

The Development of Earthquake Probability Forecasts from a Hidden Markov Model: An Example from the Killini Region, Greece

By

Katerina Orfanogiannaki (National Observatory of Athens),

David Vere-Jones (Victoria University of Wellington and Statistics
Research Associates),

and David Harte (Statistics Research Associates)

Abstract

A four-state hidden Markov model is used to develop 1-day probability forecasts for earthquakes in the Killini region of Greece. The model allows the rates in each state, and the transition probabilities between states, to be estimated from the earthquake occurrence data. Forecasts proceed by using the model to estimate the current state probabilities, and using these to estimate the forecast probabilities for the next 1-day period. The data used is provided by the National Observatory of Athens, and comprises all events over local magnitude 3.2 in the Killini seismic zone. The forecasts are prepared initially for the region as a whole, and then partitioned out over a spatial grid covering the observation region, and into magnitude classes. The final results therefore consist of daily forecasts for each grid in the spatial region and for each magnitude class. The forecasts from the hidden Markov model are compared to forecasts obtained from applying a temporal ETAS (Epidemic Type Aftershock Sequence) model to the same data and region. The hidden Markov model appears to have advantages when the data is clustered, and the clusters only partially follow the traditional main-shock, aftershock pattern.

1 Introduction

The development of earthquake forecasts on an on-going basis is a long-term project. It involves considerable resources for measurement and monitoring of the underlying physical processes, researches into the character and patterns of earthquake occurrence, and, finally, the development and testing of algorithms for earthquake probability forecasts. At the present stage in the evolution of this process, two overlapping needs are for the development of suitable models to produce forecasts from available data, and the adaption of a broadly standard forecast framework to allow different forecasting procedures to be assessed and compared.

The present paper is intended to illustrate a simple candidate model for producing earthquake probability forecasts, and to outline the steps needed to bring this model to the stage where the forecasts are available in a format similar to that being used in earthquake forecast testing centres such as the RELM centre in California (see Schorlemmer and Gerstenberger 2007b).

The model used here is simplistic in several respects, and we would not like to give the impression that forecasts from these procedures would be suitable for use in publicly released statements. On the other hand the results may be comparable with forecasts from other initial models currently being considered for the European and other testing centres, and illustrates issues which are likely to arise in most candidate procedures.

Two issues in particular are addressed in this study: the effectiveness of simple hidden Markov models as a possible class of candidate models for daily earthquake forecasts; and the problems inherent in extending forecasts from a regional character to a more detailed character in terms of space, time and magnitude requirements.

The study builds on previous work by Orfanogiannaki, Karlis and Papadopoulos (2008), denoted by OKP in the sequel.

2 The hidden Markov methodology

Hidden Markov models (HMMs) represent a flexible class of models that have been used with considerable success in a range of different application areas, such as voice recognition technology, analysis of IT traffic, mathematical genomics. Their key assumption is that there is an underlying Markov process (past history influences future development only through the current state), say $\{X_n\}$, which is not observable as such, but which influences the distributions of the quantities which can be observed. The HMM methodology allows characteristics of the hidden process, such as the transition probabilities from one state to another, to be estimated from the directly observable data. Accounts of the methodology are available for example in Macdonald and Zucchini (1997); see also the notes on associated R procedures in Harte (2008(a) (b)), and the discussion in OKP. Further applications to the earthquake context are in Granat (2002), Ebel et al (2007).

In the earthquake context the hidden process might represent the state of a regional stress field, or in the case of deep or volcanic earthquakes, the temperature or pressure fields at critical locations and depths. The directly observable data might consist of the occurrence times, locations and magnitudes of earthquakes recorded in a regional earthquake catalogue, possibly coupled with ancillary data such as GPS strain measurements.

The underlying model supposes that the hidden regime can exist in a small, finite number of states, here $k = 1, 2, 3, \dots$ and that each state is characterized by an underlying Poisson process rate, say λ_k , when the process is in state k . Progress of the hidden process X_n from state to state is controlled by a transition matrix T , where

$$T_{ij} = \Pr\{X_{n+1} = j \mid X_n = i\}.$$

Again for simplicity, we suppose here that the process evolves in discrete time (day by day), so that p_{ij} represents the probability that, given the hidden process is in state i today, it will be in state j tomorrow. λ_k represents the expected number of events that will occur during the day, given that the underlying process is in state k . Observations (for the full Killini region) comprise the numbers, u_n , of events occurring during successive days $n = 1, 2, \dots, N$.

If the state history is known (which in reality it is not), it is easy to write down the *complete likelihood* for both the set of observations u_n , and the sequence of states X_n . Several approaches are then available to obtain the actual likelihood, which is just the complete likelihood averaged over all possible state histories. The *E-M* algorithm, which was used in OKP, iteratively calculates sets of “forward, $\alpha_n(k)$, and backwards, $\beta_n(k)$, probabilities”

$$\alpha_n(k) = \Pr\{X_n = k \text{ and } U_m = u_m, m = 1, 2, \dots, n\},$$

$$\beta_n(k) = \Pr\{X_n = k \text{ and } U_m = u_m, m = n + 1, n + 2, \dots, N\},$$

based on a current set of parameter values λ_k, p_{kj} . The likelihood, assuming this set of parameter values, is then given, for any n , by

$$L = \sum_k \alpha_n(k) \beta_n(k) = \sum_k \alpha_N(k).$$

Expectation (E-) and *maximization (M-)* steps are then used to produce an updated set of parameter values, the process being repeated until the likelihood stabilizes at its maximum.

An alternative method is based on a matrix representation of the likelihood. We shall not give further details of the algorithms here, but refer to the references already quoted above.

Both methods yield formulae for parameter estimates. An important point for the sequel is that the forward and backward probabilities may also be used to produce estimates of the state probabilities

$$\pi_m(j) = \Pr\{X_m = j \mid u_1, \dots, u_N\} = \frac{\alpha_m(j) \beta_m(j)}{\sum_k \alpha_m(k) \beta_m(k)} \quad (1)$$

for both current and past times. Such estimates may be used to gain an approximate representation of the state history $X(t)$. They also provide the basis for calculating probability forecasts based on the model, as we briefly consider.

Given data up to time point N , the probability $\hat{\phi}_{N+1}$ that an event will occur in day $N + 1$ can be found from averaging the probabilities that an event will occur when the system is in state k over the probabilities that the system is in state k in day $N + 1$. Denoting the probability that an event will occur during the day when the system is in state k by

$$\hat{p}_k = 1 - e^{-\lambda_k},$$

we have

$$\begin{aligned} \hat{\phi}_{N+1} &= \sum_k \pi_{N+1}(k) \hat{p}(k) \\ &= \sum_j \sum_k \pi_N(j) T_{jk} \hat{p}(k), \end{aligned}$$

the probabilities $\pi_N(j)$ being found from (1) with $m = N$.

3 Data and fitted results for the Killini region

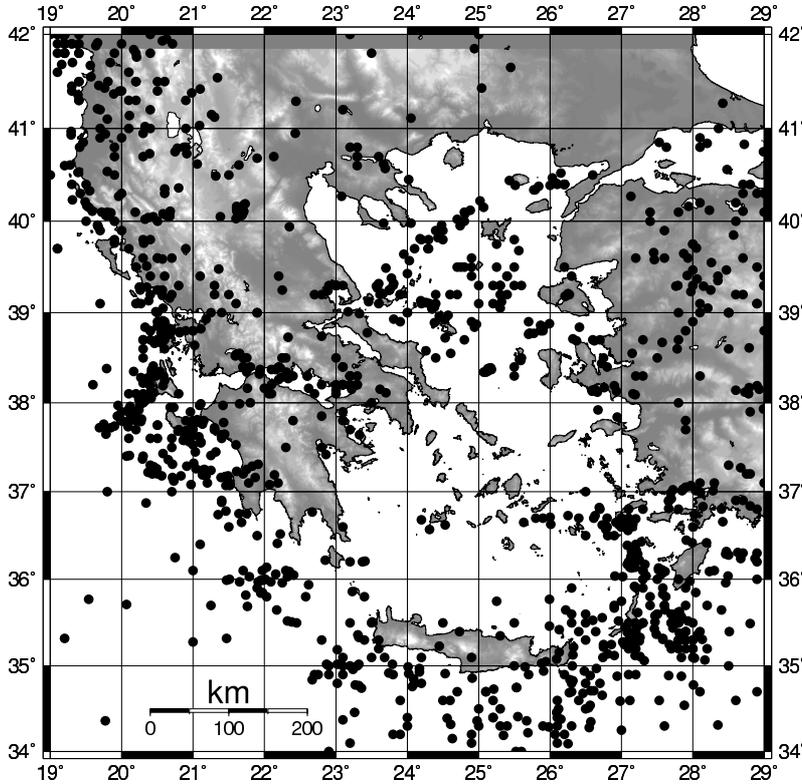


Figure 1: Greek Earthquakes, $M \geq 5$, 1990-2004

Figure 1 shows the general seismicity rate for Greece for the period of the study, 1990-2004. Here the data are restricted to events with $M_L \geq 5$. Note that the magnitudes here are the local magnitudes determined by the National Observatory of Athens, Institute of Geodynamics, Greece (<http://www.gein.noa.gr>).

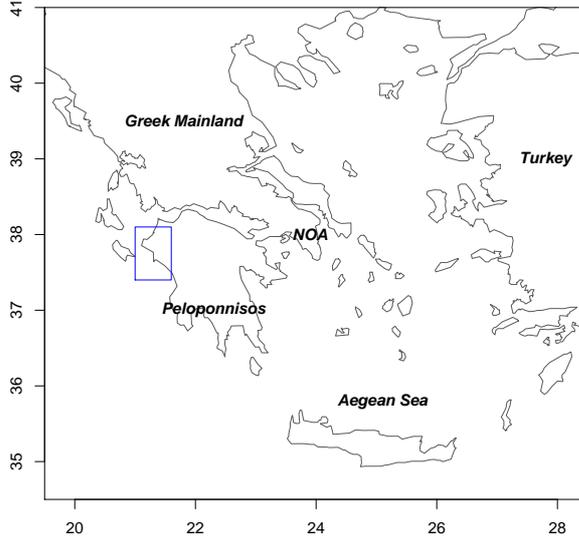


Figure 2: Overall map of Greece; rectangle denotes the Killini region.

The basic data used in the present study are the same as those used in OKP (2008). They are restricted to shallow events (depth $\leq 60km$) having magnitudes $M_L \geq 3.2$ from the Killini area in Greece. Figure 2 shows the Killini area, as defined by the grid points used in the forecasts. The actual region used in the original study is somewhat larger and more irregular than that suggested by the rectangle (see Fig 1 of OKP). The epicentral map for the events in the Killini region that were used in the present analysis is shown in Figure 3. The different circle sizes represent different magnitude classes. The 4 magnitude classes that appear on the map from smaller to bigger circle size are: $[3.2,4.0)$, $[4.0,4.5)$, $[4.5,5.0)$, $[5.0,5.5]$.

Monthly counts for events from the Killini region with magnitudes $M_L \geq 3.2$ are shown in Figure 4. The average occurrence rate over this period for events with $M_L \geq 3.2$ is 0.1474 events/day, or about 50 events/year.

Several HMM models were tried for this data, with varying numbers of states. The HMM model selected (using the Akaike Information Criterion (AIC); see OKP for further details) had 4 states, with vector of intensity rates

$$\lambda = (0.0595 \quad 0.2319 \quad 1.7838 \quad 11.5862)$$

The mean interval lengths vary from 16 days in State 1 to just over 2 hours in the most active state, State 4.

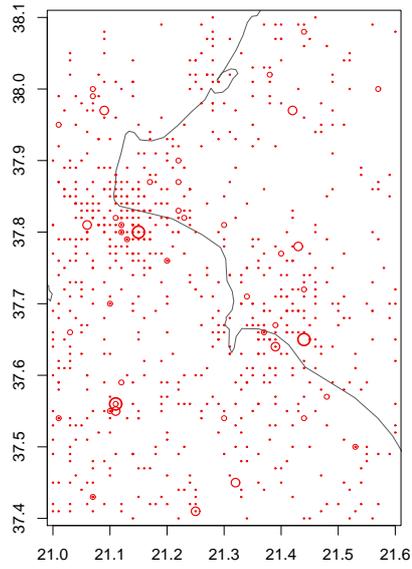


Figure 3: Epicentral map for the data in the Killini region

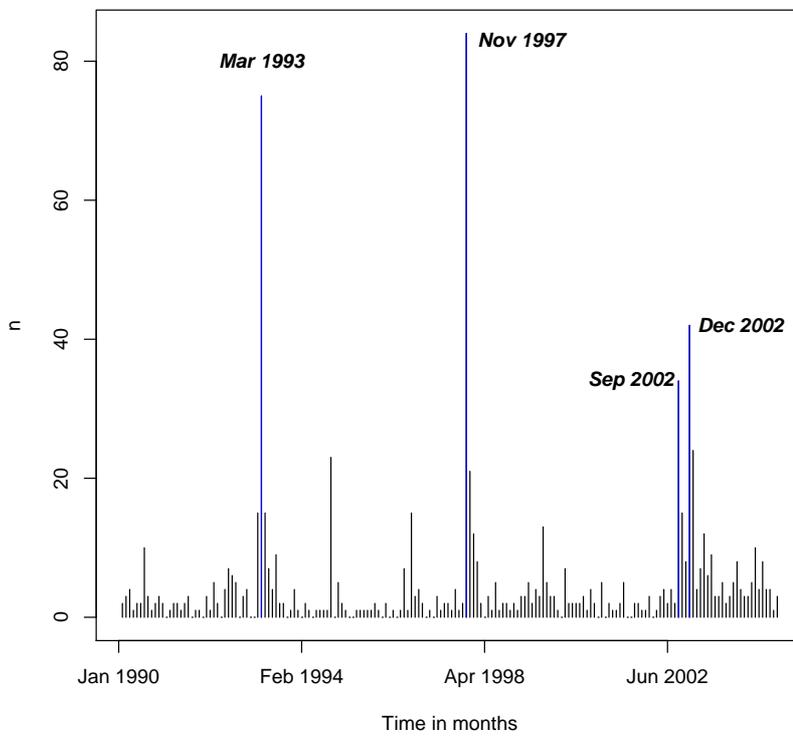


Figure 4: Monthly Count Data for Killini Region, $\text{Mag} \geq 3.2$, 1990-2004

The estimated matrix of 1-day transition probabilities T is

	[, 1]	[, 2]	[, 3]	[, 4]
[1,]	0.9953	0.0002	0.0045	0.0000
[2,]	0.0290	0.9210	0.0477	0.0023
[3,]	0.0000	0.4780	0.4990	0.0230
[4,]	0.0000	0.0000	0.3600	0.6400

The elements in the first row represent the probabilities, given that the process starts from state 1, of remaining in state 1 the next day, or transferring to one of the states 2,3,4. The entries in each row sum to unity. The high value for $T_{1,1}$ implies that the process has a high probability of remaining in that state. The sojourn times are geometrically distributed with means determined by the diagonal entries. The mean sojourn times (in days) $\bar{\tau}_j$ in the different states are given by

$$\bar{\tau}_1 = 212.765957 \quad \bar{\tau}_2 = 12.658228 \quad \bar{\tau}_3 = 1.996008 \quad \bar{\tau}_4 = 2.777778$$

The average numbers of events occurring while the system remains in a given state are given by

$$(12.659574 \quad 2.935443 \quad 3.560479 \quad 32.183889)$$

We see that the majority of events are generated either when the process is in a quiet state (for a long time) or in the very active state.

After many days, the n -day transition matrix T^n converges towards the matrix of stationary probabilities. This matrix is given by

	[, 1]	[, 2]	[, 3]	[, 4]
[1,]	0.8395	0.1361	0.0221	0.0023
[2,]	0.8395	0.1361	0.0221	0.0023
[3,]	0.8395	0.1361	0.0221	0.0023
[4,]	0.8395	0.1361	0.0221	0.0023

Note that the entries are independent of the initial state: after a long time, memory of the initial state gets lost. The entries represent the stationary probability distribution for the chain,

$$\mathbf{p}_\infty = (.8395, .1361, .0221, .0023).$$

The average daily rate, when the process is in the stationary regime, is

$$(\mathbf{p}_\infty)^T \boldsymbol{\lambda} = 0.1474,$$

in keeping with our overall estimate.

4 1-day forecasts from the HMM model

Producing forecasts from such models involves first finding the forecasts conditional on the process being in a given state, and then averaging these over the current state probabilities. We consider first the forecasts for the Killini region as a whole.

As a simple but key example, consider the probability that no event occurs during the coming day. Interpreting the vector of rates as the parameters in a set of Poisson distributions, the 1-day no-event probabilities for the different states are given by the values $e^{-\lambda_k}$, namely the vector

$$\mathbf{p}_f = (0.9422 \quad 0.7930 \quad 0.1680 \quad 0.00000929).$$

We suppose that forecasts are given at midnight, at which time the probabilities of the process being in any given state for the day following are known. For a particular example, suppose that the current state probabilities are those for the stationary distribution. Then the no-event probability for the coming day is given by the inner product

$$p_0(1) = (\mathbf{p}_\infty)^T \mathbf{p}_f = 0.9026,$$

which is the stationary value of the no-event probability over a single day.

The no-event probabilities for the next day are given by

$$(\mathbf{p}_\infty)^T T \mathbf{p}_f$$

which has the same value 0.9026 since the distribution \mathbf{p}_∞ is stationary, invariant under T . To find the stationary value of the probability $p_0(2)$ that there are no events either today or tomorrow, write D for the diagonal matrix with entries given by the elements of \mathbf{p}_f . Then we have

$$p_0(2) = (\mathbf{p}_\infty)^T D T \mathbf{p}_f = 0.8268948,$$

Similarly, the stationary value of the probability that there are no events for a full week is given by

$$p_0(7) = (\mathbf{p}_\infty)^T (D T)^6 \mathbf{p}_f = 0.5647341.$$

The probability that at least one event occurs during the current day is $1 - p_0(1) = 0.0974$ and the probability that there is at least one event during the coming week is $1 - p_0(7) = 0.4353$. Probabilities of other combinations of events and no-events can be calculated in similar ways.

Replacing the stationary distribution \mathbf{p}_∞ by the current set of state probabilities, we obtain the probability forecast for no events during the day, given the state probabilities just after midnight. The next day, the EM algorithm can be used to update the state probabilities to include the most recent information, and the process repeated. Thus, starting from some initial set of probabilities, we can systematically build up a continuing sequence of forecasts.

For many purposes it may be more convenient to replace the no-event forecasts or their complement, by forecasts of the expected number of events. Much as in

computing the no-event probabilities, we take the rate vector $\boldsymbol{\lambda}$, and average these by the current state probabilities. In the stationary regime, the expected number of events for the current day is $m_1 = 0.1474$. The expected number of events in a week, starting from an initial probability vector \mathbf{p}^* , is the sum

$$m_7 = \sum_{r=1}^7 (\mathbf{p}^*)^T T^r \boldsymbol{\lambda}.$$

Again, starting from some set of initial probabilities, we can build up a continuing sequence of forecasts for the expected numbers of events by updating the state probabilities.

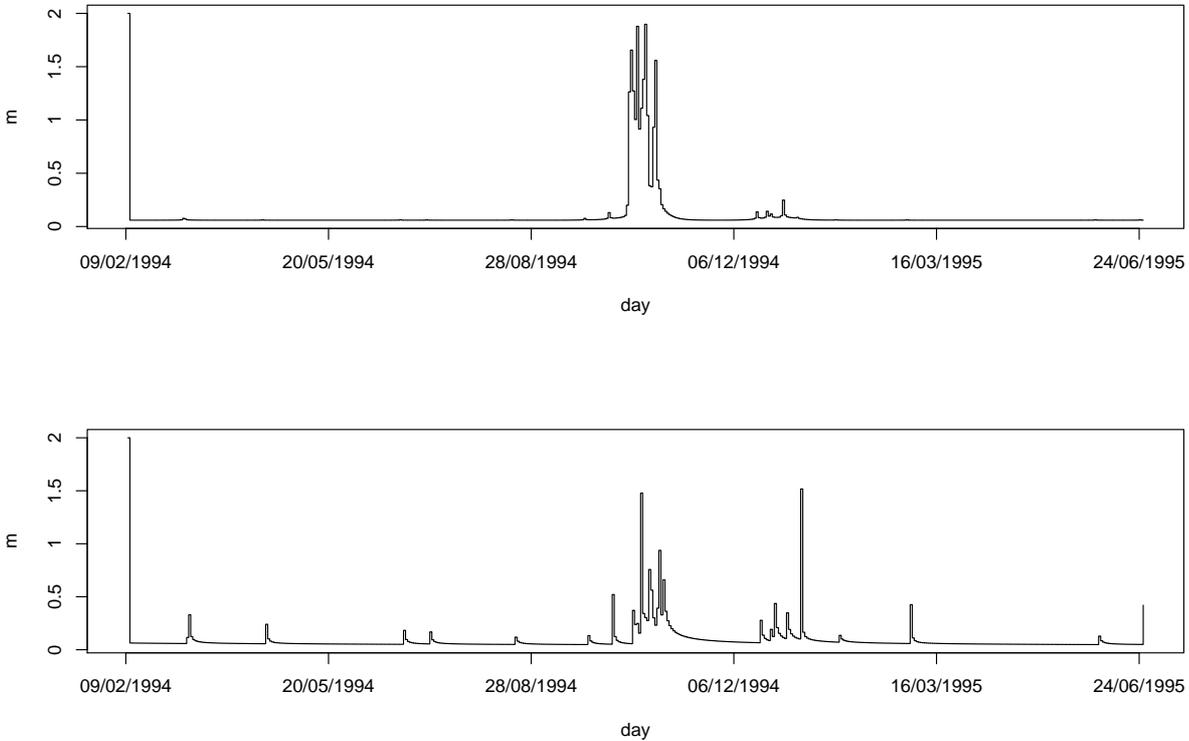


Figure 5: HMM (top) and ETAS (bottom) daily expected numbers for Killini region, Jan 1994-June 1995

HMM forecasts for the expected numbers of events have been built up in this way for the entire observation period. Some portion of these, namely for the time period from 09/02/1994 to 24/06/1995 are displayed in the upper part of Figure 5.

A shorter set of day-by-day forecasts, for the week that immediately follows the observation period, are listed in Table 1. The initial set of state probabilities that were used to produce the forecasts are:

$$(9.950675 \cdot 10^{-01} \quad 4.467819 \cdot 10^{-03} \quad 4.647253 \cdot 10^{-04} \quad 2.389992 \cdot 10^{-11})$$

The upper row gives the expected numbers of events; the lower row gives the non-event probability.

Table 1: Daily forecasts of probabilities and expected numbers of events for the Killini region, 01/01/2005 – 07/01/2005.

	01/01/05	02/01/05	03/01/05	04/01/05	05/01/05	06/01/05	07/01/05
m	0.937475	0.9353893	0.9338857	0.9326691	0.9316067	0.9306363	0.9297275
p_0	0.06962336	0.07505072	0.07915733	0.08245557	0.08523447	0.08766497	0.0898522

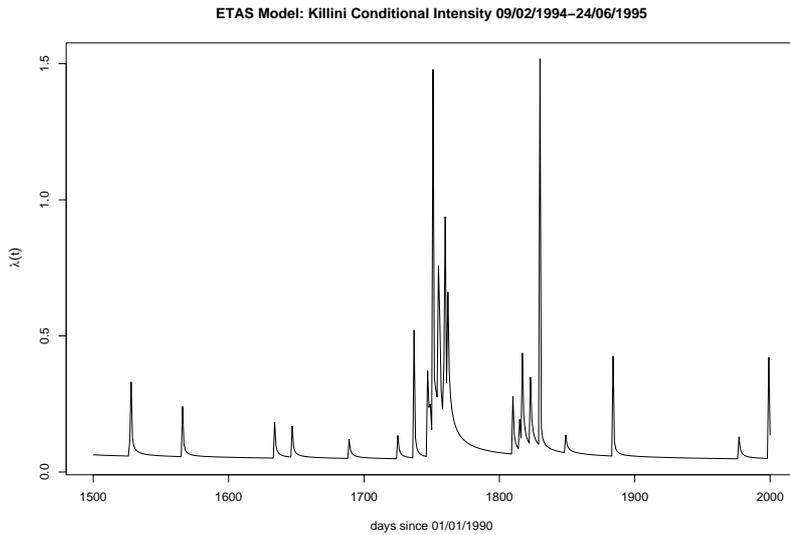


Figure 6: ETAS Intensity for Killini data, Jan 1994-June 1995

5 Comparison with the ETAS model

The lower plot in Figure 5 is included by way of comparison. It was obtained from fitting an ETAS model to the same data and using the fitted ETAS model to produce the forecasts. The ETAS (Epidemic Type Aftershock Sequence) model is a continuous-time point process model with conditional intensity (rate conditioned by the past history)

$$\lambda^*(t | \mathcal{H}_t) = \mu + \sum_{i:t_i < t} A e^{\alpha(M_i - M_c)} f(t - t_i). \quad (2)$$

This represents the current value of the intensity as the sum of a background intensity μ and contributions from previous events, weighted in accordance with their magnitudes M_i , and decaying in time after their occurrences by a power law probability density

$$f(x) = (p/c)/(1 + x/c)^{1+p}.$$

The quantities μ , A , α , p , c are parameters of the model. Many further details about the model, likelihood estimation methods, etc, can be found, for example, in Ogata (1988), or Daley and Vere-Jones (2003, Chapter 6).

To produce 1-day forecasts from this model, it was first fitted to the Killini data, taking $M_c = 3.2$, and giving the parameter values

$$\mu = 0.029545, A = 1.35743, \alpha = 1.54556, p = 1.11751, c = 0.05886632.$$

Then the expected number of events for a given day was first estimated by taking

$$m_1(n) = \int_n^{n+1} \lambda^*(u) du,$$

where we interpret the time n as midnight before the day for which the forecasts are required, and $n + 1$ as midnight the following day. These are the quantities shown in the lower panel of Figure 5. The conditional intensity itself is shown in Figure 6.

An interesting feature is that during the active periods, the HMM model seems to produce better forecasts than the ETAS model. This is shown more clearly in Figure 7, which shows the log probability gain for the HMM as compared to the ETAS model. By the probability gain we mean the ratio of the probabilities of the actual outcome as forecast by the two models. For example, if the actual outcome on a given day is that no event occurs, and the HMM probability forecast of no event is $p_0(HMM)$, and the ETAS probability forecast of no event is $p_0(ETAS)$ the log probability gain for that day is

$$\log \left[p_0(HMM) / p_0(ETAS) \right]$$

Successive ratios of this general type are shown in Figure 7.

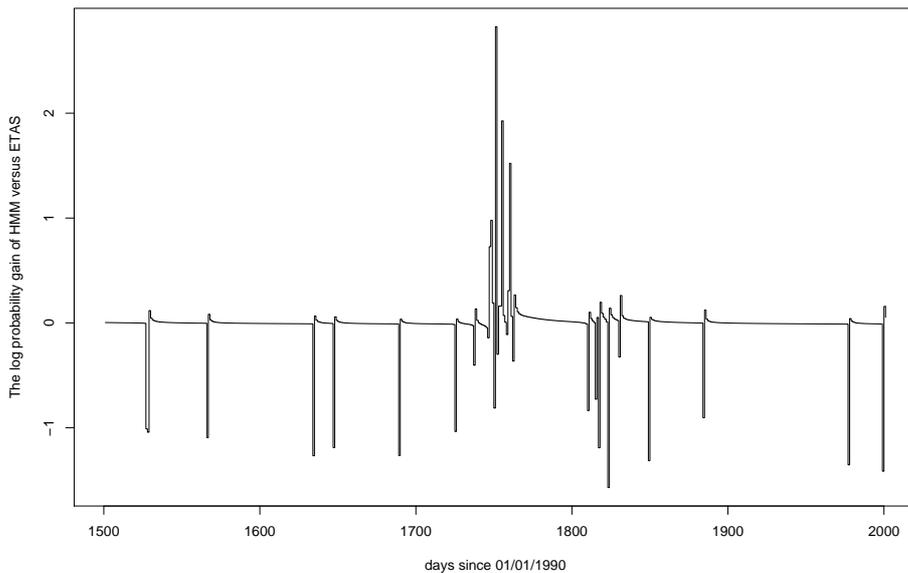


Figure 7: Time variations of the log probability gain of HMM versus ETAS

The good performance of the HMM model is the more surprising in that the method we have used for obtaining the forecast probability for the ETAS model offers an advantage to the ETAS model insofar as, being a continuous time model, it can incorporate into its forecasts information arising during the current day. In particular if an event occurs during the day, it will cause an immediate increase in the conditional intensity, which then give a high probability for short-term aftershocks occurring during the same day. No such possibility of incorporating information during the day is available to the HMM model. In such a situation we would expect the ETAS model to outperform the HMM model and this is shown clearly in the short-term downwards spikes in Figure 7. The surprising fact is that no similar advantage appears for the major sequence indicated by the increased intensity in the center of Figure 6.

An alternative method for producing the ETAS forecasts is to base the forecasts only on the information available when the forecast is made, that is to say restricting the summation in (2) to events before midnight on the day in question while still integrating over the 24-hour period. This eliminates the advantage to the ETAS model, but replaces it with an even more serious disadvantage, as the continuous time version used here optimizes the parameters for the short-term, continuous-time situation, and a different optimization procedure would have to be used to find the best ETAS model for such 1-day forecasts. This reversal of circumstances is shown in Figure 8. The short-term advantages of the ETAS model have been lost, and its disadvantages greatly emphasized.

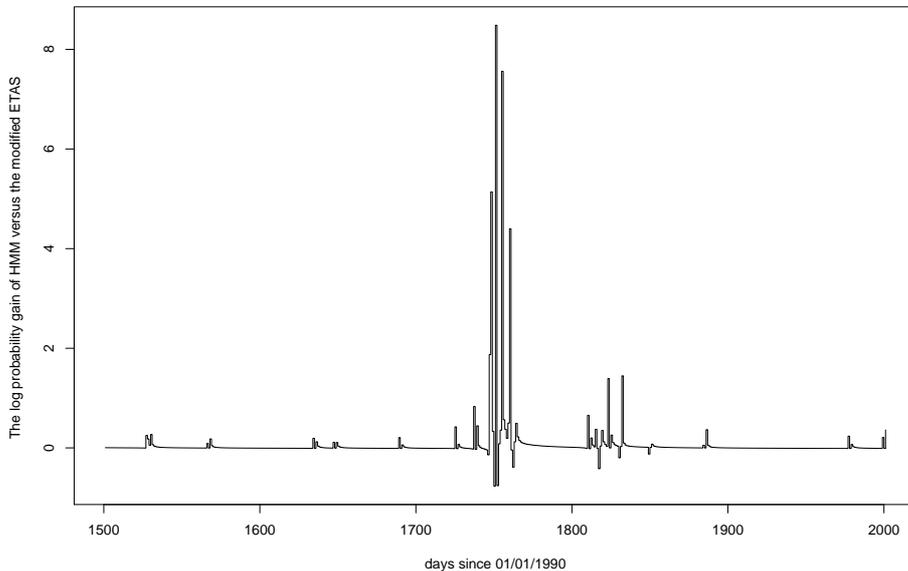


Figure 8: Time variations of the log probability gain of HMM versus the modified-ETAS

Overall, the relative performance of the two models presumably lies between the two extremes illustrated in the Figures 7 and 8. At the very least, and for this data in particular, the HMM model shows itself to be a candidate worth further exploration.

There are other factors in the data which may also be at play here. Not all the

high activity periods in the Killini data are associated with traditional aftershock sequences. There are several examples where the largest event does not occur at or near the beginning of the sequence. The sequence shown in the middle of Figures 7 and 8 is of such non-standard type. The ETAS model is not designed to fit such variable clustering patterns. The best method for producing forecasts for such non-standard or mixed systems is not clear, and requires further investigation. All we can say here is that the HMM model seems to handle such non-standard sequences better than the ETAS model.

6 Forecasts by grid and magnitude cells

To adopt the forecasts to the format similar to that required for the forecast testing centres in California or New Zealand (see for example Schorlemmer et al 2007a, 2007b), spatial locations and magnitudes have to be figured into the forecasts. The only simple way of doing this with the current model is to assume that the space-time-magnitude intensity can be represented as a product of time, space and magnitude factors.

Let i, j denote the coordinates of the grid cell, and M_k a half-unit magnitude cell (e.g. $5.5 < M \leq 6$). We then assume that we can write

$$\lambda(n, i, j, k) = \lambda^*(n)f(i, j)g(M),$$

where $\lambda^*(n)$ is the overall forecast rate (expected number of events) for day n , as described in the previous section, and $\{f(i, j)\}$ and $\{g(M)\}$ are probability distributions over the grid cells and the magnitude cells respectively. In each case we assume that f, g are independent of time and hence of the state of the hidden Markov chain.

This spatial information is summarized by the average values at the points of the Killini spatial grid, shown in Figure 9. A smoothed version of the spatial distribution over the period of the study is shown in Figure 10 for the region around and including Killini. The spatial probabilities $f_{i,j}$ are condensed from this smoothed seismicity map, and summarized as a set of proportions $\{f_{ij}\}$ in Table 2.

Table 2: Spatial Proportions f_{ij}

	37.4	37.5	37.6	37.7	37.8	37.9	38.0	38.1
21.0	0.0235	0.0257	0.0310	0.0355	0.0373	0.0292	0.0096	0.0022
21.1	0.0201	0.0314	0.0295	0.0503	0.0688	0.0476	0.0157	0.0031
21.2	0.0148	0.0221	0.0173	0.0347	0.0479	0.0274	0.0090	0.0045
21.3	0.0113	0.0170	0.0189	0.0372	0.0280	0.0105	0.0078	0.0061
21.4	0.0043	0.0112	0.0252	0.0413	0.0232	0.0091	0.0081	0.0050
21.5	0.0038	0.0088	0.0167	0.0162	0.0113	0.0060	0.0065	0.0044
21.6	0.0022	0.0036	0.0029	0.0026	0.0034	0.0034	0.0036	0.0021

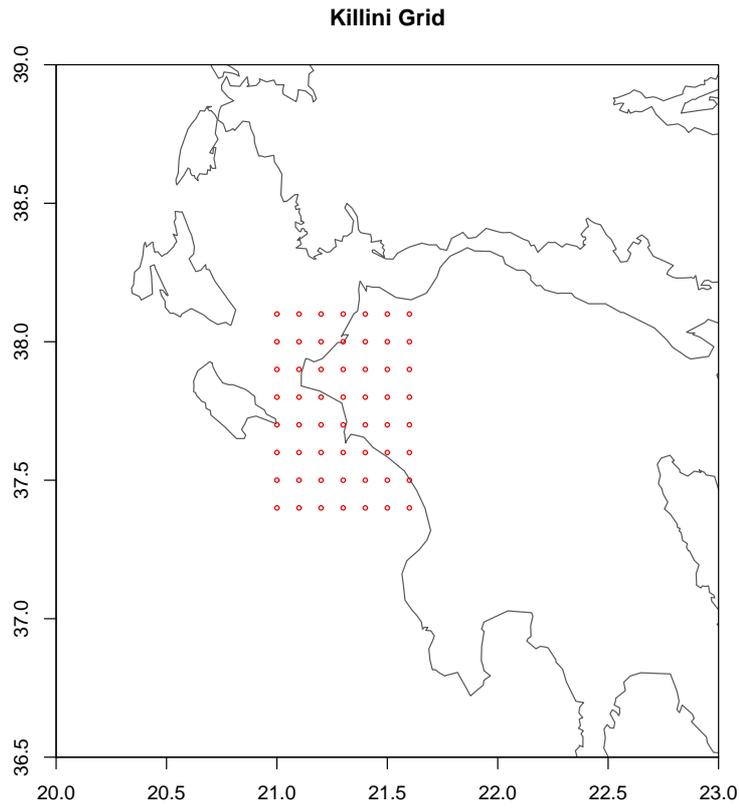


Figure 9: Forecast Grid for Killini Region

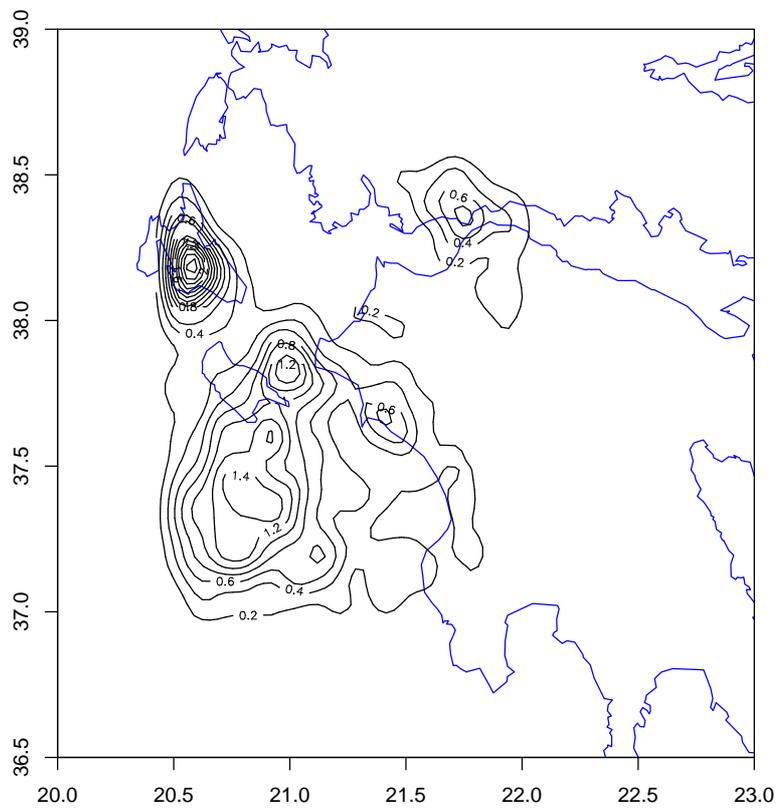


Figure 10: Smoothed seismicity map for Killini region

Examples of the daily forecasts, refined down to specific spatial and grid cells, are included in the next two tables. The second of these illustrates also the further effect of cutting down the forecasts to specific magnitude cells.

Table 3: 1-day probability forecasts for the space grid cells defined in Figure 9

	37.4	37.5	37.6	37.7	37.8	37.9	38.0	38.1
21.0	0.0220	0.0241	0.0291	0.0333	0.0349	0.0274	0.0090	0.0020
21.1	0.0189	0.0295	0.0276	0.0472	0.0645	0.0446	0.0147	0.0029
21.2	0.0138	0.0208	0.0162	0.0325	0.0449	0.0257	0.0084	0.0042
21.3	0.0106	0.0159	0.0177	0.0349	0.0262	0.0098	0.0073	0.0058
21.4	0.0040	0.0105	0.0236	0.0387	0.0217	0.0086	0.0075	0.0047
21.5	0.0035	0.0082	0.0156	0.0152	0.0106	0.0056	0.0061	0.0041
21.6	0.0020	0.0034	0.0027	0.0024	0.0032	0.0032	0.0034	0.0020

Table 4: Forecasting rates for 1 day ahead at each grid point for the magnitude range: $Pr(4.0 < M \leq 4.5)$

	37.4	37.5	37.6	37.7	37.8	37.9	38.0	38.1
21.0	$2.06 \cdot 10^{-6}$	$2.26 \cdot 10^{-6}$	$2.73 \cdot 10^{-6}$	$3.12 \cdot 10^{-6}$	$3.27 \cdot 10^{-6}$	$2.56 \cdot 10^{-6}$	$8.41 \cdot 10^{-7}$	$1.89 \cdot 10^{-7}$
21.1	$1.77 \cdot 10^{-6}$	$2.76 \cdot 10^{-6}$	$2.59 \cdot 10^{-6}$	$4.42 \cdot 10^{-6}$	$6.04 \cdot 10^{-6}$	$4.18 \cdot 10^{-6}$	$1.38 \cdot 10^{-6}$	$2.74 \cdot 10^{-7}$
21.2	$1.30 \cdot 10^{-6}$	$1.94 \cdot 10^{-6}$	$1.52 \cdot 10^{-6}$	$3.04 \cdot 10^{-6}$	$4.21 \cdot 10^{-6}$	$2.41 \cdot 10^{-6}$	$7.91 \cdot 10^{-7}$	$3.96 \cdot 10^{-7}$
21.3	$9.96 \cdot 10^{-7}$	$1.49 \cdot 10^{-6}$	$1.66 \cdot 10^{-6}$	$3.27 \cdot 10^{-6}$	$2.46 \cdot 10^{-6}$	$9.18 \cdot 10^{-7}$	$6.83 \cdot 10^{-7}$	$5.39 \cdot 10^{-7}$
21.4	$3.73 \cdot 10^{-7}$	$9.87 \cdot 10^{-7}$	$2.21 \cdot 10^{-6}$	$3.62 \cdot 10^{-6}$	$2.04 \cdot 10^{-6}$	$8.01 \cdot 10^{-7}$	$7.07 \cdot 10^{-7}$	$4.36 \cdot 10^{-7}$
21.5	$3.32 \cdot 10^{-7}$	$7.69 \cdot 10^{-7}$	$1.46 \cdot 10^{-6}$	$1.42 \cdot 10^{-6}$	$9.95 \cdot 10^{-7}$	$5.26 \cdot 10^{-7}$	$5.72 \cdot 10^{-7}$	$3.86 \cdot 10^{-7}$
21.6	$1.91 \cdot 10^{-7}$	$3.20 \cdot 10^{-7}$	$2.54 \cdot 10^{-7}$	$2.29 \cdot 10^{-7}$	$2.97 \cdot 10^{-7}$	$3.02 \cdot 10^{-7}$	$3.17 \cdot 10^{-7}$	$1.83 \cdot 10^{-7}$

7 Concluding remarks

The study presented here was prepared rather as a case study than to claim especial advantages for the HMM procedures, particularly in the elementary form presented here. Nevertheless the following points may be made.

- The hidden Markov model is based on well-established and flexible procedures, including software and software packages.
- It is easy to use in predictive mode, and as the basis for regular predictions.
- Although the particular model illustrated here is limited, extensions to the basic form to include joint space-time and space-time-magnitude dependencies, for example, are the subject of further research.
- The comparisons with the ETAS model raise issues about the scope of the ETAS model and its effectiveness as a forecasting tool. The character of the seismicity and the use of continuous-time models to produce regular, discrete-time forecasts are among these. Further research is needed to clarify these points.
- After this study was largely completed, we became aware of the work by Ebel et al (2007), in which a similar hidden Markov model is used to produce daily forecasts for events in the California region. The main difference is that Ebel et al apply the model to a declustered data set, whereas our study includes the clustered data as a major element of the data and the model.

Acknowledgements

The work undertaken in this study was supported by the New Zealand Earthquake Commission, under EQC Grant 6TVH/1. International travel for the first author was covered by the National Observatory of Athens. During her stay in New Zealand she was given office and computing support as a visitor to the School of Mathematics, Statistics and Computer Science, Victoria University of Wellington. Additional support was provided by Statistics Research Associates. The assistance of these organizations is gratefully acknowledged.

References

1. Daley, D. J. and Vere-Jones, D. (2003). An Introduction to Theory of Point Processes: Elementary Theory and Methods, Volume I, 2nd edition, Springer, (New York).
2. Ebel, J.E., Chambers, D.W., Kafka, A.L. and Baglivo, J.A. (2007). Non-Poissonian earthquake clustering and the hidden Markov model as bases for

- earthquake forecasting in California, *Seismological Research Letters*, **78**(1), 57-65.
3. Granat, R. and Donnelan, A. (2002). A hidden Markov model-based tool for geophysical data exploration, *Pure and Applied Geophysics*, **159**, 2271-2283.
 4. Harte, D.S. (2008). Users Guide for the Statistical Seismology Library, Statistics Research Associates, (Wellington), (<http://www.statsresearch.co.nz/software.html>).
 5. MacDonald, I.L. and Zucchini, W. (1997). Hidden Markov models and other models for discrete-valued time series, Chapman and Hall, (London).
 6. Ogata, Y. (1988). Statistical models for earthquake occurrence and residual analysis for point process, *Journal of the American Statistical Association*, **83**, 9-27.
 7. Orfanogiannaki, K., Karlis, D. and Papadopoulos, G.A. (2008). Identification of seismic patterns using hidden Markov models, Technical report 237, July 2008, Department of Statistics, Athens University of Economics and Business.
 8. Schorlemmer D. and Gerstenberger, M.C. (2007b). RELM Testing Center, *Seismological Research Letters*, **78**, 30-36.
 9. Schorlemmer, D., Gerstenberger, M.C., Wiemer, S., Jackson, D.D. and Rhoades, D.A. (2007a). Earthquake Likelihood Model Testing, *Seismological Research Letters*, **78**, 17-29.