Development, installation and testing of new models in the New Zealand Earthquake Forecast Testing Centre

D.A. Rhoades
A. Christophersen
M.C. Gerstenberger

GNS Science Consultancy Report 2010/253
November 2010

DISCLAIMER

The data presented in this Report are available to GNS Science for other use from November 2010.

BIBLIOGRAPHIC REFERENCE

Rhoades, D.A.; Gerstenberger, M.C.; Christophersen, A. 2010. Development, installation and testing of new models in the New Zealand Earthquake Forecast Testing Centre, *GNS Science Consultancy Report* CR 2010/253. 55 p.

CONTENTS

FIGURES

TABLES

APPENDICES

NON-TECHNICAL SUMMARY

The New Zealand Earthquake Forecast Testing Centre is one of four existing regional testing centres so far established around the globe to undertake rigorous tests of proposed earthquake forecasting models. The primary purpose of the testing centres is to carry out prospective testing but, because implementation of models for the testing centre is technically demanding, it is also useful to undertake retrospective testing to determine whether models have been properly implemented and installed.

In this study we report on the development and/or implementation for the New Zealand testing centre of six new models, including one new long-term model (a 2010 major revision of the New Zealand national seismic hazard model), three new medium-term models (which are updated at 3-month intervals) and two new short-term models (which are updated daily). Except for one of the new short-term models, these models have all been installed in the testing centre. The intention is to test all classes of model prospectively over an extended period of five or more years.

The new models and ten previously-installed models have been tested retrospectively using two simple statistical tests. The first test compares the total number of earthquakes expected under each model with the number observed over the test period. The second test estimates the overall information gain of one model over another. We use new versions of these tests which are easier to interpret than comparable tests which are part of the standard testing centre software. The retrospective testing has allowed for problems with the implementation of several of the models to be identified now rather than after years of computer-intensive testing.

Fifteen models were retrospectively tested. For seven models there is no indication of incorrect implementation. However, for five models there is a weak indication of incorrect implementation, and for three models, including two of the new models, there is a strong indication of incorrect implementation. The results will be referred back to the modellers, so that the implementation can be corrected if necessary.

The results of retrospective testing have thus been highly instructive in revealing teething problems with implementation of both new and existing models. The incidence of model implementation errors revealed in this study is similar to that found in retrospective testing of models submitted to other regional testing centres.

TECHNICAL SUMMARY

The New Zealand Earthquake Forecast Testing Centre is one of four existing regional testing centres around the globe that are under the umbrella of the Collaboratory for the Study of Earthquake Predictability (CSEP). These centres were established to undertake verifiable and transparent tests of proposed earthquake forecasting models, so that the whole science community can have confidence in the results. The Centre is designed to test regional earthquake likelihood models for shallow earthquakes (h < 40 km) occurring within the New Zealand test region primarily on three timescales (5-years, 3-months and 1-day). In order to reliably measure forecast performance, an extended prospective testing period of five or more years is required for all classes of models. The tests and software used are compatible with other CSEP regional testing centres.

In this study, six new models have been developed, implemented and/or installed for testing in the centre, including two 1-day models for clustering of aftershocks, three new 3-month models – two being elaborations of the Every Earthquake a Precursor According to Scale (EEPAS) model and the other a double-branching process model proposed by Italian researchers – and one new 5-year model – the 2010 revision (NZNSH2010) of the New Zealand national seismic hazard model.

The new and previously existing models have been tested retrospectively using a new presentation of the N-test comparing the expected and actual number of earthquakes, and the recently proposed T-test for estimating the information gain of one model over another. These efficient and straightforward tests give a much clearer indication of possible implementation errors than the standard CSEP test outputs have provided in previous studies. This has allowed for problems with the implementation of several of the models to be identified now rather than after years of computer-intensive testing.

Tests of the 5-year models over 25 years show that the new NZNSH2010 model is significantly more informative than the old NZNSHM model over this period, although it slightly over-estimates the number of earthquakes and appears not to have been implemented over all bins in the test region.

Tests of the 3-month models over a two-year period have shown that one of the new EEPAS models is not correctly implemented in the testing centre, but the other models are performing approximately as expected. However the precision of the estimates of the expected number of earthquakes needs to be checked for three of the models previously installed.

Tests of the short-term (1-day) models over six months have revealed problems with a new implementation of the Short-Term Earthquake Probability (STEP) model in the JAVA programming language. It underestimates the number of earthquakes in the test region. The tests show that neither the original STEP model nor the new JAVA version of this model are properly implemented for the testing centre.

The results of retrospective testing have thus been highly instructive in revealing teething problems with implementation of both new and existing models. The incidence of model implementation errors revealed in this study is similar to that found in retrospective testing of models submitted to other regional testing centres.

1.0 INTRODUCTION

The New Zealand Earthquake Forecast Testing Centre was recently established with support from the EQC Research Foundation (Rhoades et al., 2008; Gerstenberger and Rhoades, 2010), with official prospective testing of models beginning in 2008. The Centre tests regional earthquake likelihood models for shallow earthquakes (h < 40 km) occurring within the New Zealand test region on four different time scales. Short-term models are tested using 24-hour time bins at magnitude M ≥ 4, intermediate-term models are tested using 3-month or 6-month time bins at M ≥ 5, and long-term models are tested using 5-year time bins at M ≥ 5. In order to reliably measure forecast performance, an extended test period of five or more years is required for all classes of models. The tests and software used are compatible with other regional testing centres of the Collaboratory for the Study of Earthquake Predictability (CSEP) (Jackson, 1996; Schorlemmer and Gerstenberger, 2007; Schorlemmer et al., 2007).

The purpose of the regional earthquake forecast testing centres, of which the New Zealand centre is one, is to undertake verifiable and transparent tests of proposed models, so that the whole science community, and not just the individual scientists and groups developing the models, can have confidence in the results. In order to provide verifiability and transparency it is necessary that the forecasts be completely specified on a grid of magnitude-location bins for each test period.

Before any model can undergo testing in a regional testing centre, (at least) three steps are required. These are referred to here as *development*, *implementation* for testing, and *installation* in the testing centre. Each of these steps involves a significant amount of work. The first and largest step – model development – involves the total specification of the model, including optimisation of its parameters. In most cases, this is done without reference to the testing centre grid cells, because it would be computationally inefficient to do it that way. The second step – implementation – involves adapting the model to the testing centre requirements, i.e., computing the expected number of earthquakes in each magnitude-and-location grid cell for any specified time period, given an earthquake catalogue. The third step – installation – involves linking that program into the testing centre software, so it can be run automatically at the beginning of each time step during the tests. Installation can only be done by a close collaboration between the testing centre scientists and the model developers.

Three long-term, two intermediate-term and two short-term forecasting models were initially installed in the testing centre. The five-year models are NZNSHM – the New Zealand National Seismic Hazard Model (Stirling et al., 2002), PPE – a quasi-time-invariant model based on Proximity to Past Earthquakes (Jackson and Kagan, 1999; Rhoades and Evison, 2004), and SUP – a Stationary Uniform Poisson model, included as a model of least information. The intermediate-term models are PPE and EEPAS – a time-varying model, based on the precursory scale increase (Ψ) phenomenon (Evison and Rhoades, 2004) in which Every Earthquake is a Precursor According to Scale (Rhoades and Evison, 2004, 2005, 2006; Console et al. 2006; Rhoades, 2007). The short-term models are STEP – a Short Term Earthquake Probability model (Gerstenberger et al., 2005), and ETAS – a version of the space-time Epidemic Type AfterShock model (Ogata, 1998). In 2009, a Regional Earthquake Likelihood Model developed from the M8 algorithm (M8RELM) by

David Harte and Ray Brownrigg (Harte et al., 2003) was installed in the six-month model class (Gerstenberger et al., 2009).

Since that time, other models have been under development. A new Short-Term Forecasting Model (STFM), which combines a measure of earthquake abundance with STEP, has been developed (Christophersen and Gerstenberger, submitted). In addition, two Italy-based modellers have adapted their double-branching process model (DBPM) (Marzocchi and Lombardi, 2008) to the New Zealand region, and have submitted it for testing as an intermediate-term model. This project includes the installation of these models in the testing centre.

A new software implementation of the STEP model (STEP-JAVA) has also been produced using the Java programming language within the OpenSHA seismic hazard library. The previous implementation in Matlab was dependent on licensed software. Ensuring shared licences would always be available when needed was a problem for the testing centres, and further downstream problems were envisaged with wide distribution of the testing centre software. Therefore the developers of the CSEP system are intent on replacing elements of the system dependent on licensed proprietary software with independently executable code.

A new intermediate-term model – EAS – an elaboration of the EEPAS model to allow for aftershocks of forecasted earthquakes, has also been recently developed (Rhoades, 2009), but not adapted to the New Zealand region. Another variant of the EEPAS model under development is ERDEEP – an Earthquake-Rate-Dependent version of the EEPAS model which embodies the hypothesis that the precursor time T_P and area A_P occupied by the precursory scale increase phenomenon are both dependent on the localised long-term rate of earthquake occurrence (Rhoades, 2008). This project includes further development of the ERDEEP model, and adaptation of the EAS and ERDEEP models to the New Zealand test region, implementation of both models for CSEP testing, and their installation in the testing centre.

A new version of the New Zealand National Seismic Hazard model (NZNSH2010) has been under development for several years and has been finalised this year. It was referred to as NZNSH08 in the project proposal. It includes a modified set of fault sources and associated parameters and a modified distributed-seismicity background model. It is important for this model to be submitted to formal testing so that, over time, the value of the modifications can be quantified. This project includes the implementation of the NZNSH2010 model for CSEP testing and its installation in the testing centre in the long-term (five-year) model class.

Whenever new models are installed, it is helpful to run retrospective tests, using the same tests as will be used to test them prospectively. The main purpose of such tests is to ascertain whether the models are properly implemented, according to their authors' expectations, in CSEP-compatible software. Experience with models submitted to the various regional testing centres shows that many if not most models are improperly implemented or installed at the first attempt (Werner et al., in press). Conforming to the elaborate CSEP testing system places high demands on model developers, and a high error rate is therefore to be expected in initial submissions. Retrospective testing at the outset can thus avoid much wasted effort in prospective testing of models which are not properly implemented or installed. Therefore we have carried out retrospective testing of all the new models installed as part of this project.

2.0 SHORT-TERM EARTHQUAKE PROBABILITIES – JAVA VERSION (STEP-JAVA) MODEL

The Short-Term Earthquake Probability (STEP) model (Gerstenberger, 2003; Gerstenberger *et al*., 2005) is an aftershock model based on the idea of superimposed Omori-type sequences (Ogata, 1988, 1998) in which the number of aftershocks decay with time according to an inverse power law with an exponent not much different from 1. The model comprises two components: a background model, and a time-dependent clustering model. As implemented in the testing centre, the national seismic hazard model NZNSHM, a modified version of the model described by Stirling et al. (2002), is applied as the background model. The clustering model is based on the work of Reasenberg and Jones (1989) which defines aftershock forecasts based on the *a*- and the *b*- value from the Gutenberg-Richter relationship (Gutenberg & Richter, 1944) and the *p*-value from the modified Omori law (Ogata, 1983).

The clustering model itself combines three different approaches to forecast aftershocks. The first is based on the average (“generic”) behaviour of aftershock sequences in New Zealand and uses the median Reasenberg and Jones parameter values for New Zealand aftershock sequences, with parameter estimates from Pollock (2007). The second approach uses the development of the ongoing aftershock sequence to refine the forecast. In this component the Reasenberg and Jones parameters are estimated for each individual aftershock sequence as it develops. The third component refines the forecast further by allowing for spatial heterogeneities within an aftershock sequence. It computes the Reasenberg and Jones parameters on a 0.1 degree by 0.1 degree grid within the aftershock sequence.

The model was previously implemented in the testing centre using the Matlab software package. New software to implement the STEP model has now been developed and installed in the testing centre using the Java programming language rather than Matlab. Java is an object-oriented language that is specifically designed to have as few implementation dependencies as possible, and one of the main rationales for using this language is to be able to run the model without using licensed proprietary software. However, the result is a code that runs many times faster than previously. This is a major advantage, since it is a short-term model which needs to be updated daily. The STEP-JAVA model is installed in both the New Zealand and California testing centres. Also, it has been implemented in the China test region, and is being implemented in the Japan test region.

Although there are no intentional differences between STEP-JAVA and the original STEP model, it is important to ascertain whether any such differences exist. This can be checked through the retrospective testing described in section 8.0 below.

3.0 SHORT-TERM FORECASTING MODEL (STFM) BASED ON A NEW GENERIC MODEL FOR AFTERSHOCK OCCURRENCE

A new short-term forecasting model has been developed based on a new generic model for aftershock occurrence (Christophersen and Gerstenberger, submitted). The paper describing this STFM model and its application to New Zealand, California and Italy is reproduced in the Appendix of this report.

Like the STEP model, the STFM model forecasts the future aftershocks of previously occurring earthquakes, grouped into main shock – aftershock clusters. The key difference is that the forecast abundance of aftershocks is different from that in the Reasenberg and Jones (RJ) (1989) formulation that is used in the STEP model. In the RJ formulation, aftershock occurrence is scale-invariant, in the sense that for a main shock of any magnitude M and for a fixed positive number δM, the expected number of aftershocks exceeding the lower magnitude $M - \delta M$ is independent of the actual value of M. In the STFM model, the expected number is allowed to depend on M. This is achieved by estimating a growth exponent α of aftershock abundance which is such that the log of the expected number of aftershocks exceeding $M - \delta M$ is proportional to $(\alpha - b)(M - \delta M)$, where b is the Gutenberg-Richter b-value.

Christophersen and Gerstenberger (submitted) found that $\alpha > b$ for all regions studied (Appendix, Table 2) implying that, relative to a scale-invariant formulation, aftershock abundance increases with the magnitude of the main shock.

The STFM model is essentially a modification of the STEP model in which the first approach to estimating the clustering component of the model (the “generic” or mean behaviour of aftershock sequences) is modified to accommodate the growth of aftershock abundance with magnitude. The second and third approaches, in which the forecast is refined based on the development of an individual aftershock sequence, are the same as in the STEP model.

The final installation and testing of the STFM model have been delayed due to issues related to the estimation of α that have arisen during the review process of the Christophersen and Gerstenberger paper. The model is not yet being tested.

4.0 EEPAS ALLOWING FOR AFTERSHOCKS (EAS) MODEL

In the EEPAS (Every Earthquake a Precursor According to Scale) intermediate-term forecasting model, based on the precursory scale increase (Ψ) phenomenon and associated predictive scaling relations (Rhoades and Evison, 2004), the model is supposed to apply to all earthquakes above a given magnitude threshold m_c, but actually no specific allowance is made for aftershocks of earthquakes that the model forecasts. The model could only effectively forecast aftershocks in general if aftershocks had their own precursory scale increase, preceding the main shock and with a shorter precursor-time than that of the main shock. To correct this deficiency, Rhoades (2009) proposed modifications to the EEPAS model to allow for the occurrence of aftershocks of predicted events. Using earthquake catalogues of California and the Kanto region, central Japan, he fitted versions of the modified (EAS) and original EEPAS model to a period of about 10 years and independently tested them on a later period of about 10 years of each catalogue. He found that allowing for aftershocks of predicted events increased the log likelihood by an average of about 0.1 for every target earthquake.

Here we describe the adaptation of the EAS model to the New Zealand CSEP testing region.

The aftershocks that are expected to follow any shallow major earthquake can be described, to a good approximation, in time by the Omori-Utsu relation for aftershock-rate decay (Ogata, 1983), in magnitude by Båth's law (Båth, 1965) and the Gutenberg-Richter (G-R) frequency-magnitude relation, and in location by the Utsu (1961) relation between earthquake magnitude and aftershock area. In the EAS model, the Omori-Utsu relation is disregarded and the aftershocks are assumed to have the same time distribution as the main shock. This is because the timescale of aftershock occurrence, although it can extend for months, is generally much shorter than the time between the onset of the precursory scale increase and the occurrence of a related mainshock. The effect of Båth's law is incorporated directly into the EAS model, through an extra fitted parameter, γ , which represents the minimum difference between the magnitude of the mainshock, m, and that of any of its aftershocks, m_a. The G-R law is allowed for by two extra fitted parameters, α and θ, with α representing the slope of the G-R frequency-magnitude relation and θ reflecting the average abundance of aftershocks. The conditional probability density for the magnitude m_a of aftershocks of a mainshock with given magnitude m is assumed to be of the form

$$g_A(m_a \mid m) = \alpha \exp\left[-\alpha\left(m_a - m + \theta\right)\right] H(m - m_a - \gamma) \tag{1}$$

where $H(s) = 1$ if $s > 0$ and 0 otherwise. Here $\alpha/\ln(10)$ is the Gutenberg-Richter b-value of aftershock sequences, b_{AS}. If $\beta/\ln(10)$ is the Gutenberg-Richter b-value of the whole earthquake catalogue, b_{EEPAS}, then we require $\alpha < \beta$ to avoid a mathematical contradiction. For practical reasons, it is convenient to fit the parameters ζ and ω, rather than γ and θ where

$$\zeta = \gamma - \sigma_M^2 \alpha\,; \quad \omega = \theta - \sigma_M^2 \alpha / 2\,; \tag{2}$$

where σ_M is the standard deviation of the conditional distribution for the magnitude of a mainshock given a precursor in the EEPAS model.

The Utsu (1961) relation, $\log A = M - 4$, for the aftershock area A in km^2 of an earthquake of magnitude M, implies that the spread of the location distribution scales with magnitude. In the EAS model, the distribution for aftershock locations is taken as a bivariate normal distribution with circular symmetry, centred on the location of the mainshock, i.e.,

$$h_A(x_a, y_a \mid m, x, y) = \frac{1}{2\pi\sigma_V^2 10^m} \exp\left[-\frac{(x_a - x)^2 + (y_a - y)^2}{2\sigma_V^2 10^m} \right], \tag{3}$$

where σ_V is a parameter controlling the variance of aftershock locations.

The EAS model has one more extra parameter, p_m, which is the proportion of earthquakes in the target magnitude range that are mainshocks, as opposed to aftershocks. This parameter is necessary to simplify the normalisation of the model. The model assumes that this proportion is constant for all magnitudes in the target range. This assumption represents a pragmatic but somewhat unrealistic compromise since it is not expected to be borne out in practice; in real earthquake catalogues the proportion of mainshocks tends to increase with earthquake magnitude.

In adapting the EAS model to the New Zealand test region, a number of the parameters were fixed or estimated prior to optimization of the model. These parameters and their fixed or estimated values are listed in Table 1. The parameter b (of the PPE model) is the Gutenberg-Richter b-value estimated from earthquakes with magnitude M ≥ 4.95 in the New Zealand test region over the period 1964-2006. This and the other PPE model parameters are the same as those in PPE model already installed in the testing centre. The EAS model parameter b_{EEPAS} is the Gutenberg-Richter b-value optimised for M ≥3.95 in the New Zealand test region over the period 1987-2006. The parameter b_M was fixed at 1 in the interests of simple magnitude scaling, as in nearly all recent applications of the EEPAS model. The parameter b_{AS} was arbitrarily set at 0.9, to satisfy the requirement that $b_{AS} < b$ (i.e., $\alpha < \beta$). The parameter σ_V was set at to be consistent with Utsu's areal relation. The parameter p_m was estimated from the aftershock model that is used to down-weight aftershocks in the EEPAS_1r and EEPAS_1f models (Rhoades et al., 2008).

Table 1 Fixed parameters, and parameters fitted independently of other model parameters in adaptation of EAS mode1 to the New Zealand test region.

Parameter	**Value**
b	1.16
PPE model	
a	0.35
d	3.02 km
s	8.09×10^{-12} d^{-1}km^{-2}
EAS model	
b_{EEPAS}	1.10
b_M	1.00
b_{AS}	0.90
σ_V	0.0056 km
p_m	0.70

The best-fitting version of the EEPAS model, EEPAS_0f, in which all precursory earthquakes are weighted equally and eight EEPAS model parameters are fitted, was used as the basis for the EAS model. Hence we may call this the EAS_0f model. The eight standard EEPAS model parameters and two additional parameters were simultaneously optimised for the period 1987-2006. The minimum magnitude for precursory earthquakes was taken as m_0 = 2.95, and the minimum magnitude for target earthquakes as m_c = 4.95. The fitted values are shown in Table 2.

Table 2 EAS_0f model parameters fitted to New Zealand test region over the period 1987–2006.

Parameter	Restriction	Fitted value
a_M	1.0 – 2.0	1.38
σ_M	0.2-0.6	0.25
a_T	1.0 – 2.5	1.69
b_T	0.3-0.7	0.41
σ_T	0.2-0.6	0.56
b_A	0.3-0.7	0.30
σ_A (km)	0.5 – 5.0	1.62
μ	0.0 – 0.2	0.0003
ζ	0.5 – 1.5	0.57
ω	0.1 – 1.5	0.21

5.0 EARTHQUAKE-RATE-DEPENDENT EEPAS (ERDEEP) MODEL

The ERDEEP model is an elaboration of the EEPAS model in which the possibility is entertained that the scale parameters for time and area depend on the local seismicity rate.

The time density of an individual earthquake's contribution to the EEPAS model (Rhoades and Evison, 2004, 2005, 2006) is of the form:

$$f_{1i}(t) = \frac{H(t-t_i)}{(t-t_i)\sigma_T \ln(10)\sqrt{2\pi}} \exp\left[-\frac{1}{2}\left(\frac{\log(t-t_i) - a_T - b_T m_i}{\sigma_T}\right)^2\right], \qquad (4)$$

where m_i is the magnitude of a precursory earthquake, and $H(s) = 1$ if $s > 0$ and 0 otherwise. The scale parameter for time is a_T. The expected precursor time is proportional to 10^{a_T}.

The location density in the EEPAS model is of the form:

$$h_{1i}(x, y) = \frac{1}{2\pi\sigma_A^2 10^{b_A m_i}} \exp\left[-\frac{(x-x_i)^2 + (y-y_i)^2}{2\sigma_A^2 10^{b_A m_i}}\right] \qquad (5)$$

where (x_i, y_i) is the location of the precursory earthquake. The scale parameter is σ_A. The area of the location density is proportional to σ_A^2.

The ERDEEP model is based on the observation that, using data from several different catalogues with different average seismicity rates, the Ψ scaling relation between mainshock magnitude M_m and $\log(A_P T_P)$ has a small scatter compared to those between M_m and $\log T_P$ and $\log A_P$ separately (Rhoades, 2010), where A_P and T_P are the precursor time and area,respectively. In catalogues with low seismicity rates, the precursor times are expected to be relatively long because of a low tectonic stressing rate. The small scatter in the relation between M_m and $\log(A_P T_P)$ then implies that there is a compensating decrease in A_P in such a way that the product $A_P T_P$ is constant. Hence, in equation (4), the constant parameter a_T is replaced by a variable $a_T(i)$ that depends on the local seismicity rate $\rho(i)$ in the location where the *i*th earthquake occurs:

$$a_T(i) = a_T - c_\rho \log \rho(i), \qquad (6)$$

where c_ρ is an adjustable parameter. In equation (5), the parameter σ_A is replaced by a variable $\sigma_A(i)$ that also depends on the seismicity rate:

$$\sigma_A(i) = \sigma_A \rho(i)^{c_\rho/2}. \qquad (7)$$

We note that the product of the area of the location density and the expected precursor time is a constant, independent of the seismicity rate, i.e.

$$\sigma_A^2(i) 10^{a_T(i)} = \sigma_A^2 10^{a_T}. \qquad (8)$$

In applying the ERDEEP model to the New Zealand test region, we use the PPE model rate density at the end of 1986 to estimate the local seismicity rate at grid-points spaced at 0.1 degree intervals of latitude and longitude, and interpolate between grid-points to estimate $\rho(i)$ at the locations of precursory earthquakes.

The ERDEEP model with all precursory earthquakes weighted equally was used, and eight EEPAS parameters were optimised as for the EEPAS_0f. Hence we could refer to the model as ERDEEP_0f model. A number of the parameters were fixed or estimated prior to optimization of the model. These parameters, b, a, d, s, b_{EEPAS} and b_M and their fixed or estimated values are listed in Table 1. See the discussion of the EAS model for further explanation. The other nine ERDEEP parameters were simultaneously fitted to minimise the log likelihood of the model over the period 1987-2006. The minimum magnitude for precursory earthquakes was taken as m_0 = 2.95, and the minimum magnitude for target earthquakes as m_c = 4.95. The fitted values are given in Table 3.

Table 3 ERDEEP_0f model parameters fitted to New Zealand test region over the period 1987–2006.

Parameter	Restriction	Fitted value
a_M	1.0 – 2.0	1.00
σ_M	0.2-0.6	0.41
a_T	1.0 – 2.5	1.65
b_T	0.3-0.7	0.40
σ_T	0.2-0.6	0.60
b_A	0.3-0.7	0.33
σ_A (km)	0.5 – 5.0	1.62
μ	0.0 – 0.2	0.0003
c_ρ	0.5 – 1	0.50

6.0 DOUBLE BRANCHING PROCESS MODEL (DBPM)

A double branching model for earthquake occurrence was proposed by Marzocchi and Lombardi (2008). The model is a two-step branching process. The first-step branching describes short-term clustering as a spatial ETAS model (Ogata, 1998). In this model, the seismic rate is the sum of two components: the rate of "spontaneous" events, that refers to activity which is not triggered by precursory events, and the rate of events internally triggered by previous earthquakes. The total intensity function of the first-step branching model is given by

$$\lambda_1(t,x,y) = \mu_1(x,y) + \sum_{t_i<t}\left[\frac{K_1}{(t-t_i+c)^p}\exp\left[\alpha_1(M_i - M_{\min})\right]\frac{C_{d_1,q_1}}{\left(r_i^2+d_1^2\right)^{q_1}}\right], \qquad (9)$$

where $\mu_1(x,y)$ is the probability density function of locations of spontaneous events, $M_{\min}$ is the minimum magnitude of the catalogue, C_{d_1,q_1} is a normalisation constant of the triggering spatial function, and r_i is the distance between locations (x,y) and (x_i, y_i). The value $q_1 = 1.5$ is imposed to reflect the theoretical decay of static stress with the inverse cube of the epicentral distance.

A "background" process is derived from the fitted first-step branching process by applying the procedure of stochastic declustering (Zhuang et al., 2002), and then by selecting a number of earthquakes equal to the expected number of spontaneous events as those with the highest probability of belonging to the background.

The second-step branching involves fitting a branching process to the background events obtained from the first-step branching model. The purpose of the second-step branching is to capture long-term clustering of earthquakes. The conditional rate of earthquakes for second-step branching is of the form:

$$\lambda_2(t,x,y) = \mu_2(x,y) + \sum_{t_i<t} K_2 \exp\left[\alpha_2(M_i - M_{\min}) - \frac{(t-t_i)}{\tau}\right]\frac{C_{d_1,q_1}}{\left(r_i^2+d_2^2\right)^{q_2}}. \qquad (10)$$

The difference between the first-step and second-step branching is thus in the time-decay of clustering, which is a power-law decay following the modified Omori law in the case of first step branching (9) and an exponential decay in the case of second-step branching (10).

The DBPM model has now been implemented by its authors for the earthquake forecast testing centres and installed in the New Zealand testing centre in the three-month model class. It is also installed in the European testing centre covering the Italy test region.

7.0 2010 REVISION OF THE NEW ZEALAND SEISMIC HAZARD MODEL (NZNSH2010)

A team of earthquake geologists, seismologists and engineering seismologists from GNS Science, NIWA, University of Canterbury and Victoria University of Wellington have collectively produced a major update of the national probabilistic seismic hazard model for New Zealand (Stirling et al., 2002). The 2002 model has been subject to minor updates in the intervening years, and one of those updates has been implemented in the testing centre as the NZNSHM model. Like the earlier model, the new model sits within the standard framework of probabilistic seismic hazard analysis, which incorporates recurrence of characteristic earthquakes on known active faults and widely distributed background seismicity conforming to the Gutenberg-Richter frequency-magnitude law.

The new model incorporates over 200 new onshore and offshore fault sources, and utilises newly developed New Zealand-based scaling relationships and methods for the parameterisation of the faults and subduction interfaces. These relationships are used to derive characteristic earthquake magnitudes from the measured or estimated source dimensions.

The background seismicity component of the model allows for the occurrence of earthquakes away from the known fault sources, modelled as a grid of earthquake sources with rate parameters estimated from the historical seismicity data. This component of the model has also been updated to include new seismicity data, a new seismicity regionalisation, and a simpler methodology for calculation of the seismicity parameters, which uses less of the earthquake catalogue, but better accounts for uncertainties.

The NZNSH2010 model will be described fully in a forthcoming paper (Stirling et al., in prep.). It has been installed in the testing centre as a five-year model. Retrospective tests of this model over 25 years are described in section 8.0 below. These tests and comparison with similar tests of the earlier national model (NZNSHM) and other five-year models installed in the centre (SUP and PPE), will contribute to an evaluation of the impacts of the revision.

For both models, the implemented model is an interpretation of the original model with a difference due to the requirements for grid-based rates in the Testing Centre. To accomplish this, the rates of the fault sources are evenly distributed over the grid cells they pass through. This ensures that the overall rate of expected events is retained; however individual rates of larger events per cell can differ from those in the original model.

8.0 RETROSPECTIVE TESTS

Retrospective tests have been carried out for all the new models and some of the models previously installed in the testing centre. The standard tests installed in the CSEP software are the N-test comparing the total number of target earthquakes expected under each model with the observed number, the L-test comparing the expected log likelihood expected with the observed log likelihood for each model, and the R-test which compares the observed log likelihood-ratio of pairs of models with its expected value under each model. The standard CSEP presentation of the test results is given in terms of statistical significance of differences, and is aimed at a decision of whether or not to reject a given model. Such a presentation is not helpful when the aim is to determine whether a model is correctly implemented, because the results cannot readily be related to statistics that modellers can compute for themselves. In addition, the interpretation of the L- and R-tests is difficult for models that fail the N-test.

In order to provide useful feedback to modellers, we have carried out two tests. These are the N-test, but presented in a different way from the standard CSEP presentation, and the T-test, recently proposed by Rhoades et al. (submitted). The T-test is based on the classical paired t-test. It gives a direct comparison of the log likelihoods of two models. It places a confidence interval on the information gain per earthquake $I(A,B)$ of model A over model B. $I(A,B)$ is the limiting value, as the number of target events tends to infinity, of the sample information gain per earthquake $I_N(A,B)$:

$$I_N(A,B) = \frac{1}{N}\left[\sum_{k=1}^{N}\left(\ln\lambda_A(i_k) - \ln\lambda_B(i_k)\right) - \left(\hat{N}_A - \hat{N}_B\right)\right] \tag{11}$$

where there are N target earthquakes, occurring in bins $(i_k, k = 1,\cdots,N)$; $\lambda_A(i)$ is the expected number of earthquakes in the ith bin under model A; and $\hat{N}_A$ is total expected number of earthquakes in the test domain under model A.

Although approximate and predicated on a normality assumption, this T-test is not computer-intensive, is easier to interpret than the R-test, and becomes increasingly dependable as the number of earthquakes increases. Unlike the N-test, L-test and T-test, which are all test of consistency between a model and the data, the T-test is based on the variability in the ratios of the rates in the bins in which the target earthquakes fall.

8.1 Tests of 5-year models

As in Gerstenberger et al. (2009), we test the 5-year models retrospectively on 25-years of the New Zealand earthquake catalogue from 1984-2009.

We present results for the mainshock-only models, which are designed for application to a declustered catalogue, because this is the category to which the new 5-year model NZNSH2010 belongs. The N-test results are shown in Figure 1. The error bars on the expected number of earthquakes represent a 95% tolerance interval for the number of earthquakes in the test domain, given the model. If the error bar intersects the dashed line representing the actual number of earthquakes in the test domain, there is no significant difference between the model expectation and the data and the model passes the N-test. Otherwise it fails the N-test.

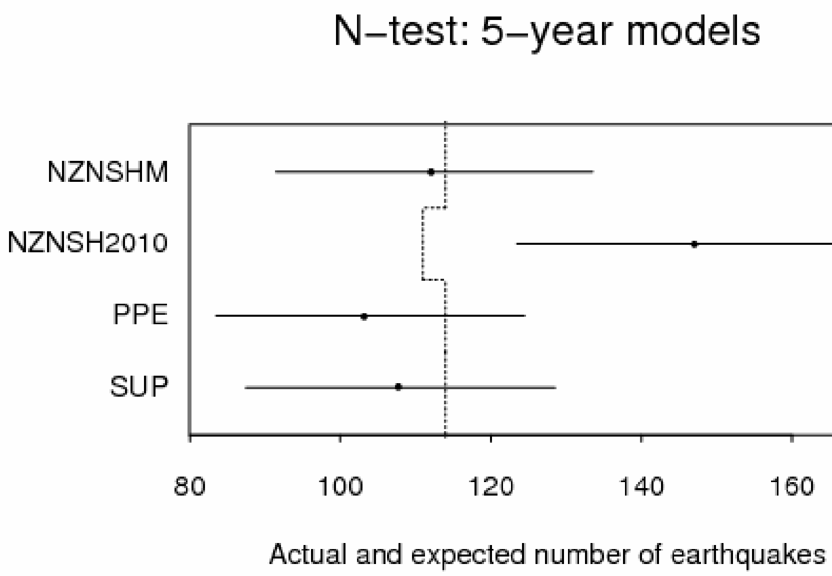


Figure 1 N-test of 5-year mainshock-only models in the New Zealand test region over the period 1984-2009. A model is consistent with the data if the 95% tolerance interval as shown encompasses the actual number of earthquakes shown by the dotted line.

The N-test immediately indicates a problem with implementation of the NZNSH2010 model, since the observed number of earthquakes is less for this model than the otherwise. This is because the model has an expected number of zero in the bins to which several of the observed earthquakes belong. Since such an eventuality causes a model to have likelihood zero and log likelihood undefined, these bins were excluded from the evaluation of the NZNSH2010 model. The reason for these zeroes can be investigated; they should not occur if the model is implemented as intended. It is possible that the model has not been defined out to the limits of the test region.

Figure 1 shows that, with the zero-rate bins excluded, the NZNSH2010 model fails the N-test; it overestimates the number of earthquakes in the test domain. This is probably due to the increased number of fault sources in the NZNSH2010 model over the NZNSHM model. The failure of an N-test in a particular period does not necessarily indicate a serious deficiency of a model. It is well-known that any model is bound to fail the N-test during certain time-periods because the distribution of the number of earthquakes occurring during successive periods is not Poissonian, but is better fitted by the negative binomial distribution. This is true for declustered catalogues as well as for unfiltered catalogues (Schorlemmer et al. 2010). Because the NZNSH2010 model incorporates contributions from fault sources derived from paleo-seismological studies as well as observed earthquakes, it is not expected to conform as closely to the past instrumental earthquake catalogue as models, such as PPE, which are based entirely on that catalogue. It is aimed at a longer-term estimate of the hazard than 25 years.

Figure 2 shows the results of the T-tests, with each panel showing the comparisons between one particular model and each of the others. The error bars on the sample information gain per earthquake are approximate 95% confidence intervals for the true information gain per earthquake of one model over another. If the error bar intersects the dashed “zero” line, there is no significant information gain of one model over the other. If the error bar lies entirely to the right of the zero line, the model named in the heading of the plot is more informative than that named in the side of the panel. If it lies entirely to left of the zero line, then opposite is true.

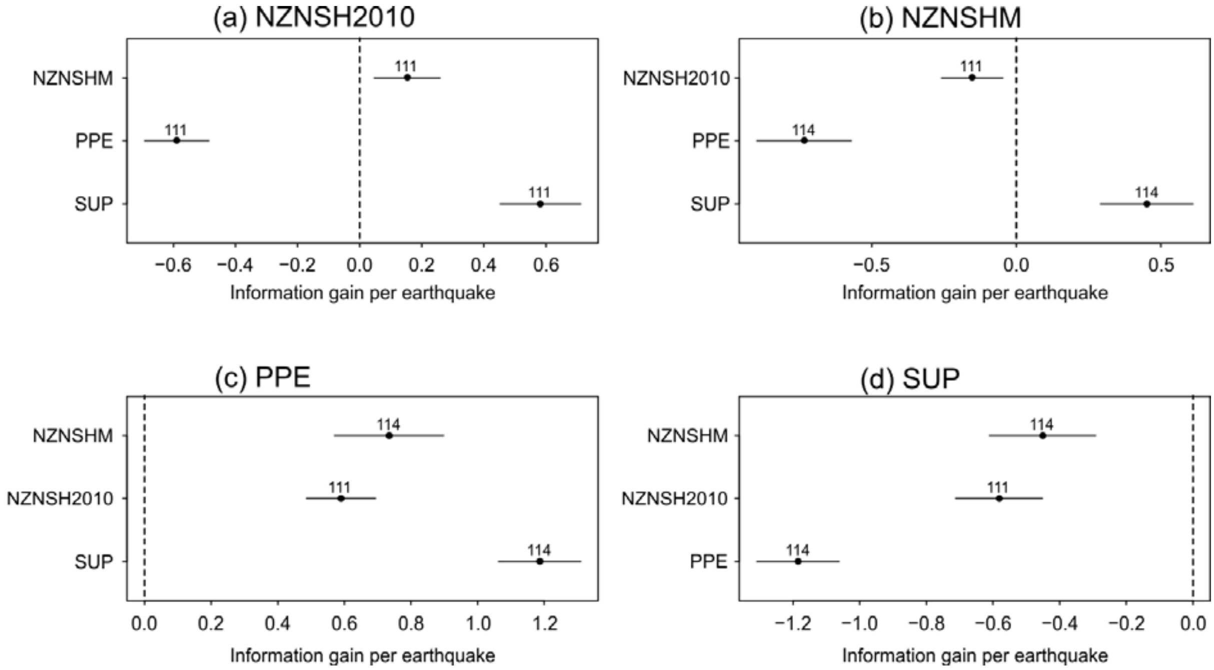


Figure 2 T-test of 5-year mainshock-only models over the period 1984-2009. The model named in the heading of each plot is significantly more informative than the one listed down the side if the associated 95% confidence limits lie entirely to the right of the dashed zero line, and vice versa.

In the case of the NZNSH2010 model, we see from Figure 2b that this model is significantly more informative than the SUP and NZNSHM models, but significantly less informative than the PPE model over the 25-year test period. The result with respect to the SUP model is to be expected, because the SUP model is a model of least information that includes no spatial variation in the rate of earthquake occurrence. The result with respect to NZNSHM shows that in some respects the revised seismic hazard model is a better fit to the recent catalogue than the old model, in spite of it failing the N-test.

The PPE model performs significantly better than all other models over the test period. This is not surprising, since it was estimated by smoothing the observed earthquake locations over almost the same time-period. This shows that the implementation of the PPE model is substantially as intended, but by no means indicates that it will outperform the other models in prospective testing.

On the other hand, the SUP model performs significantly worse than all of the others over the test period. This is entirely consistent with its role as a model of least information.

The T-test results are more clear-cut and easier to interpret than the R-test results which form part of the standard CSEP output. R-test results for the same time period were given for the NZNSHM, PPE and SUP models by Gerstenberger et al. (2009), and were much more equivocal and difficult to interpret. They showed, for instance, that the log likelihood-ratio between the NZNSHM and SUP models is consistent with neither model being the correct one, as evidenced by both plots lying in the rejection region in Figure 3. In the original interpretation of the R-test by Schorlemmer et al. (2007), the conclusion that would be drawn is that each of these two models is rejected in favour of the other. This conclusion is not accurate because the R-test is only a test of consistency of a particular statistic with a model. Gerstenberger et al. (2009) interpreted the result as implying that each model contained some information that the other did not. However this may be, it is difficult to see what information the SUP model could be providing that the NZNSHM model does not, except perhaps that it conforms more closely to the Gutenberg-Richter frequency magnitude law. The first order conclusion provided by the T-test, that the NZNSHM model is more informative than the SUP model, is a more important one, and it is a conclusion that the R-test does not provide.

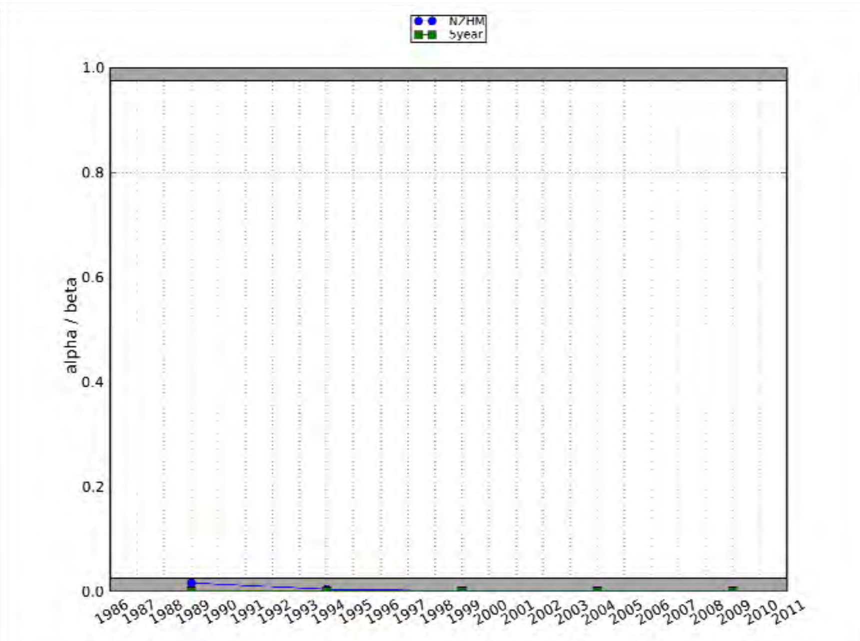


Figure 3 Cumulative R-test comparing the NZNHSM and SUP models over the period 1984-2009. After Gerstenberger et al. (2009). Note that both sets of results plot in the bottom rejection bar.

Summarising the results of the tests of the 5-year models, the models all perform much as expected. The failing of the N-test by the NZNSH2010 model is not necessarily a concern, but the reasons for it should be thoroughly investigated. However, the zero expected values that the model assigns to some cells certainly need to be investigated and the implementation of the model corrected in this respect.

8.2 Tests of 3-month models

The 3-month models were tested retrospectively over the two-year period from 1 Jan 1996 to 31 Dec 1997. This limited time-span was used because of the total time taken to compute the bin-rates for these models, which is of the order of one day for each 3-month period.

The results of the N-tests are shown in Figure 4. This shows that one new model, EAS_0f, is overestimating the total number of earthquakes by a wide margin (about a factor of 10). This indicates that the EAS_0f model has been incorrectly implemented in the testing centre. The other models all have a tendency to overestimate the number of earthquakes in this period, and for all but two of the models – DPBM and PPE – the tendency is significant at the 95% confidence level. This does not necessarily mean, however, that any of these models are incorrectly implemented. We note that the 1996-97 year had a lower than average number of target events (only 9), and as already mentioned above the number of earthquakes in successive years does not follow the Poisson distribution exactly and therefore occasionally any model will under- or over-predict the number of earthquakes.

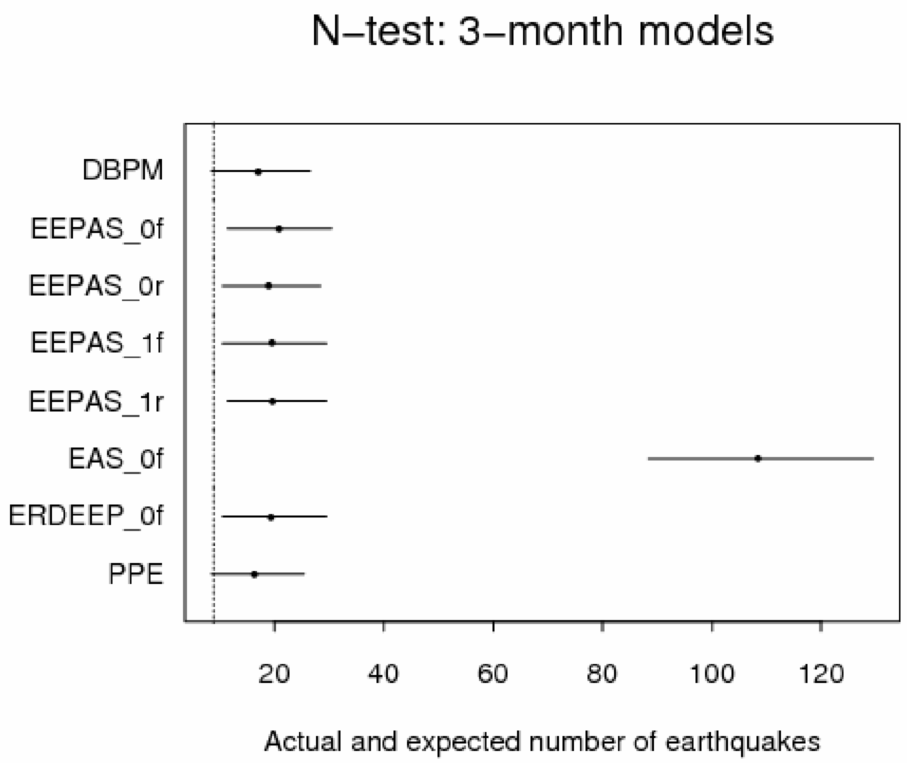


Figure 4 N-test of 3-month models in the New Zealand test region over the period 1996-1997. See Figure 1 caption for details.

In any case, the numbers of events expected in any time period can be checked for the PPE model and all forms of the EEPAS model against the numbers estimated independently in the EEPAS software package which does not involve binning of earthquake times-of-occurrence, magnitude and location. When such checks are done it is found that for the EEPAS_0f, EEPAS_0r and ERDEEP_0f models there is hardly any discrepancy (<0.3) between the expected number of earthquakes calculated from the bin expectations and those in the EEPAS software package. For two models, EEPAS_1f and PPE there is a moderate discrepancy of about 2 (or 10%). For one model, EEPAS_1r, there is a large discrepancy of 5.7 (more than 25%). Both the software used to compute bin-rates and that used to estimate total expected numbers in a test domain need to be checked. In both cases, numerical approximations are involved and the precision is controlled by parameter settings. These settings should be examined to see if the precision of the approximations should be increased, or whether there is a programming error responsible for the discrepancies seen.

The general tendency for the EEPAS and PPE models to overestimate the number of earthquakes in this particular two-year period is confirmed by the results using the EEPAS software package. In fact, for the three models where a moderate or large discrepancy exists, the estimate of the expected number of earthquakes is higher using the EEPAS software package than by summing the bin rates.

The T-test results are shown in Figure 5(a-h). Because of its gross over-estimation of the number of earthquakes the EAS_0f model is seen to be much less informative than every other model. The other models mostly differ insignificantly from each other on the T-test, the only exceptions being the EEPAS_0r model is significantly more informative than the EEPAS_0f and ERDEEP_0f models in this period. Taking into account the discrepancies already noted in the expected numbers of earthquakes for some models, the information gains among the PPE and EEPAS models (excluding EAS_0f) are very similar to those computed using continuous time, magnitude and location in EEPAS software package. Therefore, apart from the discrepancies in total expected numbers already noted, there is nothing in these results to suggest that any of these models are not correctly implemented.

Summarising the results for the tests of the 3-month models, there is a serious error in the implementation of the EAS_0f model which must be corrected. There is a possible implementation error affecting the calculation of total expected number of earthquakes for the PPE, EEPAS_1f and, especially, the EEPAS_1r model.

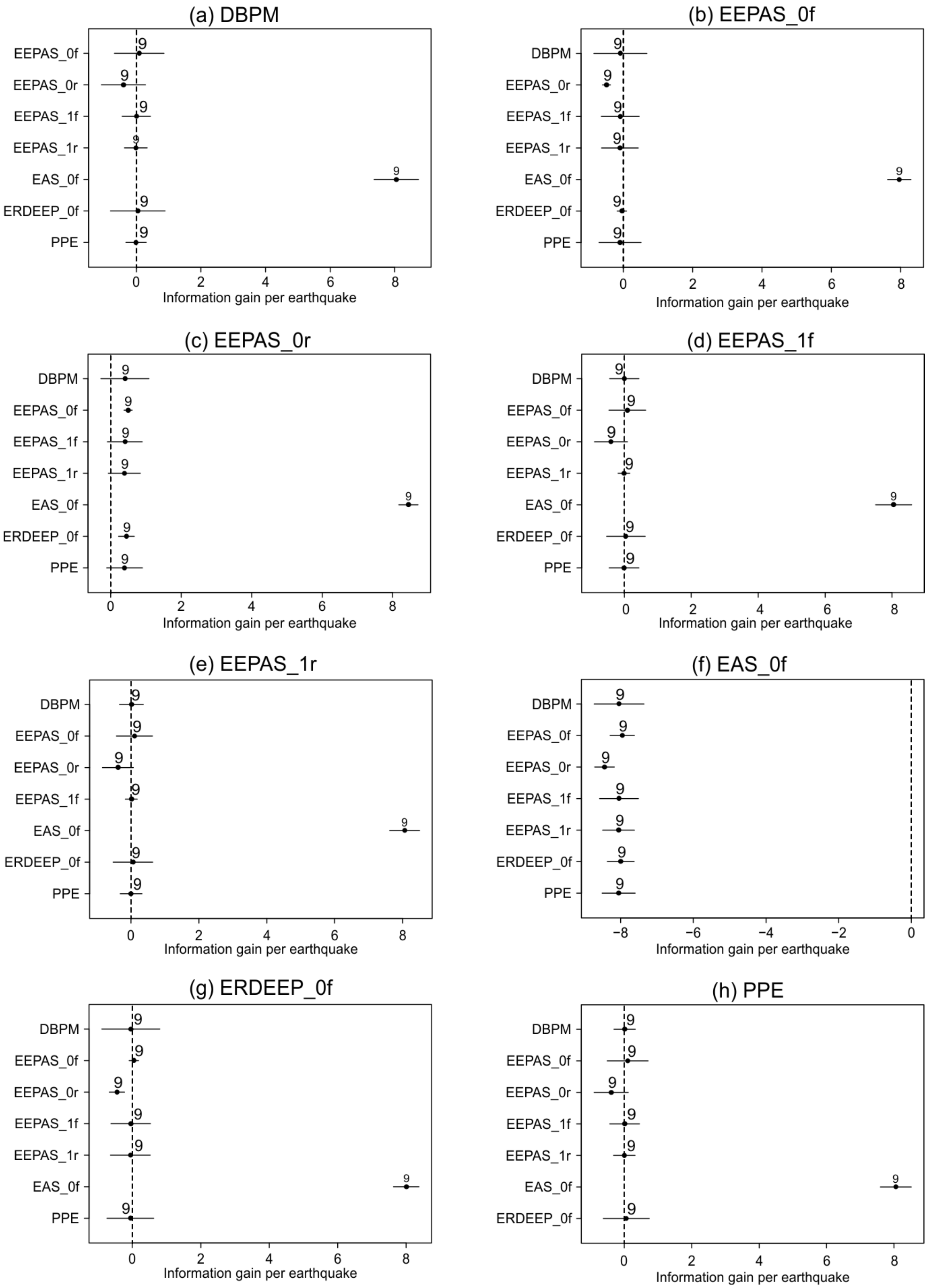


Figure 5 T-tests of 3-month models in the New Zealand test region over the period 1996-1997. See Figure 2 caption for details.

8.3 Tests of 1-day models

The 1-day models were tested retrospectively over a six-month period starting 1 Jan 2006. In this period there were 36 target earthquakes at M > 3.95 in the test region.

The N-tests are shown in Figure 6. All three models fail the N-test. As noted above, merely failing the N-test for one period does not necessarily indicate a serious flaw in model implementation. These clustering models are based partly, or in the case of PPE-ETAS, totally, on average aftershock rates and aftershock rates of individual sequences can vary markedly from the average. Nevertheless, these results need to be compared with modeller's independent computations for the same periods to see if there are any discrepancies. The original STEP model was known to over-estimate the number of earthquakes (Gerstenberger et al., 2009) as seen also here. The STEP-JAVA version is seen here to underestimate the number by a similar margin and the reason for this should be investigated.

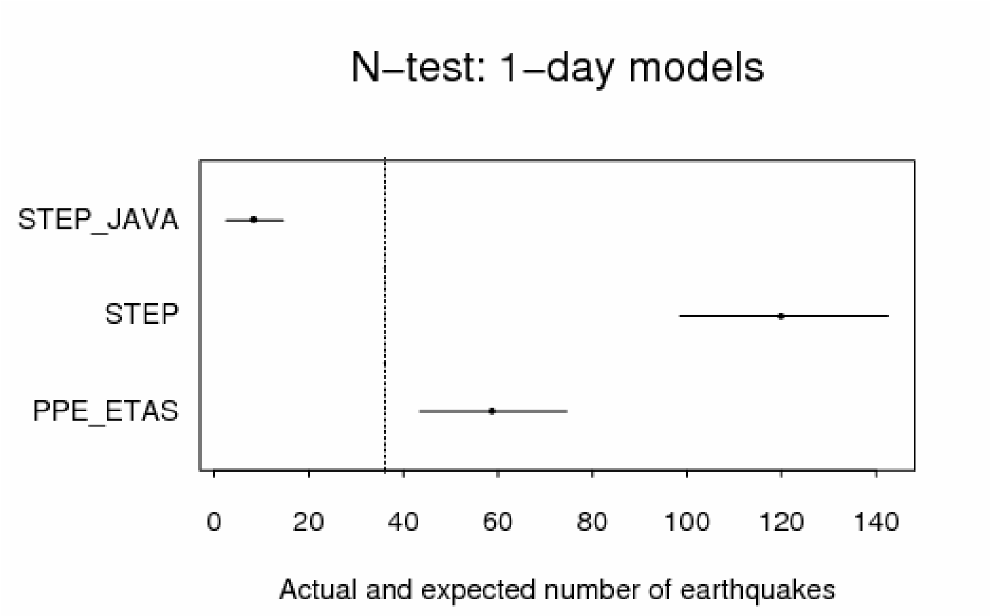


Figure 6 N-test of 1-day models in the New Zealand test region over the period January-June 2006. See Figure 1 caption for details.

The T-test results are given in Figure 7(a-c). These show that the PPE-ETAS model is significantly more informative than the STEP model, which is in turn significantly more informative than the STEP-JAVA model. It is clear from the size of the information gains that neither the STEP model nor the STEP-JAVA model are implemented as intended in the testing centre. If all models were working as intended, the information gains per earthquake would be relatively small, no greater than about 0.2 for these models, which are all based on the same basic laws for aftershock behaviour (Omori-Utsu law and Gutenberg-Richter law).

Summarising the results of the tests on the one-day models, there are errors in the implementation of the STEP and STEP-JAVA models which need to be identified and corrected. The results for PPE-ETAS should be checked against the modeller's code.

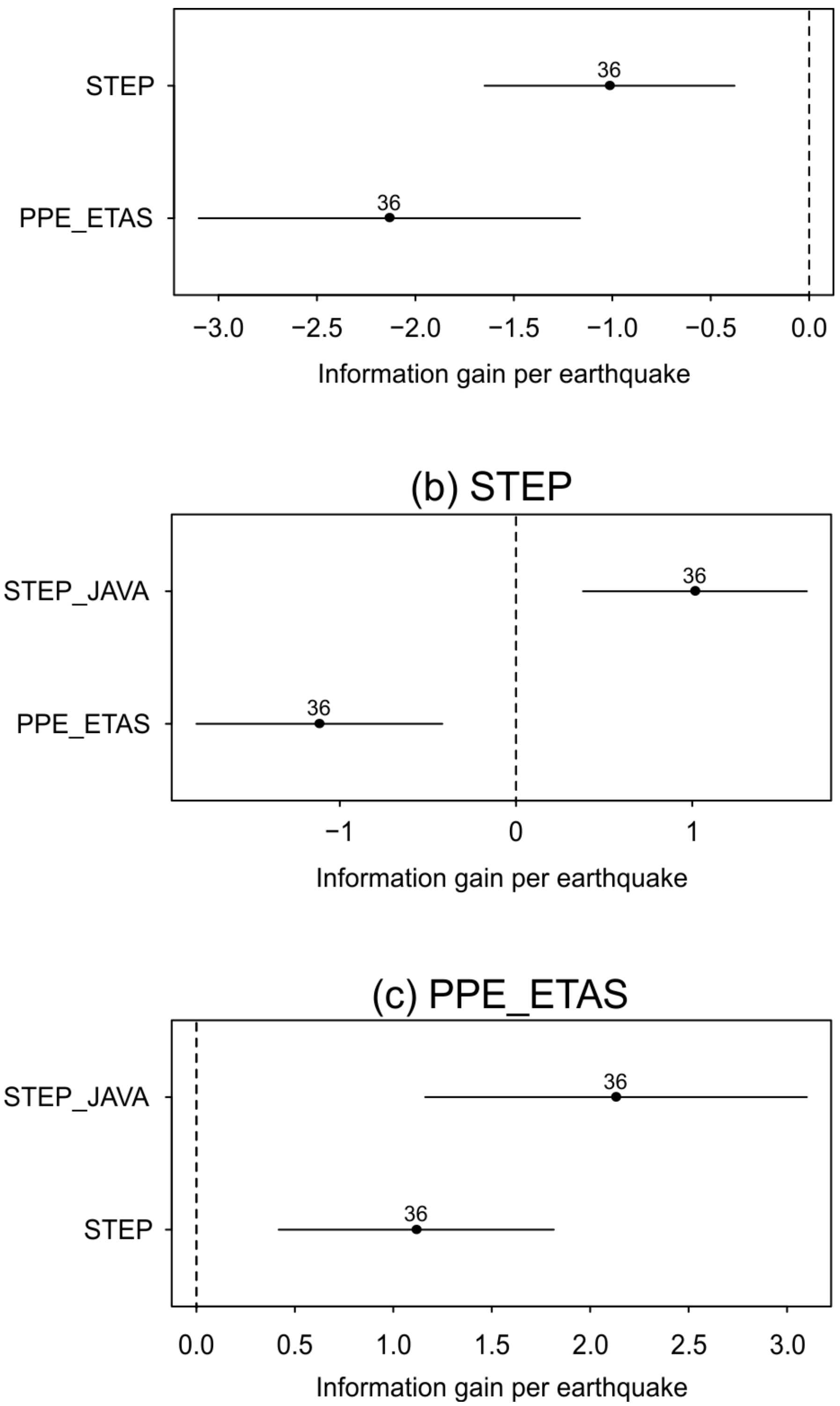


Figure 7 T-tests of 1-day models in the New Zealand test region over the period January-June 2006. See Figure 2 caption for details.

8.4 Summary of model status

In Table 4 we summarize the status of the new models developed and/or installed in the Testing Centre as part of this study, and of both new and existing models retrospectively tested here. The status is summarised by the indication from the tests that a given model may be incorrectly implemented or installed. The indication for each model is assigned, on the basis of the test results, to one of three classes (“no”, “weak” or “strong”).

The classification “no” implies that there is no indication that that the model is incorrectly implemented or installed, i.e., that its retrospective performance is consistent with what is known about the model and the intentions of its developer(s). Note that a model receiving this classification does not necessarily have to pass the N-test or perform well relative to other models in the T-test.

The classification “weak” implies that there is some indication that the model may be incorrectly implemented or installed, i.e., that its retrospective performance does not appear to be entirely consistent with what is known about the model and the intention of its developer(s). Models in this class should be referred back to the developer(s) for reconciliation of the results or possibly correction of the model’s implementation for the Testing Centre.

The classification “strong” implies that there is a strong indication that the model is incorrectly implemented or installed, i.e., that its retrospective performance is clearly inconsistent with what is known about the model and the intention of its developer(s). Models in this class should be referred back to the developer(s) for correction of the model’s implementation for the Testing Centre.

Table 4 Summary of the status of models in the New Zealand earthquake forecast testing centre

Model	**Installed?**	**Indication of incorrect implementation?**
5-year models		
NZNSHM	yes	no
NZNSH2010	yes	weak
PPE	yes	no
SUP	yes	no
3-month models		
DBPM	yes	no
EEPAS_0f	yes	no
EEPAS_0r	yes	no
EEPAS_1f	yes	weak
EEPAS_1r	yes	weak
EAS_0f	yes	strong
ERDEEP_0f	yes	no
PPE	yes	weak
1-day models		
STEP_JAVA	yes	strong
STEP	yes	strong
STFM	no	NA*
PPE_ETAS	yes	weak

* not applicable

Table 4 indicates that out of 15 models retrospectively tested here, for seven models there is no indication of incorrect implementation. However, for five models there is a weak indication of incorrect implementation, and for three models, including two of the newly installed models, there is a strong indication of incorrect implementation.

9.0 CONCLUSIONS

Six new models have been implemented and/or installed for testing in the New Zealand earthquake forecast testing centre, including two new short-term (1-day) models, three new intermediate-term (3-month) models and one new long-term (5-year model), the latter being the 2010 revision of the earthquake-occurrence component of the New Zealand national seismic hazard model.

These and previously existing models have been tested retrospectively using a new presentation of the N-test comparing the expected and actual number of earthquakes, and the recently proposed T-test for estimating the information gain of one model over another, which is not yet installed as a standard test in the CSEP software. The use of these efficient tests has given a much clearer indication of possible implementation errors than previous studies using the standard CSEP test outputs. This has allowed for problems to be identified at the outset rather than after years of computer-intensive testing.

The tests of the long-term (5-year) models over 25 years have shown that the new NZNSH2010 model is significantly more informative than the old NZNSHM model over this period, although it slightly over-estimates the number of earthquakes and appears not to have been implemented over all bins in the test region.

The tests of the intermediate-term (3-month) models over two years have shown that one model, EAS_0f, is not correctly implemented in the testing centre. The technical reason for its gross over-estimation of the number of earthquakes needs to be identified. The results for the other models show there is not much difference between their information values for that period. The tendency of all models to overestimate the number of earthquakes in this period is mainly due to the unusually low number of events in this test period than to any problem with the models. There is a concern with discrepancies in the estimates of the expected number of earthquakes between the binned testing centre version and original continuous version of the model for three of the previously installed models developed using the EEPAS software package, namely PPE, EEPAS_0F and EEPAS_0r. The precision of these estimates needs to be checked.

The tests of the short-term (1-day) models over six months have revealed problems with the new STEP-JAVA implementation. It underestimates the number of earthquakes. On the other hand, the old STEP code, written in Matlab, overestimates the number of earthquakes. The large information gain of the PPE-ETAS model over both versions of the STEP model suggests that neither version is working as intended.

The results of retrospective testing have thus been highly instructive in revealing problems and potential problems with implementation of not only some of the new models but also some of the models previously installed in the testing centre. The incidence of model implementation errors revealed in this study is similar to that found retrospective testing of models of earthquake occurrence in the Italian test region.

10.0 ACKNOWLEDGEMENTS

This work was funded by the EQC Research Foundation under project number 09/TV580. Software developed by the California CSEP Earthquake Forecast Testing Center has been extensively used in establishing the New Zealand Testing Centre. Masha Liukis at the California Testing Center has assisted with installation of models and system software issues. Graeme McVerry and Charles Williams have provided helpful reviews of the report.

11.0 REFERENCES

Båth, M. (1965). Lateral inhomogeneities of the upper mantle, *Tectonophysics*, *2*, 483-514.

Christophersen, A.; Gerstenberger, M.C. (submitted). A new generic model for aftershock occurrence, *Bull. Seismol. Soc. Amer.*

Console, R.; Rhoades, D.A., Murru, M.; Evison, F.F.; Papadimitriou, E.E.; Karakostas, V.G. (2006). Comparative performance of time-invariant, long-range snd short-range forecasting models on the earthquake catalogue of Greece. *Journal of Geophysical Research 111*, B09304, doi:10.1029/2005JB004113.

Evison, F.F.; Rhoades, D.A. (2004) Demarcation and scaling of long-term seismogenesis. *Pure and Applied Geophysics, 161(1):* 21-45.

Gerstenberger, (2003). Earthquake clustering and time-dependent probabilistic seismic hazard analysis for California. *Dissertation submitted to the Swiss Federal Institute of Technology for the degree of Doctor of Science, Zurich.*

Gerstenberger, M.C.; Rhoades, D.A. Stirling, M.W; Christophersen, A.; Brownrigg, R. (2009). Continued development of the New Zealand Earthquake Forecast Testing Centre, GNS Science Consultancy Report CR 2009-182, EQC Research Report 08/BIE35.

Gerstenberger, M.C.; Rhoades, D.A. (2010). New Zealand Earthquake Forecast Testing Centre. *Pure and Applied Geophysics*, *167(8/9*), 877-892.

Gerstenberger, M.; Wiemer, S.; Jones, L.M., Reasenberg, P.A (2005). Real-time forecasts of tomorrow's earthquakes in California, *Nature 435***,** 328-331.

Gutenberg, B., and Richter, C. F., (1944). Frequency of earthquakes in California, Bull. Seismol. Soc. Amer. 34, 185-188.

Harte, D.; Li, D.-F.; Vreede, M. & Vere-Jones, D. (2003). Quantifying the M8 prediction algorithm. Reduction to a single critical variable and stability results. NZ Journal of Geology and Geophysics 46(1), 141-152. URL: www.rsnz.org/publish/nzjgg/2003/010.php

Jackson, D.D. (1996) Hypothesis testing and earthquake prediction. Proceedings of the National Academy of Sciences of the United States of America, 93, 3772-3775.

Jackson, D.D. and Kagan, Y. (1999). Testable earthquake forecasts for 1999, *Seismological Research Letters 70*(*4*), 393-403.

Marzocchi, W, Lombardi, A.M. (2008). A double branching model for earthquake occurrence, *Journal of Geophysical Research*, *113*, B08317, doi:10.1029/2007JB005472.

Ogata, Y. (1983). Estimation of the parameters in the modified Omori formula for aftershock frequencies by the maximum likelihood method. *Journal of Physics of the Earth*, *31*, 115-124.

Ogata, Y. (1988). Statistical models for earthquake occurrences and residual analysis for point processes., *Journal of the American Statistical Association*, *83*, 9-27.

Ogata, Y. (1998). Space-time point process models for earthquake occurrences. *Annals of the Institute of Statistical Mathematics*, *50*, 379-402.

Pollock (2007). Aspects of short-term and long-term seismic hazard assessment in New Zealand, Diploma Thesis, ETH-Zürich.

Reasenberg, P. A., and L. M. Jones (1989). Earthquake hazard after a mainshock in California. *Science,* 243**,** 1173 -1176.

Rhoades, D.A. (2007). Application of the EEPAS model to forecasting earthquakes of moderate magnitude in southern California. *Seismological Research Letters 78(1)*, 110-115.

Rhoades, D.A. (2008). Dependence of precursor time and area on local seismicity rate in long-term seismogenesis. Asian Seismological Commission General Assembly, Tsukuba, Japan, November 24-28 2008. Abstracts Volume.

Rhoades, D.A. (2009). Long-range earthquake forecasting allowing for aftershocks. *Geophysical Journal International*, *178(1):* 244-256.

Rhoades, D.A. (2010). Lessons and questions from thirty years of testing the precursory swarm hypothesis, Pure and Applied Geophysics, 167(6/7), 629-644.

Rhoades, D.A. and Evison, F.F. (2004). Long-range earthquake forecasting with every earthquake a precursor according to scale. Pure Appl. Geophys. 161(1), 47-71.

Rhoades, D.A.; Evison, F.F. (2005). Test of the EEPAS forecasting model on the Japan earthquake catalogue. *Pure and applied geophysics, 162(6/7):* 1271-1290.

Rhoades, D.A.; Evison, F.F. (2006). The EEPAS forecasting model and the probability of moderate-to-large earthquakes in central Japan. *Tectonophysics, 417(1/2):* 119-130.

Rhoades, D.A.; Gerstenberger, M.C. (2009). Mixture models for improved short-term earthquake forecasting. Bull. Seismol. Soc. Amer., *99(2A):* 636-646.

Rhoades, D.; Gerstenberger, M., Christophersen, A; Savage, M.; Zhuang, J. (2008). Testing and Development of Earthquake Forecasting Models. EQC Research Foundation report 06/510. GNS Science Consultancy Report 2008/70.

Rhoades, D.A, Schorlemmer, D., Gerstenberger, M.C., Christophersen, A., Zechar, J.D. and Imoto, M. (submitted). Efficient testing of earthquake forecasting models. Acta Geophysica (special issue on Statistical Seismology).

Schorlemmer, D.; Gerstenberger, M.C. (2007) RELM Testing Center, *Seismological Research Letters, 78(1)*, 1-16.

Schorlemmer, D.; Gerstenberger, M.C.; Wiemer, S.; Jackson, D.D.; Rhoades, D.A. (2007) Earthquake likelihood model testing. *Seismological Research Letters 78(1)*, 17-29.

Stirling, M.W.; McVerry, G.H.; Berryman, K.R. (2002). A new seismic hazard model for New Zealand. Bull. Seismol. Soc. Amer. 92, 1878 – 1903.

Stirling, M.W. and 21 other authors (in prep.). National seismic hazard model for New Zealand: 2010 update.

Utsu, T. (1961). A statistical study on the occurrence of aftershocks, *Geophys. Mag*. *30*, 521-625. A double branching model for earthquake occurrence

Werner, M.J., Zechar, J.D., Marzocchi, W., Wiemer, S. and the CSEP-Italy Working Group (in press). Retrospective evaluation of the five-year and ten-year CSEP-Italy earthquake forecasts. *Annals of Geophysics.*

Zhuang, J., Ogata, Y., and Vere-Jones, D. (2002). Stochastic declustering of space-time earthquake occurrence, *Journal of the American Statistical Association*, *97*, 369-380.

APPENDIX 1 A NEW GENERIC MODEL FOR AFTERSHOCK OCCURRENCE

A new generic model for aftershock occurrence

Annemarie Christophersen and Matt (Matthew Charles) Gerstenberger

GNS Science

PO Box 30368, Lower Hutt 5040, New Zealand

D+64 4 570 4772, F+64 4 570 1440

Email: A.Christophersen@gns.cri.nz

Manuscript submitted to the Bulletin of the Seismological Society of America

25 June 2010

Abstract

We introduce a new generic model for aftershock decay and propose it as an alternative to an earlier aftershock model that has become known as the Reasenberg and Jones model. The Reasenberg and Jones model assumes that the growth parameter for productivity is the same as the *b*-value of the magnitude frequency distribution of the aftershocks. We develop an alternative description for an average productivity based on how the mean number of aftershocks increases with main shock magnitude. We derive parameters for our new model for California, Italy and New Zealand, and describe in detail the factors that influence the model parameters. We show that the growth parameter α for the productivity depends on the aftershock selection and is significantly larger than the *b*-value for all regions. The average productivity differs by more than a factor of 10 between the new and the old model. The new generic model and the Reasenberg and Jones model have both been implemented in the Short Term Earthquake Probability (STEP) model are currently tested in the testing centers of the Collaboratory for the Study of Earthquake Predictability (CSEP).

Introduction

The U. S. Geological Survey (USGS) has forecast the probability of aftershock occurrence following a major earthquake in California since 1989 [*Reasenberg and Jones*, 1989; 1990; *Reasenberg and Jones,* 1994]. Since 2005, the USGS has provided maps that forecast the probability of strong shaking at any location in California within the next 24-hours based on the Short Term Earthquake Probability (STEP) model [*Gerstenberger et al.*, 2004; *Gerstenberger et al.*, 2005]. Both forecasts rely on an aftershock decay model, which has become known as the Reasenberg and Jones model. The model is commonly applied as a reference model for aftershock decay, e.g. in estimating the risk [*Stiphout et al.*, 2010] or to compare to induced seismicity [*Bachmann et al.*, 2009]

According to the Reasenberg and Jones model, the rate of earthquakes of magnitude M_{min} and larger, following a main shock of magnitude M_m is assumed to decay with time t according to a power law

$$\dot{n}(M_{\min}, t) = \frac{10^{a' + b(M_m - M_{\min})}}{(t + c)^p} \tag{1}$$

Equation (1) is identical to the Omori-Utsu law for aftershock decay (see [*Utsu et al.*, 1995] for a review paper), except that the productivity parameter for an individual main shock was replaced by the numerator in equation (1). The meaning of the constant c is still being debated (e.g.[*Enescu et al.*, 2009]; [*Narteau et al.*, 2009]) but its presence allows for a time delay that is either due to physical properties, network properties, or both; the parameter p is the power law exponent that controls the temporal decay. The numerator of equation (1) describes how the aftershock rate increases with increasing main shock M_m and decreasing minimum magnitude M_{min} as a function of the productivity parameter a' and the growth exponent b. The parameter b is derived from the Gutenberg-Richter b-value of the magnitude-frequency distribution of aftershock sequences. Thus the Reasenberg and Jones model implicitly assumes that aftershock rate increases as a function of main shock magnitude with the same parameters as the frequency of large earthquakes decreases in an aftershock sequence. Other aftershock studies refer to the growth exponent of aftershock number with main shock magnitude as α (e.g. [*Felzer et al.*, 2004; *Helmstetter*, 2003; *Zhuang et al.*, 2004]). While it is a common assumption that $\alpha = b$ [e.g. *Felzer et al.*, 2003], the two parameters can differ significantly depending on the definition of aftershocks in space, time and magnitude (Christophersen and Smith, 2008). For main shocks in the magnitude range 2.0 to 6.0 in Southern California the parameter α ranges from 0.71±0.05 to 0.96±0.08 depending on the clustering algorithm [*Christophersen and Smith*, 2008]. As α and b are exponents, any differences in their values can have a large effect on forecast aftershock rates. We derive a new model for an average productivity, which is based on how the mean number of aftershocks grows with main shock magnitude. The new model has three parameters, which depend on a number of factors, in particular completeness of the data set. We determine parameters for the new equations with two aims in mind: comparing the new model parameters with the Reasenberg and Jones model in California and deriving parameters for the STEP model in several regions. Consequently we apply two different clustering mechanisms for defining aftershocks, the Reasenberg algorithm [*Reasenberg*, 1985] used by Reasenberg and Jones and a method, which we have developed to be consistent with the STEP model selection criteria for aftershocks in space. We explore the effect of the different clustering algorithm for California, in particular the dependence on minimum magnitude M_{min} to highlight potential biases due to data selection. Finally we compare the STEP clustered results for aftershock productivity for California, Italy and New Zealand with previously derived results for the Reasenberg and Jones model. We find that the Reasenberg and Jones model with the generic parameters for each region over-predicts aftershock rate in comparison to our new model by a factor of up to 10. We note here that Reasenberg and Jones also presented methods to up-date the model parameters for an on-going sequence (1989, 1990). However, we only compare the generic parameters with new generic parameters from the mean abundance model. The STEP model with Reasenberg and Jones parameters is under evaluation in the earthquake forecast centers in California, New Zealand and Italy and with the new generic model parameters for the later two regions, as part of the Collaboratory for the Study of Earthquake Predictability (CSEP; www.cseptesting.org).

A new generic model for aftershock decay

The decay of aftershock rate with time t after a main shock is described empirically by the Omori-Utsu law for an individual aftershock sequence

$$\dot{n}(t) = \frac{K}{(t+c)^{p}} \, , \tag{2}$$

with the parameters, c and p already explained above. The parameter K represents the productivity and depends on the definition of earthquake clusters in time and space. Furthermore, K is typically calculated using only events within some area A and above a minimum magnitude M_{min} that often is identical to the completeness magnitude above which all aftershocks have been recorded. The number of aftershocks in area A and time interval [S, T] can then be derived from the Omori-Utsu law as follows:

$$N = \int_S^T \frac{K(A, M_{\min})}{(t+c)^p}\, dt = K(A, M_{\min}) \int_S^T (t+c)^{-p}\, dt = K(A, M_{\min}) I_{OU}(S,T) \tag{3}$$

We call $I_{OU}(S,T)$ the Omori-Utsu integral where

$$I_{OU}(S,T) = \ln\left(\frac{T+c}{S+c}\right) \text{ for } p=1 \text{ and } I_{OU}(S,T) = \frac{(T+c)^{1-p} - (S+c)^{1-p}}{(1-p)} \text{ for } p \neq 1. \tag{4}$$

To derive an average K, we relate equation (3) to mean abundance, i.e. the mean number of aftershocks per main shock of a given magnitude. Mean abundance grows exponentially with main shock magnitude and can be described as [*Christophersen and Smith*, 2008]

$$N_{ave}(M_m, M_1) = 10^{\alpha(M_m - M_1)} \tag{5}$$

The parameter α is the growth parameter. The magnitude M_1 corresponds to the main shock magnitude that has a mean abundance of 1.0, or in other words to the main shock magnitude that on average has at least one aftershock above the minimum magnitude M_{min} within the selected aftershock area A and the time interval [S, T]. The magnitude M_1 is a function of the minimum magnitude, the aftershock area and the time interval. It increases with increasing minimum magnitude and generally decreases with increasing area and/or time interval. The parameter α also depends on the selection of aftershocks in space as we will illustrate with our two clustering algorithms. However, the minimum magnitude and the selected time interval have no effect on α as long as they apply to all main shock magnitudes uniformly.

By replacing N in equation (3) with N_{ave} above, we can derive an average productivity K_{ave}.

$$K_{ave}(M_{\min}, A, S, T) = \frac{10^{\alpha(M_m - M_1)}}{I_{OU}} \tag{6}$$

The Omori-Utsu integral depends on the parameters p and c, as well as the time period [S, T]. In table 1 we show the percentage change of the Omori-Utsu integral for a number of p and c values compared to p=1.0 and c=0.01 days for the time interval [0.1, 30] days. These examples illustrate that for a c-value of 0.01, a change in p by ± 0.1 causes I_{OU} to change around 6%, and a change in p to the very small value of 0.7 causes the I_{OU} to changes by around 30%. They also show that the c-value has no large effect on I_{OU} as long as it is smaller than the start time S for counting aftershocks. Therefore the exact values of p and c are not so crucial for determining an average K from mean abundance data. We select p=1.0 and c=0.01.

To determine mean abundance, we first apply a clustering algorithm to an earthquake catalog to define earthquake sequences in time and space. For each main shock magnitude we then count the number of aftershocks above a selected minimum magnitude of interest M_{min} and within a selected time interval [S, T] following the main shock. Single earthquakes are included in this method as main shocks without aftershocks.

While counting the number of aftershocks per main shock and averaging the number for each main shock magnitude to determine mean abundance sounds simple, the process soon becomes complex when considering data consistency, homogeneity and completeness. As the mean abundance parameters depend on the selection of aftershocks in space and time, one aspect of consistency is to select aftershocks according to the application of parameters to be derived.

Completeness problems arise mainly because smaller earthquakes cannot be detected in the coda of larger events. Furthermore, large earthquakes often generate so many aftershocks that the coda of all earthquakes cannot always be resolved and not all earthquakes can be processed. For Southern California, an estimate of the completeness magnitude as a function of time t following a main shock of magnitude M_m is given by

$$M_C(M_m, t) = M_m - 4.5 - 0.75 * \log_{10}(t) \tag{7}$$

with a minimum magnitude of completeness m_0 of 2.0 [*Helmstetter et al.*, 2006]. Thus a magnitude 6.5 main shock should reach aftershock detection at m_0 after one day. For the largest earthquake in the catalogue, the 1992 Landers event of magnitude 7.3, reaching m_0 would take more than 10 days according to equation (7). Due to the nature of the Omori-Utsu law most aftershock activity happens shortly after the main shock. Thus, using an optimal completeness threshold in the initial days of a sequence is critical and we need to find a compromise between a low minimum magnitude and an early start time S.

Figure 2 gives an example of mean abundance for California with minimum magnitude 2.0 and time interval [0.1,30] for the Reasenberg and STEP clustering regardless of general magnitude completeness considerations. The minimum magnitude M_{min} is chosen equal to the smallest main shock magnitude. Due to magnitude uncertainties and the relatively small magnitude range available to select aftershocks from, a significant number of aftershocks might be missed for the smallest main shocks. As a consequence the mean abundance for main shocks rolls off close to the minimum magnitude. We avoid this bias by excluding main shock magnitudes within 0.3 magnitude units of the minimum magnitude from the mean abundance fitting.

Once all completeness and data consistency issues are taken care of, the logarithm of mean abundance is fitted by a robust linear regression. The algorithm uses iteratively reweighted least squares with a bisquare weighting function.

The clustering algorithms

We have two goals: deriving parameters for the mean abundance model that are comparable to the Reasenberg and Jones model in California and deriving parameters to employ in the STEP model for earthquake forecasting in different testing regions. Therefore we use two different clustering algorithms: the Reasenberg (1985) clustering algorithm and a method that we have developed to be consistent with the STEP model.

Reasenberg clustering

We use the Reasenberg declustering algorithm, a common linking algorithm based on a simple interaction radius. We use the default parameters for California [*Reasenberg*, 1985].

STEP clustering

The second clustering method is a method that is selected to match the spatial smoothing used by the STEP model [*Gerstenberger et al.*, 2004; *Gerstenberger et al.*, 2005]. The catalogue is processed chronologically and each earthquake qualifies as a potential main shock. A magnitude dependent area is searched with the radius chosen as the larger of 5 km or the equivalent of the magnitude dependent sub-surface rupture area according to [*Wells and Coppersmith*, 1994]

$$r(M) = 10^{0.59M - 2.44} \tag{9}$$

The minimum of 5km was selected to optimize location errors and minimum grid spacing as used by the STEP model. The problem is that, for earthquakes of magnitude 5.3 and below, the search radius is always 5 km; thus the smaller the main shock magnitude is, the more likely background earthquakes that are not part of the cluster are to be included in a cluster. As a consequence smaller main shocks have more aftershocks in relation to their size and this can affect some scaling relationships.

In time, the STEP model continues to identify earthquakes with an aftershock sequence until the forecast rate as calculated from previous events falls below the background rate. We have simplified the temporal definition of aftershocks by using rolling time windows of length ΔT. Each time a new earthquake above the cut-off magnitude is associated with a cluster, the time window to associate further earthquakes is extended by ΔT. We selected ΔT=30 days.

Discussion of the two search algorithms

Figure 1 shows the spatial extent of the earthquake clusters for the ANSS catalog in California cut at magnitude 2.0, again, regardless of completeness magnitude considerations. For each cluster, we found the largest distance between the main shock, and all other earthquakes in the cluster. We plotted the mean of this distance per main shock magnitude, including all the zero distances for single event clusters. The solid line in figure 1 is the STEP search radius. For main shocks of magnitude 5.5 and larger, the data are close to the line, indicating that the search radius constrains the growth of the aftershock area. The Reasenberg data extend above the STEP search radius in the magnitude range 5.0 to 7.0. For the Reasenberg linking algorithm the mean spatial extent increases roughly linearly as a function of main shock magnitude. However, there is a trend for main shocks below about magnitude 2.8 to fall below this linear trend. This is most likely due to lack of completeness for all magnitudes in the data set. The STEP data trend upwards in relation to the Reasenberg data below magnitude 4, as, most likely, unrelated background seismicity is included in the relatively large search radius of 5 km. The mean abundance is consequently biased towards higher productivity for smaller main shocks. Figure 2 gives an example of mean abundance in the time interval 0.1-1 day with minimum magnitude 2.0 for California. The pattern seen in the spatial extent of the clusters is also seen here; the mean abundance for STEP trends upwards for main shocks below magnitude 4.0 relative to the Reasenberg derived abundance. As a consequence, the slope α and the parameter M_1 are reduced for smaller main shocks and the STEP parameters have a stronger dependence on the main shock magnitude than the Reasenberg declustered data, as we will demonstrate in detail for California.

New generic aftershock parameters for California

The earthquake catalog

We downloaded the ANSS catalogue for California for the time period 1 January 1984 to 31 December 2009 for earthquakes with minimum magnitude 0.1 We removed 21 nuclear explosions including one aftershock following the procedure of [*Werner et al.*, 2009]. We selected earthquakes within the CSEP collection area as defined by [*Schorlemmer and Gerstenberger*, 2007]. The resulting catalog has 1,008,894 earthquakes. We assumed a completeness magnitude of 3.0. We confirmed this completeness magnitude by spatially analyzing the completeness magnitude of the complete and the Reasenberg declustered catalog. While some areas have good detection capability below magnitude 1.0, the southern and north western parts of the region have completeness magnitudes around 3.0. The catalog of earthquakes with magnitude 3.0 and larger has 17,630 earthquakes.

Mean abundance in California

To study the dependence on the main shock magnitude in detail, we determined mean abundance as a function of minimum magnitude. Figure 3 shows mean abundance for two selected minimum magnitudes for the Reasenberg clustering. The mean abundance was determined in the time interval 0.1-30 days. Only the fitting range is shown. We excluded mean abundance data within 0.3 magnitude units from the minimum magnitude to account for incompleteness due to magnitude uncertainty. We further excluded mean abundance data for main shock magnitudes larger than 3.75 magnitude units above the minimum magnitude to account for completeness issues according to equation (7). The figure illustrates that there is some scatter in the data. Even though the minimum magnitudes of the two samples are 1.5 magnitude units apart, the mean abundance data are close around magnitude 6.1-6.3. For minimum magnitude 4.5, some main shock magnitudes have abundance zero and therefore are not included in the linear regression of the logarithm of mean abundance. As a consequence the mean abundance fit is biased towards smaller M_1, because sequences with low abundance are likely to be dropped from the linear regression analysis first. The slope α also tends to be smaller when data points are excluded from the fitting due to a mean abundance of zero. We note this as problem of the linear regression and in the analysis below exclude minimum magnitudes that have main shocks with zero abundance. To avoid this problem in future, we would need to fit an exponential distribution directly to the mean abundance data.

Figure 4 shows the slope α as a function of minimum magnitude for the Reasenberg (triangles) and STEP (circles) clustering and the 95% confidence intervals. As expected due to the spatial selection criteria, α is systematically smaller for the STEP method compared to Reasenberg. However, except for the smallest minimum magnitude, the 95% confidence intervals of α overlap for both methods. For minimum magnitude larger than 4.0 for Reasenberg and 4.2 for STEP, α decreases. This is caused by some data points dropping from the logarithmic fit. Averaging α and its 95% confidence intervals for the Reasenberg algorithm in the magnitude range 3.0 – 3.6, where no mean abundance equals zero, results in α = 1.07±0.13. Figures 5 and 6 show the mean abundance parameter and the difference of M_1 and the minimum magnitude for both clustering algorithms. As expected, the STEP clustering results in smaller M_1 for smaller minimum magnitudes. For minimum magnitude 3.5 and above, both clustering algorithms have similar M_1 values. For the Reasenberg clustering algorithm, the difference Δ between M_1 and minimum magnitude is not magnitude dependent, except for the bias at higher minimum magnitudes when data points drop from the linear regression. Averaging Δ and its 95% confidence interval in the magnitude range 3.0 – 3.6, where no mean abundance equals zero, results in Δ= 1.65±0.09. Therefore the parameter M_1 for the time interval 0.1 – 30 days is M_1=M_{min}+1. 65±0.09. The STEP clustering algorithm has a magnitude dependent Δ, again due to the constant search radius in space for smaller main shocks. For the target magnitude 4.0, α = 1.01±0.18 and M_1 = 5.62±0.12. Figure 7 compares the *K*-value for minimum magnitude 4.0 for the Reasenberg and Jones model (solid line) with the mean abundance model for STEP (solid line with circles) and Reasenberg (solid line with traiangles) clustering. The dashed line shows a variation of the Reasenberg-Jones model with a *b*-value of 1.0 and a *K*-value of 0.1163 for a mean main shock magnitude of 6.04 and M_{min}=4.8 from 73 stacked aftershock sequences in California [*Felzer et al.*, 2003]. Also included are data from the 62 sequences that Reasenberg and Jones (1989) used to determine the first set of generic Californian parameters. The model parameters were published by [*Gasperini and Lolli*, 2006]. We scaled the *K*-value for each cluster with the cluster-specific *b*-value of the magnitude-frequency relation. In the magnitude range where data is available, the data scatter more than the difference between the models. However, the Reasenberg and Jones model exceeds the abundance model by about a factor of 3 in the magnitude range 5.5 – 7.0. Due to the difference in slope between the models, the *K*-value of the Reasenberg and Jones model exceeds the new mean abundance *K*-value by a factor of around 10 for main shocks as small as magnitude 3.0. A bias of the Reasenberg and Jones model to higher aftershock rates is caused by using only well-recorded aftershock clusters for the parameter determination and therefore neglecting clusters with few or no aftershocks. The bias was previously thought to be in the order of a factor of 2 [*Felzer et al.*, 2003]. The difference in the slope parameters between the Reasenberg and Jones model and the mean abundance model enhances the discrepancy, especially for small main shocks. This pattern can also be seen in other regions.

New generic parameters for other regions

Versions of STEP that include the new generic abundance model have already been installed in the testing centers in Italy and New Zealand. In both cases, mean abundance parameters were derived in the time interval 0.1 – 30 days for a minimum magnitude of 3.95. Figure 8 compares the Reasenberg and Jones model (dashed lines) for New Zealand (triangles), California (squares) and Italy (circles) with the new generic model (dashed lines). All model parameters are listed in table 2. Quantitative comparisons between regions need to be done with care as we have not investigated how the magnitude scales compare between regions. However, qualitatively all regions show the same relationship between the Reasenberg and Jones and the new generic model; for most of the magnitude range, the Reasenberg and Jones model has a higher productivity than the generic model. This is a general bias of the Reasenberg and Jones model and comes from two aspects of the model: 1) only aftershock sequences with sufficient events to fit the parameters of the Omori-Utsu law are typically used when fitting the Reasenberg and Jones model as has been described by [*Felzer et al.*, 2003].; and 2) the slope of the new generic model is steeper in all regions than the slope of the Reasenberg and Jones model.

Discussion and conclusions

We have derived a new generic description of the average aftershock productivity parameter in the Omori-Utsu law for aftershock decay. We calculate aftershock productivity from mean abundance, the mean number of aftershocks per main shock magnitude. Mean abundance has two parameters: the growth parameter α that describes how the mean number of aftershocks increases with main shock magnitude and the magnitude M_1 that on average has one aftershock above completeness. We have discussed in detail the factors that influence mean abundance, in particular completeness of the data set. We showed that the minimum search radius of 5 km in the STEP declustering, which is selected to account for uncertainty in earthquake location and for grid spacing in earthquake forecasting, has a systematic effect on the mean abundance parameters. Therefore deriving parameters for a small minimum magnitude, e.g. 3.0, will overpredict aftershock rates for a larger minimum magnitude, e.g. 4.0. However, if we want to forecast aftershocks of magnitude 3.0 and larger, then it would be appropriate to use the same minimum magnitude for deriving parameters because we want to know about the earthquakes that happen within the search radius. Apart from having to take care of sensitivity to data selection and ensuring the data for deriving parameters are selected consistent to their application, the parameters of the mean abundance model are easy to derive. We propose our model as an alternative to the commonly used Reasenberg and Jones model in modeling generic aftershock occurrence. The key difference between the two models is twofold: 1) the mean abundance model derives parameters from all earthquake clusters, including single earthquakes and thus avoids a bias towards large aftershock sequences, and; 2) the growth parameter of the aftershock productivity is derived from mean abundance rather than the magnitude-frequency distribution of aftershocks. For our three sample regions, California, Italy and New Zealand, α exceeds the b-value by 0.1-0.3. As a consequence larger main shocks are relatively more productive in the mean abundance model than in the Reasenberg and Jones model, and smaller main shocks are more productive in the Reasenberg and Jones model. We have shown that the assumption that that b equals α does not hold but that α depends strongly on the definition of aftershocks in space.

Data and Resources

The Californian earthquake catalog was searched on the ANSS site (http://www.ncedc.org/anss/catalog-search.html; last accessed on 13 February 2010).

For Italy, the CSI1.1 catalog was used (http://csi.rm.ingv.it/; last accessed in September 2009).

For New Zealand, geonet catalog was used (http://www.geonet.org.nz/earthquake/; last accessed March 2010).

Acknowledgments

The research for this paper was supported by the European project NERIES (026130), by the New Zealand Earthquake Commission and the New Zealand Foundation of Research, Science and Technology.

References

Bachmann, C., J. Woessner, and S. Wiemer (2009), A new probability-based monitoring system for induced seismicity: Insights from the 2006-2007 Basel earthquake sequence, *Seism. Res. Lett.,*, *80*(2), 327.

Christophersen, A., and E. G. C. Smith (2008), Foreshock rates from aftershock abundance, *Bull. Seism. Soc. Am.*, *98*(5), 2133-2148.

Enescu, B., J. Mori, M. Masatoshi, and K. Yasuyuki (2009), Omori_Utsu law c-values associated with recent moderate earthquakes in Japan, *Bull. Seismol. Soc. Am.*, *99*(2a), 884-891.

Felzer, K. R., R. E. Abercrombie, and G. Ekstrom (2003), Secondary aftershocks and their importance for aftershock forecasting, *Bull. Seismol. Soc. Am.*, *93*(4), 1433-1448.

Felzer, K. R., R. E. Abercrombie, and G. Ekstrom (2004), A common origin for aftershocks, foreshocks, and multiplets, *Bull. Seismol. Soc. Am.*, *94*(1), 88-98.

Gasperini, P., and B. Lolli (2006), Correlation between the parameters of the aftershock rate equation: Implications for the forecasting of future sequences, *Phys. Earth Planet. Int.*, *156*, 41-58.

Gerstenberger, M., S. Wiemer, and L. Jones (2004), Real-time Forecasts of Tomorrow's Earthquakes in California: a New Mapping Tool, *United States Geological Survey Open-File Report,, 2004-1390*.

Gerstenberger, M. C., S. Wiemer, L. M. Jones, and P. A. Reasenberg (2005), Real-time forecasts of tomorrow's earthquakes in California, *Nature*, *435*(7040), 328-331.

Helmstetter, A. (2003), Is earthquake triggering driven by small earthquakes?, *Phys. Rev. Lett.*, *91*(5).

Helmstetter, A., Y. Y. Kagan, and D. D. Jackson (2006), Comparison of short-term and time-independent earthquake forecast models for Southern California, *Bull. Seismol. Soc. Am.*, *96*, 90-106.

Lolli, B., and P. Gasperini (2003), Aftershock hazard in Italy Part I: Estimation of time-magnitude distribution model parameters and computation of probabilities of occurrence , *J. Seismol.*, *7*, 235-257.

Narteau, C., S. Byrdina, P. Shebalin, and D. Schorlemmer (2009), Common dependence on stress for the two fundamental laws of statistical seismology, *Nature*, *462*(7273), 642-645.

Pollock, D. (2007), Apects of short-term and long-term seismic hazard assessment in New Zealand.

Reasenberg, P. A. (1985), Second-order moment of Central California Seismicity, *J. Geophys. Res.*, *90*, 5479-5495.

Reasenberg, P. A., and L. M. Jones (1989), Earthquake hazard after a mainshock in California, *Science*, *243*, 1173-1176.

Reasenberg, P. A., and L. M. Jones (1990), California aftershock hazard forecast, *Science*, *247*, 345-346.

Reasenberg, P. A., and L. M. Jones (1994), Earthquake aftershocks: Update, *Science*, *265*, 1251-1252.

Schorlemmer, D., and M. Gerstenberger (2007), RELM Testing Centre, *Seism. Res. Lett.,*, *78*(1), 30-36.

Stiphout, T., S. Wiemer, and W. Marzocchi (2010), Are short-term evacuations warranted? The case of the 2009 L'Aquila earthquake, *Geophys. Res. Lett.*, *submitted*.

Utsu, T., Y. Ogata, and R. S. Matsu'ura (1995), The centenary of the Omori formula for a decay law of aftershock activity, *Journal of the Physics of the Earth*, *43*, 1-33.

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., A. Helmstetter, D. D. Jackson, and Y. Y. Kagan (2009), High Resolution Short-Term and Long-Term Earthquake Forecasts for California, *Manuscript in preparation for submission to BSSA.*

Zhuang, J. C., Y. Ogata, and D. Vere-Jones (2004), Analyzing earthquake clustering features by using stochastic reconstruction, *J. Geophys. Res.-Solid Earth*, *109*(B5).

Authors' affiliations, addresses

GNS Science

PO Box 30368, Lower Hutt 5040, New Zealand

Tables

Table 1: Percentage change of the Omori-Utsu integral I_{OU} relative to p=1.0 and c=0.01 for a range of p- and c-values for S=0.01 days, T=30 days.

c\p	0.70	0.80	0.90	1.10	1.20	1.30
0.001	35%	20%	9%	-3%	-4%	-3%
0.01	34%	19%	8%	-5%	-7%	-6%
0.02	33%	18%	6%	-6%	-9%	-9%
0.03	33%	17%	5%	-8%	-11%	-12%
0.05	31%	15%	3%	-11%	-15%	-16%
0.10	28%	12%	-1%	-17%	-22%	-25%
0.20	24%	6%	-7%	-26%	-32%	-36%
0.50	15%	-4%	-18%	-39%	-46%	-52%

Table 2: Model parameters for the new generic and the Reasenberg and Jones model for California [*Reasenberg and Jones,* 1989], Italy [*Lolli and Gasperini*, 2003] and New Zealand [*Pollock*, 2007]

	Abundance model with STEP clustering (30 days)			Reasenberg and Jones model			
	α	M_1	I_{OU}	a’	b	p	c
California	1.01±0.18	5.62±0.12	5.6088	-1.67	0.91	1.08	0.05
Italy	1.31±0.21	5.39±0.11	5.4680	-1.84	0.98	0.92	0.09
New Zealand	1.25±0.21	5.4±0.21	5.1355	-1.59	1.03	1.07	0.04

Figure Captions

Figure 1: The average spatial cluster extension measured by determining the maximum distance between main shock and all aftershock for each cluster. The average cluster extension includes single main shocks. The circles represent the STEP clustered data, the triangles the Reasenberg clustered data. The solid line represents the STEP search radius as function of main shock magnitude.

Figure 2: One example of mean abundance in the time interval 0.1 – 10 days with minimum magnitude 2.0 for California. Again, the circles represent the STEP clustered data, the triangles the Reasenberg clustered data.

Figure 3: Two examples of mean abundance for Reasenberg clustered data in the time interval 0.1-30 days for minimum magnitude 3.0 and 4.5 in California including the best fitting model. Only data in the fitting interval are plotted which starts 0.3 magnitude units to account for magnitude uncertainties and ends 3.7 magnitude units above the minimum magnitude to account for the completeness of aftershocks. Six data points are lost to the fitting for M_i=4.5 because the mean abundance is zero. The resulting model is therefore biased towards smaller alpha and M_1.

Figure 4: The mean abundance parameter α as a function of minimum magnitude for Reasenberg and STEP declustering methods including the 95% confidence intervals.

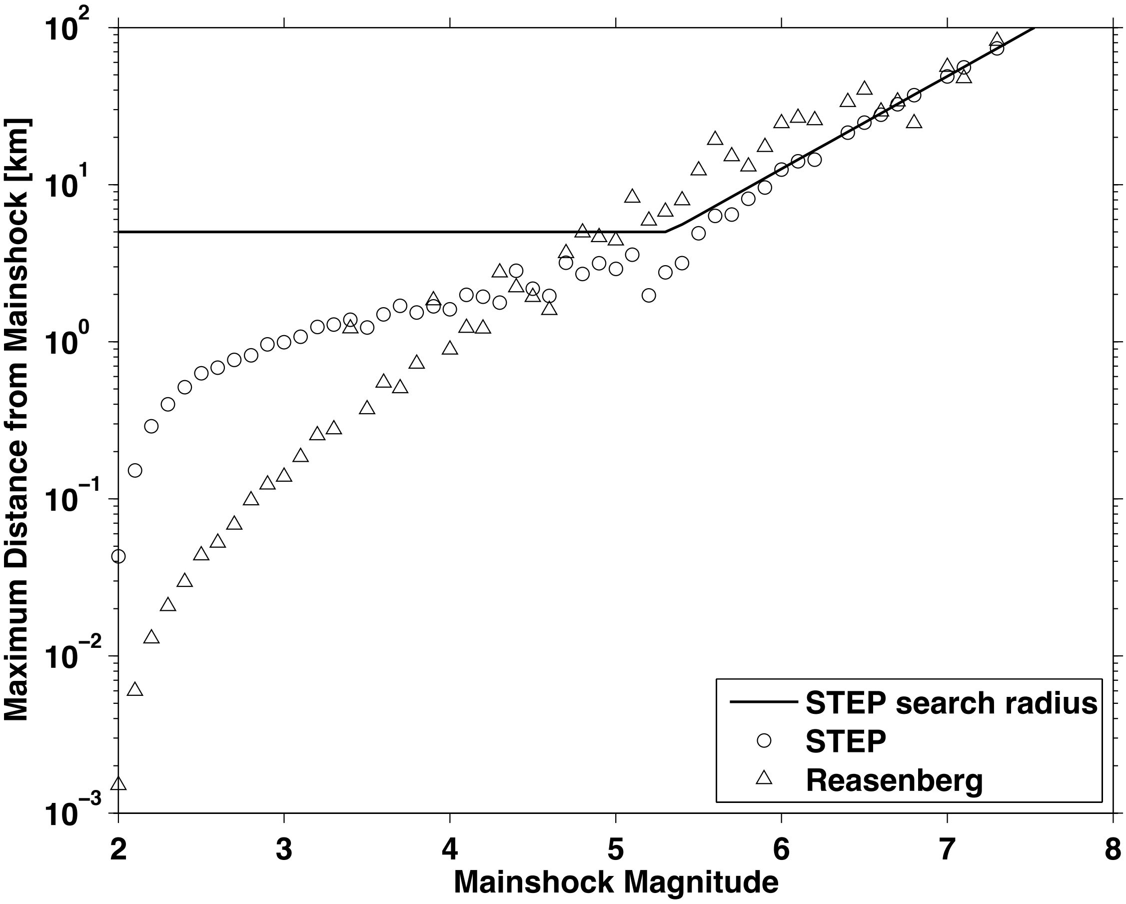
Figure 5: The mean abundance parameter M_1 as a function of minimum magnitude for Reasenberg and STEP declustering methods including the 95% confidence intervals.

Figure 6: The difference between the mean abundance parameter M_1 and the minimum magnitude as a function of minimum magnitude for Reasenberg and STEP declustering methods including the 95% confidence intervals. If the parameter M_1 was independent of minimum magnitude the data would fall on a uniform line. For the Reasenberg clustering the data are uniform within their 95% confidence intervals until some magnitudes drop from the mean abundance fitting because their mean abundance is 0 and therefore the fitted data is biased towards more productive sequences in the difference M_i- M_1 decreases. For the STEP data the difference M_i- M_1 increases with increasing minimum magnitude reflecting the bias of larger sequences for small M_i given by the fixed search radius.

Figure 7: Comparison of the *K*-parameter for the minimum magnitude 4.0 for the Reasenberg and Jones model, a variation of it and the abundance model for STEP and Reasenberg clustering. The circles represent the 62 sequences used by Reasenberg and Jones but adjusted for the minimum magnitude by using the *b*-value of the magnitude frequency distribution fitted for each cluster.

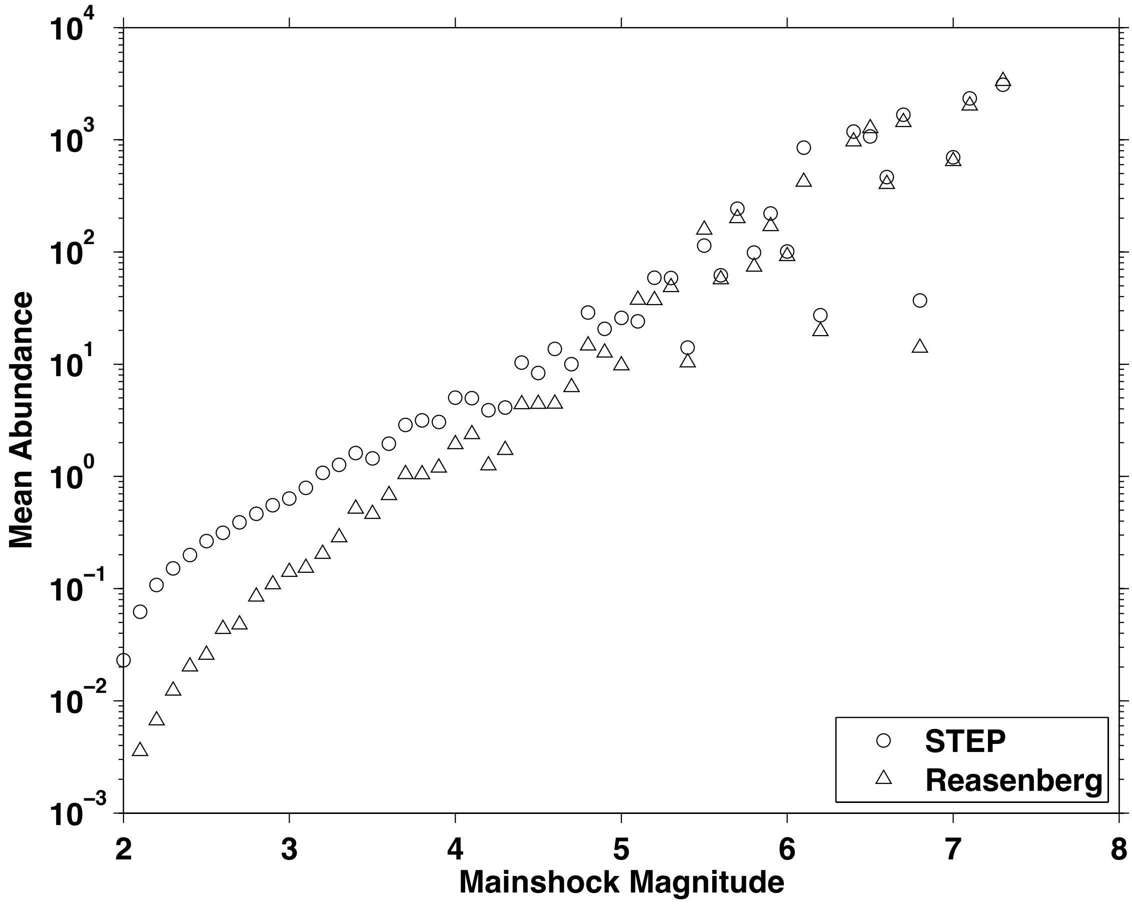
Figure 8: Regional comparison of Reasenberg and Jones models (dashed lines) and new generic model (solid lines) for New Zealand (triablges), California (squares) and Italy (circles).

Figure 1:



Maximum Distance from Mainshock [km]
10^2
10^1
10^0
10^−1
10^−2
10^−3
2
3
4
5
6
7
8
Mainshock Magnitude
STEP search radius
STEP
Reasenberg

Figure 2:



Mean Abundance
10^4
10^3
10^2
10^1
10^0
10^{-1}
10^{-2}
10^{-3}
STEP
Reasenberg
2
3
4
5
6
7
8
Mainshock Magnitude

Figure 3:

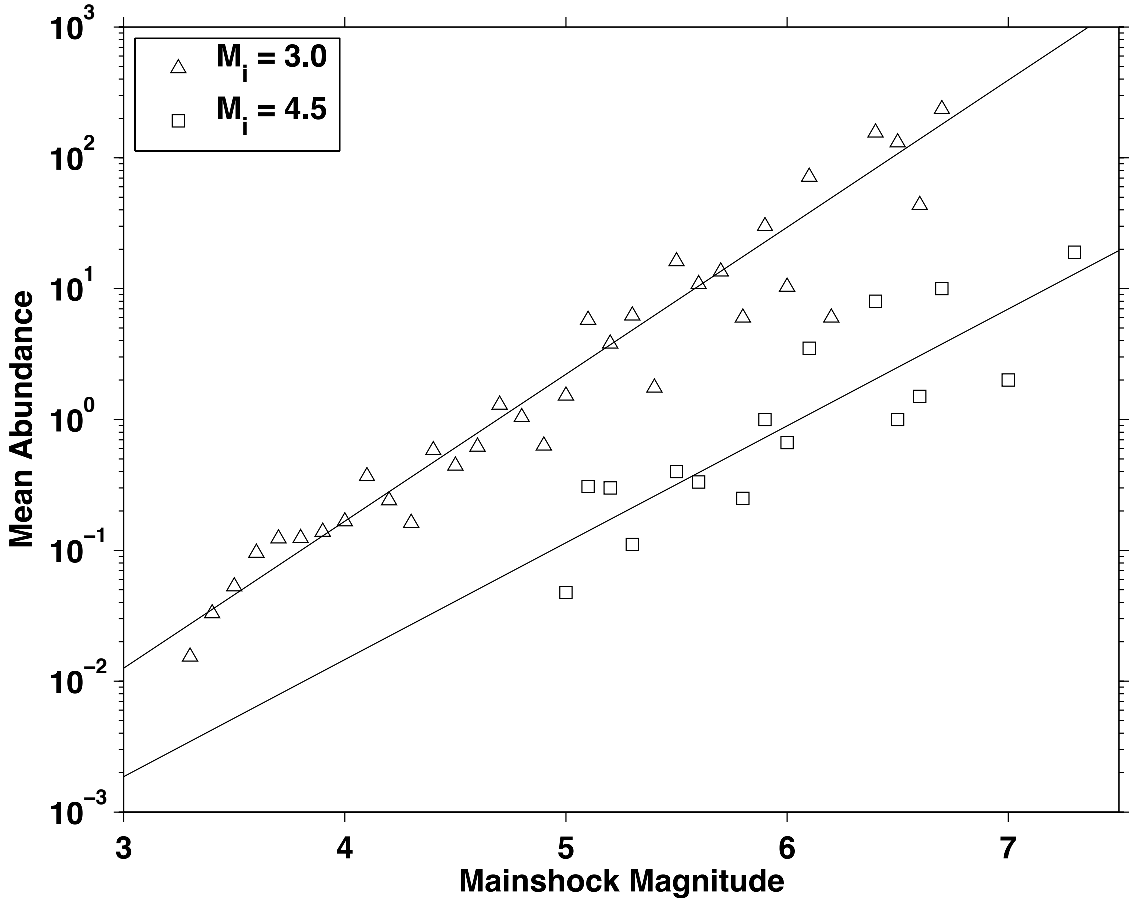
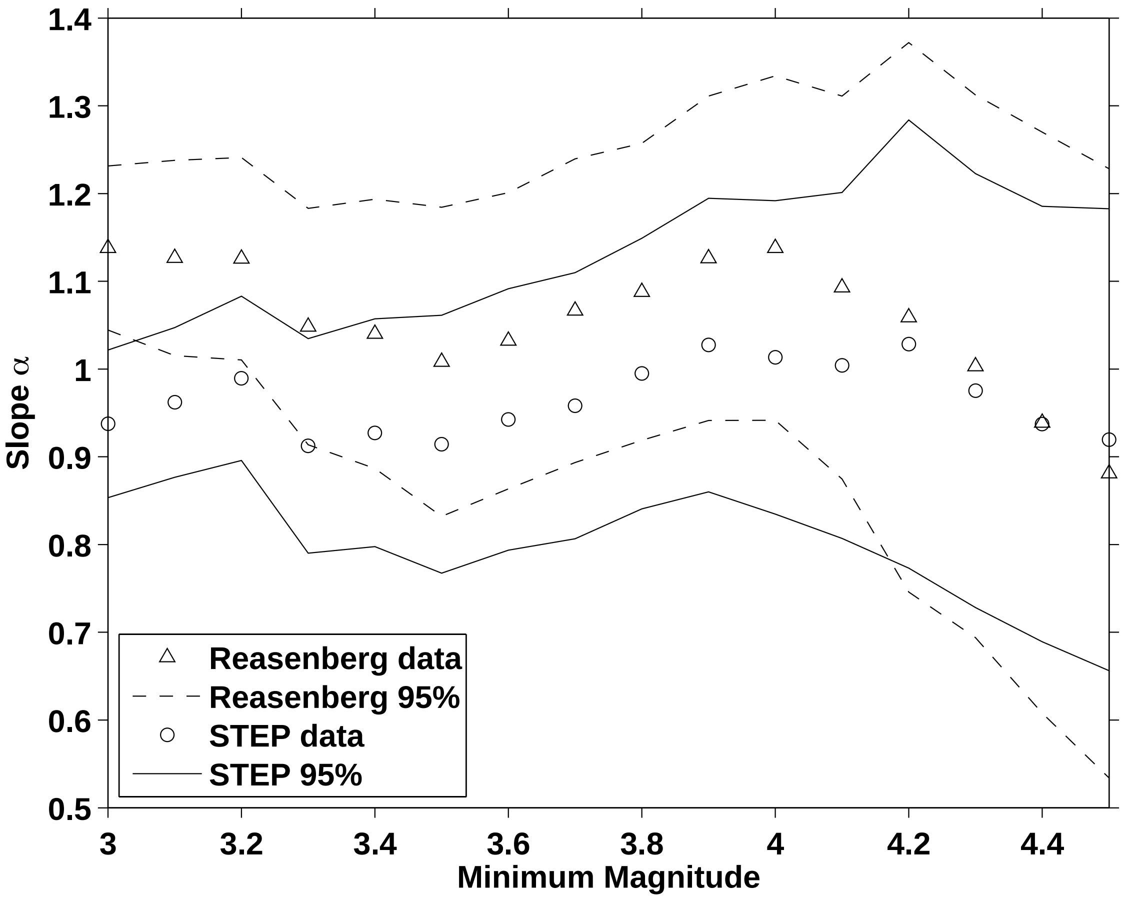
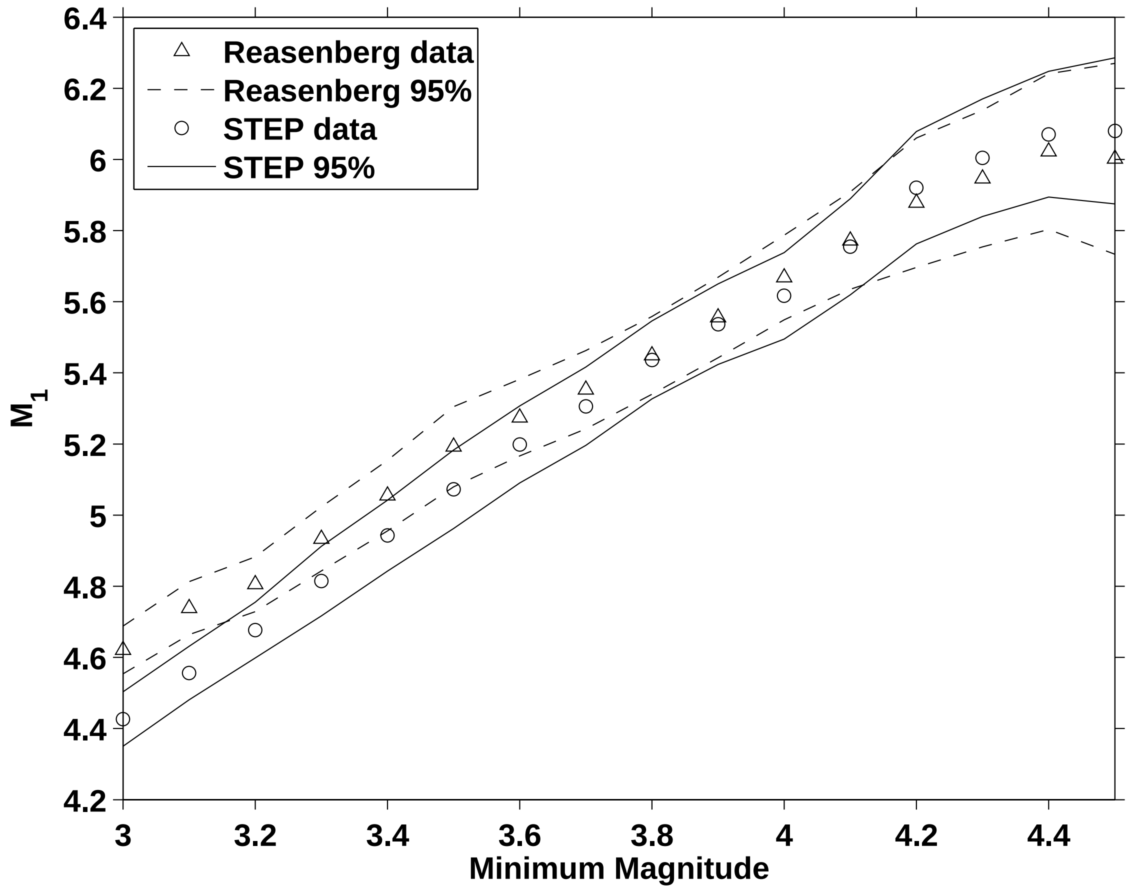

$M_i = 3.0$
$M_i = 4.5$
10^3
10^2
10^1
10^0
10^{-1}
10^{-2}
10^{-3}
Mean Abundance
3
4
5
6
7
Mainshock Magnitude

Figure 4:



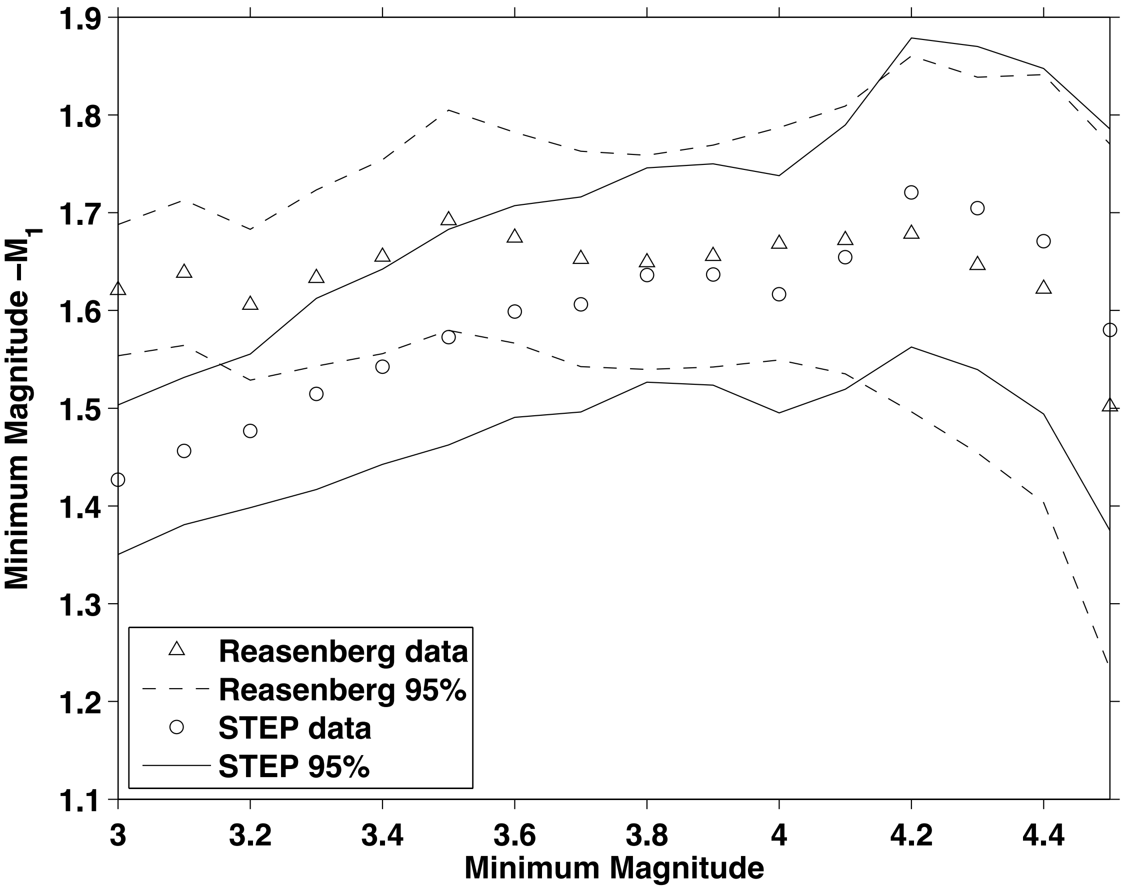
1.4
1.3
1.2
1.1
1
0.9
0.8
0.7
0.6
0.5
Slope α
3
3.2
3.4
3.6
3.8
4
4.2
4.4
Minimum Magnitude
Reasenberg data
Reasenberg 95%
STEP data
STEP 95%

Figure 5:



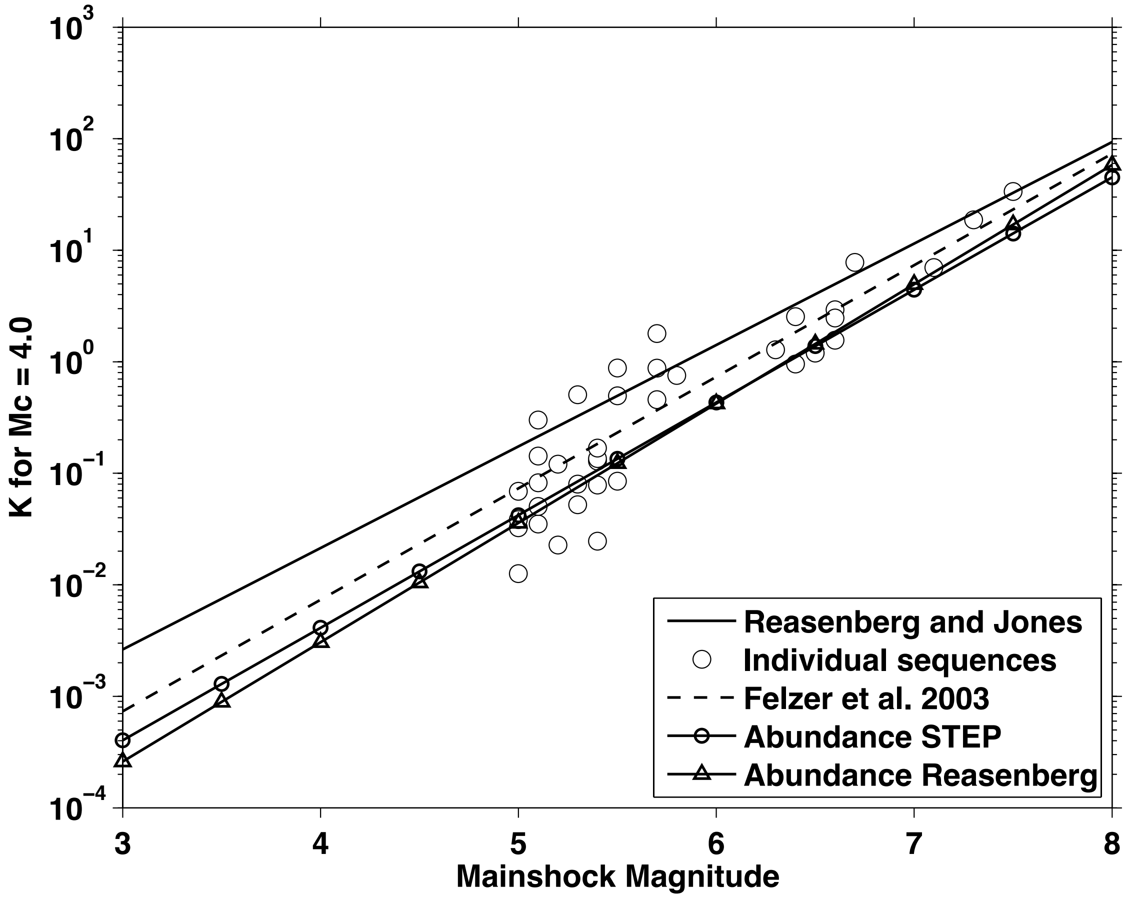
Reasenberg data
Reasenberg 95%
STEP data
STEP 95%
6.4
6.2
6
5.8
5.6
5.4
5.2
5
4.8
4.6
4.4
4.2
M1
3
3.2
3.4
3.6
3.8
4
4.2
4.4
Minimum Magnitude

Figure 6:



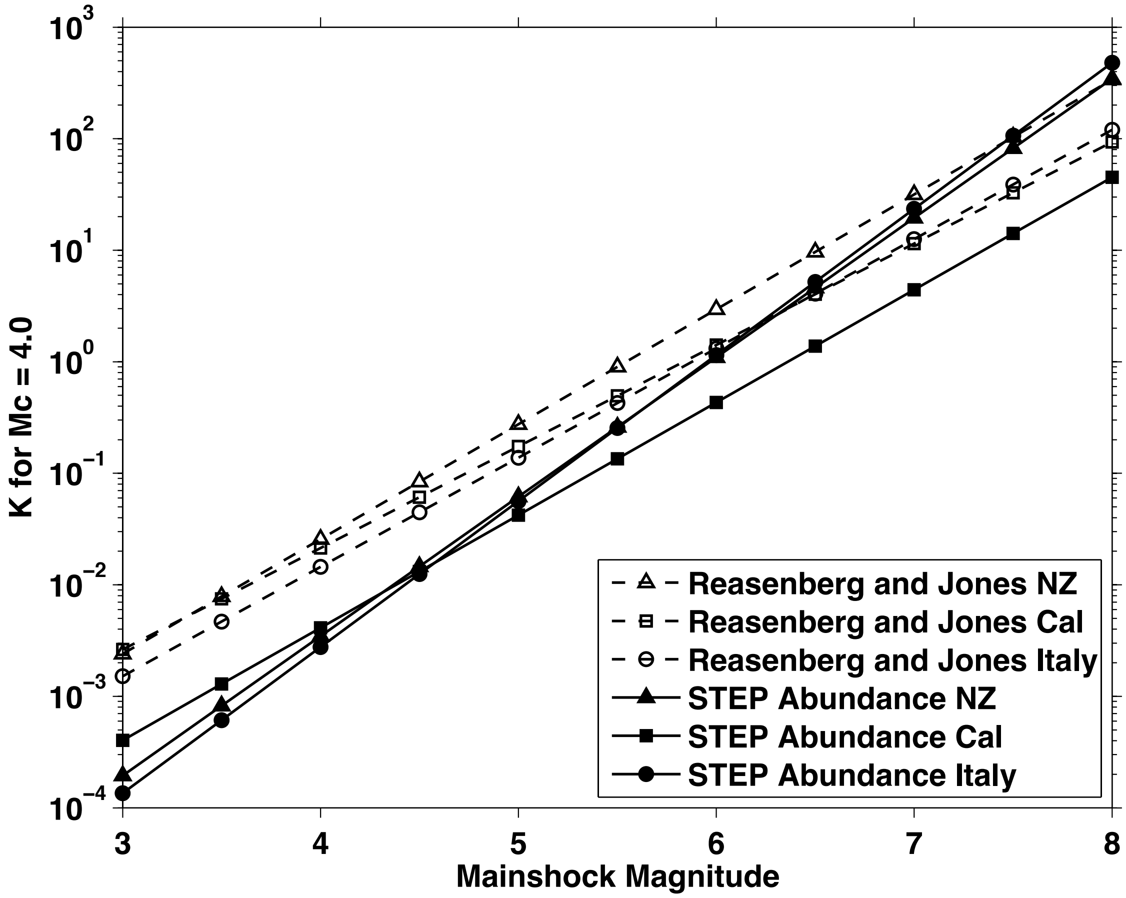
Reasenberg data
Reasenberg 95%
STEP data
STEP 95%
Minimum Magnitude −M1
Minimum Magnitude
1.9
1.8
1.7
1.6
1.5
1.4
1.3
1.2
1.1
3
3.2
3.4
3.6
3.8
4
4.2
4.4

Figure 7:



Reasenberg and Jones
Individual sequences
Felzer et al. 2003
Abundance STEP
Abundance Reasenberg
K for Mc = 4.0
Mainshock Magnitude
10^{3}
10^{2}
10^{1}
10^{0}
10^{-1}
10^{-2}
10^{-3}
10^{-4}
3
4
5
6
7
8

Figure 8:


K for Mc = 4.0
Mainshock Magnitude
10^3
10^2
10^1
10^0
10^-1
10^-2
10^-3
10^-4
3
4
5
6
7
8
Reasenberg and Jones NZ
Reasenberg and Jones Cal
Reasenberg and Jones Italy
STEP Abundance NZ
STEP Abundance Cal
STEP Abundance Italy

www.gns.cri.nz

Principal Location

1 Fairway Drive
Avalon
PO Box 30368
Lower Hutt
New Zealand
T +64-4-570 1444
F +64-4-570 4600

Other Locations

Dunedin Research Centre
764 Cumberland Street
Private Bag 1930
Dunedin
New Zealand
T +64-3-477 4050
F +64-3-477 5232

Wairakei Research Centre
114 Karetoto Road
Wairakei
Private Bag 2000, Taupo
New Zealand
T +64-7-374 8211
F +64-7-374 8199

National Isotope Centre
30 Gracefield Road
PO Box 31312
Lower Hutt
New Zealand
T +64-4-570 1444
F +64-4-570 4657