

Copyright is owned by the Author of the thesis. Permission is given for a copy to be downloaded by an individual for the purpose of research and private study only. The thesis may not be reproduced elsewhere without the permission of the Author.

Statistical Models for Earthquakes Incorporating Ancillary Data

A thesis presented in partial fulfilment of the requirements for the degree of

Doctor of Philosophy
in
Statistics

at Massey University, Palmerston North,
New Zealand.

Ting Wang

2010

In memory of Professor Ma Li

Abstract

This thesis consists of two parts. The first part proposes a new model – the Markov-modulated Hawkes process with stepwise decay (MMHPSD) to investigate the seismicity rate. The MMHPSD is a self-exciting process which switches among different states, in each of which the process has distinguishable background seismicity and decay rates. Parameter estimation is developed via the expectation maximization algorithm. The model is applied to data from the Landers earthquake sequence, demonstrating that it is useful for modelling changes in the temporal patterns of seismicity. The states in the model can capture the behavior of main shocks, large aftershocks, secondary aftershocks and a period of quiescence with different background rates and decay rates. The state transitions can then explain the seismicity rate changes and help indicate if there is any seismicity shadow or relative quiescence.

The second part of this thesis develops statistical methods to examine earthquake sequences possessing ancillary data, in this case groundwater level data or GPS measurements of deformation. For the former, signals from groundwater level data at Tangshan Well, China, are extracted for the period from 2002 to 2005 using a moving window method. A number of different statistical techniques are used to detect and quantify coseismic responses to P, S, Love and Rayleigh wave arrivals. The P phase arrivals appear to trigger identifiable oscillations in groundwater level, whereas the Rayleigh waves amplify the water level movement. Identifiable coseismic responses are found for approximately 40 percent of magnitude 6+ earthquakes worldwide. A threshold in the relationship between earthquake magnitude and well–epicenter distance is also found, satisfied by 97% of the identified coseismic responses, above which coseismic changes in groundwater level at Tangshan Well are most likely.

A non-linear filter measuring short-term deformation rate changes is introduced to extract signals from GPS data. For two case studies of a) deep earthquakes in central North Island, New Zealand, and b) shallow earthquakes in Southern California, a hidden Markov model (HMM) is fitted to the output from the filter. Mutual information analysis indicates that the state having the largest variation of deformation rate contains precursory information that indicates an elevated probability for earthquake occurrence.

Acknowledgements

I wish to acknowledge all those who have helped me through the three years of my Ph.D. study. First and foremost, I would like to express my gratitude to my supervisor Mark Bebbington. Three years ago, I followed my dream and came to Massey University to become a research professional with a special interest in probabilistic models for geophysical hazard but with little background knowledge about geophysics. Without Mark's kind guidance and motivation, nothing could have been accomplished. It's from Mark that I learnt first how to crawl, then how to walk and finally how to run in academic research. During the three years, I have never found a proper way to thank him. In completion of this thesis I would like to thank Mark for being patient when I was not making good progress, inspiring and motivating me when I got lost, kindly encouraging me when I was frustrated, making my research a top priority, and showing me the beauty of the New Zealand countryside.

I am very grateful to my co-supervisors David Harte and David Vere-Jones for their continuing support, guidance and encouragement, as well as all their helpful discussions for this research. David H always came to my rescue whenever I was faced to programming issues and patiently explained everything in detail. I also want to thank him for teaching me how to write an R package. David VJ was also inspirational for my undertaking of this adventure and provided insightful comments for this work. I would also like to thank David VJ for inviting me to the King Lear play, to the benefit of my cultural education, and encouraging me to enjoy the world beyond academia. My gratitude goes to all three supervisors for thoroughly reviewing and improving this manuscript.

My special thanks also go to two of the most important people in my life, Professor Ma Li and Jill Bebbington. If it is Mark who made me love statistics-related geophysics, it was Professor Ma Li who inspired my first interest in this field. She was always willing to ask for and provided the best for her students. For her having been truly inspirational, I would like to dedicate this thesis to her memory. By having Mark as my principal supervisor, I was actually very lucky to find myself two teachers. Jill taught me how to live a valuable life while Mark was trying to inspire and improve my research skills. If there was a Nobel Prize in cooking and raising a family, I wish Jill would be awarded one. My love also goes to my two little friends, Anna and Craig Bebbington for all the joy they have brought to me.

I also want to acknowledge all those who have provided insightful comments on my research or otherwise helped me, particularly Marco Brenna, Cheryl and Ray Brownrigg, Professor Huizhong

Chen, Robert Davies, Alistair Gray, Jianping Huang, Takaki Iwata, Professor Chin-Diew Lai, Wenjing Li, Nicholas Look, Dejun Luo, Tim O’Dea, Professor Yoshihiko Ogata, Patrick Rynhart, Professor Martha Kane Savage, Peter Thomson, Baojun Yin, Jiancang Zhuang and Walter Zucchini, and my apologies to all those whom I failed to mention. I would also like to express my appreciation to Takaki Iwata, Professor Chin-Diew Lai, and Roger Littlejohn for their thorough review of my thesis.

It was indeed delightful to work with a group of ever friendly statisticians, the Statistics Research Group at Massey University. I would like to thank all of them for all their help and encouragement. The postgraduate students in the Stats PostGrads Office AHA2.74 accompanied me through these three years in the office and made my student life much more enjoyable. I am grateful to all of them for cheering me up when I was down and offering help whenever I needed it.

My appreciation also goes to those friends who have always accompanied and supported me outside of the academic world, Valentine Borges, Boon Feei Chong, Beibei Feng, Kasey Gordon, Marissa Isidro, Wirya Khim, Xiaojing Peng, Bandeth Ros, Zhongwei Xing, Zhiwen Yu, and many others.

This research project was financially supported by Marsden Fund administrated by the Royal Society of New Zealand. I am grateful to the Tangshan Earthquake Administration for providing the well data.

Last but not the least, to my parents, my brother and my sister-in-law, goes my deepest love for always being there to support me.

Contents

Abstract	i
Acknowledgements	iii
List of Figures	ix
List of Tables	xv
Glossary	xvii
1 Introduction	1
1.1 Motivation	1
1.1.1 Earthquake Cycles	1
1.1.2 Clustering	2
1.1.3 Ancillary Data	4
1.2 Overview	5
I HIDDEN MARKOV TYPE MODELS FOR EARTHQUAKES	9
2 A General Class of Discrete-time Models	11
2.1 A General Formulation for Discrete-time Models	11
2.2 Models without Ancillary Data	12
2.3 Transition Probabilities Depending on Ancillary Data	17
2.4 Observation Mean Depending on Ancillary Data	21
3 Markov-modulated Hawkes Process with Stepwise Decay	27
3.1 Introduction	27

3.2	Markov-modulated Hawkes Process with Stepwise Decay	28
3.2.1	Hawkes Process with Stepwise Decay	28
3.2.2	Markov-modulated Hawkes Process with Stepwise Decay	29
3.3	The Complete Likelihood	33
3.4	An EM Algorithm for Parameter Estimation	34
3.4.1	Implementation	36
3.5	Goodness-of-fit	42
3.5.1	Estimated Intensity Function of the Observed Process	42
3.5.2	Residual Analysis	43
3.6	Simulation Algorithm	45
4	Simulation Study and an Application of the MMHPSD	47
4.1	Introduction	47
4.2	Evaluation of Parameter Estimation Algorithm	49
4.3	Simulation Study Using a Simulated ETAS Sequence	50
4.3.1	Fitting MMHPSDs to Simulated ETAS Sequence	51
4.3.2	Consistency of the Parameter Estimation	58
4.4	MMHPSD Investigation of Earthquake Data around Landers	59
4.4.1	Earthquake Data around Landers	59
4.4.2	Exploratory Data Analysis Using MMHPSDs	61
4.4.3	Comparison with the ETAS Model	71
4.5	Conclusion and Discussion	73
II	STATISTICAL ANALYSIS OF EARTHQUAKES	
	WITH ANCILLARY DATA	77
5	Statistics on Association Between Series of Events	79
5.1	Introduction	79
5.2	Coherence	81
5.3	Mutual Information	83
5.4	Ogata's Lin-Lin Model	85

6	Transient Coseismic Responses at Tangshan Well	89
6.1	Introduction	89
6.2	Data	91
6.2.1	Well Data Levelling	95
6.2.2	Earthquake Data	97
6.3	Identifying Signals in Well Data	97
6.3.1	Exponential Decay of Well Oscillations	97
6.3.2	A Moving-window Detection Algorithm	99
6.3.3	Determining the Detection Threshold	101
6.4	Tests of Dependence Between Earthquakes and Well Responses	102
6.4.1	Clustering Tendencies in Well Signals and the Earthquake Catalogue	102
6.4.2	Correlations Between Earthquakes and Lagged Well Responses: (a) Coherence Analysis	104
6.4.3	Correlations Between Earthquakes and Lagged Well Responses: (b) Mutual Information Analysis	105
6.4.4	Earthquakes as an Explanatory Variable for Well Responses	107
6.5	Coseismic Responses and Detection Probability	108
6.5.1	A Magnitude-distance Threshold for Detection	108
6.5.2	Quantifying the Effects of Magnitude, Distance, Depth and Azimuth	113
6.6	An Exploratory Analysis of Earthquake–well Signal Interaction	116
6.6.1	Correlations Between Earthquake and Well Signal Characteristics	117
6.6.2	Metamodels for Delay, Amplitude and Length of Well Signals	118
6.7	Discussion	120
7	HMM and MI on GPS Measurements and Earthquakes	125
7.1	Introduction	125
7.2	Methodology	129
7.2.1	Hidden Markov Model and Mutual Information Analysis	129
7.2.2	Probability Forecast Using a Logistic Probability Model	131
7.3	Data from Central North Island, New Zealand	132
7.3.1	NHMM and MMGLM Analyses of the Earthquake Data with Ancillary GPS Measurements	136
7.3.2	Hidden Markov Model and Mutual Information Analyses	136

7.3.3	Probability Forecast Using Logistic Probability Model	158
7.4	Data from Southern California	167
7.4.1	Hidden Markov Model and Mutual Information Analyses	169
7.4.2	Probability Forecast Using Logistic Probability Model	179
7.5	Conclusion and Discussion	183
8	Conclusions and Future Research	185
8.1	Conclusions	185
8.2	Future Research	188
8.2.1	Markov-modulated Hawkes Processes with Time-varying Decay	189
8.2.2	MMHPSD with Marks	190
8.2.3	NHMM and MMGLM Analysis of Earthquakes with Ancillary GPS Data	191
8.2.4	Testing of HMM Analysis of GPS Data for Earthquake Forecasting	193
	Appendix	195
	A. Dispersion Test (Cox and Lewis, 1966)	195
	B. Logistic Regression Analysis	195
	C. Multiple Regression Analysis	197
	D. Stationary Distribution of a Markov Chain	197
	Bibliography	199

List of Figures

4.1	Cumulative curve of a simulated MMHPSD sequence	49
4.2	Histograms of the parameter estimates for the simulated MMHPSDs	50
4.3	Cumulative curve of the simulated ETAS events.	52
4.4	Log-scaled power-law decay rate and log-scaled exponential decay rate	54
4.5	Estimated probability of the hidden state occupying State 2, and estimated intensity function for the two-state MMHPSD fitted to the simulated ETAS events	56
4.6	Estimated probability of the hidden state occupying each state, and estimated intensity function for the three-state MMHPSD fitted to the simulated ETAS events	57
4.7	Estimated probability of the hidden state occupying each state, and estimated intensity function for the four-state MMHPSD fitted to the simulated ETAS events	58
4.8	Estimated intensity function for the MMHPSDs fitted to the simulated ETAS events with 2, 3 and 4 hidden states minus the true ETAS intensity function	59
4.9	Boxplot of the estimated parameters for the consistency test of the parameter estimation for the MMHPSD	60
4.10	Frequency–magnitude plot of earthquakes around Landers	61
4.11	Location map of the earthquakes around Landers	62
4.12	Cumulative curve of the earthquakes around Landers	63
4.13	Cumulative number of the residual process versus the transformed time for the fitted MMHPSD with 3 states to the earthquakes around Landers	64
4.14	Cumulative number of the residual process versus the transformed time for the fitted MMHPSD with 4 states to the Landers earthquakes	65
4.15	Kolmogorov–Smirnov test for the inter-arrival times of the residual process for the MMHPSD with 4 states fitted to the earthquakes around Landers	66
4.16	Test of correlation between inter-arrival times of the residual process for the MMHPSD with 4 states fitted to the earthquakes around Landers	67

4.17	Test of correlation between inter-arrival times of the residual process for the MMH-PSD with 4 states fitted to the earthquakes around Landers	67
4.18	Illustration of the two-state MMHPSD fitted to the data around Landers	68
4.19	Illustration of the three-state MMHPSD fitted to the data around Landers	69
4.20	Illustration of the four-state MMHPSD fitted to the data around Landers	70
4.21	Cumulative curve of the simulated MMHPSD events using estimated parameters for the four-state MMHPSD fitted to the earthquakes around Landers	71
4.22	Illustration of the ETAS model fitted to the data around Landers.	72
4.23	Cumulative number of the residual process versus the transformed time for the ETAS model fitted to the data around Landers	73
6.1	Tangshan Well and its columnar section	91
6.2	Groundwater level record at Tangshan Well from January 1, 2002 to December 31, 2005	93
6.3	Example of the oscillations of the groundwater level at Tangshan Well	95
6.4	Cumulative distributions of the first differences of the groundwater level and of the lengths of the non-missing intervals of the data at Tangshan Well	96
6.5	Two examples of cumulative first difference data	98
6.6	Illustration of calculating the weighted variance	100
6.7	Poisson tests for well signals and earthquakes	103
6.8	Time-occurrence histograms of the earthquake occurrence times, well signals, and the survival functions of the well signal inter-event time and well signal length . . .	103
6.9	Coherence plots for well signals and the earliest P phase, the earliest S phase, Love wave and Rayleigh wave arrival times	104
6.10	Mutual information for well signals and the earliest P phase, the earliest S phase, Love wave and Rayleigh wave arrival times	105
6.11	Example of an S phase arrival initiating well oscillations	106
6.12	Estimated Lin–Lin response functions of well signals with the seismic phase arrival times	107
6.13	Magnitude versus well–epicenter distance plot	109
6.14	Global map and earthquakes with minimum magnitude 6.0 from January 1, 2002 to September 30, 2005	111
6.15	Enlarged version of the squared region in Figure 6.14	112

6.16	Magnitude versus well–epicenter distance plot	113
6.17	Magnitude versus well–epicenter distance plot	114
6.18	A schematic illustration of the azimuth angle from an earthquake to the well	115
6.19	Proportion of earthquakes with coseismic responses and the fitted probability that an earthquake produces a coseismic response	116
6.20	Well signals’ lengths versus well–epicenter distances	117
7.1	Classification results for a seven-state HMM applied to the GPS data set collected in the city of Claremont, California	126
7.2	GPS measurements of deformation at Alice Springs, Australia	127
7.3	Digital elevation map of central North Island, New Zealand and its tectonic settings	133
7.4	Location map for the GPS stations and earthquakes around Lake Taupo	134
7.5	Frequency–magnitude plot for the earthquakes around Taupo	135
7.6	Earthquake occurrences and Viterbi paths from the HMM, NHMM and MMGLM analyses	137
7.7	GPS movements of HAMT, NPLY and WANG relative to TAUP	137
7.8	Mutual information for the largest state in each of the HMMs with up to 10 hidden states for the entire data (Taupo)	140
7.9	Mutual information for the state which accounts for the north–component movement in each of the HMMs with 7, 8, 9 and 10 hidden states for the entire data (Taupo) .	141
7.10	Illustration of state transitions for HMMs on the data around Taupo	142
7.11	Sojourn time distribution for the fitted five-state HMM to the entire data around Taupo	143
7.12	Histograms of the deformation rate ranges along with the estimated mixed normal density (Taupo)	145
7.13	Scatter plot of the deformation rate ranges of the east component versus that of the north component (Taupo)	145
7.14	Scatter plot of the deformation rate ranges of the up component versus that of the north component (Taupo)	146
7.15	Scatter plot of the deformation rate ranges of the up component versus that of the east component (Taupo)	146
7.16	Viterbi path and trend changes of the GPS movements in the previous 10 days (Taupo)	147
7.17	Mutual information between the Viterbi path of the fitted five-state HMM and the earthquakes for the entire data (Taupo)	151

7.18	Mutual information between the Viterbi path of the fitted five-state HMM and the earthquakes for the second half of the data (Taupo)	152
7.19	Mutual information between the Viterbi path of the fitted five-state HMM and the earthquakes for the first half of the data (Taupo)	153
7.20	Illustration of state transitions for HMMs on data around Taupo, using the north and east components only	154
7.21	Mutual information between the Viterbi path of the fitted five-state HMM and the earthquakes for data around Taupo only using the north and east components	155
7.22	Scatter plots of the deformation rate ranges of the east component versus that of the north component (Taupo)	156
7.23	Viterbi path and trend changes of the GPS movements (north and east components) in the previous 10 days (Taupo)	157
7.24	Sojourn time distribution for the fitted five-state HMM using the north and east components (Taupo)	158
7.25	Mutual information between the Viterbi path of the fitted five-state HMM and the earthquakes for the Euclidean distance calculated from the north and east components around Taupo	159
7.26	Illustration of probability forecast of earthquakes around Taupo with minimum magnitude 5.1 using the entire data	163
7.27	Illustration of probability forecast of earthquakes around Taupo with minimum magnitude 5.0 using the entire data	164
7.28	Scatter plot of the time to the next event versus the sojourn time of the HMM in State 5 (Taupo)	164
7.29	Scatter plot of the time to the next event versus the minimum distance of the deformation rate ranges to the origin when the HMM is sojourning in State 5 (Taupo)	165
7.30	Illustration of probability forecast of earthquakes around Taupo with minimum magnitude 5.1 using the north and east components only (Taupo)	166
7.31	Location map for the GPS stations CHIL and LBC2 and earthquakes in Southern California	167
7.32	Frequency–magnitude plot for the earthquakes in Southern California	168
7.33	Baseline between the GPS stations CHIL and LBC2	169
7.34	Illustration of state transitions for HMMs on data in Southern California	170

7.35	Scatter plots of the trend ranges of the east components versus that of the north component (Southern California)	171
7.36	Scatter plots of the trend ranges of the up components versus that of the north component (Southern California)	172
7.37	Scatter plots of the trend ranges of the up components versus that of the east component (Southern California)	172
7.38	Viterbi path and trend changes of the GPS movements in the previous 10 days (Southern California)	174
7.39	Mutual information between the Viterbi path of the fitted five-state HMM and the earthquakes for data in Southern California	174
7.40	Sojourn time distribution for the fitted five-state HMM (Southern California)	175
7.41	Illustration of state transitions for HMMs on data in Southern California, using the north and east components only	176
7.42	Mutual information between the Viterbi path of the fitted five-state HMM and the earthquakes for data in Southern California only using the north and east components	177
7.43	Scatter plot of the trend ranges of the east component versus that of the north component (Southern California)	178
7.44	Viterbi path and trend changes of the GPS movements (north and east components) in the previous 10 days (Southern California)	179
7.45	Sojourn time distribution for the fitted five-state HMM (Southern California)	180
7.46	Illustration of probability forecast of earthquakes in Southern California with minimum magnitude 4.5 using the entire data	181
7.47	Illustration of probability forecast of earthquakes in Southern California with minimum magnitude 4.5 using the north and east components only	182

List of Tables

4.1	The sample means and standard deviations of the estimated parameters for the simulated MMHPSDs	50
4.2	Estimated parameters of the MMHPSDs fitted to the simulated ETAS events	53
4.3	Estimated parameters of the MMHPSDs fitted to the data around Landers	65
4.4	Log likelihood and BIC for the MMHPSDs and the ETAS model fitted to the data around Landers	68
4.5	Estimated parameters for the ETAS model fitted to the data around Landers	71
6.1	Global earthquakes obtained from the USGS-NEIC catalogue with minimum magnitude 7.5 from January 1, 2002 to December 31, 2005	94
6.2	The 30 events denoted in Figure 6.13 by circles above the threshold for which there are no coseismic responses identified	110
6.3	Chronology of well signals and seismic wave arrivals	110
6.4	Correlation coefficients between well signal characteristics and earthquake characteristics	118
6.5	Contingency table for the seismic wave arrival immediately preceding the onset of well oscillations cross referencing azimuth	119
7.1	Information of the GPS measurements around Taupo	135
7.2	Log likelihood and BIC for the fitted models to the entire data around Taupo and in Southern California	138
7.3	Proportion of data in the precursory state(s) for data around Taupo	139
7.4	Parameter estimates of the fitted five-state HMM for the entire data around Taupo	141
7.5	Parameter estimates of the fitted five-state HMM to the first half of the GPS data (Taupo)	149

7.6	Parameter estimates of the fitted five-state HMM to the second half of the GPS data (Taupo)	149
7.7	Parameter estimates of the fitted five-state HMM for the case when only the north and east components are considered (Taupo)	150
7.8	Contingency table for probability forecast using all three components (Taupo) . . .	159
7.9	State transition pattern for earthquakes with magnitude 5.1 or larger (Taupo)	160
7.10	Frequency of state transitions for Table 7.9	161
7.11	Contingency table for probability forecast using the transition pattern State 2–5 . .	161
7.12	Average number of entries to each state between two consecutive earthquakes (Taupo)	161
7.13	Logistic regression result for earthquakes with minimum magnitude 5.1 (Taupo) . .	162
7.14	Logistic regression result for earthquakes with minimum magnitude 5.0 (Taupo) . .	162
7.15	Contingency table for probability forecast using only north and east components (Taupo)	165
7.16	Logistic regression result for earthquakes with minimum magnitude 5.1, using north and east components only (Taupo)	166
7.17	Parameter estimates of the fitted five-state HMM (Southern California)	171
7.18	Parameter estimates of the fitted five-state HMM for the case when only the north and east components are considered (Southern California)	177
7.19	Contingency table for probability forecast using all three components (Southern California)	180
7.20	Logistic regression result for earthquakes with minimum magnitude 4.5 (Southern California)	181
7.21	Contingency table for probability forecast using only north and east components (Southern California)	181
7.22	Logistic regression result for earthquakes with minimum magnitude 4.5, using north and east components only (Southern California)	182

Glossary

AMR	Accelerated moment release model
Craton	An old and stable part of the continental lithosphere
ETAS	Epidemic type aftershock sequence model
Declustered data	Earthquake sequences with aftershocks removed
Gutenberg-Richter law	The total number of earthquakes in a population that are larger than or equal to some magnitude M varies as 10^{bM}
HMM	Hidden Markov models
Hypocenter	The point within the earth where an earthquake rupture starts
LSRM	Linked stress release model
MMGLM	Markov-modulated generalized linear model
MMHPSD	Markov-modulated Hawkes process with stepwise decay
MMPP	Markov-modulated Poisson process
NHMM	Nonhomogeneous hidden Markov model
Relative quiescence	A significant decrease of earthquakes compared with the occurrence rate expected from a point-process model for ordinary seismic activity
Seismicity/stress shadows	Seismicity rate decreases
SRM	Stress release model