Utilising short-term and medium-term forecasting models for earthquake hazard estimation in the wake of the Canterbury earthquakes

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

GNS Science Consultancy Report 2013/141
August 2013

GNS
SCIENCE
TE PŪ AO

DISCLAIMER

This report has been prepared by the Institute of Geological and Nuclear Sciences Limited (GNS Science) exclusively for and under contract to the Earthquake Commission. Unless otherwise agreed in writing by GNS Science, GNS Science accepts no responsibility for any use of, or reliance on any contents of this Report by any person other than the Earthquake Commission and shall not be liable to any person other than the Earthquake Commission, on any ground, for any loss, damage or expense arising from such use or reliance.

The data presented in this Report are available to GNS Science for other use from July 2013.

BIBLIOGRAPHIC REFERENCE

Rhoades, D. A.; Gerstenberger, M. C.; Christophersen, A.; Liukis, M. 2013. Utilising short-term and medium-term forecasting models for earthquake hazard estimation in the wake of the Canterbury earthquakes, *GNS Science Consultancy Report* 2013/141. 52 p.

CONTENTS

EQUATIONS

FIGURES

TABLES

EXECUTIVE SUMMARY

Following the damaging Canterbury earthquake sequence, which commenced with the September 2010 M7.1 Darfield earthquake, a hybrid operational earthquake forecasting model is being used for decision-making on building standards and urban planning for the rebuilding of Christchurch city. The model estimates occurrence probabilities of magnitude M ≥ 5.0 for the Canterbury region for each of the next 50 years. It combines short-term, medium-term and long-term forecasting models. Short term models include the STEP (Short-Term Earthquake Probability) and ETAS (Epidemic-Type AfterShock) models, which incorporate the Utsu-Omori inverse power law for decay of aftershock activity. Medium-term models include two versions of the EEPAS (Every Earthquake a Precursor According to Scale) model with different weighting strategies. These are based on the precursory scale increase phenomenon – an increase in the rate of minor earthquake activity which typically precedes major earthquakes – and associated predictive relations for the magnitude, precursor time and precursor area. The long-term models include several different smoothed seismicity models, some designed to forecast main shocks only and others to forecast all earthquakes, including aftershocks. The weight accorded to each individual model in the operational hybrid (the “EE” hybrid model) was determined by an expert elicitation process. Another hybrid model (the “AVMAX” hybrid model) involving only one model from each class (long-term, medium-term, and short-term) was used prior to the expert elicitation process. For both hybrid models, the annual rate of earthquake occurrence in a particular spatial cell is defined as the maximum of a long-term rate and a time-varying rate. The difference is that in the EE model a weighted average of more individual models is used to compute the long-term and time-varying rates.

In this study we test the individual and hybrid models by comparing their performance over 26 years in the whole New Zealand region (the period during which the earthquake catalogue is adequate for this purpose). We also estimate optimal hybrid model combinations over the same period. Accordingly, the individual models and hybrid models have been installed in the New Zealand Earthquake Forecast Testing Centre, and used to make retrospective annual forecasts of earthquakes with magnitude M > 4.95 from 1986 on, for time-lags ranging from zero up to 25 years. The number of target earthquakes decreases as the time-lag increases, from 303 at a time-lag of zero down to 19 at a time-lag of 25 years.

We report on the performance of the individual and hybrid models for each time-lag, using two standard statistical tests adopted by the Collaboratory for the Study of Earthquake Predictability (CSEP): the N-test, which compares the observed number of earthquakes with the number forecast by the model, and the T-test, which measures the information gain per earthquake (IGPE) of one model over another.

The N-tests show that all models tend to under-predict the number of earthquakes in the test period. This is shown to be mainly due to the unusually large number of earthquakes with M > 4.95 that have occurred in the test region since the M7.8 Dusky Sound earthquake of 15 July 2009, including the Canterbury earthquakes. The T-tests show that both hybrid models are more informative than most of the individual models for all time-lags. Using data from the full 26-year test period, the IGPE relative to a stationary and spatially uniform reference model (a model of “least information”) drops off steadily as the time-lag increases, to become zero at a time-lag of about 20 years. When the unusual period since the Dusky Sound earthquake is removed from the tests, the hybrid models both show a significant positive IGPE over the model of least information at all time-lags, but do not outperform all the

individual long-term models at long time-lags. The test results are therefore seen to be sensitive to unusual features of the test catalogue, and a much longer catalogue would be needed to obtain robust results.

An optimal hybrid model with the same general form as the EE and AVMAX hybrid models is computed for each time-lag from the 26-year test period. In the optimal hybrid model, the time-varying component is dominated by the medium-term models, with hardly any contribution from the short-term models for time-lags up to 12 years. The short-term and medium-term model rates diminish with increasing time-lag, with the result that the time-varying component as a whole has hardly any impact on the optimal hybrid model for time-lags greater than 12 years.

For short time-lags up to one year, the long-term component of the optimal hybrid model is dominated by a smoothed seismicity model designed to forecast main shocks only – the National Seismic Hazard Model Background model (NSHMBG). For intermediate time-lags it is dominated by a smoothed seismicity model designed to forecast all earthquakes, computed from the locations of earthquakes with magnitude M > 4.95 since 1950 (PPE). For long time-lags greater than 17 years, the long-term rate is dominated by a similar model computed from the locations of earthquakes with M > 5.95 between 1840 and 1950 in the historical and early instrumental catalogue (PPE1950). At long time-lags the optimal hybrid model is considerably more informative than the EE Hybrid model, with an IGPE close to 1.0 for a time-lag of 25 years. This is because the PPE1950 model is the best individual model for forecasting the Canterbury earthquakes at long time-lags. When the years including the Canterbury earthquakes are removed from the test data set, the contribution of the PPE1950 model to the optimal long-term model vanishes and PPE dominates the optimal model for long, as well as intermediate, time-lags. For this reduced data set, the IGPE of the optimal hybrid model over the EE hybrid model is only moderate, in the range 0.2-0.3, for all time-lags.

A three-component hybrid model, with cell rates defined as the maximum of long-term, medium-term and short term rates, was also optimised for each time-lag from the 26-year test period. This model was found to be less informative than the two-component model for zero time-lag, but slightly more informative than the two-component at longer time-lags.

The Canterbury earthquakes are an unusual feature in New Zealand seismicity. Never before in the instrumental period has a large earthquake occurred in a region of such low seismicity and low crustal deformation rate as the Darfield earthquake. Although an optimal hybrid model for New Zealand as a whole would have a higher contribution to the time-varying component from medium-term models, the same does not necessarily apply to the Canterbury region. Tests of one-day, three-month and five-year models installed in the New Zealand earthquake forecast testing centre show that, after the Darfield earthquake, the Canterbury earthquakes are well described by short-term one-day models. The 1891 Great Nobi earthquake in Japan, which like the Darfield earthquake was located in a lower seismicity zone slightly away from a plate boundary, had 100 years of aftershocks decaying regularly according to the Omori-Utsu law. This aftershock decay, captured by the short-term models, is the most predictable component of future Canterbury seismicity. The medium-term component is less predictable, because the precursor time parameters are not well established for low seismicity regions. There is considerable variation in the estimates from the individual long-term models contributing to the EE Hybrid model. But, based on the tests carried out here, the EE hybrid, which gives appreciable weight to four different long-term models, is likely to outperform most of the individual models in the next 50 years.

NON-TECHNICAL SUMMARY

Following the Canterbury earthquake sequence, a specially developed operational earthquake forecasting model is being used for decision-making on building standards and urban planning for the rebuilding of Christchurch city. This is a hybrid model, with contributions from two short-term, two medium-term and four long-term forecasting models. It estimates the earthquake occurrence in the Canterbury region for the next 50 years. The models contributing to the hybrid model are individually very different, but are designed to capture different known features of earthquake occurrence. The short-term models describe the way in which aftershock rates typically decay following a large event. The medium-term models exploit the observation that small earthquakes often precede larger ones in the medium term. The long-term models describe the average rate at which main shocks or all earthquakes of magnitude 5 or greater are expected to occur in the long term. The weighting that each individual model received in the hybrid model was decided by eliciting the opinions of an international panel of experts.

We tested the individual and hybrid models over 26 years in the whole New Zealand region, starting in 1986 when there was a major improvement in the quality of the earthquake catalogue. We also estimated the best hybrid model that could be fitted to this part of the earthquake catalogue.

We examined how well the various models predicted the number of earthquakes in the test period, and also compared the overall information value of each model with the others. Overall, the models tended to under-predict the number of earthquakes. This was found not to be a failing of the models, but rather to be due to the unusually large number of earthquakes that occurred towards the end of the test period. These earthquakes started with a large magnitude 7.8 earthquake near Dusky Sound in 2009, which triggered many aftershocks. A year later the magnitude 7.1 Darfield earthquake initiated the Canterbury earthquakes.

The hybrid model was shown to be more informative than most of the individual models for most of the time, even when looking ahead for 25 years. However, the further ahead one looks, the less informative it becomes. It is more informative looking further ahead when the earthquakes since 2009 are left out of the tests, because the average occurrence rate since 2009 has been well above the long-term average.

The best fitting hybrid model gives greater weight to the medium-term models and less weight to the short-term models than the operational model does. But the best model for the whole of New Zealand is not necessarily the best for Canterbury. The gradual decay of aftershocks over many years is the most predictable feature of future earthquakes in Canterbury. The long-term rate at which the crust is deforming in Canterbury is lower than in places nearer to the plate boundary. The time scales for medium-term clustering are not well established for such slowly deforming regions. Therefore, the weights used in the operational model seem appropriate.

The long-term models in this study vary considerably in their estimates of future earthquakes in Canterbury. However, based on this study, the operational hybrid model, which has contributions from four very different long-term models, is likely to be more informative than most of the individual models over the next 50 years.

1.0 INTRODUCTION

The Canterbury earthquakes, which began with the Darfield M7.1 earthquake of September 2010, are unique amongst recent New Zealand earthquakes in that they constitute a major earthquake sequence occurring in a region that was previously perceived to have a rather low earthquake hazard. In the wake of these earthquakes, it was recognized that the standard probabilistic seismic hazard method would provide little relevant information about the earthquake hazard in the Canterbury region for the next several decades. This is because the method delivers very low estimates of earthquake hazard from both fault sources and distributed main shock sources, although the hazard is considered to actually be rather high due to the high probability of on-going aftershocks and the possibility of other triggered moderate-to-large earthquakes.

As a practical response to this unprecedented circumstance, a hybrid operational earthquake forecasting model was developed for decision-making on building standards and urban planning for the rebuilding of Christchurch city. This is a time-varying model of earthquake occurrence constructed from a number of short-term, medium-term and long-term forecasting models. The hybrid model was defined by combining the opinions of an international panel of experts invited to a workshop held at GNS Science in late 2011, and is accordingly known as the “Expert Elicitation” (EE) model. Prior to the workshop a simpler hybrid model was used. The latter “AVMAX” model was defined as the maximum of two components – the first component being the average of a single short-term model and a single medium-term model, and the second component being a single smoothed-seismicity model. The EE model is also the maximum of two components, but the first component is a weighted average of two different short-term and two medium-term models and second component is a weighted average of four long-term models.

The EE model has been, or is being, used for a variety of purposes, including: re-assessment of the anti-seismic provisions in the building code that will be adopted for the rebuilding of Christchurch; assessment of the potential for further liquefaction events in Christchurch in the coming decades; assessment of the viability of rebuilding in particular suburbs; assessment of future risk to life and property due to rock-falls in the Port Hills suburbs; assessment of the viability of continued occupation of certain buildings in close proximity to the Port Hills cliffs; and informing insurers of the risks now faced in providing future insurance cover for properties in Christchurch. Therefore, it is desirable to find out, insofar as is possible from the existing earthquake catalogue, how well it is likely to perform relative to the component models from which it has been constructed.

The short- and medium-term component models contributing to the EE hybrid have previously been optimised and tested with time horizons of one-day, in the case of short-term models, or a few months to a few years, in the case of medium-term models (Gerstenberger and Rhoades, 2010). In the hybrid model, they are being used with much longer time horizons ranging from one to fifty years. Their likely performance with long time horizons is hitherto unknown, as is the likely performance of the EE and AVMAX hybrid models. The aim of this project is to assess the information value of the component models and the hybrid models with time horizons up to 25 years. This is the longest time horizon which can be considered in retrospective testing, given the quality of the New Zealand earthquake catalogue.

The AVMAX model was the one in use at the time when this project was proposed. The objectives have been re-ordered and adjusted to accommodate the fact that the EE model has now superseded the AVMAX model, and that all of the presently viable component models have been considered in the expert elicitation process. However, the major objective of this project remains to assess how well the individual component and hybrid models can be expected to perform with a long time horizon in view, and to consider what model weightings would provide the most informative hybrid model, based on the existing earthquake catalogue.

2.0 INDIVIDUAL MODELS

Here, we briefly describe each of the individual models considered in this study, including the component models of the EE hybrid earthquake occurrence model. The models fall into three classes: short-term, medium-term and long-term.

2.1 SHORT-TERM MODELS

2.1.1 STEP

The Short-term Earthquake Probabilities (STEP) model was defined by Gerstenberger et al. (2005; 2007). This is an aftershock clustering model based on the idea of superimposed Omori-type sequences (Ogata, 1988,1998). It estimates the future decay of activity in existing earthquake clusters using the inverse power law known as the Omori-Utsu law (Utsu et al., 1995). The model normally comprises two components: a background model and a time-dependent clustering model. For the testing of the model here, we consider both the clustering component by itself (STEP_TV) and the whole STEP model. The background model contributes to the long-term component of the EE hybrid model, and the clustering model to the time-varying component. 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 Omori-Utsu law (Ogata, 1983).

The clustering model 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 (1989) 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. When the model is applied with a long time-lag, there is little opportunity for the latter two approaches to contribute much to the forecast.

2.1.2 ETAS

The Epidemic-type aftershock (ETAS) model (Ogata, 1989; 1998) is another aftershock clustering model, which now has a wide acceptance amongst researchers. Like the STEP model, the ETAS model has two components: a background model and a time-dependent clustering model. Again, the background model (PPE) contributes to the long-term component of the EE hybrid model, and the clustering model (ETAS_TV) to the time-varying component. In the clustering model, each earthquake is regarded as having its own aftershock sequence decaying in time according to the Omori-Utsu law. Several different versions of the ETAS model have been proposed by different researchers. The version considered here is the one that is installed in the New Zealand Earthquake Forecast Testing Centre (Gerstenberger and Rhoades, 2010). The version is different in several details from some other published versions. In particular, a fixed set of parameters controls the decay of each sequence, with no spatial or temporal variation of these parameters. Also the spatial distribution of aftershocks is according to a bivariate normal distribution with circular symmetry, and the productivity parameter α is set equal to the Gutenberg-Richter slope parameter *b* so that the expected number of aftershocks of a parent earthquake at a given magnitude *m* depends only on the difference between *m* and the magnitude of the parent earthquake.

For a number of reasons, including latencies in the catalogue and model installation issues, no reliable formal comparison of the ETAS and STEP models has yet been possible in New Zealand. However, in the California CSEP testing centre, preliminary tests show that the STEP model is superior to the ETAS model in one-day-ahead forecasting, possibly because it is a more elaborate model that includes internal adaptation of parameters to the characteristics of each aftershock cluster. Whether or not this superiority would carry over to longer forecast time-horizons remains to be demonstrated.

2.2 MEDIUM-TERM MODELS

The Every Earthquake a Precursor According to Scale (EEPAS) model of Rhoades and Evison (2004, 2005, 2006) is a medium-term clustering model based on the precursory scale increase (Ψ) phenomenon – an increase in the magnitude and rate of occurrence of minor earthquakes that precedes most major earthquakes in the long run (Evison and Rhoades, 2002; 2004). It adopts the notion that the precursory scale increase occurs at all scales in the seismogeneic process. Each earthquake is considered to generate a transient medium-term increase in the future earthquake rate with the duration indicated by the Ψ predictive relation for precursor time, at magnitudes indicated by the Ψ predictive relation for mainshock magnitude, and covering an area indicated by the Ψ relation for precursor time. All of these relations depend on the magnitude of the generating earthquake. The EEPAS model normally includes a background model (PPE). Again, the background model contributes to the long-term component of the EE hybrid model, and the clustering model to the time-varying component. Two versions of the EEPAS model, with different weighting strategies, are considered here.

2.2.1 EEPAS_0F

EEPAS_0F is a version of the EEPAS model which adopts the equal-weighting strategy (w_i = 1, for all i). In this acronym, the “0” implies equal weighting of precursory earthquakes and the “F” indicates that a rather full set of parameters has been optimized. This is the version of EEPAS that gives the highest likelihood in fitting to the past New Zealand catalogue with m_c = 4.95. With the same value of m_c, it is also the most informative version of EEPAS after several years of prospective testing in the California CSEP testing test region. The clustering component of EEPAS_0F is denoted EEPAS_0F_TV.

2.2.2 EEPAS_1F

EEPAS_1F is a version of the EEPAS model which adopts an unequal-weighting strategy in which aftershocks are down-weighted. For details, see Rhoades and Evison (2004). In this acronym, the “1” implies unequal weighting of precursory earthquakes and the “F” again indicates that a rather full set of parameters has been optimized. The unequal-weighting strategy not only works best with higher threshold magnitudes, such as m_c = 5.75 in New Zealand and m_c = 6.75 in Japan (Rhoades and Evison 2004; 2005), but also appears to work better than the equal-weighting strategy when EEPAS and ETAS are combined in a simple additive hybrid at time-horizons up to 3000 days in both New Zealand and California (Rhoades, 2013). The clustering component of EEPAS_1F is denoted EEPAS_1F_TV.

EEPAS_0F and EEPAS_1F often give similar forecasts. However, in the Canterbury region following the 2011 Darfield earthquake, these two models give very different forecasts, with the EEPAS_0F forecasts being much higher because it treats aftershocks as precursors of larger earthquakes to follow in the medium term.

2.3 LONG-TERM MODELS

The long-term models considered are all "smoothed seismicity" models, which depend on the spatial distribution of earthquakes in the previous catalogue. They vary with time only because new earthquakes are continually being added to the catalogue as time passes.

2.3.1 NSHMBG_B_POLY

The NSHMBG_BVAL_POLY model uses the method of the "distributed source" component of the 2010 update of the New Zealand national seismic hazard model (Stirling et al., 2012). It applies a Gaussian kernel with a 50 km standard deviation to smooth the *a*-value of the Gutenberg-Richter frequency magnitude relation across the 0.1 degree x 0.1 degree cells of the model. The *a*-value is estimated using the catalogue from 1840 on. The Gutenberg-Richter *b*-value is calculated for each polygonal region defined to group together zones of similar seismotectonic type and the values are smoothed across polygon boundaries. The model uses a declustered New Zealand earthquake catalogue from 1964 on with magnitudes M ≥ 4.0 to estimate the *b*-value. There are some differences in the depth and magnitude ranges from the NSHM. Specifically, a depth range of 0 – 40 km is applied, with the deeper earthquakes in subduction regions being ignored, and the upper magnitude cutoff of M 7.2 from the NSHM is relaxed, so that forecasts extend up to the same upper magnitude limit as for other models considered. This is the background model used for the STEP model installed in the New Zealand CSEP test region.

2.3.2 NSHMBG_B_1

The NSHM_BVAL_1 model is similar to NSHMBG_BVAL_POLY except for one aspect: the Gutenberg-Richter *b*-value is taken as 1 everywhere, instead of being estimated separately in each seismotectonic zone. This model is used only as a reference model here, since it is not included in either the AVMAX or EE hybrid models.

2.3.3 PPE

The "Proximity to Past Earthquakes" (PPE) model is a smoothed seismicity model which uses the earthquakes with magnitude M > 4.95 at depths up to 40 km from 1951 on to estimate the future rate of the same class of earthquakes. No declustering is employed. It uses an inverse power law smoothing kernel, based on the model described by Jackson and Kagan (1999). For details see Rhoades and Evison (2004). The PPE model is the background model normally used for the EEPAS and ETAS models in the New Zealand CSEP test region.

2.3.4 PPE_FROM_1840

The PPE_FROM_1840 model is a convex linear combination of two PPE models – the first constructed from earthquakes from 1840-1950 with M > 5.95 (PPE1950), and the second constructed from the earthquakes from 1951 to the date of the forecast with M > 4.95 (standard PPE). In PPE1950, the earthquake occurrence rate density is extrapolated downwards from M5.95 to magnitudes in the range 4.95 – 5.95 using a fixed Gutenberg-

Richter *b*-value. The two models are weighted in proportion to the length of the catalogues involved. For example, for a forecast using data up to 3 September 1986 the weight accorded to PPE1950 is 101/(101+35.7) and the weight accorded to the standard PPE model is 35.7/(101+35.7).

2.3.5 PPE_DECLUS

The PPE_DECLUS model is version of the PPE model which down-weights the contributions of aftershocks in a similar way that EEPAS_1F does. Like the NSHMBG models it aims to forecast the future rate of mainshocks only, although the declustering algorithm is different from that used in the NSHMBG models. This model is included in the retrospective testing as a substitute for another model (PPE_PRE_DARFIELD), which could not be retrospectively tested, because it is not well-defined prior to the Darfield earthquake.

The PPE_PRE_DARFIELD model was defined as the PPE model forecast up to the time just prior to the initiation of the Canterbury sequence. The rationale for the inclusion of this model in the EE hybrid model is that the Canterbury earthquake sequence is viewed as a rarely occurring aberration from normal seismicity and should therefore be excluded from the long-term estimate of the earthquake occurrence rates. According to this rationale, the earthquakes prior to the Darfield earthquake give a better estimate of the long-term earthquake occurrence rate as a function of location; once the aftershocks of the Darfield earthquake have concluded, the background rate would be expected to return to a background rate similar to this previously estimated long-term rate. Allowing the large number of aftershocks of the Darfield earthquake with M > 4.95 to contribute to the long-term estimates of future earthquake occurrence (as the standard PPE model does) would be seen as distorting the long-term earthquake rate estimates for many years to come.

The PPE_DECLUS model is substituted for PPE_PRE_DARFIELD because it down-weights the contribution of aftershocks (not only of the Darfield earthquake but of all main shocks), consistent with the rationale behind the PPE_PRE_DARFIELD model. Unlike the latter model, it can be retrospectively applied and tested.

2.3.6 SUP

The spatially uniform Poisson (SUP) model is included in the comparisons as a reference model of least information, although it plays no part in the hybrid models. It recognises only that earthquakes follow the Gutenberg-Richter law with a fixed *b*-value. The earthquake rate is the same at all locations, and the total rate within the testing region is estimated from the total number of earthquakes with M > 4.95 from the beginning of the year 1951 up to the time at which a forecast is made. Although this model generally has a low information value, it can outperform more detailed models at times when innovations occur in the locations of earthquakes in the catalogue. The time of occurrence of the Darfield main shock was such a time.

3.0 HYBRID MODELS

Two hybrid models are considered in the analyses that follow. These are the AVMAX hybrid and the EE Hybrid. Both hybrid models are defined as the maximum of two components – a long-term component and a time-varying component.

3.1 AVMAX HYBRID

In the AVMAX hybrid the long-term component is the PPE model. This choice was based on the knowledge that the PPE model was the best performing 5-year model in retrospective tests in the New Zealand CSEP testing region (Rhoades et al., 2010). The time-varying component is the average of the STEP_TV and EEPAS_0F_TV models. This choice was based on the knowledge that a nearly equal weighting of these two models was found to be optimal for one-day-ahead forecasting in the California test region.

3.2 EE HYBRID

The EE hybrid component models were determined by combining the opinions of an international panel at an expert elicitation workshop held at GNS Science in late 2011. The long-term component is a weighted average of four long-term models discussed in the previous section, and the time-varying component is a weighted average of the time-varying components of the two short-term and two medium-term models discussed in the previous section. The weights accorded to each model are given in Table 1.

Table 1 Model weightings in the AVMAX and EE hybrid models.

Model	Component	AVMAX weight	EE weight
STEP_TV	Time-varying	0.5	0.36
ETAS_TV	Time-varying		0.19
EEPAS_0F_TV	Time-varying	0.5	0.24
EEPAS_1F_TV	Time-varying		0.21
NSHMBG_B_POLY	Long-term		0.58
PPE	Long-term	1.0	0.13
PPE_FROM_1840	Long-term		0.16
PPE_DECLUS	Long-term		0.12

This page is intentionally left blank.

4.0 DESIGN OF RETROSPECTIVE FORECASTING EXPERIMENT

The models discussed in the previous two sections were installed in the CSEP New Zealand Earthquake Forecast Testing Centre, and used to make one-year forecasts with time-lags ranging from zero up to 25 years. The testing centre uses data within a search polygon and tests the models on their ability to predict earthquakes of M > 4.95 in a test region (Figure 1).

The experiment was designed taking into account the changing quality of the New Zealand earthquake catalogue over time (Gerstenberger and Rhoades, 2010), and its suitability for fitting to all of the models in different time periods. From 1964 to about 1986, the catalogue within the CSEP testing region was complete down to about magnitude 4 or a little lower. After that time it has been complete down to about magnitude 3 (the smallest earthquakes used by the models considered here).

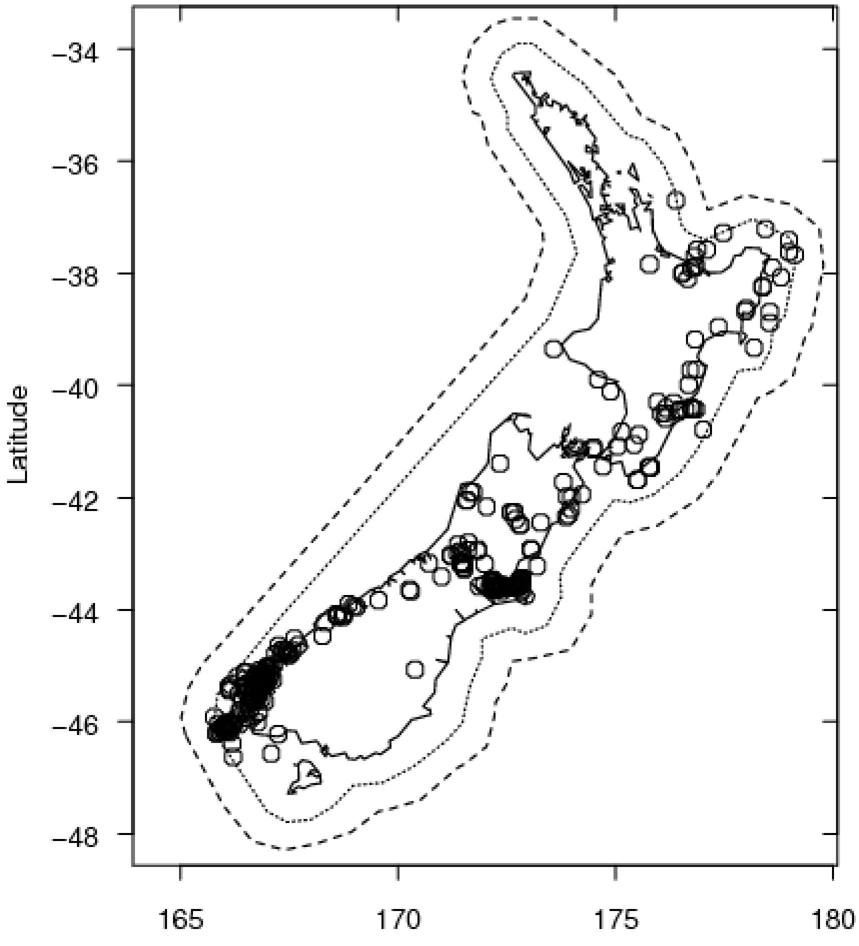

Figure 1 Map showing the test region (dotted polygon), the search region (dashed polygon) and earthquakes with M > 4.95 in the period 4 September 1986 to 3 September 2012 – the target earthquakes for the retrospective tests.

Our experiment therefore begins in 1986 and relies on approximate completeness of the catalogue for magnitude 6 and above from 1840 (PPE_FROM_1840), for magnitude 5 and above from 1951 (PPE models), from magnitude 4 and above from 1964 (NSHMBG models), and for magnitude 3 and above from 1986 on (STEP, ETAS and EEPAS models). To optimise aftershock forecasts of the Darfield earthquake, the annual periods begin on 4

September and end on 3 September. Absolute completeness above these thresholds cannot be guaranteed, of course, but it is not required for a statistical experiment of this sort.

The experimental design is illustrated in Figure 2. In each year a set of one year forecasts is created using each model with lags of zero, 1 year, 2 years, etc., with the final forecast being for the year beginning 4 September 2011. The set of forecasts with the same time-lag are then analysed separately.

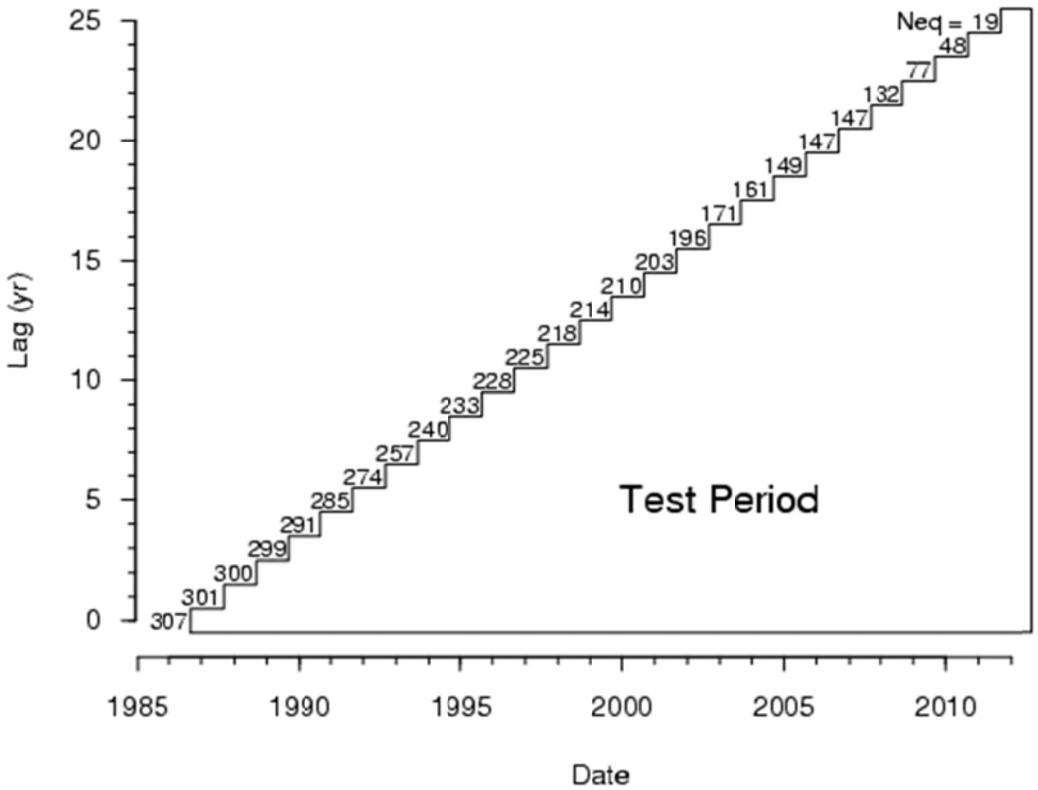

Figure 2 Experimental design for retrospective tests, showing test period for lagged forecasts and number of target earthquakes in the polygonal test region for each lagging class.

The targeted earthquakes are those in the gridded test region with magnitude M > 4.95. The number of earthquakes in the test period depends on the time-lag. For example, the zero-lag forecasts date from 3 September 1986 and begin on 4 September 1986 and are updated annually until 4 September 2011. The total period covered by these forecasts is from 4 September 1986 to 3 September 2012 – a period of 26 years, which includes 307 target earthquakes in the test region polygon (the number shown in Figure 2). The test region grid has rectangular edges and so the number of earthquakes inside the grid can differ by a small amount from the number inside the polygon. In this case the number of earthquakes in the test region grid is only 303. At the other extreme, the forecasts with a time-lag of 25 years also date from 3 September 1986 but – with a time-lag of 25 years – do not begin until 4 September 2011. The test set in this case consists of a single year of data from 4 September 2011 to 3 September 2012, which includes only 19 target earthquakes. The number of target earthquakes for each lag is shown in Figure 2. The number of target earthquakes is in excess of 100 except for time-lags greater than 22 years.

The annual number of target earthquakes varies markedly throughout the test period (Figure 3). It ranges from 0 (in the year beginning 4 September 2006) to 55 (in the year beginning 4 September 2008, which included the M7.8 Dusky Sound earthquake of 15 July

2009 and its early aftershocks), with most of the observations falling in the range 0 – 10 events and more than half of the earthquakes occurring in only five of the 26 years. Figure 3 illustrates that even when looking at the whole New Zealand region, earthquake statistics are dominated by time-varying features, such as those that the short-term and medium-term clustering models are designed to describe. Stationarity of the earthquake process, which the long-term models tend to rely on, is not evident on this scale.

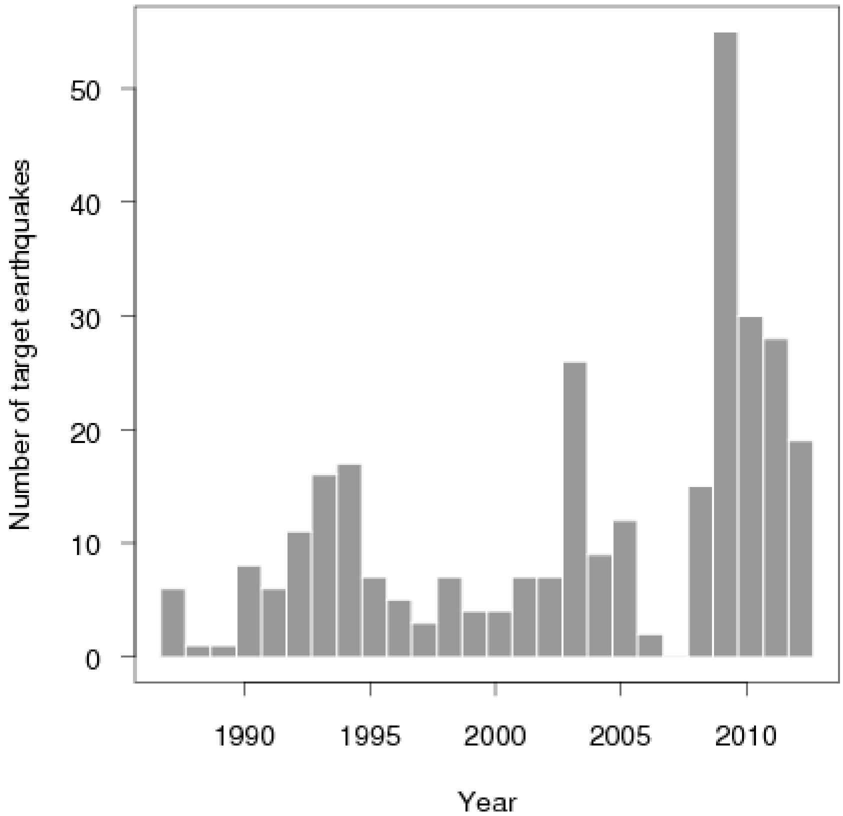

Figure 3 Histogram of the number of target earthquakes (M > 4.95) in individual years from 4 September 1986 to 3 September 2012.

This page is intentionally left blank.

5.0 RESULTS

5.1 STATISTICAL TESTS OF MODEL PERFORMANCE

We evaluate the performance of the models using two standard CSEP tests: the N-test and the T-test.

The N-test (Schorlemmer et al., 2007; Zechar et al., 2010a) measures the consistency of the number of earthquakes expected by a model with the number observed. The test assumes a Poisson distribution of errors, which is unachievable for models which cannot anticipate the episodes of intense clustering, including both time-invariant models and short-term aftershock models that are not continuously updated. For this reason, the Poisson assumption has been criticised by a number of authors (Schorlemmer et al., 2010; Werner et al., 2010), but there is presently no consensus for an alternative.

The T-test (Rhoades et al., 2011) is designed to compare the log likelihood of two models over the test period. It uses the classical paired Student's t-test to provide a confidence interval on the information gain per earthquake (IGPE) of one model *A* over another model *B* (Imoto, 2007), where

$$IGPE = \frac{\ln L_A(x) - \ln L_B(x)}{N_{eq}}$$ **Equation 1**

where ln $L_A(x)$ is the log-likelihood of the target earthquakes *x* under model *A*, and N_{eq} is the number of target earthquakes.

5.2 NUMBER OF EARTHQUAKES (N-TESTS)

The number of earthquakes predicted by the AVMAX and EE hybrid models at each time-lag is shown in Figure 4, and compared to the actual number using the T-test. It can be seen that both models under-predict the number of earthquakes at each time-lag, with the only exception being for the AVMAX model at zero time-lag. It is important to investigate whether the under-prediction is a generic feature of the hybrid models or of some peculiar feature of the test data.

The hybrid models have a long-term component and a short-term component. In the case of the AVMAX model, the long-term component consists of the single model PPE. The PPE model is intended to forecast all earthquakes, not just the main shocks. The N-tests for the PPE model are shown in the upper panel of Figure 5. It can be seen that it under-predicts the number of earthquakes at each time-lag to a greater extent than the AVMAX hybrid model and to a similar extent as the EE hybrid model. Another long-term model intended to predict all earthquakes in the target set, including aftershocks, is the SUP model. The N-tests for the SUP model are shown in the lower panel of Figure 5. It under-predicts the number of earthquakes to an even greater extent than the PPE model. A fortiori, it under predicts to a greater extent than the AVMAX and EE hybrid models do.

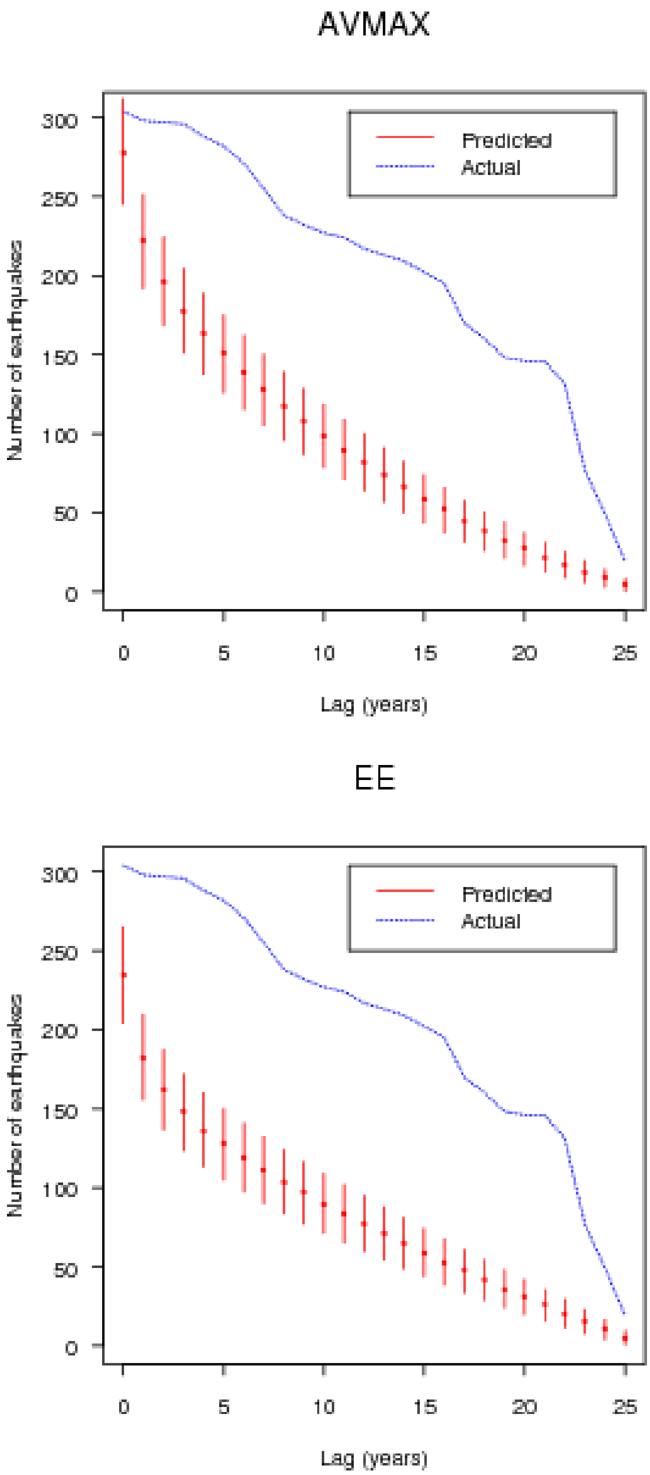

Figure 4 Number of earthquakes predicted and observed at each time-lag by AVMAX hybrid model (upper panel) and EE hybrid model (lower panel). Error bars are Poisson 95% confidence limits.

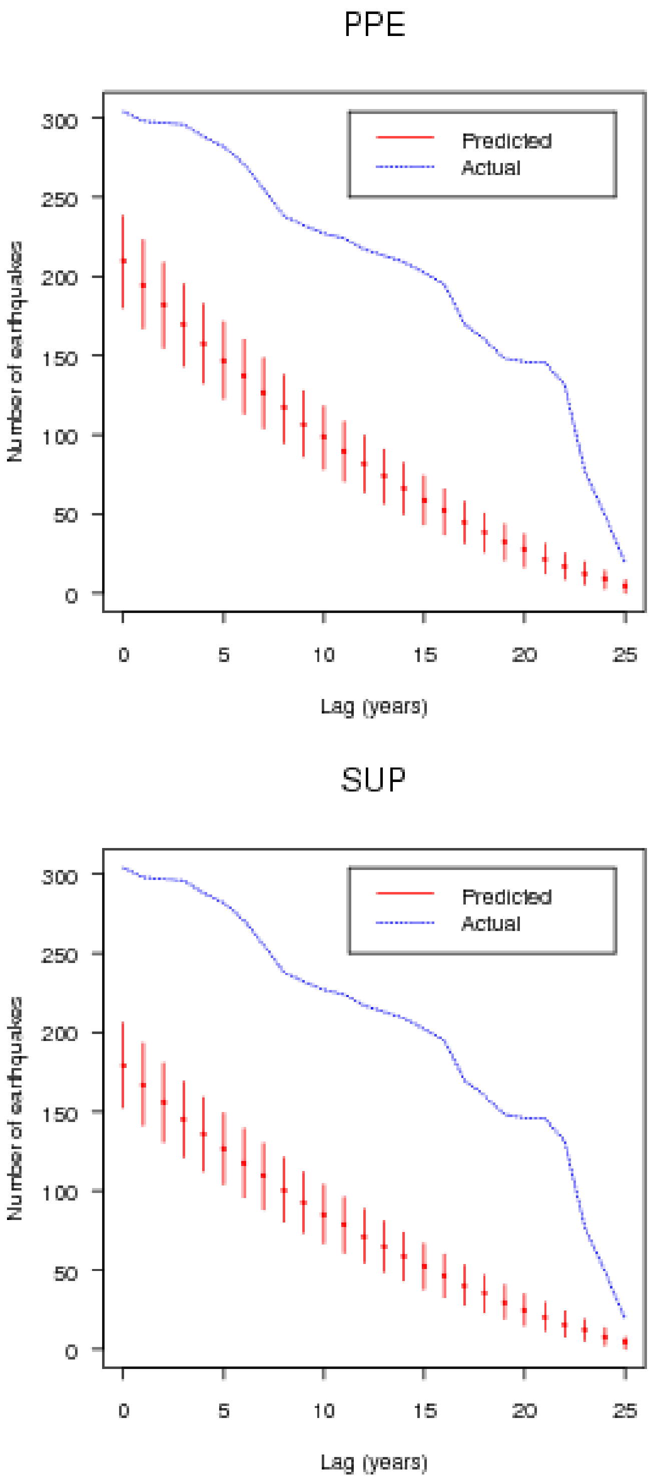

Figure 5 Number of earthquakes predicted and observed at each time-lag by PPE model (upper panel) and SUP model (lower panel). Error bars are Poisson 95% confidence limits.

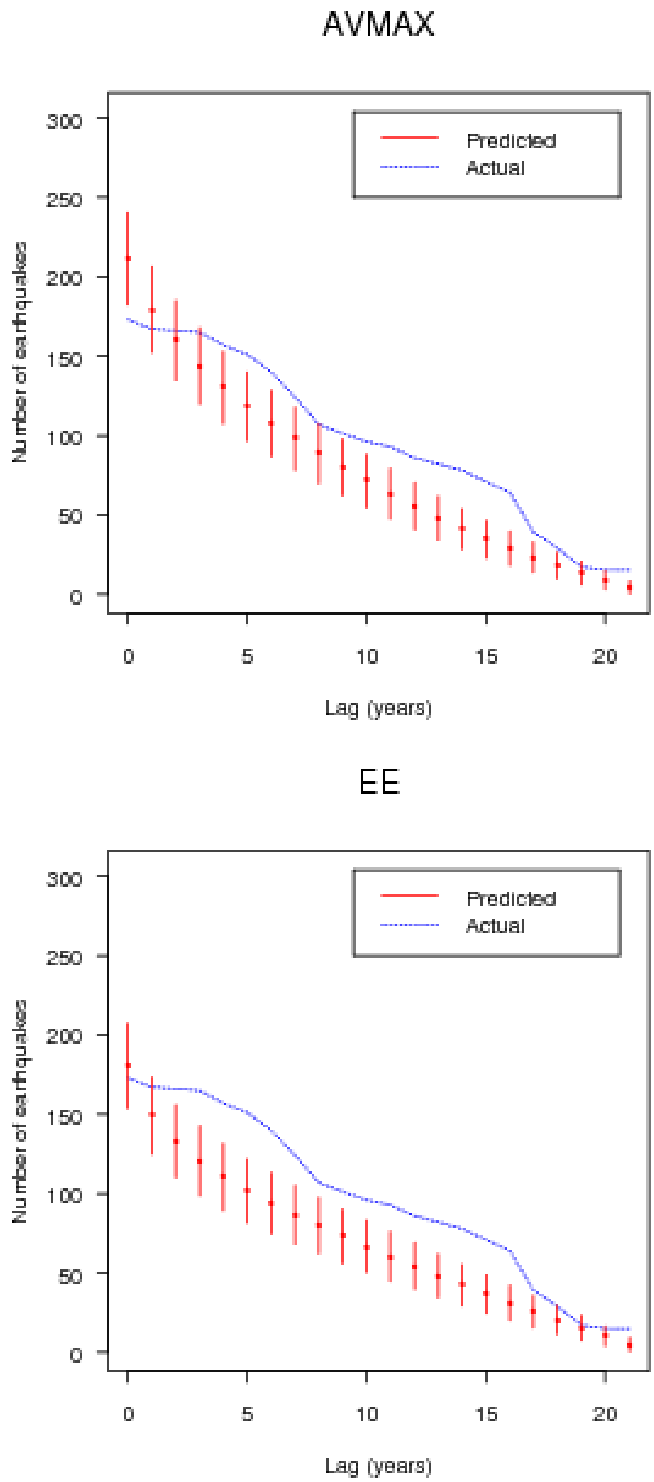

Figure 6 Number of earthquakes predicted and observed at each time-lag by AVMAX hybrid model (upper panel) and EE hybrid model (lower panel), excluding data from 4 September 2008 to 3 September 2012. Error bars are Poisson 95% confidence limits.

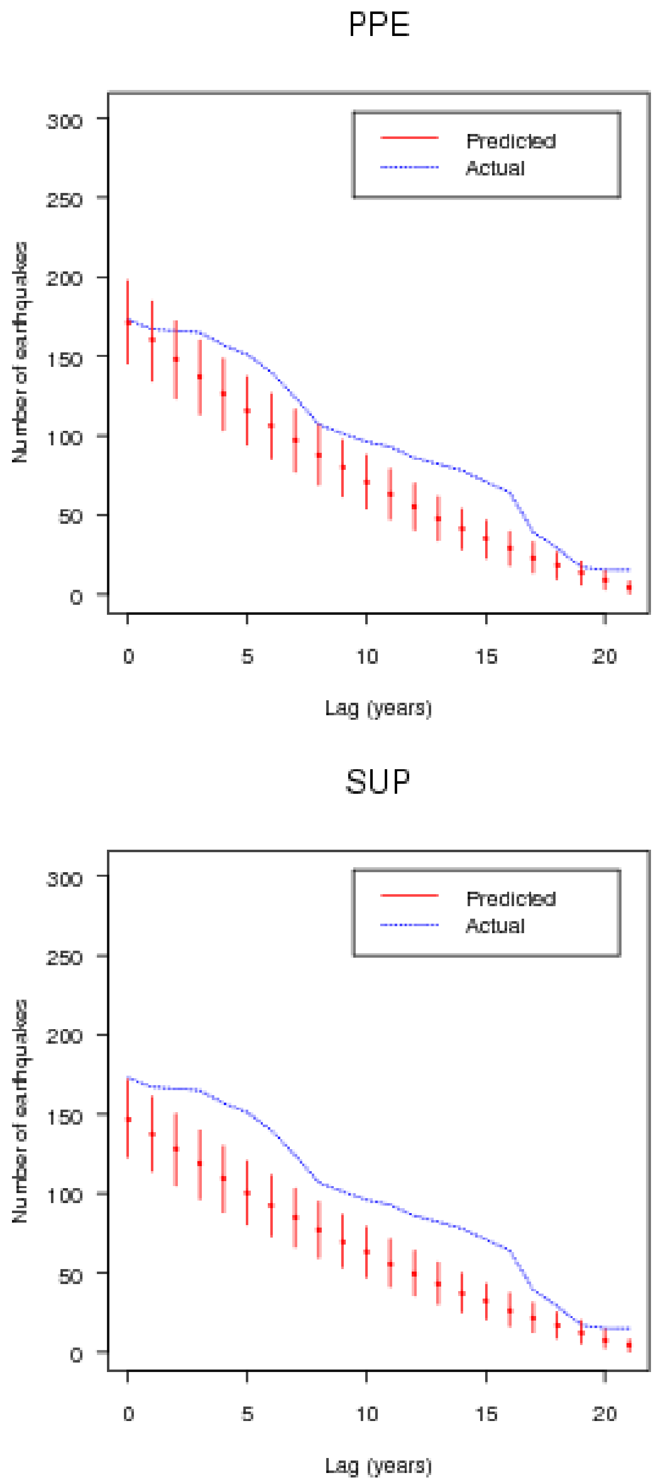

Figure 7 Number of earthquakes predicted and observed at each time-lag by PPE model (upper panel) and SUP model (lower panel), excluding data from 4 September 2008 to 3 September 2012. Error bars are Poisson 95% confidence limits.

The number of target earthquakes in the four years from 4 September 2008 to 3 September 2012 was markedly above average, as Figure 3 shows. Therefore, it is interesting to consider whether the consistent under-prediction seen in Figure 4 and Figure 5 could be attributable to the large number of target earthquakes in these four years. Accordingly, we have performed N-tests for the same models excluding these years. With this reduced data set, the maximum time-lag is only 21 years.

The results are shown in Figure 6 for the AVMAX and EE hybrid models, and in Figure 7 for the PPE and SUP models. It can be seen that there is still a tendency for all models to under-predict the number of target earthquakes. However, this tendency is much reduced, and is eliminated for time-lags less than 2 years. The AVMAX hybrid model actually over-predicts at a zero time-lag. This is not unexpected, because the long-term component (the PPE model) is already normalised to predict the average rate of all earthquakes with M >4.95, including both main shocks and aftershocks. The time-varying component (made up of STEP and EEPAS_0F) contributes much more to the earthquake rate at short time-lags than at long time-lags. At long time-lags, the long-term component is expected to dominate the short-term component almost everywhere, but at short time-lags the short-term component will be the dominant one in some locations with active earthquake clusters. Because of these features, the AVMAX model is expected to over-predict the number of earthquakes at short time-lags and to neither under- nor over- predict at long time-lags. For the test data set, it tends to under-predict at long time-lags, and this may be attributed to the under-prediction of the PPE model for the same data set.

The long-term component of the EE hybrid model is a mixture of models, some of which are normalised to predict both main shocks and aftershocks, and others to predict main shocks only. Therefore, the long-term component of the EE hybrid model is expected to under-predict the number of earthquakes. The time-varying component has some capability to predict short- and medium- term clustering, but this capability diminishes with increasing time-lag. Therefore, the EE hybrid model as a whole is expected to neither over- nor under-predict the number of earthquakes at short time-lags, but to under-predict at long time-lags. This is broadly the pattern seen in Figure 6. However, it appears that the under-prediction at long time-lags is exacerbated in this case by the nature of the experimental data, because it is also seen for the PPE and SUP models, which are normalised to predict both main shocks and aftershocks.

The under-prediction of the PPE and SUP models, even with the reduced data set, is consistent with the observation that the average rate of occurrence of target events in the test period from 1986 to 2008 was 7.9 per year compared to the lower rate of 6.2 per year during the prior period of 1951 to 1986 used to warm up these models.

In summary of the N-tests, there is a general tendency for the long-term models and the hybrid models to under-predict the number of earthquakes during the test period for a range of time-lags. However, this under-prediction is consistent with the temporal distribution of earthquakes during the test period and the time period used for model warm-up. Given this distribution, the models performed as expected in the retrospective tests. The results of the N-tests do not raise any new concerns about the AVMAX or EE hybrid models. Nevertheless, the under-prediction illustrates the difficulties in forecasting even the long-term rate of earthquake occurrence over several decades in a wide area, with temporal and spatial clustering being such a strong feature of the earthquake process.

5.3 INFORMATION GAIN PER EARTHQUAKE (T-TESTS)

A large number of pairwise comparisons of individual models at a variety of time-lags have been performed by the testing centre. It is impractical to present all of them here, and not all of them are of equal interest. Here we primarily examine the performance of the EE hybrid model relative to other models at each time-lag, and the change in performance as a function of the time-lag relative to selected reference models. As for the N-tests discussed above, the performance needs to be interpreted with regard to the particular characteristics of the seismicity during the experimental period of the retrospective experiment.

Figure 8 shows the information gain of the EE hybrid model over other models for zero time-lag. This is an indication of how the model performs relative to the other models when forecasting the earthquakes expected in the next year, using the past earthquake catalogue up to the beginning of the year. In this plot, a confidence interval entirely to the right of the zero line indicates an information gain significantly greater than zero with 95% confidence. The EE hybrid model is thus seen to be more informative than all the individual models it is compared to, and significantly so for all models except EEPAS_0F and EEPAS_0F_TV. However, it is less informative than the AVMAX hybrid. The relatively large information gains of the EE hybrid model over the short-term model clustering components ETAS_TV, STEP_TV, and both forms of the NSHMBG model are worthy of note. These large gains are easy to understand, because ETAS_TV and STEP_TV set out only to forecast aftershocks and the NSHMBG models only to forecast main shocks. The EE hybrid model is designed to forecast both of these, as well as events displaying the precursory scale increase phenomenon.

Figure 9 shows a comparable plot for a time-lag of 1 year. The results are generally similar to Figure 8, but with some notable differences. First, the information gain of the EE hybrid model over ETAS_TV and STEP_TV is increased, and the gain over the NSHM_BG models is reduced. Both of these effects are consistent with the expectation that aftershocks tail off in an inverse power law from the time of the main shock and, therefore, aftershock model forecasts should become less informative as the time-lag is increased. The large gain of the EE model over the STEP model (which includes the NSHMBG_B_POLY model as background) is more difficult to understand, and may indicate a problem with the installation of this model. If installed correctly, this model should perform similarly to the ETAS model, which also comprises short-term clustering and background components.

Figure 10 to Figure 15 show comparable plots for time-lags from 2 years up to 25 years. For a 2 year lag Figure 10(a), we note that the EE hybrid model is no longer more informative than all of the individual models. It is slightly less informative than both the PPE and the EEPAS_1F models, although not significantly so. It must be borne in mind, however, that the EE model was intentionally given a large aftershock component. The aftershock sequence in Canterbury is expected to be a long one, because the main shock occurred away from the active plate boundary, in a region of low prior seismic activity and low strain rate. The general seismicity in the wider New Zealand region is much more heterogeneous, includes a variety of tectonic processes such as subduction and continental collision, but is mostly associated with physical processes close to the boundary between the Australian and Pacific Plates.

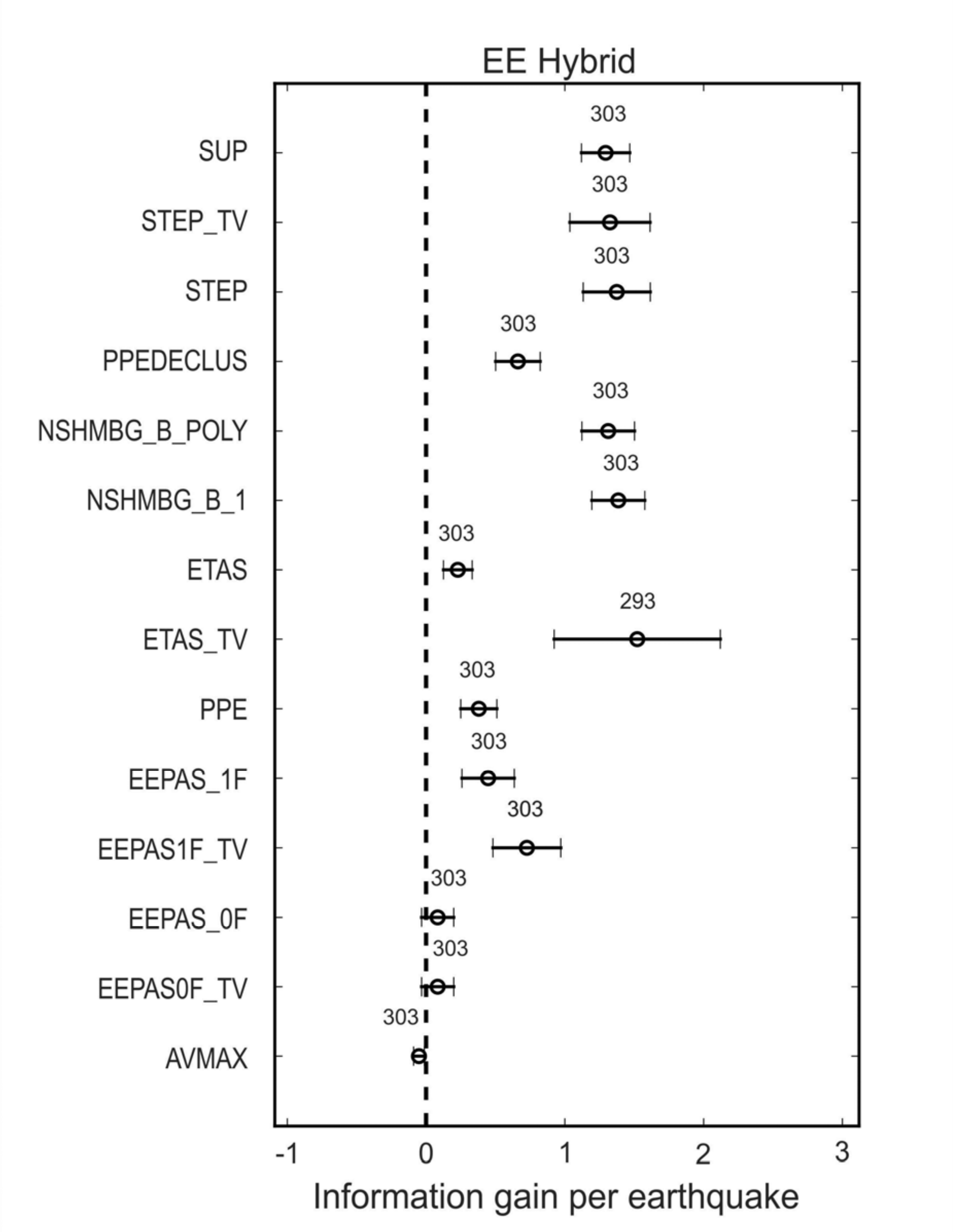

Figure 8 Information gain per earthquake of the EE hybrid models over other models in one year forecasts with zero time-lag in the New Zealand CSEP testing region from 1986 to 2012. Error bars are 95% confidence intervals according to the T-test. The number of target earthquakes is given above the error bars. Target earthquakes for which a model gave a zero rate are excluded from the paired comparisons.

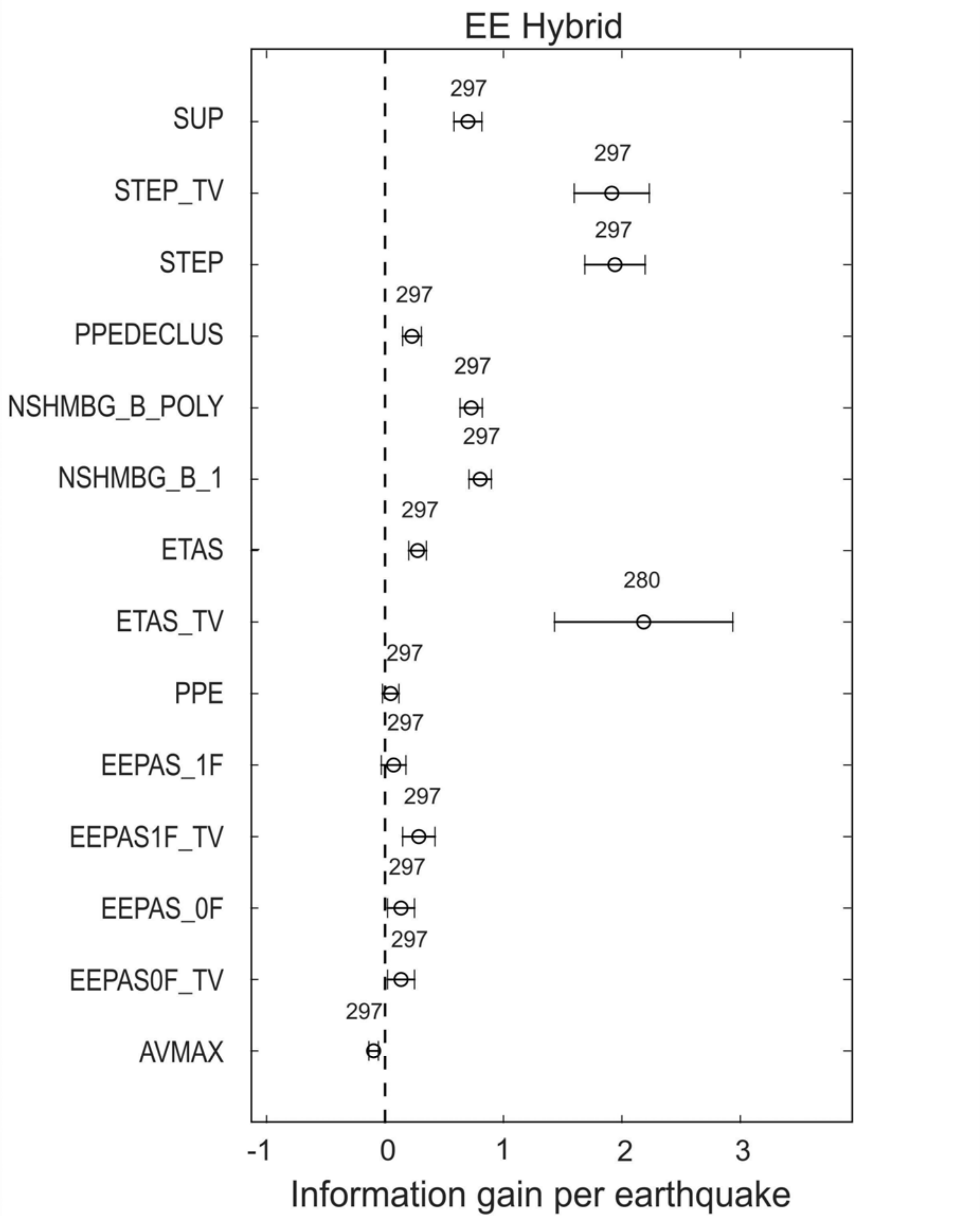

Figure 9 Information gain per earthquake of the EE hybrid models over other models in one year forecasts with 1-year time-lag in the New Zealand CSEP testing region from 1986 to 2012. Error bars are 95% confidence intervals according to the T-test. The number of target earthquakes is given above the error bars. Target earthquakes for which a model gave a zero rate are excluded from the paired comparisons.

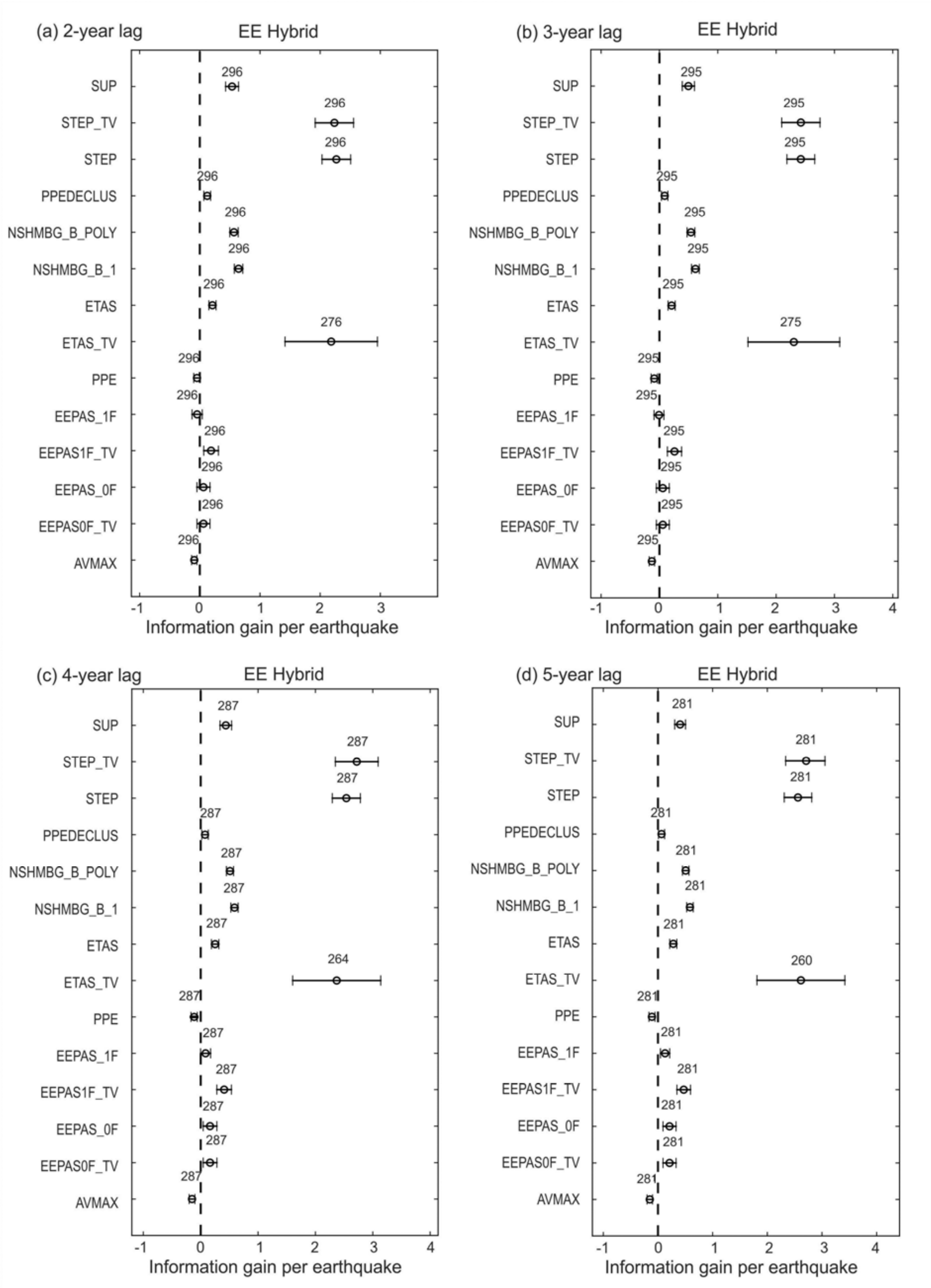

Figure 10 Information gain per earthquake of the EE hybrid models over other models in one year forecasts with (**a**) 2-year (**b**) 3-year (**c**) 4-year (**d**) 5-year time-lag in the New Zealand CSEP testing region from 1986 to 2012. Error bars are 95% confidence intervals according to the T-test. The number of target earthquakes is given above the error bars. Target earthquakes for which a model gave a zero rate are excluded from the paired comparisons.

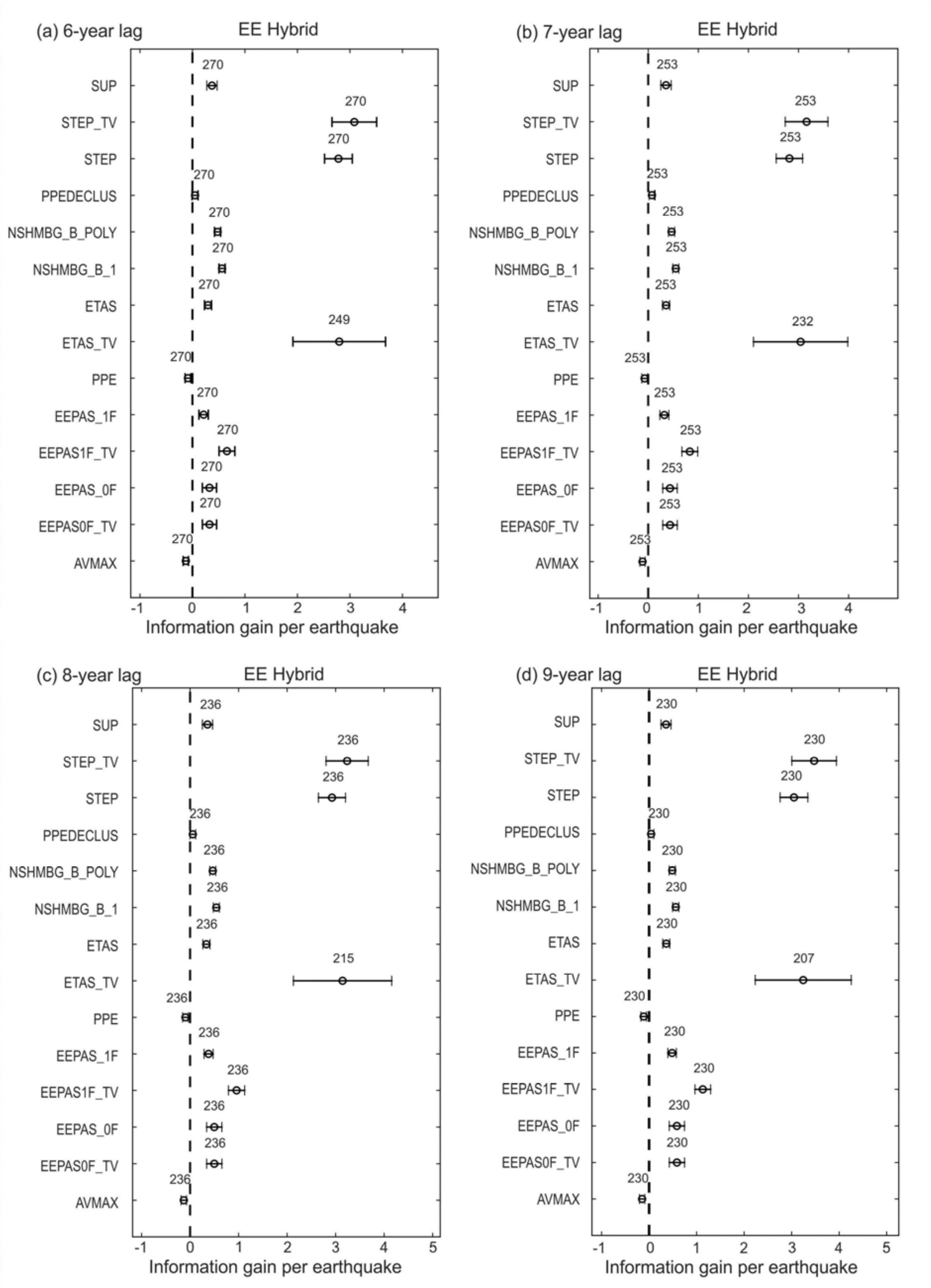

Figure 11 Information gain per earthquake of the EE hybrid models over other models in one year forecasts with (**a**) 6-year (**b**) 7-year (**c**) 8-year (**d**) 9-year time-lag in the New Zealand CSEP testing region from 1986 to 2012. Error bars are 95% confidence intervals according to the T-test. The number of target earthquakes is given above the error bars. Target earthquakes for which a model gave a zero rate are excluded from the paired comparisons.

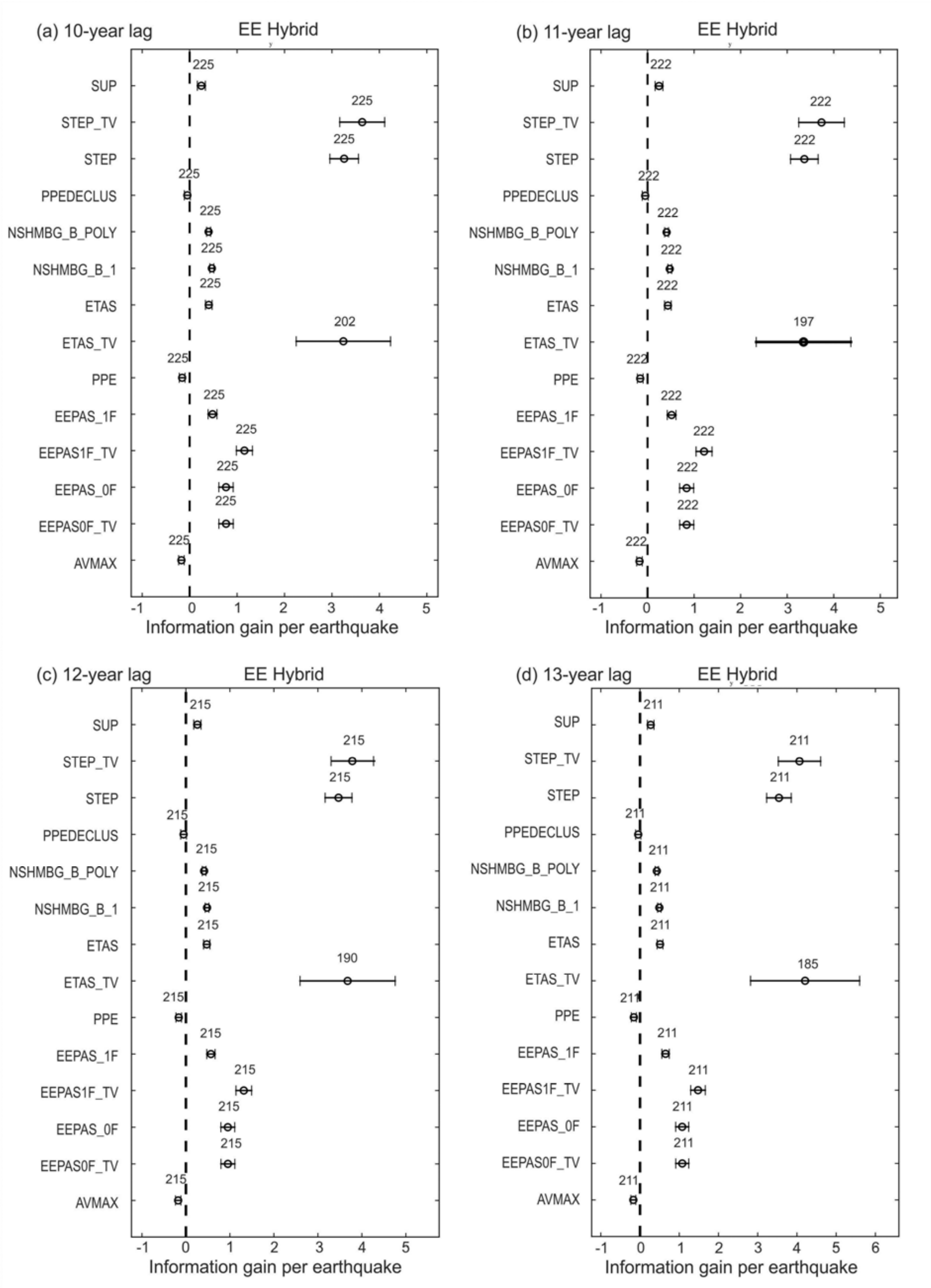

Figure 12 Information gain per earthquake of the EE hybrid models over other models in one year forecasts with (**a**) 10-year (**b**) 11-year (**c**) 12-year (**d**) 13-year time-lag in the New Zealand CSEP testing region from 1986 to 2012. Error bars are 95% confidence intervals according to the T-test. The number of target earthquakes is given above the error bars. Target earthquakes for which a model gave a zero rate are excluded from the paired comparisons.

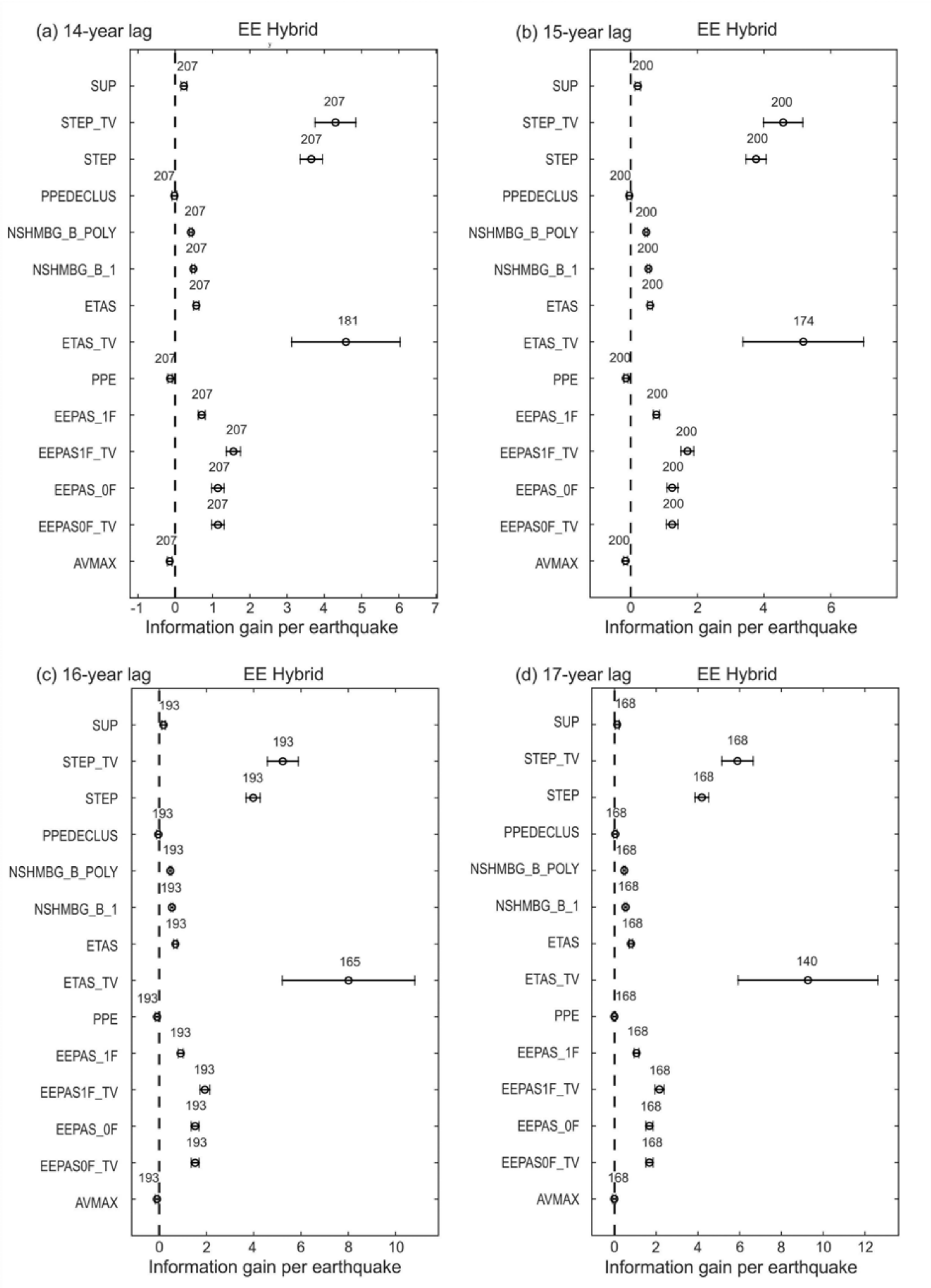

Figure 13 Information gain per earthquake of the EE hybrid models over other models in one year forecasts with (**a**) 14-year (**b**) 15-year (**c**) 16-year (**d**) 17-year time-lag in the New Zealand CSEP testing region from 1986 to 2012. Error bars are 95% confidence intervals according to the T-test. The number of target earthquakes is given above the error bars. Target earthquakes for which a model gave a zero rate are excluded from the paired comparisons.

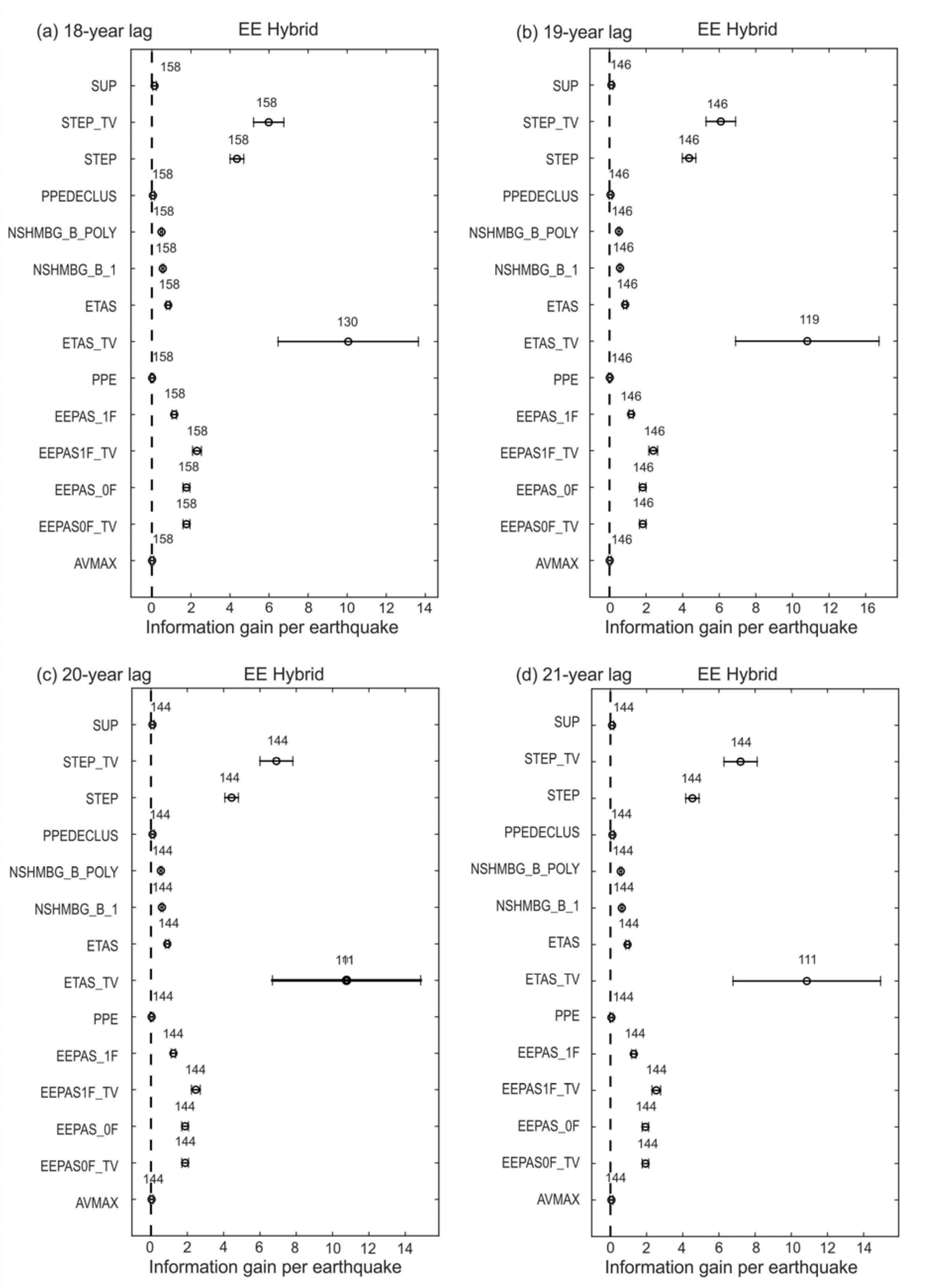

Figure 14 Information gain per earthquake of the EE hybrid models over other models in one year forecasts with (**a**) 18-year (**b**) 19-year (**c**) 20-year (**d**) 21-year time-lag in the New Zealand CSEP testing region from 1986 to 2012. Error bars are 95% confidence intervals according to the T-test. The number of target earthquakes is given above the error bars. Target earthquakes for which a model gave a zero rate are excluded from the paired comparisons.

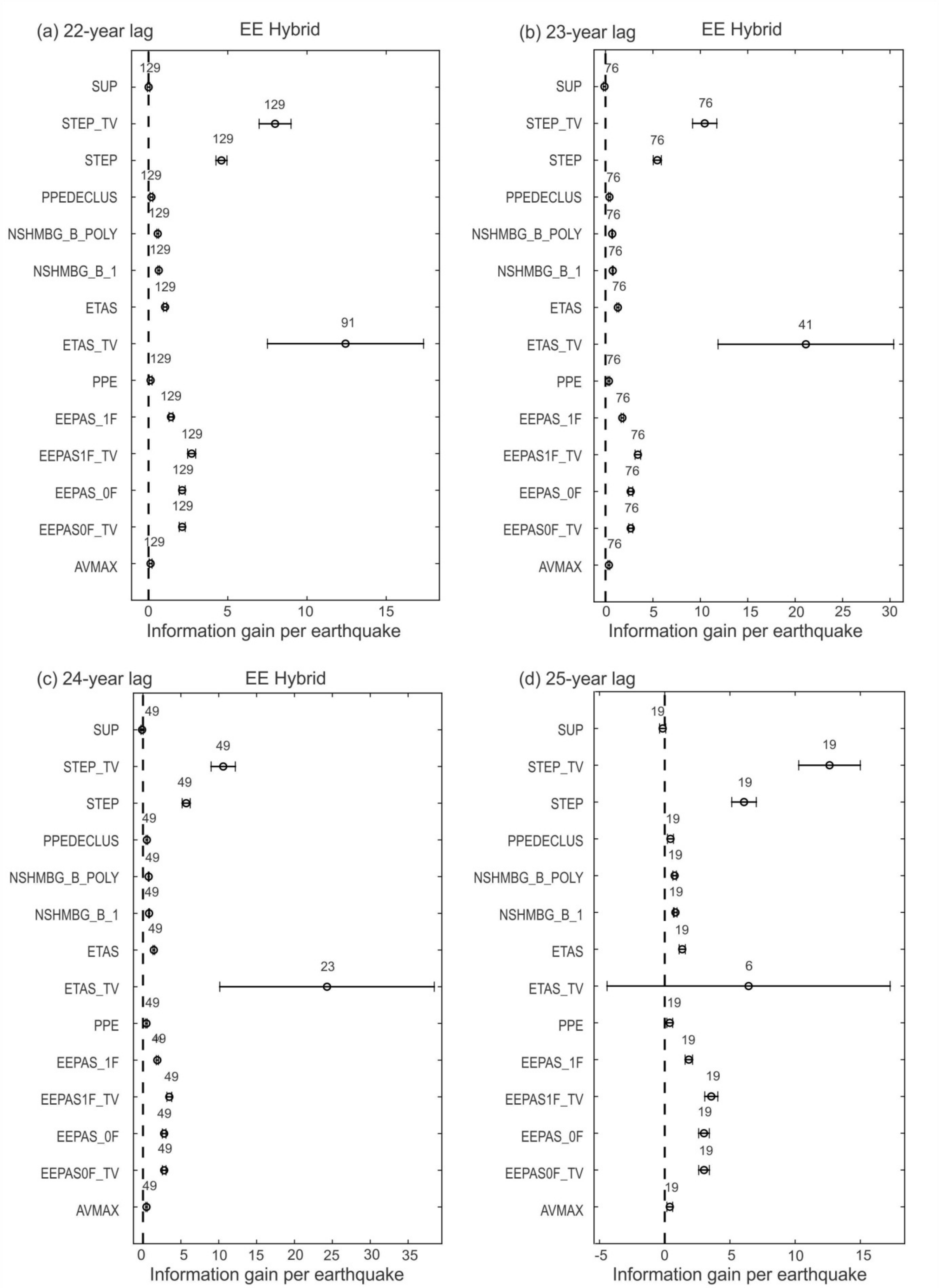

Figure 15 Information gain per earthquake of the EE hybrid models over other models in one year forecasts with (**a**) 22-year (**b**) 23-year (**c**) 24-year (**d**) 25-year time-lag in the New Zealand CSEP testing region from 1986 to 2012. Error bars are 95% confidence intervals according to the T-test. The number of target earthquakes is given above the error bars. Target earthquakes for which a model gave a zero rate are excluded from the paired comparisons.

Attention is now turned to comparisons of the hybrid models with selected reference models. As reference models, we use the SUP model, as a model of least information, and the PPE model, as a spatially varying smoothed seismicity model. Both of these models are normalised to forecast both main shocks and aftershocks above the minimum target magnitude of 4.95. Figure 16 and Figure 17 show the information gain per earthquake of the EE and AVMAX hybrid models over the SUP and PPE models, respectively, as a function of the time-lag.

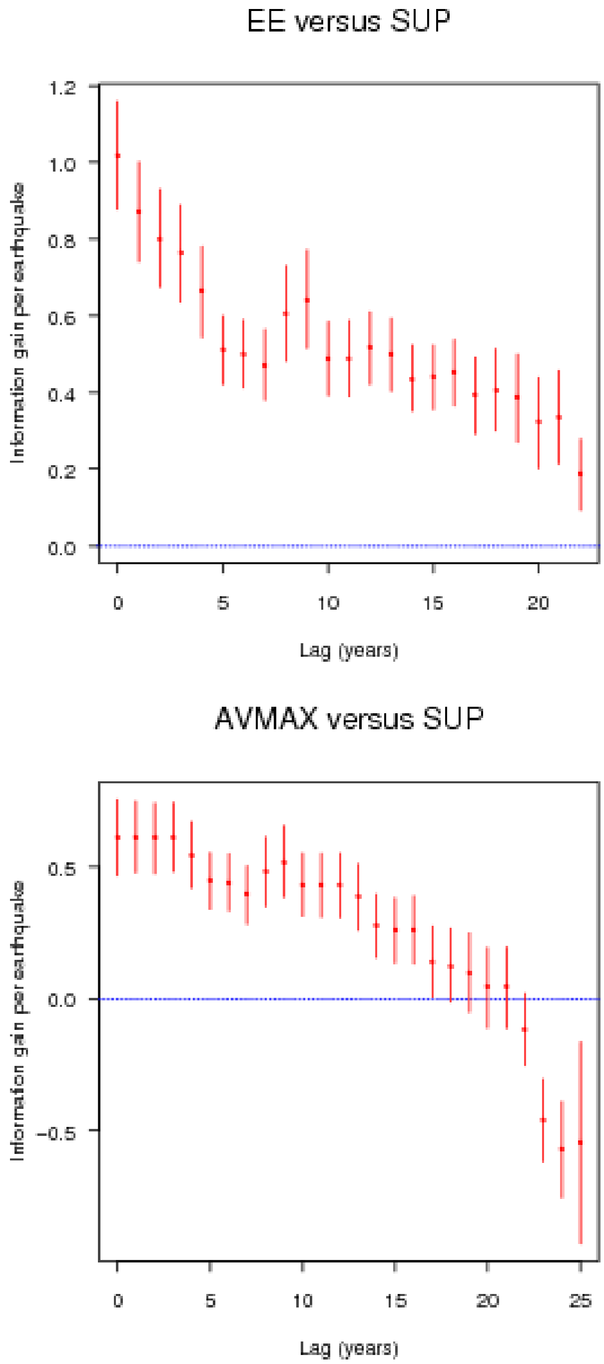

Figure 16 Information gain per earthquake of the EE (upper panel) and AVMAX (lower panel) hybrid models over the SUP model, as function of time-lag. Error bars are 95% confidence limits.

Both the EE and AVMAX models are significantly more informative than the SUP model at short time-lags (Figure 16), with the information gain dropping off as the time-lag increases, and becoming negative for time-lags greater than 20 years (but not significantly so in the case of the EE model). Again, both the EE and AVMAX models are more informative than the PPE model at short time-lags (Figure 17). The information gain of the AVMAX model over PPE drops off to be close to zero for time-lags beyond about 15 years – a result which is easy to understand from the definition of the model, in which the PPE model is likely to dominate the time-varying component at long time-lags. The information gain of the EE model over PPE drops significantly below zero at a five year time-lag, but then begins to recover at a 14 year lag to be significantly above zero for time-lags beyond 21 years.

The abnormally high number of earthquakes from the year beginning 4 September 2008 will have influenced these results. In particular, the Canterbury earthquakes beginning in 2010 tend to favour the SUP model over all other models, because they occurred in a location where few earthquakes have occurred before. It should be borne in mind also that the number of target earthquakes reduces as the time-lag increases (Figure 2), so that the results at long time-lags become increasingly less robust. This is not only because a small number of target earthquakes is likely to be a less representative sample, but also because the error bars on the information gains are based on the normal distribution approximation, which is only justified by the central limit theorem when the number of target earthquakes is large.

The influence of the Canterbury earthquakes is clearly illustrated by Figure 18, which shows the comparable results to Figure 16 when the target earthquakes from September 2008 on are excluded. Now the information gain of both hybrid models over the SUP model holds up well out to time-lags of 22 years. The increase in information gain for time-lags greater than about 17 years is almost certainly an artefact due to a small sample size. What we would expect to see is the information gain reducing initially as the time-lag increases, and then tailing off to an approximately constant level beyond 10-15 years as the information in the time-varying component of the hybrid models runs out, but the spatial information in the long-term component persists.

Figure 19 shows the information gain of the EE and AVMAX models over the PPE model excluding the target earthquakes from September 2008 on. Both hybrid models show a reducing information gain as the time-lag increases. In the case of the EE hybrid, the information gain becomes negative for time-lags greater than about five years. This contrasts with Figure 17, which shows positive information gains for time-lags greater than 20 years. These conflicting results underline that some of the effects seen at longer time-lags are not robust. The comparison of AVMAX with PPE in the lower panel of Figure 19 is, however, easily understood, and generally similar to the corresponding comparison in Figure 17. The information gain is initially positive at short time-lags due to the information provided by the short-term and medium-term clustering models, but is zero beyond about 15 years because the time-varying information does not extend beyond this time-lag.

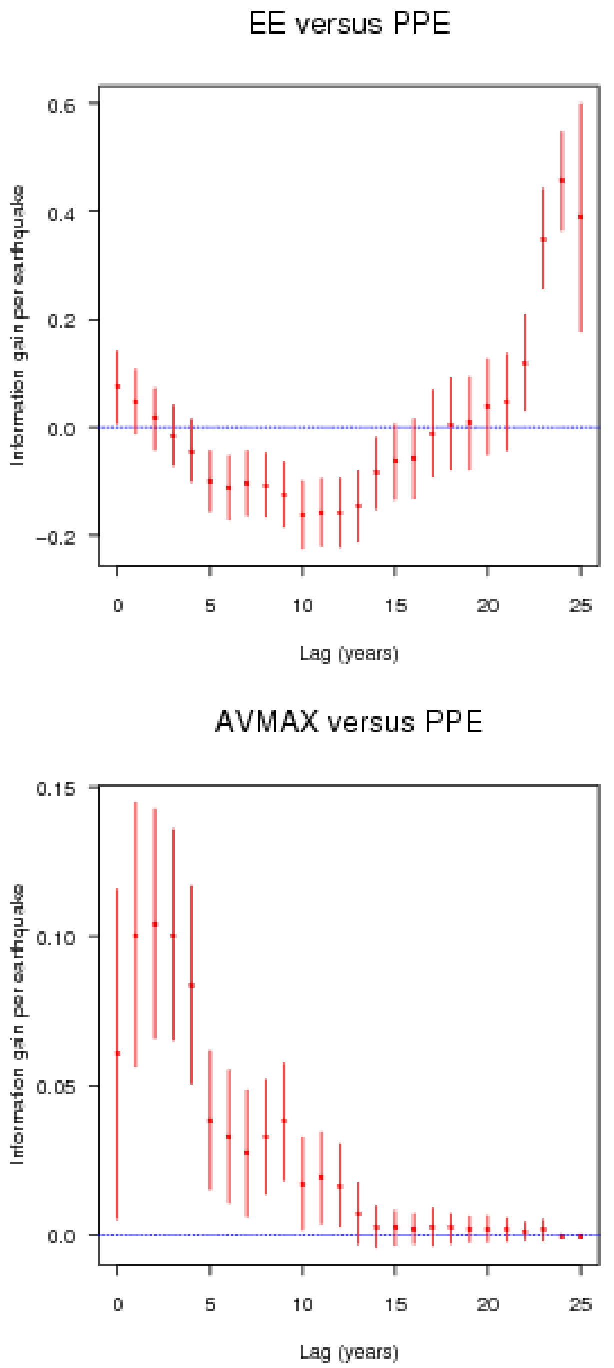

Figure 17 Information gain per earthquake of the EE (upper panel) and AVMAX (lower panel) hybrid models over the SUP model, as function of time-lag. Error bars are 95% confidence limits.

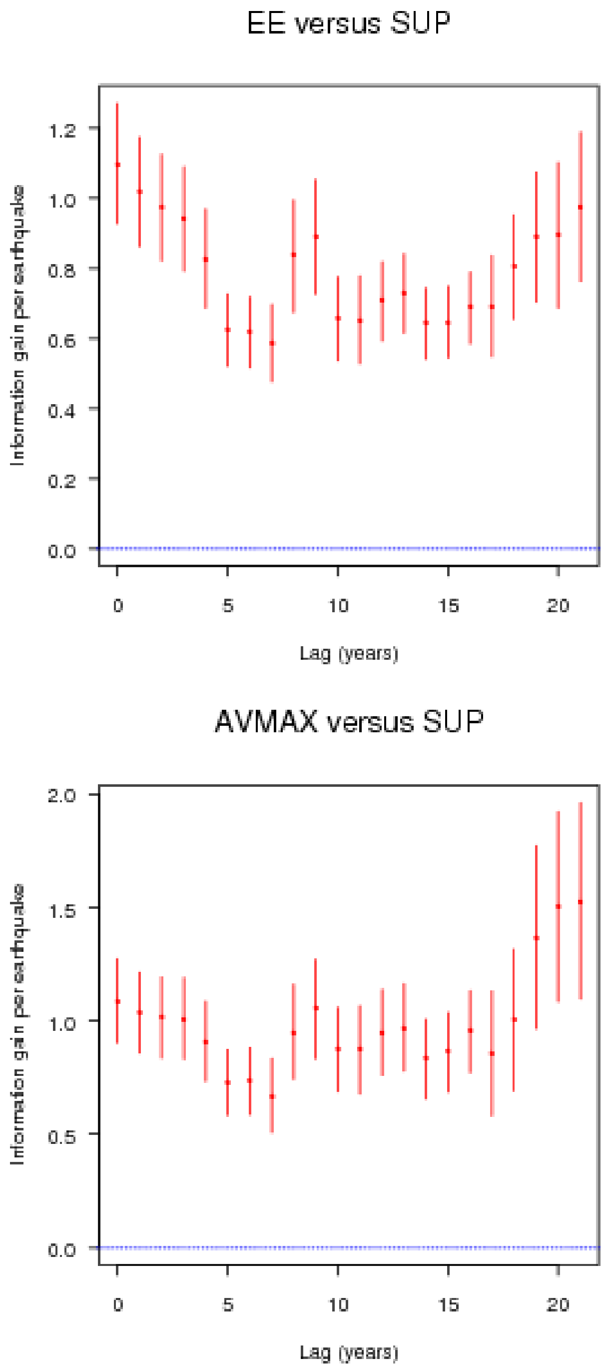

Figure 18 Information gain per earthquake of the EE (upper panel) and AVMAX (lower panel) hybrid models over the SUP model, as function of time-lag, excluding target earthquakes from 4 September 2008 to 3 September 2012. Error bars are 95% confidence limits.

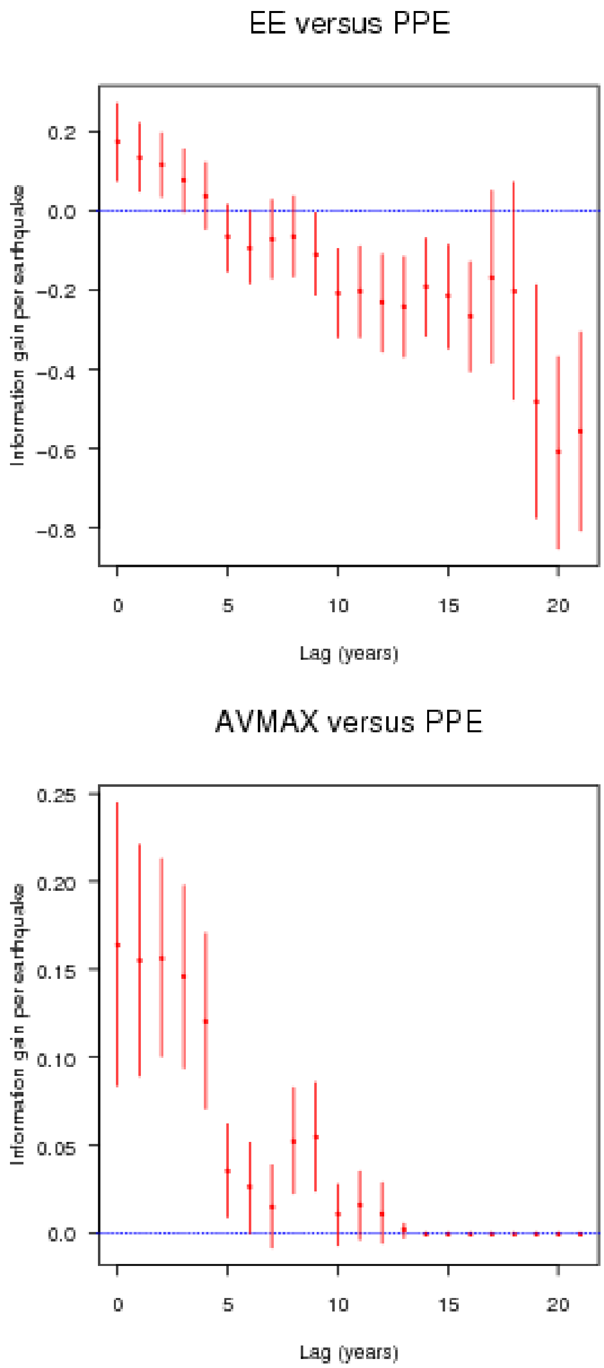

Figure 19 Information gain per earthquake of the EE (upper panel) and AVMAX (lower panel) hybrid models over the PPE model, as a function of time-lag, excluding target earthquakes from 4 September 2008 to 3 September 2012. Error bars are 95% confidence limits.

6.0 ESTIMATION OF OPTIMAL HYBRID MODELS

Optimal Long-term hybrid model

For each time-lag from 0 – 25 years, we find the optimal (maximum likelihood) convex linear (cell-by-cell) combination of the long-term model rates of the PPE, PPEDECLUS, NSHMBG, SUP and PPE1950 models. That is, we find the coefficients (a_i: i = 1, …, 5) to maximise the likelihood of LT_OPT, where

$$\mathrm{LT\_OPT} = a_1\mathrm{PPE} + a_2\mathrm{PPEDECLUS} + a_3\mathrm{NSHMBG} + a_4\mathrm{SUP} + a_5\mathrm{PPE1950}$$ **Equation 2**

and

$$\sum_{i=1}^{5} a_i = 1.$$ **Equation 3**

In Equation 2, the name of each model represents the rate of that model in any given cell. The SUP model is included in this analysis because it has already been seen to be one of the more informative long-term models at long time-lags. For simplicity, the PPE1950 model is included here rather than PPE_FROM_1840 model because the latter model is just a linear combination of PPE and PPE1950. Not much is lost, and some additional flexibility is gained, by overlooking the specific weights that the PPE_FROM_1840 model assigns to its two components at different times. Note that the NSHMBG model referred to here, and in what follows, is the NSHMBG_B_POLY model of the previous sections.

The optimal coefficients for each discrete time-lag are shown in Figure 20. It can be seen that the two individual models contributing most to the optimal combination are PPE and PPE1950. The PPE model is the dominant component for time-lags up to 15 years, and PPE 1950 is the dominant component at time-lags greater than 15 years. Note that LT_OPT is the combination that best forecasts all the earthquakes above the threshold magnitude 4.95, including both main shocks and aftershocks.

The PPE, PPE1950 and SUP models were normalised to forecast all earthquakes, including aftershocks, whereas PPEDECLUS and NSHMBG were designed to forecast only the mainshocks. It is not surprising that the models designed to forecast all the target earthquakes would contribute most to the optimal mixture. The significant contribution of PPE1950 to the optimal mixture at long time-lags reflects the fact that earthquakes are clustered in time and space, on time scales up to a few decades. The PPE model has an advantage at short time-lags of only a few years because of this clustering, but is disadvantaged relative to PPE1950 at longer time-lags, because the latter model relies on a longer 110-year catalogue, albeit a catalogue of lower quality which only included earthquakes of magnitude 6 and above. By comparison the catalogue on which the PPE model is based covers a period of time ranging from 36 to 51 years during this retrospective experiment.

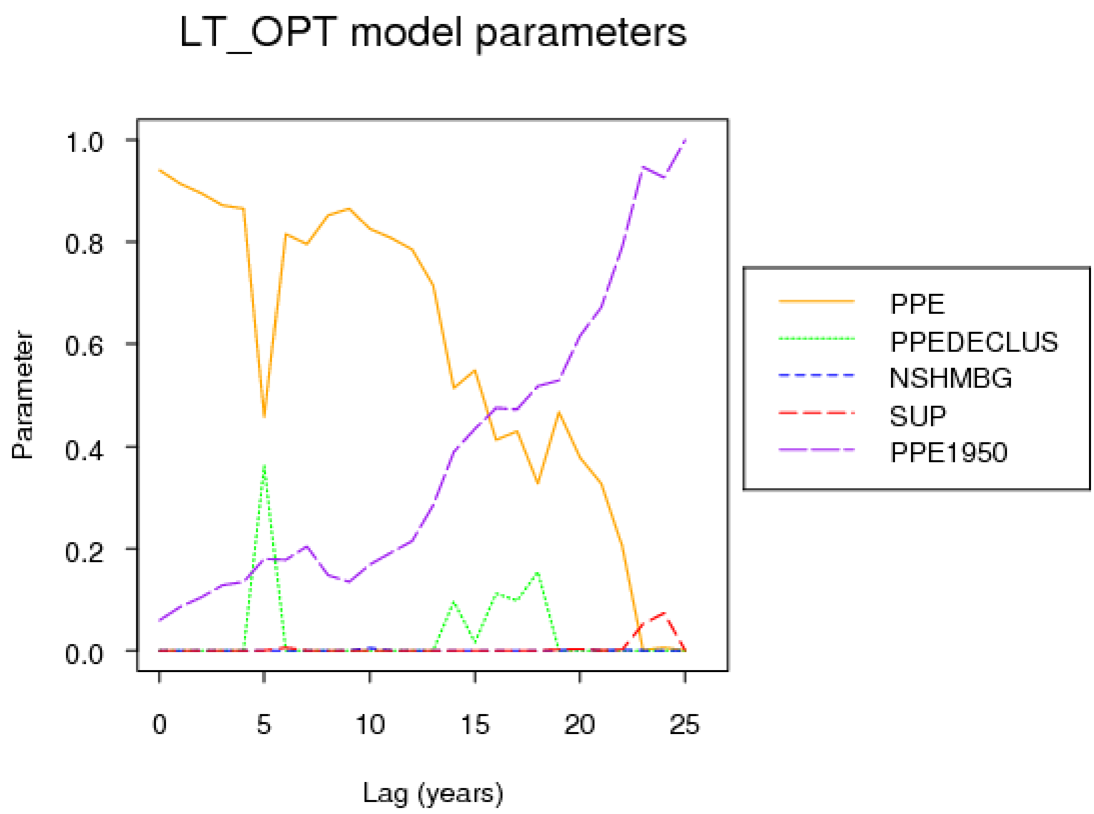

Figure 20 Parameters of the LT_OPT model when fitted with different time-lags.

When comparing two models, one or more of which has been optimised on the data at hand, we use an estimate of the information gain per earthquake (IGPE) based on the Akaike Information Criterion (AIC) (Akaike, 1965). Specifically, the IGPE of model *A* over model *B* is given (Rhoades and Gerstenberger, 2009) by

$$IGPE(A,B) = \frac{AIC_B - AIC_A}{2N_{eq}}$$ **Equation 4**

where N_{eq} is the number of target earthquakes and AIC_A is the AIC statistic for model A, given by:

$$AIC_A = -2\ln L_A(x) - 2p_A,$$ **Equation 5**

where ln $L_A(x)$ is the log-likelihood of the target earthquakes *x* under model *A* and p_A is the number of free parameters involved in fitting model *A*. For model LT_OPT, the number of free parameters is 4. Although 5 coefficients are involved (Equation 4), the number of free parameters is reduced by 1 by the convexity constraint (Equation 3).

Figure 21 shows the IGPE of the LT_OPT model over the SUP, PPE and EE Hybrid models as a function of the time-lag. The error bars in this and subsequent figures are based on a difference of 6 in the numerator of Equation 4. Allowing for the number of free parameters fitted, the error bars are such that the likelihood corresponding to any IGPE value within the error bars is within a factor of 20 of the estimated value.

Figure 21(a) shows there is a solid, but not large, gain over SUP at every time-lag less than 25 years. The negative IGPE at a lag of 25 years is due to the correction for the number of free parameters; the log-likelihood of the LT_OPT being only slightly greater than that of the

SUP model. Note that the SUP model was the most informative of all the models plotted at a time-lag of 25 years in Figure 19(d).

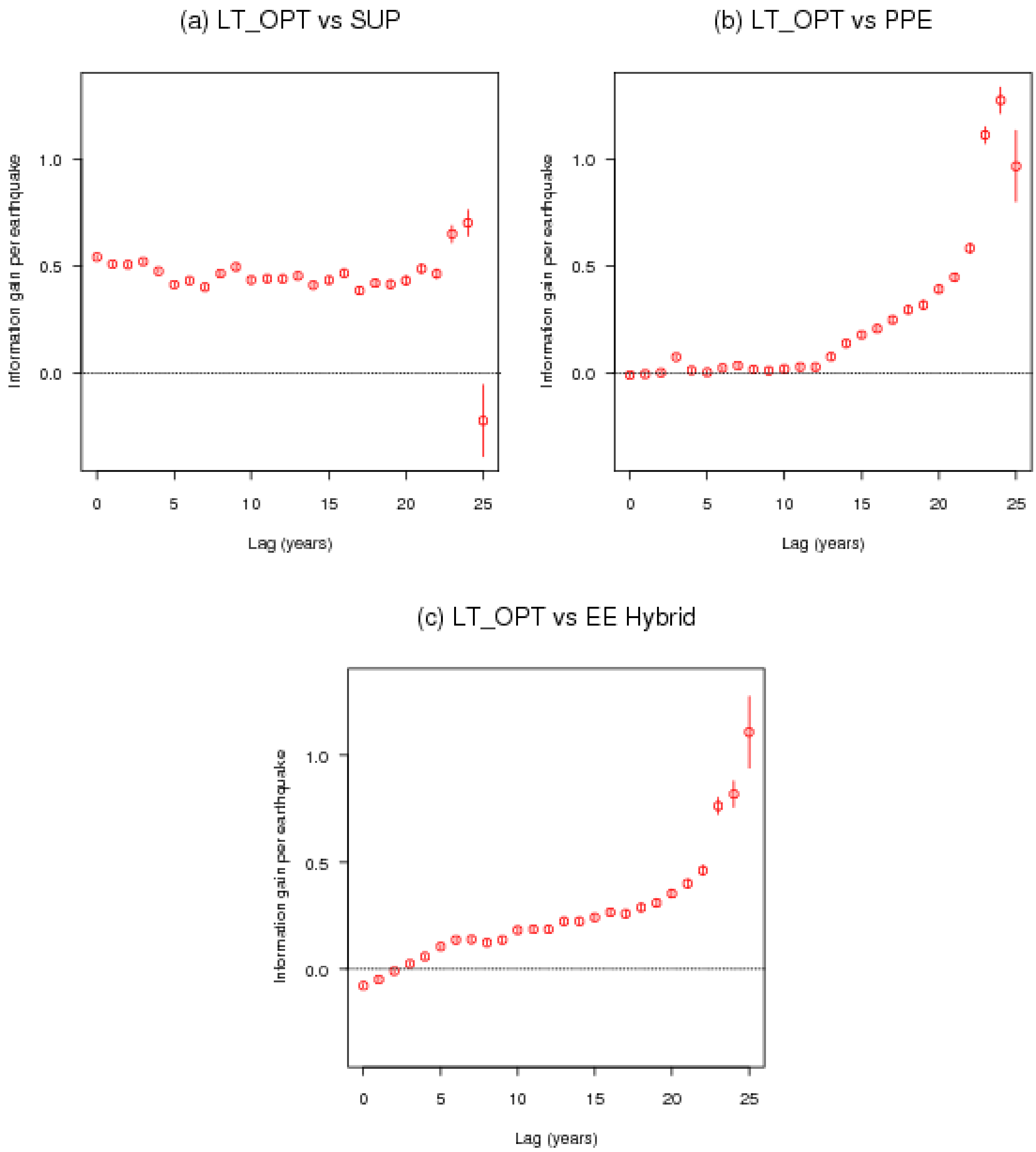

Figure 21 Information gain per earthquake of the LT_OPT model over the (a) SUP model; (b) PPE model; (c) EE Hybrid model, as a function of time-lag. Error bars are based on a likelihood ratio of 20 (see text).

Figure 21(b) shows the IGPE of LT_OPT relative to PPE is close to zero for time-lags out to 12 years but for longer time-lags increases rapidly to about 1 at a lag of 25 years. This trend is consistent with the fact that the weight of PPE in LT_OPT is greater than 0.8 for time-lags up to 12 years, but drops off steadily beyond that to become zero beyond 22 years (Figure 20).

Figure 21(c) shows that the LT_OPT model is less informative than the EE Hybrid model for time-lags up to two years. For longer time-lags it becomes more informative, and for lags greater than 20 years much more informative than the EE Hybrid model. This result is expected to be partly due to the particular characteristics of the seismicity during the experimental period of the retrospective experiment, including the occurrence of the Darfield earthquake and its aftershocks in a region of previously low seismicity, as already discussed above.

6.1 OPTIMAL LONG-TERM AND TIME-VARYING HYBRID MODEL

We now optimise the maximum of convex linear combinations of long-term and time-varying model components at each time-lag. The long-term component models considered are the same as in the previous section. The time-varying component models considered are those included in the EE hybrid model, namely STEP_TV, ETAS_TV, EEPAS0F_TV and EEPAS1F_TV. In what follows, we drop the "_TV" appendage to keep the nomenclature simple.

The new optimum model is denoted LT_TV_OPT. In this model, the long-term and time-varying components are simultaneously optimised, so that the weights accorded to the long-term component models are not assumed to be the same as in LT_OPT. The long-term (LT) model is defined by Equation 2 and Equation 3 above. The time-varying (TV) model is of the form

$$\mathrm{TV} = b_1\mathrm{STEP} + b_2\mathrm{ETAS} + b_3\mathrm{EEPAS0F} + b_4\mathrm{EEPAS1F}, \qquad \textbf{Equation 6}$$

where the coefficients b_1, …, b_4 are constrained to sum to 1. The latter constraint reduces the number of free parameters by 1, so the TV model has three free parameters. The LT_TV_OPT model forecast in a given time-space-magnitude cell is the maximum of the LT and TV forecasts in that cell. The LT_TV_OPT model has a total of seven free parameters – four associated with the LT model and three associated with the TV model.

The LT_TV_OPT model parameters are shown in Figure 22 as a function of the time-lag. The long-term model parameters are generally similar to those of the LT_OPT model (Figure 20), but with some notable differences. We should expect to see the greatest differences for time-lags less 15 years, at which the time-varying model will make its main contribution. First, note that the models designed to forecast only the main shocks now make a larger contribution to the optimal mixture. In particular, the NSHMBG model is the dominant contributor to the optimal mixture at a zero time-lag, when the time-varying contribution is greatest. Also, the PPEDECLUS model makes occasional contributions to the optimal mixture for time-lags up to 13 years, with correspondingly reduced contributions from the PPE model. Surprisingly, the SUP model makes a small contribution to the optimal mixture at time-lags up to three years. The reasons for all of these differences are not clear, but they do illustrate that even models that are individually of low information value can sometimes make useful contributions to an optimal hybrid model, depending on what other models are included in the mix. Certainly, there is no direct relation between the weight accorded to a particular model in an optimal hybrid model and its information value as a stand-alone model.

The time-varying model parameters are shown in the lower panel of Figure 22. The main interest is in the parameter values up to about 12 years, because beyond this time-lag the forecast rates of all the component time-varying models are low. In this range, medium-term clustering models dominate the optimal time-varying model, with the greatest contribution coming from EEPAS0F and a minor contribution coming from EEPAS1F at certain time-lags. There is hardly any contribution from the short-term clustering models, STEP and ETAS, at time-lags up to 10 years. There is a minor contribution at 11 years, but their contributions become comparable with those of the medium-term models only at time-lags greater than 15 years.

Figure 23 shows the IGPE of the LT_TV_OPT model over the LT_OPT, SUP, PPE and EE Hybrid models as a function of the time-lag. Figure 23(a) shows that the information gain of LT_TV_OPT over LT_OPT drops from about 0.23 at a zero time-lag to zero at time-lags greater than 12 years. This indicates that time-varying information only lasts out to a time horizon of about 12 years. Beyond 12 years, any increase in the log-likelihood due to the inclusion of the time-varying model is less than the penalty in the AIC criterion for the three extra parameters fitted. The IGPE of LT_TV_OPT over the SUP, PPE and EE Hybrid models Figure 23(b - d) can be compared to that of the LT_OPT model in Figure 21(a - c), respectively. In each case, the main difference seen is due to the information gain of LT_TV_OPT model over the LT_OPT model for time-lags from zero up to 12 years.

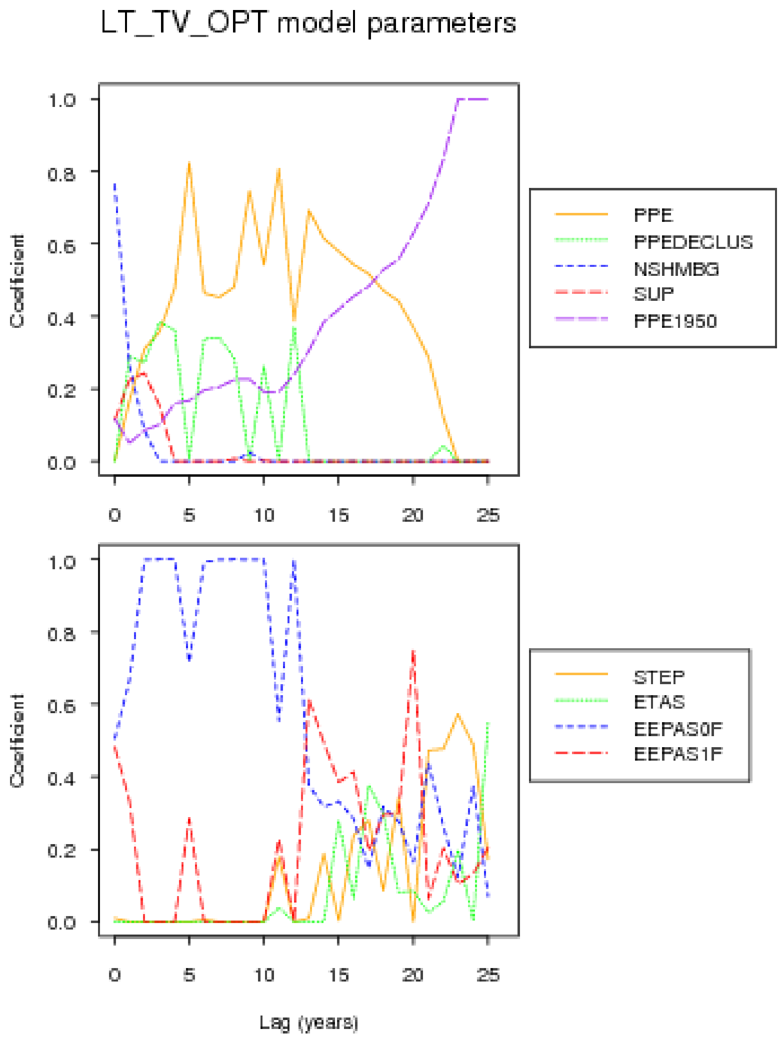

Figure 22 Fitted parameters of the LT_TV_OPT model as a function of time-lag. Upper panel: Coefficients of long-term models; Lower panel: Coefficients of time-varying models.

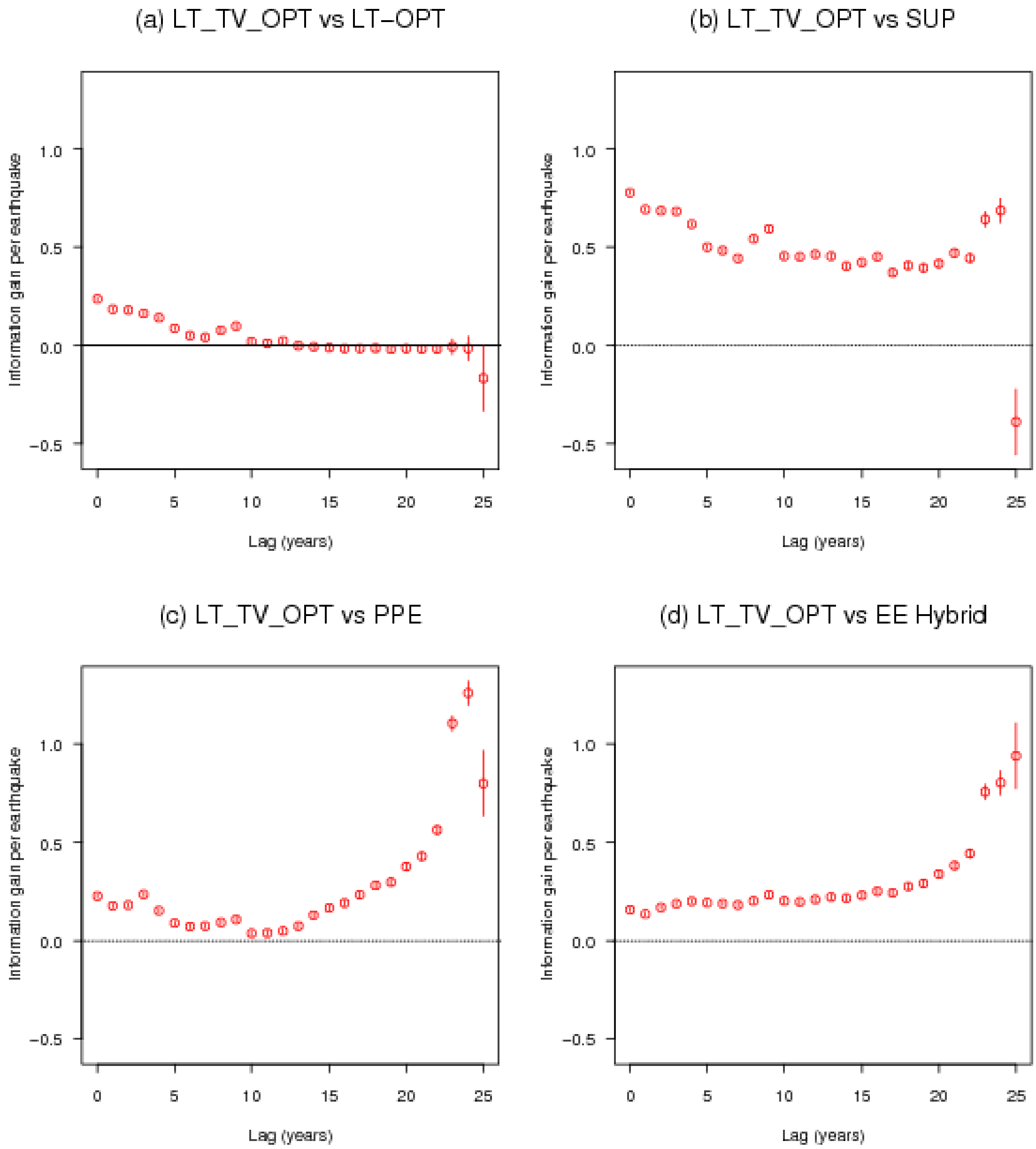

Figure 23 Information gain per earthquake of the LT_TV_OPT model over the (**a**) LT_OPT model; (**b**) SUP model; (**c**) PPE model; (**d**) EE Hybrid model, as a function of time-lag.

6.2 OPTIMAL LONG-TERM, MEDIUM-TERM AND SHORT-TERM HYBRID MODEL

In order to further examine whether the short-term models could contribute anything to an optimal hybrid, we split the models into three classes – long-term, medium-term and short-term – and optimise the maximum of a convex linear combination the models in each class. The long-term component models considered are the same as in the previous section. The time-varying component models are split into a short-term model, consisting of a convex linear combination STEP and ETAS, and a medium term model, consisting of a convex linear combination of EEPAS0F and EEPAS1F. The new optimum model is denoted LT_MT_ST_OPT. All of these models are simultaneously optimised.

To be precise, the long-term (LT) model is defined by Equation 2 and Equation 3 above. The medium-term (MT) model is of the form

$$\mathrm{MT} = c\mathrm{EEPAS0F} + (1-c)\mathrm{EEPAS1F}, \qquad \textbf{Equation 7}$$

and the short (ST) model is of the form

$$\mathrm{ST} = d\mathrm{STEP} + (1-d)\mathrm{ETAS} \qquad \textbf{Equation 8}$$

The LT_MT_ST_OPT model forecast in a given cell is the maximum of the LT, MT and ST forecasts in that cell. Since the MT and ST models each have one free parameter, the total number of parameters in the LT_MT_ST model is six, one less than the LT_TV model. The optimal parameters are shown in Figure 24.

Comparing the upper panel of Figure 24 with the upper panel of Figure 22, we note that the trend of each parameter against time-lag is very similar in both Figures. Therefore the optimal long-term component is virtually unaffected by the different treatment of the time-varying models.

The middle panel of Figure 24 shows a fluctuating dominance of one short-term model over the other, with ETAS dominating STEP at a zero time-lag, but STEP dominating ETAS at time-lags of three years, seven years and eleven years. On the other hand, the lower panel shows that EEPAS0F consistently dominates EEPAS1F for lags between zero and 12 years and is completely dominant between 2 years and 12 years.

Figure 25 shows the IGPE of the LT_MT_ST_OPT model over the LT_OPT, LT_TV_OPT, EE Hybrid and SUP models. The plots involving LT_OPT, EE Hybrid and SUP, when compared with the corresponding plots in Figure 23, show that the information value of the LT_MT_ST_OPT model is not much different on the whole from the LT_TV_OPT model. However, Figure 25(b) shows that there are small but significant differences. Most notably, LT_MT_ST_OPT is less informative than LT_TV_OPT at a zero time-lag, with an IGPE of about -0.05. This shows that forcing a contribution from the short-term models is not helpful. However, at time-lags from one to twelve years, the forced contribution from the short-term models is beneficial and gives a small but sometimes significant IGPE up to about 0.015, involving increases in the log likelihood and not only in the AIC. On the other hand, the insignificantly positive IGPEs for time-lags greater than 12 years can be attributed entirely to the reduction in the number of parameters by one from LT_TV_OPT to LT_MT_ST_OPT.

6.3 SENSITIVITY OF RESULTS TO THE CANTERBURY EARTHQUAKES

The Canterbury earthquake sequence occurred in a region which had previously seen a low level of seismic activity in recent historical times (from 1840) and an even lower level during the period in which seismograph networks have provided catalogues with adequate completeness over the test region (from 1950 for magnitudes $M \geq 5$ and from 1964 for $M \geq 4$). In this section we examine the effect of the Canterbury earthquakes on the LT_TV_OPT model, which has a similar form to the EE Hybrid model now being used as an operational forecasting model for the Canterbury region. Accordingly, we refit the LT_TV_OPT model considering only the years prior to the one in which Darfield earthquake mainshock occurred.

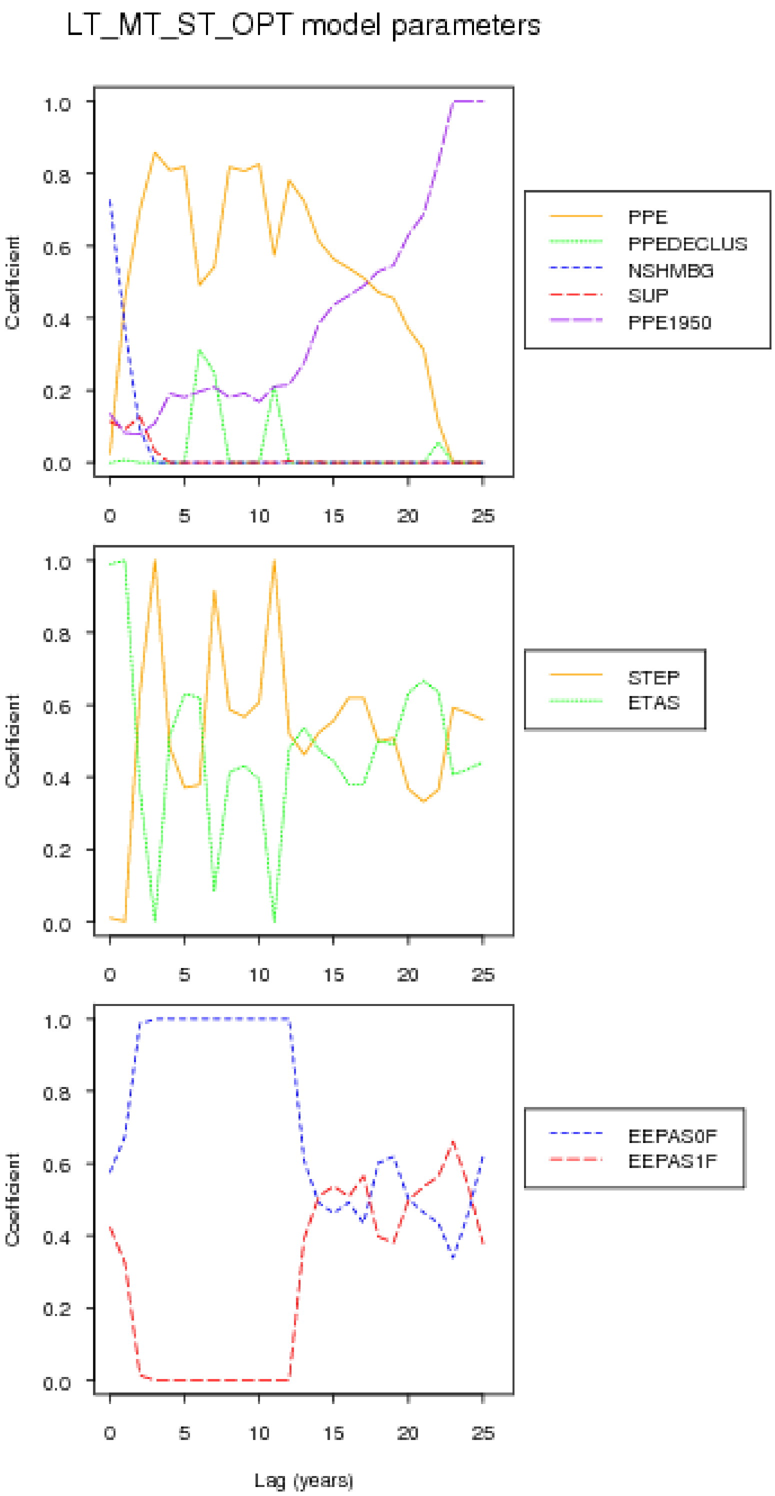

Figure 24 Fitted parameters of the LT_MT_ST_OPT model as a function of time-lag. Upper: Coefficients of long-term models; Middle: Coefficients of short-term models; Lower: Coefficients of medium-term models.

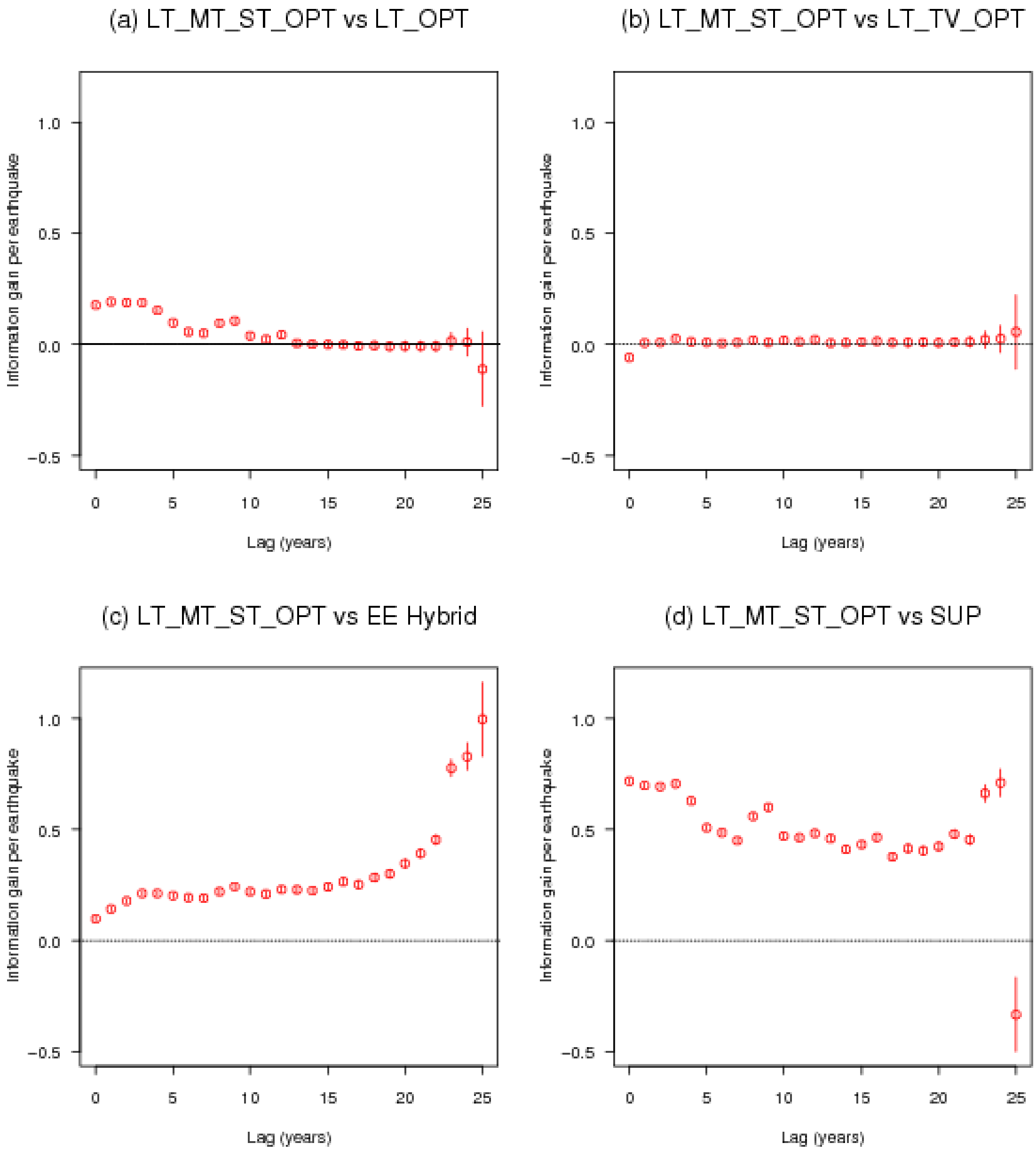

Figure 25 Information gain per earthquake of the LT_MT_ST_OPT model over the (**a**) LT_OPT model; (**b**) LT_TV_OPT model; (**c**) EE Hybrid model; (**d**) SUP model, as a function of time-lag.

Figure 26 shows the parameters of the LT_TV_OPT model optimised for target earthquakes in the period from 4 September 1986 to 3 September 2009. The maximum lag is now 22 years because of the data restriction. The upper panel shows the parameters of the optimal long-term model. Comparison with the upper panel of Figure 22 points to one major difference brought about by restricting the target period. The PPE1950 model now makes no significant contribution to the optimal model; in particular, its dominant contribution at long time-lags has vanished. Instead, PPE is the dominant model at time-lags greater than 10 years. The dominance of the NSHMBG model at short time-lags is enhanced by removing the Canterbury sequence from the target set; at a zero lag this model is now completely dominant. The minor contributions from the PPEDECLUS model are also increased both at short and long time-lags. Also, the minor contributions from the SUP model at short time-lags have vanished.

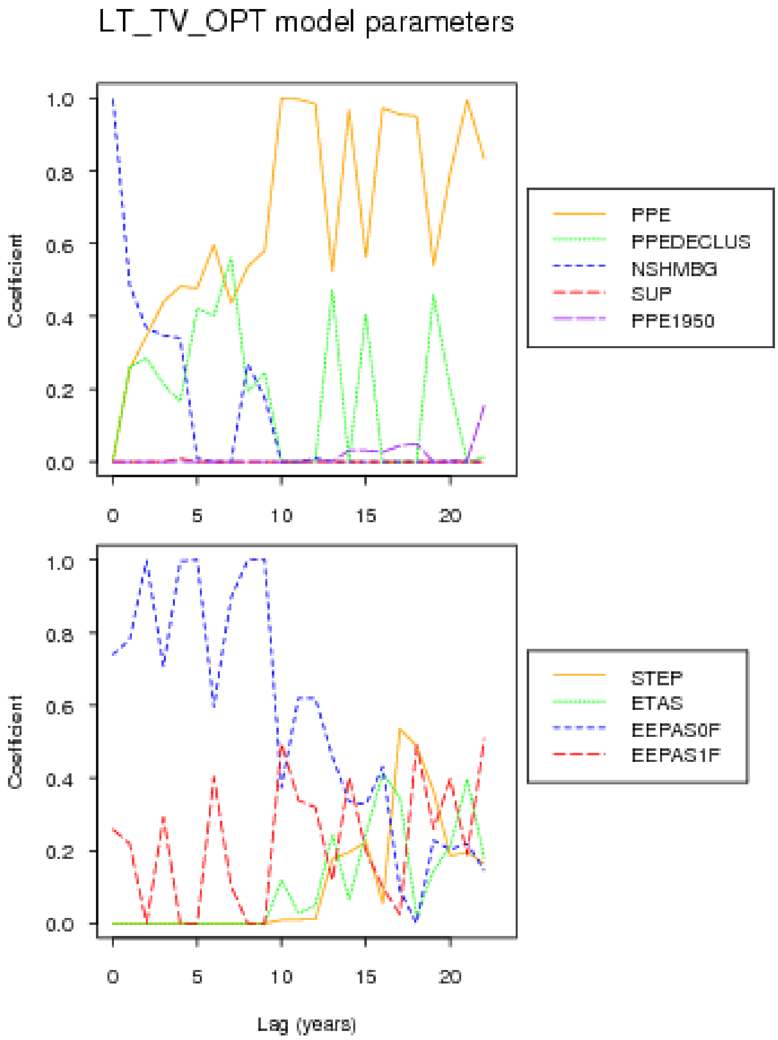

Figure 26 Parameters of the LT_TV_OPT model as a function of time-lag, using target earthquakes up to 2009 September 3. Upper panel: Coefficients of long-term models; Lower panel: Coefficients of time-varying models.

The optimal parameters of the time-varying model are shown in the lower panel of Figure 26. On the whole, these are not much changed from those in Figure 22. The EEPAS models still contribute most to the optimal mixture out to time-lags of 12 years, with the dominant contribution coming from the EEPAS0F model.

Figure 27 shows the IGPE of the LT_TV_OPT model optimised for target earthquakes in the period from 4 September 1986 to 3 September 2009 over the SUP, PPE and EE Hybrid models. This can be viewed in comparison with Figure 23(b-d), which shows the corresponding plots for the model fitted to the whole test period.

Large differences are evident between Figure 27 and Figure 23. The information gain over the SUP model (Figure 27a; c.f. Figure 23b) is now larger at short time-lags and generally declines as the time-lag increases, though remaining positive – above 0.4 – out to 22 years. The information gain over the PPE model (Figure 27b; c.f. Figure 23c) is also larger at short time-lags and decreases to zero at about 10 years, marking the limit of the contribution from time-varying models. As noted above, the PPE model now dominates long-term model at long time-lags. The negative IGPE of LT_TV_OPT model relative to PPE at long time-lags can be attributed mainly to the penalty for fitting seven additional parameters without increasing the log-likelihood by much. In contrast, the rising IGPE with increasing time-lags in

Figure 23c is due to the dominance of PPE1950 in the long-term component of that model at long time-lags. The information gain of LT_TV_OPT over the EE Hybrid model (Figure 27c; c.f. Figure 23d) is now fairly consistent across time-lags, varying mostly in the range between 0.2 and 0.3. Similarly to the other comparisons, the IGPE is now higher than before at short time-lags and lower than before at long time-lags, and for the same reasons. Importantly, the information gain of the optimal model over the EE Hybrid model is only moderate for all time-lags.

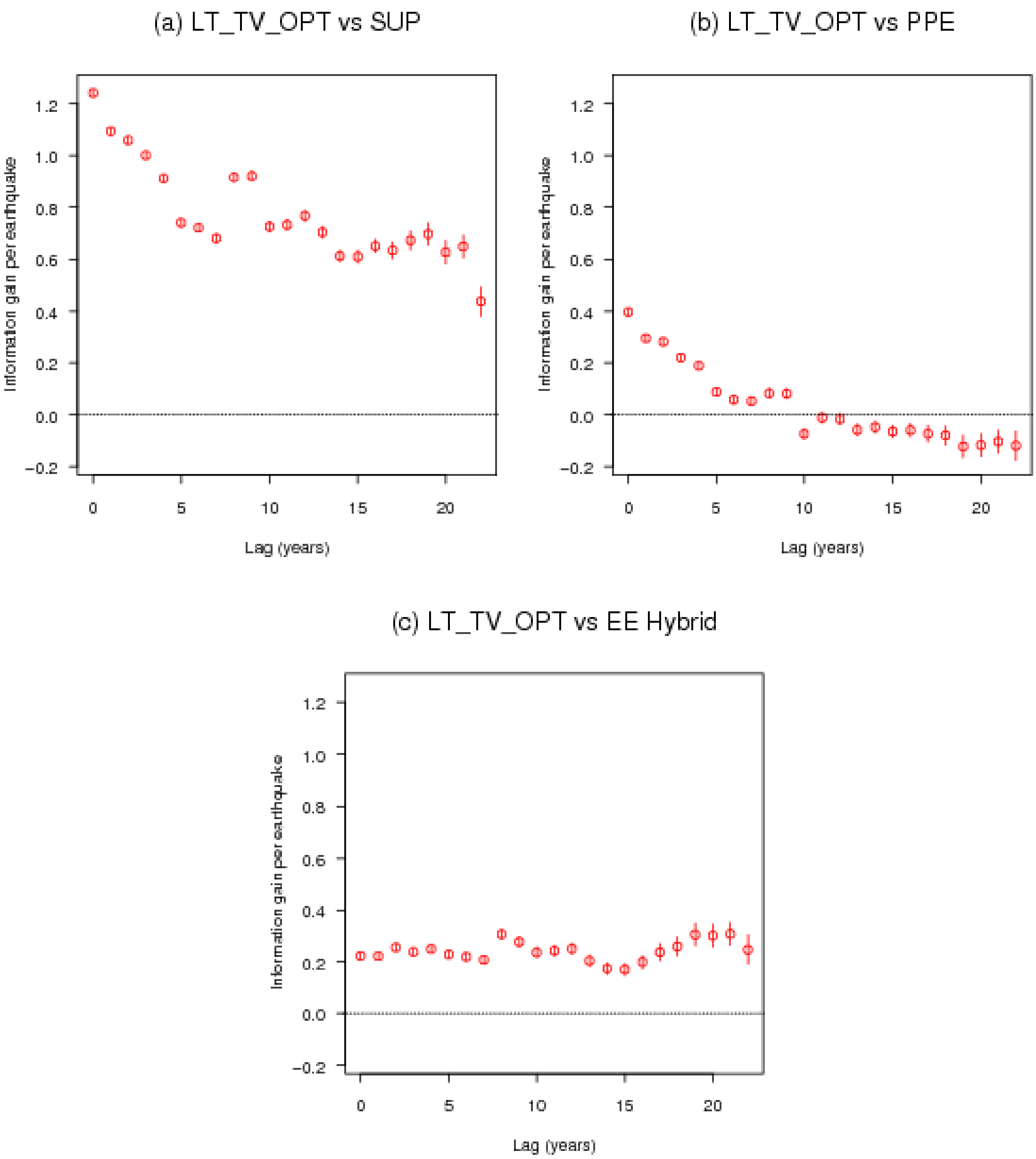

Figure 27 Information gain per earthquake of the LT_TV_OPT model over the (**a**) SUP model; (**b**) PPE model; (**c**) EE Hybrid model, as a function of time-lag, using target earthquakes up to 2009 September 3.

7.0 DISCUSSION

The analyses with restricted target earthquakes have clearly shown that the present earthquake catalogue is not long enough to give robust measurements of model performance in all respects. In section 5.2 it was shown that a period of very high seismicity near the end of the 25-year test period, starting with the M7.8 Dusky Sound earthquake of 17 July 2008 and continued by the Canterbury earthquake sequence, caused all of the models considered in this study to severely under-predict the number of earthquakes during the test period. In section 5.3 it was shown that the information gains of the EE and AVMAX hybrid models are also strongly influenced by this period of high activity. Further, in section 6.4 it was shown that both the parameters and information gains of optimal time-varying hybrid models are strongly affected by the inclusion of the Canterbury earthquake sequence in the target earthquake data set. In the light of such sensitivities, a caution against drawing overly strong conclusions from the results of this study is advisable. Nevertheless there is much to be learned from the analyses presented here.

First, consider the long-term models. There is evidence from this study that long-term models which aim to forecast all earthquakes above the minimum magnitude threshold tend to outperform those which aim to forecast only the main shocks. This is particularly true when comparing the performance of individual models (Figure 8 to Figure 15) or when fitting an optimal long-term model combination (Figure 20). It is also true when fitting a maximum hybrid of a long-term model combination and one or more time-varying model combinations (Figure 22, Figure 24 and Figure 26) for long time-lags, but not for short time-lags (0-1 years), where the NSHMBG model is the dominant contributor to the optimal long-term model combination. The latter result is easy to understand, because it is at short time-lags that the time-varying models tend to be greatest and most informative. These results for long-term models may have implications for the estimation of the distributed seismicity component in traditional probabilistic seismic hazard analysis.

Next consider the medium-term models. This study has added observations supporting evidence that the EEPAS0F_TV outperforms the EEPAS1F_TV model in most situations for a target magnitude threshold of 4.95 (Figure 8 to Figure 15, Figure 22, Figure 24, Figure 26). This suggests that aftershocks can also sometimes be medium-term precursors. i.e., form part of the precursory scale increase to a future major earthquake. At the moment, it appears that this result is dependent on the target magnitude, because it conflicts with previous observations supporting the reverse conclusion for threshold magnitudes of 5.75 in New Zealand and California (Rhoades and Evison, 2004) and 6.75 in Japan (Rhoades and Evison, 2006), but is consistent with previous observations for southern California for a threshold magnitude of 4.95 (Rhoades, 2007) and for the Kanto region of central Japan for a threshold magnitude of 4.75 (Rhoades and Evison, 2006). However, even for a threshold magnitude of 4.95 the situation is somewhat unclear as to the relative value of EEPAS0F and EEPAS1F in hybrid combinations. Rhoades (2013) found that an optimal weighted average of ETAS_TV, EEPAS0F_TV and PPE did not perform as well as optimal weighted average of ETAS_TV, EEPAS1F_TV and PPE in both New Zealand and California for instantaneous forecasts with time-lags ranging from zero out to 3000 days, despite the advantage of EEPAS0F over EEPAS1F in the same regions. The difference in the present study is that there are more models involved in the hybrid, the hybrid is not a simple additive mixture, and the forecasts are for one-year periods. It is understandable that these differences could affect the results.

Thirdly, we consider the short-term models. A surprising result of this study is the small contribution that the short-term models make to the optimal hybrid model LT_TV_OPT (Figure 22 and Figure 26). Further, in the hybrid model LT_MT_ST_OPT there is actually a loss of information by forcing the inclusion of a contribution from the short-term models at a zero time-lag (Figure 25b). It seems that the time steps of one year are too long for the benefits of the short-term models to be realised. However, their value in short-term forecasting with one-day updating has been well demonstrated (Gerstenberger et al. 2005, 2007; Rhoades and Gerstenberger, 2009), and not least in the CSEP tests during the Darfield aftershock sequence (Figure 28).

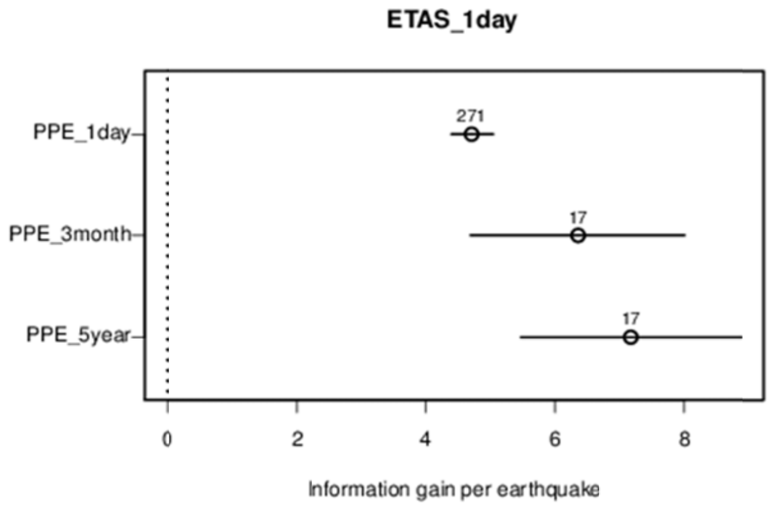

Figure 28 Information gain of the ETAS one-day model over the PPE month with updating at intervals of one day, three months and five years, for earthquakes with magnitude M > 3.95 in the CSEP New Zealand test region during the period from 4 September 2010 to 8 March 2011. About 75% of the earthquakes in this period were in the Darfield aftershock region.

Rhoades (2013) found that in an optimal weighted average of ETAS, EEPAS0F and PPE in the New Zealand CSEP testing region, the ETAS model makes the dominant contribution to the optimal mixture for time horizons less than 10 days, but for longer time-lags the EEPAS0F model contribution is dominant. Nevertheless, the ETAS model makes a significant minor contribution out to time-lags of 3000 days. Therefore, taking the Rhoades (2013) study into account, the very low contributions from the short-term models in the LT_TV_OPT model are still surprising, and are most likely to the use of one-year forecasts instead of instantaneous forecasts, although we note also that a different form of hybrid is employed here and that additional models are included in the hybrid.

It is clear from this study that no individual model consistently outperforms all others. As a consequence, hybrid models tend to outperform most of the individual models most of the time. This is certainly true of both the EE and AVMAX hybrid models (Figure 8 to Figure 15) and supports the use of a hybrid model for decision-making in the wake of the Canterbury earthquakes.

The parameters of the optimal hybrid model have been shown to be sensitive to the selection of the test period (Figure 22 and Figure 26). Therefore, we should not expect that the optimal hybrid for the test period will also be optimal for the next 25 years or 50 years, let alone for a particular region such as Canterbury. The information gain of the optimal hybrid over other models is also sensitive to the test period (Figure 23 and Figure 27). However, we note that when the Canterbury earthquakes are excluded from the test period, the information gain of the optimal hybrid LT_TV_OPT over the EE Hybrid model is modest at all time-lags (Figure 27).

When the whole test period is included, the optimal hybrid model outperforms the EE hybrid by a wide margin (Figure 23), especially at long time-lags where these models are dominated by their long-term components. The wide margin at long time-lags is attributable entirely to the dominance of the PPE1950 model in the long-term component of the optimal hybrid model. This dominance is due to the single fact that the PPE1950 model, which makes only a minor contribution to the long-term component of the EE Hybrid model, has a much higher rate than any other individual long-term model in the region where the Canterbury earthquakes occurred. The contribution of the PPE1950 model to future earthquake hazard in Canterbury is not so important for the future, now that the Canterbury earthquakes have occurred and are part of the learning set for most of the models included in the EE Hybrid looking forward.

If a hybrid forecasting model similar to the EE model were being designed for the whole New Zealand region, it is clear from this study that the weights in the time-varying component should be different from those in the EE model. In particular the contributions from the medium-term models should be greater, and the contributions from the short-term models smaller, than those in the EE model. This conclusion cannot be particularised to the Canterbury region however. The Canterbury earthquake sequence is very unusual in the context of historical and instrumental seismicity in New Zealand. The Darfield earthquake occurred without any appreciable previous seismic activity nearby in the short or medium term, and in a region with low long-term fault slip rates and low present-day crustal strain rates. The subsequent earthquakes in the Canterbury sequence so far are consistent with standard aftershock models like the short-term models in this study (Figure 28). Empirical studies of the precursory scale increase phenomenon, and using the associated EEPAS model, do not give a clear indication whether the medium-term precursors in such a region, if they occur at all, would have a similar time scale to those in more seismically active regions. Therefore, it would be naïve to simply adopt the optimal hybrid model for the whole New Zealand test region for operational forecasting in Canterbury.

Empirical EEPAS modelling of seismicity in the low seismicity intra-plate region of Australia (Somerville et al., 2006; Swindon et al., 2007) has produced some examples of longer precursor times (e.g. Tasmania) and some examples of similar precursor times to New Zealand and other plate boundary regions (e.g. Victoria and New South Wales). A problem for EEPAS studies in low seismicity regions is that such regions are seldom well-catalogued over long-periods and the shortness of the adequate catalogue can bias estimation of the precursor-time parameters.

Major earthquakes occurring elsewhere in low seismicity regions have been found to have aftershock sequences lasting for many decades and conforming well to the Omori-Utsu inverse power law for aftershock decay rates over a long period. The most well-known example of this phenomenon is the aftershock sequence of the M8.0 Mino-Owari (or Great Nobi) earthquake of 1891, in the Gifu prefecture of central Japan, which has closely followed the Omori-Utsu relation for more than 100 years (Utsu et al., 1995). This earthquake, the largest known inland earthquake in Japan, is in some ways a good analogue of the Darfield earthquake. It occurred near to, but somewhat distant from, an active plate boundary in a region which is on the whole much less seismically active than the plate boundary region. A difference is that the plate boundary near to Gifu is a subduction zone rather than a continental collision zone.

If the Canterbury earthquakes continue to follow the pattern of those of the Great Nobi earthquake, or even if a small proportion of the Darfield aftershocks are long-term precursors, then the time-varying component of the EE model, which is dominated by the contributions from the short-term models but with significant minor contributions from the medium-term clustering models, should perform satisfactorily in the coming decades.

There is a great deal of variation in the estimates from the individual long-term models contributing to the EE Hybrid model. In particular, the PPE model forecasts much higher rates than the models that aim to forecast only main shocks, such as NSHMBG and PPEDECLUS. At short time-lags the NSHMBG model (designed to forecast main shocks only) makes a dominant contribution to the optimal long-term model based on earthquakes prior to Darfield (Figure 26) and at long time-lags the PPE model (designed to forecast all earthquakes) is dominant. If a hybrid model were being designed for the whole New Zealand region, it would be best at long time-lags to give more weight to long-term models designed to forecast all earthquakes. Again, this conclusion cannot be particularised to the unusual situation in the Canterbury region. There is a danger that the PPE model could grossly overestimate the number of earthquakes expected in Canterbury in the long-run because of the effect of the large number the Darfield aftershocks with M > 4.95 in boosting its estimates. In the face of the wide variation between models, the EE hybrid, which gives appreciable weight to four different models, is likely to outperform most of the individual models.

8.0 CONCLUSION

In this study we tested individual and hybrid models by comparing their performance over 26 years in the whole New Zealand region, in annual forecasts with time-lags ranging from zero up to 25 years. We also estimated optimal hybrid model combinations over the same period and with the same time-lags. The number of target earthquakes varied from 303 at a time-lag of zero down to 19 at a time-lag of 25 years.

The N-tests showed that all models tend to under-predict the number of earthquakes in the test period. This was shown to be mainly due to the unusually large number of earthquakes with M > 4.95, including the Canterbury earthquakes, that have occurred in the test region since the M7.8 Dusky Sound earthquake of 15 July 2009. The T-tests showed that both the EE and AVMAX hybrid models are more informative than most of the individual models for all time-lags. Using data from the full 26-year test period, the IGPE relative to a spatially uniform reference model drops off steadily as the time-lag increases, to become zero at a time-lag of about 20 years. When the unusual period since the Dusky Sound earthquake is removed from the tests, both hybrid models show a significant positive IGPE over the spatially uniform model at all time-lags, but do not outperform every individual long-term model at long time-lags. The test results are therefore seen to be sensitive to unusual features of the test catalogue, and a much longer catalogue would be needed to obtain robust results.

An optimal hybrid model with the same general form as the EE and AVMAX hybrid models was computed for each time-lag from the full test period. In the optimal hybrid model, the time-varying component is dominated by the medium-term models, with hardly any contribution from the short-term models for time-lags up to 12 years. The short-term and medium-term model rates diminish with increasing time-lag, with the result that the time-varying component as a whole has hardly any impact on the optimal hybrid model for time-lags greater the 12 years.

The long-term component of the optimal hybrid model is dominated by the NSHMBG model. For intermediate time-lags it is dominated by the PPE model, and for time-lags greater than 17 years by the PPE1950 model. At long time-lags the optimal hybrid model is considerably more informative than the EE Hybrid model, with an IGPE close to 1.0 for a time-lag of 25 years. This is because the PPE1950 model is the best individual model for forecasting the Canterbury earthquakes at long time-lags. When the years including the Canterbury earthquakes are removed from the test data set, the contribution of the PPE1950 model vanishes and the PPE model dominates the optimal model for long, as well as intermediate, time-lags. For this reduced data set, the IGPE of the optimal hybrid model over the EE hybrid model is only moderate, in the range 0.2-0.3, for all time-lags.

A three-component hybrid model, with cell rates defined as the maximum of long-term, medium-term and short term rates, was also optimised for each time-lag from the 26-year test period. This model was found to be less informative than the two-component model for zero time-lag, but slightly more informative than the two-component model at longer time-lags.

As a large earthquake in a region of previously low seismicity and low crustal deformation rate, the Darfield earthquake is unique in the New Zealand instrumental earthquake catalogue. Although an optimal hybrid model for New Zealand as a whole would have a higher contribution to the time-varying component from medium-term models, the same does not necessarily apply to the Canterbury region. So far, the Canterbury earthquake sequence appears to be a well-behaved aftershock sequence, similar to that of the 1891 Great Nobi earthquake in central Japan, which decayed according to the Omori-Utsu law for more than 100 years. This aftershock decay, captured by the short-term models, is the most predictable component of future Canterbury seismicity. The medium-term component is less predictable, because the precursor time parameters of the EEPAS model are not well established for low seismicity regions. There is considerable variation in the estimates from the individual long-term models contributing to the EE Hybrid model. But, based on the tests carried out here, the EE hybrid, which gives appreciable weight to four different long-term models, is likely to outperform most of the individual models in the next 50 years.

9.0 ACKNOWLEDMENTS

This research was funded by the Earthquake Commission Research Foundation under grant number BIE 12/633. The report has been internally reviewed by Mark Stirling and David Harte.

10.0 REFERENCES

Evison, F.F.; Rhoades, D.A. 2002. The precursory scale increase and long-term seismogenesis in California and northern Mexico. Annals of Geophysics, 45: 479-495

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

Gerstenberger, M.; Wiemer, S.; Jones, L.; Reasenberg, P. 2005. Real-time forecasts of tomorrow's earthquakes for California. Nature, 435: 328-331

Gerstenberger, M.C.; Jones, L.M.; Wiemer, S. 2007. Short term aftershock probabilities : case studies in California. Seismological Research Letters, 78(1): 66-77

Gerstenberger, M.C.; Rhoades, D.A. 2010. New Zealand Earthquake Forecast Testing Centre. Pure and Applied Geophysics, 167(8/9): 877-892; doi:10.1007/s00024-010-0082-4

Imoto, M. 2007. Information gain of a model based on multidisciplinary observations with correlations. Journal of Geophysical Research, 112: B05306, doi:10.1029/2006JB004662

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

Ogata, Y. 1989. Statistical models for standard seismicity and detection of anomalies by residual analysis. Tectonophysics, 169: 159-174

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

Pollock, D., 2007. Aspects of short-term and long-term seismic hazard assessment in New Zealand. Diploma Thesis, Department of Earth Sciences, ETH Zurich.

Reasenberg, P.A. and Jones, L.M., 1989. Earthquake hazard after a main shock 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. 2013. Mixture models for improved earthquake forecasting at short-to-medium time horizons. Bulletin of the Seismological Society of America, 103(4), doi: 10.1785/0120120233.

Rhoades, D.A.; Evison, F.F. 2004. Long-range earthquake forecasting with every earthquake a precursor according to scale. Pure and Applied Geophysics, 161(1): 47-72

Rhoades, D.A.; Evison, F.F. 2005. Test of the EEPAS forecasting model on the Japan earthquake catalogue. Pure and Applied Geophysics, 162: 1271-1290; doi: 10.1007/s00024-004-2669-0

Rhoades, D.A., and Evison, F.F., 2006. The EEPAS forecasting model and the probability of moderate-to-large earthquakes in central Japan, Tectonophysics. 417, 119-130, doi: 10.1016/j.tecto.2005.05.051.

Rhoades, D.A.; Gerstenberger, M.C. 2009. Mixture models for improved short-term earthquake forecasting. Bulletin of the Seismological Society of America, 99(2A): 636-646; doi: 10.1785/0120080063

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, EQC Research Project 09/TV580. 55p.

Rhoades, D.A.; Schorlemmer, D.; Gerstenberger, M.C.; Christophersen, A.; Zechar, J.D.; Imoto, M. 2011. Efficient testing of earthquake forecasting models. Acta Geophysica, 59(4): 728-747; doi:10.2478/s11600-011-0013-5

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, doi: 10.1785/gssrl.78.1.116.

Schorlemmer, D., Zechar, J.D., Wiemer, S.,Werner, M.J., Field, E.H.,Jackson, D.D. and Jordan, T.H., 2010. First results of the Regional Earthquake Likelihood Models experiment. Pure Applied Geophysics, 167, 859-876.

Swindon, A.; Gillon, M.; Clark, D.; Somerville, P.; Van Dissen, R.J.; Rhoades, D.A. 2007 Seismic hazard assessment of the Lake Edgar Fault. p. 281-289 IN: McSaveney, E.R. (ed.) Dams, securing water for our future: proceedings of NZSOLD ANCOLD 2007 Conference, Queenstown, 19-20 November 2007: technical sessions. Wellington, NZ: Institution of Professional Engineers New Zealand. Proceedings of technical groups / Institution of Professional Engineers New Zealand 33(1LD).

Somerville, P.; Fisher, J.; Rhoades, D.A.; Leonard, M. 2006 Preliminary test of the EEPAS long term earthquake forecast model in Australia. p. 117-122 IN: McCue, K.; Lenz, S. (eds) Earthquake engineering in Australia : conference 2006 : proceedings of a conference held by the Australian Earthquake Engineering Society, Canberra, ACT, 24-26 November 2006. Jamison, ACT: Australian Earthquake Engineering Society

Stirling, M., McVerry, G., Gerstenberger, M., Litchfield, N., Van Dissen, R., Berryman, K., Barnes, P., Wallace, L., Villamor, P., Langridge, R., Lamarche, G., Nodder, S., Reyners, M., Bradley, B., Rhoades, D., Smith, W., Nicol, A., Pettinga, J., Clark, K., and Jacobs, K. 2012. National Seismic Hazard Model for New Zealand: 2010 Update. Bulletin of the Seismological Society of America, 102(4), 1514–1542, doi: 10.1785/0120110170

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

Werner, M.J., Zechar. J.D., Marzocchi, W. and Wiemer, S. 2010. Retrospective evaluation of the five-year and ten-year CSEP-Italy earthquake forecasts. Annals of Geophysics, 53(3), 11-30, doi:10.4401/ag-4840.

Zechar, J.D.; Gerstenberger, M.C.; Rhoades, D.A. 2010a. Likelihood-based tests for evaluating space-rate-magnitude earthquake forecasts. Bulletin of the Seismological Society of America, 100(3): 1184-1195; doi: 10.1785/0120090192

Zechar, J.D.; Schorlemmer, D.; Liukis, M.; Yu, J.; Euchner, F.; Maechling, P.J.; Jordan, T.H. 2010b. The Collaboratory for the Study of Earthquake Predictability perspective on computational earthquake science. Concurrency and Computation: Practice and Experience, 22(12): 1836-1847.