with execution times often growing exponentially with the number of data points. In many analyses studying temporal or spatial clustering, a common approach to deal with the large amount of data is to introduce threshold criteria. These can be minimum magnitude, maximum location uncertainty or a specific time or space window. While large magnitude events are of greater importance for risk assessment studies, events with small location uncertainties are crucial for accurately mapping the active/potential unknown structures [Wang, 2013; Wang *et al.*, 2013]. Some studies prefer limiting their data to the most recent periods, however, there is strong evidence that the notion of time invariance does not hold for seismicity distributions, at least on the time scales covered by existing catalogs. Apart from containing a degree of arbitrariness, implementing these thresholds discards data that clearly contain some information, and could potentially be useful.

In this study, we present a novel method that (i) assesses the relative importance of each earthquake hypocenter location data point using its uncertainty as a metric, (ii) reduces the size of the dataset, (iii) preserves the total number of events and (iv) helps improving the signal to noise ratio for statistical analyses such as the multifractal analysis of the spatial distribution of

hypocenters. Our motivation stems from the fact that the geometrical information contained in a seismic catalog is not optimally encoded, is redundant and thus requires unnecessary memory. In other words, the same spatial information can be stored allocating fewer resources. This encoding inefficiency is a result of the sequential data entry where new events are recorded without taking into account the information contained in previous ones. As an extreme example, consider several events with identical parameters occurring at the same location. For the purpose of spatial clustering, it would be optimal to group together and represent them by a single entry with a multiplier (weight). Instead, they each occupy a memory space as if they provided distinct spatial information. We can generalize this idea for events with locations that are not identical but relatively close: given its mean vector and covariance matrix of position errors, each event can be regarded as a representation of the distribution of its possible locations (i.e. a probability density function, PDF). We propose to implement the re-grouping at this micro-scale in order to optimize the encoding of the joint spatial distribution of all events in a catalog. In this way, we also minimize biases of statistical estimators for variables depending on the whole spatial information.

The method does not rely on any assumptions regarding the physical processes generating the point distribution. As such, the method can be readily applied to other point process datasets featuring location uncertainty (e.g. meteorite impacts, explosions, sunspots...). To facilitate a non-arbitrary implementation, the following physical analogy is useful: if each event entry describes the possible location distribution of the event (i.e. microstates), the logarithm of the squared deviations can be regarded as its entropy. The higher the entropy is, the wider is the scatter of microstates. With this definition in mind, the condensation phenomena can be viewed as water vapor (high entropy state) liquefying upon encountering a low entropy setting (lower temperature). Following the same analogy, we can imagine events with high location uncertainties as clouds of water vapor and the ones with more certain locations as droplets of cold water. Under such a setting, the vapor would condense onto the nearest droplets reducing the overall volume (i.e data size) while conserving the total weight content. It is important to note that in this thermodynamic analogy the entropy is related to the actual scatter/disorder of the microstates of gases and liquids. In the case of earthquake locations, the entropy is an expression of the lack of precise knowledge due to observational limitations.

Similar concepts of spatial clustering and entropy have been used in previous studies [*Jones and Stewart, 1997; Nicholson et al., 2000*], with the main difference that these aimed at relocation of events to reveal underlying structures. The collapsing method introduced by Jones and Stewart [*Jones and Stewart, 1997*] iteratively moves each event towards the centroid of all events enclosed by its uncertainty ellipsoid. The process stops when the distribution of all movements becomes comparable to the variance of the spatial distribution of the catalog. This method successfully highlights structures by collapsing multiple events onto each other. However, synthetic tests with uniformly distributed random points produced similar linear structures that could be proven as artifacts of the method. To counter this downside, Nicholson et al.[*Nicholson et al., 2000*] introduced a weighting function in the calculation of the centroid. However, their synthetic test showed that the artifacts were still present, only reduced in amplitude. The method presented in this paper 1) preserves the total seismicity rate of a catalog while minimizing the data size without altering the location of the events, 2) illuminates possible structures as well-located events are highlighted by the condensation of the poorly located ones in their vicinity and 3) regularizes the

catalogs with respect to the spatially varying location quality. This provides an efficient encoding of the location quality for further analysis such as inter-event distances.

The paper is structured as follows. First, we present an illustrative description of the condensation method and its application to Southern Californian seismicity. In Section 3, we investigate quantitatively and visually the effect of condensation in terms of likelihood gain and weight transfer. In Section 4, we perform a comparative multifractal analysis on the original and condensed catalogs. We conclude our paper with a discussion of our findings and their implications for interpreting previous studies and for future studies.

5.3 The Condensation Method

5.3.1 Description and illustration of the method

Condensation transfers weight from events with large location uncertainty (“sources”) to events with smaller location uncertainty (“targets”). For this purpose, reported horizontal and vertical standard location errors are used to estimate an isotropic overall variance for each event in the catalog (see Equation (5-1)).

$$\sigma_{iso}^2 = \sigma_x^2 + \sigma_y^2 + \sigma_z^2 \quad 5-1$$

The implementation of the method follows the following steps:

- 1) Assign a unit weight to each event and sort them in descending order according to their overall variance. Select the event with the highest variance as a source.
- 2) Sample the source’s location P_F using a large number of points (typically 1,000; labeled as offsprings).
- 3) Compute the responsibility of each event (targets and current source) in generating each offspring defined in 2). This is done by estimating the likelihood of the offspring conditioned on the target.
- 4) Each event gains a weight proportional to the ratio of the number of offspring for which it had the largest responsibility.
- 5) We consider the next source event and go to step 2.

The condensation stops when all events except the ones with the lowest variance have been processed, i.e., when there are no target events for the latter to condense on. For a better understanding, we present a step by step illustration of the method applied to a set of 1D normal distributions representing location of five events (Figure 5-1). The distributions have respectively variances of 0.5, 1, 1.5, 2, and 2.5 and are labeled with letters A to E. The event E, having the largest variance, becomes the initial source. Condensation continues until all events except A, which has the lowest variance, have transferred their weight to targets providing the higher likelihoods (Figure 5-1). At the end of the condensation, one observes that the weights of events E and C decrease while those of events A and B increase, the total sum of weights being preserved. The weight of event D remains almost unchanged since no event with better or worse variance is present in its vicinity. For applications to seismic catalogs, this indicates that, in areas with sparse seismicity, condensation will tend to conserve the initial weight distribution (i.e. the initial structure of seismicity). However, many catalogs feature pronounced spatial clustering of events, which may differ significantly in their location accuracies. Therefore, condensation leads to a large fraction of

events ending up with vanishing weights. Discarding these events allows for a reduction in the overall data size of the catalog without any loss in the overall information on the spatial structure encoded in the catalog, as we demonstrate below.

The output of catalog condensation is a vector of weights assigned to each individual event in the original input catalog. The sum of these weights is equal to the total number of events in the input catalog. Events can be classified into the following four categories according to their final condensation weights:

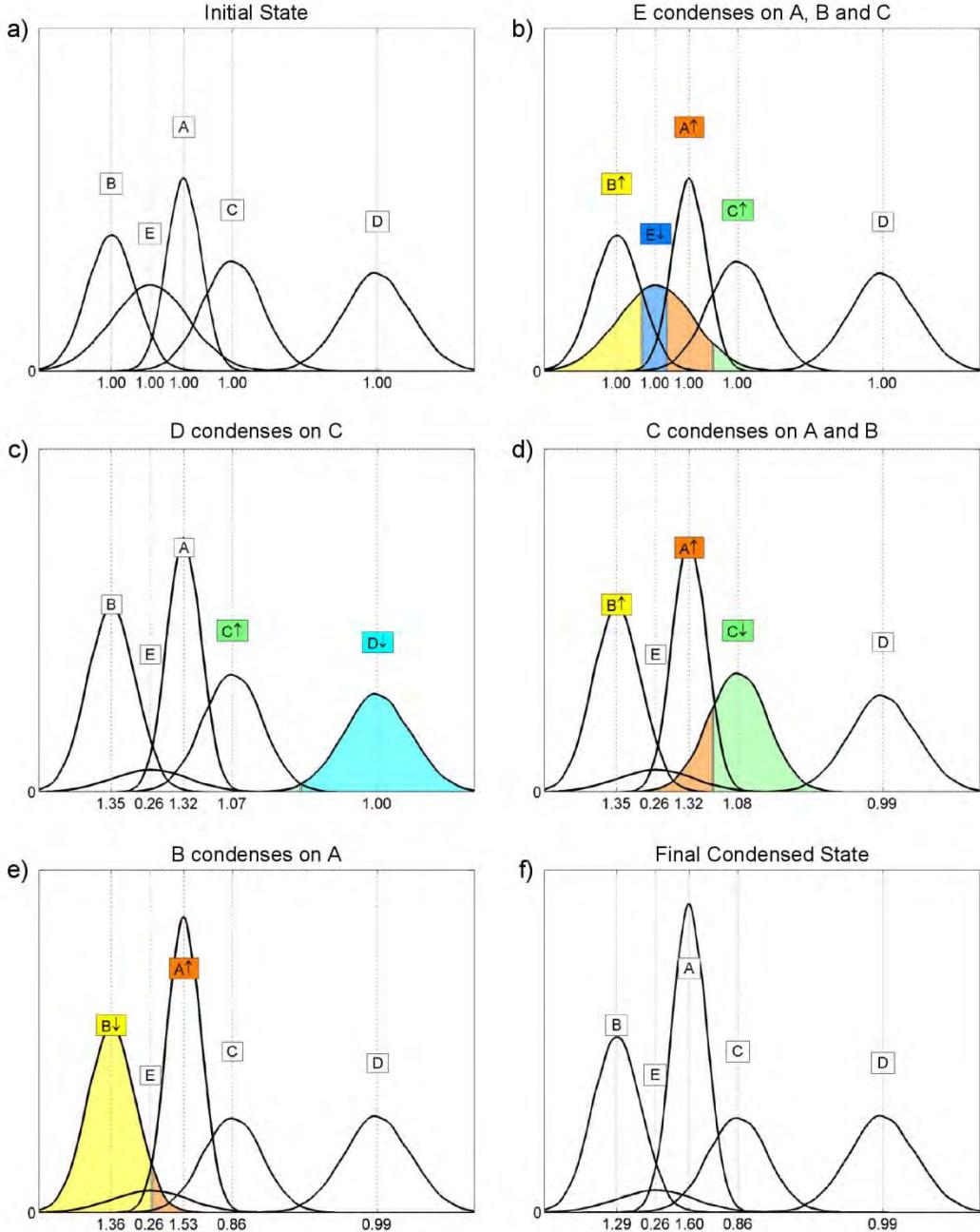


Figure 5-1 Illustration of the condensation procedure for a set of 1D distributions. These are labeled with letters A to E according to their standard deviations [0.5, 1, 1.5, 2, and 2.5]. At each step, source events losing mass are represented by downward arrows while target events gaining mass are labeled with upward arrows. The portion of the probability weight assigned to each event is depicted with its respective color

- Weight >1:* Events that are better located with respect to their neighbors and have thus increased their weights. Note that good location accuracy does not necessarily lead to increased weights since this depends on the local presence of events with higher location

uncertainty. Being selected as a source or a target depends on the respective accuracies of the events with overlapping location error PDFs.

- b. *Weight < 1*: Events that have relatively poor location accuracy and are in the close proximity of better located events. It is likely that these events will further decrease their weights in the future, as location capabilities of the seismic network improve and as new events with better accuracy are recorded in their vicinity.
- c. *Weight = 1*: This occurs for spatially isolated events that neither gained nor lost weight. This can also be observed when several events are close together and have identical variances. Another possibility is that during condensation, an event acquires the exact weight that it loses and hence ends up with a weight ≈ 1 . However, due to computational precision, the final weight is unlikely to be precisely 1.
- d. *Weight = 0*: Events whose spatial PDF information can be virtually expressed as a combination of other better located events. These events can be discarded and hence reduce the catalog's length significantly.

The weights of the condensed catalog can be regarded as coefficients of spatial importance optimized with respect to the spatially varying location capability of the seismic network.

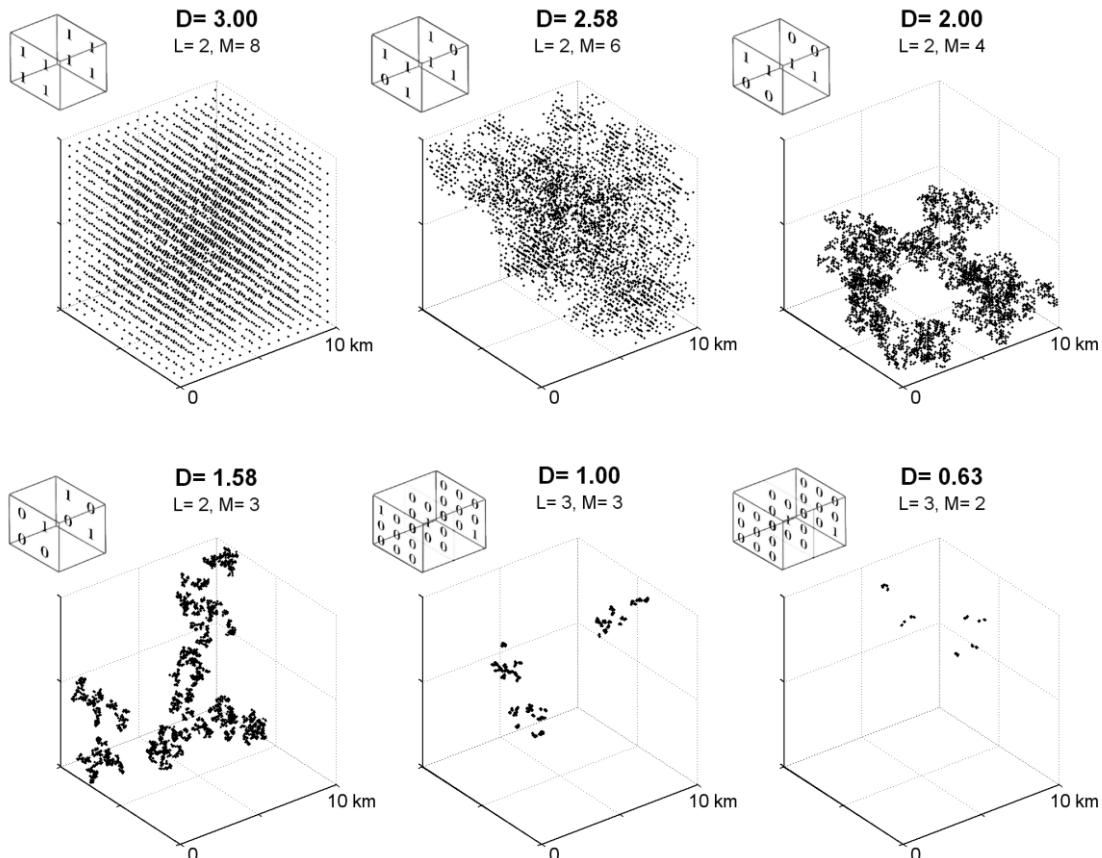


Figure 5-2 Synthetic distributions of 3360 points with different fractal dimension (D). Each distribution is generated by iteratively replicating and permuting the 3D template matrices given in the upper left corners. L and M denote the factors of length reduction and mass increment per iteration, where $\log(M)/\log(L)=D$. See Kamer et al. (2013) for details.

5.3.2 Synthetic Test with Fractal Distributions

In this subsection, we generate synthetic fractal distributions, perturbed with realistic locations errors, and compare their spatial information content before and after condensation. We

consider seven distributions (Figure 5-2), with different fractal dimensions $D=[3.00, 2.58, 2.00, 1.58, 1.00, 0.63, 0]$ generated by recursive replication of template 3D density matrices (for a detailed description of this process the reader is referred to Figure 3 of [Kamer *et al.*, 2013a]). The fractals are generated within a cube of 10km side length and contain 3360 points. For a realistic representation of location uncertainties, we use the covariance matrices of 3360 aftershock events following the 1992 M7.3 Landers earthquake [Wang *et al.*, 2013]. The location uncertainties associated with each of these events has been determined using a probabilistic inversion and expressed by six independent elements in their covariance matrix. The median values of the square roots of the eigenvalues are $\sigma_I=0.89$ km, $\sigma_{II}=0.39$ km and $\sigma_{III}=0.27$ km. The initial locations are stored as the true locations and then perturbed using the randomly assigned covariance matrices. The resulting perturbed locations and the covariance matrices become the perturbed catalog. This perturbed catalog is then condensed to obtain a vector of weights. We then calculate the likelihood of the true (unperturbed) locations with respect to the perturbed (L_p) and condensed catalog (L_c) according to the following formulation:

$$L_p = \prod_{i=1}^N \sum_{j=1}^N \frac{1}{N} f(t_i, r_j, \hat{\sigma}_j) \quad 5-2$$

$$L_c = \prod_{i=1}^N \sum_{j=1}^N \frac{w_j}{N} f(t_i, r_j, \hat{\sigma}_j)$$

where N is the number of points, f is the multivariate normal probability density function, t_i is the true location of point i , r_j is the perturbed location, σ_j is the covariance matrix and w_j is the condensation weight of point j . Thus, we consider each catalog as a Gaussian mixture model represented by a weighted sum of the multivariate normal distributions associated with each event [Bishop, 2007]. We calculate the probability that the true locations were generated by either of these models. This formulation allows us to quantify the likelihood gain for any given arbitrary set of true locations. Figure 5-3a shows the log likelihood gain per point ($\ln(L_c/L_p)/N$) for the seven fractal distributions as a function of the number of data-points. We observe significant likelihood gains for all cases except the uniform ($D=3.00$) case. The gain increases with both the number of samples and the degree of clustering (decreasing D), thus suggesting that it is governed by the minimum spacing (Δd) between the points. To investigate this hypothesis, we calculate analytically Δd for each fractal distribution using the following formula:

$$\Delta d = 10L^{-\log(N)/\log(M)} km \quad 5-3$$

$$D = \log(M)/\log(L)$$

where L and M are the unit length and unit mass of the template matrices given in Figure 5-2. As expected, the likelihood gain increases as Δd decreases (Figure 5-3b). However the fact that the data do not collapse on a single curve suggests that the minimum spacing is not the sole controlling parameter. Spatial clustering of small spacings, is another factor, as well as the interplay between the local anisotropy of the events' locations and the one of the uncertainty ellipsoids.

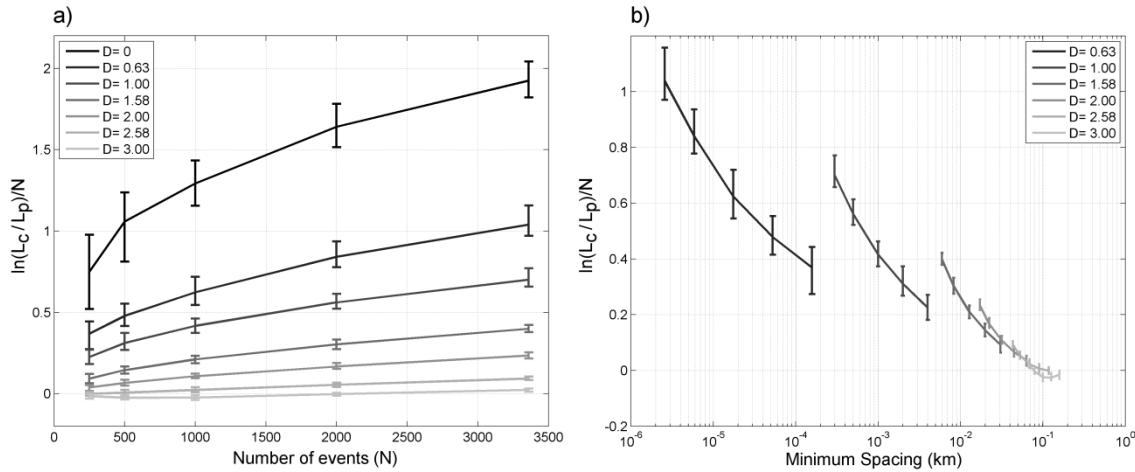


Figure 5-3 Log likelihood gain of the condensed catalog with respect to the original catalog. Each curve corresponds to a different distribution with a given fractal dimension D , calculated for an increasing number of events. b) All curves, except $D=0$, plotted against the minimum spacing calculated from Equation (5-3).

The likelihood gains obtained through condensation can be understood in terms of information retrieval. Each time an earthquake is located, the true location is perturbed with an error vector due to instrumental (signal to noise ratio) and modeling (Earth structure) errors. For a single event, having multiple wave arrival time observations allows one to estimate the amplitude of this error vector (expressed as the standard error) but not its orientation. Thus, each time we locate an earthquake, this information is lost. Condensation facilitates the retrieval of this orientation by exploiting the mutual information of proximate events and giving preference to the more certain ones.

5.4 Application to South Californian Seismicity

5.4.1 Condensing the catalog

As a first application of the method, we use the waveform relocated Southern California earthquake catalog of Hauksson et al. [Hauksson et al., 2012]. Most of the events in this catalog have been relocated using differential travel times and a 3D velocity model. Since condensation is based on the absolute location quality of all events, we consider the absolute location errors that are provided as one-sigma errors for the horizontal and vertical components. As a pre-filter to reduce the size of the catalog, we exclude events with horizontal or vertical error larger than 20 km, resulting in a total of 493,025 events. Keeping these events would not change our results as their large location errors implies that the condensation method would make them disappear by distributing their mass approximately uniformly to a large number of neighboring events. The 493,025 events are ranked by their descending isotropic variances. Since the one-sigma errors are reported with a 100m resolution, the isotropic variance distribution becomes discrete and results in a total of 5,651 distinct groups. The events in each group are then used in turn as sources, transferring weights to target events in lower variance groups. The condensation reduces the weights of 111,487 poorly located events to zero, while increasing correspondingly the weights of better located events in their vicinity. This corresponds to an overall data length reduction (i.e. compression) of 22.6%. An investigation of the temporal distribution of events with zero weights reveals that they are most numerous ($\approx 10,900$) in 1992, the year of the Landers earthquake, which triggered a large number of aftershocks. The year with the second largest number of events with

ero weight is 1994 (Northridge earthquake) with $\approx 9,700$ events (40% of all events recorded that year). This implies that many aftershocks of these two major earthquakes in Southern California have been recorded with rather poor location quality relative to the other events, notwithstanding the relocation procedure.

The probability density distributions of the vertical, horizontal and isotropic errors over all events belonging to the original and condensed catalogs are shown in Figure 5-4. The change in the distributions depicts the weight transfer occurring between events with high and low spatial variance governed by their proximity and relative location error distributions.

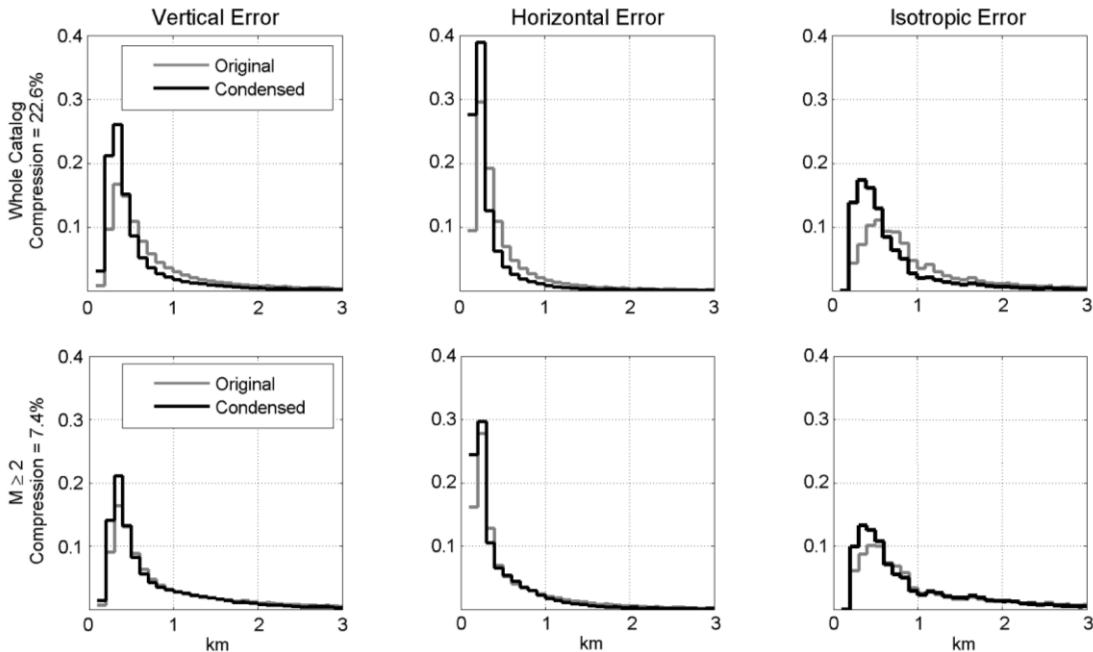


Figure 5-4 Probability density distributions of vertical, horizontal and isotropic errors of the original and condensed catalogs: first row of panels for the whole catalog; second row of panels for events with $M \geq 2$

In the presented study the condensation method is applied to a set of hypocenter (i.e. nucleation points), and yields a new catalog of nucleation points (which is a subset of the original one). The analysis of nucleation points can be regarded as a robust approach due to the fact that any statistical analysis of seismicity is dominated by the more numerous small magnitude events [Sornette *et al.*, 1996]. For instance, constraining the analysis to events smaller than M5 (rupture length of ~ 1.4 km [Wells and Coppersmith, 1994]) reduces the number of data points only by 107 ($\sim 0.02\%$ of all events) and thus the results remain unaffected. Furthermore, the extended geometry of large earthquakes is always illuminated by nucleation points of small magnitude aftershocks.

5.4.2 Visualizing and quantifying weight transfer due to condensation

Since seismicity is distributed within a volume, visualizing the weight transfer occurring as a result of the condensation would require volumetric density plots and slice planes at various angles, which become difficult to interpret. For a simpler illustration, in this section, we project all events onto the surface and thus omit the vertical dimension. With this simplification, condensation is performed in 2D, relying only on the horizontal location errors. In this 2D setting, the compression rate increases to 41.1%. To estimate the probability density distributions, the spatial extend of the catalog is discretized using a 100x100 m grid resulting in a 8052 x 8341 matrix. The

probability density functions for each catalog are estimated at each of these grid points by summing up the weighted contributions of the PDFs of all individual events. Notice that the two catalogs differ only in the weights assigned to the events (a constant of 1 for the original and varying weights for the condensed one). The total PDF of the original catalog is then subtracted from the total PDF of the condensed one. The total displaced weight is calculated by integrating the absolute differences divided by two, in order to account for the fact that transferred weight is double-counted as it is reported as negative at the source location and as positive at the target. At the initial resolution of 100 m, we calculate that 14.7% of the total weight has been displaced due to condensation. We note that the percentage of displaced mass and the compression level (41.1%) are not directly comparable because the former is coarse grained at the scale of the grid cells (100m) while the latter is estimated on a pointwise basis (i.e. corresponding to a cell size $\rightarrow 0$).

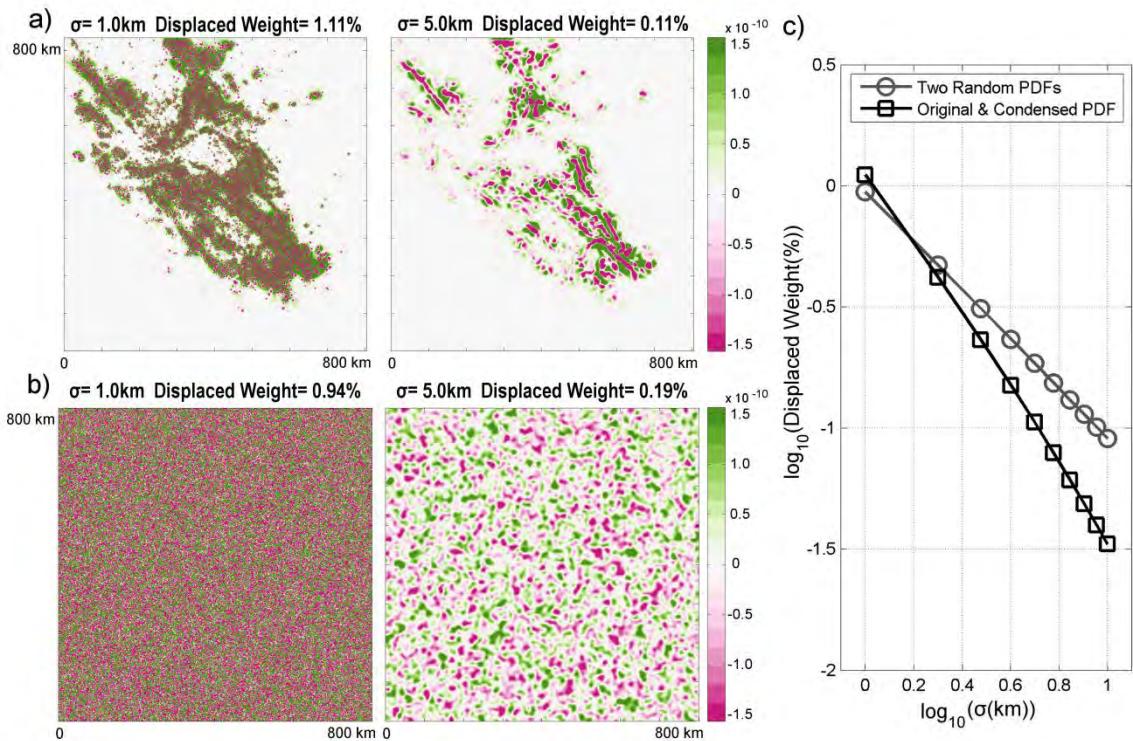


Figure 5-5 Areas of weight enrichment (red or dark gray) and depletion (green or light gray) comparing a) the original and condensed catalogs of Southern California, b) two uniform random spatial PDFs with similar extends for Gaussian filters with bandwidths of $\sigma=1$ and $\sigma=5$ m. c) Percent of displaced weight as a function of filter bandwidth σ .

The percentage of displaced weight includes all displacement occurring at scales larger than the grid resolution (100m). To depict the effect of condensation more accurately, we quantify weight displacement as a function of distance. For this purpose, we apply a rotationally symmetric Gaussian low-pass filter to the total PDFs of the two catalogs (original and condensed). By varying the bandwidth of the filter, we can effectively coarse-grain the PDFs, so that the measure is insensitive to weight displacement occurring at scales smaller than the filter bandwidth. We observe that the total displaced weight decreases significantly as the scale of interest is increased (Figure 5-5c).

To put these results in perspective, we compare the displaced weight as a function of the filter bandwidth curve with the same curve obtained by subtracting two random PDFs spanning the same extent as the seismicity catalog (Figure 5-5). The random PDFs were generated by assigning

each cell with a random number drawn from a uniform distribution [0 1] and then normalizing the whole matrix so that it integrates to one. The difference of two standard independent uniform random variables (X_1, X_2) follows the standard triangular distribution given as:

$$Y = X_1 - X_2$$

$$f_Y(y) = \begin{cases} y+1 & -1 < y < 0 \\ 1-y & 0 \leq y < 1 \end{cases} \quad 5-4$$

As discussed above, the analysis of displaced weight considers half of the absolute differences, whose distribution is given as:

$$f_Y(z) = 2(1-z) \quad 0 \leq z < 1 \quad 5-5$$

Equation (5-5) represents a right-angled triangle and as such this distribution has its mean at $z=1/3$. Thus, for two random PDFs, the expected displaced weight at the original resolution is 33%, which is considerably larger than the 14.7% observed for the original and condensed catalogs of Southern California seismicity. For increasing spatial bandwidths, the two curves given in Figure 5-5c show that the weight transfer due to condensation is focused at small scales (~1km) and decreases more rapidly as a function of scale for the natural catalog than for a random displacement process. The limited scale of mass transfer is also in agreement with the distributions of horizontal location errors given in Figure 5-4. These results show that condensation is consistent with the overall location error distribution and preserves the overall spatial features.

For a qualitative inspection, in Figure 5-6, we superimpose the fault traces obtained from the Community Fault Model [*Plesch et al.*, 2007] onto the map of weight depletion/enrichment coarse-grained at bandwidth $\sigma=3\text{km}$. This value is chosen in agreement with previous fault observations of apparent low-velocities zones with similar widths [*Feng and McEvilly*, 1983]. Maps obtained for $\sigma=[1-10\text{km}]$ are presented in the electronic supplement. Notice that the total weight transfer at this spatial scale ($\sigma=3\text{km}$) is merely 0.23% and the maximum weight transfer at each grid cell does not exceed 1.5×10^{-10} . Due to the omission of the depth component in this 2D illustration, weight transfers on dipping faults are projected on the surface and hence the depletion/enrichment regions close to these faults may be exaggerated, but nevertheless limited at the scale of analysis (σ). In many places, such as the Brawley and Laguna Salada fault zones in the South-East, the structures highlighted by weight increase coincide with the observed and extruded fault traces. It is possible to infer larger structures such as the San Jacinto Fault and the San Andreas Fault. In other parts of the map, regions of weight accumulation seem to produce patch-like features without a linear structure. This is mostly observed on the San Clemente Fault where condensation highlights small features perpendicular to the fault trace. One explanation could be that this is due to issues of offshore network coverage; however for the Santa Cruz - Santa Catalina Ridge Fault Zone, which is also offshore, we observe a good correspondence. These results, together with the likelihood gains reported in the previous section, suggest that the condensation method can be used to complement fault network reconstruction applications, which rely on high quality location data [*Ouillon and Sornette*, 2011; *Wang et al.*, 2013].

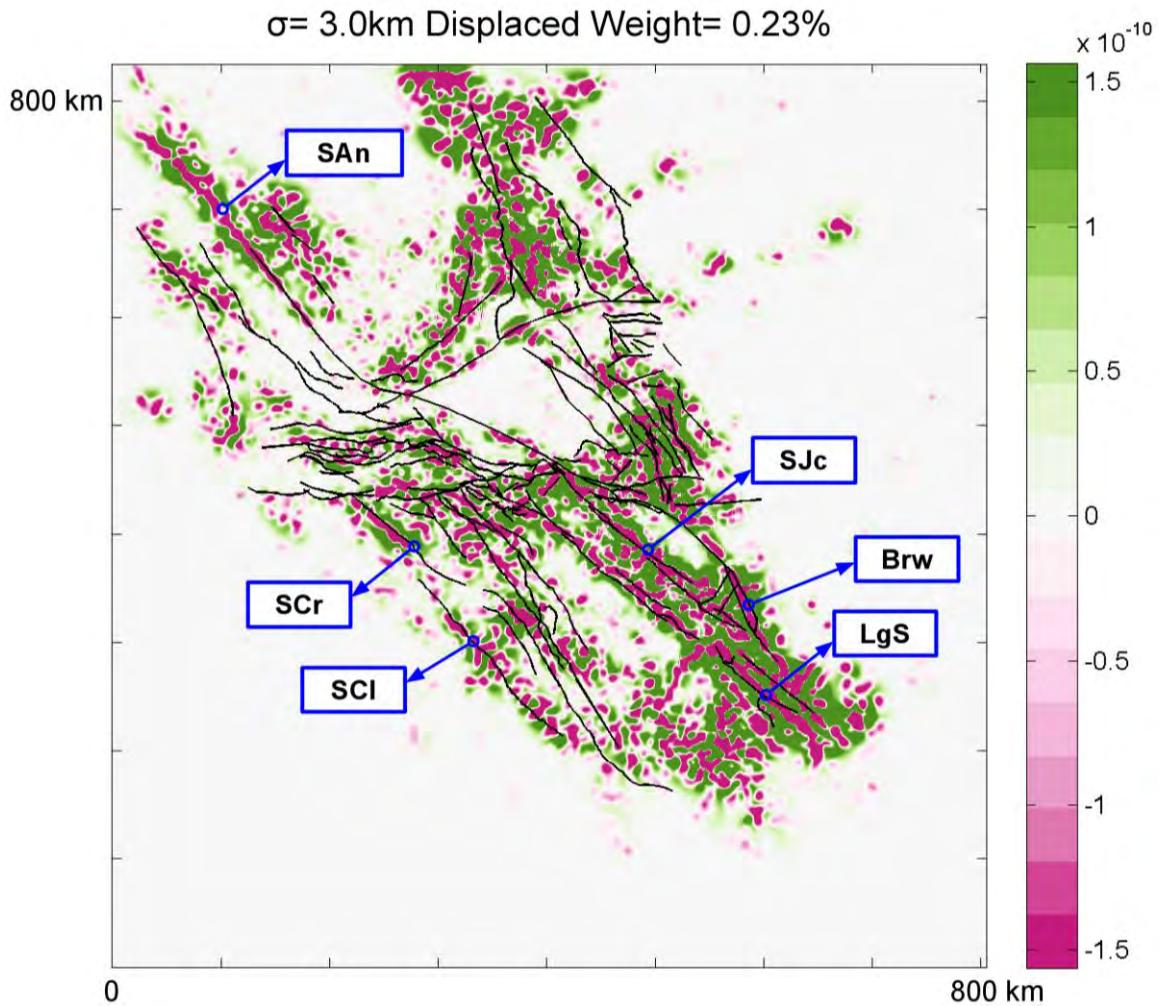


Figure 5-6 Areas of weight enrichment (red or dark gray) and depletion (green or light gray) resulting from condensation at bandwidth of 3km, superimposed with the fault traces obtained from the Community Fault Model. Individual faults are labeled with the following abbreviations: San Andreas (SAn), Santa Cruz (SCr), San Clemente (SCI), San Jacinto (SJc), Brawley (Brw) and Laguna Salada (LgS).

5.5 Multifractal properties of the original and condensed catalogs

In this section, we investigate the implications of the condensation method for the multifractal properties of the seismicity distribution. Such an investigation is of particular interest because it provides a quantitative description of the spatial patterns at various scales. Since condensation allows us to incorporate the location quality information in the form of a scalar weight, we are interested in whether we find the same values in the appropriate scaling regimes for both original and condensed catalogs. Additionally, we are interested in re-evaluating scaling regimes at small scales that until now have remained concealed due to the effect of location errors [Marsan, 2005; Kagan, 2007].

5.5.1 The multifractal formalism and classical estimation methods

Even a simple question such as “how long is the coast of Britain” [Mandelbrot, 1967] can become problematic since the answer changes as a function of the observation scale. The formalism of Fractals aims to extract this functional form (i.e recursive scaling regime) and provide insight into the underlying phenomena. The term fractal (monofractal, or homogeneous fractal) implies that

the scaling can be quantified by a single exponent, the fractal dimension. Multifractal distributions can be seen as different sets with different scaling properties interwoven altogether. The consequence is that the singularity of the underlying distribution fluctuates from place to place, the complexity of the structure being fully encoded by the multifractal spectrum. A common way to quantify the fractal or multifractal properties of a given set of data points is to calculate its generalized (Renyi) dimensions [Renyi, 1970], given as:

$$D_q = \lim_{\varepsilon \rightarrow \infty} \frac{\frac{1}{1-q} \log(\sum_i p_i^q)}{\log(\frac{1}{\varepsilon})} \quad 5-6$$

where ε is the scale of observation, $p_i(\varepsilon)$ is the fraction of data points (e.g, estimated measure) within box i of size ε , q is a real-valued moment order and the sum is performed over all boxes covering the data set under investigation. Varying the q parameter, D_q characterizes the scaling of the underlying measure within the distribution. Thus, D_∞ and D_0 respectively correspond to the local scaling of the lowest and highest density areas, i.e. to the weakest and strongest singularities of the distribution. For monofractal sets, D_q is a constant independent of q . For multifractal distributions, D_q decreases monotonically with q .

The commonly used multifractal analysis methods can be classified into two broad classes, called fixed-size and fixed-mass methods respectively. Fixed-size methods (FSMs) [Jensen *et al.*, 1985; Tél *et al.*, 1989] estimate D_q via the scaling of the total mass M (i.e number of data points) within a constant r -sized sphere, as r is increases:

$$\log \langle M(<r)^{q-1} \rangle \approx (q-1)D_q \log(r) \quad 5-7$$

Fixed-mass methods (FMMs) estimate D_q via the scaling of the smallest radius r to include a fixed mass m , as m is increases:

$$\begin{aligned} \log \langle R(<m)^{-(q-1)D_q} \rangle &\approx -(q-1)\log(m) \\ \tau(q) &= (q-1)D_q \\ \log \langle R(<m)^{-\tau(q)} \rangle^{1/\tau(q)} &\approx \log(m)^{-1/D_q} \end{aligned} \quad 5-8$$

Several studies report FMMs to be superior to FSMs [Badii and Broggi, 1988; Grassberger *et al.*, 1988]. For a detailed review of both FSMs and FMMs, the reader is referred to [Theiler, 1990].

Many previous studies have undertaken the task of estimating fractal dimensions for seismic catalogs. However, most of the published results are questionable because I) the used methods are prone to finite size and edge effects that have not be adequately addressed [Theiler, 1990]; and II) of lack of benchmarks with synthetic fractals with analytically derivable fractal dimension. For instance, many studies use the correlation integral [Grassberger and Procaccia, 1983] to estimate the fractal dimension (D_2) of hypocenter or epicenter sets [Helmstetter *et al.*, 2005; Marsan, 2005]. The observed scale ranges in such studies are usually quite limited and, due to the inherent finite size and edge effect, it is difficult to quantify the quality of the measure. Phase transitions, observed

in the form of change of slopes, are usually identified manually and their attribution to physical dimensions (such as seismogenic layer thickness) or locations errors (horizontal or depth) remains somewhat speculative. Kagan [Kagan, 2007] has made notable efforts to facilitate the applicability of the correlation dimension measure by characterizing and correcting for edge effects and location errors. In the same study, Kagan casted doubt on the significance of fractal analysis performed under these conditions and proposed that studying the higher order point configurations might be a better option. Interestingly, Hirabayashi et al [Hirabayashi et al., 1992] reached the same conclusion when they showed that using fixed-mass methods provides more reliable results in multifractal analysis of seismic catalogs. Nevertheless, even with the use of FMMs, dealing with edge-effects remains a problematic task often tackled by introducing scaling limits or data censoring [Hirabayashi et al., 1992; Zamani and Agh-Atabai, 2011].

5.5.2 The Barycentric Fixed Mass estimation Method

To address the problems associated with the commonly used methods, we have previously introduced a new non-parametric method for multifractal analysis [Kamer et al., 2013a]. The so-called Barycentric Fixed Mass (BFM) method incorporates two criteria aimed at reducing edge effects, improving precision and decreasing computation time; a) barycentric pivot point selection and b) non-overlapping coverage. As most fixed mass methods, the BFM method has more stable results at small scales, since it avoids sampling empty spaces by extending the measuring scale to the next neighboring point. Figure 5-7 compares the performance of the BFM method with two commonly used FMM and FSM.

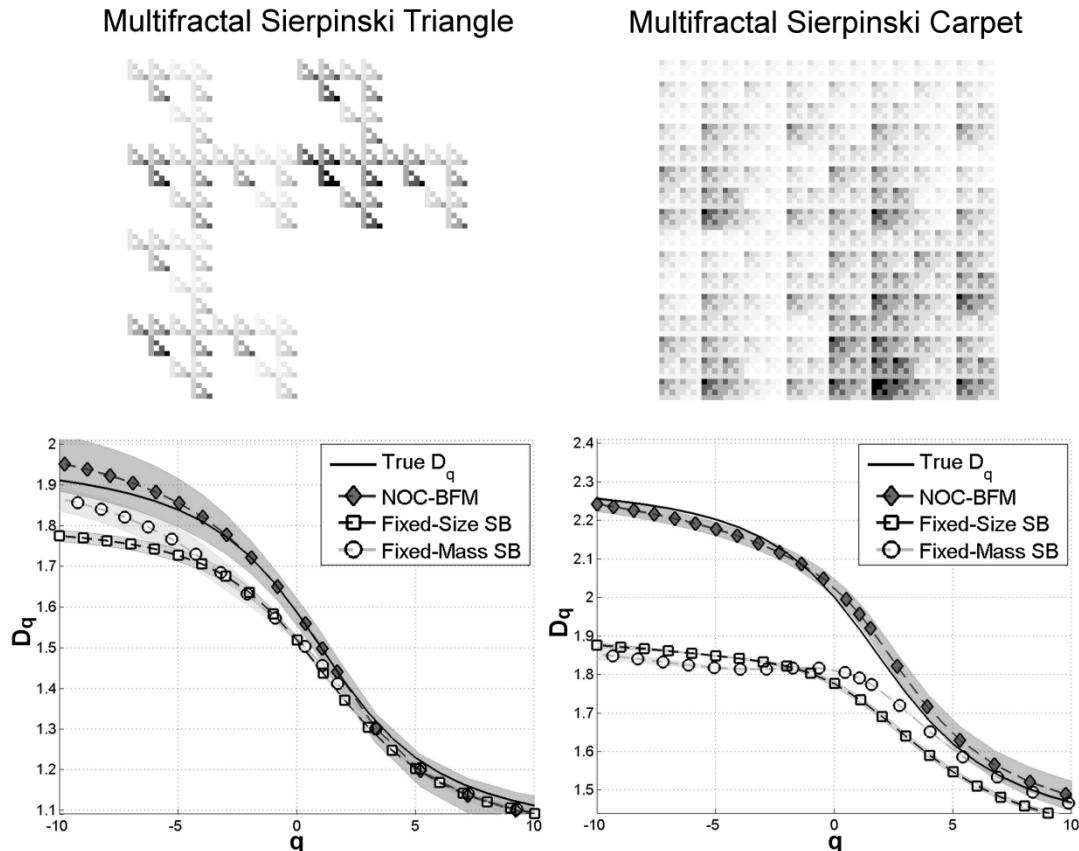


Figure 5-7 D_q vs q curves for two synthetic multifractal point distributions given in the upper right insets: A multifractal Sierpinski triangle (left) and Sierpinski carpet (right). Analytically derived true D_q (solid line) is shown together with

the BFM method (diamonds), fixed-size (squares) and fixed-mass (circles) methods. Reproduced from Figure 6 of *Kamer et al. [2013]*

5.5.3 Multifractal analysis of the Southern Californian Seismicity with robust estimation of the different scaling regimes

For a given dataset, the D_q and q values are estimated from the moment curves $\log\langle R(m)^{-\tau} \rangle^{1/\tau}$ calculated from fixed-mass spheres covering the point distribution. Incrementing the exponent τ allows sweeping q in the range of interest. Previous studies have found that, for negative q values, the D_q measure becomes unstable due to the inherent undersampling of the emphasized regions [Amitrano et al., 1986; Hirabayashi et al., 1992; Ouillon et al., 1996]. That is why we focus our attention on dimensions D_0 to D_5 . D_q is estimated from the $\tau(q)$ -moment curve's slope, while q is obtained subsequently as $q=1+\tau(q)/D_q$, and we use $\tau(q)$ values of [-3, -2, -1, 0.1, 1, 2, 3, 4, 5, 6]. This allows to roughly cover the interval $q \in [0;5]$. The mass range m is sampled at logarithmically spaced steps rounded to their closest integer value, given by $m_i = m \cdot 10^{i\alpha}$, with $\alpha=0.05$, where the smallest possible mass value is $m=2$. The curves of averaged radii versus fixed-mass for the different τ values are given in Figure 5-8. Since the configuration of covering spheres is stochastic, we can reduce the variance in the curves by repeating the measurement multiple times and averaging the resulting curves (Figure 5-8). Averaging multiple realizations reduces the variability observed in the large mass ranges.

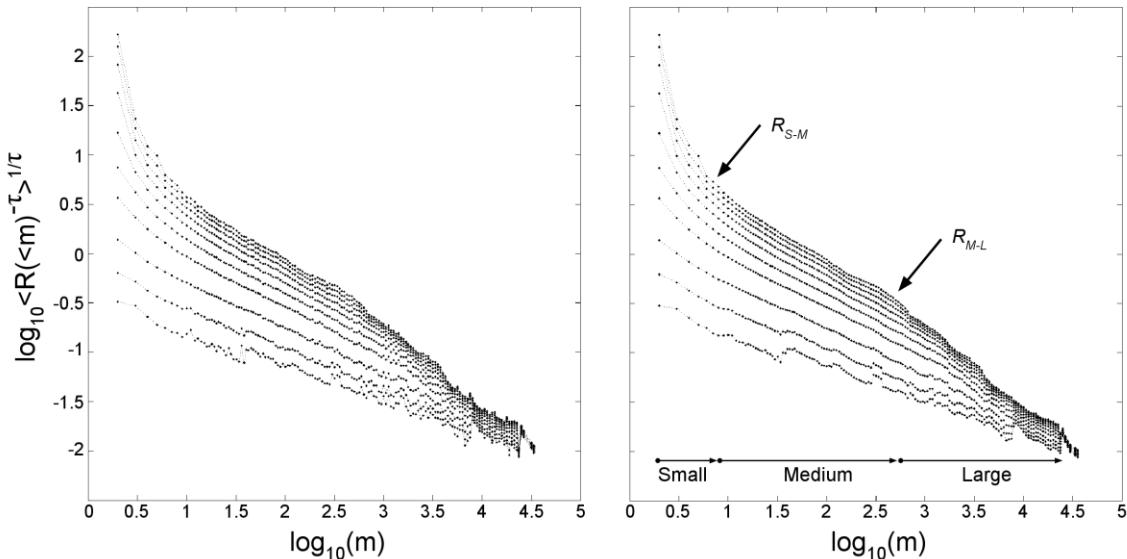


Figure 5-8 Averaged radii versus fixed-mass for increasing τ obtained from a single measurement (left) and averaged over 100 measurements (right) of the ≥ 2 Southern Californian seismicity. The two arrows mark the transitions from small to medium scales (R_{S-M}) and medium to large scales (R_{M-L}).

Both D_q and q are estimated via the local slope for each τ exponent. For relatively simple multifractals, such as the widely studied growth process of Diffusion Limited Aggregation, the entire m range can be characterized by a single set of slopes (see Figure 1 of [Vicsek et al., 1990]). However, in the case of the Southern Californian hypocenter distribution, we observe phase transitions highlighted by changes in these slopes as a function m . The visually identifiable breakpoints in Figure 5-8 mark the transitions from small to medium scales (R_{S-M}) and medium to

large scales (R_{M-L}). These are of particular interest as they might provide insight into different characteristic length scales that govern the seismogenic processes. For instance, in their multifractal analysis of fault networks in Saudi Arabia, Ouillon et al. [Ouillon et al., 1996] showed that such characteristic length scales might correspond to the rheological stratification of the crust.

The task of identifying the number of observed scaling regimes and their effective ranges can be viewed as an optimization problem. In this setting, we model the curves as a set of discontinuous and piecewise linear functions where each segment is characterized by its slope and intercept. The segments, defined by their breakpoints, are imposed on all the curves simultaneously. The sum of squared errors ($Q(s)$) to be minimized is given as:

$$Q(s) = \sum_{i=1}^t \sum_{j=1}^s (R_{obs}(m) - R_{Mod}(m, i, j))^2 \quad 5-9$$

$$R_{Mod}(m, i, j) = a_{ij}m + b_{ij} \quad p_j < m \leq p_{j+1}$$

where T is the number of curves corresponding to each τ exponent, S is the number of segments, R_{Obs} and R_{Mod} represent the observed and modeled ordinates of the curves, a_{ij} and b_{ij} are the slope and intercept estimated for segment j of curve i and p_j is the breakpoint between segments j and $j+1$. Equation (5-9) implies that the SSE would tend to zero as the number of segments (i.e. the complexity of the model) is increased. Thus, it becomes essential to include a regularizing term that penalizes the goodness of fit for the complexity of the model. In their study, Seidel et al. [Seidel et al., 2004] address a similar problem of investigating the number of linear trends in the global atmospheric temperature record by comparing different models, using the Bayesian Information Criterion (BIC) [Schwarz, 1978]. In this study, our goal is to represent the continuous multifractal spectra in an interpretable form, rather than identifying the single best model to describe the data. For this purpose, we use an ensemble approach where we consider best fitting models with different number of segments (see Figure 5-9). Each model is essentially a staircase function defining constant slopes in each segment. By averaging the different staircase functions obtained for each curve, we are able to obtain a continuous set of slopes. These are used in Equation (5-8) to obtain the continuous multifractal spectra that specifies D_q for any given q and R . The proposed approach is much simpler in the sense that it can be applied without the need to account for correlation between data points, which can become problematic for BIC [Seidel et al., 2004], and it can handle the irregular sampling intervals at small m values.

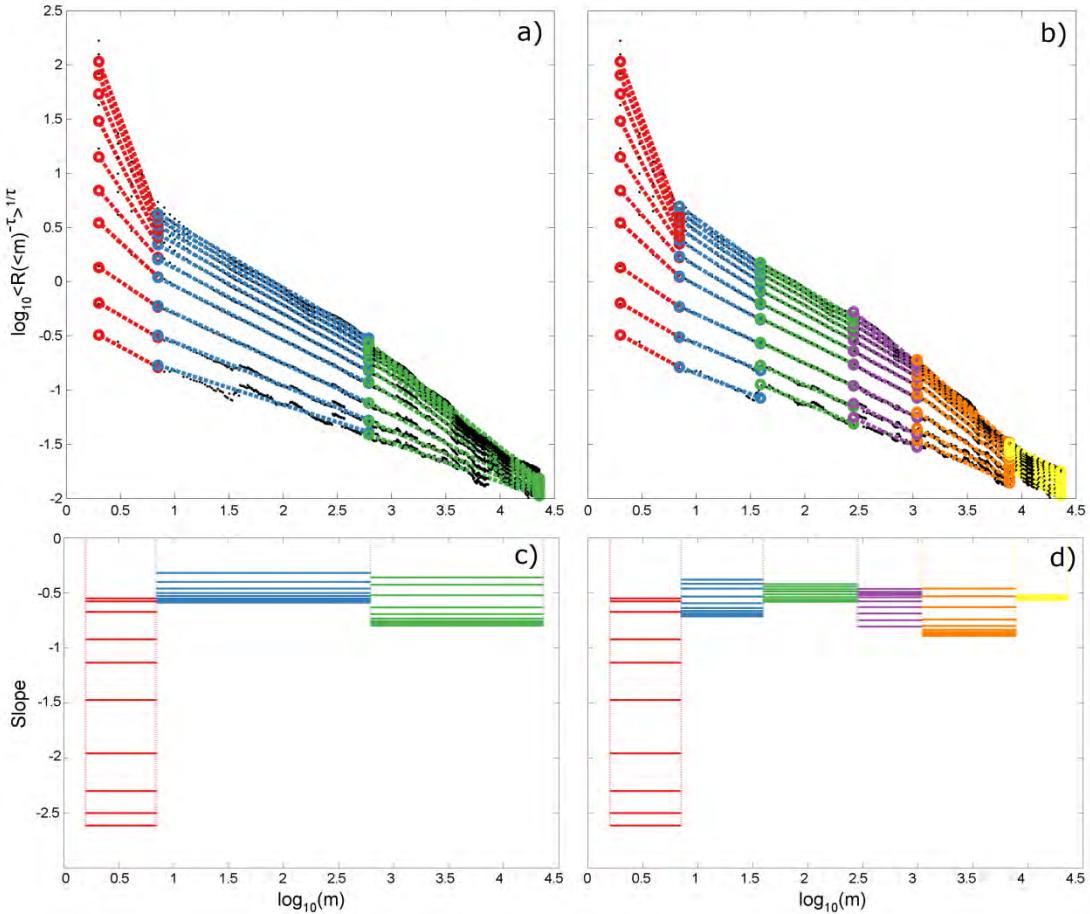


Figure 5-9 Best fitting piecewise linear models with a) 3 and b) 6 segments fitted to the moment curves shown in Figure 5-8. Different segments are represented by different colors (shading). The staircase slope functions obtained for the best fitting models shown in a,b) are shown in c,d).

The ensemble is obtained by averaging the slopes of all the best fitting models with 3 to 10 segments. Our choice for the minimum of 3 segments is based on the general shape of the curves and on previous studies reporting the presence of similar numbers of apparent scaling regimes [Marsan, 2005; Kagan, 2007]. We confirmed that our results are stable with respect to these initial choices by varying the minimum and maximum number of segments.

5.5.4 Multifractal analysis of the Southern Californian Seismicity: condensed versus original catalogs

The continuous multifractal spectra obtained for the original and condensed Southern Californian seismicity with ≥ 2 are given in Figure 5-10. For the original catalog, we observe changes in the scaling regimes occurring at several scales. The first scale is $R_{M-L} \approx 10$ m ($m \approx 400$): this scale can be inferred as the effective thickness of the crustal seismogenic width ($2R \approx 10-25$ km [Nazareth and Hauksson, 2004]). The decrease of dimensions at this scale can be understood by considering the case of a plate with finite thickness a that is sampled with a uniform point distribution; spheres with $R < a$ will report $D=3$ while, for $R > a$, the spheres will be insensitive to the thickness and thus report $D=2$. We observe a similar decrease of the dimensions $D_{q>2}$ beginning at $R_{M-L} \approx 10$ m. This scale is also consistent with the depth distribution of the catalog (mean $\mu = 7.95$ m, standard deviation $\sigma = 4.44$ m; $2R_{M-L} \approx 20$ m corresponds to the 0.98 percentile). Another transition between small and medium distances is $R_{S-M} \approx 1.5$ m ($m \approx 6$): *a priori* it is difficult to

conclude if it stems from a genuine physical process or from location uncertainties. However, if the latter is true, we would rather expect generalized dimensions close to 3 at $r < R_{S-M}$.

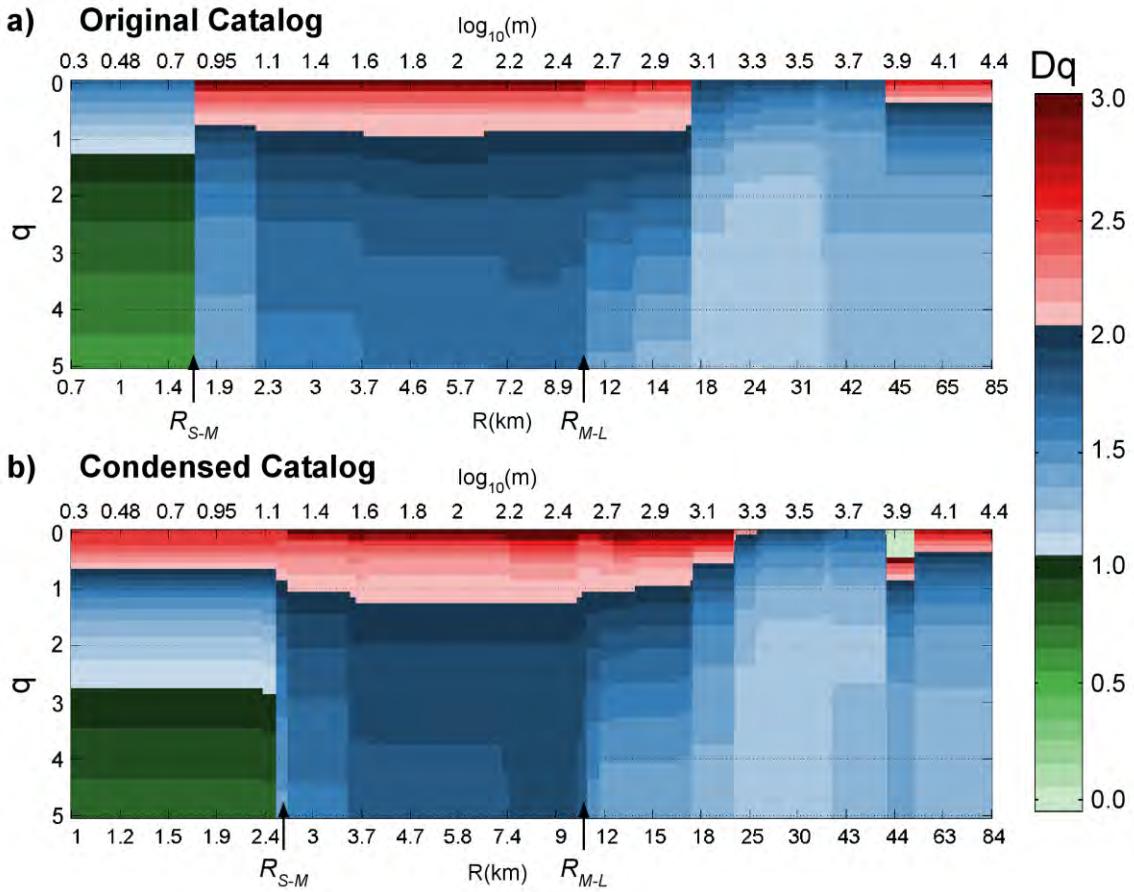


Figure 5-10 Continuous multifractal spectra of the a) original and b) condensed catalogs. The horizontal axis indicates the increasing radius (bottom) and mass (top); the vertical axis represents the q value and individual colors represent D_q within the range [0-3]. The two arrows mark the transitions from small to medium scales (R_{S-M}) and medium to large scales (R_{M-L}).

For the sake of clarity, we shall investigate the clustering properties in terms of D_2 and the multifractality in terms of $\Delta D = D_1 - D_5$. At small scales ($r < R_{S-M}$), the clustering is more pronounced with low $D_2 = 0.9$, while for medium scales ($R_{S-M} < r < R_{M-L}$), it is significantly reduced as evidenced from the high values of $D_2 = 1.8-1.9$. At large scales ($r > R_{M-L}$) we observe $D_2 = 1.2-1.3$. The multifractality is strongest at the small scales with $\Delta D = 0.5$, decreasing at medium scales to $\Delta D = 0.3$ and decreasing even further to $\Delta D = 0.2$ at large scales.

Similarly, for the condensed catalog (Figure 5-10b), we observe two scaling breaks close to the ones reported above. The gradual transition at large scales becomes more pronounced and remains at $R_{M-L} \approx 10$ m; however, the small to medium scale transition is shifted from $R_{S-M} \approx 1.5$ m to $R_{S-M} \approx 2.5$ m. Furthermore, we observe a significant decrease of clustering within the small scale regime ($r < R_{S-M}$) as D_2 increases to 1.2. A similar increase of the fractal dimensions is also observed for the medium scale range ($R_{S-M} < r < R_{M-L}$): $D_2 = 1.9-2.0$. At large scales ($r > R_{M-L}$), the D_2 values remain similar to the values observed in the original catalog. In terms of multifractality, we observe a significant increase at small scales with $\Delta D = 0.9$, while the values at medium and large scales remain similar.

The difference observed in the fractal dimensions of the original and condensed catalogs indicates that the location uncertainty information, which is the basis of condensation, should be an important factor in the spatial clustering analysis of any seismic catalog. As a reminder, the only cause for the different results in the multifractal analyses is the consideration of location error information. Both catalogs have the same event locations and the BFM analyses are conducted in the exact same ways. Previous studies have disregarded location uncertainties in the analysis step only to introduce it to help the interpretation of the scales where the phase transitions occur [Marsan, 2005]. It is important to note that, for both the condensed and the original catalog, the small to medium phase transition occurs at distances that are much larger compared to the vertical and horizontal errors of the catalog (see Figure 5-4). To verify that the observed scaling breaks are not due to location errors, we perturbed the events in the catalog according to their confidence ellipsoids. We repeated the analysis by scaling up the ellipsoids with a factor of 10 and observed that R_{S-M} increased only by a factor of about 1.5, R_{M-L} remains unaffected while D_q values on all scales increased. This indicates that the observed R_{S-M} cannot be due to location errors since one would expect a higher degree of dependency between the two.

5.5.5 Multifractal analysis of a multiscale synthetic dataset

The results suggest that the Southern Californian seismicity catalog features a distinctive scaling regime at small scales. In order to demonstrate that the methods used in this study are indeed capable of correctly detecting phase transitions and the respective scale at which they occur, we conduct the same analysis on a synthetic dataset. Our goal is to create a dataset that is the result of two different scaling regimes effective on different scales. For this purpose, we generate a spatial density distribution by recursive replication of a 2 by 2 density matrix $[2\ 0;\ 0\ 1]$. After a number of replications, we modify the density matrix, which is now $[2\ 0;\ 1\ 1]$ (see insets of Figure 5-11c), and continue the replication process to obtain a multiscale fractal. We then sample the resulting spatial density distribution with weighted points. Figure 5-11 illustrates the whole process and the obtained multifractal spectra. Since we replicate each matrix 4 times, the phase transition occurs when the linear regime extends to a length of $2^4=16$ units. Thus, the corresponding radius of a covering sphere is $R = 16\sqrt{2}/2 = 11$. The method not only detects the phase transition correctly but is also able to estimate the fractal dimensions for the two scaling regimes accurately.

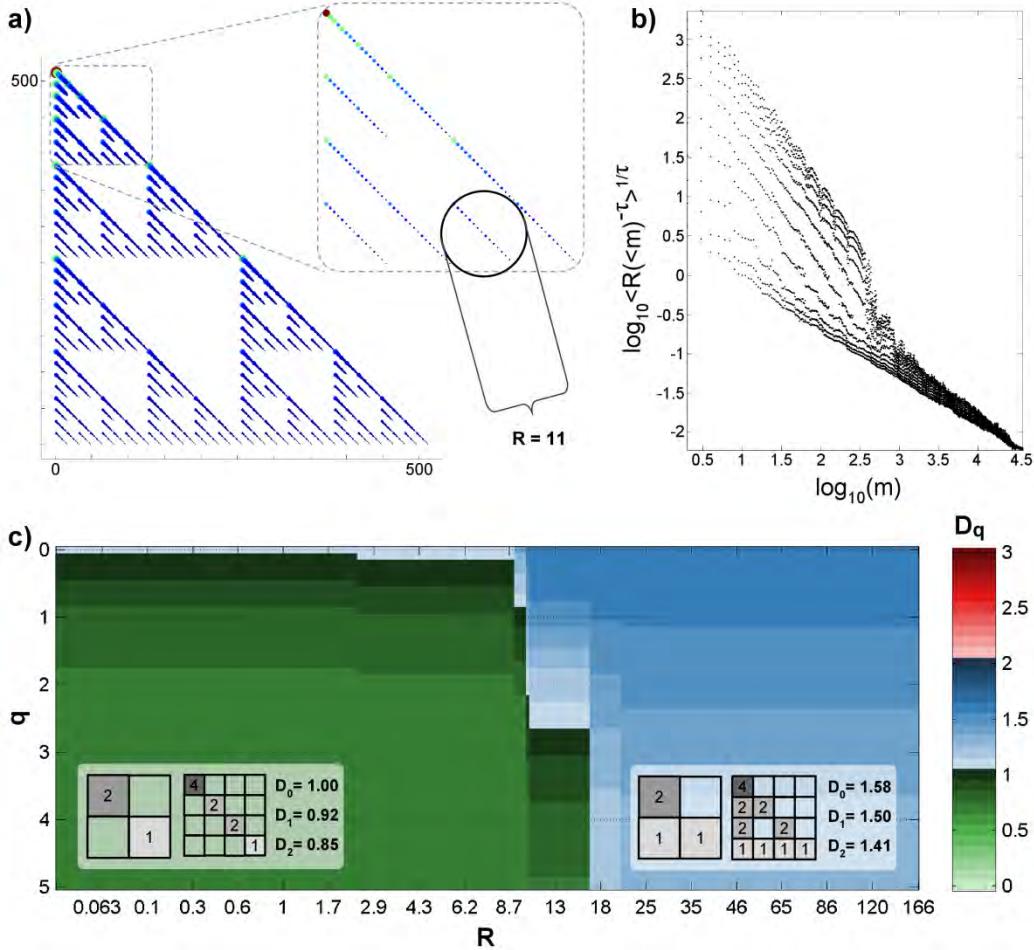


Figure 5-11 a) Spatial distribution of the synthetic distribution, inset shows a close up with the small scale linear features. b) Radii vs mass curves for different increasing τ exponents c) Continuous multifractal spectra of the distribution (similar to Figure 10); insets show the replicating density matrices of each scaling regime together with the analytically calculated D_0 , D_1 and D_2 values.

5.6 Discussion

5.6.1 Consequences for the spatial distribution of earthquake loci

The results of our analysis reveal the multifractal characteristics of hypocenter distributions, which are evidenced by different scaling regimes holding at different scales. An important question arises regarding the origin of these distinct scales. The largest of these scales (approximately 10km), is common to both the original and condensed catalog and can safely be interpreted as the typical thickness of the seismogenic crust in Southern California. Beyond this scale, seismicity becomes a 2D process, while it is a 3D one at smaller scales. Another phase transition is observed at a smaller scale $R_{S-M} \approx 1.5$ m for the original catalog. Accounting for the location uncertainties via the condensation method, this transition is offset to $R_{S-M} \approx 2.5$ m. Below this scale, the effect of the condensation process is also to decrease the strength of the clustering. This can be rationalized by the observation that the compression achieved during condensation (about 7.4% for all events with $M > 2$) is performed by assigning a zero weight to repeating events and thus by removing them (and their associated bias). The change in the location of the scaling break R_{S-M} may then be also partly a consequence of the change of slope of the moment curves at small and medium scales (see Figure 5-8). In other words, if we have a bilinear curve and change the slope of only one of the segments

the intersection point would shift. A possible physical constraint for R_{S-M} can be the width of the fault gouge zone. However, Sammis and Biegel [Sammis and Biegel, 1989] found that the particle size distribution within gouge zones is likely to be a power law with exponent 2.6, suggesting a similar fractal dimension for the set of ruptures bounding the grains and blocks. We observe a substantially smaller value $D_2=1.2$, suggesting that seismicity is indeed much more clustered. Interestingly, such a high degree of clustering seems in agreement with the reports of narrow, quasi-linear seismicity streaks along several faults in California [Rubin *et al.*, 1999; Waldhauser *et al.*, 1999; Shearer, 2002].

A different explanation for the origin of $R_{S-M}\approx 2.5$ m can be the earthquake relocation process itself. The relocation is based on a double-difference method using cross correlation of events that are initially clustered according to multiple criteria. In their paper [Hauksson *et al.*, 2012], the authors report one of these criteria to be a maximum separation distance of 2.5km. A conclusive analysis would require repeating the relocation procedure varying this arbitrary distance criterion and repeating the multifractal analysis. The transition scale may also be controlled by the initial clustering criterion such as the correlation coefficient threshold. Although such an analysis is beyond the scope of this study, we also note that double-difference methods have been reported to be highly susceptible to biases resulting from velocity structure errors [Michelini and Lomax, 2004]. These can strongly affect the shape and inner structure of the relocated clusters themselves, hence their associated scaling properties.

5.6.2 Consequences for earthquake triggering models

The spatial distribution of earthquakes plays an important role in understanding their interactions. Previous studies investigating the importance of small earthquakes in triggering have been mostly limited in reporting only the capacity (D_0) [Helmstetter, 2003] or correlation dimensions (D_2) [Helmstetter *et al.*, 2005; Marsan and Lengliné, 2008] and using these two dimensions interchangeably (which is valid only under the assumption of monofractality). We argue that the reported values certainly feature strong biases resulting from the applied methods. Moreover, from those values, they were able to draw important conclusions about the triggering properties of events with different magnitudes (and assuming that seismicity can be modeled as an Epidemic Type Aftershock-Sequence (ETAS) process). Within the ETAS formalism [Ogata, 1988], the number of aftershocks following a magnitude M event is assumed to scale as:

$$n(M) \sim 10^{\alpha M} \quad 5-10$$

where α is the productivity parameter. For example, Helmstetter (2003) estimates $\alpha=0.8$. The magnitude-frequency distribution of these aftershocks obeys the Gutenberg-Richter law:

$$P(M) \sim 10^{-bM} \quad 5-11$$

In the following, we use $b=1$ as evidenced by global and regional analyses [Kagan, 2010; Godano *et al.*, 2014; Kamer and Hiemer, 2015]. Since each individual aftershock can trigger its own aftershocks, the total number of aftershocks triggered collectively by all magnitude M events scale as

$$N(M) = n(M)P(M) \sim 10^{(\alpha-b)M}$$

5-12

Equation (5-12) implies that, if $\alpha > b$, the triggering is dominated by the largest earthquakes, while, if $\alpha < b$, then it is controlled by the smallest ones. The latter case would have serious implications for understanding earthquake interactions, and hence advances in earthquake prediction, since the Gutenberg-Richter holds up till very small magnitudes [Boettcher et al., 2009], meaning that the majority of the small events are below the detection threshold of current seismic networks [Sornette and Werner, 2005]. Measuring α from a seismic catalog is problematic not only because it involves subjective definitions of time and space windows for aftershocks, but also because of the inherent incompleteness due to missed events following large main shocks. Here, we argue that fractal analysis can be used to obtain more reliable estimates of α . To provide a link between α and the fractal dimension(s) of seismicity, we use the empirical observation that the rupture length L of a magnitude M event is given by [Kanamori and Anderson, 1975]:

$$L(M) \sim 10^{\theta M}$$

5-13

where regression analyses show that θ depends on the faulting style [Wells and Coppersmith, 1994]. We also notice that, in case of a fractal distribution, the average number of events within a domain of size L scales as:

$$n(L) \sim L^{\theta D_2}$$

5-14

Combining Equation (5-13) and (5-14) yields the scaling of the average number of aftershocks within a domain with the size of the mainshock:

$$n(M) \sim 10^{\theta M D_2}$$

5-15

so that we can identify $\alpha = \theta D_2$. Triggering properties can thus be inferred by comparing θD_2 and b .

Using a correlation integral (i.e. a fixed scale approach), Helmstetter et al.[Helmstetter et al., 2005] estimated $D_2=1.5$ (also reported by Kagan[Kagan, 2007]) and $D_2=1.74$ for two different catalogs in Southern California for $0.1 \leq r \leq 5$ m. However, they chose to use $D_2 \approx 2$, estimated for inter-event times larger than 1000 days, in order to remove the distortion of the scaling due to the triggering itself. Citing [Kanamori and Anderson, 1975], they chose $\theta=0.5$, and concluded that small earthquakes are as important as big ones for triggering, as $0.5D_2 \approx b$. We notice that, without any constraint on inter-event times, they would have obtained $0.5D_2 < b$, so that small events would be predicted to dominate the triggering. We also notice that Helmstetter[Helmstetter, 2003] uses $\theta=0.5$ while[Kanamori and Anderson, 1975] give four possible values of θ ranging from 0.33 to 1 based on theoretical derivations. Similarly Marsan and Lengliné [Marsan and Lengliné, 2008] used 6190 ≥3 earthquakes in Southern California to estimate individually $\alpha=0.6$ and $\theta=0.43$. They reported $D_2=1.17$ and reached the conclusion that small earthquakes have a greater effect on triggering.

We argue that the ETAS formulations can be extended by accounting for the multiscale multifractal characteristics presented in this study and hence provide more rigorous inferences about the earthquake interactions process. As Southern California seismicity is largely dominated by strike slip events, we suggest to use the value $\theta=0.74$. This value has been derived from 43 global strike-slip events with rupture lengths $1.3 < L < 432 \text{ km}$ [Wells and Coppersmith, 1994]. To illustrate the impact of this scaling parameter, we conduct the following exercise: We calculate the predicted rupture lengths for the 1999 Hector Mine and 2014 M6.0 West Napa earthquakes, based on the rupture length observed during the 1992 M7.3 Landers earthquake. We employ $L(M) = 85 \text{ km} \times 10^{\theta(M-7.3)}$ using the three different θ values discussed above (see Table 1). We note that small θ values will exaggerate the triggering effect of small magnitude events since they overestimate observed rupture lengths significantly as the magnitude decreases.

Observed Rupture Length		Predicted Rupture Length using 1992 M7.3 Landers (85 km*)		
		$\theta=0.43$	$\theta=0.50$	$\theta=0.74$
1999 M7.1 Hector Mine	41 km*	69.7 km	67.5 km	60.4 km
2014 M6.0 West Napa	12 km**	23.5 km	19.0 km	9.3 km

* <http://www.data.scec.org/significant/chron-index.html>

**<http://www.eqclearinghouse.org/2014-08-24-south-napa/files/2014/08/EERI-Special-Eq-Report-2014-South-Napa-versionOct19web.pdf>

Our results indicate that, for distances $r < 10 \text{ km}$, the relocated catalog of Southern California exhibits two different scaling regimes with a transition at $R_{S-M} \approx 2.5 \text{ m}$. If this scaling break is due to the parameter choices of the relocation procedure, we conclude that, using $D_2=1.9-2.0$, we get $0.74D_2 \approx 1.44 > b$, so that large events dominate triggering. On the other hand, if this scale is physical, earthquakes with magnitudes $M < M_5.7$ (approximate rupture length of $2R=5 \text{ km}$) induce a triggered seismicity with $D_2=1.2$. The inequality $0.74D_2 < b$ then suggests that at this scale the triggering is controlled by the smallest earthquakes. Yet, for larger magnitudes, the largest events dominate triggering. We anyway underline that such scaling breaks are not compatible with the definition of the ETAS model, which does not feature any transition scale, so that we should be cautious when drawing such conclusions.

The implications of our result, specifically the case of $\alpha > b$, has been previously investigated by Sornette and Helmstetter [Sornette and Helmstetter, 2002] in terms of the branching ratio:

$$n = \frac{kb}{b-\alpha} \left(\frac{1-10^{-(b-\alpha)(m_{\max}-m_0)}}{1-10^{-b(m_{\max}-m_0)}} \right) \quad 5-16$$

where k is a normalization constant of the productivity rate given in Equation (5-10), m_{\max} is the maximum earthquake magnitude (in the range of $M8-M9.5$ [Kagan, 1999b], but see improved methods of determination of m_{\max} [Pisarenko et al., 2008, 2010, 2014]) and m_0 is the minimum magnitude of an event that can trigger its own aftershocks. The authors showed that the subcritical, stationary behavior of the earthquake process (i.e. that aftershock sequences die out within a finite time length) requires that $n < 1$. Furthermore they demonstrate that the case of $\alpha > b$ in which m_{\max} is

infinite leads to explosive seismicity dynamics, in the form of stochastic finite-time singularity [Sammis and Sornette, 2002]. Such transient dynamics can actually be observed in various aftershock sequences and also instances of accelerated seismicity. For finite m_{max} , such explosive dynamics are transient and taper off before any mathematical divergence, when the largest events of the distribution are sampled.

It is important to note that the conclusions of Helmstetter and Marsan, arguing for the importance of small earthquakes, have spawned case studies aiming to quantify this claim in terms of static stress triggering (e.g [Cocco *et al.*, 2010; Meier *et al.*, 2014]). Such studies rely on the limited number of available focal mechanisms, various assumptions and large uncertainties to compute Coulomb stresses [Felzer *et al.*, 2002]. Due to these limitations, the results are often inconclusive and difficult to generalize. Here, we showed that a purely statistical and robust approach based on empirical laws can provide rigorous answers to such questions. Although our answers depend on the origin of R_{S-M} , this ambiguity can be resolved by relocating the catalog directly from the waveforms using the fully probabilistic approach NonLinLoc, which gives a more realistic representation of the location PDFs [Lomax *et al.*, 2009]. We shall then be able to conclude on the existence and properties of such a scaling break.

5.7 Conclusion

We have introduced a novel condensation method that improves the spatial information content of seismicity catalogs by accounting for the heterogeneity of the reported location qualities. We obtain significant likelihood gains in synthetic datasets perturbed with realistic location uncertainties, and expect the same to hold for natural ones. Qualitative comparison with mapped fault traces in Southern California indicates that condensation highlights active fault structures. The method also reduces the length of the catalogs significantly and allows the location uncertainty information to be taken into account in spatial analyses. Using this information and the state of the art BFM method, we have performed multifractal analyses on the last 20 years of Southern Californian seismicity. Our analysis reveals a phase transition occurring at $R_{S-M} \sim 2.5$ m, which is most likely due to the relocation procedure rather than a genuine physical process. We use the correlation dimension D_2 obtained in our analysis, together with observations on rupture length scaling with magnitude, to make inferences on earthquake triggering models. Contrary to previous studies, our results suggest that large earthquakes dominate the earthquake triggering process. We thus conclude that the limited capability of detecting small magnitude events cannot be used to argue that earthquakes are unpredictable in general.

We envision that the proposed condensation method will become an essential preprocessing tool in the field of seismicity-based fault network reconstruction, which had significant advances in the recent years[Ouillon *et al.*, 2008; Ouillon and Sornette, 2011; Kamer *et al.*, 2013b; Wang *et al.*, 2013]. These studies employ clustering methods to infer fault structures illuminated by past seismicity. The vast amount of events contained in the seismicity catalogs hinders the large-scale application of these methods because the computation time increases exponentially with the number of data points. For instance, Ouillon and Sornette [Ouillon and Sornette, 2011] used 2747 events of the 1986 Mount Lewis sequence while Wang *et al.*[Wang *et al.*, 2013] used 3013 aftershocks of the 1992 Landers event. The condensation method reduces the number of data points significantly and should thus allow for faster computations. With the help of these improvements, we were recently able to perform a clustering analysis on the Southern Californian catalog containing $\sim 500,000$

events [Kamer *et al.*, 2014]. Condensation also ingrains the information of location uncertainty into the weight of each event, providing an efficient representation of the relative location quality. Furthermore, the weight of each event can be regarded as a multiplier quantifying repeating occurrences at individual locations, which can be inferred as a representation of repeated slip. By equalizing all the individual condensed weights to 1, we can get better insights into the underlying fault structure. This should be of particular interest because individual faults often have regions with different seismicity rates that hinder a holistic clustering inference.

5.8 Acknowledgments

We would like to acknowledge Shyam Nandan for his valuable suggestions and discussions. The Matlab implementation of the catalog condensation and the BFM method for fractal analysis can be downloaded on the following URLs: <http://www.mathworks.com/matlabcentral/fileexchange/48702> and <http://www.mathworks.com/matlabcentral/fileexchange/50621>. The Southern Californian seismicity catalog used in this study can be downloaded from the following URL: <http://www.data.scec.org/research-tools/alt-2011-dd-hauksson-yang-shearer.html>

Chapter 6

Probabilistic, non-linear, absolute relocation of the 1981-2011 South Californian seismicity

by

Yavor Kamer, ETH Zürich

Edi Kissling, ETH Zürich

Guy Ouillon, Lithophyse

Didier Sornette, ETH Zürich

to be submitted to

Bulletin of the Seismological Society of America

6.1 Abstract

Earthquake catalogs are one of the most essential elements in the study of earthquake phenomena. They provide the origin time, location and magnitude information that constitutes the basis for earthquake interaction analysis, as well as physics and statistics based earthquake forecasting models. In this paper, we locate the Southern Californian seismicity using the state-of-the-art probabilistic and non-linear method NonLinLoc. We use only the P wave picks in order not to introduce the velocity model errors of the S phase, which is also harder to detect and thus less constrained. Using a subset of the best well-locatable earthquakes, we then conduct a joint inversion using the VELEST software to obtain a minimum 1D velocity model and station corrections. With this 1D velocity model and the inferred model uncertainties, we use the NonLinLoc software to obtain realistic location distributions for each event. We compare the resulting catalog with the current state-of-the-art obtained using double difference methods. Our results indicate that the absolutely located seismicity exhibits a single multifractal regime and is much less clustered compared to its relatively located counterpart, which features a distinct scaling break and a higher degree of clustering. We speculate that this discrepancy is likely to have significant effects on physical and statistical models that rely on consistent seismicity catalogs as their input data.

6.2 Introduction

Earthquake records provide pick times at which the earthquake phases of interest exceed the background noise. These arrival times are then used together with the previously determined velocity structure (Earth model) to constrain the origin time and location of the event. There has been a continuous effort for more accurate and precise locations, with more stations being deployed to increase station coverage, decrease the minimum earthquake detection threshold and improve the location quality. In parallel to these instrumental improvements, different location methods have also been proposed aiming to extract additional information from the available recordings. Observed waveform similarities between records of earthquakes within proximity of each have motivated the development of cross-correlation based relocation methods [*Waldhauser and Ellsworth, 2000; Hauksson and Shearer, 2005*]. These methods use the waveform similarity (measured by of cross-correlation coefficient) to constrain the relative location between earthquake pairs or between a master event and a set of events associated with it. The resulting catalogs often feature linear structures (streaks) of tightly grouped hypocenters that are often associated with fault geometries.

Studies have demonstrated that these so called double difference (DD) methods may result in biased relocations in the presence of velocity modeling errors [*Michelini and Lomax, 2004*]. These effects can be aggravated and modulated through the parameter choices employed in the procedure (such as the cross-correlation coefficient threshold, the minimum event separation distance between correlation pairs, etc...). In our recent multifractal analysis of the relocated Southern Californian seismicity catalog (hereafter referred to as the HYS-12 catalog) we observed the presence of a scaling break at 2.5km that coincides with the minimum distance criteria used in their DD relocation procedure [*Hauksson et al., 2012*]. The resulting earthquake locations can thus be seen as the realization of a particular parameter set, which is only a single point in a large

parameter space. The inevitable need for absolute locations catalog with uniformly consistent uncertainty estimations has motivated us to relocate the last 30 years of Southern Californian seismicity using a probabilistic non-linear method.

The paper is organized as follows: we start by developing a minimum one dimensional velocity model from the best constrained earthquakes using the VELEST routine [Kissling *et al.*, 1995]. We then employ this velocity model in the probabilistic non-linear location software NonLinLoc [Lomax *et al.*, 2000] to obtain maximum likelihood hypocenters together with sampled probability density functions (PDFs) of the location distributions. We compare this new catalog, which we name KaKiOS-15, with the HYS-02 catalog in terms of location shifts, depth distributions and multifractal spectra analysis. We finally summarize our findings and their implications for previous and future studies.

6.3 A Minimum 1D Velocity model for Southern California

It is important to note that all earthquake parameters contained in a catalog (location, time and magnitude) are jointly inferred from a limited set of seismic observations under the assumption of a previously inferred earth model. Thus, the observational and modeling errors affect all of them through different non-linear couplings. For instance, a misidentified phase arrival can bias the hypocentral location that in turn will lead to a bias in both the time and magnitude of the earthquake. Similarly, a velocity model biased towards faster seismic wave velocities will tend to cause overestimated depths. Thus, it becomes apparent that a well-constrained earth model is the key ingredient for a uniformly consistent earthquake location catalog.

There exist many methods in the literature for the inversion of a velocity structure. These methods vary in their model complexities (1D, 3D and number of layers), formulation and inversion techniques. It has been shown that simple 1D models with parsimonious number of layers combined with station corrections, which account for shallow local velocity variations, can perform as well as or even better than their complex 3D counterparts. Such models are called *minimum 1D velocity models*. Their advantage is that they are less prone to the under-sampled ray coverage and, having few free parameters, they facilitate a more efficient and robust global optimization. In this paper, we used the VELEST inversion tool [Kissling *et al.*, 1995] to obtain a minimum 1D model for Southern California together with the corresponding station corrections. VELEST employs an iterative approach by jointly adjusting the layer velocities, earthquake times and locations, and station corrections so as to minimize the root mean squared (RMS) residual between the observed and modeled arrival time.

6.3.1 Data selection

It is imperative to use a large number of earthquakes that are well constrained by multiple observations. Otherwise, the inversion procedure will become unstable as parameters compensate for each other and the obtained minima will not be stable when posed against a different dataset.

For this purpose, we obtained all individual waveforms recorded between 1981-2011 from the Southern California Earthquake Data Center (SCEDC). We decided to use only the P picks in order to avoid introducing velocity model errors of the S phase, which is harder to detect and thus less constrained. We considered all recordings within 100km radius of the routine SCEDC epicenters that yields a total of ~10 million waveforms. Figure 6-1 shows the spatial distribution of all events based on number of P picks (N_{obs}) and largest azimuthal gap ($AzGap$). Earthquakes

recorded at the boundaries of the network tend to have fewer P picks and hence larger azimuthal gaps. This implies that if the selection is done solely on these two criteria there would be a selection bias towards the central regions; hence the obtained velocity model would be biased towards these local structures.

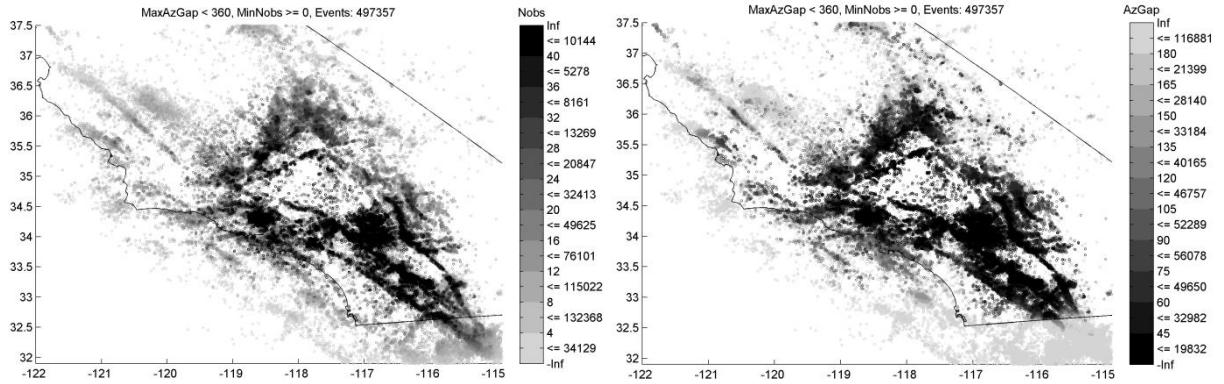


Figure 6-1(Left) Routine epicenter locations colored according to the number of P observations within 100 km of their epicenters. (Right) Colored according to the azimuthal gap

Since we intend the resulting minimum 1D velocity model and station corrections to be used for the whole region we aimed at obtaining a good areal and station coverage without compromising the well-locatability of the selected events. For this purpose, we grid the region into fixed size cells and consider the top well-locatable events (no worse than $N_{obs} > 12$ $AzGap < 150$) within each one. To optimize coverage, we chose the best event in each 6.3km cell for the training set, giving us a total of ~2500 events with ~73,000 P picks (See Figure 6-2). We also obtained the 2nd-6th best events in each cell. These were shuffled and then used to construct four different validation sets with ~2000 events each.

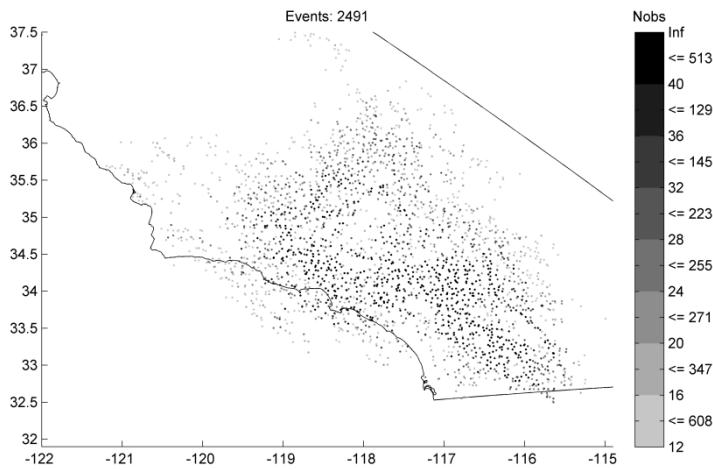


Figure 6-2 Training set of events selected as the most well-locatable events within a 6.3km cell grids

6.3.2 Inversion results and validation

We begin the inversion procedure by selecting a reference station that provides coupling between the minimum 1D model and the station corrections. To obtain station corrections that are representative of the local velocity structure, the reference station should have numerous observations and be situated in the middle of the network. We chose the station CI XTL (Crystal

Lake; 34.29° , -117.86°), which is on a Mesozoic granite site, as a reference station since it complies with these criteria. Next, we investigate different depth parameterizations and initial velocity values by performing multiple inversions initialized with different published 1D models [Hadley and Kanamori, 1977; Dreger and Helmberger, 1990]. Figure 6-3 shows the progression and results of 4 such inversions.

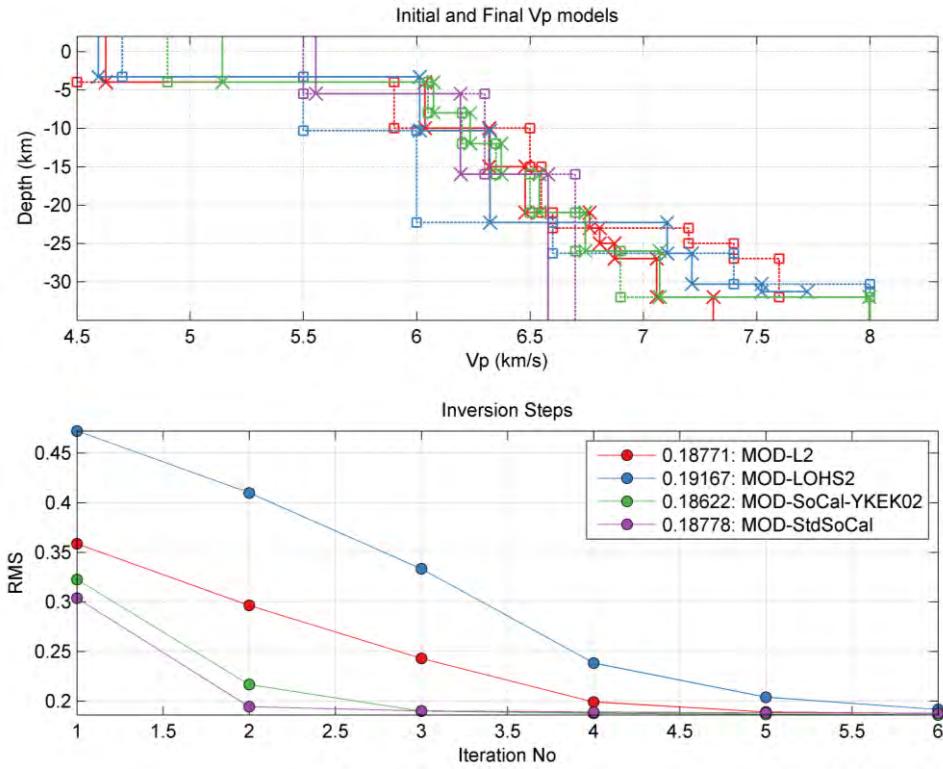


Figure 6-3 Top: Initial (squares) and final (crosses) values of V_p velocities for different models. Bottom: Progression of the RMS with increasing iteration numbers. Final RMS values are shown in the legend.

The results in Figure 6-3 indicate a fairly smooth increase of velocity in the mid depth section. Following the general rules of [Kissling *et al.*, 1995] and after several trials with different parameterizations, initial values based on previous convergences and investigating the hypocentral adjustments ,we converged to a stable minimum 1D model with the lowest RMS value. To test the robustness of the obtained minimum 1D model, we initiated four inversions with the four different validation sets that we generated earlier. All the models were initialized at the same velocity layer. The results in Figure 6-4 show that the obtained velocity model is stable throughout the depth ranges of 2-23km. The variations below and above this range are to be expected due to the subvertical and subhorizontal ray path coverage.

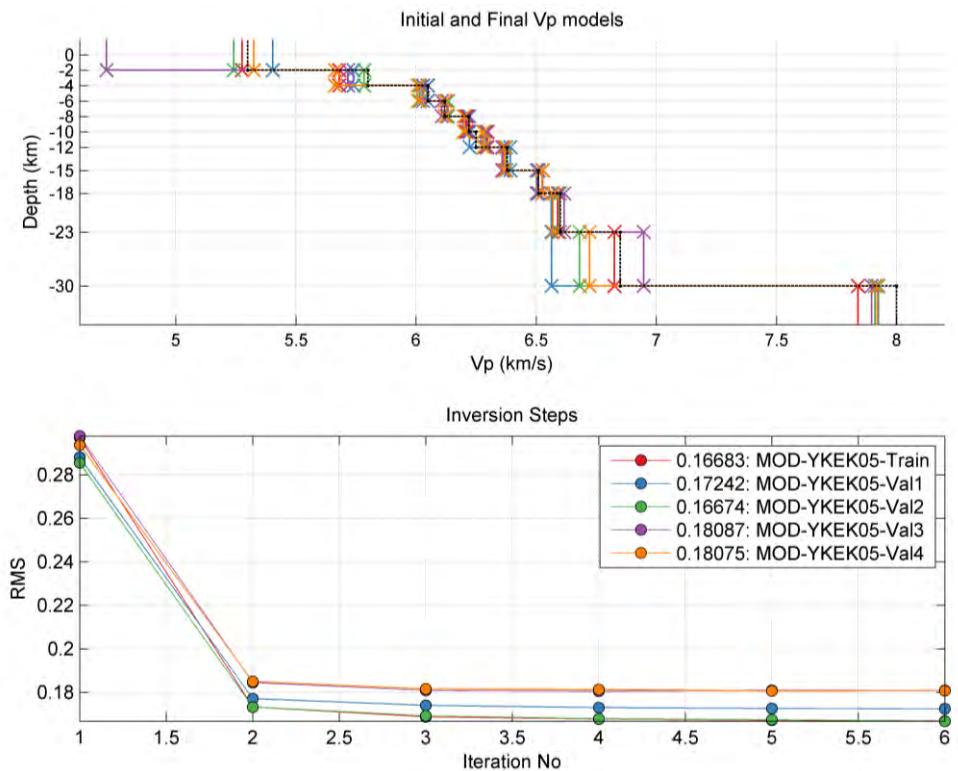


Figure 6-4 Same as Figure 6-3 obtained for the YKEK05 initial model using the four validation sets

Another important validation is the correspondence between the obtained station corrections and the surface geology. In Figure 6-5 we show the station corrections individually constrained by more than 100 observations and superimposed onto the surface geology. Since the reference station is on Mesozoic granite, stations on softer (slower) structures have positive corrections while the ones on harder (faster) structures have negative corrections with respect to the reference. The first order correspondence between these values and the underlying surface geology should be regarded as a testament to the physical consistency of the obtained model.

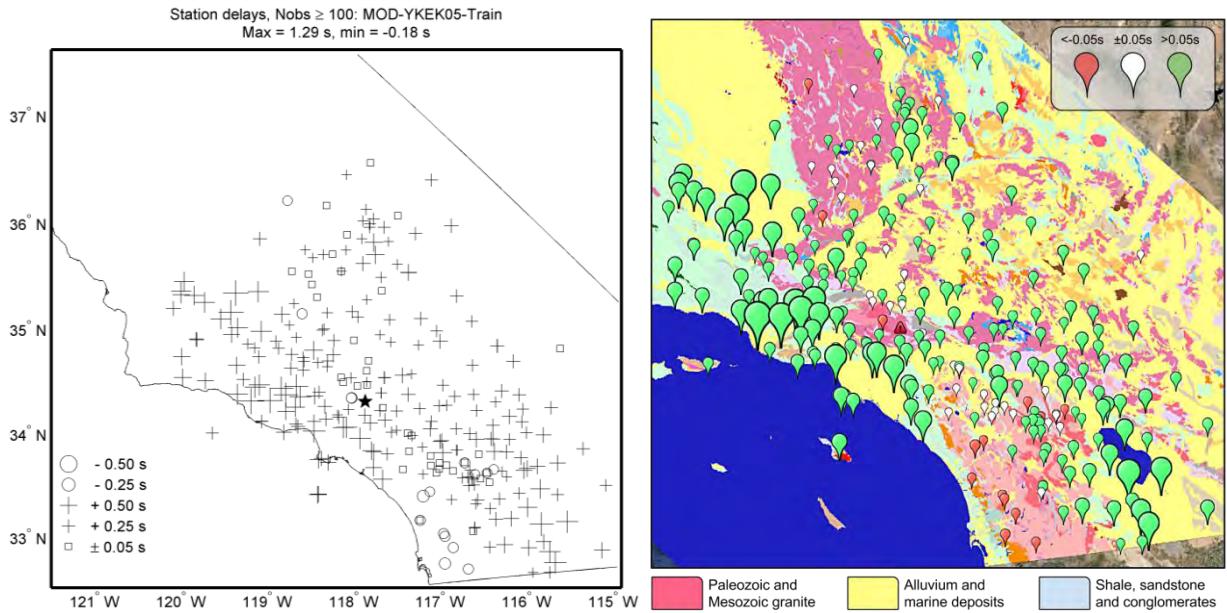


Figure 6-5 Left: Station corrections shown as O (negative) and + (positive) signs scaled according to their amplitude.
Right: The same station corrections superimposed over the geological map of California.

6.4 Hypocenter relocation using NonLinLoc

Having obtained and validated our minimum 1D model, we proceed to employ it in the non-linear, probabilistic and absolute earthquake location routine NonLinLoc. For this purpose, we consider all the events with at least 4 P phase picks within 100 km radius of their routinely determined epicenters. The number of observations, their spatial distribution and their uncertainty coupled through the velocity model and its modeling error determine the probability density function of the absolute location (for illustrative figures see [Lomax *et al.*, 2009]). The velocity modeling error was estimated from the station residuals averaged over the validation sets. For information regarding the observational error (i.e the picking time errors) we investigated the quality classes provided in the earthquake phases files. The quality class in these files varies in the range of [0-1] with a precision of 0.1. However we found that usage of this class is not consistent through time; with different sets of three to four values being used throughout the catalog time span. For this reason, we considered instead the onset type (impulsive/emergent) as a binary indicator of picking time quality. We used standard deviations of 0.05s and 0.1s for impulsive and emergent P picks respectively. To test if the pick quality provides an additional constrain for the earthquake locations, we conducted a controlled test using VELEST. We conducted three separate inversions using the same training set and the minimum 1D velocity model as initial conditions. The tests differed from each other only in terms of the provided quality information: a) no quality information, all observations weighted equally; b) quality classes assigned based on impulsive (0) and emergent (1) onsets; c) randomly assigned classes of quality 0 and 1. The results of these tests are shown in Figure 6-6. The lowest RMS value was obtained in case c) which indicates that onset type is a reliable first order indicator of observation uncertainty. These results also suggest that the pick quality has little effect on the final velocity model which can be attributed to the well-locatability of the events used in the inversion.

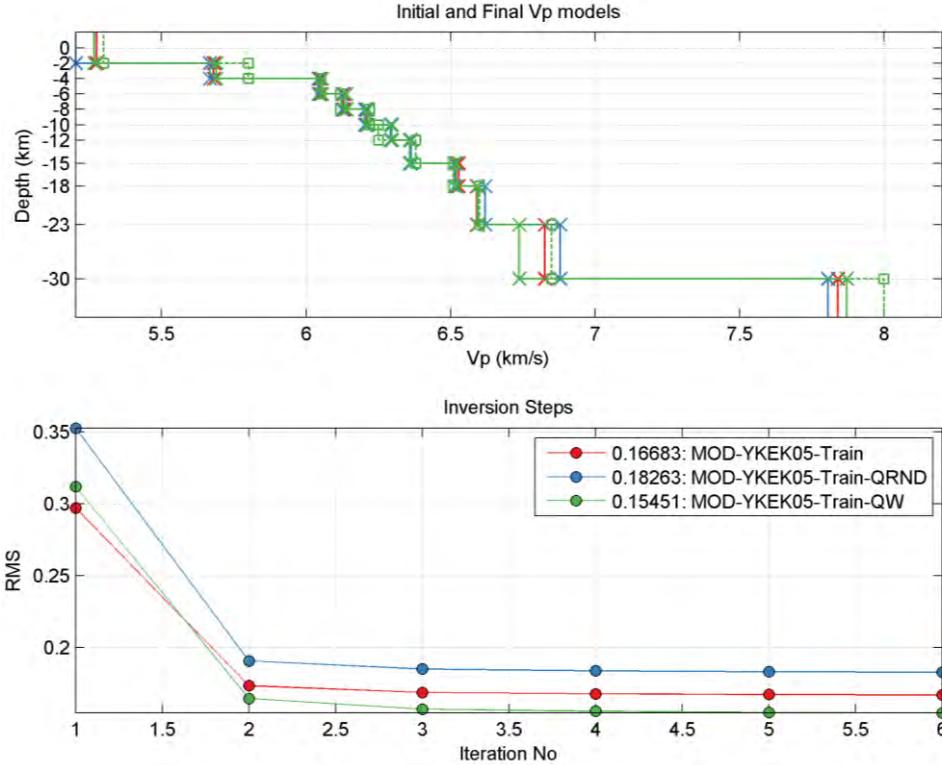


Figure 6-6 Same as Figure 6-3. Red: No pick quality information; blue: randomly assigned pick quality; green: qualities assigned according to onset type (impulsive=0, emergent=1)

The global optimization search is conducted using the Oct-tree approach in a 100x100x40 km volume centered on the routinely reported epicenter. We used the Equal Differential Time (EDT) norm as the likelihood function because it has been demonstrated to be more robust in the presence of outlier (such as misidentified picks or timing errors) and provide more realistic location PDFs [Lomax *et al.*, 2009]. The entire computations were performed in parallel on the high performance computing cluster Brutus of ETH Zurich. The output of these computations is a summary file containing the earthquake parameters, stations residuals and particularly the sampled location PDF (using 1000 points in our case).

We believe that one of the main issues hindering a wider spread and acceptance of NonLinLoc is the general reluctance of the seismological community to deal with complex non-linear uncertainties. For this purpose, in this study we represent the complex location PDFs as mixtures of Gaussians, i.e. Gaussian Mixture Models (GMM). In this formulation, the spatial distribution of PDF sample points is represented as the combination of a finite number of three dimensional Gaussian kernels, each with an individual mean vector, covariance matrix and responsibility weights which all sum up to 1 [Bishop, 2007]. To find the optimal model complexity (i.e. the number of GMM components) we employ the Bayesian Information Criteria (BIC) which penalizes the likelihood of the model according to the number of free parameters it contains [Schwarz, 1978]. For each location PDF we initiate several GMMs with different initial conditions and employ the expectation maximization algorithm [Bishop, 2007] to converge to a final model. We keep the model with the best BIC score, increment the number of components and restart the modeling. The process continues until the current BIC score performs worse than the last two steps (i.e until a stable minima is reached). Using this scheme, we find that only 19% of the events are modeled by a single Gaussian, while on average of 3.24 components are needed to model the

location PDFs of the rest of the events. Figure 6-7 shows an example of an event location PDF and its corresponding GMM.

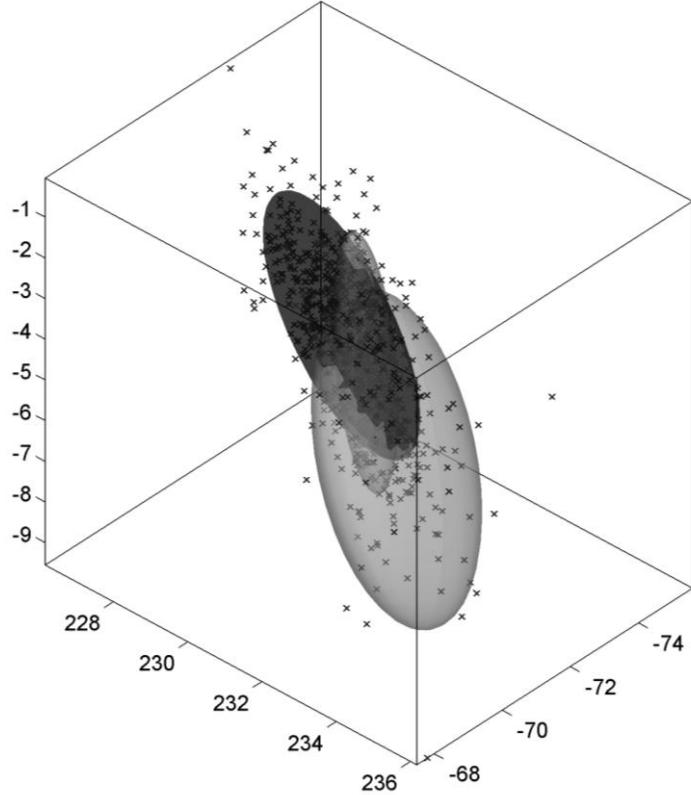


Figure 6-7 A NonLinLoc location PDF sampled with 1000 points shown as crosses superimposed on the optimal BIC Gaussian mixture model with components weights 0.59, 0.31 and 0.09 (darker shading for larger weights). Gaussian ellipsoids are plotted at $\pm 1\sigma$ level.

6.5 Comparison of the KaKiOS-15 and HYS-12 catalogs

6.5.1 Vertical and Horizontal shifts

To ensure the consistency of the comparisons, we limit the following analysis only to events contained in both catalogs. From a total of 502,962 events contained in the HYS-12 catalog we were able to match 476,157 to the KaKiOS-15 catalog based on their unique event IDs. The reduced events number in the KaKiOS-15 catalog is due to the more stringent location criteria that we employ (at least four P picks within 100km). Since in the KaKiOS-15 catalog each hypocenter is represented by varying number of Gaussians, we use the maximum likelihood (ML) location for the following comparison (i.e. the mean location of the Gaussian component with the largest weight).

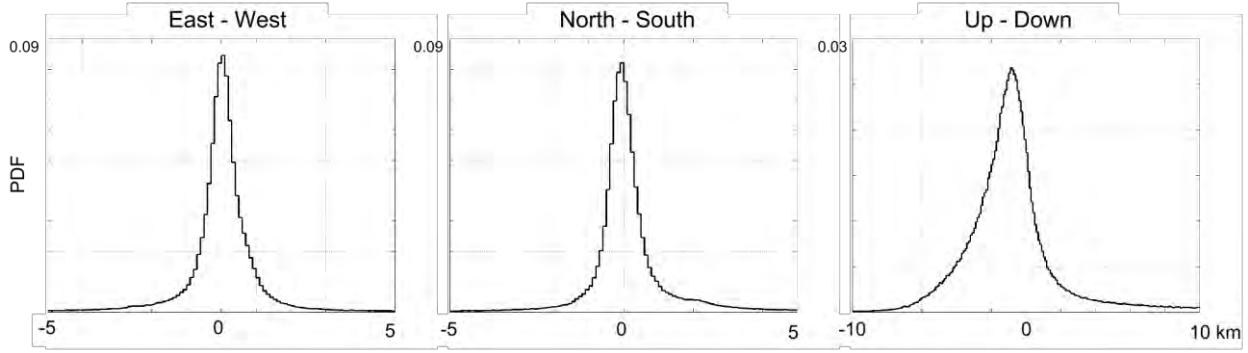


Figure 6-8 Horizontal and vertical shifts PDFs between the HYS-12 and KaKiOS-15 catalogs, calculated with respect to the later.

In Figure 6-8 we show the distribution of the horizontal and vertical displacements between the two catalogs. The lateral shifts do not show first order systematic effects, while the vertical shifts exhibit a systematic pattern indicating that HYS-12 catalog is deeper with respect to KaKiOS. This effect can be attributed to the difference in the velocity models. Figure 6-9 shows the depth distribution of the two catalogs. Although several studies assert that the DD methods reduce the effect of the velocity modeling errors, here we observe a direct dependence between the hypocentral depth distribution and the assumed velocity model. This dependence can be inferred in terms of the velocity model parameterization. In our VELEST inversions, we observed a rather smoothed velocity gradient within the 2-23km range (see Figure 6-4), whereas in the standard South California model this transition is expressed by two steps located respectively at 5.5km and 16 km depth. This leads to a strong bias in the depth distributions as hypocenters tend to avoid these abrupt steps, leading to the bimodal distribution observed in Figure 6-9. Although far less prominent, a similar effect can also be observed in the KaKiOS-15 catalog at a depth of ~2km.

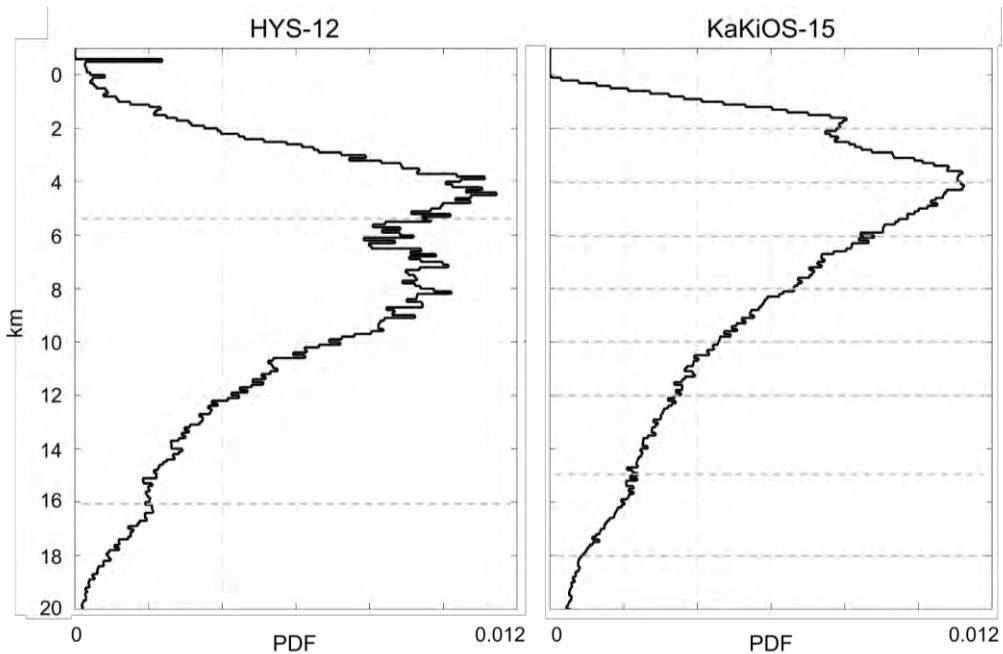


Figure 6-9 Hypocentral depth distribution of the HYS-12 and KaKiOS-15 catalogs. Gray dashed lines indicate the velocity model layer boundaries.

6.5.2 Multifractal analysis of the two catalogs

The comparison presented in the previous section provides a first order glimpse into the differences between the two catalogs relative to each other. However it does not reveal the statistical properties (e.g. degree of clustering, inter-event distance scaling) of these spatial distributions. For a more informative comparison, we employ the recently introduced multifractal analysis method of [Kamer et al., 2013a, 2015] to better assess the differences in the clustering properties of the two catalogs. The spatial scaling relationships in a catalog affect any analysis based on inter-event distances (e.g. declustering, static/dynamic stress transfer, pattern recognition etc...). The multifractal analysis consist in covering the spatial distributions with spheres containing the same number of points (mass) and then investigating how the average radii of these spheres scales as the mass is gradually increased. The average radii is calculated for different moments and thus the slope of each radii-mass curve corresponds to the different generalized fractal dimensions D_q , where large and small q s represent the scaling in dense and sparse parts of the distribution. For detailed information about the multifractal formalism and the numerical implementation, the reader is referred to Kamer et al. (2015).

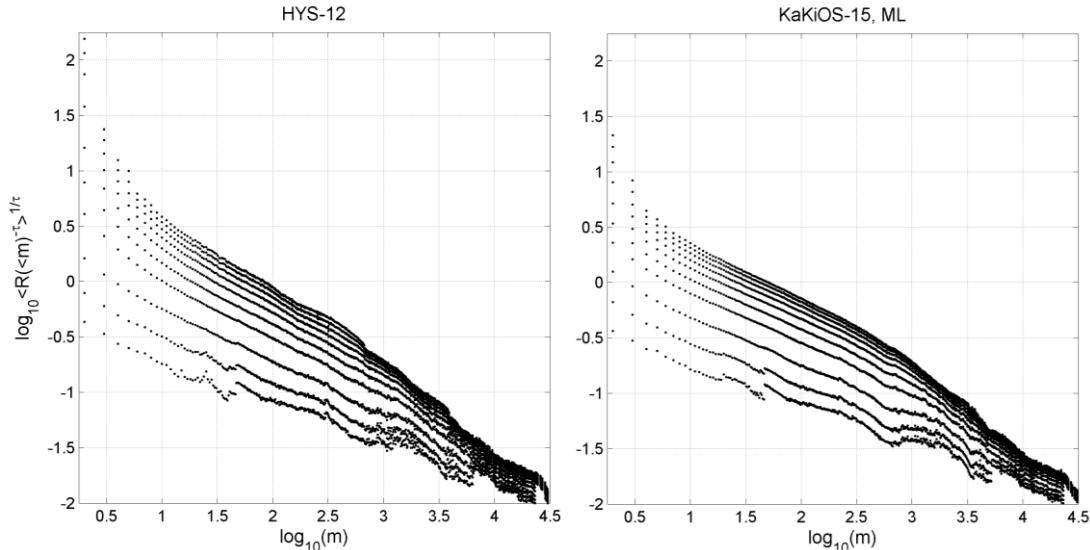


Figure 6-10 Averaged radii versus fixed-mass curves for $\tau = [-3, 6]$ averaged over 100 measurements. Left panel: the HYS-12 catalog; right panel: the KaKiOS-15 catalog.

We consider only events with magnitudes $M \geq 2$ in order to reduce the effect of variations in the regional detection threshold. This leaves us with 98,050 events in both catalogs. In Figure 6-10, we show the radii versus mass for the two catalogs, using the ML locations for the KaKiOS-15. We observe that there are significant variations in the scaling regimes of the two catalogs. To facilitate the visual representation, Figure 6-11 shows a plot of D_q in the range of $q=[0,5]$ for an increasing radius.

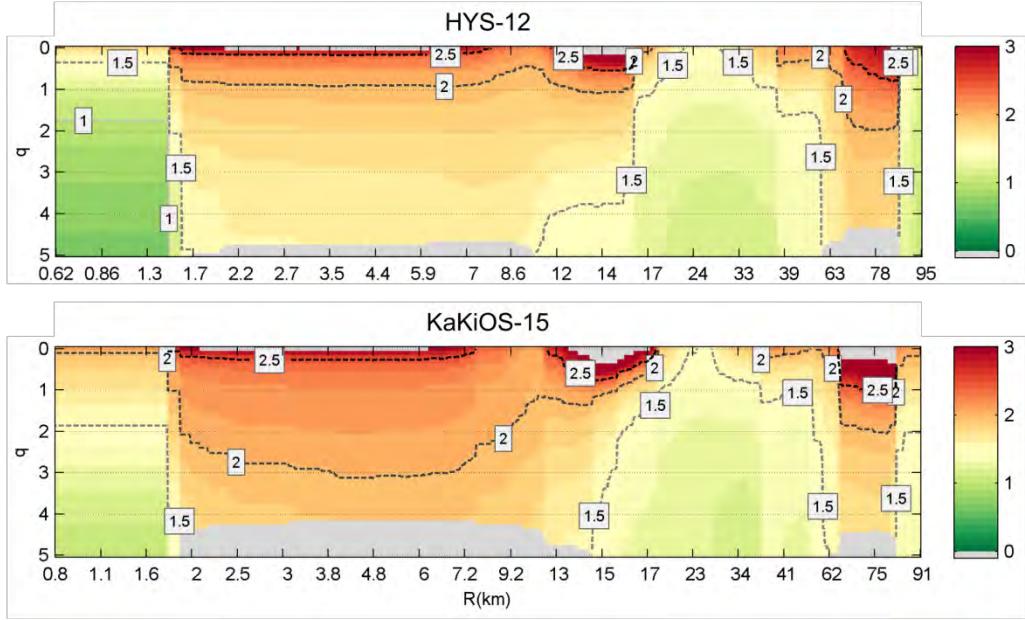


Figure 6-11 Continuous multifractal spectra of the HYS-12 (top) and KaKiOS-15 (bottom) catalogs. The horizontal axis indicates the increasing radius; the vertical axis represents the q value and individual colors represent D_q within the range [0-3]

For the sake of clarity, we shall investigate the clustering properties in terms of D_2 and the multifractality in terms of $\Delta D = D_1 - D_4$. We observe that overall HYS is significantly more clustered than KaKiOS which indicates that the effects of the cross-correlation relocation procedure are not limited to small scales. We also observe a distinct scaling break at ~ 1.5 km for HYS and at ~ 1.8 km for KaKiOS. The break is more abrupt for the HYS where the change is from $D_2 = 0.95$ to $D_2 = 1.78$, compared to $D_2 = 1.45$ to $D_2 = 2.05$ for KaKiOS. There is also a gradual dimension reduction for both catalogs beginning at ~ 10 km for HYS and at ~ 8 km for KaKiOS. This can be attributed to the seismogenic zone thickness (see Kamer et al. (2015)). In terms of multifractality, the main feature in HYS-12 ($1.5 \text{ km} < r < 15 \text{ km}$) has a larger $\Delta D = 0.34$ compared to KaKiOS-15 ($1.8 \text{ km} < r < 13 \text{ km}$) where $\Delta D = 0.25$. This indicates that the differences in densities between individual localities is less pronounced in the KaKiOS-15 catalog (i.e. relatively more homogeneous).

To recreate the dimension reduction effect and to investigate the possible origin of the small scale behaviour in both catalogs, we conduct the following synthetic test: we generate 80,000; 160,000 and 320,000 uniform random points inside a volume with dimensions $400 \times 400 \times 16$ km, which is compatible with the spatial extent of the Southern California. As with the seismicity catalogs, we conduct 100 repetitions of MFA and average the obtained results (see Figure 6-12).

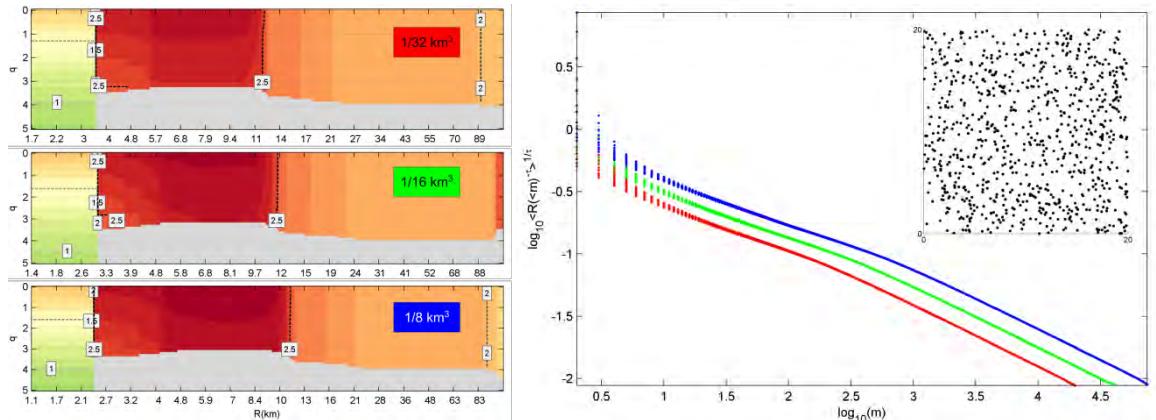


Figure 6-12 (Left) Multifractal spectra of three different uniform synthetic distributions with dimensions $400 \times 400 \times 16 \text{ km}$. Insets show the point densities with color correspondence in the right panel. (Right) The radius versus mass curves plotted with their respective colors. Inset shows a $20 \times 20 \text{ km}$ sample region of densest catalog.

Interestingly, the synthetic distributions exhibit a similar small scale feature of increased clustering and decreasing D_q as a function of q . The spatial extent of this feature is governed by the density of the distributions. As the density increases from $1/32 \text{ km}^3$ to $1/8 \text{ km}^3$ the average smallest measured radius decreases from 1.7 km to 1.2 km while the extent of the high clustering is reduced from 3.5 km to 2.4 km . We interpret this persistent feature as being the result of the seemingly coherent alignments created by the uniform random distribution that can be highly clustered at small scales. This can be regarded as a correlated spatial noise effect. We thus interpret the small scale feature in both catalogs as possibly being due to the location error that tends to create artificial structures at small scales. The effect is more pronounced in the HYS-12 catalog where the cross-correlation approach tends to amplify the random clustering by collapsing the hypocenters on streak-like features.

To further investigate the effect of the location uncertainties on the MFA, we repeat the analysis on the KaKiOS-15 catalog; but this time, instead of using the single ML hypocenter location for each earthquake, we use the mean locations of all Gaussian components used in the GMM modelling of its non-linear location PDF. This increases the number of points from 98,050 to 240,070. Each such mean location represents a mass that is equal to the relative weight of its Gaussian kernel, hence the total mass remains 98,050. Naturally, increasing the number of points while preserving the overall mass should lead to a decrease in the overall clustering (i.e increase in D_q). This is also what we observe in the results shown in Figure 6-13. The interesting result is that by considering the location uncertainty information, we observe a more or less single scaling regime that extends till the width of the seismogenic zone thickness. The D_q values in this analysis should be expected to be higher than the true values associated with the real seismicity because the multi-modal representation of the hypocentral PDF does not have a physical origin, but is due to our observational limitations. However, by taking these limitations into consideration, we are able to conclude that the spatial distribution of earthquakes in South California is likely to follow a single scaling law with $D_2 = 2.05$ and $\Delta D = 0.25$ (see Figure 6-11).

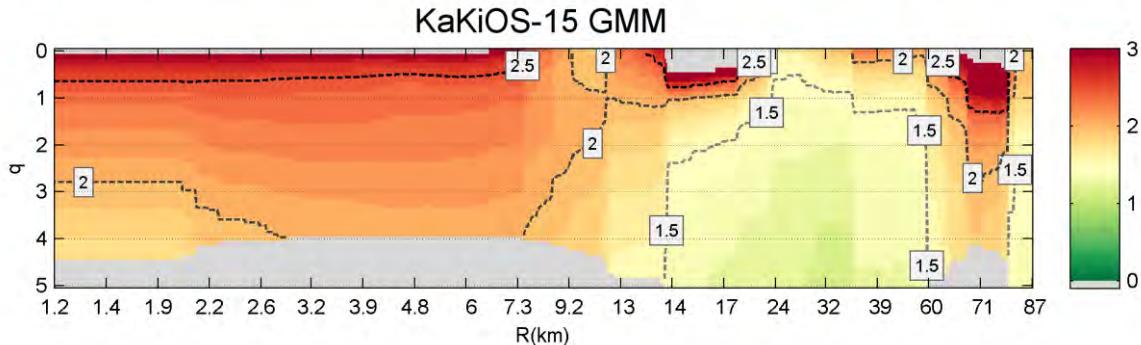


Figure 6-13 Continuous multifractal spectra of the KaKiOS-15 catalog, using all the individual Gaussian components of each earthquake.

6.6 Conclusion

In this paper, we developed an absolute seismicity catalog spanning three decades of Southern California seismicity. The main advancement of the KaKiOS-15 catalog is that it features uniformly consistent location uncertainty information. This achievement can be seen as the net result of employing the fully probabilistic location algorithm NonLinLoc and the development of a minimum 1D model using the most well-locatable portion of the dataset, hence allowing for the constraining and quantification of the modeling errors. Compared to the HYS-12 catalog that has been obtained using a DD methods, the KaKiOS-15 catalog is significantly less clustered, not only at small scales (expected to be the primary target of the cross-correlation based similarity measures) but also at relatively large scales too. Throughout this paper, we showed evidences that in the case of Southern California, the DD method has led to the emergence of highly clustered small scale features that can be regarded as an artifact of the method. These features are much less prevalent in the KaKiOS-15 catalog and they become insignificant when the analysis accounts for the non-linear location uncertainties. The discrepancy between the results of these two methods (relative and absolute) will undoubtedly have significant effect on the consequent efforts regarding physical and statistical modeling of earthquakes. For instance, in Coulomb stress analysis, where the stress amplitude decreases as r^{-3} , an inflated clustering will lead to higher stresses for pairs of events that can in reality be further away. Similarly, de-clustering methods employed in many earthquake forecast methods calculate a probability of being an independent event based on a space-time distance. Thus, such artificial clustering can lead to an underestimation of the true background seismicity rate.

In this respect, we expect the KaKiOS-15 catalog to reduce the amount of bias in such studies. Nevertheless, there is certainly room for further improvements. The recent developments in automated pickers [*Di Stefano et al.*, 2006; *Lomax et al.*, 2012] can allow us to re-pick the waveforms recordings of all events in a systematic manner, providing more accurate pick times with more consistent uncertainty quantifications. Such an approach may complement NonLinLoc's EDT norm in reducing the effect of misidentified picks or similar blunders. The training of such an automated algorithm might require considerable effort; however, the minimum 1D model and the well-locatable events can be readily employed in VELEST to provide a robust benchmark for the quality-weighted observation residuals

Chapter 7

Fault Network Reconstruction using Agglomerative Clustering: Applications to South Californian Seismicity

by

Yavor Kamer, ETH Zürich

Guy Ouillon, Lithophyse

Didier Sornette, ETH Zürich

to be submitted to

Journal of Geophysical Research: Solid Earth

7.1 Abstract

We present applications of a new clustering method for fault network reconstruction based on the spatial distribution of seismicity. Current methods start from the simplest large scale models and gradually increase the complexity trying to explain the small scales, whereas the method introduced in this paper uses a bottom-up approach, by an initial sampling of the small scales and then reducing the complexity by optimal local mergers. We describe the implementation of the method through illustrative synthetic examples. We then apply the method to the absolute, non-linear KaKiOS-15 catalog, which consists of three decades of South Californian seismicity. To reduce data size and increase computation efficiency the new approach builds upon the recently introduced catalog condensation method that exploits the location uncertainty associated with each event. We validate the obtained fault network through a retrospective spatial forecast test and discuss possible improvements for future studies.

The performance of the introduced model attests to the importance of the location uncertainty information, which is a crucial input for the large scale application of the method. We envision that the results of this study could be used to construct improved models for temporal evolution of seismicity.

7.2 Introduction

The continuing advances in instrumentation and the improved seismic network coverage is decreasing earthquake detection thresholds and increasing the number of recorded events. Recent studies suggest that the Gutenberg-Richter law might hold down to very small magnitudes corresponding to molecular dislocations [Boettcher *et al.*, 2009]. This implies that there is practically no upper limit on the amount of seismicity we can expect to record as our instrumentation capabilities continue to improve. Although considerable funding and efforts are being channeled into recording seismicity, when we look at the uses of the end product (i.e. seismic catalogs) we often see that the vast majority of the data (i.e. events with small magnitudes) are omitted. For instance, probabilistic seismic hazard studies rely on catalogs with large durations and increases the minimum magnitude that can be considered due to the high completeness magnitude levels of the past. Similarly, earthquake forecasting models are based on the complete part of the catalogs. For instance, [Helmstetter *et al.*, 2007] use only $M > 2$ events, which corresponds to only $\sim 30\%$ of the recorded seismicity. The forecasting skills of the current state-of-the-art models can well be suffering not only from our limited physical understanding of earthquakes, but also from this data starvation.

In this conjecture, fault network reconstruction can be regarded as an effort to tap into this seemingly neglected but vast data source, and extract information in the form of fault patterns. We are motivated by the ubiquitous observations that large earthquakes are followed by aftershocks that sample the main rupturing faults, and conversely that these faults become the focal structures of following large earthquakes. In other words, there is a relentless cycle; earthquakes occur on faults that grow by accumulating earthquakes, which nucleate on these faults... By using each earthquake, no matter how big or small, as a spark in the dark we aim to illuminate and reconstruct the underlying fault network. The emerging structure should allow us to better forecast the spatial distribution of future seismicity and also to investigate possible interactions between its constituents.

The paper is structured as follows. First, we give an overview of recent developments in the field of fault network reconstruction and spatial modeling of seismicity. In Section 2 we describe our new clustering method and demonstrate its performance on various synthetic point distribution. In Section 3 we apply the method to the recently relocated KaKiOS-15 catalog and discuss the obtained fault networks. In Section 4 we perform a prospective forecasting test using the seismicity of last four years, which is not included in the KaKiOS-15 catalog. In the final section we present a sample temporal analysis application using the reconstructed fault network and conclude with outlook on future developments.

7.3 The agglomerative clustering method

7.3.1 Recent developments in fault reconstruction

Fault network reconstruction based on seismicity catalogs was introduced by [Ouillon *et al.*, 2008] where they develop a dynamical clustering method based on fitting the point distribution with a plane that was then iteratively split into an increasing number of subplanes to provide better fits. They use the overall location uncertainty as a lower bound on the fit residuals to avoid over fitting. [Ouillon and Sornette, 2011] then extended this method by incorporating the pattern recognition technique of mixture modeling [Bishop, 2007] using 3D Gaussian kernels. Notable improvements in this method are the introduction of an independent validation set used to constrain the number of kernels and diagnostic based on nearest-neighbors tetrahedra volumes aiming to eliminate singular clusters. Recently [Wang *et al.*, 2013] made further improvements by accounting for the individual location uncertainties of the events. While our method is inspired by these studies, and in several aspects builds upon their findings, we also notice an inherent deficiency of the iterative splitting approach that is common in all the previously mentioned methods. This can be observed when an additional plane (or kernel), introduced by splitting, fails to converge to the local points and is instead attracted to the regions of high variance (see Figure 7-1 for an illustration).

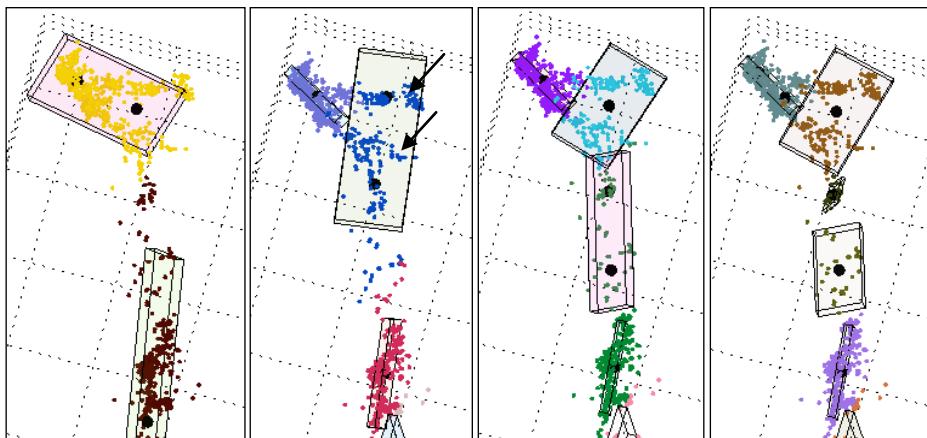


Figure 7-1 Iterative splits on the 1992 Landers aftershock data. Points with different colors represent seismicity associated with each plane. Black dots show the center points of the planes resulting from the next split. Notice how in the 2nd step the planes fail to converge to the local branches (shown with arrows).

This deficiency has motivated us to introduce an important conceptual change. Instead of starting with the simplest model (i.e a single plane or kernel) and increasing the complexity progressively by iterative splits, we propose just the opposite: start at the highest possible

complexity level (as many kernels as possible) and proceed to a simpler structure by iterative merging of the individual kernels. In this respect our new method is a “bottom-up” approach, while the ones previous are “top-down” approaches.

7.3.2 Method description

The method implementation follows the basic principles of agglomerative clustering with additional improvements to suit the specific characteristics of seismicity data. We start with the most detailed structure, i.e. the model with the highest complexity, and then employ penalized likelihood based information criterion to determine the best merger, i.e. simplification. The process continues till there is no merger that can be considered as improvement in terms of the information criterion.

For a given dataset with N points, we first construct an agglomerative hierarchical cluster (AHC) tree based on Ward’s minimum variance lin a ge method [Ward, 1963]. Such a tree starts out with a cluster for each data-point (zero variance) and then branches into a decreasing number of clusters following the minimum distance D_w criterion given in Equation (7-1). In this equation, C_{ij} is the cluster formed by merging clusters C_i and C_j , x represents the dataset and r is the centroid of each cluster. Hence in this formulation two clusters are favorable to merger if the sum of squares does not increase too much after they are merged. This merging procedure continues till there is a single cluster (maximum variance). Thus the constructed AHC can be used to create a clustering with any number of clusters from N to 1.

$$D_w(C_i, C_j) = \sum_{x \in C_{ij}} (x - r_{ij})^2 - \sum_{x \in C_i} (x - r_i)^2 - \sum_{x \in C_j} (x - r_j)^2 \quad 7-1$$

Notice that the clusters formed at any branching level of the AHC will be more or less isotropic due to the sum of squared errors criteria in Equation (7-1). Such a simplistic formulation would not be relevant for faults, which are known to be highly anisotropic [Ouillon et al., 1996], thus we model each cluster with a three dimensional Gaussian kernel expressed by its mean vector and covariance matrix. This motivated decision gives rise to the problem of singular kernels, with rank deficient covariance matrices. Such instances are problematic because the likelihood data-points associated with such kernels become degenerate and hinders the calculation of the objective function. Our intuitive solution is to discard such clusters all together. Thus we add an additional constrain to the general method of Ward, implying that that number of clusters is constrained by the spatial coherence of the dataset. Hence, for any data set we can estimate the so called holding capacity (i.e maximum number of undiscarded kernels) for a given coherence threshold. Since our aim is to detect 3D structures, the most non-arbitrary criterion is to discard all clusters with less than 4 points. Essentially, for any number of clusters we use the AHC tree at this level as the initial configuration. We discard all clusters with less than 4 points and obtain the final number of clusters. We continue the search for the maximum final clusters by incrementing the number of initial clusters. Figure 7-2 shows the estimated holding capacity under this criterion for a synthetic dataset in the presence and absence of uniform background noise. We call this procedure *atomization* because it aims in finding the smallest coherent elements, *protoclusters*, of a complex structure. We shall use this synthetic dataset for applications presented in this section.

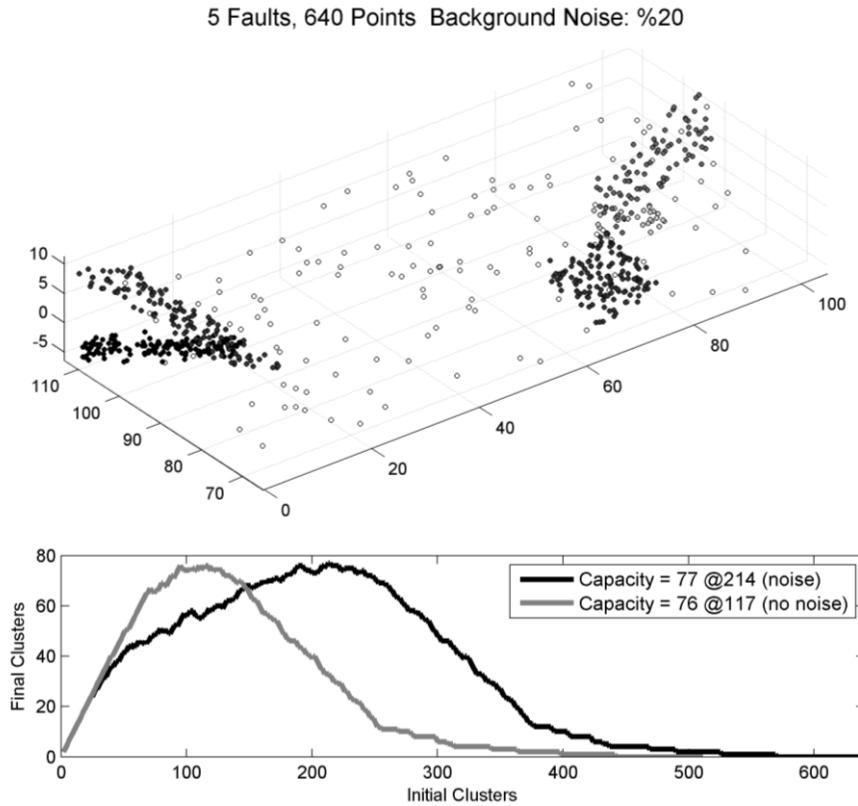


Figure 7-2 Top: Synthetic fault network created by uniform sampling of 5 faults, each shown with a different shaded according to its total number of points. Empty dots represent the uniform random background. Bottom: The number of initial and final (undiscarded) cluster for the case with and without background noise.

Since a holistic approach necessitates accounting for all observed data, rather than removing the data-points associated with the discarded clusters, they are assigned to a uniform background kernel that encloses the whole dataset. The density of this kernel is simply the number of its events divided by its volume (see Figure 7-3). The Gaussian kernels together with the uniform background kernel represent a mixture model where each kernel has a contributing weight proportional to the number of points that are most likely to be associated with it [Bishop, 2007]. This representation facilitates the calculate of an overall likelihood and allows us to use the Bayesian Information Criteria (BIC) [Schwarz, 1978] given in Equation (7-2), where L is the likelihood of each data point, k is the number of free parameters and N is total number of data-points. The value of k is calculated as $k=10N_C-1$ since each kernel requires 3 (mean vector) + 6 (symmetric covariance matrix) + 1 (weight) = 10 free parameters. k is reduced by one due to the fact that the weights have to sum to unity and hence knowing N_C-1 of them is sufficient.

$$BIC = -\sum_i^N \log(L) + \frac{k}{2} \log(N) \quad 7-2$$

Similarly to the Ward's minimum variance criteria, we use the measure of information gain in terms of BIC to select the best merger. For any given pair of clusters the BIC gain is calculated using Equation (7-3) where L_{int} is the likelihood of each data-point for the initial (unmerged) model and L_{mrg} is likelihood in the case where the two candidate clusters are merged. Notice that each merger

decreases k by 10, thus a merger can be favourable only if the reduction of the penalty term is greater than the decrease of likelihood (i.e. $BIC_{Gain} > 0$)

$$\begin{aligned}
 BIC_{Gain} &= BIC_{int} - BIC_{mrg} \\
 BIC_{Gain} &= -\sum_i^D \log(L_{int}) + \frac{k}{2} \log(N_d) \\
 BIC_{mrg} &= -\sum_i^D \log(L_{mrg}) + \frac{k-10}{2} \log(N_D) \\
 BIC_{Gain} &= \sum_i^D \log(L_{mrg}) - \sum_i^D \log(L_{int}) + 5 \log(N_D)
 \end{aligned} \tag{7-3}$$

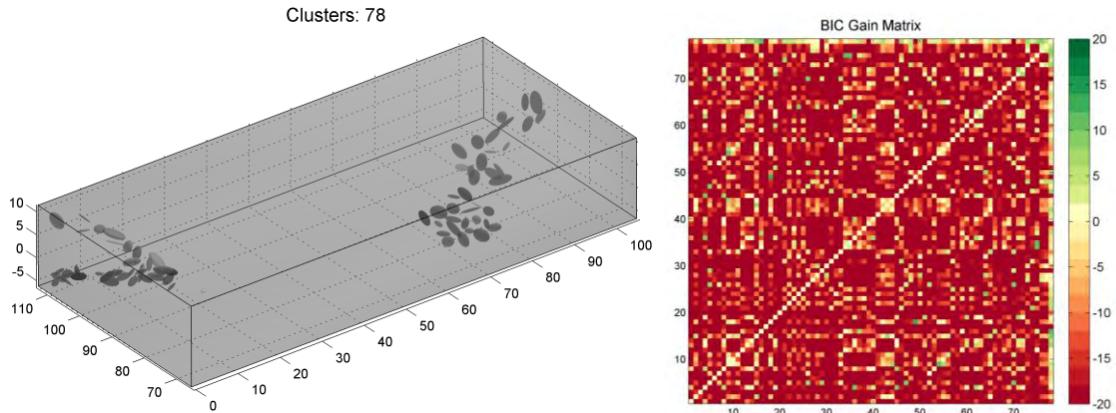


Figure 7-3 Left: The initial *protoclusters* for the synthetic dataset given in Figure 7-2. Notice that the number of clusters includes the uniform background kernel as well. Right: The *BIC* gain matrix calculated for all possible mergers.

Using this formulation we calculate a matrix where the value at i^{th} row and j^{th} column corresponds to the BIC gain for a merger between cluster i and j . We merge the pair with the maximum BIC gain and then recalculate the matrix. At each step the complexity of the model is reduced by one cluster and this continues until there is no merger with a positive BIC gain. Figure 7-3 shows such a BIC gain matrix calculated for the initial model with 77 clusters. Notice that mergers involving the background kernel are not considered. The $BIC_{Gain} > 0$ criteria, which essentially drives and terminates the merging procedure, is similar to a likelihood ratio test [Neyman and Pearson, 1933; Wilks, 1938] with the advantage that the models tested need not be nested. The computational demand of the BIC gain matrix increases quadratically with the number of data-points, hence to make our approach feasible for seismic datasets, we introduce a preliminary check that considers clusters as merger candidates only if they are overlapping within confidence interval of $\sqrt{12}\sigma$ in any principal component direction. The factor $\sqrt{12}$ is due to the variance formulation of the assumed uniform distribution of events over seismogenic planes (for details see [Ouillon et al., 2008]).

During all steps of the merger procedure the data-points are in the state of *soft clustering*, meaning that they have a finite probability to belong to any given kernel. If a deterministic assignment is needed, then each point can be assigned to the cluster that provides the highest probability, this is referred to as *hard clustering*. The dichotomy of stochastic and deterministic inference gives rise to two different implementations of the merging criteria: *a) local* criterion:

considering only the two candidate clusters and the data-points assigned to them through hard-clustering and *b) global* criterion: considering the likelihood of all data-points for all clusters. In essence, the local criterion tests the information gain of for the case of two kernels versus one kernel on a subset, whereas the global criterion considers N_G versus N_G-1 kernels on the whole dataset.

After all the mergers are complete, the weights of the kernels are updated based on the number of their hard clustered events. Figure 7-4 shows the resulting reconstructions for the two criteria.

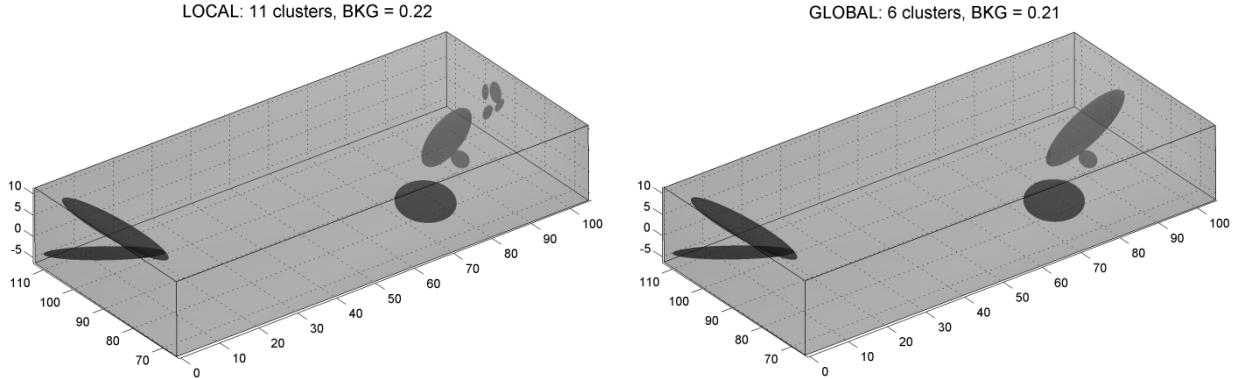


Figure 7-4 The final models obtained using the local (left) and global (right) criteria. Notice that the number of clusters includes the uniform background kernel as well

For this synthetic dataset, we observe that both the local and global criteria converge to a similar final structure. The global criterion yields a model the same number of clusters as the input synthetic, while the local criterion produces four additional clusters in the undersampled part of a part of a fault. For most pattern recognition applications that deal with a robust definition of noise and signal, the global criterion may be the preferred choice since it is able to recapture the input complexity level. However, considering the nature of our application, we see potential in the local criterion as well. Several studies provide evidence for the notion of fractal structure of fault networks [Hirata, 1989; Mandelbrot, 1989; Sammis and Biegel, 1989; Hirabayashi *et al.*, 1992; Quillon *et al.*, 1995b]. Thus, if a fault is the collection of a set of smaller self-similar faults, one can expect to recognize a fault only if it is sufficiently sampled. Additionally, during its growth a fault network may undergo expansion and reduction of structural complexity as small faults emerge and then coalesce to form larger faults. In our synthetic case, the dataset is created by random uniform sampling of individual planes. The random uniform sampling can produce coherent small scale structures that become less significant as the sample size is increased. Thus we can consider the results obtained using the local criterion as expressions of a locally coherent structure that can be simplified if viewed in the context of more numerous regional observations.

7.4 Application to seismicity

In this section, we apply our method to real seismicity data. For this purpose, we use the KaKiOS-15 catalog that was obtained by probabilistic absolute location of nearly 475,000 Southern Californian events spanning the time period 1981-2012. We consider all recorded events since they each sample parts of the fault network. Before tackling this vast dataset, however, we first consider a smaller dataset to assess the overall performance and computational demands.

7.4.1 Small Scale application to Landers aftershocks

We use the same dataset as [Wang *et al.*, 2013] that consists of 3360 aftershocks of the 1992 Landers earthquake. The initial atomization step produces a total of 394 protoclusters that are iteratively merged using the two different criteria. The resulting fault networks are given in Figure 7-5. Comparing the two fault networks, we observe that the local criterion provides a much detailed structure that is at large scales consistent with the global. We also observe that in the northern end the global criterion lumps together small features with seemingly different orientations producing thick clusters. These small scale features have relatively few points and thus low contribution to the overall likelihood, thus the global criterion favors merger to benefit from the reduced penalty, which scales with the total number of points. In the local case, however, because each merger is evaluated only based on the points of the merging clusters, the likelihood of such features can overcome the penalty.

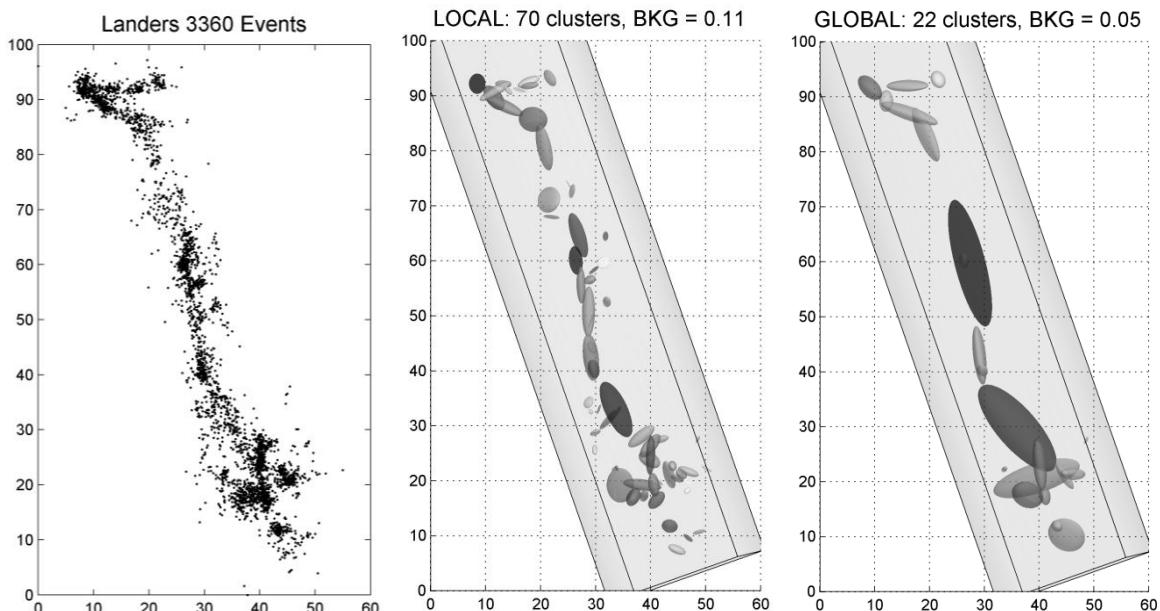


Figure 7-5 Left: Top view of the 1992 Landers aftershocks. Fault networks obtained from these event using the local (center) and global (right) merging criterion.

Our second observation is that the background kernel attains a higher weight using the local criterion compared to the global one. Keeping in mind that both criteria are applied on the same initial set of protoclusters and that there are no mergers with the background kernels, the disparity between the background weights can be understood as being due to density differences in the tails of the kernels. This is demonstrated in Figure 7-6 where see that a merged kernel has higher densities in its tails compared to its constituents. This effect is amplified when the distance between the merging clusters is increased. Hence, in the local case, the peripheral points are more likely to be associated with the background due to the lower densities at the tails of the small clusters.

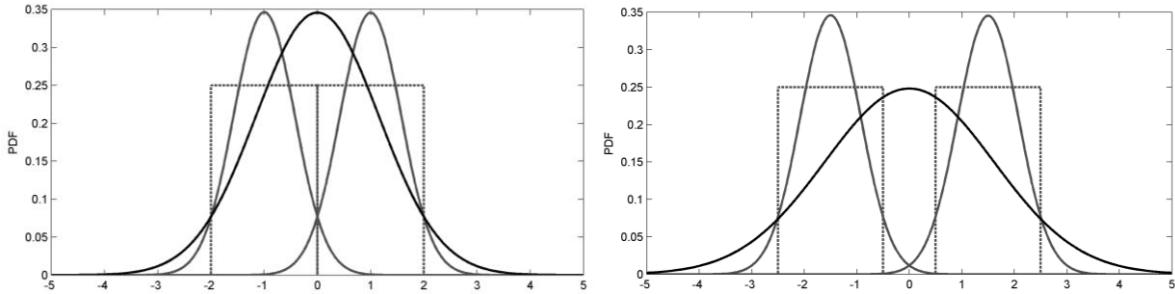


Figure 7-6 Two uniform distributions (dotted gray lines), their Gaussian approximations (solid gray lines) and the Gaussian resulting from their merger (solid black line). Notice that the joint Gaussian has higher densities at the tails compared to its constituents.

Another important insight from this sample case was related to the feasibility of a large scale application. As pointed out here and in previous studies, the computational demands for such pattern recognition methods escalate rapidly with the number of data-points. The Landers case with 3360 points took ~5 minutes on a 4-core, 2.2GHz machine with 16GB memory. Considering that our target catalog is nearly ~145 times larger, a quadratic increase would put the expected computation time at more than two months. Even with a high performance computing cluster, we would have to tackle memory management and associated overhead issues. Although technically feasible, pursuing this path would limit the use of our method to only the few privileged with access to such computing facilities. In our previous work we proposed a different solution, catalog condensation, that uses the location uncertainty to reduce the length of a catalog while preserving its information content [Kamer *et al.*, 2015]. In the following section we will show how this method was applied to the KaKiOS-15 catalog.

7.5 Condensation of the KaKiOS-15 catalog

The initial formulation of the condensation method was developed considering the state of the art catalogs of the time. Location uncertainties in these catalogs are assumed to be normally distributed and hence expressed either in terms of a horizontal and vertical error or with a symmetric 3x3 covariance matrix. With the development of the KaKiOS-15 catalog, we extended this representation to allow for arbitrarily complex location PDFs modelled with mixtures of Gaussians. This was found to be the optimal representation for 81% percent of the events, which were expressed by an average of 3.24 Gaussians. The condensation methodology can be easily generalized to accommodate this broader representation. In the normal version, all events begin with equal weights of 1. They are then ranked according to their isotropic variances and weights are progressively transferred from the high variance to low variance events. In the generalized version, each Gaussian kernel starts with its own mixture weight, these kernels are then ranked according to their isotropic and the weight transfers are carried out as in the previous method except for the single constrain that transfer between kernels of the same event are not allowed (see Figure 7-7a,b). This constrain is motivated by the fact that the kernels representing each event location PDF are already optimized and hence a weight transfer between those can only lead to a sub-optimal model.

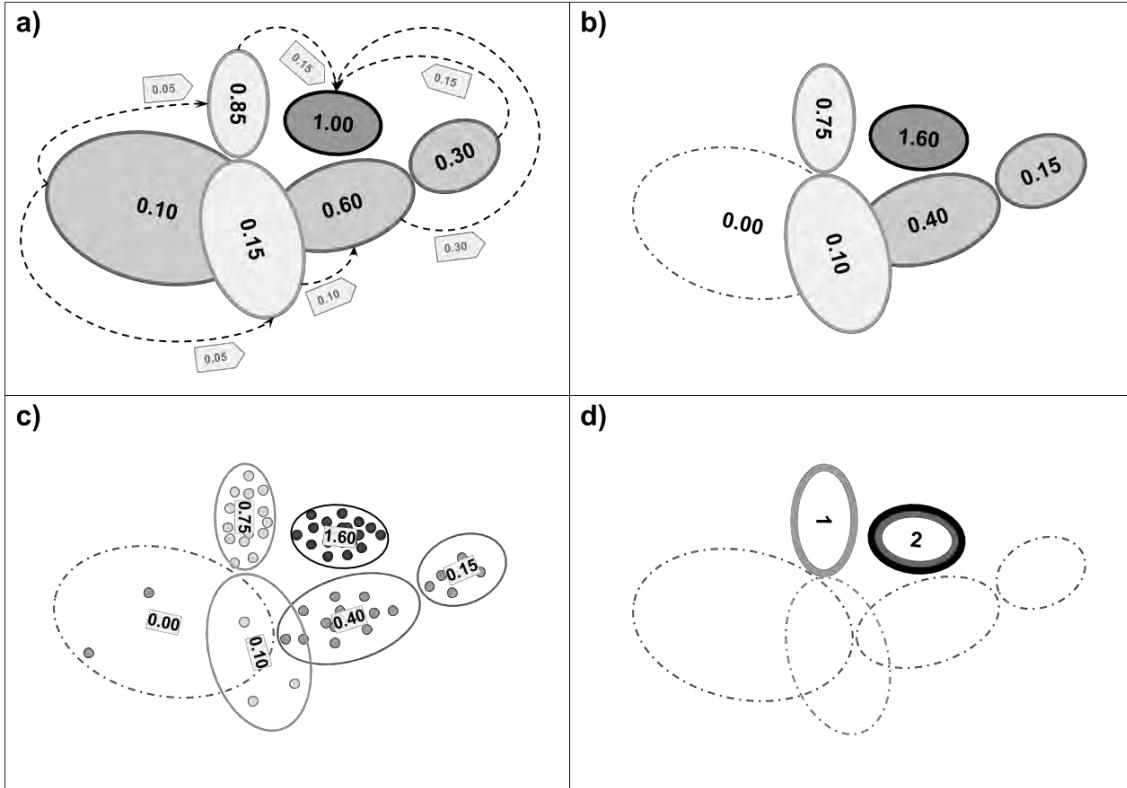


Figure 7-7 Idealized schematic representations of a) Condensation: each event is represented by a different shade, weight is transfer is represented by the arrows, notice that there are no intra-event weight transfers b) Final condensed catalog: total weight sum is preserved, one component is discarded. c) Sampling of the event PDFs: this step is independent of the condensed catalog d) Maximum likelihood assignment of the three events onto two of the condensed.

The KaKiOS-15 catalog contains 475,371 events whose location PDFs are represented by a total of 1,336,322 Gaussian components. Condensation reduces this number to 600,463 as weights are transferred to better located ones. Nevertheless, in Figure 7-8 we see that nearly half of these components amount to only 10% of the total event weight. One way to proceed in such a case would be to take the components with the largest weights with a cumulative weight amounting to a 90% or 95% of the total mass.

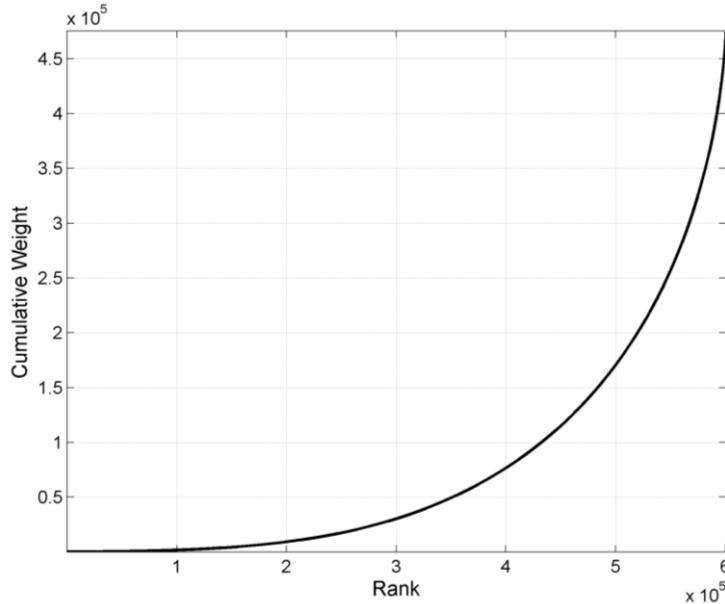


Figure 7-8 Cumulative weights of the 600,463 condensed KaKiOS-15 components representing a total of 475,371 events. The components are ranked according to increasing weights.

We avoid such arbitrary cut-offs by noticing that the condensed catalog is nothing more than a Gaussian mixture model representing the spatial PDF of earthquake occurrence in South California. We can then, in the same vain as the hard clustering described previously, assign each event to its most likely GMM component. Since each event is represented by a location PDF, we sample it with 1000 points and then calculate the kernels with the highest likelihoods for each sample point. The event is assigned to the kernel that provides the highest likelihood for the highest number of sample points (see Figure 7-7c,d). As a result, we see that the 475,371 events are assigned to 93,149 distinct kernels. The spatial distribution of all condensed kernels and the ones assigned with at least one event are given in Figure 7-9. Essentially, we are using the condensed catalog as a prior on the individual event locations.

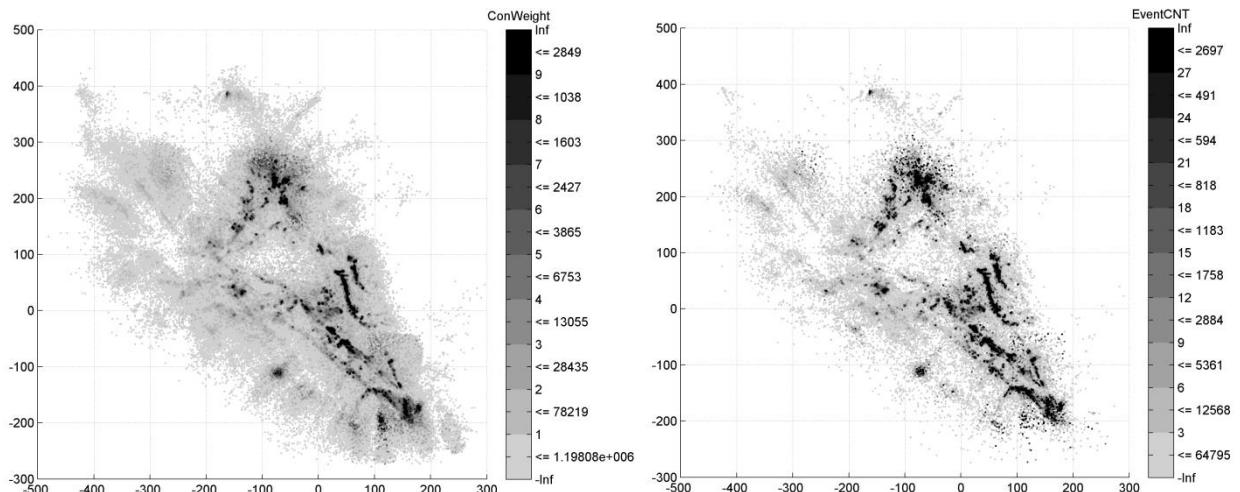


Figure 7-9 Left: Mean locations of condensed 600,463 Gaussian components shaded according to their weights. Right: The same components shaded according to the total number of events assigned to them after the maximum likelihood assignment

7.6 Large scale application to Southern California

Our previous analysis concluded that the spatial distribution of southern California seismicity is multifractal, i.e. it is an inhomogeneous collection of singularities [Kamer *et al.*, 2015]. We see the distributions given in Figure 7-9 as expressions of these singularities. Since we are interested in the general form of the fault network rather than the second order features (e.g inhomogeneous seismicity rates) we consider all the 93,149 kernels as individual points without taking into account their event counts. Considering these counts would result in more complex and singular structures that can be associated with the multifractal slip distribution of big events modulated through the non-uniform network detection capabilities. These two different options can be understood as mimicking the definitions of the capacity (D_0) and information dimensions (D_I), which are identical in monofractals whereas $D_0 > D_I$ in multifractals.

Another important aspect, which deserves further attention in the case of such a large scale application, is the uniform background kernel. The assumption of a single background kernel defined as the minimum bounding box of the entire dataset seems to be a fair assumption for the case of Landers aftershocks, however it becomes evident that in the case of Southern California such a box would overestimate the data span and hence lead to an underestimated density. We can also expect the background density to vary regionally. The densities might be higher for regions in the periphery of the network, swarms or geothermal fields compared to well-defined active faults. We thus extend our approach by allowing for multiple background kernels. For this purpose we make use of the AHC tree of that is already calculated for the atomization of the whole dataset. We divide the dataset into a small number of subsets based on this tree. Each of these subsets then is atomized individually and hence in the process gets its own background kernel. The atomized subsets are then joined to form a single structure which is then progressively merged. Naturally, we have no objective way of knowing how many background kernels a dataset may feature. However, in various synthetic tests (involving cuboid backgrounds) we observe that inflating this number has no effect on the densities, whereas a too low value causes underestimation. Apart from the justifications mentioned above, we are motivated to divide this large dataset into subsets for purely computational reasons as this allows for improved parallelization and computational efficiency.

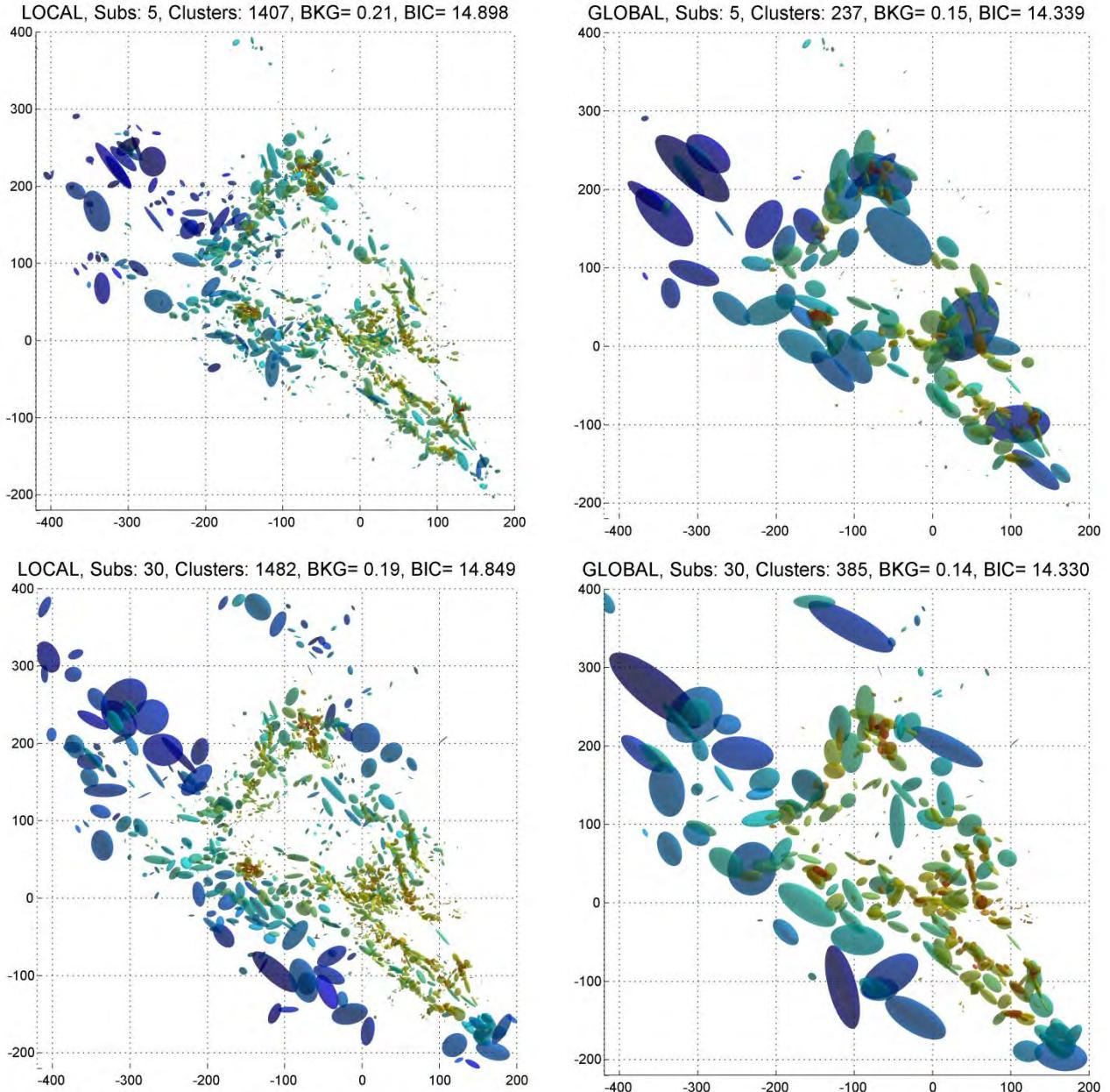


Figure 7-10 Fault network reconstructions for the KaKiOS-15 catalog. Top row shows results for the case of 5 initial subsets with local (left) and global (right) merging criterion. Bottom row shows the same for 30 initial subsets. The number of clusters, background weight and BIC per data point is given in the titles.

In Figure 7-10 we present two fault networks obtained for two initial divisions with 5 and 30 subsets. For each choice, we show the result of the local and global criteria, the background cuboids are not plotted to avoid clutter. Our immediate observation is related to the events associated with the 1986 Oceanside sequence. The kernel associated with these events is virtually absent in the fault networks reconstructed from 5 initial subsets. This can be explained in terms of the atomization procedure. In the case of 5 initial subsets, the offshore Oceanside seismicity falls in a subset containing onshore faults (e.g the Elsinore fault). Because these faults have a more coherent spatial structure compared to the diffused Oceanside seismicity, their protocluster capacity is higher. Hence the atomization procedure continues increasing the number of clusters while the Oceanside seismicity has actually reached its capacity. As a result, this causes nearly all of the protoclusters

within this region to become singular and hence be discarded. In the case of 30 subsets, the Oceanside seismicity is in a smaller region and thus has a better capacity estimation.

At this point the natural question would be: which of these fault networks is a better model? The answer to this question would naturally depend on the application. If one is interested with correspondence of focal mechanisms or high resolution fault traces, which are expressions of local stress/strain conditions, then the ideal choice would be the local criterion. However, if the application of interest is forecasting then one might consider the global criterion as it provides a lower BIC value due to being formulated with respect to the overall likelihood. We leave the statistical investigation of the fault network parameters as a subject for a separate study and instead focus on immediate uses of the obtained fault network.

7.7 Validation through a spatial forecast test

Several methods can be proposed for the validation of a reconstructed fault network. One way could be to project the faults on the surface and check their correspondence with the mapped fault traces. This would be a tedious task since it would involve a case by case qualitative analysis. Furthermore, many of the faults illuminated by the seismicity might not have been mapped or they may simply have no surface expressions. In the recent case of the 2014 Napa earthquake, there was also a significant disparity between the aftershock distributions and the observed surface trace [Brocher *et al.*, 2015]. Another option would be to compare the accordance between the reconstructed faults and the focal mechanisms of the events associated with them. With many of the metric already developed [Wang *et al.*, 2013], this would allow for a systematic evaluation. However, the current focal mechanism catalog for Southern California is based on the HYS-12 catalog obtained by relative double-difference techniques. As previously discussed in our studies [Kamer *et al.*, 2015], we believe this catalog exhibits artificial clustering effects at different scales. Hence, any focal mechanism based on hypocenters from this catalog would be inconsistent with our absolute KaKiOS-15 catalog.

Hence we are left with the ultimate option: validation by 3D spatial forecasting. For this purpose, we will use the global criterion model obtained from 30 subsets because it has the lowest BIC value. Our fault reconstruction is based on all events in the KaKiOS-15 catalog, regardless of their magnitude. The last event in this catalog occurred on June 30th 2011. Thus, we consider all routinely located events by the Southern California Earthquake Data Center between the July 1st 2011 and July 1st 2015 with magnitudes larger than M2.5 as our target events. We limit our volume of interest arbitrarily to the region limited by latitudes [32.5, 36.0], longitudes [-121, -115] and depths in the range 0-20km. The likelihood scores of the target events are calculated directly from the fault network, which is essentially a weighted mixture of Gaussian PDFs and uniform backgrounds. The only modification done to accommodate the forecast is aggregating all background kernels into a single cuboid covering the volume of interest. The weight of this cuboid is equal to the sum of all individual background kernel weights. To compare the performance of our fault network we also need a reference model. The trivial choice with zero prior information would be a single uniform background. We also consider a more competent model, the TripleS [Zechar and Jordan, 2010a] that is often used in as forecasting benchmark. This model is obtained by replacing each event with an isotropic, constant bandwidth Gaussian kernel. The bandwidth is then optimized for by dividing the dataset into training and validation sets. As already pointed out by [Zechar and Jordan, 2010b] this model involves several choices (e.g. choice of optimization

function, choice of candidate bandwidth, etc...). Rather than spending time on trying to construct a fairly competitive TripleS model, we side-step the issue by allowing optimization of the bandwidth parameter directly on the target set. Given such a privilege of foresight, this should correspond to the upper bound of the TripleS forecast skill. In Figure 7-11 we show the results for varying magnitude cut-offs in terms of negative log likelihood per event. Our fault reconstruction performs better for all magnitude cut-off levels. We also see a consistent relative increase in its performance with increasing magnitude cutoff.

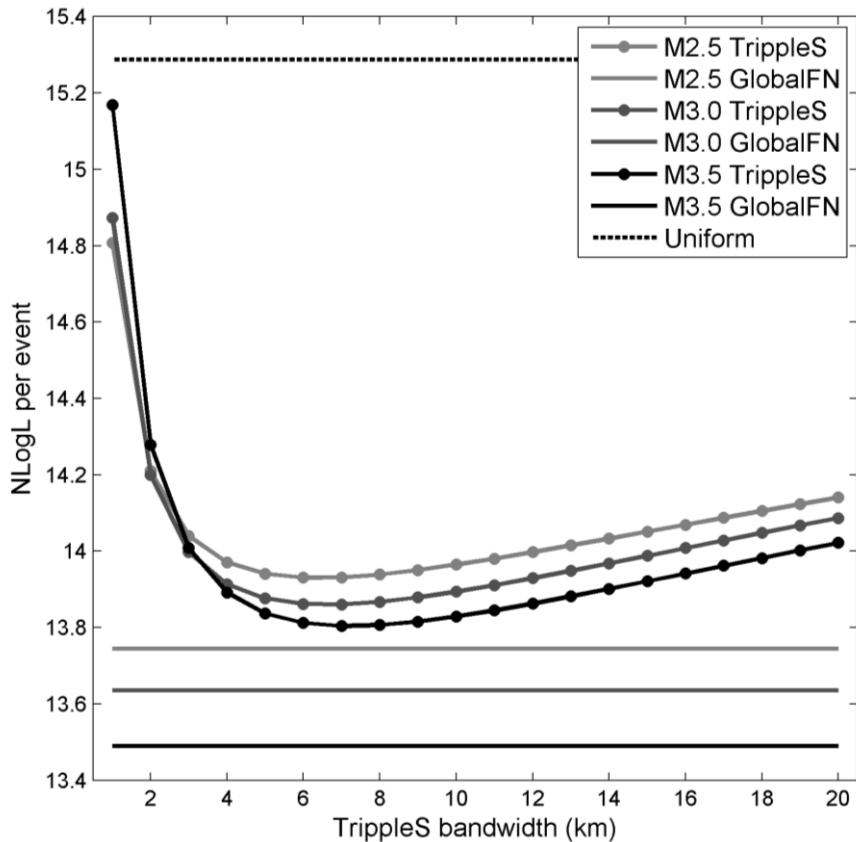


Figure 7-11 Average Negative Log Likelihood for the target dataset limited to events above M2.5 (light gray), M3.0 (dark gray) and M3.5 (black). Performance of the TripleS models is evaluated as function of the isotropic kernel bandwidth (dotted lines). The fault network performance is shown with constant level solid lines.

The superiority of our model with respect to TripleS can be understood in terms of model parameterization. There is a general misconception regarding the “complexity” of a model in the earthquake forecasting community. Sadly, we often see this misconception being perpetuated by the very people involved in testing and evaluating forecasts. For instance, in their 2010 paper, Zechar and Jordan refer to the TripleS model as “a simple model” compared to models employing anisotropic or adaptive kernels [Kagan and Jackson, 1994, 2007]. It is likely that, owing to such allegations, some of our readers might believe that the model presented in this study is far more complex than TripleS. However, this is not true. The complexity of a model is independent of the algorithmic procedures undertaken to obtain it. What matters is the number of its free parameters or in other words its minimum description length [Rissanen, 1978; Schwarz, 1978]. TripleS is essentially a GMM model expressed by the 3D locations of its components and a constant kernel bandwidth. Hence it has a total of $(3*475,371)+1=1,426,114$ free parameters compared to the $(10*93,149)-1=931,489$ of our fault network. The fact that the TripleS kernels are co-located with

hypocenters of previous earthquakes does not reduce the complexity of the model; it only amounts to a suboptimal modeling choice. Thus, the performance difference between the two models can be understood in terms of the TripleS' build-in tendency to overfit. Another contributing factor could be regarded as the location uncertainty information that facilitates condensation and consequently puts more emphasis on the fault structure rather than the aftershock singularities.

7.8 Conclusion

In this paper, we introduced an agglomerative clustering method for seismicity-based fault network reconstruction. The method provides the following advantages: 1) a bottom-up approach that explores all possible paths at each step and moves coherently towards a global optimum; 2) introduction of an optimized atomization scheme that aims to isolate the background/uncorrelated points; 3) improved performance due to added geometrical merging constrains. We were able to reconstruct a very large dataset consisting of 30 years of South Californian seismicity by taking into account the non-linear location uncertainties of the events and condensing the catalog ~20% of its initial size. We validated the reconstructed Southern California fault network through a retrospective 3D spatial forecast test, targeting the last 4 years of seismicity.

Notwithstanding these encouraging results, there several aspects in which the proposed methodology can be further improved and extended. In the current formulation, the background kernels are represented by the minimum bounding box of each subset, thus they tend to overlap and bias the overall background density. This can be improved by employing convex hulls, alpha shapes [*Edelsbrunner and Mücke, 1994*] or a Voronoi tessellation [*Voronoi, 1908*] optimized to match the subset borders. The background kernel could also be adapted to the specific application; for induced seismicity catalogs, it can be a minimum bounding sphere or an isotropic Gaussian since the pressure field diffuses radially from the injection point. Different types of protocluster such as Student-t kernels or copulas can be used in the atomization step or they can be introduced at various steps by hypothesis testing.

An important implication of the reconstructed fault network is its potential in modeling the temporal evolution of seismicity. The Epistemic Type Aftershock Sequence (ETAS) model can be simplified significantly in the presence of optimally defined spatial kernels. Rather than expressing the whole catalog sequence as the weighted combination of all previous events, we can instead have multiple sequences corresponding to each kernel and hence model each of those as a partial combination of the others. Such a formulation would eliminate the need for the commonly used isotropic distance kernel. This single degree kernel induces essentially the same deficiencies discussed in the case of the TripleS model. Thus, we can expect such an ETAS model, based on our fault network, to perform significantly better than its isotropic variant.

Chapter 8

Conclusion and Future perspectives

8.1 Achieved results

The main goal of this thesis was to enhance our understanding of earthquake occurrence and ultimately improve the current state-of-the-art long term seismicity forecasting models. Under the current testing regulations, a skillful long term forecast requires precision in both magnitude and space dimensions. For this purpose, in Chapter 2 we initially focused our attention on the spatial variation of the frequency magnitude distribution (FMD). We introduced a non-parametric approach based on Voronoi tessellation and penalized likelihood in order to investigate models with different spatial complexities. Our results indicated significant spatial variations in the FMD, expressed in terms of Gutenberg-Richter's b-value. Consequent investigations about the origin of these variations revealed that while some were geophysical in nature, the majority was due to magnitude errors and inconsistencies. These conclusions are significant not only for California but also for other seismically active regions (e.g Japan, Turkey, Italy, Greece...) where b-values are extensively being mapped and associated with physical properties such as normal/shear stresses.

It is difficult to assess the origin of b-value variations in the absence of magnitude uncertainty estimations. To our best knowledge, currently only the New Zealand GeoNET catalog reports consistent magnitude errors. Our results can be used to argue that, under these conditions, resources allocated to b-value studies may be diverted to other fields that are less prone to such biases, or at least where data errors are assessed and reported.

To understand to what degree these b-value variations may influence typical earthquake forecasting models, in Chapter 3 we investigated the performance of our non-parametric approach in a retrospective test. The results of this test suggested that the method achieves significant information gains with respect to current models. This can be considered as an indication that the performance of current forecasting models may just as well be driven by seismic network effect. Unsettled by this conclusion and in the absence of consistent data on magnitude error, we turned our attention to the second component of long term forecast models; i.e the spatial distribution of seismicity. Just as in the case of FMD, the spatial distribution of earthquakes also follows a self-similar scaling. However, unlike in the b-value, whose estimation methods are well established, estimates of the multifractal spectrum vary depending on the used method. This motivated us, in Chapter 4, to develop a new method based on barycentric pivot point selection and the non-overlapping coverage criteria. We tested our method's performance against various multifractal distributions to confirm its improved performance. It is interesting to notice that, similarly to the state-of-the art FMD modeling methods that try to avoid dropping below the detection threshold,

our multifractal analysis method tries to avoid the edges of the spatial distribution. In both cases the constraints can be considered as instrumental rather than physical.

The Barycenteric Fixed Mass (BFM) method was able to successfully account for the effect of the network coverage geometry, which governed the spatial boundaries, however another important factor remained neglected: the location uncertainty. We realized that for a spatial analysis, the catalog can be viewed as the discrete sampling of an underlying spatial PDF. This gave rise to the notion of an optimal encoding for such a PDF. We were then able to demonstrate in Chapter 5, that taking into account the location uncertainty information of all events allows for the formulation of a similar PDF but much more concisely; with only a fraction of the initial events. Furthermore, we demonstrated through synthetic tests that such a condensed catalog shows greater correspondence with respect to a known generating distribution. The condensation method provides not only a reduction of overall data-length, but also a natural way of accounting for the location uncertainty of the events. Employing these two methods in the spatial analysis of the Southern Californian seismicity (HYS-12 catalog) revealed a robust multifractal scaling. We also observed distinct transitions in this scaling. While we were able to associate some of these breaks to a physical attribute (e.g the thickness of the seismogenic zone), interpretation of other transitions remained more elusive. We decided to investigate further, since such transitions could have implications for the general understanding of earthquake triggering. We once again turned our attention to the consistency of the data and decided to undertake the task of relocating the last three decades of Southern Californian seismicity, rather than basing our conclusions on speculations regarding the possible caveats of the currently available catalogs.

Our main goal in this endeavor was to develop a catalog with uniformly consistent location uncertainty information. For this purpose, in Chapter 6, we first developed a minimum 1D Vp velocity model with station corrections using VELEST. We then used this model in the non-linear, absolute relocation of the whole catalog using NonLinLoc. Our comparison of this new absolute location catalog (KaKiOS-15) with its relatively relocated counterpart (HYS-12) revealed that there are significant differences in the observed multifractal spectra. We also found strong evidence that the previously observed scaling break at small scales is due to location errors. The absolute locations of the KaKiOS-15 catalog provide evidence for a single multifractal regime and hence they are consistent with the overall scale-invariant nature of earthquake phenomena.

Empowered by this improved catalog and the understanding of multifractal scaling, we moved on to a more detailed spatial analysis based on pattern recognition techniques. In Chapter 7, we developed a new fault network reconstruction method. It is important to note that this last chapter was actually the initial goal of this thesis. Thus while developing all the other presented methods, we were simultaneously laying the foundations of the new agglomerative clustering method. For instance, the spatial b-value variation results in Chapter 2 motivated us to choose Southern California as our area of investigation due to its uniform b-value distribution. If we had chosen to investigate the north, we would have to deal with the distinctively non-tectonic seismicity in Geysers and the poor network coverage in Mendocino Triple Junction. Building on the cumulative results of Chapter 4, 5 and 6, we saw the single multifractal regime and the uniform b-value in Southern California as an indication that the seismogenic process could be seen, and hence modeled, as emerging from simple structures that intermingle to form a large scale, self-similar network. This understanding inspired the bottom-up, agglomerative approach presented in Chapter 7. Similarly, its large scale applications to the KaKIOS-15 catalog were only made possible by the

data-length reduction achieved with the condensation method. We were able to benefit from our multifractal understanding also in terms of improving spatial forecasts. Unlike models based on the higher order dimension D_I , we argued that earthquake nucleation essentially samples the capacity dimension D_0 . This motivated us to disregard the condensation weights in our application. The information gain achieved by the reconstructed fault network can be attributed to this basic understanding.

8.2 Future perspectives

The methods and data sources developed in this thesis have led to a better understanding of the statistical properties of earthquake phenomena and data inconsistencies associated with its measurement. Also, as demonstrated on several occasions, we were able to exploit these understandings in order to improve current forecasting models. Thus, we are confident that efforts aimed at improving and extending the proposed methods will lead to further insights and better forecasts. Here we shall point to some of the possible paths to pursue in this direction.

The data-driven spatial b-value estimation method, presented in Chapter 1, can be extended by allowing for segmentations in the time-domain as well. This would facilitate the temporal analysis that is often found to be affected by aftershock incompleteness issues. The Voronoi tessellation method can also be improved to allow for more sophisticated and realistic segmentations. This can be achieved fairly easily by implementing nested Voronoi tessellations. Another alternative could be to build up on the clusters obtained from our fault network reconstruction in Chapter 7. We could keep the spatial structure constant and consider mergers in the magnitude domain that are conditioned on nested hypothesis tests. This would allow different spatial clusters to be modeled with similar FMDs. This would automatically constitute a time-independent forecast model. However, the application of the bottom-up approach might not be so straightforward. Mergers in the magnitude domain would require the use of entire range FMD models, since the likelihood has to be defined for all events, regardless of the varying completeness levels during the mergers. Other issues might arise if clusters span across different networks, in which case we can expect consistency in earthquake locations but not in magnitudes.

Although we have spent considerable effort in improving the consistency of the Southern Californian seismicity locations, we have not tackled the issue of magnitudes. In Section 2.7.2 we showed how location errors can propagate into magnitude errors. Now that we have the complete location PDFs of the events in the KaKiOS-15 catalog, we can do a similar estimation for those magnitude errors. We might have to make certain assumptions in the absence of information about the actual stations used in magnitude estimation; however, this would still provide us with a lower bound on the realistic magnitude uncertainties. The current version of the KaKiOS-15 catalog can also benefit from further improvements. The waveforms for most of the events are available and we can use automatic picking algorithms to provide not only better pick times but also uncertainties.

The fault reconstruction methodology introduced in Chapter 7 is probably the topic that provides the largest possibilities for future investigation and refinements. One of them is certainly the immediate need for a consistent focal mechanism catalog that would allow us to make physical validations in terms of fault orientations. With such insights we should be able to reconstruct better fault networks consistent with observed faulting styles. If substantiated, the correspondence between the fault geometries and focal mechanism can be used to improve Coulomb stress models. The reconstruction procedure can also be extended to allow for different types of kernels. For

instance, a Student-t kernel can allow for heavier tails whereas an Epanechnikov kernel can provide a stronger taper. Moreover, the kernel type does not have to be specified *a priori*, it can vary through optimization at each step of the merger procedure. This would result in a reconstruction that is more data-adaptive. In Chapter 7, we discussed how the fault network can allow for a simplified yet more robust re-formulation of the ETAS model. Another possible way to model such a multiple-input-multiple-output system would be to use artificial neural networks (ANNs). This would allow for the use of different types of non-linear interconnections with customizable complexities. The ANN field can provide well established training and optimization methods (such as back propagation of error and neuron pruning) that could be used to complement our current statistical forecast models.

In the light of these exciting prospects, we believe that fault network reconstruction based on a consistent, absolute catalog opens a whole new dimension that awaits to be explored. Since data is so vast, this needs to be done with rigor and efficiency in a holistic way. Yet, we should not let this endeavor degrade itself into a petty, mundane task of feeble refinements that never amount to nothing. We should be able to see our achievement for what it is, and if need be, leave it; for if it can't save lives, then it is worthy of none. We then move on to new frontiers, new data sources and push on.

References

- Agnew, C. (2010), Comment on “Changes of Reporting Rates in the Southern California Earthquake Catalog, Introduced by a New Definition of “L” by Thessa Tormann, Stefan Wiemer, and Egill Hauksson, *Bull. Seismol. Soc. Am.*, 100(6), 3320–3324, doi:10.1785/0120100027.
- Aki, K. (1965), Maximum Likelihood Estimate of b in the Formula $\log N = a - bM$ and its Confidence Limits, *Bull. Earthq. Res. Inst.*, 43, 237–239.
- Amitrano, C., A. Coniglio, and F. Di Liberto (1986), Growth probability distribution in kinetic aggregation processes, *Phys. Rev. Lett.*, 57(8), 1016.
- Amitrano, D. (2003), Brittle-ductile transition and associated seismicity: Experimental and numerical studies and relationship with the b value, *J. Geophys. Res.*, 108(B1), 2044, doi:10.1029/2001JB000680.
- Amitrano, D. (2012), Variability in the power-law distributions of rupture events, *Eur. Phys. J. Spec. Top.*, 205(1), 199–215, doi:10.1140/epjst/e2012-01571-9.
- Amorese, D., J.-R. Grasso, and P. a. Rydelek (2010), On varying b -values with depth: results from computer-intensive tests for Southern California, *Geophys. J. Int.*, 180(1), 347–360, doi:10.1111/j.1365-246X.2009.04414.x.
- Argoul, F., A. Arneodo, G. Grasseau, and H. Swinney (1988), Self-Similarity of Diffusion-Limited Aggregates and Electrodeposition Clusters, *Phys. Rev. Lett.*, 61(22), 2558–2561, doi:10.1103/PhysRevLett.61.2558.
- Argoul, F., A. Arneodo, and J. Elezgaray (1990), Wavelet analysis of the self-similarity of diffusion-limited aggregates and electrodeposition clusters, *Phys. Rev. A*, 41(10).
- Arneodo, A., G. Grasseau, and M. Holschneider (1988), Wavelet transform of multifractals, *Phys. Rev. Lett.*, 61(20), 2281.
- Bachmann, C. E., S. Wiemer, B. P. Goertz-Allmann, and J. Woessner (2012), Influence of pore-pressure on the event-size distribution of induced earthquakes, *Geophys. Res. Lett.*, 39(9), 1–7, doi:10.1029/2012GL051480.
- Barbosa, R., and A. Protti (1988), Measurement of the dimension spectrum $f(\hat{I}^\pm)$: Fixed-mass approach, *Phys. Lett. A*, 131(6), 339–343.
- Bak, P., and C. Tang (1989), Earthquakes as a self-organized critical phenomenon, *J. Geophys. Res.*, 94(B11), 15635, doi:10.1029/JB094iB11p15635.
- Bender, B. (1983), Maximum likelihood estimation of b values for magnitude grouped data, *Bull. Seismol. Soc. Am.*, 73(3), 831–851.

- Bird, P. (2003), An updated digital model of plate boundaries, *Geochemistry, Geophys. Geosystems*, 4(3).
- Bird, P., and Y. Y. Kagan (2004), Plate-Tectonic Analysis of Shallow Seismicity: Apparent Boundary Width, Beta, Corner Magnitude, Coupled Lithosphere Thickness, and Coupling in Seven Tectonic Settings, *Bull. Seismol. Soc. Am.*, 94(6), 2380–2399, doi:10.1785/0120030107.
- Bishop, C. M. (2007), *Pattern Recognition and Machine Learning*, Springer.
- Bodin, T., and M. Sambridge (2009), Seismic tomography with the reversible jump algorithm, *Geophys. J. Int.*, 178(3), 1411–1436, doi:10.1111/j.1365-246X.2009.04226.x.
- Boettcher, M. S., A. McGarr, and M. Johnston (2009), Extension of Gutenberg-Richter distribution to $W -1.3$, no lower limit in sight, *Geophys. Res. Lett.*, 36(10), L10307, doi:10.1029/2009GL038080.
- Boulalet, H., and G. Albinet (1996), Frozen and active regions in diffusion-limited aggregation clusters., *Phys. Rev. E*, 53(5), 5106–5110.
- Boyle, K., and M. Zoback (2014), The Stress State of the Northwest Geysers, California Geothermal Field, and Implications for Fault-Controlled Fluid Flow, *Bull. Seismol. Soc. Am.*, 104(5), 2303–2312, doi:10.1785/0120130284.
- Bray, A., and F. P. Schoenberg (2013), Assessment of Point Process Models for Earthquake Forecasting, *Stat. Sci.*, 28(4), 510–520.
- Brocher, T. M. et al. (2015), The Mw 6.0 24 August 2014 South Napa Earthquake, *Seismol. Res. Lett.*, 86(2A), 309–326, doi:10.1785/0220150004.
- Burnham, K., and D. Anderson (2002), *Model selection and multimodel inference: a practical information-theoretic approach*, Springer New York.
- Clauset, A., C. Shalizi, and M. Newman (2009), Power-law distributions in empirical data, *SIAM Rev.*, 51(4), 661–703.
- Clements, R., F. P. Schoenberg, and D. Schorlemmer (2011), Residual analysis methods for space-time point processes with applications to earthquake forecast models in California, *Ann. Appl. Stat.*, 5(4), 2549–2571.
- Coase, R. H. (1995), *Essays on Economics and Economists*, University of Chicago Press.
- Cocco, M., S. Hainzl, F. Catalli, B. Enescu, A. M. Lombardi, and J. Woessner (2010), Sensitivity study of forecasted aftershock seismicity based on Coulomb stress calculation and rate- and state-dependent frictional response, *J. Geophys. Res.*, 115(B5), B05307, doi:10.1029/2009JB006838.
- Deluca, A., and Á. Corral (2013), Fitting and goodness-of-fit test of non-truncated and truncated power-law distributions, *Acta Geophys.*, 61(6), 1351–1394, doi:10.2478/s11600-013-0154-9.

- Dreger, D. S., and D. V. Helmberger (1990), Broadband modeling of local earthquakes, *Bull. Seismol. Soc. Am.*, 80(5), 1162–1179.
- Eberhard, D. A. J., J. D. Zechar, and S. Wiemer (2012), A prospective earthquake forecast experiment in the western Pacific, *Geophys. J. Int.*, 190(3), 1579–1592, doi:10.1111/j.1365-246X.2012.05548.x.
- Edelsbrunner, H., and E. Mücke (1994), Three-dimensional alpha shapes, *ACM Trans. Graph.*, 13(1), 43–72.
- Efron, B. (1979), Bootstrap methods: another look at the jackknife, *Ann. Stat.*, 7(1), 1–26.
- str m, ., . Ne ttles, and A. . iewoński (2012), The global C T pr o ect 2004–2010: Centroid-moment tensors for 13,017 earthquakes, *Phys. Earth Planet. Inter.*, 200-201, 1–9, doi:10.1016/j.pepi.2012.04.002.
- Enescu, B., and K. Ito (2001), Some premonitory phenomena of the 1995 Hyogo-Ken Nanbu (Kobe) earthquake: Seismicity, b-value and fractal dimension, *Tectonophysics*, 338(3-4), 297–314, doi:10.1016/S0040-1951(01)00085-3.
- Enescu, B., and K. Ito (2002), Spatial analysis of the frequency-magnitude distribution and decay rate of aftershock activity of the 2000 Western Tottori earthquake, *Earth, planets Sp.*, 847–859.
- Felzer, K. R. (2006), Calculating the Gutenberg-Richter b value, *AGU Fall Meet. Abstr.*, S42C-08.
- Felzer, K. R., T. W. Becker, R. E. Abercrombie, G. Ekstrom, and J. R. Rice (2002), Triggering of the 1999 M W 7.1 Hector Mine earthquake by aftershocks of the 1992 M W 7.3 Landers earthquake, *J. Geophys. Res.*, 107(B9), 2190, doi:10.1029/2001JB000911.
- Feng, R., and T. V. McEvilly (1983), Interpretation of seismic reflection profiling data for the structure of the San Andreas fault zone, *Bull. Seismol. Soc. Am.*, 73(6A), 1701–1720.
- Field, E. H. (2007), Overview of the Working Group for the Development of Regional Earthquake Likelihood Models (RELM), *Seismol. Res. Lett.*, 78(1), 7–16, doi:10.1785/gssrl.78.1.7.
- Field, E. H. et al. (2014), Uniform California Earth- quake Rupture Forecast, Version 3 (UCERF3)- The Time- Independent Model, *Bull. Seismol. Soc. Am.*, 104(3), 1122–1180, doi:10.1785/0120130164.
- Frohlich, C., and S. Davis (1993), Teleseismic b values; or, much ado about 1.0, *J. Geophys. Res.*, 98, 631–644.
- Furlong, K. P., and S. Y. Schwartz (2004), Influence of the Mendocino Triple Junction on the Tectonics of Coastal California, *Annu. Rev. Earth Planet. Sci.*, 32(1), 403–433, doi:10.1146/annurev.earth.32.101802.120252.
- Gawthrop, W. (2014), Corporate Money Trumps Science, in *Geoethics: Ethical Challenges and Case Studies in Earth Sciences*, vol. 13, p. 450, Elsevier Science.

Geist, E. L., and T. Parsons (2014), Undersampling power-law size distributions: effect on the assessment of extreme natural hazards, *Nat. Hazards*, 72(2), 565–595, doi:10.1007/s11069-013-1024-0.

Godano, C., E. Lippiello, and L. de Arcangelis (2014), Variability of the b value in the Gutenberg-Richter distribution, *Geophys. J. Int.*, 199(3), 1765–1771, doi:10.1093/gji/ggu359.

Grassberger, P., and I. Procaccia (1983), Measuring the strangeness of strange attractors, *Phys. D Nonlinear Phenom.*, 9, 189–208.

Grassberger, P., R. Badii, and A. Politi (1988), Scaling laws for invariant measures on hyperbolic and nonhyperbolic attractors, *J. Stat. Phys.*, 51(1-2), 135–178, doi:10.1007/BF01015324.

Greenside, H., A. Wolf, J. Swift, and T. Pignataro (1982), Impracticality of a box-counting algorithm for calculating the dimensionality of strange attractors, *Phys. Rev. A*, 25(6), 3453–3456.

Guilia, L., S. Wiemer, and D. Schorlemmer (2010), Asperity-based earthquake likelihood models for Italy, *Ann. Geophys.*, 53(3), 63–75, doi:10.4401/ag-4843.

Gutenberg, B., and C. F. Richter (1954), *Seismicity of the earth and associated phenomena*, [2d. ed.], Princeton University Press, Princeton N.J.

Hadley, D., and H. Kanamori (1977), Seismic structure of the Transverse Ranges, California, *Geol. Soc. Am. Bull.*, 88(10), 1469–1478, doi:10.1130/0016-7606(1977)88.

Hanan, W. G., and D. M. Heffernan (2012), Multifractal analysis of the branch structure of diffusion-limited aggregates, *Phys. Rev. E*, 85(2), 021407, doi:10.1103/PhysRevE.85.021407.

Hauksson, E., and P. M. Shearer (2005), Southern California Hypocenter Relocation with Waveform Cross-Correlation, Part 1: Results Using the Double-Difference Method, *Bull. Seismol. Soc. Am.*, 95(3), 896–903, doi:10.1785/0120040167.

Hauksson, E., W. Yang, and P. M. Shearer (2012), Waveform relocated earthquake catalog for Southern California (1981 to June 2011), *Bull. Seismol. Soc. Am.*, 102(5), 2239–2244.

Hausdorff, F. (1918), Dimension und äußeres Maß, *Math. Ann.*, 68.

Helmstetter, A. (2003), Is Earthquake Triggering Driven by Small Earthquakes?, *Phys. Rev. Lett.*, 91(5), 058501, doi:10.1103/PhysRevLett.91.058501.

Helmstetter, A., and M. J. Werner (2012), Adaptive Spatiotemporal Smoothing of Seismicity for Long-Term Earthquake Forecasts in California, *Bull. Seismol. Soc. Am.*, 102(6), 2518–2529, doi:10.1785/0120120062.

Helmstetter, A., and M. J. Werner (2014), Adaptive Smoothing of Seismicity in Time, Space, and Magnitude for Time-Dependent Earthquake Forecasts for California, *Bull. Seismol. Soc. Am.*, 104(2), 809–822, doi:10.1785/0120130105.

- Helmstetter, A., Y. Y. Kagan, and D. D. Jackson (2005), Importance of small earthquakes for stress transfers and earthquake triggering, *J. Geophys. Res.*, 110(B5), B05S08, doi:10.1029/2004JB003286.
- Helmstetter, A., Y. Y. Kagan, and D. D. Jackson (2007), High-resolution Time-independent Grid-based Forecast for M5 Earthquakes in California, *Seismol. Res. Lett.*, 78(1), 78–86, doi:10.1785/gssrl.78.1.78.
- Hirabayashi, T., K. Ito, and T. Yoshii (1992), Multifractal analysis of earthquakes, *Pure Appl. Geophys.*, 138(4), 591–610, doi:10.1007/BF00876340.
- Hirata, T. (1989), Fractal dimension of fault systems in Japan: Fractal structure in rock fracture geometry at various scales, *Pure Appl. Geophys. PAGEOPH*, 131(1-2), 157–170, doi:10.1007/BF00874485.
- Howell, B. F. (1985), On the effect of too small a data base on earthquake frequency diagrams, *Bull. Seismol. Soc. Am.*, 75(4), 1205–1207.
- Imoto, M. (1987), A Bayesian method for estimating earthquake magnitude distribution and changes in the distribution with time and space in New Zealand, *New Zeal. J. Geol. Geophys.*, 30(2), 103–116, doi:10.1080/00288306.1987.10422177.
- Imoto, M. (1991), Changes in the magnitude—frequency b-value prior to large ($M \geq 6.0$) earthquakes in Japan, *Tectonophysics*, 193(4), 311–325, doi:10.1016/0040-1951(91)90340-X.
- Imoto, M., D. A. Rhoades, H. Fujiwara, and N. Yamamoto (2011), Conventional N-, L-, and R-tests of earthquake forecasting models without simulated catalogs, *Earth, Planets Sp.*, 63(3), 275–287, doi:10.5047/eps.2010.08.007.
- Ishimoto, M., and K. Iida (1939), Observations sur les seismes enregistres par le microseismographe construit dernierement, *Bull. Earthq. Res. Inst. Univ. Tokyo*, 17, 443–478.
- Jensen, M., L. Kadanoff, and A. Libchaber (1985), Global universality at the onset of chaos: results of a forced Rayleigh-Bénard experiment, *Phys. Rev. Lett.*, 55(25), 2798–2801.
- Jones, R. H., and R. C. Stewart (1997), A method for determining significant structures in a cloud of earthquakes Simplifying the Earthquake Cloud, *J. Geophys. Res.*, 102(134), 8245–8254.
- Jordan, T. H. (2006), Earthquake Predictability, Brick by Brick, *Seismol. Res. Lett.*, 77(1), 3–6, doi:10.1785/gssrl.77.1.3.
- Kagan, Y. Y. (1999a), Is Earthquake Seismology a Hard, Quantitative Science?, *Pure Appl. Geophys.*, 155(2-4), 233–258, doi:10.1007/s000240050264.
- Kagan, Y. Y. (1999b), Universality of the Seismic Moment-frequency Relation, *Pure Appl. Geophys.*, 155(2-4), 537–573, doi:10.1007/s000240050277.
- Kagan, Y. Y. (2002), Seismic moment distribution revisited: I. Statistical results, *Geophys. J. Int.*, 148(3), 520–541, doi:10.1046/j.1365-246x.2002.01594.x.

Kagan, Y. Y. (2007), Earthquake spatial distribution: the correlation dimension, *Geophys. J. Int.*, 168(3), 1175–1194, doi:10.1111/j.1365-246X.2006.03251.x.

Kagan, Y. Y. (2010), Earthquake size distribution: Power-law with exponent beta=1/2?, *Tectonophysics*, 490(1-2), 103–114, doi:10.1016/j.tecto.2010.04.034.

Kagan, Y. Y., and D. D. Jackson (1994), Long-term probabilistic forecasting of earthquakes, *J. Geophys. Res.*, 99(B7), 13685–13700, doi:10.1029/94JB00500.

Kagan, Y. Y., and . . . Jackson (2007) , Forecast for ≥ 5 Earthquakes in California, , 78(1).

Kagan, Y. Y., and D. D. Jackson (2012), Whole Earth high-resolution earthquake forecasts, *Geophys. J. Int.*, 190(1), 677–686, doi:10.1111/j.1365-246X.2012.05521.x.

Kagan, Y. Y., and L. Knopoff (1980), Spatial distribution of earthquakes: the two-point correlation function, *Geophys. J. Int.*, 62(2), 303–320, doi:10.1111/j.1365-246X.1980.tb04857.x.

Kamer, Y. (2014), Comment on “Systematic survey of high-resolution b value imaging along Californian faults: Inference on asperities” by T. Tormann et al., *J. Geophys. Res. Solid Earth*, 119(7), doi:10.1002/2014JB011147.

Kamer, Y., and S. Hiemer (2013), Comment on “Analysis of the b-values before and after the 23 October 2011 Mw 7.2 Vanuatu, Turkey, earthquake” by the margin, *Tectonophysics*, 608, 1448–1451, doi:10.1016/j.tecto.2013.07.040.

Kamer, Y., and S. Hiemer (2014a), To b or not to b: Top 100 Voronoi Tessellation [2004-2014], https://www.youtube.com/watch?v=5Rc6_0P477A.

Kamer, Y., and S. Hiemer (2014b), To b or not to b: Top 100 Voronoi Tessellations [1984-2014], <https://www.youtube.com/watch?v=RPbFaWf2KAA>.

Kamer, Y., and S. Hiemer (2015), Data-driven spatial b value estimation with applications to California seismicity: To b or not to b, *J. Geophys. Res. Solid Earth*, 120(7), 5191–5214, doi:10.1002/2014JB011510.

Kamer, Y., G. Ouillon, and D. Sornette (2013a), Barycentric fixed-mass method for multifractal analysis, *Phys. Rev. E*, 88(2), 022922.

Kamer, Y., G. Ouillon, D. Sornette, and J. Woessner (2013b), Towards a time-dependent fault network reconstruction, in *40th Workshop of the International School of Geophysics on Properties and Processes of Crustal Fault Zones*, Erice, Italy.

Kamer, Y., G. Ouillon, D. Sornette, and J. Woessner (2014), Fault Network Reconstruction using Agglomerative Clustering: Applications to South Californian Seismicity, in *EGU General Assembly 2014*.

Kamer, Y., G. Ouillon, D. Sornette, and J. Wössner (2015), Condensation of earthquake location distributions: Optimal spatial information encoding and application to multifractal analysis of south Californian seismicity, *Phys. Rev. E*, 92(2), 022808, doi:10.1103/PhysRevE.92.022808.

- Kanamori, H., and D. Anderson (1975), Theoretical basis of some empirical relations in seismology, *Bull. Seismol. Soc. Am.*, 65(5), 1073–1095.
- Kissling, U. Kradolfer, and A. L. Sturzbecher (1995), V. L. ST us er's guide-short introduction, *Inst. Geophys. Swiss Seismol. Serv.*, 25.
- Kisslinger, C., and L. M. Jones (1991), Properties of aftershock sequences in southern California, *J. Geophys. Res.*, 96(B7), 11947, doi:10.1029/91JB01200.
- Lee, Y.-T., D. L. Turcotte, J. R. Holliday, M. K. Sachs, J. B. Rundle, C.-C. Chen, and K. F. Tiampo (2011), Results of the Regional Earthquake Likelihood Models (RELM) test of earthquake forecasts in California., *Proc. Natl. Acad. Sci. U. S. A.*, 108(40), 16533–8, doi:10.1073/pnas.1113481108.
- Li, G., L. Sander, and P. Meakin (1989), Comment on Self-similarity of diffusion-limited aggregates and electrodeposition clusters", *Phys. Rev. Lett.*, 63(12), 2558.
- Lomax, A., J. Virieux, P. Volant, and C. Berge-Thierry (2000), *Advances in Seismic Event Location*, Modern Approaches in Geophysics, edited by C. H. Thurber and N. Rabinowitz, Springer Netherlands, Dordrecht.
- Lomax, A., A. Zollo, P. Capuano, and J. Virieux (2009), Earthquake Location, Direct, Global-Search Methods, in *Encyclopedia of Complexity and Systems Science*, edited by R. A. Meyers, pp. 2449–2473, Springer New York, New York, NY.
- Lomax, A., C. Satriano, and M. Vassallo (2012), Automatic Picker Developments and Optimization: FilterPicker--a Robust, Broadband Picker for Real-Time Seismic Monitoring and Earthquake Early Warning, *Seismol. Res. Lett.*, 83(3), 531–540, doi:10.1785/gssrl.83.3.531.
- Lombardi, A. M., and W. Marzocchi (2010), The ETAS model for daily forecasting of Italian seismicity in the CSEP experiment, *Ann. Geophys.*, 53(3), 155–164, doi:10.4401/ag-4848.
- Lynch, S. (2004), *Dynamical systems with applications using MATLAB*, Birkhauser.
- Mandelbrot, B. (1967), How long is the coast of britain? Statistical self-similarity and fractional dimension., *Science* (80-.), 156(3775), 636–8, doi:10.1126/science.156.3775.636.
- Mandelbrot, B. B. (1977), *Fractals: Form, Chance and Dimension*, W.H.Freeman & Company.
- Mandelbrot, B. B. (1982), *The Fractal Geometry of Nature*, W. H. Freeman and Company.
- Mandelbrot, B. B. (1989), Multifractal measures, especially for the geophysicist, *Pure Appl. Geophys.*, 131(1-2), 5–42, doi:10.1007/BF00874478.
- Márquez-Rámirez, V. H., F. A. Nava Pichardo, and G. Reyes-Dávila (2012), Multifractality in Seismicity Spatial Distributions: Significance and Possible Precursory Applications as Found for Two Cases in Different Tectonic Environments, *Pure Appl. Geophys.*, 169(12), 2091–2105, doi:10.1007/s00024-012-0473-9.

- Marsan, D. (2005), The role of small earthquakes in redistributing crustal elastic stress, *Geophys. J. Int.*, 163(1), 141–151, doi:10.1111/j.1365-246X.2005.02700.x.
- Marsan, D., and O. Lengliné (2008), Extending earthquake's reach through cascading., *Science*, 319(5866), 1076–9, doi:10.1126/science.1148783.
- Marzocchi, W., and L. Sandri (2003), A review and new insights on the estimation of the b-value and its uncertainty, *Ann. Geophys.*, 46(6), 1271–1282.
- Marzocchi, W., J. D. Zechar, and T. H. Jordan (2012), Bayesian Forecast Evaluation and Ensemble Earthquake Forecasting, *Bull. Seismol. Soc. Am.*, 102(6), 2574–2584, doi:10.1785/0120110327.
- Meakin, P. (1983), Formation of fractal clusters and networks by irreversible diffusion-limited aggregation, *Phys. Rev. Lett.*, 51(13), 1119–1122.
- Meakin, P., and S. Havlin (1987), Fluctuations and distributions in random aggregates, *Phys. Rev. A*, 36(9).
- Meier, M.-A., M. J. Werner, J. Woessner, and S. Wiemer (2014), A search for evidence of secondary static stress triggering during the 1992 M w 7.3 Landers, California, earthquake sequence, *J. Geophys. Res. Solid Earth*, 119(4), 3354–3370, doi:10.1002/2013JB010385.
- Michelini, A., and A. Lomax (2004), The effect of velocity structure errors on double-difference earthquake location, *Geophys. Res. Lett.*, 31(9), n/a–n/a, doi:10.1029/2004GL019682.
- Mignan, A. (2012), Functional shape of the earthquake frequency-magnitude distribution and completeness magnitude, *J. Geophys. Res.*, 117(B8), B08302, doi:10.1029/2012JB009347.
- Mignan, A., and J. Woessner (2012), Estimating the magnitude of completeness for earthquake catalogs, *Community Online Resour. Stat. Seism. Anal.*, doi:10.5078/corssa-00180805.
- Mogi, K. (1962), Magnitude–frequency relation for elastic shocks accompanying fractures of various materials and some related problems in earthquakes, *Bull. Earthq. Res. Inst. Tokyo Univ.*, 40, 831–853.
- Molchan, G. (2012), On the testing of seismicity models, *Acta Geophys.*, 60(3), doi:10.2478/s11600-011-0042-0.
- Mori, J., and R. E. Abercrombie (1997), Depth dependence of earthquake frequency-magnitude distributions in California: Implications for rupture initiation, *J. Geophys. Res.*, 102(B7), 15081, doi:10.1029/97JB01356.
- Nagatani, T. (1988), Multifractal structures of mass and growth probability distributions in diffusion-limited aggregation on hierarchical lattices, *Phys. Rev. A*, 38(5), 2632–2640.
- Nanjo, K. Z. (2010), Earthquake forecast models for Italy based on the RI algorithm, *Ann. Geophys.*, 53(3), 117–127, doi:10.4401/ag-4810.

Nazareth, J. J., and E. Hauksson (2004), The Seismogenic Thickness of the Southern California Crust, *Bull. Seismol. Soc. Am.*, 94(3), 940–960, doi:10.1785/0120020129.

Neyman, J., and E. Pearson (1933), On the Problem of the Most Efficient Tests of Statistical Hypotheses, *Philos. Trans. R. Soc. London*, 231, 289–337.

Nicholson, T., M. Sambridge, and Ó. Gudmundsson (2000), On entropy and clustering in earthquake hypocentre distributions, *Geophys. J. Int.*, 142(1), 37–51, doi:10.1046/j.1365-246x.2000.00113.x.

Nittmann, J., H. Stanley, E. Touboul, and G. Daccord (1987), Experimental evidence for multifractality, *Phys. Rev. Lett.*, 58(6), 1987.

Ogata, Y. (1988), Statistical models for earthquake occurrences and residual analysis for point processes, *J. Am. Stat. Assoc.*, 83(401), 9–27.

Ogata, Y., and K. Katsura (1993), Analysis of temporal and spatial heterogeneity of magnitude frequency distribution inferred from earthquake catalogues, *Geophys. J. Int.*, 113(3), 727–738, doi:10.1111/j.1365-246X.1993.tb04663.x.

Ogata, Y., K. Katsura, G. Falcone, K. Nanjo, and J. Zhuang (2013), Comprehensive and Topical Evaluations of Earthquake Forecasts in Terms of Number, Time, Space, and Magnitude, *Bull. Seismol. Soc. Am.*, 103(3), 1692–1708, doi:10.1785/0120120063.

Olami, Z., H. Feder, and K. Christensen (1992), Self-organized criticality in a continuous, nonconservative cellular automaton modeling earthquakes, *Phys. Rev. Lett.*, 68(8), 1244–1247, doi:10.1103/PhysRevLett.68.1244.

Ouillon, G., and D. Sornette (1996), Unbiased multifractal analysis: Application to fault patterns, *Geophys. Res. Lett.*, 23(23), 3409–3412.

Ouillon, G., and D. Sornette (2011), Segmentation of fault networks determined from spatial clustering of earthquakes, *J. Geophys. Res.*, 116(B2), 1–30, doi:10.1029/2010JB007752.

Ouillon, G., D. Sornette, and C. Castaing (1995a), Organisation of joints and faults from 1-cm to 100-km scales revealed by optimized anisotropic wavelet coefficient method and multifractal analysis, *Nonlinear Process. Geophys.*, 2(3/4), 158–177.

Ouillon, G., D. Sornette, and C. Castaing (1995b), Organisation of joints and faults from 1-cm to 100-km scales revealed by optimized anisotropic wavelet coefficient method and multifractal analysis, *Nonlinear Process. Geophys.*, 2, 158–177.

Ouillon, G., C. Castaing, and D. Sornette (1996), Hierarchical geometry of faulting, *J. Geophys. Res.*, 101(B3), 5477, doi:10.1029/95JB02242.

Ouillon, G., C. Ducorbier, and D. Sornette (2008), Automatic reconstruction of fault networks from seismicity catalogs: Three-dimensional optimal anisotropic dynamic clustering, *J. Geophys. Res.*, 113(B1), 1–15, doi:10.1029/2007JB005032.

Pickering, G., J. M. Bull, and D. J. Sanderson (1995), Sampling power-law distributions, *Tectonophysics*, 248(1-2), 1–20, doi:10.1016/0040-1951(95)00030-Q.

Pisarenko, V. F., and D. Sornette (2003), Characterization of the Frequency of Extreme Earthquake Events by the Generalized Pareto Distribution, *Pure Appl. Geophys.*, 160(12), 2343–2364, doi:10.1007/s00024-003-2397-x.

Pisarenko, V. F., A. Sornette, D. Sornette, and M. V. Rodkin (2008), New Approach to the Characterization of M_{max} and of the Tail of the Distribution of Earthquake Magnitudes, *Pure Appl. Geophys.*, 165(5), 847–888, doi:10.1007/s00024-008-0341-9.

Pisarenko, V. F., D. Sornette, and M. V. Rodkin (2010), Distribution of maximum earthquake magnitudes in future time intervals: application to the seismicity of Japan (1923–2007), *Earth, Planets, Sp.*, 62, 567–578.

Pisarenko, V. F., A. Sornette, D. Sornette, and M. V. Rodkin (2014), Characterization of the Tail of the Distribution of Earthquake Magnitudes by combining the GEV and GPD descriptions of Extreme Value Theory, *Pure Appl. Geophys.*, 171(7).

Plesch, a. et al. (2007), Community Fault Model (CFM) for Southern California, *Bull. Seismol. Soc. Am.*, 97(6), 1793–1802, doi:10.1785/0120050211.

Reasenberg, P. (1985), Second-order moment of central California seismicity, 1969–1982, *J. Geophys. Res.*, 90(B7), 5479, doi:10.1029/JB090iB07p05479.

Renyi, A. (1970), *Probability Theory*, North-Holland, Amsterdam.

Rhoades, D., and D. Dowrick (2000), Effects of magnitude uncertainties on seismic hazard estimates, in *12th World Conf. on Earthquake Engineering*.

Rhoades, D., D. Schorlemmer, M. Gerstenberger, A. Christophersen, J. Zechar, and M. Imoto (2011), Efficient testing of earthquake forecasting models, *Acta Geophys.*, 59(4), doi:10.2478/s11600-011-0013-5.

Rhoades, D. A. (1996), Estimation of the Gutenberg-Richter relation allowing for individual earthquake magnitude uncertainties, *Tectonophysics*, 258(1-4), 71–83, doi:10.1016/0040-1951(95)00182-4.

Rhoades, D. A., M. C. Gerstenberger, A. Christophersen, J. D. Zechar, D. Schorlemmer, M. J. Werner, and T. H. Jordan (2014), Regional Earthquake Likelihood Models II: Information Gains of Multiplicative Hybrids, *Bull. Seismol. Soc. Am.*, 104(6), 3072–3083, doi:10.1785/0120140035.

Richter, C. F. (1935), An instrumental earthquake magnitude scale, *Bull. Seismol. Soc. Am.*

Rissanen, J. (1978), Modeling by shortest data description, *Automatica*, 14(5), 465–471, doi:10.1016/0005-1098(78)90005-5.

Rubin, A. M., D. Gillard, and J.-L. Got (1999), Streaks of microearthquakes along creeping faults, , 400(6745), 635–641, doi:10.1038/23196.

Sachs, M., Y. Lee, and D. Turcotte (2012), Evaluating the RELM Test Results, *Int. J. Geophys.*, 1(8).

Sambridge, M., K. Gallagher, A. Jackson, and P. Rickwood (2006), Trans-dimensional inverse problems, model comparison and the evidence, *Geophys. J. Int.*, 167(2), 528–542, doi:10.1111/j.1365-246X.2006.03155.x.

Sambridge, M., T. Bodin, K. Gallagher, and H. Tkalcic (2013), Transdimensional inference in the geosciences, *Philos. Trans. A. Math. Phys. Eng. Sci.*, 371(1984), 20110547.

Sammis, C., and R. Biegel (1989), Fractals, fault-gouge, and friction, *Pure Appl. Geophys.*, 131(1), 255–271.

Sammis, C. G., and D. Sornette (2002), Positive feedback, memory, and the predictability of earthquakes., *Proc. Natl. Acad. Sci. U. S. A.*, 99 Suppl 1(90001), 2501–8, doi:10.1073/pnas.012580999.

Schapire, R. E. (1990), The strength of weak learnability, *Mach. Learn.*, 5(2), 197–227, doi:10.1007/BF00116037.

Scholz, C. (1968), The frequency-magnitude relation of microfracturing in rock and its relation to earthquakes, *Bull. Seismol. Soc. Am.*

Schorlemmer, D., and S. Wiemer (2005), Microseismicity data forecast rupture area, *Nature*, 434(7037), E1–2; discussion E2, doi:10.1038/nature03581.

Schorlemmer, D., G. Neri, S. Wiemer, and A. Mostaccio (2003), Stability and significance tests for b-value anomalies: Example from the Tyrrhenian Sea, *Geophys. Res. Lett.*, 30(16), SDE 3–1–SDE 3–4, doi:10.1029/2003GL017335.

Schorlemmer, D., S. Wiemer, and M. Wyss (2004), Earthquake statistics at Parkfield: 1. Stationarity of b values, *J. Geophys. Res.*, 109(B12), B12307, doi:10.1029/2004JB003234.

Schorlemmer, D., S. Wiemer, and M. Wyss (2005), Variations in earthquake-size distribution across different stress regimes., *Nature*, 437(7058), 539–42, doi:10.1038/nature04094.

Schorlemmer, D., M. C. Gerstenberger, S. Wiemer, D. D. Jackson, and D. A. Rhoades (2007), Earthquake Likelihood Model Testing, *Seismol. Res. Lett.*, 78(1), 17–29, doi:10.1785/gssrl.78.1.17.

Schwarz, G. (1978), Estimating the Dimension of a Model, *Ann. Stat.*, 6(2), 461–464.

Seidel, D. J., Lanzante, and J. R. (2004), An assessment of three alternatives to linear trends for characterizing global atmospheric temperature changes, *J. Geophys. Res.*, 109(D14), D14108, doi:10.1029/2003JD004414.

Shapiro, S. A., O. S. Krüger, and C. Dinske (2013), Probability of inducing given-magnitude earthquakes by perturbing finite volumes of rocks, *J. Geophys. Res. Solid Earth*, 118(7), 3557–3575, doi:10.1002/jgrb.50264.

Shcherbakov, R., D. L. Turcotte, and J. B. Rundle (2005), Aftershock Statistics, *Pure Appl. Geophys.*, 162(6-7), 1051–1076, doi:10.1007/s00024-004-2661-8.

Shearer, P. M. (2002), Parallel fault strands at 9-km depth resolved on the Imperial Fault, Southern California, *Geophys. Res. Lett.*, 29(14), 1674, doi:10.1029/2002GL015302.

Shebalin, P. N., C. Narteau, J. Zechar, and M. Holschneider (2014), Combining earthquake forecasts using differential probability gains, *Earth, Planets Sp.*, 66(1), 37, doi:10.1186/1880-5981-66-37.

Shi, Y., and B. Bolt (1982), The standard error of the magnitude-frequency b value, *Bull. Seismol. Soc. Am.*, 72(5), 1677–1687.

Silver, E. (1971), Tectonics of the Mendocino triple junction, *Geol. Soc. Am. Bull.*, 82(11), 2965–2978.

Silverman, B. W. (1986), *Density Estimation for Statistics and Data Analysis*, CRC Press.

Smyth, C., M. Yamada, and J. Mori (2012), Earthquake forecast enrichment scores, *Res. Geophys.*, 2(1), 2, doi:10.4081/rg.2012.e2.

Sornette, D. (1998), Discrete-scale invariance and complex dimensions, *Phys. Rep.*, 297(5), 239–270.

Sornette, D., and A. Helmstetter (2002), Occurrence of Finite-Time Singularities in Epidemic Models of Rupture, Earthquakes, and Starquakes, *Phys. Rev. Lett.*, 89(15), 158501, doi:10.1103/PhysRevLett.89.158501.

Sornette, D., and M. Werner (2009), Seismicity, Statistical Physics Approaches to, in *Encyclopedia of Complexity and Systems Science*, edited by R. A. Meyers, pp. 7872–7891, Springer New York, New York, NY.

Sornette, D., and M. J. Werner (2005), Apparent clustering and apparent background earthquakes biased by undetected seismicity, *J. Geophys. Res.*, 110(B9), B09303, doi:10.1029/2005JB003621.

Sornette, D., L. Knopoff, Y. Y. Kagan, and C. Vanneste (1996), Rank-ordering statistics of extreme events: Application to the distribution of large earthquakes, *J. Geophys. Res.*, 101(B6), 13883, doi:10.1029/96JB00177.

Di Stefano, R., F. Aldersons, E. Kissling, P. Baccheschi, C. Chiarabba, and D. Giardini (2006), Automatic seismic phase picking and consistent observation error assessment: application to the Italian seismicity, *Geophys. J. Int.*, 165(1), 121–134, doi:10.1111/j.1365-246X.2005.02799.x.

Stock, C., and E. Smith (2002), Adaptive Kernel Estimation and Continuous Probability Representation of Historical Earthquake Catalogs, *Bull. Seismol. Soc. Am.*, 92(3), 904–912, doi:10.1785/0120000233.

- Stumpf, M., and M. Porter (2012), Critical truths about power laws, *Science* (80-.), 335(6069), 665–6, doi:10.1126/science.1216142.
- Taroni, M., J. D. Zechar, and W. Marzocchi (2013), Assessing annual global M6+ seismicity forecasts, *Geophys. J. Int.*, 196(1), 422–431, doi:10.1093/gji/ggt369.
- Tél, T., Á. Fülöp, and T. Vicsek (1989), Determination of fractal dimensions for geometrical multifractals, *Physica A*, 159, 155–166.
- Tence, M., J. P. Chevalier, and R. Jullien (1986), On the measurement of the fractal dimension of aggregated particles by electron microscopy: experimental method, corrections and comparison with numerical models, *J. Phys.*, 47(1986), 1989–1998.
- Theiler, J. (1990), Estimating fractal dimension, *J. Opt. Soc. Am. A*, 7(6), 1055, doi:10.1364/JOSAA.7.001055.
- Tinti, S., and F. Mulargia (1987), Confidence intervals of b values for grouped magnitudes, *Bull. Seismol. Soc. Am.*, 77(6), 2125–2134.
- Torcini, A., A. Politi, P. Puccioni, and . . . 'Alessandro (1991), Fractal dimension of spatially extended systems, *Phys. D*, 53, 85–101.
- Tormann, T., and S. Wiemer (2014), Reply to Comment by Kamer on “Systematic survey of highresolution b value imaging along Californian faults: Inference on asperities,” *J. Geophys. Res. Solid Earth*, 119(7), 5834–5837, doi:10.1002/2014JB011269.
- Tormann, T., S. Wiemer, and A. Mignan (2014), Systematic survey of high-resolution b -value imaging along Californian faults: inference on asperities, *J. Geophys. Res. Solid Earth*, n/a–n/a, doi:10.1002/2013JB010867.
- Tricot, C. (1982), Two definitions of fractional dimension, *Math. Proc. Cambridge Philos. Soc.*, 91(01), 57, doi:10.1017/S0305004100059119.
- Uhrhammer, R. (1986), Characteristics of northern and southern California seismicity, *Earthq. Notes*, 57(21).
- 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.
- Utsu, T. (1966), A statistical significance test of the difference in b-value between two earthquake groups, *J. Phys. Earth*, 14(2), 37–40.
- Utsu, T. (1992), On seismicity, *Rep. Jt. Res. Inst. Stat. Math.*, 34, 139–157.
- Vicsek, T. (1990), Mass multifractals, *Physica A*, 168, 490–497.
- Vicsek, T., F. Family, and P. Meakin (1990), Multifractal geometry of diffusion-limited aggregates, *Europhys. Lett.*, 12(3), 217–222.

Voronoi, G. F. (1908), Nouvelles applications des paramètres continus à la théorie de formes quadratiques, *J. für die reine und Angew. Math.*, 134, 198 – 287.

Waldhauser, F., and W. L. Ellsworth (2000), A Double-difference Earthquake location algorithm: Method and application to the Northern Hayward Fault, California, *Bull. Seismol. Soc. Am.*, 90(6), 1353–1368.

Waldhauser, F., W. L. Ellsworth, and A. Cole (1999), Slip-parallel seismic lineations on the Northern Hayward Fault, California, *Geophys. Res. Lett.*, 26(23), 3525–3528, doi:10.1029/1999GL010462.

Wang, Y. (2013), Automatic reconstruction of fault networks from seismicity catalogs including location uncertainty, ETH Zurich.

Wang, Y., G. Ouillon, J. Woessner, D. Sornette, and S. Husen (2013), Automatic reconstruction of fault networks from seismicity catalogs including location uncertainty, *J. Geophys. Res. Solid Earth*, 118(11), 5956–5975, doi:10.1002/2013JB010164.

Ward, J. H. J. (1963), Hierarchical Grouping to Optimize an Objective Function, *J. Am. Stat. Assoc.*, 58(301), 236–244.

Weitz, D., and M. Oliveria (1984), Fractal structures formed by kinetic aggregation of aqueous gold colloids, *Phys. Rev. Lett.*, 52(16), 1433–1437.

Wells, D. L., and K. J. Coppersmith (1994), New empirical relationships among magnitude, rupture length, rupture width, rupture area, and surface displacement, *Bull. Seismol. Soc. Am.*, 84(4), 974–1002.

Werner, M. J., and D. Sornette (2008), Magnitude uncertainties impact seismic rate estimates, forecasts, and predictability experiments, *J. Geophys. Res.*, 113(B8), B08302, doi:10.1029/2007JB005427.

Werner, M. J., A. Helmstetter, D. D. Jackson, Y. Y. Kagan, and S. Wiemer (2010), Adaptively smoothed seismicity earthquake forecasts for Italy, *Ann. Geophys.*, 53(3), 107–116, doi:10.4401/ag-4839.

Wiemer, S., and K. Katsumata (1999), Spatial variability of seismicity parameters in aftershock zones, *J. Geophys. Res.*, 104(B6), 13135, doi:10.1029/1999JB900032.

Wiemer, S., and D. Schorlemmer (2007), ALM: An asperity-based likelihood model for California, *Seismol. Res. Lett.*, 78(1), 134–140.

Wiemer, S., and M. Wyss (1997), Mapping the frequency-magnitude distribution in asperities: An improved technique to calculate recurrence times?, *J. Geophys. Res.*, 102(B7), 15115, doi:10.1029/97JB00726.

Wiemer, S., and . W yss (2000), i nimum agnitude of Completeness in a rthqua e Catalogs : Examples from Alaska , the Western United States , and Japan , , (August), 859–869.

- Wiemer, S., and M. Wyss (2002), Mapping spatial variability of the frequency-magnitude distribution of earthquakes, *Adv. Geophys.*, **45**, 259–302.
- Wilks, S. S. (1938), The Large-Sample Distribution of the Likelihood Ratio for Testing Composite Hypotheses, *Ann. Math. Stat.*, **9**(1), 60–62.
- Woessner, J., and S. Wiemer (2005), Assessing the Quality of Earthquake Catalogues: Estimating the Magnitude of Completeness and Its Uncertainty, *Bull. Seismol. Soc. Am.*, **95**(2), 684–698, doi:10.1785/0120040007.
- Woo, G. (1996), Kernel estimation methods for seismic hazard area source modeling, *Bull. Seismol. Soc. Am.*, **86**(2), 353–362.
- Wyss, M., A. Hasegawa, S. Wiemer, and N. Umino (1999), Quantitative mapping of precursory seismic quiescence before the 1989, M 7.1 off-Sanriku earthquake, Japan, *Ann. Di Geofis.*, **42**(5), 851–869.
- Wyss, M., D. Schorlemmer, and S. Wiemer (2000), Mapping asperities by minima of local recurrence time: San Jacinto-Elsinore fault zones, *J. Geophys. Res.*, **105**, 7829–7844.
- Wyss, M., C. G. Sammis, R. M. Nadeau, and S. Wiemer (2004), Fractal dimension and b-value on creeping and locked patches of the San Andreas fault near Parkfield, California, *Bull. Seismol. Soc. Am.*, **94**(2), 410–421, doi:10.1785/0120030054.
- Zamani, A., and M. Agh-Atabai (2011), Multifractal analysis of the spatial distribution of earthquake epicenters in the Zagros and Alborz-Kopeh Dagh regions of Iran, *Iran. J. Sci. Technol.*, **35**, 39–51.
- Zechar, J. D., and T. H. Jordan (2010a), Simple smoothed seismicity earthquake forecasts for Italy, *Ann. Geophys.*, **53**(3), 99–105, doi:10.4401/ag-4845.
- Zechar, J. D., and T. H. Jordan (2010b), Simple smoothed seismicity earthquake forecasts for Italy, *Ann. Geophys.*, **53**(3), 99–105, doi:10.4401/ag-4845.
- Zechar, J. D., D. Schorlemmer, M. Liukis, J. Yu, F. Euchner, P. J. Maechling, and T. H. Jordan (2009), The Collaboratory for the Study of Earthquake Predictability perspective on computational earthquake science, *Concurr. Comput. Pract. Exp.*, **22**(12), 1836–1847, doi:10.1002/cpe.1519.
- Zechar, J. D., M. C. Gerstenberger, and D. A. Rhoades (2010), Likelihood-based tests for evaluating space-rate-magnitude earthquake forecasts, *Bull. Seismol. Soc. Am.*, **100**(3), 1184–1195, doi:10.1785/0120090192.
- Zechar, J. D., D. Schorlemmer, M. J. Werner, M. C. Gerstenberger, D. A. Rhoades, and T. H. Jordan (2013), Regional Earthquake Likelihood Models I: First-Order Results, *Bull. Seismol. Soc. Am.*, **103**(2A), 787–798, doi:10.1785/0120120186.