Predicting New Zealand Earthquakes Using the 'M8' Algorithm

A Thesis presented

in partial fulfilment of the requirements

for the Degree of

Master of Applied Statistics

at Massey University

Palmerston North, New Zealand

Maaike Jane Vreede

December 1999

Abstract

The M8 algorithm uses seven time series to issue intermediate term earthquake predictions, stipulating which areas have an increased probability of a strong earthquake in the next five years. The series measure the frequency of earthquakes, change in frequency, energy release, and numbers of aftershocks. If six out of seven series (including series 7) cross the thresholds specified a 'Time of Increased Probability' is declared. This thesis takes the series as given and examines the role of each series. It also explores using alternative ways of summarising the predictive information in the series, using linear combinations of them rather than considering the number of series which cross certain thresholds.

It was found that the maxima of the series, rather than linear combinations of them, are related to future earthquakes. It was also found that the series have no predictive power unless they are all considered. Both of these findings are consistent with M8s own treatment of its series. The models using the M8 series that were constructed here can not be generalised to data they were not constructed with, so they cannot be used as a prediction tool. However, because of the scarcity of large earthquakes, earthquakes targeted for prediction have magnitudes less than the magnitudes that M8 is intended to predict. When data is available with enough target magnitudes that the algorithm is designed to predict the algorithm may yet be proven to be a successful tool for intermediate term earthquake prediction.

Acknowledgments

My thanks go to Greg Arnold, my long-suffering supervisor. He was always ready to do battle with red tape on my behalf, and he kept my thesis from straying down the wrong track. He has always been willing to give help, above and beyond the call of duty, and he helped me to keep things in perspective. Whenever my sanity was endangered, he was always on hand for a medicinal cup of coffee.

I wish to thank David Harte for taking the time to teach me about the curious phenomenon of earthquakes. His discussions with me were always interesting. I also thank him for giving me valuable tuition with the S+ programming language.

David Vere-Jones provided me with many ideas. His enthusiasm was infectious, and I was encouraged by his perseverance with difficult problems.

Mark Bebbington started me off on earthquake research. By his making certain that I got to meet the right people, I started off on the right footing.

Contents

Chapter 1 ~ Introduction

Several strong earthquakes have been recorded in New Zealand. The Hawkes Bay earthquake (magnitude 7.9 on the Richter scale) resulted in 256 deaths. The 1855 earthquake of South-West Wairarapa was in a less populated area (causing 5 deaths) but its magnitude was approximately 8.1, causing extensive faulting and coastal uplift. It was 'destructive' in Wellington (Eiby, p 150) – so if Wellington had been as populated as it is now many more lives would have been lost. Given that New Zealand as a whole is now much more populated, many lives could be saved if it were possible to predict earthquakes.

The M8 algorithm is a step in this direction. It aims at 'intermediate term' predictions, stipulating which areas are likely to have a strong earthquake in the next 5 years. The authors of the M8 algorithm consider earthquake prediction to be a process of narrowing down the time and place. First, only areas with a history of large earthquakes are considered. Then the area can be narrowed down further using M8, even further with another algorithm they have produced, and hopefully further still with methods for short term earthquake prediction that have yet to be developed.

M8 was developed by a group of Russian scientists. Their first article on intermediate term earthquake prediction appeared in 1980 (Keilis-Borok, Knopoff & Rotvain, 1980) and M8 was published in its final form in 1990 (Keilis-Borok & Kossobokov, 1990). They have made the algorithm available worldwide, with an implicit invitation to trial it on local earthquake data. It was first trialed in New Zealand by Li and Vere-Jones (Li and Vere-Jones, 1997), and the results were promising enough to warrant further investigation.

The algorithm uses frequency of earthquakes, change in frequency, energy release and aftershocks to predict future earthquakes, and these features are summarised in seven time series. This thesis takes these series as given and examines the role of each series, as well as alternative ways of summarising the predictive information in the series.

The authors of the M8 algorithm give very little motivation for their choice of series. Therefore a graphical analysis of New Zealand data is used to explore the role of each series, and their sensitivity to small changes in the position of the geographical area they are calculated over. The decision to declare a 'Time of Increased Probability' (or 'TIP') is based on the percentile of just one series at any given time. The graphical analysis investigated how important each series is – if a series is never, or hardly ever used in the decision to declare a TIP it could be omitted from the analysis. Important series could be studied in greater detail, and given more weight in predictions.

Reducing the number of series to be considered is important in view of the high degree of correlation between the first 6 series, which causes complications in linear models. Principal components analysis was also used to reduce the number of dimensions to be considered, and to investigate the inter-relationships between series.

The next approach to the series is to put them in to the framework of linear models, and this also sheds light on exactly how the series are related to future earthquakes. In these models the range of variables for which an attempt at prediction is made is expanded. The response variables include earthquake depth, the probability of an earthquake of a certain magnitude, and the expected magnitude of the largest earthquake in the next year.

The M8 algorithm is still in its development stage, and is currently being tested world-wide. This thesis will examine the working of the M8 algorithm, and attempt to extract more information out of its series.

Chapter 2 ~ Progress in the Task of Earthquake Prediction

2.1 Introduction

Earthquakes must certainly be one of the more difficult things to measure, and it is perhaps their rarity and uncontrollable nature that makes them an inviting challenge for statisticians. Things that are easily measured are, as a rule, less interesting. However, getting the measurements of earthquakes is the domain of seismologists, and statisticians wishing to study earthquakes rely very heavily on them. Eiby attributes the beginnings of seismology to John Michell (of the 18^{th} century) who realised that earthquakes are the result of elastic waves passing through the earth. It was his idea to trace the waves back to the origin, in order to find the cause of the disturbance. This led to a whole realm of discovery about the nature of the earth's interior, as well as that of earthquakes. However, the study of earthquakes goes right back to antiquity, and there may be some valuable insights underlying some of the tall tales.

2.2 Earthquake Folklore

Many strange and fantastic events are popularly supposed to preclude an earthquake. These include lights in the sky; purple, or 'blood red' rain; and strange animal behaviour, to name but a few. Recently, some of these events have been photographed, but many come to us only through eyewitness reports. Because of the trauma that the earthquake survivors have experienced, scientists are reticent about giving these reports much credence. Should the reports be accurate, the phenomena could have preceded the earthquake merely by chance. However, despite the controversy, there is an uncanny consistency in the reports of precursory phenomena. Close examination of historic and contemporary accounts of earthquakes also provides possible explanations for discrepancies. In all likelihood, there is some truth in earthquake folklore.

Unusual behaviour of animals, insects and fish before the onset of an earthquake is one of the most well known examples of earthquake folklore. Most animals panic. Snakes can come out of hibernation in the middle of winter, and many species simply leave the area. Lomnitz (1994, pp 130-133) downplays the value of animals responses before an earthquake, because of the problems involved with using them for prediction. Observer bias is a problem; once the earthquake has happened, everyone remembers strange behaviour in animals. Without doubt this bias does occur, but this does not preclude that animals do react strangely before an earthquake.

In some ways the 'quality of data' about anomalous animal behaviour is poor, since no systematic data is kept on their behaviour. However, reports of strange reactions of animals before an earthquake go back thousands of years, and in many different cultures. The expectation that animals will react before an earthquake could continue the 'myth' over thousands of years through observer bias, but for this to happen in unconnected cultures is unlikely.

The mechanism which causes animals to react the way the do is a bit of a mystery. This may explain some of the reluctance to take their reactions seriously. However, some plausible hypotheses have been put forward. Tributsch (1982, pp 108-121) speculates that positively charged ions emitted from the ground before an earthquake could explain the reactions of the animals. Electromagnetic effects have been observed during earthquakes, and sufficient exposure to positive ions would be sure to produce stress in animals with dry skin such as cats, birds, dogs and snakes. These happen to be the animals traditionally believed to predict earthquakes. Typical animal reactions such as snakes coming out of the ground, the flight of chickens into tree-tops, and the high circling of large flocks of other birds may be to escape the charge coming from the ground. The increased amount of suspended particles and/or aerosol particles indoors lead to a heightened electrical charge, which could actually be uncomfortable for some animals. This could explain why animals so fortuitously head outdoors before an earthquake. The animals could also be responding to ground deformation, which sometimes happens before an earthquake. Reactions just seconds before the earthquake could be due to sensing the earthquake's P-waves (sound waves) which arrive before the waves that set the ground in motion.

The best argument for taking animal warnings seriously is that they have been found very effective in China and Kamchatka. There are isolated incidents in which other countries have made use of animal precursors, but in China and Kamchatka people make a point of watching for signs from the animals. In Kamchatka, no bear is ever supposed to have come to grief through an eruption or earthquake. This peninsula is one of the worlds most active, volcanically, and 'the natives depend on the bears for timely information about an impending volcanic eruption' (Tributsch, p 64). Hundreds of people lost their lives during 1955 and 1956, but the bears came out of hibernation and escaped.

Humans can also be sensitive to an approaching earthquake. The nervous system goes into a state of extreme agitation. A writer in Copiapo, Chile, in 1882 describes his feelings before an earthquake:

> "Before we hear the sound, or at least are fully conscious of hearing it, we are made sensible, I do not know how, that something uncommon is going to happen; everything seems to change colour; our thoughts are chained immovably down; the whole world appears to be in disorder, all nature looks different to what it is wont to do; and we feel quite subdued and overwhelmed by some invisible power, beyond human control or apprehension."

An increase in the number of people suffering from nervous irritability, anxiety, breathing difficulties, restlessness, nausea, and inexplicable fear has been documented before several earthquakes (Tributsch, p 170). Strangely enough, some people are affected by contrasting symptoms: a general malaise, and closer to the quake, weakness in the legs and the perception that one's body is very heavy. People whose health has been undermined by illness or addiction are also more sensitive to an impending earthquake. Tributsch writes, 'the Japanese tradition… is haunted by reports of abnormal people predicting severe earthquakes…for example, the well known story of the man said to be possessed by the soul of a fox. On the day of the great Edo earthquake of 1855 this mentally ill man is supposed to have warned his neighbours of a severe earthquake that was to happen that evening, then he began to run back and forth' (p 171). Tributsch also tells the story of an alcoholic teacher who predicted a great disaster and the end of the world, 'around 6 on the evening before the earthquake. When the earthquake happened he jumped from the window and died'. Whether the predictions of the teacher and 'possessed' man from Edo were merely chance is hard to gauge. The reports of increases in certain types of illnesses, however, are harder to ignore. Also, the descriptions of panic are in terms of the bodily sensations, not with references to 'fear of open spaces' or any other type of panic. This would suggest that it is a type of panic specific to earthquakes. That people are experiencing this panic at the same time as animals also makes it more likely that the animal's panic is 'earthquake panic'.

Earthquake lights also cause considerable alarm, though not through mysterious causes. Beside their ghostly appearance, they have long been known as heralds of disaster. Shakespeare and other writers referred to them as 'strange eruptions', or 'portents in the heavens' (Lomnitz, p 137). However, it was only when they were photographed during the Matsushiro earthquake series in Japan (1965-67) that their existence was taken seriously (Tributsch, pp 141-141). They sometimes appear as 'hemispherical, glowing white apparitions that touch the ground, and that have diameters of 20-200 metres' (Tributsch, p 144). There are also balls of light in blue, red, white, and orange. During the Idu Peninsula earthquake (Japan, 1930) lights were seen that resembled the beams of searchlights (Tributsch, p 145). Occasionally they can occur a couple of days before the earthquake, indicating danger.

In earthquake prone areas, there is usually some popular conception of 'earthquake weather'. For example, in the Hawke's Bay, earthquake weather is described as clear and still, and oppressively hot (Eiby, p 90). Hot dry winds are also associated with earthquakes; they have 'a stimulating effect on the nerves and sap the body's vitality' (Tributsch, p 140). Eiby also notes that 'In his *French Revolution* Carlyle remarks that "Hope ushers in a revolution as earthquakes are preceded by bright weather", but most Englishmen...have pictured earthquakes as the fitting climax to a storm'. Such contradictory descriptions of earthquake weather would erode its credibility, except that all the descriptions have something in common: electrical charging. Dry winds that sap people's strength are known to contain lots of positive ions (Tributsch, pp 114-118). There is also a lot of electrical charging during a thunderstorm. The 'oppressive' weather in the Hawke's bay is also suggestive of electrical charging. The particular kind of earthquake weather in a region may or may not actually precede its earthquakes. However, electrically charged weather would understandably remind people of the time before an earthquake, when the building electrical charge can make people feel nauseous and sap their strength. Electrical charging also explains the reports of ancient Greek scholars, who said that air was 'elastic', or 'heavy' before an earthquake.

Underground disturbances are also common in earthquake folklore. During the first century AD 'earthquake noise' is described by Pliny the elder, and this phenomenon has been used for earthquake prediction in China for 1500 years. Earthquake noise sometimes occurs without any earthquake and is 'a terrible sound that sometimes resembles a rumble, sometimes the lowing of cattle or the shouts of human beings or the clash of weapons' (Pliny in Tributsch, p 78).

If the earthquake noise is produced at the same time as an earthquake, the sound waves travel faster than the shock waves. For centuries, 'country folk' have used this to save lives at the last moment before an earthquake. In Cumana, Venezuela, 'it is widely believed... that the most destructive earthquakes are announced by very weak oscillations of the ground and through a howling... and it is rare that a false alarm is set off by a native' (Von Humboldt in Tributsch, pp 81-82).

Few earthquakes are preceded by noise, but it does happen reliably in certain areas. These areas tend to have shallow earthquakes (the overlying rock would muffle preceding noises of deep ones). They also tend to be rocky, mountainous regions. The noises are easier to hear where there is a lot of ground water, lakes and rivers, since water conducts sound better than rock (Tributsch, p 84).

Disturbances in the ground are also reputed to disturb the water table. Tales are often told of wells and springs turning cloudy, and exuding a distasteful odour (Tributsch, p 89). The sudden expulsion of water from springs is also thought to herald an earthquake. In Liaotung, China, where an earthquake was expected, wells were monitored. 'Ground water rose in 55% of the wells and fell in 15%. In 30% of them the water changed colour, became muddy, or foamed, and in some of them the water boiled' (Tributsch, p 91). This was prior to the successful prediction of the Haicheng earthquake.

However, there is scepticism about the changes in ground water being connected to the earthquake. Ground water in California is being monitored. Scientists are waiting to see if the next big earthquake there produces marked changes in the ground water, to assess the truth of this bit of earthquake folklore (Tributsch, p 92). This seems a somewhat unfair test of the connection between the changes in ground water and the Haicheng earthquake, since so few of the traditional signs of an earthquake occur consistently, especially not in disparate parts of the world. However, any recollections about the well water well before, and then after the earthquake would provide a 'quasi-statistical' control. While we cannot (unfortunately) run any kind of controlled experiment on earthquakes, it is true that other earthquakes should be looked at, to parallel the replication in a controlled experiment. The problem is that it is reasonable to hypothesise that *some* earthquakes are connected to changes in ground water. If the connection is real, but occurs rarely, it cannot be established by pointing out that, for example, 5% of earthquakes are accompanied by changes in ground water. The coincidence of the earthquakes and ground water changes could then be only chance. The best compromise would be to consider only earthquakes where changes in ground water are reported, and look at the correlation of earthquakes and ground water changes over time. Earthquake noise happens consistently in certain areas; hopefully the same is true of changes in ground water. If a sufficient length of time could be studied this would actually be a good way of establishing a connection, but it is likely to be quite a while before enough objective and quantitative data can be amassed.

A coming earthquake can effect the ocean as well. In 1783, in Calabria, the sea was so turbulent that sailors had to return to port, although the sky was clear and the wind still. In the village of Catro (in Calabria) residents fled their homes when they saw the sea retreat from the bank. However, they saved themselves from a devastating earthquake *instead* of a tsunami (Tributsch, p 92). In Honshu (1872) the sea receded so far it was possible to walk out to an island 140 metres from the shore. Within an hour there was an earthquake, accompanied by great tsunamis. Tsunamis travel at great speed; if the receding of the water about an hour before the earthquake was the result of another earthquake that wasn't felt in Honshu, a tidal wave would have appeared much sooner (Tributsch, p 92). Before the Tango earthquake in 1927 the sea was supposed to have dropped about 1 metre. Scientists think that this was due to ground deformation, and that the sea was acting as a level. In Tributsch's study of 'sea observations' he also noted that many eyewitnesses say that the sea becomes absolutely calm before an earthquake (perhaps the effect of the P-waves). Pliny the Elder held that earthquakes never occurred except when the sea was calm, and this ancient bit of folklore seems to be born out in the experience of observers all over the world (Tributsch, p 93).

One of the more unusual bits of earthquake folklore is that of 'earthquake fogs'. This can be found in Egyptian mythology, Chilean and Indian proverbs, and in the writings of Aristotle and Pliny the Elder. Recent observations of them have been made in Italy, Venezuela, and Japan. It seems likely that this 'fog' is a gas emitted from the earth, often purple or 'blood-red' in colour. As the fog rises and cools, purple or blood-red rain can be precipitated. Japan and Chile have traditions of intense starlight through fog signalling the approach of an earthquake, and this phenomenon has been reported in other countries (Tributsch, pp 132-137). Tributsch speculates that the intense starlight is an illusion created by the fog – it 'should differ from an ordinary fog because of the powerful electrostatic charge of its particles' (p 137).

There are many more bits of folklore concerning earthquakes, from many countries. It would not be an exaggeration to describe most earthquake folklore as 'odd', and no doubt, most of the tales are embellished in the re-telling. However, there are certain things that appear consistently in tales from all around the world: the coloured rain; the receding of the sea; the inexplicable panic in animals and humans, to name but a few things. In conclusion, there is a lot to be learned from earthquake folklore, tried and tested over the ages.

2.3 The History of Earthquake Prediction

Over the ages, human knowledge of earthquakes has been accumulating. If one looks hard enough, documented changes in the environment prior to an earthquake can be dug up. As indicated above, many insights about earthquakes are also found in folklore. However, the leap in technological advances this century has benefited the effort to predict earthquakes. Seismographs are able to give a much more precise idea of the location and size of earthquakes, and computers have made the statistical analysis of large quantities of data possible. A deterministic approach to prediction does not look plausible, but significant advances in the direction of predictions based on stochastic methods have been made (Vere-Jones, Victoria University, personal communication, 1998).

Earthquake prediction goes back to antiquity, where it was surrounded by mysticism and tied closely to religion. The Romans took the 'auguries' (or omens) of the Etruscan people very seriously, although they may have misunderstood them somewhat. The Romans held that the auguries pointed to political unrest and war as well as natural disasters. However, Tributsch notes that 'many of the signs the Romans considered most important — the *prodigies* — seem downright tailored for earthquakes, especially if we compare them with precursors known to the Chinese' (p 196). Knowledge of the interpretation of the signs was barred to the public, but they were enlisted to keep a look out for them. If the signs looked sufficiently ominous, the existing documents tell us the senate (presumably upon hearing this) would hold its session *outdoors*. The public believed that the coming disaster was caused by their wickedness, and could be averted by religious ceremonies of atonement. These were organised by the authorities, and they too were held outdoors. There is also a record of an 'evil omen', whereby an earthquake displaced the busts of the Gods in the temple. In response to this, the authorities took the seemingly strange step of organising public games, which were to last 10 days. It is possible that 'the anointed priests knew from their secret books precisely what the real situation was, but faked religious activities to avoid causing general panic or being exposed to public criticisms if their predictions were to fail' (Tributsch, p 200).

In other countries, earthquake folklore has been used to save lives, but no orchestrated effort by authorities has been made until recent times. The models by which we now understand earthquakes first started to take shape early this century. Andrew Lawson and Harry Reid conceived the theory of 'elastic rebound' right after the 1906 earthquake in San Francisco and 'the main insight proved to be astoundingly correct. It boldly anticipated plate tectonics' (Lomnitz, p 256). The idea was that the earth's crust consisted of 'blocks' moving past each other. Strain built up as they tried to move past and was released in the earthquake, during which the 'blocks' actually moved. The 'stress release model' is still at the forefront of earthquake research.

Thinking about the earth in terms of plate tectonics officially began in 1967 (Lomnitz, p 89) but, as early as 1620, Francis Bacon was struck by the similar shapes of the coastlines on opposite shores of the Atlantic (Eiby, p 50). Plate tectonics was something of a paradigm shift; previously scientists had 'attribute[d] most tectonic deformations either to gravity or to shrinking of the earth as a whole' (Tributsch, p 89).

In a nutshell, plate tectonics elaborated on the elastic rebound theory: it pictures the earth's crust as being made up of 'plates' which ride (very slowly) on the convection currents of molten material inside the earth. When plates collide, earthquakes are one of the consequences. There is a lot of evidence to support the theory; for example, location – depth plots of earthquakes in the North Island of New Zealand *look* as if the earthquakes are occurring on the edge of a subducting plate. Also, the magnetic patterns on either side of the volcanic ridges along sea floor are mirror images. This suggests that the volcanic ridge is slowly adding to the sea floor on either side of it and pushing the existing sea floor out. This explanation of what we see along what are 'theoretically' plate boundaries fits the facts well, and the theory is now well established. Provided that this theory is correct, plate tectonics is a significant advance in the understanding (and therefore, the prediction) of earthquakes.

The next logical step was to somehow measure how much strain a fault was under, in order to assess the risk of an earthquake. As early as 1935, Hugo Benioff invented the Benioff strain meter, with an accuracy of one part in a billion — 'the sun shining on the hillside was enough to drive the instrument off-scale' (Lomnitz, p 180). Strain measurements are taken in New Zealand today; the Wellington Fault is repeatedly surveyed. Scientists have been looking for 'triggers' that release the build-up of forces.

The accurate recording of an earthquake's size and location provided by seismographs opened up new avenues for earthquake prediction. The first seismographs were only suitable for recording large earthquakes, occupied a lot of space, and were not especially portable. For example, the Wiechert instrument had a great tank filled with 17 tons of iron ore (Eiby, p 8). Most countries doing research into earthquake prediction have about 30-40 years of reasonably reliable data on earthquakes, as 'earthquake prediction was born out of the deep pockets of government science in the 1960s and 1970s' (Lomnitz, preface, p x – xi; Harte and Vere-Jones).

2.4 Seismographs

Recording an earthquake's intensity is difficult when there is no stationary point with which to compare the shaking. To get around the problem, seismographs make use of inertia. The main principle behind a seismograph can be illustrated by considering a hand-held pendulum. If the natural period of the pendulum is much shorter than the period of movement of, say, your hand, the weight at the end of the pendulum just follows your hand. If we transfer this analogy to seismograph, and replace the hand with shaking ground and the weight with the stylus of the seismograph, this would give a perfect record of the shaking in an earthquake. However, there needs to be magnification so that the differences between strong and mild earthquakes can be clearly seen. If we increase the period length of the pendulum relative to the period of the support, then the magnification of the record of shaking will increase. However, the pendulum will continue to swing after the shaking has stopped. To prevent it from doing this, it is damped: extra resistance to its swing is incorporated. Methods of damping have included suspending the pendulum in oil; increased friction at the support; or having a 'vane' with air resistance; and, more recently, a metallic vane that is damped by magnets. The damping is done so that the pendulum does not make any extra swings after the shaking has stopped.

Most modern seismographs are electromagnetic. A small pendulum with a 'pick-up coil' on the end swings between the poles of a magnet, generating an electric current. This is fed into a galvanometer, a sensitive meter with a mirror instead of a pointer. A small movement in the mirror causes a big change in the position of its reflection, and this is recorded on photographic paper. Having the mirror may seem unnecessary, but it has several advantages. The actual record is located away from the part of the instrument sensitive to movement, so 'earthquakes' caused by the instrument's attendant knocking it are less likely to occur. Also, ground tilt does not distort the records. There are many different types of seismographs to measure different types of shock-waves, and California, Japan, and New Zealand have 'seismic networks', where seismographs can be placed up to a hundred km apart and linked to the same recorder.

2.5 Finding the Geographical Origin of Earthquakes

To locate the source of the shock waves being sent through the earth in an earthquake, the wave types and the nature of the ground the waves move through needs to be analysed. There are two main types of earthquake waves; P and S waves (primary and secondary). The P-wave is essentially a sound wave, travelling faster than an S-wave. Also, P-waves 'push', and S-waves 'shake'. In a P-wave, each particle in the ground moves to and fro in the direction that the wave is travelling (P-waves are also called 'compressional waves'). It can be pictured like a set of dominos — the first domino gives the next one a push, that domino pushes the next one, and so on. A kind of 'wave' travels through the positioned dominos, and this is similar to how P-waves travel. S-waves can be compared to the kind of waves made when shaking one end of a piece of rope (with the other end fixed). The rope moves at right angles to the direction that the wave is travelling in. Thus, the S-wave is also called the 'transverse wave'.

The fact that the P-wave travels faster than the S-wave allows seismologists to calculate the distance they have travelled upon arriving at the seismic station: the further the seismograph from the source of the earthquake, the greater the distance will be between the two waves. The actual location (*ie.*, the direction that the waves are travelling in) can also be estimated by triangulation of the combined results from at least 3 stations. Each individual station has an estimate of the distance, so the earthquake could be anywhere on the boundary of a circle drawn around the station, (with its radius equal to the distance of the earthquake). Given any two stations, their circles will intersect in two places, so a third circle / station is needed to determine which of these two points is the location of the earthquake. However, with the advent of computers, this distance can be calculated iteratively. The procedure compares the arrival times of P and S-waves actually observed with what they should be if the estimate of the location is correct. Starting with a rough estimate initially, the location is re-evaluated until the position and the actual arrival times are consistent (or at least, reasonably so).

The deeper the focus (or source) of the earthquake, the less accurate this methodology is (the focus is the point under-ground where the rupture occurs, and the epicentre is the point on the earth's surface directly above this). Therefore, a further modification to the procedure is made to take account of the earthquake's depth. If circles were drawn around the stations for a very deep earthquake, there would not be a common point of intersection. The distance recorded in every case is the distance between the focus and the recording station (not the epicentre and the station) and this source of error can be very significant for earthquakes on the boundary of the subducting plate (these are the deepest type of earthquake). The deepest shocks recorded have been almost 700 km beneath the surface (in the Tonga-Kermadec Trench and in the region south of Sulawesi in Indonesia). Therefore, the radii of the circles are adjusted until they all intersect at the same point. The nearer the earthquake is to the station, the greater the effect of changing the focal depth will be on how closely the circles intersect. For a good measurement of the depth, the station needs to be at least as close to the epicentre as the distance of the focus from the surface.

2.6 The Earth's Structure and Seismic Readings

Waves actually travel faster as they penetrate deeper into the earth, so the above calculations are only an approximation. P and S-waves that have travelled a long way penetrate further into the earth, so this needs to be taken into account when calculating the distance they have travelled. Naturally, more distant earthquakes cannot be located as accurately. As well as the distorting effect of different types of earth on the waves, a small error in the direction of the wave results in very large errors in location. Eiby asserts that the location of a well recorded shock is usually estimated to the nearest 5-10 km, and greater accuracy can be obtained if the regional geology is taken into account (with reference to New Zealand, p 17). This seems to be all very well in theory: Vere-Jones and Harte found that the PDE and New Zealand local catalogues contained some serious discrepancies (Harte and Vere-Jones). This indicates that the recording is not as good as is possible, in either the PDE or New Zealand catalogues at least.

PDE stands for '*preliminary* determination of epicentres', and is recorded in the United States (Harte and Vere-Jones, abstract), so perhaps some discrepancies are to be expected. There are many more sources of error when taking teleseismic readings, and it is not just the distance that the shock waves must travel that contribute to the errors.

Probably the most useful scientific property of earthquakes is that they reveal the nature of the earth's interior. Eiby uses the analogy of a water-tank on a farm: the easiest way to see if it is full is to thump it (pp 18-19). The sound it makes tells the inquisitor how dense the material inside is — whether there is water or air in that particular place. Our interest is the other way around from that of earth scientists: we want to know how the structure of the earth affects earthquake waves, not what earthquake waves can tell us about the earth. However, the discovery that the earth has a 'core' originated from the 'shadow zone' observed on the teleseismic records of large earthquakes.

When the records of a large earthquake that can be felt all over the world are collated (from various points around the world) some anomalies with regard to P-waves show up. At around 103° from the epicentre, the signal from the P-wave fades out. It can be picked up again at 142°, but the travelling time it has infers that it is not the same P-wave. What has happened is that the wave has been refracted through the earth's core. This effect, combined with the accumulated errors over the great distance that the wave travels, make the teleseismic readings less accurate than local readings. P waves can also be reflected off the earth's core, reflected off the surface of the earth, and/or be reflected on the inside of the core. The waves can have any permutation of reflection and refraction, and the number of times the wave can be reflected/refracted varies (mainly depending on its energy). There are also waves that travel along the surface of the earth (Raleigh and Love waves, named after their respective discoverers), and they arrive after the S-wave. Understandably, this makes the seismic records quite complicated. Fortunately, it is possible to pick out P and S-waves just by looking at the record, and this gives the distance of the earthquake from the seismic station. Tables of travel times are then used to fill in the rest of the picture (Eiby, p 23). Eiby notes that New Zealand Seismologists check their readings against a very good network of stations at the antipodes, in Europe, but the readings do not line up so well with those recorded in America (as Harte and Vere-Jones show).

2.7 Measures of Earthquake Size

The first efforts to measure the magnitude scale of an earthquake depended on Wood-Anderson seismographs, but these were not installed world wide, and weren't very useful for measuring shocks more than 1000 km away. Magnitudes measured this way are denoted by the symbol 'M_L'. The next scale was based on surface waves (denoted by M_S) but deep shocks, which were discovered later, do not generate surface waves. The differences between M_L and M_S are not great for magnitudes over five (on shallow earthquakes), but below this, the magnitudes for distant earthquakes may be as much as half a magnitude too low. Also, magnitudes based on surface waves cannot deal with changes in crustal structure very well and are recorded too low.

2.8 Quality of Seismic Data

New Zealand has a reasonably accurate history of all its earthquakes from 1963 onwards; however, it is not perfect. A study by Harte and Vere-Jones compared two independent earthquake catalogues: the 'Preliminary Determination of Epicentres' (PDE) Catalogue put out by the United States, and the local New Zealand Catalogue. The PDE catalogue is recorded in the United States, and therefore consists of teleseismic records. They found that the PDE catalogue 'misses many events... with $M_L > 5$ recorded in the NZ catalogue; after 1983 most of the NZ shallow events are recorded, but some intermediate events are still missing' (Harte and Vere-Jones, 1998).

Fortunately, the NZ catalogue doesn't miss events that the PDE catalogue records, except to the North East of New Zealand and it misses a few occurring to its South West. The North East has additional problems: the events are systematically displaced to the East in the New Zealand catalogue (relative to the PDE catalogue), and the depths are sometimes recorded as intermediate in the NZ catalogue, and as shallow in the PDE catalogue. Unfortunately, this is also a very active area, and where the M8 algorithm had most of its success. The study of the NZ catalogue's quality was motivated by the paper by Ma Li and Vere-Jones which doubted whether M8's success was real. It appears that the sudden drop in activity that appears in this area is a recording bias; it is way out in the middle of the sea, and keeping a vigilant watch on it became less and less of a priority (Harte, personal communication).

Magnitudes of individual events recorded in both catalogues differ by up to 1 unit in either direction. Given that the scale of magnitudes is logarithmic, for higher magnitudes this is quite a serious discrepancy. However, there are no systematic differences in the magnitudes reported.

2.9 Conclusion

The 20th Century has seen many advances in the study of earthquakes, and of the study of the earth's interior. In theory, seismographs are able to pinpoint the location and size of earthquakes, which is a great aid to efforts at earthquake prediction and risk estimation. Being able to discern patterns in the occurrence of earthquakes is dependent on the quality of the records. Because the catalogues are not entirely complete and accurate, and also because of the short time-span of the data, there is not yet sufficient confidence in the M8 algorithm to make predictions. However, the study of the existing data has suggested future approaches to using M8's time series, which will be detailed in later chapters.

Chapter 3 ~ The M8 Algorithm. Development, Testing, and Additions.

3.1 Introduction

In 1980 an article appeared in *Nature* entitled *Bursts of Aftershocks, Long Term Precursors of Strong Earthquakes* (Keilis-Borok *et al.*, 1980). The authors included V.I. Keilis-Borok, who was to develop the M8 algorithm with Kossobokov. The basic foundation for M8 can be found in this paper. In 1988 the same authors published another article in Nature describing a pattern recognition algorithm for intermediate term earthquake prediction. This differed somewhat to M8, using quiescence, long range interaction between sources, and a different measure of earthquake size. In 1990, Keilis-Borok and Kossobokov published *Premonitory Activation of Earthquake Flow: Algorithm M8*, detailing the current form of M8 available on the IASPEI software release. In different papers they outlined additions, modifications, and tests they had done with and to M8.

3.2 The Beginnings of M8

In 1980, Keilis-Borok, Knopoff, and Rotvain put forward a number of possible precursors of large earthquakes, on a intermediate term basis. The theoretical justification for these precursors was that of stress release. The theory is that tectonic pressure increases prior to an earthquake, until a fault eventually gives way and releases most of the pressure. The signs that the pressure is building should be evident in an increase in activity along faults. Also, special types of activity are thought to indicate a higher probability of tectonic strain, for example, a large numbers of aftershocks.

The three broad categories of precursory seismic activity are though to be quiescence, swarms, and migration. Quiescence is an unusual calm in an area where there are usually a reasonable number of earthquakes. The region is though to become 'locked' — pressure can not be released through small earthquakes, so the pressure builds until a very large earthquake lets the pressure go. Swarms are clusters of (usually small) earthquakes, both in time and space. Migration (as the name implies) is the gradual movement of seismic activity (usually in a straight line) across a region. This was observed before the successfully predicted Haicheng earthquake in China.

However, these 3 precursors have never been defined in terms specific enough to be measured quantitatively. In this paper, the authors put forward mathematical definitions of swarms, quiescence, and migration, and then study their relationship with strong earthquakes. The definition of swarms is in terms of aftershocks, and is essentially 'series 7' in the M8 algorithm. The values of all the parameters involved are different - they have been estimated from Southern Californian data, and must have been refined subsequently. The 'patterns' involving quiescence and migration have not found their way into the M8 algorithm, and so are not discussed here.

As mentioned above, the parameters were estimated using only Southern Californian data. There are at least 9 parameters that need to be estimated, and, given the short supply of data on earthquakes there is a danger of 'data dredging', as the authors acknowledge. There were probably only 1 or 2 great earthquakes in the portion of the Californian catalogue that is reasonably complete and accurate, so the danger is that the authors could take chance happenings that occurred before those earthquakes, and declare them precursors. However, the authors purposely used mainly Californian data to estimate the parameters, so they could test the validity of the parameters on other areas. Where possible, parameters were given values determined in other studies to reduce the danger of data dredging. The magnitude targeted for prediction, and the lower magnitude threshold for including earthquakes into the analysis were given different values in each region, the reason given being 'lack of experience and a satisfactory theory' (p 261). Changing these parameters in a retrospective test does allow for a lot of '20/20 hindsight', and the authors acknowledge it as 'the single largest difficulty in the evaluation of the significance of our results' (p 261).

Excluding California (since it was involved in estimating the parameters) 13/17 earthquakes were predicted. Including the Californian data, 18/23 earthquakes were predicted - not surprisingly the Californian success rate was 100%. The Californian data was 'highly involved with the data fitting' (p 261), implying that the other catalogues played a minor role in it as well. The alarms covered 25% of space-time.

3.3 The Forerunner of M8

In 1988, Keilis-Borok, Knopoff, Rotwain and Allen published another article in Nature with an expanded list of precursory seismic patterns (Keilis-Borok *et al.*, 1988). The new patterns included general activity, temporal rates of change, a measure of earthquake size, and long range interaction between faults. Quiescence and swarms still featured, but migration was dropped.

The authors used a ' "subclasses" pattern recognition algorithm' (p 692). There are five regions being investigated, and in each the 2 years before a strong earthquake are designated as 'D' (or 'dangerous') times. Several functions were considered, and put through a screening process. To 'make the range of values of each function robust' (p 692) the continuous functions were made discrete, divided into small, medium and large values. The difference between the distribution of D and N values is compared; if it is sufficiently large then it has passed the first test - the variables listed above have obviously passed. The variables that have passed then contribute to defining 'D and N features'. These simply specify the range of one of the functions depicting a seismic characteristic of the area (or several together) for dangerous and safe times. D features appear 'sufficiently often in D intervals and are sufficiently rare in N intervals' (p 693). It was found that the number of mainshocks was not a good predictor, as was expected, although it still passed the screening. Changes in seismic activity, large numbers of aftershocks, and increases in the dimensions of sources were found to be very important.

The results were very good for the California and Nevada regions (13 out of 14 strong earthquakes were predicted), but the data from California and Nevada were also used to estimate the parameters. The algorithm was then tested on the Pamir-Tien Shan and Caucasus regions in Russia, which have similar seismic characteristics to California and Nevada. When no parameters were changed, 9 out of 13 strong earthquakes were predicted. On three other regions with dissimilar seismic characteristics to California and Nevada, the algorithm also produced impressive results. With 25% of the area covered by TIPs all 7 earthquakes of the target magnitude were predicted. The depth of earthquakes permitted in the analysis was changed in one region, and the target magnitude was changed in all regions, going as low as magnitude 4.5.

Varying the length of the 'lead time' or TIP also has an impact on the success of the procedure. However, with no theoretical basis for setting the TIP length it has to be estimated from the data. The authors do not give the statistical significance of their predictions, because of the small sample size. The authors make a further disclaimer, recognising that the algorithm is not optimal, and that there is no set procedure for selecting regions or fixed parameters.

The authors deliberately estimated the parameters on only 1 region so they could test them on others, a strategy that they call 'geographic transfer'. The authors write, 'the set of traits we have identified is intrinsically relevant to the approach of an earthquake in diverse tectonic settings' (p 335). However, they do not keep all of the traits in the final M8 algorithm, (although they may have been included in another algorithm that was not as successful, called 'CN'). As the authors later acknowledge, these traits do need further testing.

3.4 The Emergence of the M8 Algorithm.

In 1990, Keilis-Borok and Kossobokov published a paper detailing the M8 algorithm in its current form on the IASPEI software (discussed in Chapter 3). In it, they defend the assumption of a constant time frame for TIPs (5 years) when the magnitude targeted for prediction is changing. They say that the lead time for a smaller earthquake is expected to be smaller, but not to any great degree. The authors use the recurrence Gutenburg-Richter law in the Pamirs and in Tien Shan to argue that the recurrence time is not dependent on the magnitude. As they acknowledge, the evidence is somewhat inconclusive. The findings are based on only two regions, and the extent to which recurrence times are relevant to TIP lengths is unclear.

The parameters of M8 were estimated using data world wide to predict earthquakes of magnitude 8. When testing M8's ability to predict earthquakes of lower magnitudes in the same area, this is not strictly a test on independent data. Some of the data that is input into the algorithm would be different, but there would be a lot of overlap. Every available region was used in the data fitting, so there is no 'geographic transfer' being done. The results are given in table 4.1, below.

Table 4.1 Performance of the M8 Algorithm in Various Regions of the Globe

Region	M0	Strong Earthquakes		Space-Time Volume of TIPs	Expected # of Predictions with Random TIPs
		All	Within	(%)	
World	8.0	7	5	5	0.4
Central America	8.0	1	1	16	0.2
Kuriti Is. And Kamchatka	7.5	2	2	17	0.3
Japan and Taiwan	7.5	6	5	20	1.2
South America	7.5	3	3	18	0.5
Western U.S.A.	7.5	-	-	5	-
Southern California	7.5	1	1	12	0.1
Western U.S.A.	7.0	2	2	24	0.5
Baikai and Stanovoy range	6.7	-	-	0	-
The Caucasus	6.5	3	2	12	0.4
East Central Asia	6.5	5	4	24	1.2
Eastern Tien Shan	6.5	4	4	27	1.1
Western Turkmenia	6.5	-	-	0	-
Apennines	6.5	1	1	10	0.1
Koyna reservoir	4.9	1	1	42	0.4
Greece	7.0	3	3	18	0.5
Himalayas with	7.0	2	2	8	0.2
Vrancea	6.5	2	2	58	1.2
Vancouver Island	6.0	4	4	20	0.8

Lines 10 and 11 have an earthquake in common, as do lines 1 & 4. The Iturup swarm of March 23-24, 1978 is in lines 3 and 4

Surprisingly, the success rate was better with the smaller earthquakes. Some data does overlap but as the note below the table explains, there is not as much overlap as would be expected from the names of the regions. The number of successful TIPs far exceeds the number that would be expected if the TIPs were just randomly occurring, indicating that there are 'M8' patterns in this data.

3.5 Case Studies of M8 Applied to Different Regions Around the Globe

In the same year, the authors of the M8 algorithm also published case studies on M8's performance in Japan and California. The results of testing the algorithm in Japan and California were already published (or about to be published) when the article above was written, but in the following article they go into more detail. In the case of California, Keilis-Borok and Kossobokov had communicated a prediction of the Loma Prieta earthquake in advance, so this was an real-time prediction disentangled from the complications of data dredging. The principal contribution to the TIPs for the Loma Prieta earthquake was a marked increase in lower magnitudes from about 1980 to 1989, and the increased activity was spread wide over the whole area of the TIPs.

In the paper on the Loma Prieta earthquake, they state that they consider earthquake prediction to be a process of narrowing down the time and the place. First, places where earthquakes can occur need to be defined. The only practical way of establishing this is to define places where earthquakes have occurred in the past as places where earthquakes can occur. M8 looks at patterns in seismicity - if there is not enough of it to gauge these patterns M8 cannot be used. Therefore the problem of knowing where to expect an earthquake where there are no records of any previously is not relevant to M8. M8 will (hopefully) give intermediate term predictions, and then the time and place need to be narrowed down further.

Usually, only shallow earthquakes are used as input to the M8 algorithm (and it is only intended that shallow earthquakes should be predicted). This means that it is not necessary for M8 to predict earthquake depth, as the depth of shallow earthquake do not have much impact on their felt intensity. In effect, a deep earthquake is further away - in New Zealand shallow earthquakes are usually considered to be less than 40 km from the surface, and deep earthquakes are many times deeper. However, in Japan, there is not the usual marked distinction between shallow and deep earthquakes. Deep earthquakes usually appear on the two surfaces of a subducting plate, and so are distinct from the shallow ones near the surface. However, because there is no clear distinction in Japan, earthquakes of all depths were used in the analysis.

The one earthquake the algorithm failed to predict was a near miss - the relevant functions must exceed their thresholds twice in a row: they did so once, but not twice. Giving a probability (even a qualitative probability) would be a feasible approach to making predictions, and with this method the success rate in Japan would be even higher. However, as mentioned above, the data used to test the algorithm has also been used to estimate the parameters to a degree.

The stability of the procedure (when different data is used) was also tested. However, the lack of data for testing hypotheses is the main problem when trying to predict earthquakes. To determine whether slightly different data would affect the success of the algorithm, the lower threshold for including earthquakes in the analysis was varied. Also, the time span used was varied, as a way of varying the data set. Changing the lower magnitude threshold did not make much difference to the results.

3.6 The Mendocino Scenario Algorithm

Keilis-Borok and Kossobokov acknowledge the criticism that the circle of investigation for which predictions are made by the M8 algorithm is large. For example it is 1,334 km in diameter when predicting magnitude 8 earthquakes, but reduces to 384 km in diameter when predicting magnitude 6.5 earthquakes. The M8 algorithm's authors have produced another algorithm to refine the prediction of the geographic location of earthquakes, based on the observation of quiescence in a TIP region before an earthquake of magnitude 7.0 in 1980, off the shore of cape Mendocino, in California. This algorithm is called 'Mendocino Scenario', and has had impressive success in retrospective tests using the principal of geographic transfer.

The authors of the algorithm state that 'the idea that seismic quiescence may be related to the occurrence of future earthquakes is a well-established one', but their definition of quiescence differs from those used in other studies. Quiescence is defined as a measurable decrease in seismic activity from the previous 6 years. There is also an argument for using increases in seismic activity to localise the prediction, but according to the data they have, quiescence is a better tool. There is some debate over whether seismic quiescence really is a precursor to a strong earthquake. However, the algorithm authors' case is strengthened by Scholz who 'has shown that intermediate-term quiescence may result from regional hardening or stress relaxation over volumes that are much larger than the size of the rupture zone of the subsequent earthquake' (Kossobokov and Keilis-Borok, 1990). This provides a possible explanation for the mechanism of seismic quiescence to back up the empirical evidence.

To use seismic quiescence to localise their predictions, the authors change from using a circle to using a square, and divide it up into 256 little squares (which from here on will be referred to as 'cells'). A measure of seismic activity is taken for each 3 × 3 block of cells, so that there are 256 overlapping 3 × 3 blocks of cells (here-after referred to as 'squares'). The overlap is presumably to minimise the chance of the boundaries of the squares affecting the predictions.

Every month, from 6 years before the diagnosis of a TIP, a measure of seismic activity over the past 2 months is given for each square. Squares are said to have low activity if the activity is 'less than 90% of all the other 2-month intervals in the data set' (pp 19, 764). Squares are 'strongly connected' when their central cells have a common side, either in time or space. The earthquake is expected to occur somewhere within the clusters of strongly connected squares that contain 4 or more squares.

After estimating the parameters on the Eureka earthquake the algorithm was transferred to 17 other areas in North America, Japan and Eurasia where strong earthquakes had occurred (ranging from magnitude 6.5 to 8). The M8 algorithm was also applied to these areas, with the circle of investigation centred on the earthquake's epicentre.

Centring the square of investigation on the epicentre of the strong earthquake should mean that the activity inside the square is of greater relevance to the earthquake that is targeted for prediction. If the earthquake targeted for prediction occurred on the perimeter of the circle, accumulating strain associated with it would be distributed on all sides of the epicentre. Therefore, some of the accumulating strain would occur outside of the circle of investigation. Knowing the epicentre of the earthquake in advance and centring the circle of investigation on it should improve the performance of the M8 algorithm, but it is a luxury not available in real time predictions. In 3 cases the M8 algorithm was not applied, but where it was there the earthquake was successfully predicted. Not surprisingly, this success rate of 100% is unusual in tests of M8. Earthquakes appeared in the reduced area of prediction 16 out of 17 times. This seems a very good result, especially since the Mendocino Scenario reduced the area of prediction by between 7 and 25%.

3.7 Using Active Zone Size Instead of the Number of Earthquakes

In the paper, *Active Zone Size Versus Activity: A Study for Different Seismicity Patterns in the Context of the Prediction Algorithm M8*, (Kossobokov and Carlson, 1995), using the proportion of the region in which earthquakes occur instead of the number of earthquakes is discussed. On the theoretical side, using the proportion of the region covered by earthquakes is preferable. This is because the build-up of strain that sets off a large earthquake should affect a fairly large area. If there are a lot of earthquakes but in a very small area, this is not indicative of the build-up to a strong earthquake.

Substituting the area of the circle of investigation with earthquakes over the thresholds M20 and M10 (as defined in Chapter 4, section 4.4.3) for the number of earthquakes over these thresholds in series 1 and 2 of M8 yields better results in data from the United States. Kossobokov and Carlson acknowledge this is a limited data set, but say that the false alarm time is reduced and the Landers Earthquake is predicted (it is missed by the original M8 algorithm). A further complication is the size of the grid used to define active portions of the circle of investigation. If one earthquake occurs in a cell in the grid, that cell is counted as 'active'. At one extreme, if the grid only consisted of one cell the whole circle would be 'active' if only one earthquake larger than the lower threshold occurred. The size of the grid affects the success of the algorithm, and so is an extra parameter that has to be fitted. The authors acknowledge that this study is a pilot one, and 'in order to make a meaningful statistical statement regarding the relative performance of the precursors considered here, a much larger study must be done' (p 6440).

3.8 Conclusion

The original form of the M8 algorithm has undergone many alterations – some features have been dropped, others modified, and quiescence has been used in a separate algorithm. The retrospective test results are encouraging, although these have more chance of success than real time predictions because the location and magnitude of strong earthquakes are already known, and the data used in test was sometimes used to fit the algorithm's parameters as well. The algorithm still appears to be in its developmental stage, but does show promise.

Chapter 4 ~ The Working of the M8 Algorithm

4.1 Introduction

In order to experiment with the M8 algorithm, a number of alterations have been made to it by several researchers. This means there are several versions of the algorithm, but algorithms carrying the title of 'M8' generally allow the original procedure to be used. The version of M8 described here has been altered by the author and by Ma Li (Seismological Bureau of China) and is part of the Statistical Seismology Library at Victoria University, Wellington. It has a default mode which keeps to the original design of M8. However, parameters that are fixed in the original algorithm can be varied in this version. Also, there is the option of using different approaches to training periods, an additional variable called 'TIP.level' is generated and dates are converted to Julian dates in order to speed up processing.

4.2 Preparing Data for the M8 Algorithm

This version of the M8 algorithm is designed for use in conjunction with the Statistical Seismology Library (SSLib) at Victoria University. It is not a library in the conventional sense – it is a database of programs and data for studying earthquake prediction, or earthquake risk. It is also specific to the statistical package S+ (Mathsoft, 1995).

Earthquake catalogues for several countries are stored in the database, in the same format. The earthquakes are listed chronologically, and the data can go back hundreds of years, in which case the times of the earthquakes known only very approximately. However, earthquake records used with the M8 algorithm must be as complete and accurate as possible, so the algorithm eliminates data before the date specified by the person using it.

Earthquake data fed into M8 is 'declustered', that is, aftershocks are removed from the catalogue. However, for M8, the number of aftershocks associated with each 'mainshock' is needed. Therefore, there is an extra column in the catalogue fed into the M8 algorithm.

Before the raw catalogue from SSLib can be fed into the M8 algorithm, it must be fed through 3 functions: 'subset.circle', 'subcatalogue', and 'decluster.M8'. As mentioned in chapter 2, even recent data is not 100% complete and accurate. The effort expended in recording earthquakes can change from region to region, leading to spurious differences in seismicity between regions. For this reason, it is best to apply the declustering procedure to a subset of the catalogue, a bit larger than the circle of investigation used by M8. The reason for the subset being a bit larger is that aftershocks occurring outside the circle should still be included in the analysis where they are associated with a mainshock inside the circle. The excess radius over and above the radius of the circle of investigation depends on the magnitude being predicted. This subset is created by the function 'subset.circle'. Circles are used because all points on the perimeter are equidistant from the centre. It is difficult to justify the use of, say, a square because the rotation it has around the centre point would be arbitrary. The 'subset.circle' function doesn't actually produce a subset of the catalogue, but is a list of directions on where to find the selected events within the original catalogue. It is an S+ 'list object', a list of the row numbers of selected events, plus the arguments fed into the subset.circle function.

The function 'subcatalogue' should then be applied to the output from 'subset.circle'. This looks at the directions in the output from subset.circle and forms a subset of the catalogue. It creates an S+ dataframe (essentially, a list where all the items are the same length). Arguments passed to 'subset.circle' and miscellaneous other things are added to the data frame as 'attributes'. A table of them is shown below:

Attribute Name	Description of Attribute
subset	A summary of the output of the function subset.circle
minday, mindepth & maxdepth	The arguments (with the same names) passed into the function subset.circle
rownames	A list of row numbers in order (needed by other functions)
catname	The name of the output produced by the function subset.circle

The algorithm works through the events in order of magnitude, then time within magnitude, until no more aftershocks are found. Aftershocks are defined by the table below: an aftershock is any earthquake within the time and distance specified for the magnitude bin of the mainshock, where the mainshock magnitudes are like bin boundaries in a histogram. By default, the lowest magnitude that can be used in this procedure is 3, but this can be changed by the user. Once the raw catalogue has been processed by subset.circle, subcatalogue, and decluster.M8, it is ready to put into the M8 algorithm.

Mainshock Magnitude	≤4.4	4.5 -	5.0 -	5.5 -	6.0 -	6.5 -	7.0 -	7.5 -	8.0 -
Lag (days)	23	46	91	183	183	365	730	913	1096
Distance (km)	40	40	50	50	100	100	100	150	200

4.3 A Flow Chart of the Procedure Used by the Algorithm.

This flow chart begins at the point where the function 'M8' is called, where the declustered catalogue is fed into the algorithm. 'M8' is invoked by the user, and this in turn calls other functions. M8 Uses structured programming to a degree, but most functions have odds and ends in them and call other functions as well.

THE USER CALLS THE FUNCTION 'M8'

⇓

The function 'M8' prints details about the arguments passed into it to the screen. It then cuts out events in the catalogue that are earlier than the date specified by the user (with the argument 'minday').

⇓

THE FUNCTION 'M8' CALLS THE FUNCTION 'M8.series'

⇓

The function 'M8.series' cuts out events that are not in the circle of investigation.

⇓

Next, 'M8.series' does the subsetting of the catalogue that is required for computing the series - the subsets are 'in.CAT20', 'in.CAT10', 'in.CAT20a', 'in.CAT10a', and 'in.CATMS'. The subset 'in.CAT20' has smaller earthquakes cut out so that it has an average of 20 earthquakes per year. The subset 'in.CAT20a' is the same, except earthquakes stronger than the half a point below the target magnitude are omitted. Subsets involving 'in.CAT10' are the same, except the average number of earthquakes per year is 10. The subset 'in.CATMS' is the events in the catalogue that are between (M0 - 2) and (M0 - 0.2) in magnitude, where M0 is the magnitude targeted for prediction.

⇓

The time period that the data in these subsets come from depends on the training mode. When the training mode is 'moving' all the data is used to determine the lower magnitude cut-off point. In 'user' mode only data from the beginning of the series calculation period to the end of the training period is used. In 'all' mode, only data after the point at which the series are generated is used.

⇓

Next, a vector called 'time.breaks' is used to determine the bin lengths of the series, and the arguments 'minday', and 'maxday' determine the beginning and end of reliable data. If the user knows the time up to which the catalogue is complete, it should be specified with the argument 'maxday'. If it is not, the algorithm will assume that the date of the last earthquake is the date up to which the catalogue is complete. Similarly, if minday is not specified, it will be set to the date of the first earthquake in the catalogue.

The default option for 'time.breaks' is 6 month intervals - the M8 option. Changing this option deviates significantly from the M8 algorithm. To change 'time.breaks' from the default, a vector of break points is passed in (with the argument 'time.breaks'). Care must be taken that the arguments 'minday', 'start.series', 'end.training', and 'maxday' coincide with a breakpoint in the vector 'time.breaks'. The breaks should also be at regular intervals. As with all dates in this version of the M8 algorithm, 'time.breaks' is in Julian dates.

Using the time intervals in the vector 'time.breaks', the algorithm computes values for each time interval in the seven series. A running total is built into their calculation, and its span can be changed with the argument 'running.total'. Again, changing the running total is a significant change to the M8 procedure. The default (which is the M8 running total) is 6 years. Another subset of the original catalogue is also produced, using 'time.breaks'. This is called 'max.events', and only has the largest event in every time interval retained.

⇓

The function 'M8.series' then returns a list object to the function 'M8'. The first item in the list is an 'regular time series object', with the seven series plus an eighth column giving the Julian dates for the ends of the intervals. The origin for the Julian dates is the same as that of 'time.breaks'. The second and final item in the list is also a regular time series object, 'max.events', described above.

⇓

THE FUNCTION 'M8' CALLS THE FUNCTION 'M8.TIP':

⇓

The function 'M8' then calls the 'M8.TIP' function. First, the empirical distribution of the series is calculated. The way this is done depends on the training mode. If the training mode is 'moving', only the threshold percentiles for declaring a TIP are calculated. This is because 'moving' mode emulates the position of someone making real-time predictions - the empirical percentiles are updated with each 'new' data point. Presumably, recalculating the empirical distribution with each new point would require too much computer time, and so the TIP.level variable is not calculated in moving mode. If the training mode is 'all', the empirical distribution of the series is calculated using all available values of the series. If the mode is 'user', only values of series during the training period are used to calculate the empirical distribution. The regular time series objects 'tops' and 'exceeds' are computed. 'Tops' gives the threshold percentiles for each series relevant at time 'i', and 'exceeds' is a Boolean regular time series for whether the series exceeds the threshold percentile at time 'i'.

⇓

'M8.TIP' then looks at the decision rule for declaring a TIP. For a TIP to be declared, 5 out of the first 6 series must cross their 90th percentile, and the 7th series must cross its 75th percentile. Also, this must happen in 2 consecutive intervals. So, a 'TIP' is declared if

min(percentile of second lowest series - 0.9, percentile of 7th series - 0.75)

is ≥ 0 for two consecutive intervals. The actual value of this minimum is called the 'TIP.level', since it is a good indicator of how close the series are to declaring a TIP.

If a TIP is declared, the value of the Boolean variable 'TIP' is 'T'. If an earthquake ≥ the target magnitude occurs, the character variable 'TIP.type' is set to 'STIP' (successful TIP). If an earthquake with magnitude smaller than M0, but ≥ M0 - 0.5 occurs, 'TIP.type' is set to 'STIP-' (nearly successful TIP). If the TIP is triggered by an earthquake ≥ to the target magnitude, 'TIP.type' is set to 'c.e'. If no earthquake with magnitude ≥ M0 occurs, 'TIP.type' is set to 'FTIP' (false TIP), and if the 5 year duration of the TIP has not come to an end (and no earthquake ≥ M0 has occurred) 'TIP.type' is set to 'CTIP' (continuing TIP).

⇓

The results are then printed to the screen, and a list object returned to the function 'M8'. The list object contains the objects returned by the function 'M8.series' ('series', and 'max.events') as well as the variables 'tops', 'exceeds', 'TIP', 'TIP.type', 'TIP.level'. The training mode and the target magnitude, M0, are also returned.

⇓

THE FUNCTION 'M8' CALLS THE FUNCTION 'plot.M8'

⇓

The function 'plot.M8' produces a 4 X 2 grid of graphs. The first 7 are graphs of the seven series. These have a reference line for the 90th and 75th percentiles (for the first 6 and 7th series respectively). However, when the training mode is 'moving' these values are not constant over time, and are not plotted. The 6-monthly values of the series are joined. Larger points superimposed on the line indicate that a TIP has been declared, and their colour indicates the type of TIP. The points for the different types of TIPs are indicated below (on the graphical device 'motif'):
Successful TIP — Purple diamonds
Near Successful TIP — Cyan (light blue) triangles
False TIP — Blue 'X' marks
Continuing TIP — Red Crosses
TIP caused by quake of target magnitude — Green squares
The eighth graph is of the maximum magnitude in each time interval, and of the 'TIP.level'. The maximum magnitude is shown by a spike, and the TIP.level shown by a line

⇓

THE FUNCTION 'M8' THEN RETURNS EVERYTHING THAT 'M8.TIP' RETURNS, CONCLUDING THE CALCULATIONS.

4.4 The Algorithm in More Detail

4.4.1 Arguments Passed in to the M8 Function

As mentioned above, the arguments passed in have been changed. However, modifications that can not be counted as additions are very minor. Ma Li added the training modes 'all' and 'moving', and the variable 'TIP.level'. The author added the parameters 'time.breaks', 'running.total', 'smoother', and 'TIP.length'. These refer to the time intervals over which the time series are calculated, the running total used on the series, the length of the running maximum, and the duration of the TIP, respectively. In the original version of M8 these values were fixed at 6 months, 6 years, 3 years, and 5 years. The author also changed dates to Julian dates, to lessen the running time of the algorithm, and forced the omission of time intervals for which the earthquake catalogue is not complete. The following is a list of arguments passed in to the M8 function.

catalogue

The mainshock catalogue (created by the 'decluster.M8' function), with the number of aftershocks in following 14 days. This catalogue should not contain only the circle of interest. The catalogue is sorted by date and time, in ascending order.

M0

Earthquakes with a magnitude greater than or equal to this are being predicted.

center.longitude

The longitude of the centre of the circle of investigation.

center.latitude

The latitude of the centre of the circle of investigation.

radius

The radius of the circle of investigation. The default radius is $55.5(e^{M0-5.6}+1)$

minday

This is a 'Julian' date – the number of days since the first of January, 1960. Records earlier than this are not used in analysis. By default, all the data is used.

training

The available historical data is divided into a 'training' period, and a period where the series are calculated (which can overlap). This is to test the performance of M8 on historical data. The training period is taken to be the known history of the series, and it is imagined that a test (for the duration of the series calculation period) is being conducted in the present. The training and calculation periods can be used in three training modes: 'moving', 'user', and 'all'.

In 'moving' mode, each new data point is used to reassess the empirical distribution of the series, so the historical percentiles that the series are being compared with are constantly being re-evaluated. All available data is used to compute the lower magnitude cut-offs (the magnitudes above which there are averages of 10 and 20 earthquakes per year).

In 'user' mode the training period is set at a fixed interval by the user, with the end.training argument. Empirical distributions of the series and lower magnitude cut-offs are calculated on the overlap between the series calculation period and the training period.

In the 'all' mode, the training period is fixed as the period from the beginning of the catalogue to the date when the series start being calculated. The empirical distribution is also fixed, and all the available data is used in its computation.

start.series

This is the Julian date from which series are generated. Since the series are calculated using bins, start.series must coincide with an end point of one of the bins. The default option is six month bins, and using different bins deviates significantly from the M8 algorithm.

end.training

This is the Julian date for the end of the training period. This is only used when the training mode is 'user'.

plot.it

This is a Boolean variable, and the default is 'TRUE'. Under the default, seven plots for each of the seven smoothed series are generated. There is also a plot of the largest earthquake in each 6 month interval, overlaid with the 'TIP level' series

title

This is a character string giving a title for the plot. The default is 'M8 Series and TIPs'.

debug

This is a Boolean variable, for whether to run the algorithm in a way that aids the debugging process. The default is FALSE

time.breaks

This is a vector of (Julian) times that give the bins for calculating the series. The (M8) default bin length is 6 months, and, as mentioned above, deviating from this is a significant alteration to M8.

running.total

By default, the series are calculated with a six year running total. This is the M8 default, and changing this is also a significant variation on M8.

smoother

The M8 series are smoothed by taking their maximum value over the past 3 years. The length of time over which the maximum is taken can be changed, deviating from the original specifications of M8.

TIP.length

This is the duration of the TIP or 'Time of Increased Probability'. The M8 default is 5 years, regardless of the magnitude of the earthquake being predicted. Again, changing this is a significant alteration to M8.

4.4.2 What the Algorithm Produces

The algorithm produces a list object with the following components.

series

This is a 'regular time series object' with n rows and 8 columns. The first 7 columns are the seven M8 series, and the last is the datetime corresponding to the series values.

max.events

This is a list with the same variables as the catalogue. However, the variables are regular time series (with intervals defined by the argument 'time.breaks'). The columns contain characteristics of the largest earthquake in that 6 month period. The list items are (in order) 'longitude', 'latitude', 'depth', 'time', 'missing.time', 'magnitude', 'catname', and 'n.aftershocks'. 'Missing time' is a record of how accurately the time of the earthquake is known, for example, whether the time is known to the nearest second or the nearest day. 'Catname' is just a character string giving the catalogue's name. 'N.aftershocks' stands for 'number of aftershocks' for the earthquake for that particular time interval.

tops

This is a 'regular time series object' with n rows and 7 columns. It contains the historical top 10% (for the first 6 series) or 25% of values (for the 7th series) for each 6 month interval. The way the historical percentiles are defined depends on the training mode.

exceeds

This is a regular time series object with n rows and 7 columns. It is Boolean, indicating whether each series exceeds the relevant historical percentiles for each six month interval.

TIP

This is a Boolean variable for whether a TIP is declared in each 6 month interval.

TIP.type

This classifies TIPs according to the following criteria:

c.e

A TIP caused by an earthquake of the target magnitude, that is, an earthquake with magnitude greater than of equal to M0 in the preceding year of the TIP declaration

STIP

A 'successful TIP', where a large event occurred in the 5 year period after the TIP's declaration.

FTIP

A 'false TIP', where a large event failed to occur within five years of the TIP's declaration.

CTIP

A 'current TIP', where less than 5 years has elapsed since the TIP's declaration, and no earthquake with magnitude greater than or equal to M0 has occurred.

STIP $^{-}$

This is between an STIP and an FTIP; strictly speaking it is an FTIP. This is where an event with magnitude less than M0, but greater than or equal to M0-0.5 occurred in the ensuing five years.

TIP.level

This is a number between -0.9 and 0.1 indicating the estimated level of danger. It is percentile of the series with is the least conducive to declaring a TIP (see section 3.3 for details). Two consecutive positive values of TIP.level is equivalent to a TIP.

4.4.3 The Procedure Used by the Algorithm.

For reasons of clarity, this section assumes that all the parameters of the algorithm are set to their default values. The algorithm begins with removing events in irrelevant times and places. Then the seven time series are computed. These are regular time series, with intervals of six months, and a six year running total incorporated into their calculation. Before the series are used for prediction they are smoothed again by taking the maximum value over the past three years. Then M8's decision rule is applied: if five out of the first six series are above their 90th percentile, and the 7th series is above its 75th percentile for two consecutive intervals, a 'TIP' is declared.

Series 1 is the number of earthquakes larger than a variable called 'CAT20'. CAT20 is the magnitude over which there are an average of 20 earthquakes per year. The interval this is calculated over depends on the training mode (see the argument 'training' in section 3.3.2). Series 2 is the same as series 1, but the average is 10 per year.

The function for calculating series 1 and 2 is

$$N(t{:}\underline{M},s)$$

where N( ) means 'the number of', t and s refer to the time interval which is being looked at (t changes but s doesn't) and M refers to a lower magnitude cut-off. The lower cut-offs 'CAT20' and 'CAT10' exist partly because the data quality for most catalogues is very poor at low magnitudes. The cut-offs also provide a way to 'normalise' the series, since different areas have different levels of seismic activity. For example, an area with lots of strong earthquakes will have more of the smaller ones omitted. This means that the same amount of data is analysed, no matter how active the region is seismically, giving equal statistical power.

Series 3 and 4 are derived from series 1 and 2 respectively. They are the difference between the number of earthquakes in the past 6 years and the number expected, based on all data 6 years before the date when the series start being calculated. The function used to calculate the series is

$$L(t:\underline{M},t-t_0) = N(t:\underline{M},t-t_0) - \frac{N(t-s:\underline{M},t-s-t_0)*(t-t_0)}{(t-s-t_0)}$$

where t_0 is the time at the beginning of the catalogue. The current point in time is t, and the distance looked back over is $t - t_0$. It may not be quite clear at first how this function measures deviation from the long term average. To see how this function operates, consider a regular time series, where each value comes from a period of 6 months. 1990.0 refers to the period 1989.6-1990.0. Numbers after the decimal point denote months, so 1990.12 would denote December 1990.

1990.0 1991.0 1992.0 1993.0 1994.0

1989.6 1990.6 1991.6 1992.6 1993.6 1994.6

t-s *t*

t_0

$N(t:\underline{M},t-t_0)$ is the number of earthquakes over the whole period of investigation, up to time *t*. $N(t-s:\underline{M},t-s-t_0)$ is the number of earthquakes excluding the last six months. If you divide $N(t-s:\underline{M},t-s-t_0)$ by $(t - s - t_0)$ this gives the rate of earthquakes over the period of investigation, excluding the last six months. Multiplying this by $(t - t_0)$ gives the expected number of earthquakes over the whole period of investigation, using all the information except that from the last six months. The series as a whole is the difference between the actual number of earthquakes over the period of investigation, and the number we would expect, not having information on the last 6 months. Therefore, this is a measure of the deviation of the last 6 months from the long term behaviour of the series.

Series 5 and 6 have the same lower magnitude thresholds as the first 4 series, but earthquakes used in their calculation must also be a at least half a point less in magnitude than M0 (the magnitude targeted for prediction). Series 5 uses CAT20, and series 6 uses CAT10 (the magnitude over which there are an average of 10 events per year). These series are the average strain release multiplied by the 2/3 power of the number of events. The function used to calculate series 5 and 6 is

$$Z(t:\underline{M},\bar{M},s,\alpha,\beta)=\frac{S(t:\underline{M},\bar{M},s,\alpha,\beta)}{(N(t:\underline{M},s)-N(t:\bar{M},s))^{2/3}}$$

where $S(t:\underline{M},\bar{M},s,\alpha,\beta)=\Sigma e^{\beta(M_i-\alpha)}$

The coefficient β is b/3, where b is the coefficient in the magnitude energy relation (or Gutenburg-Richter Law)

$$\log_{10} E = 11.8 + 1.5M$$

These series may be the result of experimenting with fractals, and just feeding in whatever kind of series seems to work. Series 5 and 6 are a measure of energy release and the number of earthquakes – the formula is for energy release multiplied by the two-thirds power of the number of earthquakes. The reason for using series 5 and 6 is not quite clear from the articles and documentation written by the original authors of the algorithm.

The series must exceed certain historical percentiles for a TIP to be declared. Using historical thresholds also allows for different levels of seismic activity, 'normalising' in a second way. For a TIP to be declared, areas that are very active seismically have to show a greater increase in the 5^{th} and 6^{th} series. As an example, consider the 90th percentile of the magnitude of all the earthquakes that occur in a circle of investigation. We will call these earthquakes the 'real set of earthquakes'. When the fifth series is calculated, the lower magnitude earthquakes are discarded so that there is an average of 10 earthquakes per year. Then the 90^{th} percentile of the remaining earthquake magnitudes is calculated, and it will have a higher magnitude than the 90^{th} percentile of the real set of earthquake magnitudes. The more active a region the more small earthquakes that are excluded, and the higher the magnitude cut-off point is. The higher the cut-off point, the greater the difference between the 90th percentile of the set of real earthquakes and the 90^{th} percentile of included earthquakes. Consequently, the more active a region, the higher the percentile of the set of real earthquake magnitudes it has to cross.

Series 7 only uses events with (M0 - 2) ≤ magnitude < (M0 - 0.2). It is the maximum number of aftershocks after any event in this magnitude range over the past year.

The variable 'TIP.level' gives the percentile of the series which caused the most obstruction to the declaration of a TIP. According to M8's decision rule, if:

Min((2nd lowest percentile of series 1-6 - 0.9), (percentile of series 7 - 0.75))

is greater than zero for two consecutive intervals a TIP is declared. Therefore, min((2nd lowest percentile of series 1-6 - 0.9), (percentile of series 7 - 0.75)) is an indicator of how strong the justification for declaring a TIP is.

4.5 Conclusion

All these calculation are done by three main functions, M8, M8.series, and M8.TIP. M8.series and M8.TIP do most of the work; M8.series calculates the 7 series with a running total, and M8.TIP works out the empirical distribution of the series and applies the decision rule. The reader may well question the absolute authority of the decision rule, and it is the authority of this rule that is the subject of later chapters.

Chapter 5 ~ Issues Arising when applying M8 to New Zealand Data

5.1 Introduction

The M8 algorithm is designed to work in any region with enough earthquakes. However, the algorithm is still being tested, and it may be that it will work better in some regions than in others. Some of the seismic characteristics of a region will affect the stability of M8's predictions. Also, around the world there are different seismic mechanisms that induce earthquakes. New Zealand has many deep earthquakes caused by subducting plates, and M8's rules regarding depth have had to be bent a little. This chapter looks at how the M8 algorithm interacts with New Zealand data, and the reasons behind the predictions made by M8.

5.2 Earthquake Depth

5.2.1 The Problem of Predicting Depth

The M8 algorithm is designed to only work with shallow earthquakes, and shallow earthquakes are (usually) very distinct from deep ones. Deep earthquakes only occur along the surfaces of the subducting the tectonic plate. At first, seismologists were reluctant to believe that deep earthquakes actually existed, partly because areas with a subducting plate weren't closely covered with recording stations. Also, Professor Turner (who discovered them) reported deep-focus events, but also events 'so high indeed that on any reasonable assumption about the normal depth of earthquakes, they would have been up in the air' (Eiby, p 22), which damaged the credibility of the reports. Shallow earthquakes normally appear as a distinct group in a plot of earthquakes by depth if only a small area is considered, such as a tenth of New Zealand (and the surrounding oceans). Deeper earthquakes are usually fewer in number than shallow ones, and are spread out over a greater range of depths. Shallow earthquakes occur in the earth's crust. The thickness of the crust varies — it is much thicker where there are landmasses. In New Zealand the thickness is about 40 km.

An earthquake 700 kilometres below the surface would have a felt intensity much less than a shallow earthquake of similar magnitude. M8 is only intended to predict shallow earthquakes (using shallow earthquakes) In New Zealand there is only an area of roughly 40 km for the depth to vary within. Relative to deep earthquakes, this is a very small distance and the depth does not effect the felt intensity very much.

5.2.2 Concerns Regarding Data Dredging

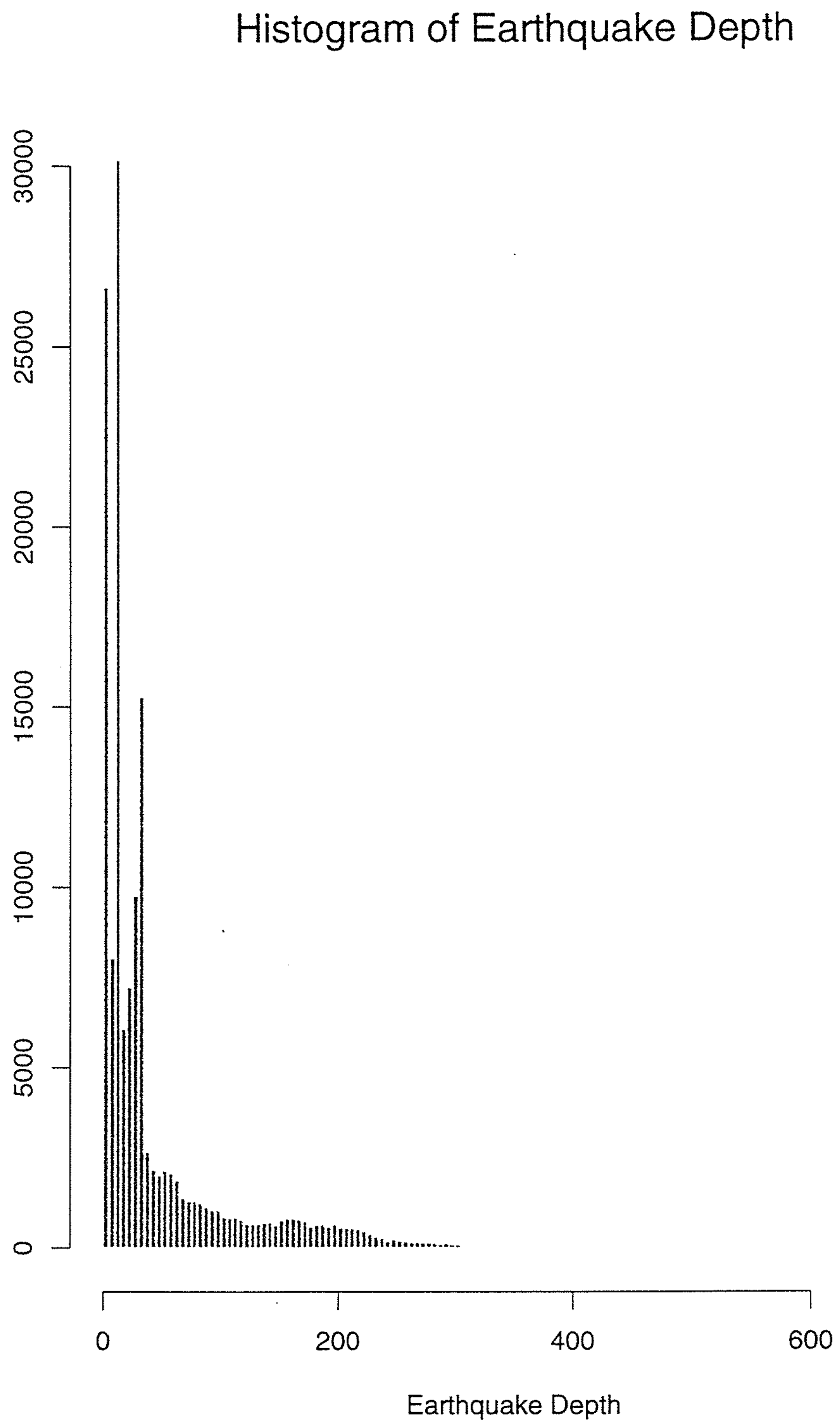
In retrospective tests of M8 in New Zealand, Ma Li and Vere-Jones found that M8 worked quite well so long as earthquakes of all depths were used. (Ma Li & Vere-Jones, p 80). M8 was not successful when only shallow events were used (or only deep events) and the authors were concerned that they were data dredging with regard to earthquake depth. Whether or not this is the case is unclear. It could be that more information is needed, and there is not enough statistical power with only shallow events included in the analysis. One possible way of clearing up the issue would be to systematically exclude deeper earthquakes, and note the effect on M8's predictions.

The authors of the M8 algorithm have included earthquakes of all depths where there was no clear distinction between shallow and deep earthquakes. However, in plots of earthquake depth, it is clear that New Zealand has a real distinction between deep and shallow events. This can be seen in figure 5.1, below. There appears to be an underlying exponential distribution with spikes where the depth is less than 40 km. Earthquakes with a depth greater than roughly 40 km are caused by one plate subducting under another, and figure 5.1 indicates that they are likely to have an exponential distribution for their depth. There are possibly some shallow earthquakes which are also due to subducting plates that are lost in the spikes.

5.2.3 Justification for Using Earthquakes of All Depths

As is explained in Chapter 3, the theoretical basis behind M8 is that large earthquakes are the result of release of tectonic stress, which builds up for years before the earthquake. M8's seven time series measure aspects of seismic activity, and when these increase the region is taken to be under a lot of tectonic stress (with a sizeable earthquake imminent). When the stress builds to a significant level, great cracks in the ground are thought to collapse under the pressure (or faults are thought to slip) resulting in an earthquake.

Figure 5.1



It is not clear why deep earthquakes would increase the stress on these faults or cracks to any great degree. These deep earthquakes are hundreds of km away from the area under stress, but perhaps there is another hidden factor that is triggering both shallow and deep earthquakes. One such factor could be the huge convection current that is driving the whole subduction process. The plates are only subducting because of convection currents inside the molten interior of the earth. The plates ride (very slowly) on these. An increase in force of the convection current would put pressure on the whole subduction area, from the surface right down to where the plate is melting as it is pushed down into the earth. This would result in an increase in deep earthquakes, as well as shallow ones.

The fact that all earthquakes are needed to successfully run the M8 algorithm in New Zealand may mean that fluctuations in strain are on a much larger scale. Instead of a plate 'getting stuck' as it is under pressure to subduct and generating a local area of stress, a much greater area of the plate may be under an increased amount of pressure. It could be expected that just the shallow earthquakes by themselves would be sufficient to make accurate predictions. However, there is not much data, and (as mentioned above) it could be that the problem with only using the shallow data is a lack of statistical power.

5.3 Analysis of New Zealand Data

5.3.1 Analysis in which Predictions are Directed at Magnitude 7 Earthquakes

An analysis similar to that done by Ma Li and Vere-Jones (1997) was carried out with the aim of studying the M8 series in greater depth and looking for patterns. In particular, an investigation into whether any series were more important than the rest was undertaken. There was a possibility that some series would not be used at all the in the predictions made by M8. This is because the M8 algorithm uses maxima and minima of the series instead of averages. If some series never had the status of being the maximum or minimum, they would be redundant.

The main reason for investigating the importance of each series was to isolate the most important series, and concentrate investigations on them. There was the possibility that in New Zealand only one or two series would actually be used by the algorithm. If this were the case, studying the M8 algorithm would be simpler as there would be fewer series to study. Studying the importance of each series also gave general information on the nature of the M8 series. It was thought that some series could be more important during TIPs. The differences between their behaviour during

TIPs and during 'safe' periods could be investigated, and the differences isolated to provide extra information on earthquake prediction.

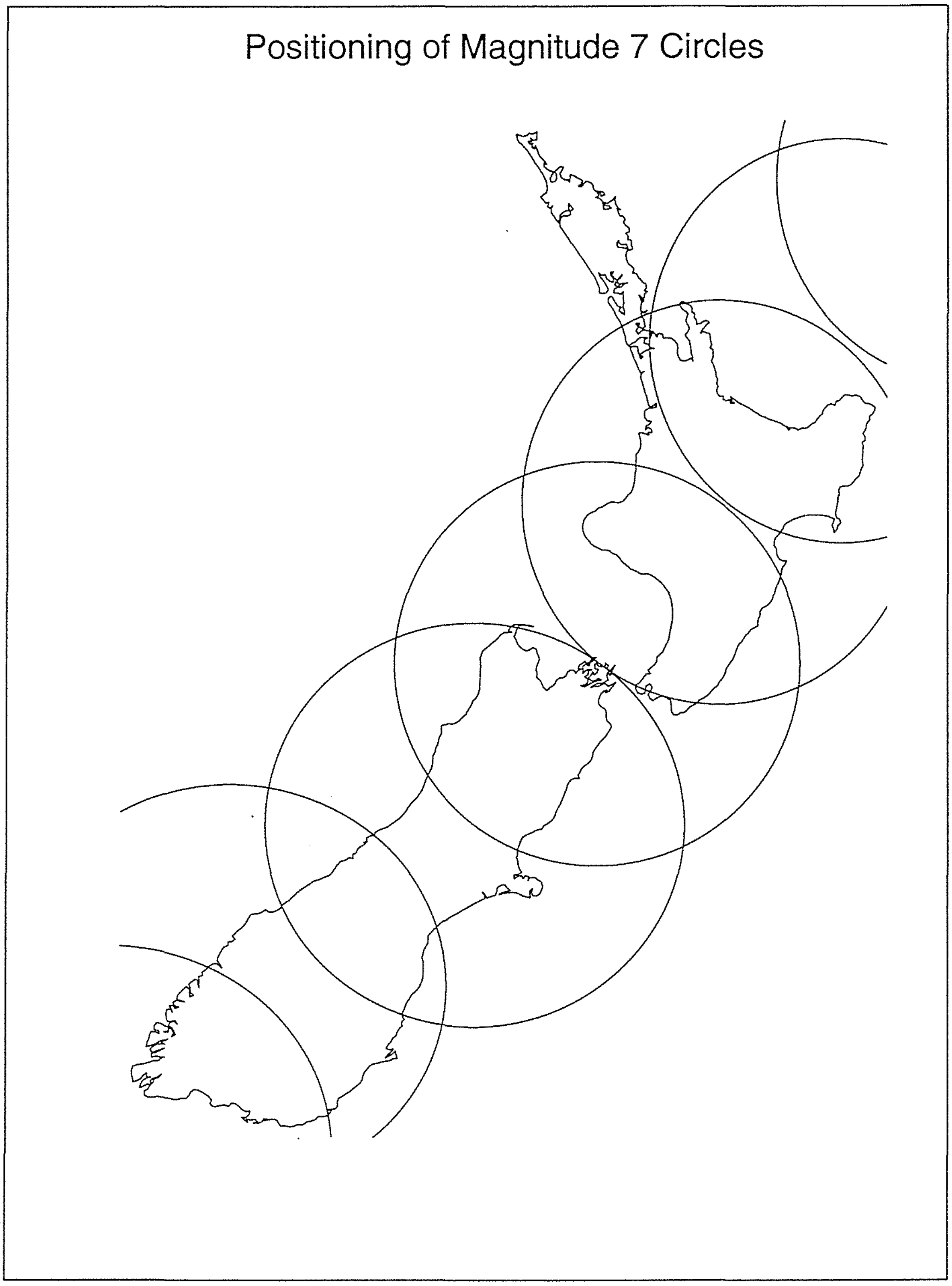
For this analysis, the magnitude at which predictions are aimed at is 7.0. Circles for predictions targeted at magnitude seven earthquakes have diameters of 562 kilometres. If the circles overlap by about one-third of the area, this maximises the information which can be squeezed out of the little data we have (Vere-Jones, personal communication, May 1998). To cover length the country, seven such circles are needed. Their positions are shown in figure 5.2 below. The sixth circle is not spaced in the same way as the others, as a circle in the normal position would not give enough data to run the M8 algorithm.

Only two magnitude 7 earthquakes occurred during the time that the series were being generated. Each circle is considered to be one unit of space, and the amount of 'space-time' is defined here to be space multiplied by time. The percentage of space-time in the seven circles covered by TIPs was 25%. Details of the success of M8 at predicting magnitude 7 earthquakes are given below, in table 5.1

Table 5.1

Success Criterion	Statistics
False TIPs	4%
Near successful TIPs	15%
Successful TIPs	3%
Continuing TIPs	2%
Number of failures to predict	1
Total number of earthquakes with magnitude ≥ 7	2

Figure 5.2



The failure to predict in the table above was quite a near miss. The TIP level series was only 0.06 below the threshold for declaring a TIP (0.06 is 6% of the TIP level series' range). Other instances of the TIP level series being between –0.063 and 0 constitute 5% of the time.

Near successful TIPs constitute quite a high percentage of the TIPs. There were19 earthquakes with magnitude greater than or equal to 6.5 but less than 7, that the algorithm was aiming to predict. With totally random TIPs, an average of 5.3 earthquakes in this magnitude range would be predicted. Since 10 were predicted, the M8 algorithm has some ability to predict magnitude 6.5 earthquakes when its target magnitude is actually 7.

As mentioned above, the decision rule about whether the tip will be declared is based on the maxima and minima of the series. The tip is declared if five out of six of series 1 to 6 are above their thresholds, and series 7 is also above its threshold. This means that the tip is based on only one number, the 'Tip Level':

Tip level at time 't' = min(V - 0.9, W - 0.75)

Where:

V = second lowest percentile of series 1 to 6,

W = percentile of series 7.

Since five out of 6 of the first six series must cross their threshold the percentile that is of interest is the percentile that is the 2nd lowest. If this percentile is greater than 0.9, 5 out of six series have crossed their thresholds. If the percentile of the series 7 is greater than 0.75 then it has crossed its threshold to. Since both V and W must be above.9 and.75 respectively, the minimum of V -0.9 and W -0.75 is the deciding number. If this is above zero twice in a row, then a TIP is declared.

This means that the decision to declare the tip is based on the percentile of one series only and the series can be said to be the 'critical series'. It is true that all the series were used in the decision, since the critical series is 'critical' when compared to the other series. However, if one series is consistently used to declare the TIP when others are not the series that are not used would be redundant. Investigation of which series are 'critical' was undertaken to see if any series were redundant. Whether any series were consistently critical was also of interest.

The graphs, in figures 5.3 and 5.4 below, show how important the first 6 series are, relative to each other. The colour coding goes from 'hot' to 'cold'. A 'hot' series (coloured red) is in its lowest percentile in comparison to the other series. Circles 1 and 3 are shown, as they have the greatest and least numbers of large earthquakes, respectively. The 7th series is not considered in these graphs, as it is of a different character and appears in a different stage of the M8 rule. The number above the series indicates the rank of the percentile that the series as in (it is ranked against the percentiles that the other series are in). For example, if a series has a 'one' at any point in time this means it is in the lowest of its percentiles relative to the rest of series 1 to 6. Figure 5.3 and 5.4 show that all of the first 6 series are needed — all series have a turn at being in the lowest (or 2nd lowest) of their percentiles compared to the other series.

At least one of series 3 and 4 is in the lowest percentile during a tip in all 7 circles. Series 3 and 4 measure the deviation from the long-term trend. Series 3 and 4 are often in lower percentiles than the other series before a TIP, but not always. They remain lowest for varying lengths of time, sometimes both series are low in comparison to the rest, and in circle 6 they are tied with the third pair of series (average strain release multiplied by the 2/3 power of the number of events). These patterns were picked out on only 7 circles (which overlap each other by roughly a third of their area). Because of the small sample size, the probability that some patterns like these could be found is high. Also, the patterns are not especially strong. For this reason, special investigation has not been undertaken with regard to any one series.

The decision on whether series 7 will be the critical series or the second lowest of the first 6 series will be the critical series depends on whether V or W is lowest. In figures 5.5 below, the identity of the critical series is shown (along with the percentile of the critical series (which is equivalent to the 'TIP level' series). Series 7 is critical for three periods in Circle 3, one period in circle 4 and quite long periods of time when the series is rising in circle 6 and 7. Therefore, it too is needed in predictions by M8. Series 7 appears to be critical when the level of the series 7 is quite high. It is unclear whether series 7 is giving the correct signal in circle 4, 6 and 7 as the TIP declared is a continuing TIP (an earthquake of the required magnitude may yet appear to validate TIP).

Figure 5.3

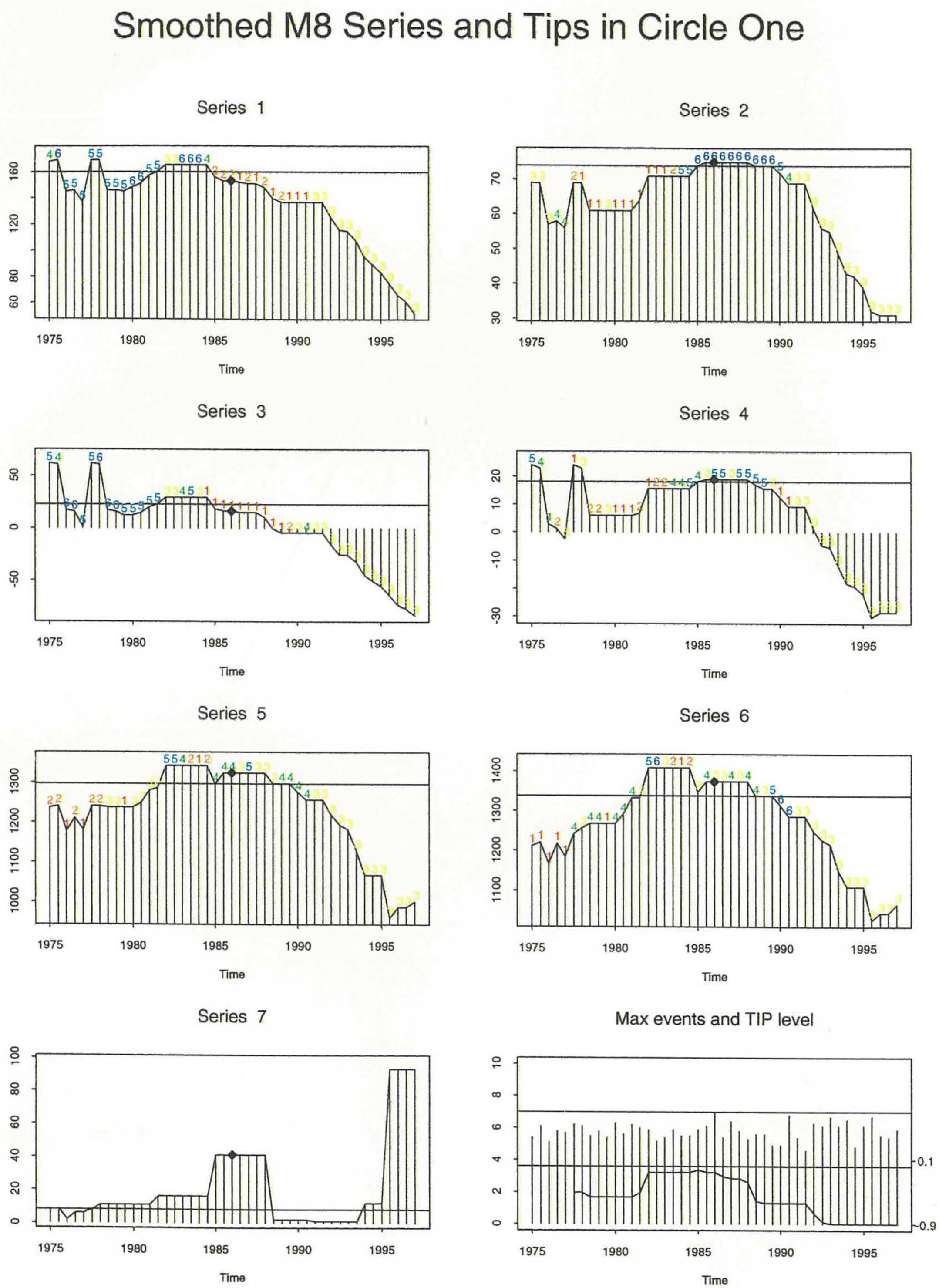

Smoothed M8 Series and Tips in Circle One
Series 1
Series 2
Series 3
Series 4
Series 5
Series 6
Series 7
Max events and TIP level
Time
1975
1980
1985
1990
1995
0.1
0.9

Figure 5.4

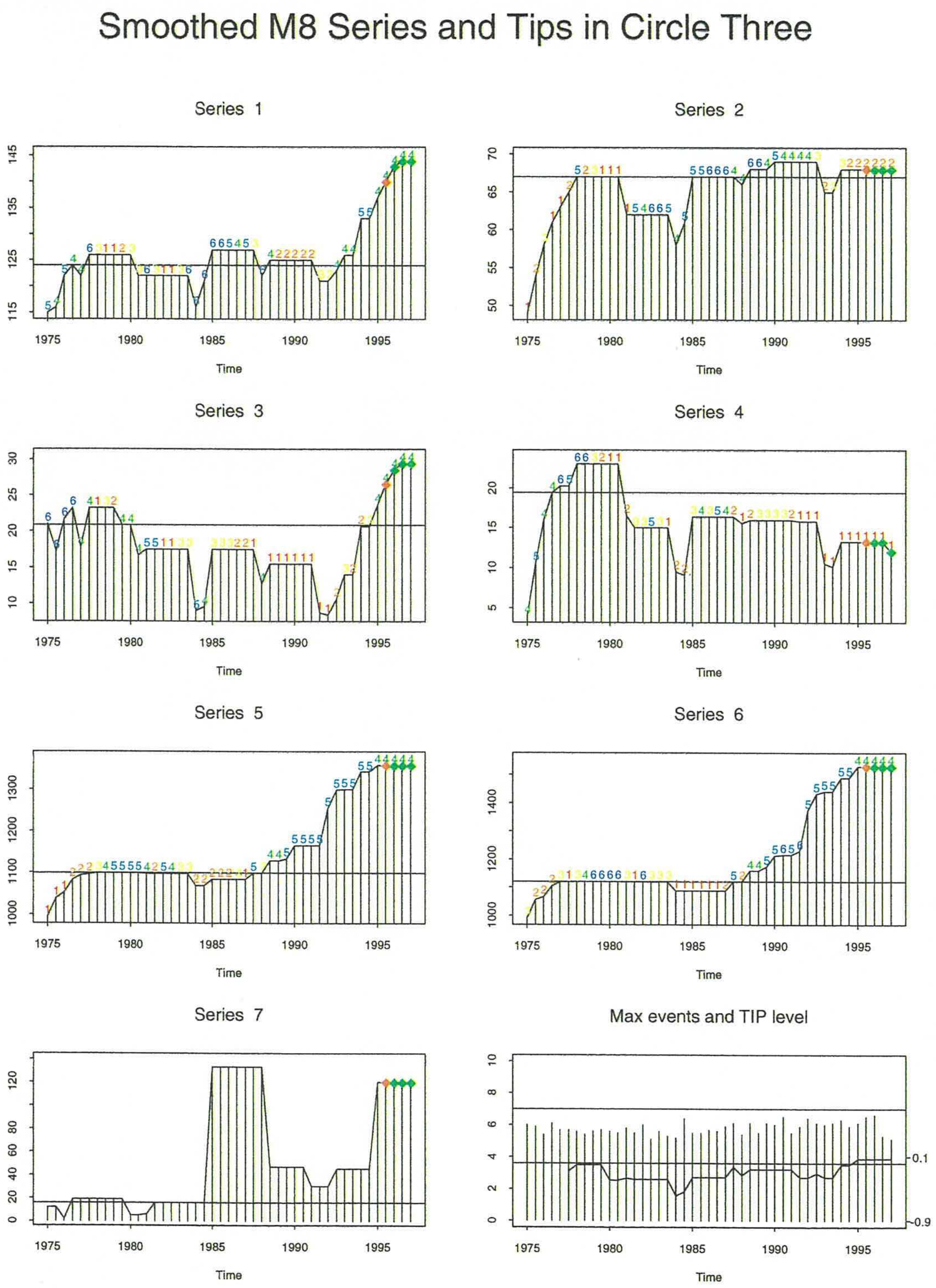
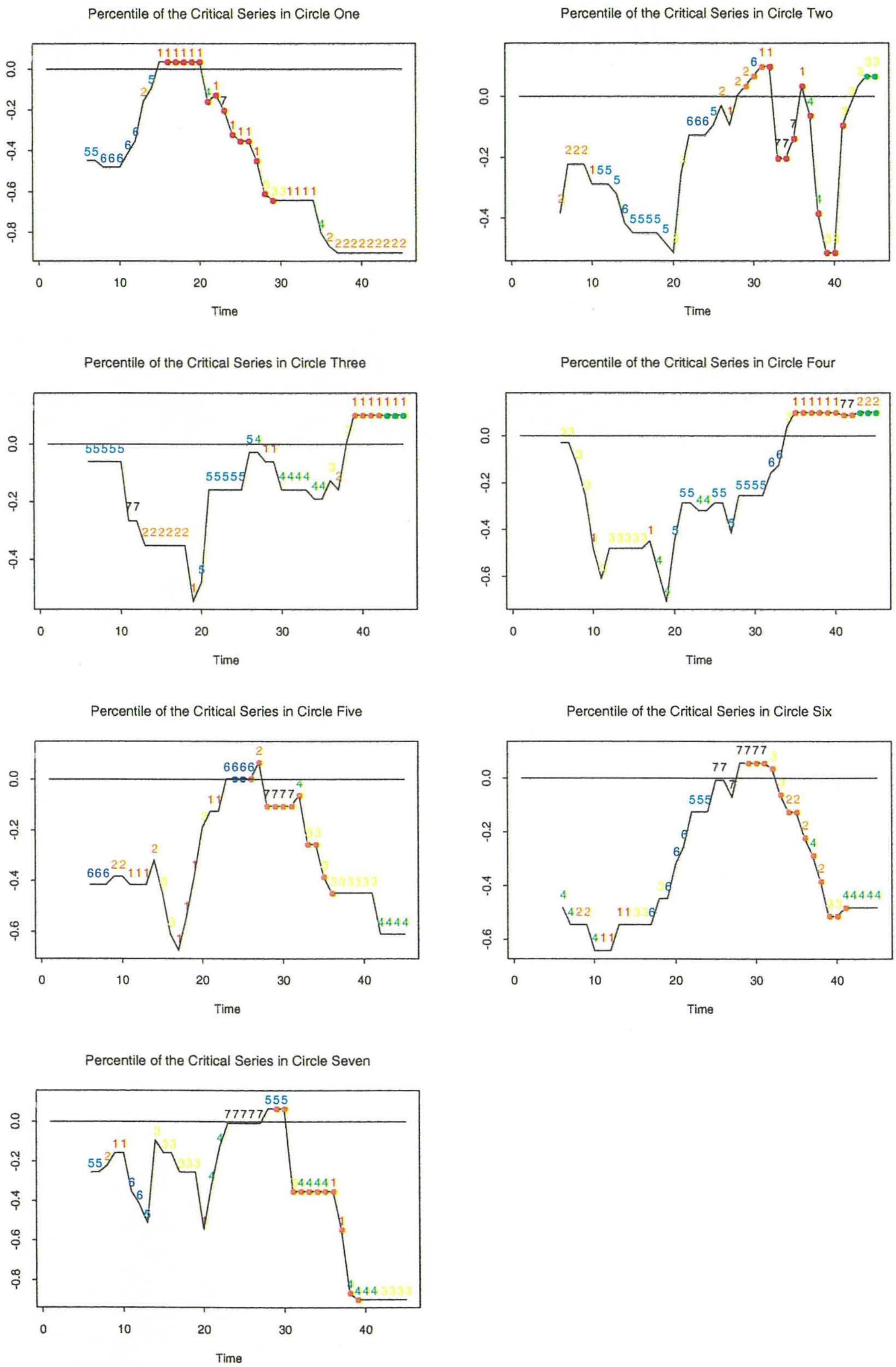

Smoothed M8 Series and Tips in Circle Three
Series 1
Series 2
Series 3
Series 4
Series 5
Series 6
Series 7
Max events and TIP level
1975
1980
1985
1990
1995
Time
0.1
0.9

Figure 5.5


Percentile of the Critical Series in Circle One
Percentile of the Critical Series in Circle Two
Percentile of the Critical Series in Circle Three
Percentile of the Critical Series in Circle Four
Percentile of the Critical Series in Circle Five
Percentile of the Critical Series in Circle Six
Percentile of the Critical Series in Circle Seven
Time

5.3.2 Predictions targeted at magnitude 6.5 earthquakes

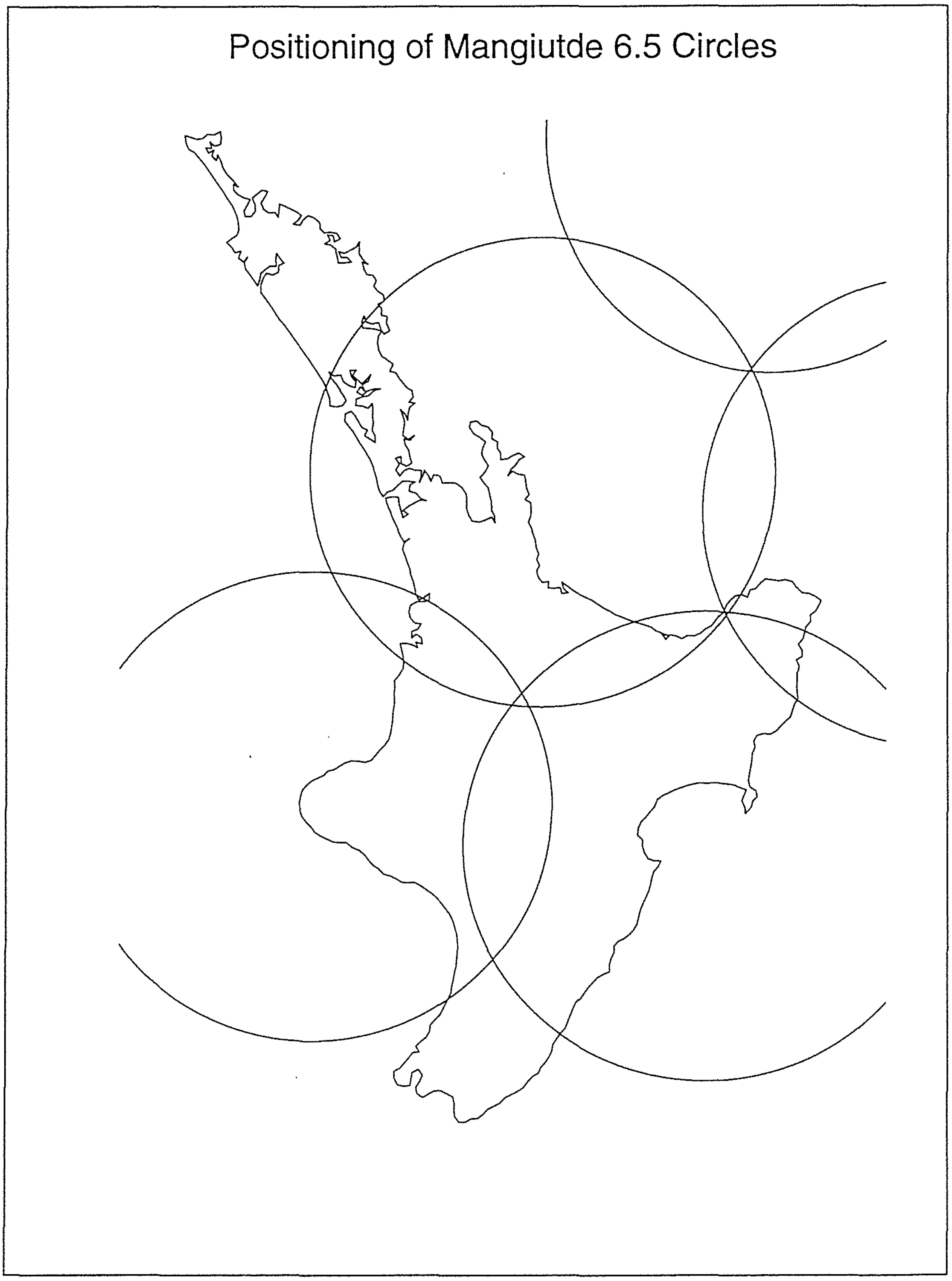
Patterns observed one circle can be checked against those observed another. If the pattern appears in all the circles, then that can be said to reflect on the nature of the M8 algorithm. In a sense, the different circles provide replication of the behaviour of the algorithm. The replication is far from ideal, given that the geological conditions vary greatly in the different circles. However, because the records do not go back very far in time replication in space is the only option. With the 7 circles for the analysis of magnitude seven predictions, there is not enough replication. However, the circles for predictions targeted at small magnitudes are smaller. This means that, potentially, more of them can fit over the length and breadth of New Zealand so there will be more replication of the behaviour of the M8 series. This is part of the rationale for predicting earthquakes of magnitude 6.5 with the M8 algorithm.

Another advantage of predictions targeted at smaller magnitudes is that there are more events to predict. With only two magnitude 7 earthquakes occurring during the time when the series are calculated there is an acute shortage of replication of the way the series behave when an earthquake of the target magnitude is imminent. There are more earthquakes of magnitude 6.5, providing replication of the way M8 Series behave when an earthquake of the target magnitude is approaching.

Unfortunately, using magnitude 6.5 did not provide more circles as replication. This was because there were not enough earthquakes in the necessary magnitude ranges to provide enough data. Details of changes to the series when the target magnitude is changed can be found in Chapter 3, section 3.4.3.

5.3.2.1 Positioning of circles for Magnitude 6.5

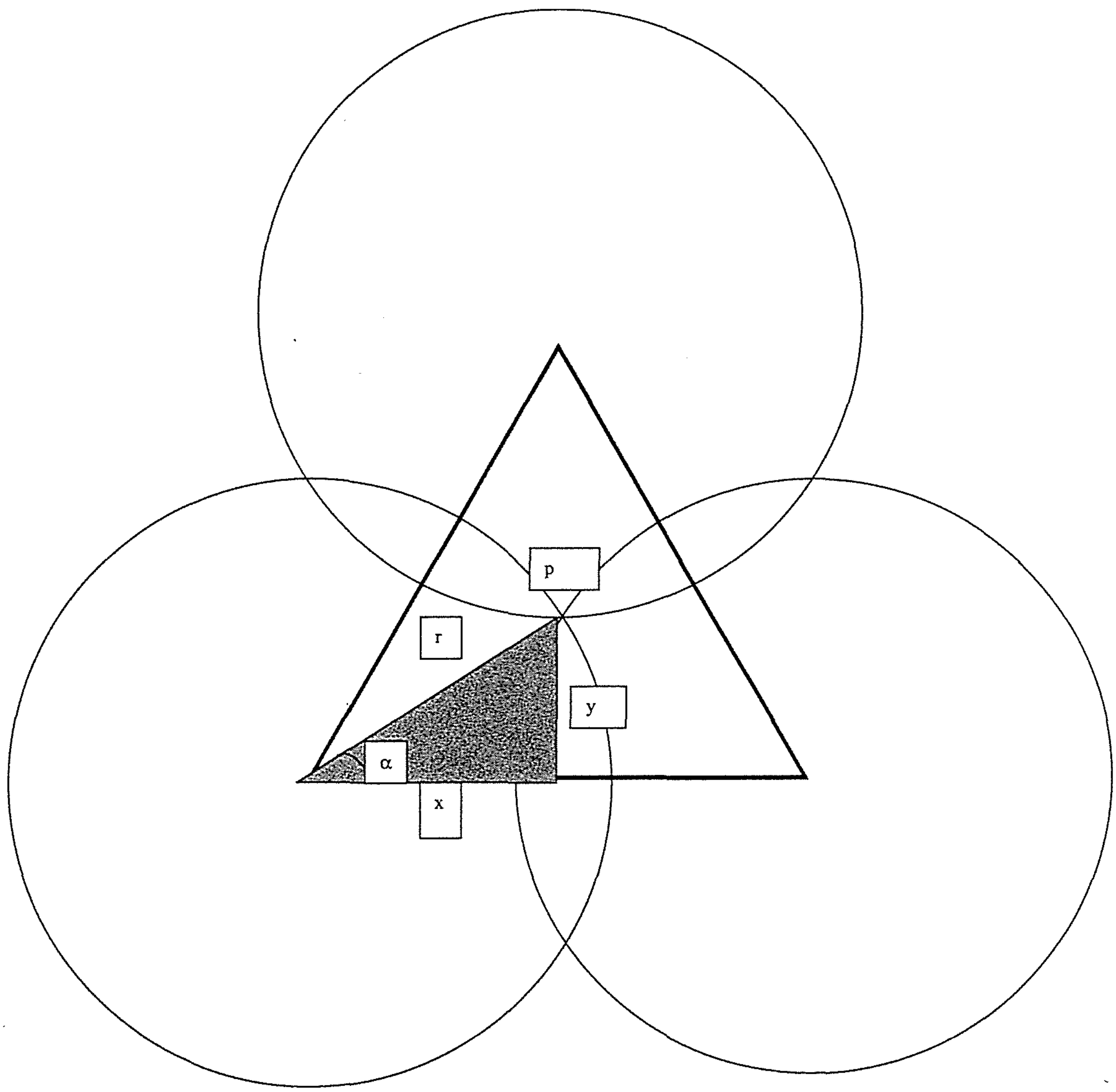
Figure 5.6



The diameter of the circles where earthquakes with magnitudes of 6.5 or more are targeted for prediction is 384 km. The angle of the column of circles in figure 5.6, above, was chosen so that the column would roughly follow the subduction zone, and so that the calculations would be simple. With this angle for the column of circles, the shift to the bottom left-hand circle to the bottom right hand circle in figure 5.6 below does not involve a shift upwards. This means there is one less distance to calculate. The position of the column over New Zealand was experimented with. The position that resulted in the most earthquakes within the column was chosen.

Note that any three adjacent circles intersect at a common point. The positioning is based around this point. Let us just consider three circles, intersecting at a common point, p, as in figure 5.7 below. If the centres of these three circles are joined together they form an equilateral triangle, with the intersection point of all three circles as the centre.

Figure 5.7



The distance of the corners from the centre of the triangle is a equal to the radii of these circles. The bottom right angle triangle that is shaded is the first triangle used to calculate the distance of the circle centres from each other. The horizontal side of the right angle triangle, x, is given by r cos α. The distance 2x gives the horizontal distance of the bottom right hand triangle from the bottom left-hand triangle. There is no need to move up to vertically. To get to the location of the centre of the top circle, it is necessary to move horizontally the distance x, and vertically the distance y + r, where $y = r \sin \alpha$.

5.3.2.2 The results of setting the magnitude targeted for prediction to 6.5

The amount of total space-time that was occupied by TIPs was 31%. There were 11 earthquakes of the target magnitude during the series generation period, which is 22.5 years for each circle. The percentages of total space-time for individual types of TIPs are shown in table 5.2 below.

Table 5.2 Results of Setting the Target Magnitude to 6.5

Success Criterion	Statistics
False TIPs	9%
Near successful TIPs	14%
Successful TIPs	3%
Continuing TIPs	2%
Number of failures to predict	6
Total number of earthquakes with magnitude ≥ 6.5	7

Setting the target magnitude to greater than or equal to 6.5 did not result in much success. This is surprising, considering that the false TIP rate is greater for Magnitude 6.5 than for magnitude 7. In Circle 1 a successful TIP is very nearly declared – the TIP level series was 0.029 less than zero in the critical place. The TIP level series needs to be above zero twice in a row for a TIP to be declared. Therefore the 'critical value' of the TIP series during the 5 years before an earthquake is the maximum of the values before and after the maximum value of the TIP level series in those 5 years. If this is greater than zero a TIP *is* declared, and the critical value of the TIP level series gives a measure of how close the series came to declaring a TIP. The critical values of the TIP level series before the 7 earthquakes of target magnitude are shown in table 5.3.

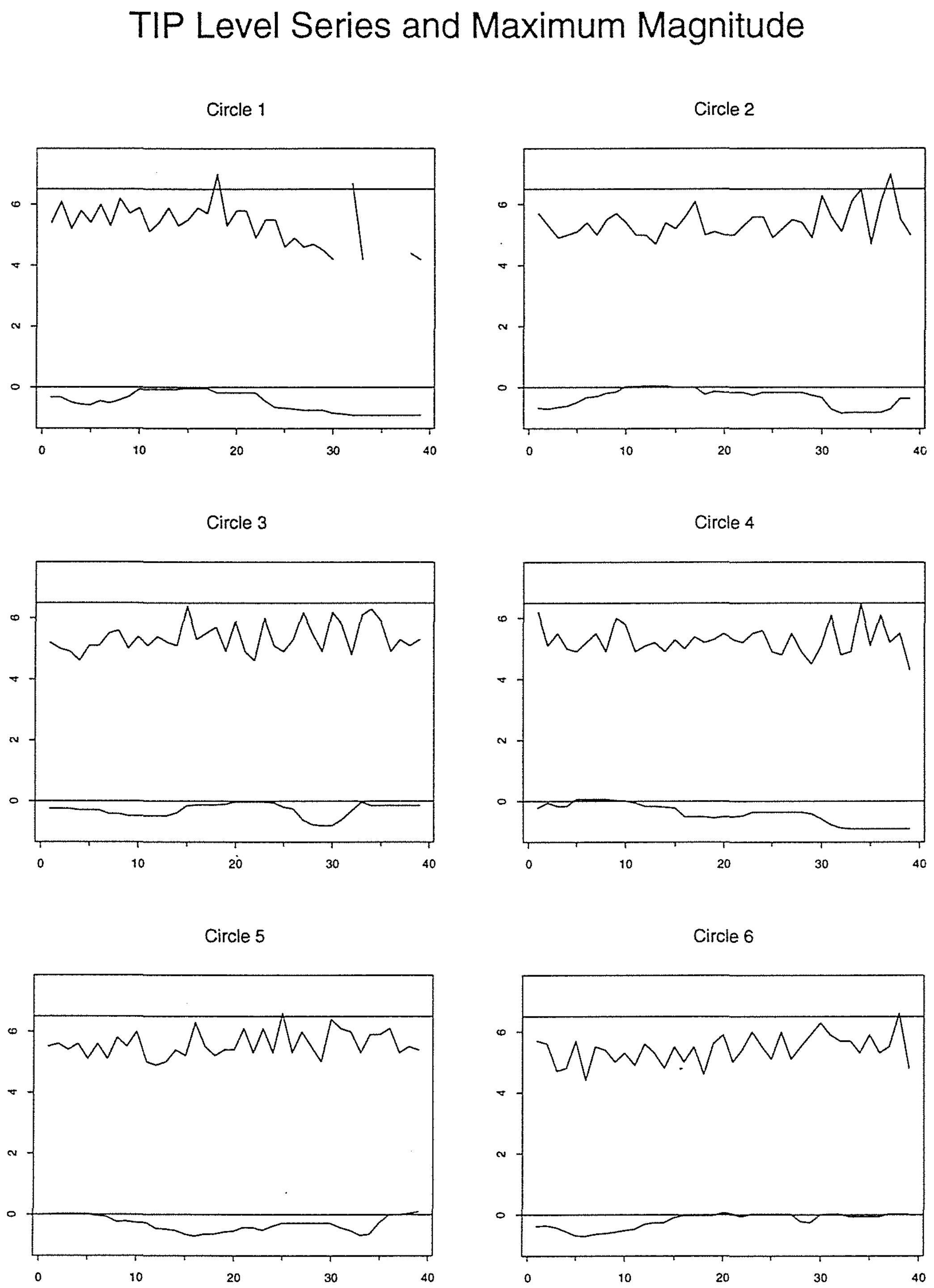
Table 5.3 Critical Values of the TIP Level Series

Earthquake	Critical Value of TIP Level Series
9 / 1985, M7, Circle 1	-0.03
12 / 1992, M6.7, Circle 1	-0.45
8 / 1993, M6.5, Circle 2	-0.16
2 / 1995, M6.5, Circle 2	-0.16
8 / 1993, M6.5, Circle 4	-0.35
5 / 1989, M6.6, Circle 5	-0.38
9 / 1995, M6.6, Circle 6	0.02

It can be seen from table 5.3 that the magnitude 7 earthquake in circle 1 was very close to being predicted. However, overall, the results are not impressive. If we round the critical value of the TIP to zero for the magnitude 7 earthquake in circle 1, that only gives 2 out of 7 earthquakes predicted. Given that 28% of space-time is covered by TIPs, randomly distributed TIPs could be expected to predict 28% of the earthquakes. This is the amount that have been predicted (the exact expected amount is 1.96).

Figure 5.8 below shows the TIP level series and maximum magnitude in each circle. If the TIP level rose to a peak during the time before an earthquake greater than magnitude 6.5 there would be some value in the TIP level series for this magnitude. However, the TIP series does not vary much, and does not show signs of rising before an earthquake, and falling when one is not imminent.

Figure 5.8


TIP Level Series and Maximum Magnitude
Circle 1
Circle 2
Circle 3
Circle 4
Circle 5
Circle 6
0
2
4
6
0
10
20
30
40

5.3.2.3 Differences in numbers of earthquakes of target magnitudes when predicting magnitudes 6.5 and 7

There are more earthquakes of M6.5, but the circles of investigation are smaller. If the number of earthquakes in each individual circle is less for magnitude 6.5 than for magnitude 7 then logistic regression by circle would still be based on one or two points. The following analysis investigates the gains made by using M6.5 circles.

The number of earthquakes decreases as one moves approximately south-west along the subduction zone. The aim of the test below is to see if having a smaller circle of investigation when predicting earthquakes ≥ magnitude 6.5 compensates for the increased amount of earthquakes in the prediction range. A reasonable estimate of differences in the number of earthquakes of target magnitude in the circles of differing size is best obtained with paired t-test, where the two circles are concentric. Whatever the direction of the trend of decreasing numbers of earthquakes, the inner circle will be a representative sample of the outer circle. This can be seen in figure 5.9, below

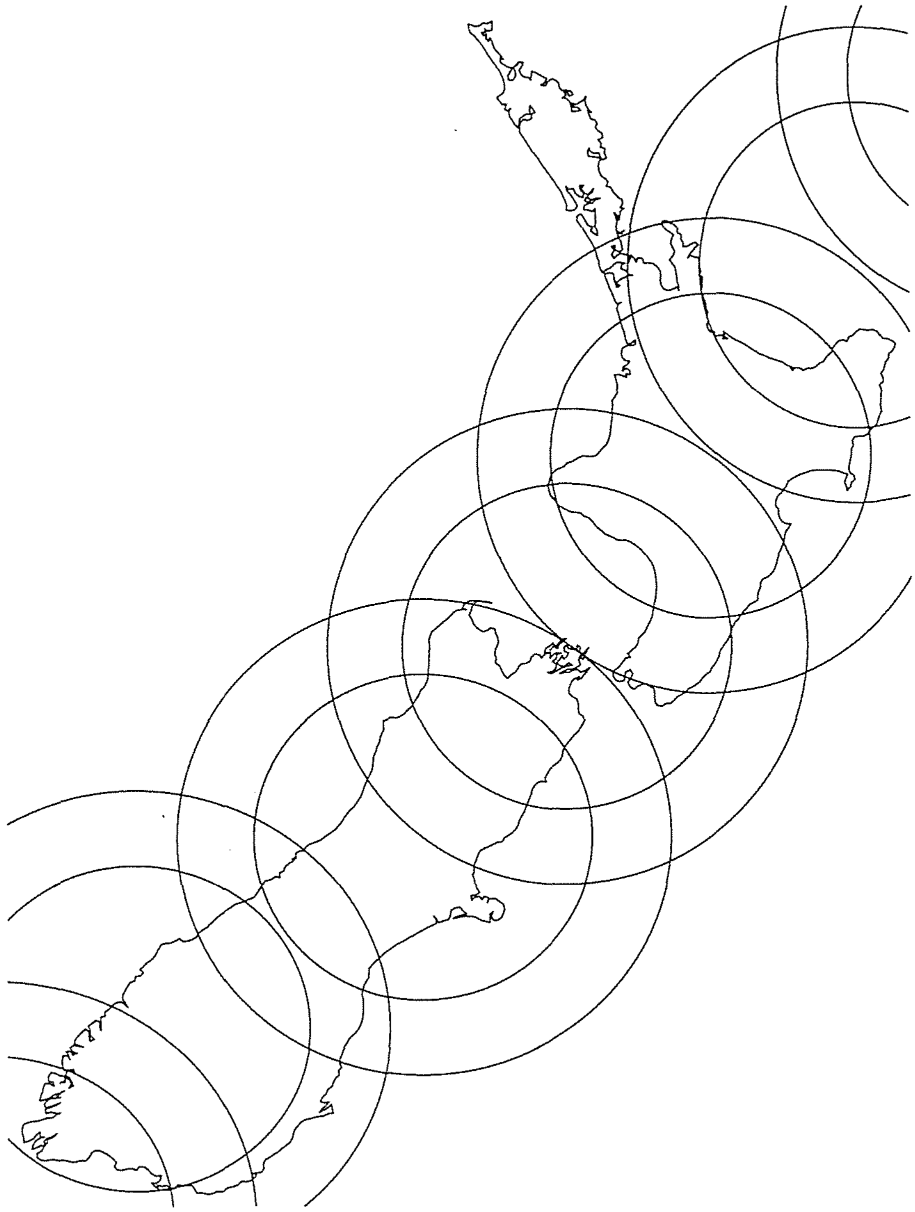
Even without doing a paired t-test it can be seen in table 5.4 below that there is a significant difference in the numbers of earthquakes of the target magnitude. To increase the power of the test, earthquakes in the training period, as well as the series calculation period of the algorithm, have been studied. The p value from the t-test is 0.0002.

Table 5.4: Number of earthquakes of the target magnitude in both the training period and series calculation period.

Circle	Magnitude 6.5, Circle radius = 384 km	Magnitude 7, Circle radius = 561 km
1	6	4
2	7	2
3	7	2
4	6	1
5	4	0
6	3	0
7	3	0

Figure 5.9

Concentric Circles for Testing Differences in Type 1 Error



5.4 The stability of M8 predictions with regard to circle position.

One of the difficulties with M8 is its tendency to ignore the existence of grey areas, and pretend that things are either black or white. It does this by dividing continuous variables and treating them as categorical. An example is the 'hard' boundary of the circle of investigation. An earthquake is either inside the boundary, or outside the boundary. But it seems reasonable for earthquakes close to the circle's boundary to affect the interior of the circle.

5.4.1 Testing the Stability of M8 when the Circle of Investigation is Moved Slightly

The stability of the M8 algorithm series was tested by observing changes in the series when the coordinates of the circle were varied are by half a degree in every direction. This was done for each of the 7 circles and using the predictions for magnitude 7. Magnitude 7 series were chosen as there is no evidence that the M8 algorithm has any predictive power when magnitude 6.5 is the target magnitude. The stability of the critical series, and series 5, 6 and 7 were examined. Series 1-4 only involve the number of earthquakes, and so cannot be affected by a large earthquake outside the boundary very much. For series 5 and 6, the differences caused by the perturbations of half a degree are hardly noticeable. However for series 7 there are differences caused by the perturbations. This can be seen in figures 5.10 and 5.11 below. Figure 5.10 shows the circle with the maximum stability, and figure 5.11 shows the circle with the minimum stability. Series 7 is defined by only one earthquake and when that earthquake moves out of the circle or a new one moves in the value of series 7 is radically altered. It can be seen how earthquakes are excluded as the circle moves in a particular direction. For example, in figure 5.11 (the circle with minimum stability) an earthquake in 1985 with a large number of aftershocks (causing series 7 to cross its threshold) disappears in the northern most and Eastern most circles. Therefore predictions resulting from series seven crossing its threshold are unstable. Series 7 was graphed for all the circles, and it was found that the threshold itself is also unstable.

Figure 5.10

Peturbations of the Circle of Investigation for Series 7, Circle 2

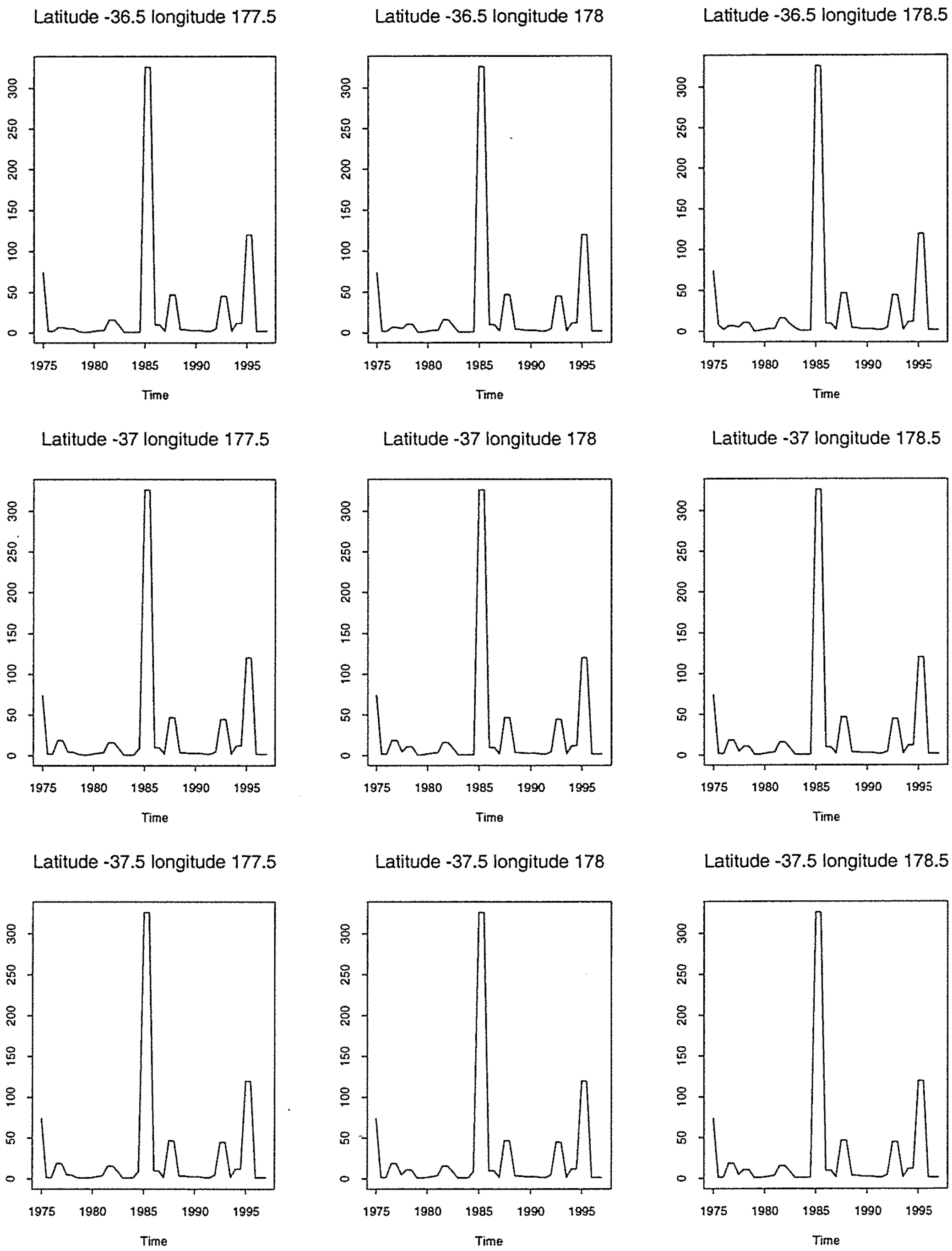
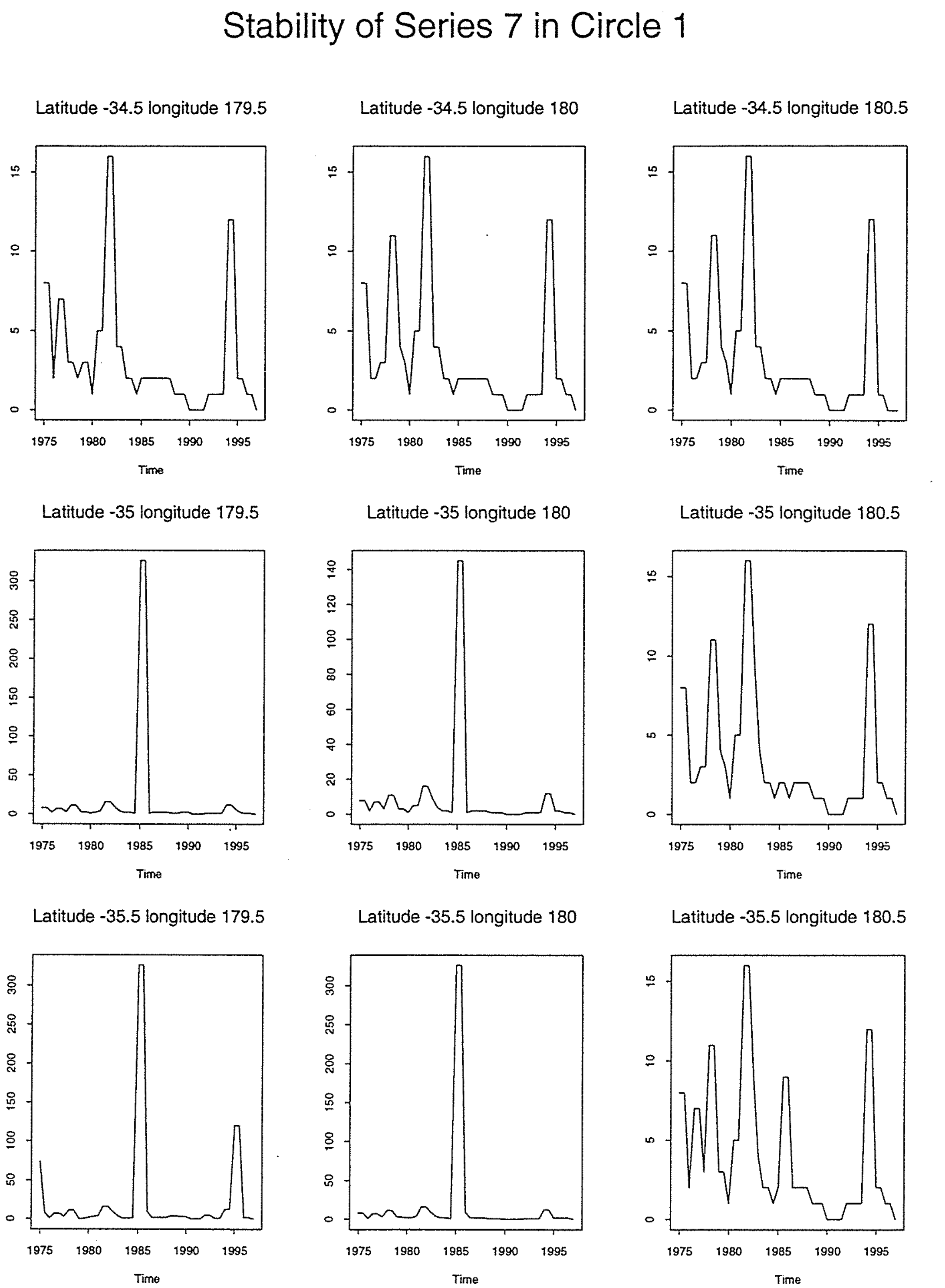


Figure 5.11


Stability of Series 7 in Circle 1
Latitude -34.5 longitude 179.5
Latitude -34.5 longitude 180
Latitude -34.5 longitude 180.5
Latitude -35 longitude 179.5
Latitude -35 longitude 180
Latitude -35 longitude 180.5
Latitude -35.5 longitude 179.5
Latitude -35.5 longitude 180
Latitude -35.5 longitude 180.5
Time

5.4.2 Discussion of the Results of the Stability Test

There are two attributes of the earthquakes that can result in instability of the predictions when the circle of investigation is moved. One is clustering of the earthquakes in space, and the other is scarcity of earthquakes. If the earthquakes are clustered there will be a significant loss of data when the cluster moves outside of the circle. If the earthquakes which are the major input into the series are quite sparse (for example, large earthquakes or earthquakes that generate large numbers of aftershocks) this will affect the stability of the predictions when the circle is moved. As the circle moves and an earthquake is excluded it is unlikely that an earthquake will be included at the same time to compensate.

The scale of clustering in the New Zealand earthquakes seems too large to affect the stability of the predictions. Series 1 through 4 do not have the magnitude of the earthquake or numbers of aftershocks as input. Consequently they are very stable when the circle moved, as there are lots of earthquakes going in to its calculation. Series 5 and 6 are also stable when the circle is moved. This is probably because the large earthquakes do not make up significant proportion of the total energy release. Also, this energy release is multiplied by the 2/3 power of the number of earthquakes, so the strain release of the large earthquakes is not a big part of series 5 and 6. Series seven is unstable, which is hardly surprising as it derives its value from only one earthquake. When that earthquake moves outside of the circle radical change in the value of series 7 can be expected. To compound this, earthquakes with very large numbers of aftershocks are scarce. This means that when an earthquake with a large number of aftershocks is moved outside of the circle, the next value for series 7 is likely to be quite low

The fact that the thresholds for series 7 are unstable can also be attributed to the unstable nature of series seven. As discussed in chapter 3 the empirical percentiles of each series are estimated from a training period that it overlaps with the part of the series 7 that is plotted. For this reason it would be advisable to set a minimum value for the threshold so that if the training period does not contain a spike the threshold will not reflect the background noise level of series seven. An alternative or additional strategy would be to base the threshold on a much larger area than the circle of investigation. The default M8 series themselves are very noisy and are subject to two lots of smoothing - this is an extra argument for also 'smoothing' the thresholds for the series.

The real test of overall stability, however, is having a stable shape for the TIP Level series. It is not at all times that series 7 is unstable, only at times when there is an earthquake with large number of aftershocks (and series 7 is therefore crossing its threshold). Based on the small sample of the seven circles for magnitude 7, series 7 has a tendency to be the critical series when it is crossing its threshold — the very time at which it is unstable. This results in instability in the TIP level series at times when series 7 if the critical series. Overall, however, the TIP level series is reasonably stable. This can be seen in figures 5.12 and 5.13, below. Figure 5.12 shows the most stable circle and figure 5.13 shows the least stable one. The most and least stable circles are not the ones that are the most and least stable with regard to circle 7. This is because the ones that were the most and least stable with regard to series 7 do not have series 7 as the critical series at all.

Figure 5.12

Stability of the TIP Level Series, Circle 2

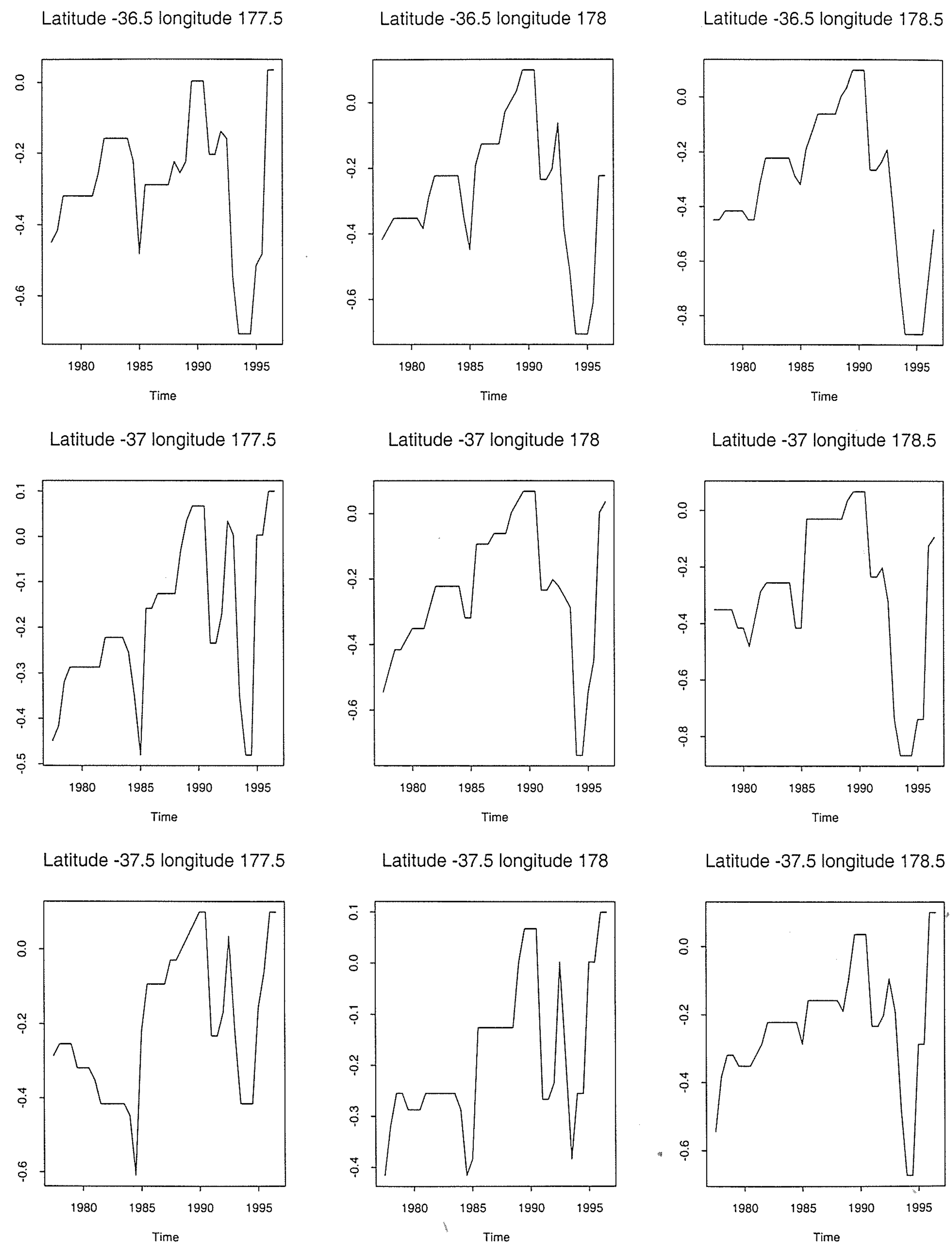
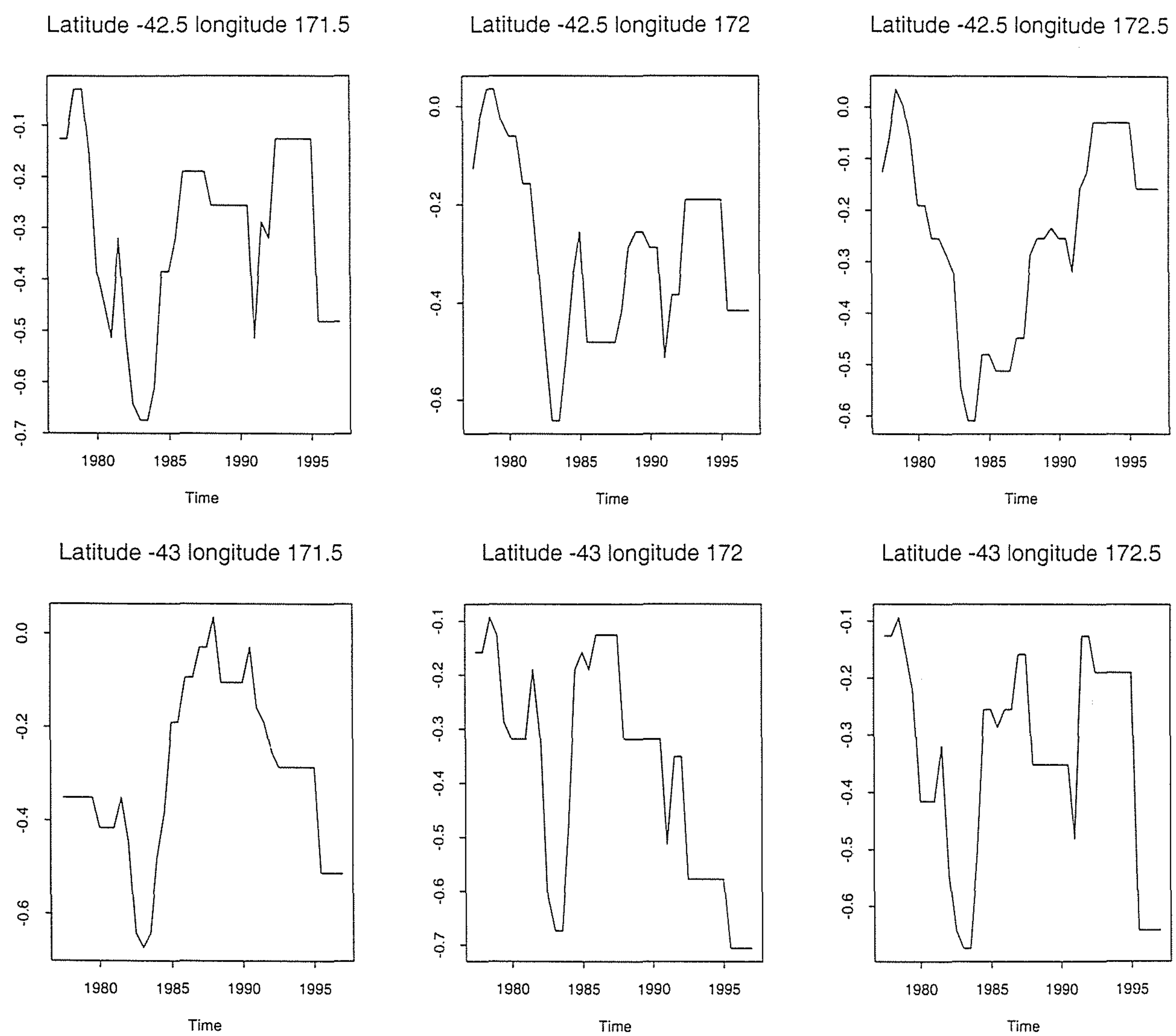


Figure 5.13

Stability of the TIP Level Series, Circle 5



5.5 Conclusion

In New Zealand, the question of what depth of earthquakes is appropriate is a significant one. The algorithm will not work unless all depths are included — but with only 7 circles for predicting magnitude 7 earthquakes more data is needed to assess whether including earthquakes of all depths is appropriate. Unfortunately, there is no evidence that using the M8 algorithm for predicting earthquakes of magnitude 6.5 is viable.

Chapter 6 ~ Exploratory Analysis using Principal Components

6.1 Introduction

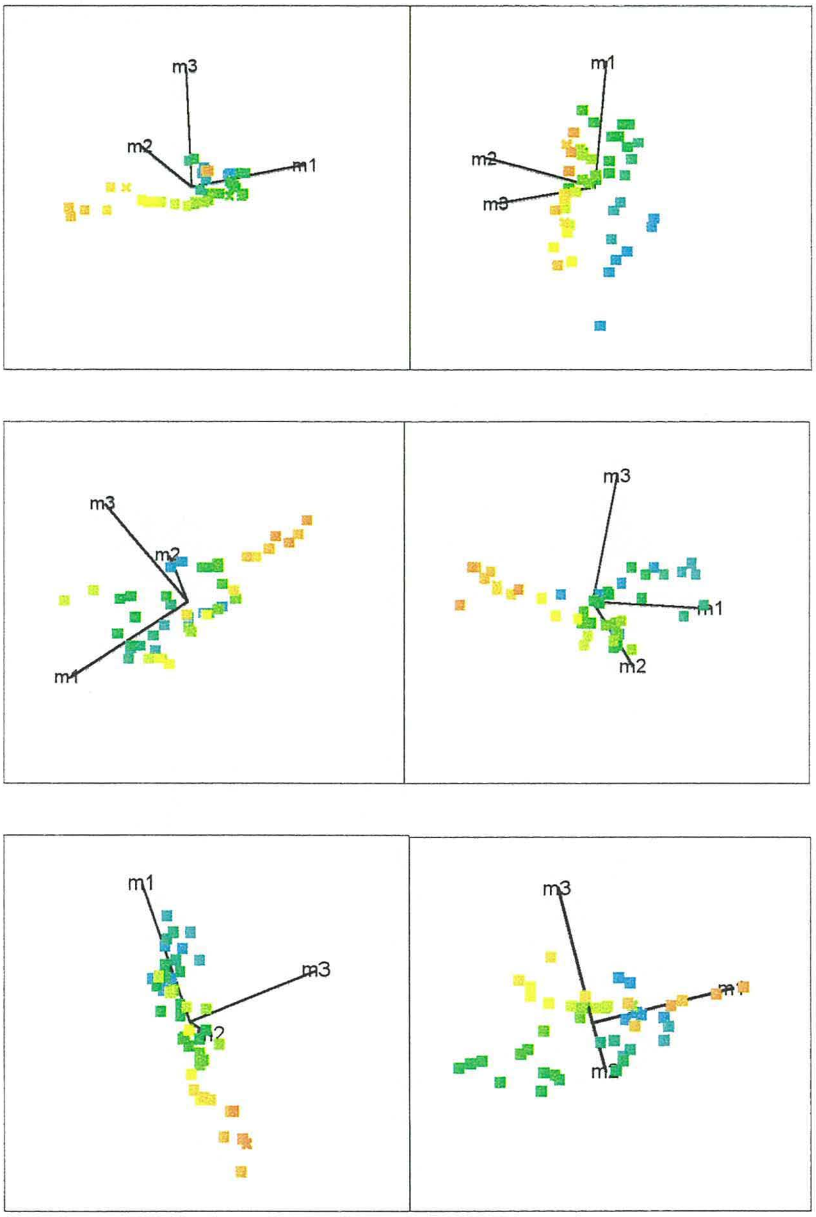
Despite correlation between the first six series none of them are redundant (in the sense of being the 'critical series'). However, there is still redundancy in the sense that they are highly correlated. Another attempt has been made to condense the series, this time using Principal Components. One aim in reducing the number of series to consider is to allow graphical analysis of the series. The other aim of using principal component analysis is to gain a better understanding of the nature of the M8 series.

The series were standardised because of their differing units and standard deviations. Here, a 3 year maximum is not taken, so more variation can be seen. Sections 6.2 and 6.3 discuss plots of the principal components, and 6.4 their analysis.

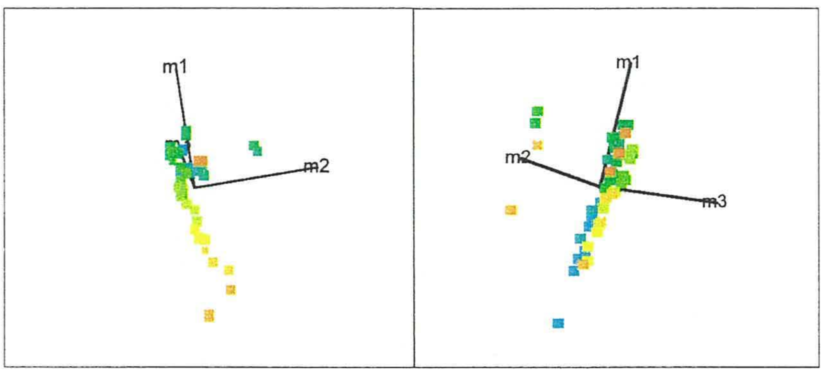
6.2 Three dimensional spin plots

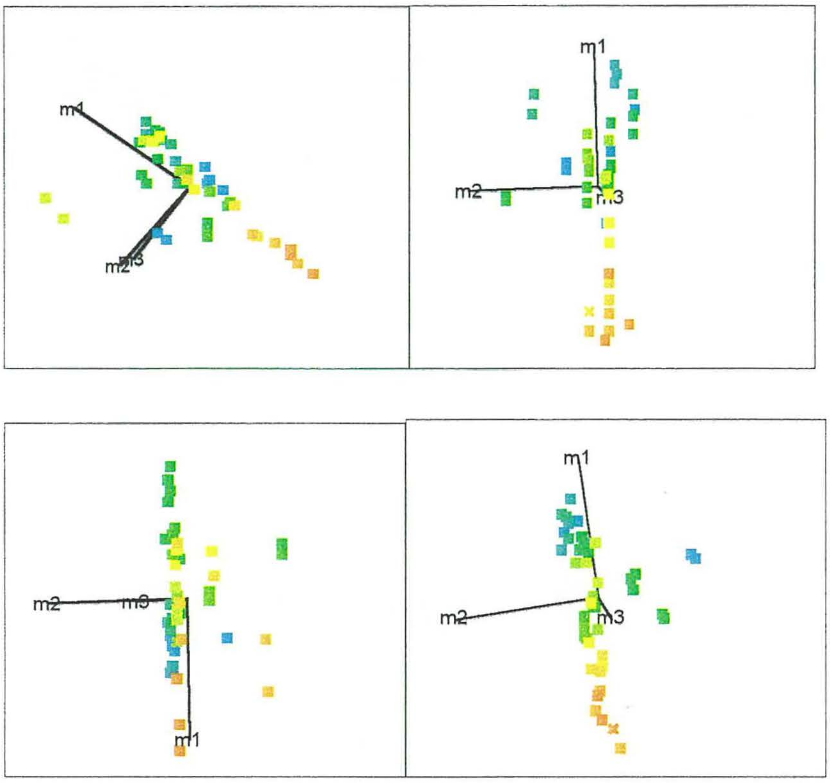
The first three principal components of the seven series were plotted in three dimensional spin plots. These plots were produced using the program Data-Desk (DataDescription Inc., 1997). Most often variation can be summarised with three principal components so most of the information contained in the series can be seen graphically in three dimensional spin plots (eigen-values of principal components can be seen in tables 6.1 – 6.3). There were two aims in creating the spin plots. The first was to look for patterns in the series connected to seismic activity that could be followed up in greater detail. The second was to see if the same pattern recurs in different circles. The colour gradation in the plots shows time and the Crosses indicate where an earthquake of magnitude 6.5 or more has occurred.

Figure 6.1 View of Spin Plots Emphasising the 'Snaky Line' Pattern



View Of Spin Plots Emphasising Outliers





In the snapshots of the spin plots shown above, the most striking feature is the snaky line that the points lie in. Two angles are shown: one is to emphasise the 'snaky line', the other is to emphasise the unusual points. The unusual points are due to series 7. As mentioned in chapter 5, series 7 consists of great bursts of aftershocks and background noise. Most circles have a principal component that is mainly series seven. The circles that show the 'snaky line' to the greatest degree are the ones with the largest earthquakes (circle 1 and circle 2 , the North East and Middle East circles). Spin plots were also drawn for magnitude 7, and a showed the same pattern. For most spin plots, the points lie in a line or in plane (this is hard to see in a static snapshot of the spin plots) which indicate that most circles only need 1 or 2 principal components to explain most of the variation. This is confirmed in the principal component analysis of section 4.

6.3 'Snaky Lines' in Seismically Active Circles

The 'snaky line' mentioned in section 6.2 is caused by auto-correlation and, less obviously, trends in the original series. Series 1 through six of the M8 algorithm have a running total over six years. Therefore, there is strong auto-correlation artificially introduced. This contributes to the 'snaky line' pattern, as the PC scores change slowly over time.

Trends in series 1 through 6 affect the 'snaky line' pattern, giving the series a large standard deviation. The series were standardised to perform principal component analysis for the reasons mentioned in section 6.2. This was done by taking away the mean and dividing by the standard deviation. The standard deviation measures the distance of individual observations from the mean, as well as the scatter about the trend line. For this reason, series with a trend have a very large standard deviation.

In the standardised series, the only part that changes with time is the numerator (containing the original value of the series). The denominator (the standard deviation) stays constant. If the denominator is large compared to the numerator, changes in the value of the original series will have only a small impact on the standardised series. This will make the standardised series smoother, contributing to the 'snaky line' pattern. Also, because of the trend the series are heading in a consistent direction, also contributing to the 'snaky line' pattern.

Conversely, series with almost no trend do not have a large denominator for the standardised version of the series. Most of the values for the series will be simply noise, and they will jump around a lot as there is no consistent direction that the values are progressing in.

6.4 Form of the Principal Components

Another advantage of performing principal component analysis is that the form of the principal components is informative. Principal components analysis was performed on magnitude 7 circles as well as magnitude 6.5 circles. This was because of the difficulties experienced with magnitude 6.5 circles referred to in chapter 5 section 5.3.2. Because of these difficulties, it is the circles for predicting earthquakes of magnitude 7 that are focussed on here. The loadings for the Principal Components are shown in tables 6.1 to 6.3.

The general form of the Principal Components is that the first principal component is a weighted average of the first six series, and either the second or third is the 7th series. The 'average' of the first six series is an average in the sense that the first six series have roughly the same coefficients. This is in and line with the M8 decision rule; part of it says that the first six series must cross their thresholds, implying that series 1-6 are of the same character.

For a TIP to be declared series 7 must also cross its threshold. This implies that series 7 is of different nature to the other series. Series 7 has a principal component all to itself, confirming M8's implication that series 7 is of a different character.

The 'basic form' mentioned above (where PC 1 is a weighed average of the first 6 series, and PC 2 or PC 3 is just series 7) is only the general pattern, and there are exceptions to it. Hereafter, this 'basic form' will be referred to as the 'M8 pattern'. In circle 5, the weights for each series differ a bit. In this circle the first six series are not so correlated. However, this is the only exception for the form of the first principal component. The general trend for the second principal component to be simply series seven is more variable. Series 7 was predominantly the second principal component in circles 1, 4, 6 and 7 and the third in circles 2 and 5.

The remaining component of the first three was different for every circle, representing the individual seismic characteristics of each circle. It can be seen that series 5 and 6 often have similar loadings. This indicates that they are highly correlated, which is of course expected. What is surprising is that the other pairs of series are not more correlated. The fourth, and higher, components showed no common pattern. Since characteristics that are specific to each circle are no use for general earthquake prediction, they are not described further.

Table 6.1 Loadings for Component One

Series.	Circle 1	Circle 2	Circle 3	Circle 4	Circle 5	Circle 6	Circle 7
% explained	0.81	0.59	0.66	0.78	0.47	0.71	0.72
1	0.408	0.451	0.442	-0.406	-0.407	-0.429	0.408
2	0.41	0.471	0.424	-0.421	-0.341	-0.425	0.43
3	0.396	0.247	0.407	-0.408	-0.455	-0.371	0.401
4	0.414	0.41	0.332	-0.402	-0.443	-0.416	0.427
5	0.415	0.419	0.412	-0.41	0.36	-0.417	0.414
6	0.405	0.393	0.401	-0.398	0.398	-0.372	0.364
7		0.121	0.137		0.164	0.113	

Table 6.2 Loadings for Component 2

Series	Circle 1	Circle 2	Circle 3	Circle 4	Circle 5	Circle 6	Circle 7
% explained	0.14	0.20	0.17	0.14	0.23	0.14	0.14
1		0.204	-0.115		0.337		
2			-0.139		0.427		
3		0.679	-0.235		0.12	-0.158	
4		0.29	-0.464		0.241		
5		-0.418	0.308		0.565		
6		-0.479	0.342		0.509		
7	0.991		0.696	-0.991	0.231	-0.98	0.992

Table 6.3 Loadings for Component 3

Series	Circle 1	Circle 2	Circle 3	Circle 4	Circle 5	Circle 6	Circle 7
1	-0.412	0.017	-0.061	0.297	0.132	0.238	0.349
2	0.327	0.111	0.053	0.137	0.012	0.078	0.086
3	-0.673	0.029	0.11	0.425	0.008	0.495	0.433
4	0.058	0.163	0.48	0.238	-0.157	0.257	0.17
5	0.251	-0.012	-0.352	-0.496	-0.215	-0.452	-0.428
6	0.436	-0.027	-0.374	-0.63	-0.209	-0.648	-0.676
7	-0.119	-0.979	0.697	-0.114	0.931	-0.038	-0.114

Table 6.4 Percentage of Space Time Taken Up By Various Types of TIPs

	Circle 1	Circle 2	Circle 3	Circle 4	Circle 5	Circle 6	Circle 7
STIP(-)	0%	27%	20%	27%	0%	27%	27%
CTIP	0%	12%	0%	4%	0%	0%	0%
FTIP	0%	0%	25%	0%	0%	0%	6%
c.e	0%	0%	0%	0%	0%	0%	0%
Total	0%	39%	45%	31%	0%	27%	35%

Table 6.5 Factors Affecting Whether the M8 Pattern is Shown in the First 2 PCs

Circle	1	2	3	4	5	6	7
Mean correlation of S7 with S1-6	-0.20	0.09	0.25	-0.08	-0.03	-0.19	-0.02
Shows M8 Pattern	y	n	n	y	n	y	y
Variance explained by PC 1(%)	0.81	0.59	0.66	0.78	0.47	0.71	0.72

Ironically, the 'M8 pattern' should not appear so strongly in circles where TIPs are declared. If there is a TIP then series 7 should peak when the other series do, causing correlation between series 7 and the other series. This would then mean that series 7 would be included in PC 1, rather than being separate in PC 2. However, in a circle where there is a TIP series 7 may also peak at a time when the other series are at a low, destroying the correlation between series 7 and the other series. Therefore, it is best to look at the actual correlations between series 7 and the other series. The percentages of space-time taken up by various types of TIPs are shown in table 6.4. e mean correlation of series 7 with the other series in each circle is shown in table 6.5, above.

Unfortunately there is not enough data from the 7 circles for predicting magnitude 7 earthquakes to back up the explanation of how the M8 pattern comes about. However, the results are in the right direction. Circles where the mean correlation of series 7 with the other series is positive do not show the M8 pattern in the first 2 PCs. There is no overlap of this mean correlation between the circles with and without the M8 pattern in the first 2 PCs. However, a t-test for a difference between the two groups gives a p-value of 0.092, giving no clear evidence for a difference. There is not much statistical power in the test with only seven data points, and the smallest difference that can be detected with 95% confidence is only 0.293. The difference between the M8 pattern circles and the no M8 pattern circles for the mean correlation of series 7 with other series is 0.226.

The percentage of variance explained by PC 1 is high for circles showing the M8 pattern in the first 2 PCs. If the first 6 series are correlated with each other a linear combination of them will have a very large variance, so this linear combination will be the first PC. Then, most of the total variance will be accounted for, and all that is left would be attributable to series 7. As far as there being a significant difference in the percentage of variance explained by PC 1 for circles showing the M8 pattern, the situation is the same as for the mean correlation of series 7 with other series. There is no overlap between the 2 groups, but the p-value for a difference in means is only 0.095. The minimum difference that can be detected with 95% confidence is 0.26 and the difference between the means is 0.19.

As mentioned above, the analysis was done with magnitude 6.5 circles as well. The results were similar: the same 'M8 pattern' emerged, for the same reasons.

6.5 Conclusion

Unfortunately, graphical analysis using principal components has not resulted in an insight linking an attribute of the M8 series to coming earthquakes. However, it has it has shed more light on the nature of the M8 series, through analysis of the form of the principal components. These insights will be used in the following chapter when fitting linear models using the M8 series.

Chapter 7 ~ Linear Models using Modified M8 Series

7.1 Introduction

There are two reasons for fitting linear models to M8's series. The first is that it allows some of the discrete variables to be replaced with continuous ones, where it is advantageous to do so. Many variables that are continuous by nature have been treated as discrete variables in the M8 algorithm. For example, the M8 series are continuous indication but if the series are slightly below the threshold for declaring a TIP no warning is given. Using logistic regression on M8's seven series would have the advantage of giving an estimate of the probability of an earthquake, rather than a warning. Regressing the maximum magnitude over a 6 month interval side-steps the problem of somewhat arbitrarily setting the target magnitude for prediction. Also, the information about magnitude size is not lost.

This chapter examines whether linear combinations of the 7 M8 series are more successful the than combinations of maxima and minima used by the M8 algorithm. The models fitted here have not managed to out-do the standard M8 procedure, although comparing them is a little like comparing apples with pears. However, the process of fitting the models has shed light on the way the series are related to the earthquake targeted for prediction, suggesting models that may be more successful.

All models were fitted using S-Plus (Math Soft Inc, 1995). Linear models were fitted to predict the magnitude of the largest earthquake in the next year. The seven M8 series were used, unsmoothed and computed on an annual basis. In a separate model the TIP level series and circle were used as predictors, with a 6 year running maximum applied to the TIP level series. Canonical correlation analysis was also used, with the same variables and the depth of the earthquake being predicted added as a response variable. Logistic regression models for the probability of an earthquake in the next 5 years were also tried, with the same predictors as for the linear models.

7.2 Data

7.2.1 Response Variables & Modifications to the M8 Series and TIP Level Series

For all models, the response variable for time $t + 1$ is associated with the predictor variable for time t, as in practice readings for a time period are only available at the end of that time period, and relate to the future.

The strong smoothing of the series in the M8 algorithm cannot be applied to the series in the linear models (for the reasons described below) and so values of the series before smoothing were used. Linear models with the 7 M8 series aim to predict the magnitude and depth of the strongest earthquake in the next 5 years to keep in line with M8, which issues a warning for the next 5 years when the series indicate danger. This means that the response variable for the models can have the same value for up to 5 years. Therefore, for up to 5 years the residuals would be a constant minus a constant times the autocorrelated values of the series, and so will be autocorrelated themselves. For this reason the smoothing has been removed from the M8 series.

Eliminating the smoothing from the 7 M8 series gives an extra 9 years of data (a 6 year running total, and then a 3 year maximum of the running total was originally taken). Originally the series were computed so that there were 2 values per year. When plotted, these series seemed to consist wholly of noise, without any trend. Therefore the series were computed on an annual basis to smooth them without introducing autocorrelation. This reduced the noise in the series so that some trends were apparent.

The canonical correlation analyses have the same response and predictor variables as the linear models, except that the response variable 'depth' is added. The reasons for having these variables are the same as for the linear models.

However, for the linear models with the TIP level series, the 5 year period is built into the TIP level series instead of the response variable. In these models, a 6 year running maximum of the TIP level series is taken (6 years, since the TIP level series must exceed zero twice in a row for a TIP to be declared). This is much closer to the procedure used by the M8 algorithm, as when the TIP level series exceeds zero for one year the M8 algorithm declares a warning for the next 5 years. With the models involving the 7 M8 series this was not possible, as it is the combination of series at any given point in time that indicate the level of earthquake risk. The response variable for the logistic regression model for the probability of an earthquake has '1's (successes) for the 5 years preceding an earthquake of the magnitude targeted for prediction, and '0's (failures) elsewhere. If the smoothing was not taken out of the M8 series this would result in much greater autocorrelation of the residuals than in the linear models. This is because there are so many zeros in a row. The response variable is '1' for the 5 years up to and including an earthquake if the 7 M8 series are used, and '1' where there is an earthquake if the TIP level series is being used. Either way, there are many '0's in a row.

Series 7 was also transformed by taking its log. This was because it had some extreme outliers with high leverage. The values of the series at this point would have a very large influence, which could cause 3rd level interactions. Magnitudes were used in the response variable rather than the amount of energy released. Energy is equivalent to $10^{\text{[illegible]}*magnitude}$. As can be seen from the formula, 'energy release' tends to have more extreme values than magnitude. These extreme values could cause the same problems as the extreme values in series 7. Since the distribution of magnitude is already skewed, using energy (which is an exponential function of magnitude) would only increase the skewness.

7.2.2 Choice of Circle Size

Magnitude 6.5 circles were used in all the models initially. The advantages and disadvantages of using magnitude 7 or magnitude 6.5 circles are discussed in Chapter 6, section 6.5. For the same reasons as in Chapter 6, the magnitude 6.5 circles were rejected for use in linear models. Instead, the magnitude 7 sized circles were used but within them a magnitude 6.5 earthquake was counted as a success in logistic regression. The M8 algorithm defines an earthquake that occurs during a TIP that is half a point less than the targeted magnitude as a 'nearly successful TIP'. As discussed in Chapter 5 (section 5.3.2) during a TIP the probability of an earthquake greater than magnitude 6.5 also increases slightly.

7.2.3 Standardisation of Series Within Circles of Investigation

In the M8 algorithm 'increased earthquake activity' is defined relative to the history of the particular circle that M8 is being used in. This goes against what might be expected, that is, an area which has had high seismic activity in the magnitude ranges used by the M8 series should have more large earthquakes. The Gutenberg-Richter Law,

$$f_m = a10^{-bm}$$

where f_m is the frequency of earthquakes at magnitude m

b = a constant (different for each region, called the 'b-value')

a = a constant

m = magnitude

is the embodiment of this expectation, where the number of large earthquakes is related to the number of small earthquakes. That is, world-wide, 'the number of shocks increases about eightfold for each decrease of one magnitude step' (Eiby, p75), where b is approximately equal to 8. The M8 algorithm assumes that the number of large earthquakes is *not* related to the number of small earthquakes. It effectively standardises the series within a circle, so that the typical level of seismicity is not considered in the predictions.

In addition to the standardisation within circles, the number of large earthquakes which have occurred have a minimal influence on the predictions. If there are lots of strong earthquakes in a particular location, it would seem reasonable to expect more strong earthquakes there, but M8 does not work on this principle. There are only two ways earthquakes $\geq$ 6.5 contribute to the predictions, and one of these ways gives them negligible weight. The way in which they do have some weight is contributing to series 7. Since series 7 is the maximum number of aftershocks after an earthquake, it is likely to take on the value of the number of aftershocks of the largest earthquakes. Series 7 must cross its threshold for a TIP to be declared, so in this way the large earthquakes are given a lot of weight. However, only earthquakes with magnitude $\leq$ 6.8 can contribute to series 7 (when the target magnitude is 7). The other way in which earthquakes with magnitude ≥ 6.5 can contribute to the predictions is adding to the number of earthquakes used in series 1-4. However, since there are not many earthquakes ≥ 6.5 in magnitude their influence here is minimal. Also, they are not distinguished from the other earthquakes since there is no indicator of earthquake size in series 1-4. In series 5 and 6, earthquakes greater than M0 – 0.5 are not included.

The Gutenburg-Richter Law leads us to expect more strong earthquakes where there is high seismicity, but strangely, the magnitude distribution of New Zealand earthquakes is not exponential, as is expected from the Gutenburg Richter law. The law only applies to earthquakes approximately greater than magnitude 2 (Eiby, p 75) but even with earthquakes less than magnitude 2 excluded, the distribution of earthquakes in all circles is bell–shaped or skewed, but still rises to a peak and then falls. This is still true if aftershocks are not included in the distribution (where aftershocks are defined by M8's declustering routine, see Chapter 3, section 3.2). The Kolmogorov-Smirnov Goodness-of-Fit Test was used to test if the distributions were exponential, and the null hypothesis (that the distributions were exponential) was rejected for raw and declustered data in all circles, with a p-value < 0.0001. Unexpectedly, the mean level of series 2, 3 and 4 is *negatively correlated* with the sum of the magnitudes over 6.5 in that circle. However, the correlation is dependent on two influential points from circles 1 and 5. The data for this is shown in table 7.2.

Table 7.2 Mean Levels of the Series and the Number of Earthquakes with Magnitude ≥ 6.5

	Series 1	Series 2	Series 3	Series 4	Series 5	Series 6	Series 7	Sum of
Circle 1	121	55	-15	-3	1173	1205	10	33.7
Circle 2	132	62	15	10	1129	1167	30	20.1
Circle 3	121	62	12	11	1135	1187	21	19.6
Circle 4	136	64	24	9	858	910	4	19.8
Circle 5	131	67	17	12	562	590	3	13.4
Circle 6	128	66	15	5	565	597	8	19.7
Circle 7	131	59	19	4	591	633	9	19.7
Correlation with Magnitudes								
	-0.573	-0.866	-0.879	-0.872	-0.600	-0.590	-0.077	
P-value	0.179	0.012	0.009	0.010	0.212	0.163	0.874	

In the context of running the M8 algorithm, an increase in the series is always an indication that the danger of a strong earthquake is increasing. A negative correlation of the level of the unstandardised series with total magnitudes over 6.5 supports M8's policy of standardising within circles. The Gutenburg-Richter law does not seem to apply in New Zealand which is also evidence against the number of large earthquakes being proportional to the number of small earthquakes.

The standardisation used medians and inter-quartile ranges, because of the asymmetric distributions involved. Ideally models would be fitted for individual circles for comparison with overall models, but the lack of data prevented this.

7.3 General Summary of Models

Table 7.3 summarises the models fitted. The canonical correlation analyses assess whether there is any relationship between the magnitude and the depth of the strongest earthquake in the next 5 years and the 7 M8 series. It is important to also relate the series to the depth of an earthquake, since this can greatly affect the earthquake's felt intensity. The linear models only aim to relate the magnitude of the strongest earthquake in the next five years to the 7 M8 series, as depth is left out of the equation. The logistic regression models estimate the probability of an earthquake with a magnitude ≥ 6.5 in the next five years, given the values of the series.

Table 7.3 General Summary of Models

Model	Response(s)	Predictors
Canonical Correlation with 7 M8 series	Running maximum of magnitude, and the associated depth	7 M8 series: annual standardised (by circle) unsmoothed all circles included in one model
Linear Models with 7 M8 series	Running maximum of magnitude	7 M8 series: annual standardised (by circle) unsmoothed all circles included in one model
Logistic Regression Models with 7 M8 series	1s 5 years before earthquake, 0s elsewhere	7 M8 series: annual standardised (by circle) unsmoothed all circles included in one model
Canonical Correlation Analysis with TIP level Series	Magnitude and depth of the strongest earthquake over a six month period.	Maximum value of the TIP level series over past 6 years, with the TIP level series begin computed bi-annually 'Circle' was tried but discarded.
Linear Model with TIP level series	Magnitude of the strongest earthquake over a six month period	Maximum value of the TIP level series over past 6 years, with the TIP level series begin computed bi-annually 'Circle 'as a blocking factor
Logistic Regression with TIP level series	Binary variable, with success being earthquake of target magnitude occurring within six month period	Maximum value of the TIP level series over past 6 years, with the TIP level series begin computed bi-annually 'Circle 'as a blocking factor

7.4 Canonical Correlation Analysis

As mentioned in Chapter 5, section 5.2.1, being able to predict the depth of an ensuing earthquake would enable a prediction of the felt intensity. This is what would determine the real danger to life and property, and so would be much more useful than a prediction of just the magnitude. To see if this is feasible, the magnitude and depth are correlated with the seven (unsmoothed) M8 series standardised by circle. In addition, depth and magnitude are correlated with the TIP level series.

7.4.1 Canonical Correlation Analysis Using the 7 M8 Series

For a correlation between a linear combination of the response variables and a linear combination of the predictor variables to be significant, the correlation needs to be quite large. The two linear combinations are created so that they have the maximum correlation. Even if there is no correlation between the groups of variables in the population under study, it is likely that some combination variables sampled will have a reasonable correlation. This is even more so when the sample size is small.

An overall canonical correlation was done pooling the data from different circles. The canonical correlations were 0.232 and 0.136. To test whether these are significantly different from zero, Bartlett's Chi-squared statistic is used:

$$\chi^2 = \{n - \tfrac{1}{2}(p+q+1)\} \times \sum_{i=1}^{r} \log_e(1-\lambda_i)$$

where n = cases of data

r = number of canonical correlations.

p = number of variables on the right hand side

q = number of variables on the left hand side

λ_i = canonical correlation i

This tests whether either of the canonical correlations is significant, and is compared to a χ^2 distribution with pq degrees of freedom. In this case the correlations are no-where near significant, with a p-value of 0.97. The canonical correlations were also calculated within each circle. Although both canonical correlations would need to be around 0.80 - 0.90 before they could be significant, if some circles had the same coefficients for their canonical variates that would be evidence that there was something real about canonical correlations. The first analysis which pooled data from the circles shows there is no wholly consistent pattern, but on the off-chance that there would be 2 or more different patterns the analysis was also done within circles. However, when the analysis was within circles every circle had a different set of coefficients for the canonical variates (and, not surprisingly, no canonical correlations were significant).

7.4.2 Canonical Correlation Analysis with the TIP Level Series

As with the 7 M8 series, the first test was done by pooling data from the 7 circles. The p-value for a test of whether the canonical correlation is zero is 0.447, indicating that there is no evidence of correlation between the TIP Level series and a linear combination of magnitude and depth. The canonical correlation was also done by circle, to see if the coefficients of the canonical variates were consistent for different groups of circles. However, they were not consistent, nor were any canonical correlations significant.

7.5 Models Using Series 1-7 Of M8

7.5.1 Regression Models Using Series 1-7 of M8

Circle was fitted as a categorical factor and a covariate, and was considered as a variable to interact with the series. There is a trend for seismic activity to decrease as circle number increases (heading approximately South East down the subduction zone). Therefore, 'circle' can be fitted as 'quantitative distance down the subduction zone' (fitted before any other variables, so it reduces the residual variation before they appear). However, it does not result in a significant reduction in the residual sums of squares (p=0.161) Fitting 'circle' as a categorical factor has a marginal effect on the residual sums of squares (p=0.09). Since the residual sums of squares are only reduced by 5% circle was not included as a categorical factor since it would not take out a reasonable amount of noise.

If an interaction term between the 'circle' and 'series' were significant, this would be very undesirable. A feature of the M8 algorithm is that it works universally. If the series×circle interaction were significant, a separate model would be needed for each location. However, while it would be interesting to find out if the series×circle interaction is significant, the benefits do not merit the cost. If interactions between circle and interactions between individual series (for example., circle×series 1×series 2) are not included then the number of parameters fitted goes from 64 (in the maximal model) to 100. Given the tendency for higher order interactions to be significant, it is likely that interactions between 'circle' and 2nd level interactions between series would also be significant.

Table 7.4 Specific Description of Models with the 7 M8 Series

Model Number	Model Description
1	All series, and 2nd and 3rd order interactions All series and 2nd order interactions All series
2	Only series 1, 3, 5 and 7, with all 2nd and 3rd order interactions Only series 1, 3, 5 and 7 with all 2nd order interactions Only series 1, 3, 5 and 7
3	a. Only series 2, 4, 6, and 7, with all 2nd and 3rd order interactions Only series 2, 4, 6 and 7 with all 2nd order interaction Only series 2, 4, 6 and 7
4	Linear combinations of series suggested by PCA in Chapter 6, namely: the sum of series 1-6 the sum of series 5-6 and series 7 3 Models were fitted, dropping 3rd and 2nd order interactions as above.
5	Only significant effects and main effects involved in interactions (and 2nd order interactions involved in 3rd order interactions). Significance depends on what effects have been entered before, because of high correlations between variables. Here, effects that were significant in model 1 are included.

Model 1a is intended as a benchmark, not as a sensible model, as only some of its terms are significant. Its residual sums of squares will be used below as a yardstick for the other models. Model 1b is also only intended for comparison with other models, as not all its terms are significant.

As mentioned in Chapter 3, section 3.4.3, series 1 & 2, 3 & 4, and 5 & 6 come in pairs. Each pair has the same formula, but has a different range of magnitudes as input. For this reason the pairs of series are expected to be highly correlated. Models 2 and 3 only use one series from each pair. This is in the hope that most of the information from a pair will be contained in just one of the series. Using only one series should reduce the multicollinearity , and simplify the model. Again, 2a is not intended as a sensible working model as it has all 2nd and 3rd order interactions. Its purpose is for comparison against model 1, to see if dropping some series reduces the predictive power.

In the principal component analysis in Chapter 6 (where the series still had a running maximum applied) series 1 & 2, and 3 & 4 showed a surprising lack of correlation relative to series from outside the pair. This suggests that series 5 and 6 should be singled out from series 1-4. Model 4 is an attempt to do this. It uses the linear combinations of series suggested by PCA in Chapter 6: the sum of the first 6 series, series 7 on its own, and the sum of series 5 and 6 since the PCA suggested they were more correlated than any other subgroup of series. The sum of series 5 and 6 is entered after the sum of series 1 to 6, to see if the sequential sum of squares for the sum of series 5 and 6 is significant. All 2nd and 3rd order interactions are included, so that it can be compared with model 1 in the same way as models 2 and 3.

Model 5 only includes significant effects, and effects that may not be significant but are involved in higher order interactions. Because there is so much correlation between series 1-6, this is by no means a unique model. There is a high degree of multicollinearity making individual terms difficult to interpret. Instead, models are assessed through groups of factors, for example, all 3rd level interactions.

Table 7.5 and 7.6 below show the effect on the residual sums of squares when different terms are included. All models are subsets of model 1c, but models 4a – 4c have constraints imposed, as explained below. In table 7.6 below, it can be seen that there are some marginally significant 3rd level interactions in model 1. However, they do not show up as significant in model 2 or 3. Combined, models 2a and 3a contain all the terms in model 1a. This hints that it is correlation with other predictor variables that is making the 3rd level interactions significant. For models 2-4, the optimum model has only 2nd level interactions included. Since the model with just main effects is not significantly different from the model with just the mean this also raises the question of whether it is correlation between the predictors that is causing some variables to be significant.

Table 7.5 Residual Sums of Squares and Degrees of Freedom

Base Model (containing only the mean): df RSS RMS	Main Effects			Second Level Interactions (and Main Effects)			Second and Third Level Interactions and Main Effects		
32.21 188 0.171	RSS	df	RMS	RSS	df	RMS	RSS	df	RMS
Model 1	30.65	181	0.169	23.33	160	0.146	16.77	125	0.134
Model 2	32.16	184	0.174	28.2	175	0.161	28.55	174	0.164
Model 3	31.44	184	0.171	28.5	178	0.16	27.98	174	0.161
Model 4	32.2	185	0.174	30.7	182	0.169	30.66	181	0.169
Model 5	RSS = 26.17, df=175, RMS = 0.150								
	Includes all series, interactions S1*S5 S2*S3 S6*S7 S1*S2 S1*S3 & S1*S2*S3								

Table 7.6 P-Values For Tests For Significance of Interactions

	Main Effects vas Mean	2nd Level Interactions vas Main Effects	3rd Level Interactions vas 2nd Level Interactions and Main Effects
Model 1	0.243	0.001	0.093
Model 2	0.990	0.010	0.629
Model 3	0.340	0.007	0.516
Model 4	0.992	0.034	0.649

Because of the correlations between series 1-6, it is not possible to find unique 3rd order interactions. There are two 3rd order interactions that are significant in model 5.

The presence of 2nd and 3rd order interactions suggests that using maxima of the series (as the M8 algorithm does) is the wisest thing to do. The interactions are the products of series, so if two series are high together, the interaction will magnify this effect. This has a parallel with the way the M8 uses the series – its when the series simultaneously cross their thresholds that TIPs are declared, what they do while below the threshold is irrelevant.

The only way that the unsmoothed series tend to change is to change in variance – their mean level does not change much at all. When the variance goes up the maximum over 5 years goes up, even though the mean level does not change. Therefore, when a running maximum over 5 years is taken the result is series that rise and fall as the variance increases and falls. As seen throughout Chapter 5, the smoothed series do have trends in them. However, if time is fitted as a third degree polynomial series 3 –7 show absolutely no time trend, and series 1 and 2 have a negligible one, as the scatter overwhelms the trend. For series 1 the quadratic term was significant, but R^2 was only 3%. Similarly, for series 2 the linear and quadratic terms were significant, but R^2 was only 4%.

In the linear models with the unsmoothed series, it is a change in variance of the series that must take the place of a rise in the series in the original M8 algorithm. This explains the significance of the interactions, and relative unimportance of the main effects in these linear models. Interactions give lots of weight to extreme values and therefore signal when there is an increase in variance.

In table 7.6, above, the tests for significant differences in residual sums of squares between pairs of models can be seen. The test is an F test, with the increase in residual sums of squares on dropping the terms from one model in the numerator, and the residual sums of squares of the full model in the denominator. The degrees of freedom for the numerator are the degrees of freedom for the residual sums of squares in the numerator. Similarly, the degrees of freedom for the denominator are from the degrees of freedom for the residual sums of squares in the denominator.

In table 7.7, below, there are comparisons of models 1-5, all with 3rd level interactions included. All models shown here are nested in model 1 with 3rd level interactions, since there is a linear combination of the terms in model 1 (with 3rd level interactions) that is equivalent to the other models. For example, model 1 with 3rd level interactions is equivalent to model 4 with just main effects if constraints are imposed. Namely, if the coefficients of all 2nd and 3rd level interactions are constrained to be zero, series 1-4 constrained to have the same coefficients, and series 5 and 6 are constrained to have the same coefficients. It can be seen in table 7.7 that only model 5 is as predictive as model 1 – the residual sums of squares are significantly higher for the other models.

Table 7.7 Comparing models with model 1:
F-tests for significant differences in residual sums of squares
All 3rd and 2nd level interactions are included

model	**p-value**
model 2	0.003
model 3	0.009
model 4	0.002
model 5	0.148

Models 2 and 3 only contain one of the series in each pair. Since the series in a pair are similar, it seems likely that series from the same pair should be related to the response in the same way. However, it was found that if one series of a pair is related to the response variable, it does not follow that the other will be. This indicates that the series are not related in the same way to the response variable, and are not particularly correlated.

Table 7.8 shows tests of significance using partial sums of squares and sequential sums of squares for models 2 and 3. For model 3, it can be seen that the series from pair 1 is not significant in the sequential sums of squares, although it is entered first. However, it is significant in the test using partial sums of squares (*i.e.*, when entered last). The series from the third pair also shows the same pattern of significance when it is entered first into the model. Together, the series from the 1st and 2nd pairs span a plane that the response vector lies within, without either of the pair actually being strongly related to the response variable.

Table 7.8 Tests Using Partial and Sequential Sums of Squares

	P-values (Model 2, Seq SS)	P-values (Model 3, Seq SS)	P-values (Model 2, Partial SS)	P-values (Model 3, Partial SS)
1st Pair (P1)	0.73	0.62	0.69	0.01
2nd Pair (P2)	0.69	0.03	0.51	0.01
3rd Pair (P3)	0.86	0.98	0.20	0.10
Series 7	0.95	0.92	0.91	0.65
P1*P2	0.29	0.09	0.08	0.06
P1*P3	0.04	0.13	0.01	0.39
P1*Series 7	0.93	0.27	0.71	0.42
P2*P3	0.00	0.45	0.03	0.46
P2*Series 7	0.25	0.61	0.23	0.70
P3*Series 7	0.14	0.00	0.18	0.00

To demonstrate that the significance of the interactions is due to the presence of other variables in the model, the two main effects and the interaction for each significant interaction in model 1 were included in a model on their own. The p-values for the significance of interactions when included with their main effects only are shown in table 7.9.

Table 7.9 Significance of Interactions From Model 5 With Only Their Own Main Effects In the Model

Interaction	p-value for interaction	Residual SS for model including interaction
S1*S5	0.216	31.92
S2*S3	0.070	31.39
S3*S4	0.389	32.04
S5*S7	0.026	31.35
S6*S7	0.010	31.06

The interaction between series 1 & 5, and the interaction between series 3 & 4 are clearly not significant without the presence of the other main effects. The change in their significance cannot be attributed to the fact that the effect of the interaction averages out over the levels of another factor (i.e., that there is a significant 3 factor interaction). There are no significant 3 factor interactions in model 1 involving both series 1 & 5 or both series 3 & 4. These 2 factor interactions are significant because together with another explanatory vector they span a plane that contains the response vector. The interaction between series 2 & 3 is marginally significant, whereas its p-value for partial sums of squares was 0.002 in model 5. When the interaction between series 2 & 3 is in a model with only series 2 and 3, its sums of squares more than halves, from 1.41 to 0.56. The residual mean square does not increase that much either, when all series except series 2, 3, and their interaction are dropped from the model. The residual mean square goes from 0.143 to 0.170. Therefore, it appears that part of the predictive power of the series2 & 3 interaction comes from its correlation with the other series.

The interactions involving series 7 cannot be expected to be due to correlation, since series 7 is not highly correlated with the other series. The possibly real interaction between series 2 & 3 is interesting, as they use different magnitude ranges, and it is the magnitude ranges that were used to split off models 2 and 3 from model 1. This argues against using models 2 & 3.

In model 4, none of the linear combinations suggested by the principal component analysis in chapter 6 were significant, except for series 7 on its own. This again suggests that it is the maxima of the series that are related to the largest earthquake coming in the next 5 years, as the average value of the series is not related to it.

A possible model is model 5, as its residual sums of squares are not significantly different from model 1, with 2nd and 3rd level interactions. Model 5 is very unlikely to be a unique model, because the series are so correlated.

The same types of linear models were tried on the series targeted at predicting magnitude 6.5. The results were very similar, i.e., the degree of correlation between the series and higher order interactions meant that the types of linear model above were not suitable.

7.5.2 Logistic Regression Using the 7 M8 Series

A logistic regression model for the probability of an earthquake was fitted (Nelder and McCullagh, 1983):

$$\log\left(\frac{p}{1-p}\right) = \underset{\sim}{\beta}\,\underset{\sim}{x}^T$$

where

p = the probability of an earthquake

β = matrix of estimated coefficients

x = matrix of the 7 M8 series

A summary of the significance of interactions is shown in table 7.10. As in the linear model using the 7 M8 series, correlation between the series makes it difficult to isolate individual interactions as having a coefficient significantly different from zero. For this reason, the interactions are considered in groups (2nd order, 3rd order, etc.) as in table 7.5 with the linear model for the M8 series. The deviance of the denominator in the F-tests in table 7.11 is the deviance of the model with more terms.

Table 7.10 Residual Deviances and Degrees of Freedom

Base (containing the mean):	Mode Main Effects onl			Second Leve Interactions (and Mai Effects)			Second and Thir Level Interactions an Main Effects		
RD = 258.7, df = 188 RMD= 1.38	RD	df	RMD	RD	df	RMD	RD	df	RMD
Model 1	252.22	181	1.39	205.94	160	1.29	131.37	125	1.05
Model 2	257.49	184	1.4	241.54	178	1.36	235.81	174	1.36
Model 3	254.63	184	1.38	237.47	178	1.33	232.45	174	1.34
Model 4	258.01	185	1.39	247.98	182	1.36	247.89	181	1.37
Model 5	Series 2, 5 and their interaction. RD = 250.18, df=185, RMD=1.35								

Table 7.11 P-Values For Tests For Significance of Interactions

	Main Effects vas Mean only	2nd Level Interactions vas Main Effects only	3rd Level Interactions vas 2nd Level Interactions and Main Effects only
Model 1	0.52	0.01	0.33
Model 2	0.88	0.02	0.27
Model 3	0.41	0.02	0.32
Model 4	0.22	0.02	0.77

In contrast with the ordinary least squares models above, no 3rd level interactions are significant in any of the models. Neither are any main effects, even with the interactions in the model, as can be seen in tables 7.10 and 7.11 above. This gives reason to suspect whether the 2nd level interactions actually are related to the response .

variable, for the same reasons indicated in the section 7.5.1, above, on ordinary least squares with the 7 M8 series.

The intercorrelation between the series caused the same difficulties in interpretation here as in section 7.5.1, that is, some coefficients of interactions are significantly different from zero, but only because of their correlation with other predictor variables. Model 5 is an attempt to create a parsimonious model, which did not work out. It was not possible to drop any non-significant variables without affecting the significance of other variables. Dropping series that were not significant left series 5, 7 and their interaction, which is not as predictive as the full model with 2^{nd} level interactions (as seen in table 7.10).

It may be correlation with series 5 and 7 that is causing the interaction between them to be significant. Neither series 5 or 7 are significant themselves, which does make this explanations seem more likely. The interaction would not make much sense in the model without its main effects. However, if it is seen not as an interaction but a variable that accentuates the instances where series 5 and 7 are high together, then it makes sense to put the product of series 5 and 7 in the model on its own. When this is done, coefficient of the product of series 5 and 7 is significantly different from zero, with a p-value 0.015. However, the residual mean deviance with series 5, 7 and their interaction is 1.36, so not a lot of deviance is explained. With a probability of 0.007 – 0.015 (from the p-value of the interaction with and without series 5 and 7 in the model) of wrongly declaring a variable significant, the expected number of variables that are wrongly declared significant when 63 are included in the model is 0.44 – 0.95. Therefore, there is not really strong evidence for the interaction between series 5 and 7 in light of the overall amount of significance tests done. Also, since series 5 is correlated with the other series, a product of series 7 and one of the other series could be just as relevant to the probability of an earthquake.

Table 7.10 Comparison Between Models 2-5 and Model 1 Without 3rd Level Interactions.

model	p-value
model 2	0.04
model 3	0.09
model 4	0.04
model 5	0.06

7.6 Models with the TIP Level Series

The TIP level series uses the M8 algorithm's rules to summarise the 7 series, as described in Chapter 5, section 5.3.1. In this section linear models and logistic regression models use the M8 series in almost the same way as the M8 algorithm, but make more ambitious predictions. The linear model predicts the maximum magnitude expected in the next year. This goes further than M8, which just indicates whether an earthquake over the target magnitude is expected. The logistic regression model gives the probability of an earthquake, which is also contains more information than just a warning.

Models with the TIP level series are made much simpler by only having one series and 'circle' as predictors. The absence of many higher order interactions make it practical to explore the models graphically, leading to suggestions for better models.

7.6.1 A Regression Model with the TIP Level Series

A regression model with the TIP level series is shown below. As mentioned in table 7.1, the response variable is the magnitude of the largest earthquake in that year. TIPs in the M8 algorithm extend for 5 years after the TIP level series exceeds zero twice in a row, regardless of subsequent values of TIP level. To transfer this rule to a linear model, a running maximum of the TIP level is taken: this is the predictor variable. 'Circle' is included as a categorical factor for several reasons:

- There is no problem with degrees of freedom. In the previous models with the 7 M8 series adding 'circle' with interactions would have reduced the residual degrees of freedom too much, but with only one other predictor variable there are no such concerns.

- In models with the TIP level series 'circle' is highly significant (p-value for significance is < 0.0001) and will reduce the residual variation. This would pose a problem if the model were to be used for prediction in another circle, but it will be useful to gain an idea of how big the differences the constant are for each circle.

The circle by TIP level interaction is also included, just to see if it is significant. If it is, a different parameter is needed in each circle. The model would be of little value as a prediction tool, as it could not be applied generally to any circle. Unfortunately the interaction is significant, with a p-value of 0.011. However, the TIP level series is not significant, so as a whole the model does not work. All p-values for models with the TIP level series are shown in table 7.11, below. It can be seen in table 7.12 that the constant does not change too much from circle to circle, but the coefficient for the TIP level series does.

Table 7.11 Tests of Significance Based on Sequential Sums of Squares

Variable	P-value
Circle	< 0.001
TIP Level	0.815
TIP level * Circle	0.011

Table 7.12 Coefficients for Regression Model with TIP Level Series

Coefficient	Value	Standard Error	t-value	p-value
(Intercept)	5.71	0.13	44.04	0
Circle 2	0.13	0.16	0.8	0.43
Circle 3	0.06	0.17	0.37	0.71
Circle 4	0.04	0.17	0.2	0.84
Circle 5	-0.94	0.22	-4.21	0
Circle 6	-0.47	0.16	-2.84	0
Circle 7	-0.53	0.16	-3.37	0
Tip Level	-0.37	0.43	-0.86	0.39
Slope C 2	1	0.65	1.55	0.12
Slope C 3	-0.09	1.37	-0.07	0.95
Slope C 4	1.82	0.7	2.61	0.01
Slope C 5	-2.18	1.18	-1.85	0.07
Slope C 6	0.39	0.55	0.71	0.48
Slope C 7	-0.6	0.8	-0.75	0.45

In figure 7.1, below, it can be seen that there is a weak relationship between the TIP level series and the maximum magnitude in some circles. However, it is only in circle 4 that there is a change in the average level of 'TIP level' with maximum magnitude. The p-value for the significance of the TIP level series in circle 4 is 0.004. In figure 7.1 it can be seen that there is a general tendency for the variance of the TIP level series to increase with maximum magnitude. This is clearer in figure 7.2 below, where data from all circles is pooled. As a rough test for increasing variance, the TIP level variable was treated as a categorical variable and Levene's test for homogeneity of variance was carried out. The data was split into four groups using the upper quartile, median and lower quartile of the TIP level variable. The variance for each quartile is shown in table 7.13 below. As can be seen from the figures in table 7.13, the real increase in the variance of the maximum magnitude is in the top quartile of the TIP level series. The variance decreases slightly in the 3rd quartile, so the variance of the maximum magnitude does not increase consistently with the TIP level.

Table 7.13 Variance of the Maximum Magnitude for Quartiles of TIP level

Range of TIP Level for Quartile	Variance of Maximum Magnitude for
-0.5770 to –0.223	0.267
-0.224 to –0.094	0.326
-0.095 to –0.008	0.264
-0.008 to 0.1	0.470

The p-value for Levene's test was 0.025 indicating that there is moderate evidence that the variance differs for the different quartiles of the TIP level series. It is typical for the TIP level series to slowly climb up to zero, and then slowly fall again. Even if it does stay above zero for a while (in the 'danger' zone), the time where it is slowly falling is also a time of increased earthquake risk (from the definition of a TIP). This is why a running maximum of the TIP level series is fed into the models. When the TIP level series is at its peak, the 6 year smoothing will make the peak last 6 years. However, only one earthquake of the target magnitude has to occur during those 6 years to count the TIP a success. In this model, if one large earthquake does occur there are still many other little ones that are associated with a high value of TIP level. This is seen in the higher variance of the maximum magnitude for the top quartile of TIP level. This standard regression model compares the mean magnitude when TIP level is high with other times, and the magnitude of the one earthquake that makes a TIP a success is lost in the average.

Figure 7.1

Smoothed TIP Level Against Maximum Magnitude, By Circle

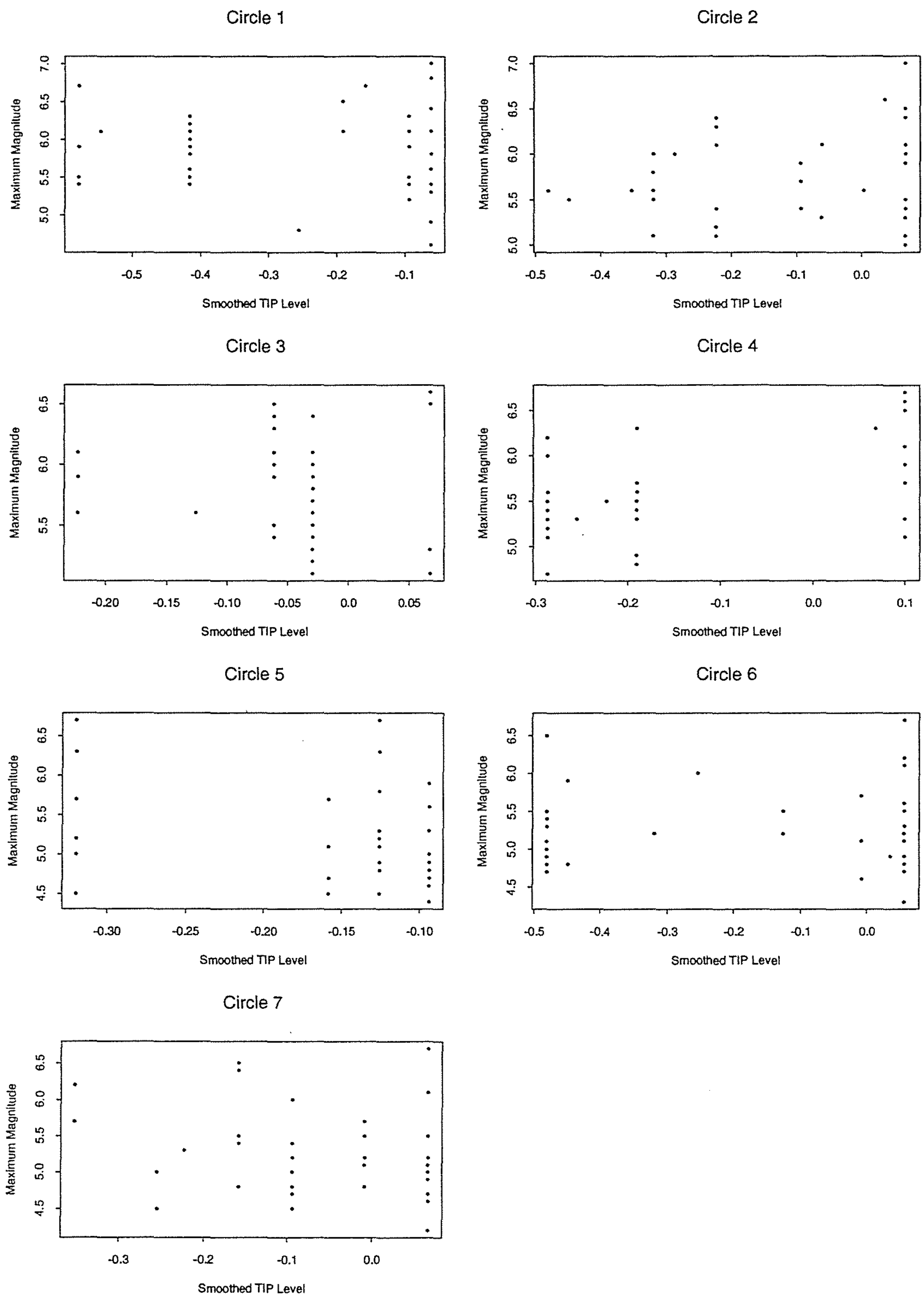
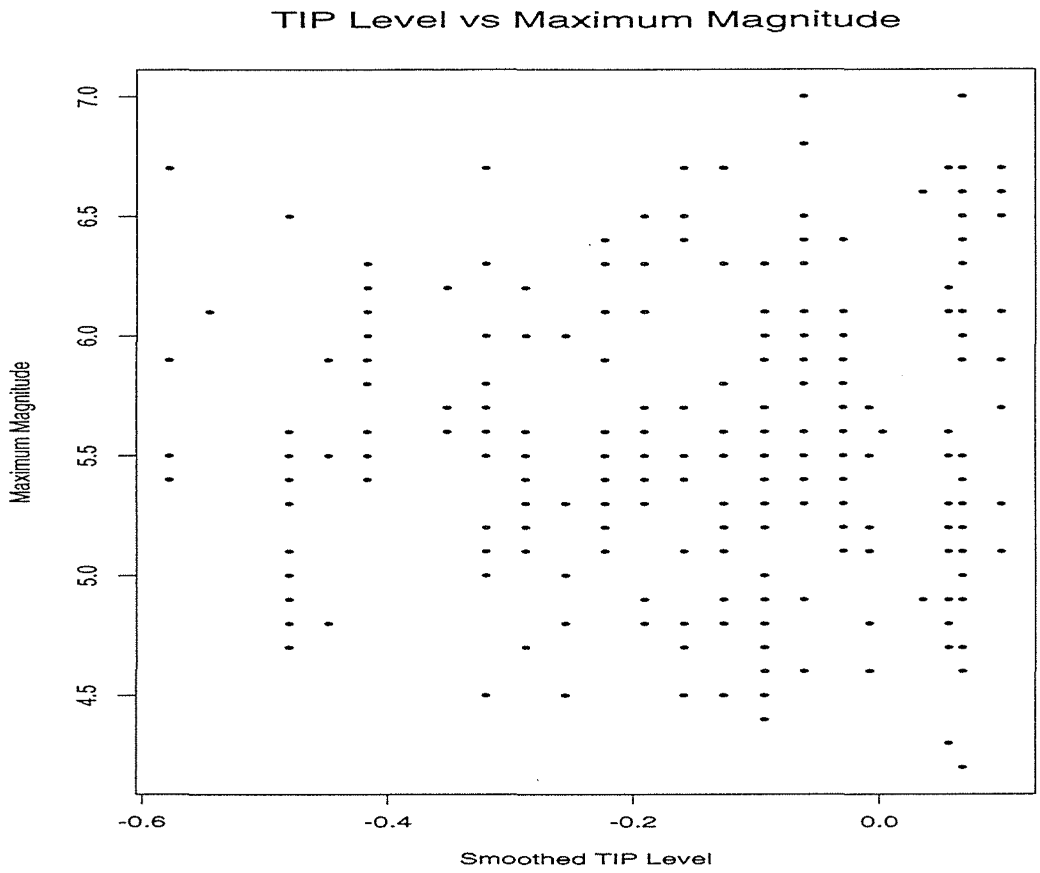


Figure 7.2



The M8 algorithm only considers the TIP level series to be indicating danger when it is above zero for 2 consecutive intervals. How far below zero the TIP level is not of concern. One of the reasons for trying the TIP level series was that lower values might also indicate the level of risk. However, it appears that the maximum magnitude over 6 months is not generally related to the TIP level. It is only the top quartile of TIP level where the variance of the maximum magnitude is different from the variance of maximum magnitude over the other 3 quartiles of TIP level. If Levene's test is done excluding the top quartile of TIP level then there is no significant difference in variance between levels. The upper quartile is –0.008, which is very close to zero (the threshold for the algorithm to declare a TIP). This suggests that values of the TIP level series below zero are not related to the expected magnitude of the largest earthquake in the coming six months. Therefore, the M8 algorithm is not loosing information by using the TIP level series to give a warning only when the TIP level exceeds zero.

7.6.2 A Logistic Model for the Probability of an Earthquake with the TIP Level Series

This logistic model gives the probability of an earthquake using the TIP level series:

$$\log_e\left(\frac{p}{1-p}\right) = \alpha_i + (\beta + \gamma_i) \times TIP.Level$$

where

p = probability of earthquake

α_i = effect of ith circle

β = coefficient of TIP level series

γ_i = difference between coefficient of ith circle and 7th circle

The variable 'circle' is included here in the same way as for the ordinary least squares model, for the same reasons. The TIP level series is significant, but unfortunately the interaction term is also significant, as it is with the regression model using the TIP level series. The interaction term can be seen as an estimate of the variability in the model's slope of the TIP level line when the model is applied to a new circle of investigation. If the interaction effect were not very large, the variability in the slope of the line in different locations would not be very large, either. Then it might be possible to apply the model to different circles, knowing that there would be a small error in the coefficient of the TIP level series due to the specific location of the circle of investigation. However, the coefficient for the TIP level series changes in sign from circle to circle indicating radical differences in the TIP level series relationship with danger times in each circle. All coefficients in the logistic regression are shown in table 7.14 , below. The fitting algorithm would not produce an estimate for circle 1 because of singularities.

Table 7.14 Coefficients in the Logistic Regression with the TIP level series

Coefficient	Value	Standard Error	t-value
circle 1	-1.78	0.70	-2.55
circle 2	-0.83	1.26	-0.66
circle 3	-0.32	0.98	-0.33
circle 4	-10.33	18.01	-0.57
circle 5	-2.85	1.90	-1.50
circle 6	-1.28	1.18	-1.08
circle 7	-1.11	0.96	-1.16
TIP level	0.63	2.39	0.26
difference, circle 1	-	-	-
difference, circle 2	15.29	15.41	0.99
difference, circle 3	26.92	12.93	2.08
difference, circle 4	111.98	180.65	0.62
difference, circle 5	-10.15	7.82	-1.30
difference, circle 6	-1.34	3.80	-0.35

The analysis of deviance table for the model is shown below (table 7.14). The interaction term in this table can be used as an estimate of the variability in the coefficient of the TIP level series when changing location. The F-statistic for variation due to the TIP level series compared with variation in the TIP level coefficient from circle to circle is 1.444. This has a p-value of 0.275, confirming what is observed in table 7.13 — that the coefficient of the TIP level series differs too widely from circle to circle to apply this model to just any circle of investigation. Therefore, it would not be practical to use this model for prediction.

Table 7.14 Analysis of Deviance Table for Logistic Regression with TIP level

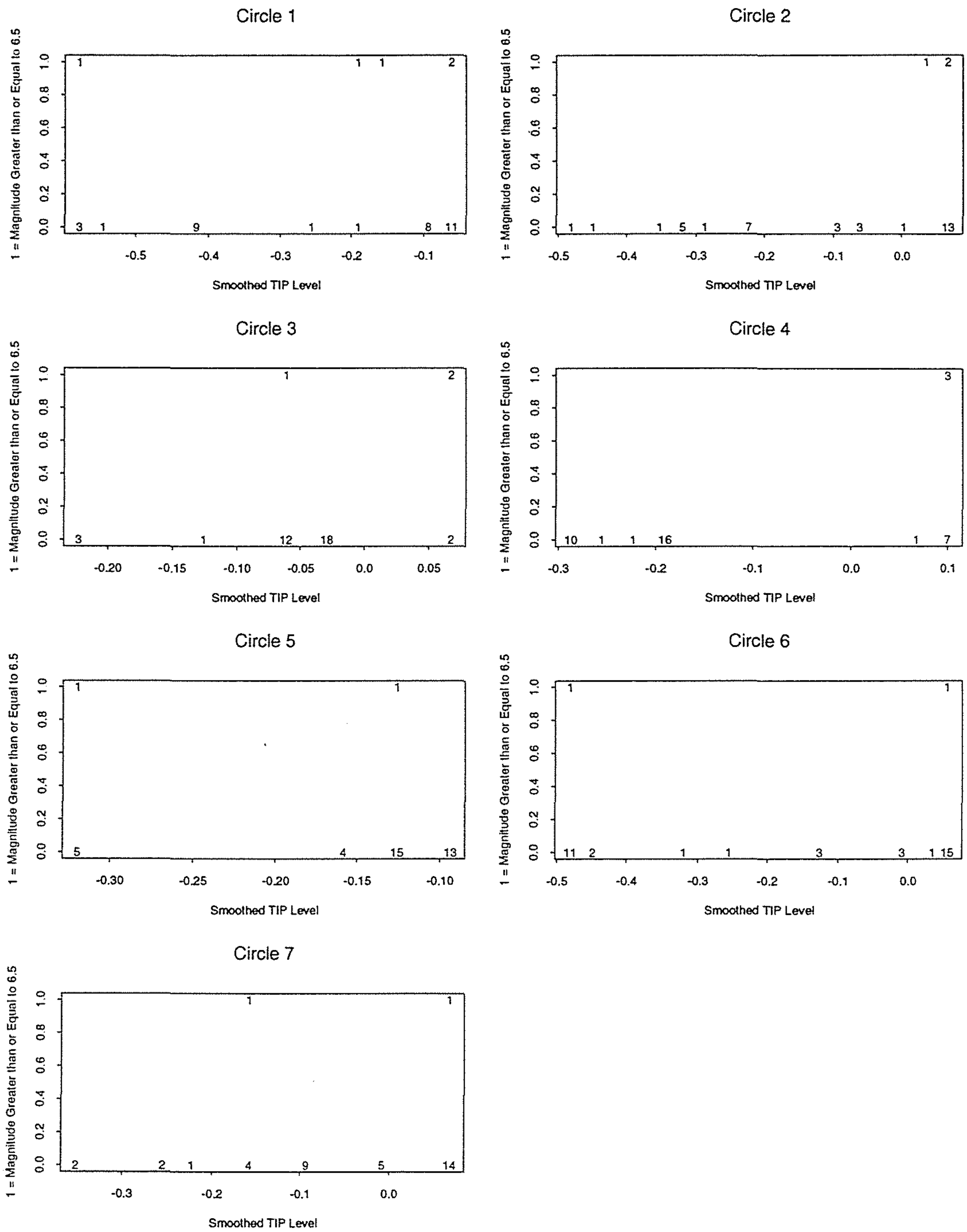
	df	Deviance	Residuals df	Residual Deviance	F statistic	p-value
NULL model	272	143.05				
circle	6	2.39	266	140.66	0.44	0.85
TIP level	1	3.96	265	136.70	4.38	0.04
circle * TIP level	6	16.45	259	120.26	3.03	0.01

In circle 4, the TIP level series is highly significant (with a p-value < 0.0001). This could be just a chance success, especially since the significance is based on 3 points, where an earthquake with a magnitude over 6.5 has occurred. Alternatively, M8 may have statistically significant success in *some* circles (but more data would be needed to test this hypothesis).

A plot of the response versus the TIP level series by circle is shown below, the response being a '1' for where an earthquake with magnitude ≥ 6.5 has occurred, and '0's elsewhere.

Figure 7.3

Earthquakes > M6.4 Against Smoothed TIP Level, By Circle



7.7 Conclusion

The M8 series are not ideally suited to linear models. The most obvious reason is the correlation between the first 6 series, which complicates the interpretation of significant effects. More importantly, it turns out that the M8 algorithm does use the series in the optimum way by considering the maxima of the series instead of concentrating on their average level, as the models fitted here do. In the models of this chapter, it is the products of the series (the interactions) that are significant, because these products accentuate the instances when the series in the product are high together. Also, all the series are needed, and most significant terms depend on other terms in the model for their significance. This parallels M8's rule of requiring 5 out of the first 6 series to cross their thresholds for a TIP to be declared. The TIP level series uses the maxima and minima of the M8 series the way that the M8 algorithm does and the logistic regression using the TIP level series had some success. However, the results could not be replicated in space. Therefore the models here can not be used for prediction. However, this does not prove that the M8 series can not be used for intermediate term prediction. Because of the scarcity of strong earthquakes, the target magnitudes have been less than those that the algorithm is designed for. When there is enough data to construct models for predicting magnitude 7 (and magnitude 8) earthquakes the results may be more encouraging. Fitting linear models to the M8 series has shown that overall, it is changes in variance that are important with the M8 series, and that a change in the variance (rather than the average) of magnitudes can be expected when the TIP level series crosses its threshold.

Chapter 8 ~ Conclusions

The aim of this thesis was to investigate how the M8 algorithm works, and explore alternative ways of using its 7 series. The algorithm is a combination of rules specifying when there is a danger of a strong earthquake. These rules are seldom explained or justified by the algorithm's authors, but almost every time any rule or procedure has been tested in this thesis, there has been a good reason for it. The result is that no significant improvements have been suggested, but a greater understanding of the algorithm has been gained.

The M8 algorithm is intended to predict shallow earthquakes, and is only meant to utilise shallow earthquakes in its series. Therefore, there is no need for the algorithm to predict the depth of an on-coming earthquake, as shallow earthquakes only occur in the crust and do not vary much in depth. However, in New Zealand the algorithm does not produce any successful predictions unless all depths are included. Since earthquakes can occur at depths of up to 700 km, this reduces the value of the predictions as the felt intensity will be greatly reduced if the earthquake has a deep focus.

At any given point in time only one series is used in the declaration of a TIP. It was hoped that some series would be redundant, in the sense that the decision to declare a TIP would never be based on them. This would have simplified the analysis, especially with linear models. However, with New Zealand data, the algorithm uses all series in its predictions.

According to the M8 algorithm an increase in activity above the normal level implies an increased likelihood of an earthquake above the specified magnitude, and this holds true whatever the normal level. The Gutenburg-Richter law provides a fixed relationship between the frequency of earthquakes of different magnitudes, so that in an area where there are normally more moderate earthquakes there will be correspondingly more large earthquakes. However, there is no evidence of a positive correlation between the mean values of the series and the total magnitude over 6.5 in a circle. Again, the evidence is in favour of M8's procedure.

In a second attempt to reduce the number of series to analyse, principal components analysis was applied to the M8 series. This resulted in a better understanding of the series. The first principal component generally consisted of a weighted average of the first 6 series, and for the second or third principal components only series 7 had a reasonable loading. This confirmed that series 1-6 are similar to each other, and different from series 7, which is implied by M8's decision rule for a TIP. Requiring that only 5 of the first 6 series cross their thresholds implies that they are similar, because they can be substituted for each other. In contrast, series 7 must cross its threshold for a TIP to be declared. Unfortunately, no changes in the principal components before an earthquake were found.

One aspect of the series found to be connected to coming earthquakes was their maxima. When the M8 series are not smoothed using running totals and running maxima they don't show any trend, or show an extremely negligible one. It is changes in the variance of the series (picked up by the running maximum) which give rise to the strong trend in the smoothed M8 series. It is the extreme points that cause the trends in the M8 series, indicating that they are the points that are related to earthquake risk. This is reflected in the linear models with the M8 series in the significance of many 2^{nd} and 3^{rd} order interactions. Being the product of two or three series, the interactions emphasise the points where those series are high together. This has a parallel with the M8 algorithm. Most of the series must cross their thresholds for a TIP to be declared. Also, most of the significant interactions are dependent upon other interactions or main effects for their significance. This shows, again, that all of the series are needed in order to make predictions. The principal components were also used as predictors in linear models. The principal components did not have any predictive value, also showing that the predictive power of the series does not lie in linear functions of them.

One of the most obvious improvements to the M8 algorithm is to change the predictions from a 'yes' or a 'no' to a continuous variable such as the probability of an earthquake or the maximum magnitude expected in the next year. The 'hard' boundary of the circle of investigation is also a kind of 'yes/no' variable: either an earthquake contributes to the series, or it doesn't, whereas in reality earthquakes just outside the circle of investigation are probably signs of stress inside the circle.

However, the use of these binary variables is surprisingly effective. When testing the stability of M8's predictions with regard to the position of the circle of investigation it was found that this did affect whether or not a TIP was declared. However, when the TIP level series was considered as a whole, rather than just whether it had reached its threshold, the algorithm was found to give a reasonably consistent signal.

It was expected that information from the whole range of values in the TIP level series would be related to impending earthquakes. The study of the stability of the predictions when the circle of investigation was shifted slightly seemed to suggest this: even if the TIP level series did not reach its threshold at the appropriate time, it was increasing, rather than decreasing. However, this relates more to the slope of the TIP level series, rather than its actual value. When the TIP level series was included in linear models it was found that the M8 algorithm extracts most (if not all) of the predictive information from the TIP level series by declaring a TIP when it crosses its threshold.

In conclusion, it is the maxima of the unsmoothed series of the M8 algorithm that are related to future earthquakes, rather than the average level of the series. Also, all series are needed to relate these maxima to coming earthquakes. The linear models using the TIP level series could not be generalised to circles other than the ones they were created with, which means that the models are not practical tools for making predictions. The fact that the patterns were different in every circle also casts doubt over whether these patterns will continue to exist in the future. However, this is not proof that M8 cannot be used for predicting earthquakes. Because of the lack of data, it has been necessary to work with earthquakes smaller than those the algorithm was originally designed for. Also, circles of investigation intended for predicting earthquakes with magnitudes greater than 7 have been used for predicting earthquakes greater than 6.5 in magnitude. This produced some successes rather than none at all, as was the case when magnitude 6.5 circles were used. However, it also produced some failures to predict which caused the lack of generality in the models with the TIP level series. The algorithm did manage to signal the approach of the two magnitude 7 earthquakes. When enough data is available to have models predicting only magnitude 7 (or magnitude 8) earthquakes the 7 M8 series may have more success with intermediate term earthquake prediction.

References

Bartlett, M. S. 1947: The general canonical correlation distribution. *Annals of Mathematical Statistics 18*: 1-17.

DataDesk 6 for Macintosh. 1997: DataDescription Inc. Ithaca, New York.

Eiby, G. A. 1989: Earthquakes. Auckland. Heinemann Reed.

Harte, D.; Vere-Jones, D. 1998: Differences in coverage between the PDE and New Zealand local earthquake catalogues. : Institute of Statistics and Operations Research, School of Mathematical and Computing Sciences, Victoria University of Wellington.

Keilis-Borok, V.; Knopoff, L.; Allen, C. R. 1980: Long-term premonitory Seismicity patterns in Tibet and the Himalayas. *Journal of Geophysical Research 85*: 813-820.

Keilis-Borok, V. I.; Knopoff, L.; Kossobokov V.; Rotwain, I. 1990: Intermediate-term prediction in advance of the Loma Prieta earthquake. *Geophysical Research Letters 17*: 1461-1464.

Keilis-Borok, V. I.; Knopoff, L; Rotwain, I. M. 1980: Bursts of aftershocks, long-term precursors of strong earthquakes. *Nature 283*:259-263.

Keilis-Borok, V. I.; Knopoff, L; Rotwain, I. M.; Allen, C. R. 1988: Intermediate-term prediction of occurrence times of strong earthquakes. *Nature 335*:690-694.

Keilis-Borok, V. I.; Kossobokov V. G. 1990: Times of increased probability of strong earthquakes ($M \geq 7.5$) diagnosed by algorithm M8 in Japan and adjacent territories. *Journal of Geophysical Research 95*: 12,413-12,422.

Keilis-Borok, V. I.; Kossobokov V. G. 1990: Premonitory activation of earthquake flow: algorithm M8. *Physics of the earth and planetary interiors 61*: 73-83.

Kossobokov, V. G.; Carlson, J. M. 1995: Active zone size versus activity: A study of different seismicity patterns in the context of the prediction algorithm M8. *Journal of Geophysical Research 100*: 6431-6441.

Kossobokov, V. G.; Keilis-Borok V. I.; Smith, S. W. 1990: Reduction of territorial uncertainty of earthquake forecasting. *Physics of the earth and planetary interiors 61*: R1-R4.

Kossobokov, V. G.; Keilis-Borok V. I.; Smith, S. W. 1990: Localisation of intermediate-term earthquake prediction. *Journal of Geophysical Research 95*: 19,763-19,772.

Kossobokov, V.G.; Romashkova, L.L.; Keilis-Borok, V.I.; Healy, J.H. 1999: Testing earthquake prediction algorithms: statistically significant advance prediction of the largest earthquakes in the Circum-Pacific, 1992-1997. *Physics of the Earth and Planetary Interiors 111*: 187-196.

Latoussakis, J.; Kossobokov, V.G. 1990: Intermediate term earthquake prediction in the area of Greece: Application of the algorithm M8. *Pageoph 134*: 261-282.

Li, M.; Vere-Jones, D. 1997: Application of M8 and Lin-Lin algorithms to New Zealand earthquake data. *New Zealand Journal of Geology and Geophysics 40*: 77-89.

Levene, H. 1960: Robust tests for equality of variance. *In*: Contributions to probability and statistics. Olkin, I; Ghanye, S. G.; Hoeffding, W.; Madow; W. G.; Mann, H. B. *eds*. Stanford University Press. California. Pp 278-292.

Lomnitz, C. 1994: Fundamentals of earthquake prediction. New York. John Wiley & Sons.

McCullagh, P; Nelder, J. A. 1983: Generalised linear models. London. Chapman Hall.

Rikitake, T., 1982: Earthquake forecasting and warning. Tokyo. Centre for Academic Publications Japan.

S-PLUS 4.3 for Unix, 1995: Math Soft Inc., Seattle.

Tributsch, H. 1982: When the snakes awake: animals and earthquake prediction (translated by Paul Langner). Cambridge, Mass. MIT Press.

Tsuneji, R. *ed.* 1981: Current research in earthquake prediction. Tokyo. Centre for Academic Publications, Japan.

Wyss, M; Dmowska, R. *eds*. 1997: Earthquake prediction — state of the art. Basel Boston. Birkhäuser Verlag.

Manly, B.F.J. 1995: Multivariate Statistical Methods. London. Chapman Hall.

Appendix ~ The M8 Algorithm

```
"subset.circle"<-
function(x, centrelat = 0, centrelong = 0, minradius = 0, maxradius = Inf,
        mindepth = 0, maxdepth = Inf, minmag = - Inf, maxmag = Inf, minday = -
        Inf, maxday = Inf, radius = 6371)
{
        attr(minday, "dp.second") <- attr(maxday, "dp.second") <- 2
        attr(minday, "origin") <- attr(maxday, "origin") <- c(month = 1, day =
          1, year = 1960)
        class(minday) <- class(maxday) <- "datetimes"
        restrict <- (x$depth >= mindepth & x$depth <= maxdepth & x$magnitude >=
          minmag & x$magnitude <= maxmag & as.numeric(x$time) >=
          as.numeric(minday) & as.numeric(x$time) <= as.numeric(maxday))
        restrict[is.na(restrict)] <- F
        indices <- seq(1, length(x$latitude))[restrict]
        #  Cartesian coordinates (unit sphere) of epicentre of reference point
        x1 <- cos((centrelat * pi)/180) * cos((centrelong * pi)/180)
        y1 <- cos((centrelat * pi)/180) * sin((centrelong * pi)/180)
        z1 <- sin((centrelat * pi)/180)
        #  Cartesian coordinates of epicentres of events
        x2 <- cos((x$latitude[indices] * pi)/180) * cos((x$longitude[indices] *
          pi)/180)
        y2 <- cos((x$latitude[indices] * pi)/180) * sin((x$longitude[indices] *
          pi)/180)
        z2 <- sin((x$latitude[indices] * pi)/180)
        #  Euclidean distance for each event from reference point
        ddist <- sqrt((x1 - x2)^2 + (y1 - y2)^2 + (z1 - z2)^2)
        #  Arc length from reference point epicentre to event epicentre
        arclength <- 2 * radius * asin(ddist/2)  #  Complete selection process
        restrict <- (arclength <= maxradius & arclength >= minradius)
        restrict[is.na(restrict)] <- F
        indices <- indices[restrict]
        n <- length(indices)
        cat("\nNumber of Selected Events =", n, "\n")
        x <- list(indices = indices, centrelong = centrelong, centrelat =
          centrelat, minradius = minradius, maxradius = maxradius,
          mindepth = mindepth, maxdepth = maxdepth, minmag = minmag,
          maxmag = maxmag, minday = minday, maxday = maxday, type =
          "Circular", catname = attr(x, "catname"))
        class(x) <- "subset"
        return(x)
}
```

```
"subcatalogue"<-
function(events, subcat.name)
{
        x <- eval(parse(text = events$catname))
        y <- list(longitude = x$longitude[events$indices], latitude = x$
          latitude[events$indices], depth = x$depth[events$indices], time
           = x$time[events$indices], missing.time = x$missing.time[events$
          indices], magnitude = x$magnitude[events$indices], catname =
          subcat.name)
        class(y) <- "data.frame"
        attr(y, "subset") <- summary(events)
        attr(y, "minday") <- events$minday
      attr(y, "mindepth")<- events$mindepth
      attr(y, "maxdepth")<- events$maxdepth
        assign(subcat.name, y, where = 1)
}

"decluster.M8"<-
function(catalogue, cutoff.mag = 4, minday = min(catalogue$time), mindepth = -
        Inf, maxdepth = Inf, debug = F)
{
        start.compute <- proc.time()[3]
        cat("--------------- decluster.M8 ---------------\n")
        del <- (is.na(catalogue[, "time"]) | is.na(catalogue[, "magnitude"]) |
          is.na(catalogue[, "longitude"]) | is.na(catalogue[, "latitude"]
          ) | is.na(catalogue[, "depth"]))
        new.catalogue <- catalogue[!del, c("time", "magnitude", "longitude",
          "latitude")]
        print(dim(new.catalogue))
        del2 <- (new.catalogue[, "magnitude"] < cutoff.mag | catalogue[!del,
          "depth"] < mindepth | catalogue[!del, "depth"] >= maxdepth)
        new.catalogue <- new.catalogue[!del2, ]
        print(dim(new.catalogue))
        del[!del] <- del2
        # removed all missing values that could not be treated
# and magnitudes below cutoff
        n0 <- nrow(new.catalogue)
        n.aftershocks <- rep(0, n0)
        order.magnitude <- order( - new.catalogue[, "magnitude"], new.catalogue[
          , "time"])
        sel <- as.integer(row.names(new.catalogue))
        print("recatalogue(): Before calling .Fortran(\"recat\", --------")
        cat("Memory used: ", sum(storage()$used)/1024^2, "MB\n")
        res <- .Fortran("recat",
          as.integer(n0),
```

```
        as.double(as.matrix(new.catalogue)),
        as.integer(order.magnitude),
        sel = as.integer(sel),
        n.aftershocks = as.integer(n.aftershocks),
        NAOK = 1)
      print("------- After calling .Fortran(\"recat\", -------")
      cat("Memory used: ", sum(storage()$used)/1024^2, "MB\n")
      if(debug) {
        print("recatalogue(): Fortran completed")
        browser()
      }
      is.main <- (res$sel == 0)
      catalogue.main <- catalogue
      catalogue.main[, "n.aftershocks"] <- 0
      if(debug)
        browser()
      catalogue.main[!del, "n.aftershocks"] <- res$n.aftershocks
      # for main shocks, n.aftershocks is number of aftershocks
# for after shocks, n.aftershocks is the
#     negative row.name of its master main shock
      catalogue.main[del, "n.aftershocks"] <- NA
      # for records with missing values, n.aftershocks is also missing
# Now keep only main shocks
      catalogue.main <- catalogue.main[!del, ][is.main, ]
      catalogue.main <- catalogue.main[catalogue.main[, "time"] >= minday, ]
      if(debug)
        browser()
      cat(paste("Magnitude selection range: [", cutoff.mag, ", Inf)\n"))
      if(minday != min(catalogue$time)) {
        class(minday) <- "datetimes"
        cat("Starting date of the mainshocks as required:\n")
        cat(format(minday), "\n")
      }
      else {
        cat("Starting date of the mainshocks as first event available:\n"
                )
        cat(format(minday), "\n")
      }
      cat(paste("Depth selection range: [", mindepth, ", ", maxdepth, ")\n"))
      cat("Names of generated data frame: \n")
      print(names(catalogue.main))
      cat(paste("\nFinished recatalogue() in ", (proc.time()[3] -
        start.compute), "seconds.\n"))
      return(catalogue.main)
}
```

```
M8<-
function(catalogue, M0, centrelong = mean(catalogue[, "longitude"]), centrelat
        = mean(catalogue[, "latitude"]), radius = radius.M8default(M0), minday
        = catalogue[1, "time"], start.series = datetimes(year = (years1(minday
        ) + 12), month = 1, day = 1, hour = 0, minute = 0), training = "user",
        end.training = NA, time.breaks = Inf, running.total = 12, smoother = 6,
        TIP.length = 10, plotit = T, title = "M8 Series and TIPs", debug = F)
{
        cat("\n")
        cat("############################################################\n")
        cat("#                M8()                       #\n")
        cat("############################################################\n")
        attributes(minday) <- attributes(catalogue[, "time"])
        attributes(start.series) <- attributes(catalogue[, "time"])
        attributes(end.training) <- attributes(catalogue[, "time"])
        cat("M0 =", M0, "\n")
        cat("(longitude, latitude) = (", centrelong, ",", centrelat, ")\n")
        cat("radius =", radius, "\n")
        cat("Reliable data start after : ", format(minday), "\n")
        cat("Series start at           : ", format(start.series), "\n")
        cat("Training method\t         : ", training, "\n")
        cat("Training end at           : ", format(end.training), "\n")
        if(!is.data.frame(catalogue) || !any(names(catalogue) ==
                "n.aftershocks")) {
                stop("Please use decluster.M8() function to get mainshocks\n")
        }
        if(time.breaks == Inf & training == "user") {
                if(!(months1(start.series) == 1 | months1(start.series) == 7) |
                        !
                        days1(start.series == 1))
                        stop("'start.series' does not correspond to the first day of Jan or July.  This is a
requirement \nfor 'start.series' AND end.training'"
                                )
                if(!(months1(end.training) == 1 | months1(end.training) == 7) |
                        !
                        days1(end.training == 1))
                        stop("'end.training' does not correspond to the first day of Jan or July"
                                )
        }
        if(time.breaks != Inf | running.total != 12 | smoother != 6)
                cat("The default parameters of M8 have been altered - please make sure the parameters are
\ncompatible\n"
```

```
            )
    if(minday != min(catalogue[, "time"])) catalogue <- catalogue[catalogue[
            , "time"] >= minday, ]
    # Get the 7 M8 series and max magnitude in every half year
    series <- M8.series(catalogue = catalogue, M0 = M0, centrelong =
        centrelong, centrelat = centrelat, radius = radius, minday =
        minday, start.series = start.series, training = training,
        end.training = end.training, running.total=running.total,
       time.breaks=time.breaks,debug = debug)    # Generate tips
    if(!is.character(series)) {
        res <- M8.TIP(series, M0, training = training, end.training =
            end.training, smoother = smoother, TIP.length =
            TIP.length, title = title, debug = debug)
        if(plotit) {
            plot.M8(res, title = title)
        }
        return(res)
    }
    else {
        cat("\n!!!Result not available.\n")
        return("Result not available.")
    }
}

"M8.series"<-
function(catalogue, M0, centrelong = mean(catalogue[, "longitude"]), centrelat
      = mean(catalogue[, "latitude"]), radius = radius.M8default(M0), minday
      = catalogue[1, "time"], start.series = datetimes(year = (years1(minday
     ) + 12), month = 1, day = 1, hour = 0, minute = 0), training = "user",
     end.training = NA, time.breaks = Inf, running.total = 12, debug = F)
{
# base level
     cat("##########################################################\n")
     cat("#              M8.series()                  #\n")
     cat("##########################################################\n")
     if(!is.data.frame(catalogue) || !any(names(catalogue) ==
       "n.aftershocks"))
       stop("Please use decluster.M8() function to get mainshocks\n")
     if(years1(max(catalogue[, "time"])) - years1(min(catalogue[, "time"])) <
       12)
       stop("Not enough years of data avalable in catalogue.\n")
     #if(years1(start.series) - years1(min(catalogue[, "time"])) < 10)
       #stop("Series generation start too early.\n")
     interval.start <- start.series
     # this is the end date of the first 6 month interval in which
```

```
#   the 7 series are calculated.
        sel <- (in.circle(catalogue, centrelong = centrelong, centrelat =
          centrelat, radius = radius))
        # Use all events in the CI(all date)
        catalogue <- catalogue[sel, ]# Only keep events in CI
        n <- nrow(catalogue)
        cat("Number of mainshocks = ", nrow(catalogue), "\n")
        cat("----- First, last event in CI -----\n")
        print.mainshocks(catalogue[c(1, n), ])
        cat("\n")
        data.start <- catalogue[1, "time"]
        # this is the start of the events in this CI
        data.end <- catalogue[n, "time"]
        # this is the end of the events in this CI
        if(training == "moving") {
#level 1
          data.length <- (data.end - data.start)/365.24220000000003
          cat("Number of years in CI: ", data.length, "\n")
          cat("Events per years in CI: ", n/data.length, "\n")
          n.CAT20 <- data.length * 20
          n.CAT10 <- data.length * 10
          if(n.CAT20 > n)
                stop(paste("M8.series(): CAT20 need ", n.CAT20,
                        "events while we have only ", n, "!!\n"))
          ordered.mags <- rev(sort(catalogue[, "magnitude"]))
          cut.CAT20 <- ordered.mags[n.CAT20]
          cut.CAT10 <- ordered.mags[n.CAT10]
        }
        else if(training == "user") {
#level 1
          in.training <- (catalogue[, "time"] >= start.series & catalogue[
                , "time"] < end.training)
          n.training <- sum(in.training)
          data.length <- (end.training - start.series)/365.24220000000003
          cat("Number of years in training CI: ", data.length, "\n")
          cat("Events per years in training CI: ", n.training/data.length,
                "\n")
          cat("----- First, last event in training CI -----\n")
          print.mainshocks(catalogue[in.training, ][c(1, n.training), ]
                )
          n.CAT20 <- data.length * 20
          n.CAT10 <- data.length * 10
          if(n.CAT20 > n.training) {
                cat(paste("M8.series(): CAT20 needs ", n.CAT20,
                        "events\n    while we have only ", n.training,
```

```
                    "in training CI!!\n"))
            return("Result not available")
        }
        ordered.mags <- rev(sort(catalogue[in.training, "magnitude"]))
        cut.CAT20 <- ordered.mags[n.CAT20]
        cut.CAT10 <- ordered.mags[n.CAT10]
    }
    else if(training == "all") {
#level 1
        in.training <- (catalogue[, "time"] >= start.series & catalogue[
            , "time"] < data.end)
        n.training <- sum(in.training)
        data.length <- (data.end - start.series)/365.24220000000003
        cat("Number of years in training CI: ", data.length, "\n")
        cat("Events per years in training CI: ", n.training/data.length,
            "\n")
        n.CAT20 <- data.length * 20
        n.CAT10 <- data.length * 10
        if(n.CAT20 > n.training) {
            stop(paste("M8.series(): CAT20 needs ", n.CAT20,
                "events while we have only ", n.training,
                "in training CI!!\n"))
        }
        ordered.mags <- rev(sort(catalogue[in.training, "magnitude"]))
        cut.CAT20 <- ordered.mags[n.CAT20]
        cut.CAT10 <- ordered.mags[n.CAT10]
    }
    cat("Cutoff magnitude of CAT20 =", cut.CAT20, "\n")
    cat("Cutoff magnitude of CAT10 =", cut.CAT10, "\n")
    in.CAT20 <- (catalogue[, "magnitude"] >= cut.CAT20)
    in.CAT10 <- (catalogue[, "magnitude"] >= cut.CAT10)
    in.CAT20a <- in.CAT20 & (catalogue[, "magnitude"] < M0 - 0.5)
    in.CAT10a <- in.CAT10 & (catalogue[, "magnitude"] < M0 - 0.5)
    in.CATMS <- (catalogue[, "magnitude"] >= M0 - 2 & catalogue[,
      "magnitude"] < M0 - 0.20000000000000001)
    # Note: we are still not quite clear if these subsets extends
#   to the whole data length or restricted to series evaluation
#   period. If training is "user" or "all", then we only
#   use training period to compute magnitude cutoff,
#   but the subset runs through all reliable data.
#   Now we've got lower cutoff magnitudes for CAT20 and CAT10
#   and boolean for CAT20 group and CAT10 group and CATMS group
#   Next we should generate a time grid,
#   using series starting time('start.series')
#   as a grid(in the middle), starting from just <= data
```

```
#   starting time, ending at just >= data ending time,
#   spaced at 6 months.
        if(time.breaks == Inf) {
#level 1
            if(!(months1(minday) == 1 | months1(minday) == 7) & !(days1(
                minday) == 1)) {
#level 2
                if(months1(minday) < 7) start.grid <- datetimes(year =
                    years1(minday), month = 7, day = 1, hour = 1,
                    minute = 1, second = 1)
                if(months1(minday) >= 7)
                    start.grid <- datetimes(year = years1(minday) +
                    1, month = 1, day = 1, hour = 1, minute = 1,
                    second = 1)
            }
            else start.grid <- minday
            if(months1(start.grid) < 7)
                start.type <- 1
            else start.type <- 2
            if(!(months1(data.end) == 12 & days1(data.end) == 31) | !(
                months1(data.end) == 6 & days1(data.end) == 30)) {
# level 2
                if(months1(data.end) < 7) end.grid <- datetimes(year =
                    years1(data.end), month = 7, day = 1, minute
                    = 1, hour = 1, second = 1)
                if(months1(data.end) >= 7)
                    end.grid <- datetimes(year = years1(data.end) +
                    1, month = 1, day = 1, minute = 1, hour = 1,
                    second = 1)
            }
            else end.grid <- data.end
            if(months1(end.grid) < 7)
                end.type <- 1
            else end.type <- 2
            n <- as.numeric(years1(end.grid) - years1(start.grid)) + 1
            cat("n is:", n, "\n")
            year.grid <- c(1:2 * n)
            month.grid <- c(1:2 * n)
            blocks <- vector("list", n)    # initialise
            for(i in 1:n)
                blocks[[i]] <- c(years1(start.grid) + (i - 1), years1(
                    start.grid) + (i - 1))
            year.grid <- unlist(blocks)
            month.grid <- rep(c(1, 7), times = n)
            time.breaks <- datetimes(year = year.grid, month = month.grid,
```

```
				day = 1, hour = 0, minute = 0, second = 0)
		if(end.type == 1)
				time.breaks <- time.breaks[-2 * n]
		if(start.type == 2)
				time.breaks <- time.breaks[-1]
	}
	n <- length(time.breaks)
	cover.breaks <- (time.breaks[1] <= catalogue[, "time"] <= time.breaks[n
		])
	catalogue <- catalogue[cover.breaks, ]
	# the above makes sure there is no data outside the range of time.breaks,
#otherwise missing values would be created in the 'group' variable.
#   ( length(time.breaks)-1 intervals )
	n.group <- length(time.breaks) - 1
	group.labels <- format(time.breaks)
	group.labels.end <- format(time.breaks - 1)
	# group.labels <- group.labels[-1]
	group.labels <- paste(substring(group.labels[ - length(group.labels)],
		1, 9), "-", substring(group.labels.end[-1], 1, 9), sep = "")
	# labels correspond to end of interval
	group <- cut(catalogue[, "time"], time.breaks, labels = group.labels,
		include.lowest = T)
cat("group", summary(group), "\n")
	# 1,2,...,n.group groups of time. We use each groups ending
#   time to represent each group. Now
#       t.grid[n.group] >= data.end
#       t.grid[1] - 6months <= data.start
# now compute the basic series
	adj.first <- 1
	adj.last <- 1
	basic.sum20 <- c(table(group[in.CAT20]))
	basic.sum20[1] <- round(basic.sum20[1] * adj.first)
	basic.sum20[n.group] <- round(basic.sum20[n.group] * adj.last)
	basic.sum10 <- c(table(group[in.CAT10]))
	basic.sum10[1] <- round(basic.sum10[1] * adj.first)
	basic.sum10[n.group] <- round(basic.sum10[n.group] * adj.last)
	basic.sum20a <- c(table(group[in.CAT20a]))
	basic.sum10a <- c(table(group[in.CAT10a]))
	# Number of events in each half year period
# in the subsets
	basic.exp20a <- tapply(catalogue[in.CAT20a, "magnitude"], list(group[
		in.CAT20a]), "M8.wgt.sum")
	basic.exp10a <- tapply(catalogue[in.CAT10a, "magnitude"], list(group[
		in.CAT10a]), "M8.wgt.sum")
	# The side half year for \sum exp(0.46M) need not adjust
```

```
        basic.exp20a[is.na(basic.exp20a)] <- 0
        basic.exp10a[is.na(basic.exp10a)] <- 0
        # sum of 10^magnitude in each half year period
#  in CAT10a and CAT20a
        basic.maxB <- tapply(catalogue[in.CATMS, "n.aftershocks"], list(group[
          in.CATMS]), "max")
        basic.maxB[is.na(basic.maxB)] <- 0
        # max of number of aftershocks in each half year period
#   in CATMS
        time.breaks <- time.breaks[-1]
        # drop the first interval starting time. Now year.grid
# has n.group rows, each is a half year ending time
        i0 <- (1:n.group)[time.breaks == start.series]

#   ym.grid[i0,] == series starting time(end of the first
#               half year series).
# i0-1 is the start of the first half year
        basic.maxM <- group.max.mag(catalogue, group)
        if(debug) {
          browser()
        }
# Now compute the seven M8 series, start generating
#  at the half year ending at ym[i0,]
        cumsum.1 <- cumsum(basic.sum20)
print(cumsum.1)
        cat("i0 is", i0, "\n")  # this is the index satisfying
        cat("running.total is:", running.total, "\n")
        cat("ngroup is" )
        series.1 <- cumsum.1[i0:n.group] - cumsum.1[(i0 - running.total):(
          n.group - running.total)]
        cumsum.2 <- cumsum(basic.sum10)
        series.2 <- cumsum.2[i0:n.group] - cumsum.2[(i0 - running.total):(
          n.group - running.total)]
        series.3 <- series.1 - cumsum.1[(i0 - running.total):(n.group - running.total)]/(((
          i0 - running.total):(n.group - running.total))/running.total)
        series.4 <- series.2 - cumsum.2[(i0 - running.total):(n.group - running.total)]/(((
          i0 - running.total):(n.group - running.total))/running.total)
        cumsum.5 <- cumsum(basic.sum20a)
        cumsum.5e <- cumsum(basic.exp20a)
        series.5 <- ((cumsum.5e[i0:n.group] - cumsum.5e[(i0 - running.total):(
          n.group - running.total)])/(cumsum.5[i0:n.group] - cumsum.5[(i0 -
          running.total):(n.group - running.total)])^(0.67000000000000004
          ))
        cumsum.6 <- cumsum(basic.sum10a)
        cumsum.6e <- cumsum(basic.exp10a)
```

```
	series.6 <- ((cumsum.6e[i0:n.group] - cumsum.6e[(i0 - running.total):(
	  n.group - running.total)])/(cumsum.6[i0:n.group] - cumsum.6[(i0 -
	  running.total):(n.group - running.total)])^(0.67000000000000004
	  ))
	series.7 <- as.vector(apply(cbind(basic.maxB[i0:n.group], basic.maxB[(
	  i0 - 1):(n.group - 1)]), 1, "max"))
	series <- rts(cbind(series.1, series.2, series.3, series.4, series.5,
	  series.6, series.7, time.breaks[ - (1:(i0 - 1))]), start =
	  years1(start.series), frequency = 2)
	dimnames(series)[[2]] <- c("F1", "F2", "F3", "F4", "F5", "F6", "F7",
	  "time")
	return(list(series = series, max.events = basic.maxM[ - (1:(i0 - 1)),
	  ]))
}

"M8.TIP"<-
function(series, M0, training = "user", end.training, smoother = 6, plot.it = T,
	title = "M8 Series and TIPs", debug = F)
{
	cat("\n\n")
	cat("##########################################################\n")
	cat("#              M8.TIP()                  #\n")
	cat("##########################################################\n")
	# The series matrix, 1:7 columns are the 7 series
	x <- series$series
	if(end(x) - start(x) < 7) stoptmp(
			"Not enough years of data to generate TIP!!\n")
	# Top quantile used as critical points for each series
	alpha <- c(0.10000000000000001, 0.10000000000000001,
	  0.10000000000000001, 0.10000000000000001, 0.10000000000000001,
	  0.10000000000000001, 0.25)	# Number of half years in the series
	n <- nrow(x)		# Maximum event in each half year
	max.mag <- series$max.events[, "magnitude"]
	tops <- x[, 1:7]	# top alpha percent of history, just initialise
	exceeds <- x[, 1:7]  # boolean for crossing history top, just initialise
# The percentile of each series value, just initialise
	percent <- x[, 1:7]
	# the maximum of percent in past 3 years(6 half years)
	percent.3years <- x[, 1:7]
	# The second smallest of 1:6 columns of percent.3years, just initialise
	percent6 <- x[, 1]
	# This is min(percent6[i]-0.9, percent.3years[,7]-0.75)
	percent.all <- x[, 1]
	percent.all[1:5] <- NA
	g <- x[, 1]  # Number different types of measures
```

```
#    exceded top in the past 3 years
        h <- x[, 1]  # Number of measures exceded top in the past 3 years
        TIP <- x[, 1]          # TIP status
        TIP[1:n] <- F
        TIP.type <- x[, 1]     # TIP type
        TIP.type[1:n] <- ""    # Non-TIP initially
        for(j in 1:7) {
          ord <- order( - x[, j])
          xj <- as.vector(x[, j])
          if(training == "moving") {
                    tops[1:6, j] <- NA
                    for(i in 7:n) {
                              tops[i, j] <- xj[ord[ord < i]][ceiling((i - 1) *
                                alpha[j])]
                    }
                    exceeds[7:n, j] <- (x[7:n, j] >= tops[7:n, j])
                    exceeds[1:6,  ] <- F
                 percent.3years<- NULL
          }
          else if(training == "all") {
                    tops[1:n, j] <- xj[ord][ceiling(n * alpha[j])]
                    exceeds[1:n, j] <- (x[1:n, j] >= tops[1:n, j])
                    percent[1:n, j] <- emp.cdf(x[1:n, j], x[1:n, j])
                    for(ii in 6:n)
                              for(j in 1:7)
                                percent.3years[ii, j] <- max(percent[(ii - (
                                  smoother - 1)):ii, j])
          }
          else if(training == "user") {
                    i <- firstGE(x[, "time"], end.training)
       # this 'i' is set to one same value
                    tops[1:n, j] <- xj[ord[ord < i]][ceiling(i * alpha[j])]
                    exceeds[1:n, j] <- (x[1:n, j] >= tops[1:n, j])
                    percent[1:n, j] <- emp.cdf(x[1:i, j], x[1:n, j])
                    for(ii in 6:n)
                              for(j in 1:7)
                                percent.3years[ii, j] <- max(percent[(ii - (
                                  smoother - 1)):ii, j])
          }
          else {
                    stop(paste("Unknown value for training: ", training))
          }
          if(debug) {
                    cat("\n\n--------- Series", j, "-----------\n")
                    print(cbind(original.series = x[7:n, j], history.tops
```

```
                    = tops[7:n, j], exceeds = exceeds[7:n, j]))
    }
  }
# for j in 1:7
  if(training == "moving") {
    g[1:6] <- 0
    h[1:6] <- 0
    TIP[1:7] <- NA
    TIP.type[1:7] <- NA
    i.TIP <- 7
  }
  else if(training == "user" || training == "all") {
    i.TIP <- 1
    TIP[1] <- NA
    TIP.type[1] <- NA
  }
  for(i in i.TIP:n) {
    options(warn = -1)
  g[i] <- ((sum(exceeds[max(c(i - 5, 1)):i, 1:2]) > 0) + (sum(
          exceeds[max(c(i - 5, 1)):i, 3:4]) > 0) + (sum(exceeds[
          max(c(i - 5, 1)):i, 5:6]) > 0) + (sum(exceeds[max(c(i -
          5, 1)):i, 7]) > 0))
    h[i] <- sum(apply(exceeds[max(c(i - 5, 1)):i, , drop = F], 2,
          sum) > 0)
  # Each row is the second smallest of percent.3years[,1:6]
# print.default(percent.3years[i,1:6])
    percent6[i] <- sort(as.vector(percent.3years[i, 1:6]))[2]
  # point i vote for TIP <==> percent.all[i]>0
    percent.all[i] <- min(c(percent6[i] - 0.90000000000000002,
          percent.3years[i, 7] - 0.75))
    options(warn = 0)
    if(i > i.TIP && g[i] >= 4 && h[i] >= 6 && g[i - 1] >= 4 && h[i -
          1] >= 6) {
          TIP[i:min(c(i + 9, n))] <- T
          if(max(max.mag[(i - 1):i]) >= M0) {
                TIP.type[i:min(c(i + 9, n))] <- "c.e."
          }
          else if(i < n && max(max.mag[(i + 1):min(c(n, i + 10))]
                ) >= M0) {
# success
                TIP.type[i:min(c(i + 9, n))] <- "STIP"
          }
          else if(i < n && max(max.mag[(i + 1):min(c(n, i + 10))]
                ) >= M0 - 0.5) {
                TIP.type[i:min(c(i + 9, n))] <- "STIP-"
```

```
				}
				else if(i + 9 >= n) {
# TIP extends after end of data
					TIP.type[i:min(c(i + 9, n))] <- "CTIP"
				}
				else {
					TIP.type[i:(i + 9)] <- "FTIP"
				}
				cat("\n***TIP Declared ", TIP.type[i],
					"after half year:", dimnames(x)[[1]][i], "\n")
				if(TIP.type[i] == "c.e.") {
					tmp.i <- ((i - 1):i)[series$max.events[(i - 1):
					  i, "magnitude"] >= M0]
					print.mainshocks(series$max.events[tmp.i, ])
				}
				else if(TIP.type[i] == "STIP") {
					tmp.i <- ((i + 1):min(c(i + 10, n)))[series$
					  max.events[(i + 1):min(c(i + 10, n)),
					  "magnitude"] >= M0]
					print.mainshocks(series$max.events[tmp.i, ])
				}
				else if(TIP.type[i] == "STIP-") {
					tmp.i <- ((i + 1):min(c(i + 10, n)))[series$
					  max.events[(i + 1):min(c(i + 10, n)),
					  "magnitude"] >= M0 - 0.5]
					print.mainshocks(series$max.events[tmp.i, ])
				}
			}
# if two succesive votes
		}
# for i
		cat("\n  =============== M8 Result ===============\n")
		cat("g:h  Mmax Half-Year    F1   F2   F3    F4  F5   F6    F7 Pct TIP\n"
		  )
		cat(paste(g, ":", h, "  ", format(max.mag), " ", dimnames(series$series
		  )[[1]], " ", format(series$series[, "F1"]), c(" ", "*")[
		  as.numeric(exceeds[, 1]) + 1], " ", format(series$series[, "F2"
		  ]), c(" ", "*")[as.numeric(exceeds[, 2]) + 1], " ", format(
		  round(series$series[, "F3"])), c(" ", "*")[as.numeric(exceeds[,
		  3]) + 1], " ", format(round(series$series[, "F4"])), c(" ", "*"
		  )[as.numeric(exceeds[, 4]) + 1], " ", format(round(series$
		  series[, "F5"])), c(" ", "*")[as.numeric(exceeds[, 5]) + 1],
		  " ", format(round(series$series[, "F6"])), c(" ", "*")[
		  as.numeric(exceeds[, 6]) + 1], " ", format(series$series[, "F7"
		  ]), c(" ", "*")[as.numeric(exceeds[, 7]) + 1], " ", format(
```

```
    round(percent.all, 2)), " ", TIP.type, "\n", sep = ""), sep =
    "")
  if(training == "user" || training == "all") {
    cat("tops                ", round(tops[1, ]), "\n")
  }
  cat("\n")
  return(list(series = series$series, max.events = series$max.events, tops =
    tops, exceeds = exceeds, TIP = TIP, TIP.type = TIP.type,
    TIP.level = percent.all, training = training, M0 = M0))
}
```

Addendum

p.21 The reference to table 4.1 is Keillis-Borok and Kossobokov, 1990 (b).

p.23 The references for the case studies in Japan and California are Keillis-Borok and Kossobokov, 1990 (b) and

p. 63 The displacement of circles was half a degree, which corresponds to 50 km

p. 71 The colour gradation goes from blue, to green, to yellow, to orange, with blue corresponding to the earliest times and orange to the latest.

p.77 A key for table 6.4:

STIP-	A 'nearly successful TIP', where a earthquake occurred that was half a point less in magnitude than the one predicted.
CTIP	A 'continuing TIP', where the duration of a TIP extends in to the future, and an earthquake of the required magnitude has not yet appeared to validate it.
FTIP	A 'false TIP', where a TIP has been declared but no earthquake of the specified magnitude (or even half a point below the specified magnitude) has occurred.
c.e.	A TIP that has been initiated by an earthquake of the magnitude that the TIP itself is intended to predict.

p. 87 The reference for Bartletts Chi-squared statistic is Manley, p. 150.

p. 101 The reference for Levene's test is Levene, 1960.