Response to ‘A statistical evaluation of a ‘stress-forecast’ earthquake’ by T. Seher & I. G. Main

Stuart Crampin,1,* Theodora Volti1,† and Ragnar Stefánsson2

1School of GeoSciences, The University of Edinburgh, Grant Institute, West Mains Road, Edinburgh EH9 3JW, UK. E-mails: scrampin@ed.ac.uk; tvolti@glg.ed.ac.uk; schastin@glg.ed.ac.uk; http://www.glg.ed.ac.uk/~scrampin/opinion/ 2Iceland Meteorological Office, Bústadvegur 9, 150 Reykjavík, Iceland. E-mail: ragnar@vedur.is.

Accepted 2003 November 11. Received 2003 November 11; in original form 2003 June 27

SUMMARY
Seher & Main claim to make a statistical evaluation of our successful stress-forecast of the time and magnitude of an earthquake in SW Iceland. A statistical evaluation of the data would be valuable, but unfortunately Seher & Main cannot do this as techniques have not yet been developed for the statistical analysis of discontinuous time-series within an overall continuous data set. Consequently, Seher & Main devise their own models of continuous time-series that are fundamentally different from those of our hypothesis. Seher & Main also briefly suggest several other concepts, which are serious misunderstandings of the stress-forecasting process. Consequently, their results and conclusions are unfortunately irrelevant to evaluating either the hypothesis or the stress-forecast.

Key words: earthquake prediction, shear-wave splitting, statistical methods.

1 INTRODUCTION
Seher & Main (2004) claim to make a ‘statistical evaluation of a ‘stress-forecast’ earthquake’ by Crampin et al. (1999), referred to here as Paper 1. We welcome the statistical testing of our successful stress-forecast of the time and magnitude of an M = 5 earthquake in SW Iceland, and provided the data for Seher & Main to analyse. Seher & Main make a purely statistical analysis of the data. We recognized that we were unable to make statistical tests (beyond fitting least-squares lines) as we were reliably informed that statistical techniques that can handle discontinuous time-series with abrupt jumps in value (the increases are episodic as and when the earthquakes occur) have not yet been developed. Such techniques are still not available and the statistical technique (the Bayesian information criteria, BIC, developed by Leonard & Hsu 1999) used by Seher & Main on a continuous data set is inappropriate for our discontinuous data. Consequently, they evaluate new hypotheses, which are fundamentally different from the hypothesis we used.

Note that magnitude ‘M’ of the SIL Seismic Network in Iceland is approximately equivalent to the body-wave magnitude ‘mb’.

2 SUMMARY OF THE STRESS-FORECASTING HYPOTHESIS
We briefly summarize the hypothesis in order to show the principles underlying stress-forecasting and the reasons for the discontinuous data sets. The underlying assumption is that the deformation of rock before fracturing and earthquakes occur (pre-fracturing deformation) is controlled by fluid movement by flow or dispersion along pressure gradients between neighbouring stress-aligned grain-boundary cracks, low aspect-ratio pores and pore throats, at different orientations to the stress field. This is the anisotropic poro-elasticity (APE) model of Zatsepin & Crampin (1997) and Crampin & Zatsepin (1997). The immediate effect of changes in low-level stress is to modify the distribution of crack aspect ratios. The range of ray-path directions, where shear wave splitting is sensitive to changing aspect ratios, is the double-leafed solid angle of ray paths in the shear wave window (known as Band-1) making angles 15°–45° either side of the average crack-plane of the distributions of vertical parallel microcracks (Crampin 1999). The average crack-plane is specified in the rose diagrams in fig. 1 of Paper 1. Believed to indicate directions of maximum compressional stress, rose diagrams suggest similar crack directions throughout much of Iceland (Volti & Crampin 2003a).

Time-delays along ray paths in the remainder of the shear wave window, the solid angle ± 15° to the average crack-plane (known as Band-2), are sensitive to crack density (Crampin 1999), but no
consistent behaviour either increasing or decreasing has yet been identified before or after earthquakes along ray paths in Band-2 (Volta & Crampin 2003b).

The observed effects of increasing stress on a fluid-saturated microcracked rockmass are illustrated in Fig. 1, which shows variations of time-delays at Station BJA in SW Iceland and includes the data which allowed the successful stress-forecast of Paper 1. The first increase of time-delays in Band-1 (dashed line) from 1996 July to October is identified as the effects of stress increases before the Vatnajökull (Gjálp) fissure eruption of 1996 October 1 approximately 200 km north–east of Station BJA. Following the eruption, the time-delays do not show an abrupt decrease as they do when stress is released in an earthquake, but display a slow decrease of approximately 2 ms km$^{-1}$ yr$^{-1}$ for the two years 1997 and 1998 (Volta & Crampin 2003b). This decrease in time-delays is seen at seismic stations ~240 km north, 200 km southwest (Fig. 1), and 240 km west southwest of Vatnajökull, and is interpreted as the effects of the Mid-Atlantic Ridge adjusting episodically to the injection of 0.4 km$^3$ of magma over the 10 km length of the Vatnajökull fissure eruption (Volta & Crampin 2003b). The stress-forecasting changes identified by Paper 1 are in this 2 yr interval where the stress appears to be relatively undisturbed by other movements of magma. The time-delay variations before the five earthquakes in this 1997–1998 interval for which there are least-squares lines are superimposed on this uniform decrease. (The 2 ms km$^{-1}$ yr$^{-1}$ decrease is visible in the data set given to Seher & Main who remove the trend without comment.) In the following year, 2000, there were several volcanic eruptions that appear to disturb the behaviour of time-delays before earthquakes, and no further stress-forecast earthquakes were possible.

Note that the likely cause of the substantial scatter in time-delays in Fig. 1, previously unexplained (Volta & Crampin 2003a), has now been identified. Crampin et al. (2002, 2004) show that in parallel vertical cracks filled with critically high pore–fluid pressures, the polarizations of the faster split shear wave are orthogonal to the polarizations in normally pressurized cracks. Along any shear wave ray path to a surface seismometer, the recorded polarizations depend on whether the normally pressurized segment is longer or shorter to the high-pressured segment. If the normally pressurized segment is longer, then typical stress-oriented polarizations are observed. If the critically high pressurized segment is longer, then 90°-flips are observed. The time delay is the difference of the (positive, say) time-delay through the normally pressured segments and the (negative) time-delay through the high-pressured segments. Since the detailed distributions of stress and pore–fluid pressure vary after the stress release at each small source earthquake, this may result in substantially different time-delays, which can easily cause the substantial scatter in observed time-delays in Fig. 1 (Crampin et al. 2004).

We list the assumptions for stress-forecasting earthquakes, where the changes in time-delays need to be examined within the range of expected behaviour. Items marked (a), in the list below, specify the geophysical assumptions and items marked (b) specify the effects of the assumptions on shear wave splitting. Paper 1, Crampin (1999)
and Volti & Crampin (2003a,b) present evidence and justification for each item.

1(a) Since rock is weak to shear stress, the build-up of stress before large earthquakes necessarily accumulates over enormous volumes of rock: probably 10–100 s of millions of cubic kilometres before the largest earthquakes. This means that the effects of the build-up of stress on shear wave splitting can be seen at substantial distances from the eventual earthquake source zone.

1(b) The stress accumulation can be monitored by increases in time-delays between the split shear waves in Band-1 directions of the shear wave window almost anywhere within this larger stressed volume (Crampin 1999). The least-squares lines are fitted to these (discontinuous) segments of the time-delay data. The start of each line is taken from a minimum in a nine-point moving average of time-delays in Band-1 of the shear wave window, and the end of the line is determined not by the time-series but by the occurrence of a larger earthquake.

2(a) The accumulating stress modifies microcrack geometry until distributions reach levels of fracture criticality when the earthquake occurs and there is an abrupt drop in the value of the time-delays.

2(b) The time of the earthquake can be stress-forecast from the estimated time that the line through the increasing time-delays reaches equivalent levels of fracture criticality, estimated from the behaviour of shear wave splitting before smaller earlier earthquakes. The level of fracture criticality at BIA varies from approximately 14 ms km$^{-1}$ following the Vatnajökull eruption to approximately 11 ms km$^{-1}$ in 1998 November. The level of time-delays decreases abruptly following larger earthquakes, although the nine-point moving average may partially disguise the abruptness of the decrease.

3(a) We assume a relatively uniform increase of overall stress associated with movement of the Mid-Atlantic Ridge. If the stress accumulates over a small volume, the effects of the increase will be rapid, but the final earthquake will be small. If similar stress accumulates over a larger volume, the effects of the increase will be slower, but the eventual earthquake will be larger.

3(b) In the only quality data set available to us, the comparatively small magnitude range $M = 3.8–5.1$ in 1997 and 1998 (Fig. 1), the relationship between magnitude and slope of fracture criticality, estimated from the behaviour of shear wave splitting before smaller earlier earthquakes. The level of fracture criticality at BIA varies from approximately 14 ms km$^{-1}$ following the Vatnajökull eruption to approximately 11 ms km$^{-1}$ in 1998 November. The level of time-delays decreases abruptly following larger earthquakes, although the nine-point moving average may partially disguise the abruptness of the decrease.

3(a) We assume a relatively uniform increase of overall stress associated with movement of the Mid-Atlantic Ridge. If the stress accumulates over a small volume, the effects of the increase will be rapid, but the final earthquake will be small. If similar stress accumulates over a larger volume, the effects of the increase will be slower, but the eventual earthquake will be larger.

3(b) In the only quality data set available to us, the comparatively small magnitude range $M = 3.8–5.1$ in 1997 and 1998 (Fig. 1), the relationship between magnitude and slope of fracture criticality, estimated from the behaviour of shear wave splitting before smaller earlier earthquakes. The level of fracture criticality at BIA varies from approximately 14 ms km$^{-1}$ following the Vatnajökull eruption to approximately 11 ms km$^{-1}$ in 1998 November. The level of time-delays decreases abruptly following larger earthquakes, although the nine-point moving average may partially disguise the abruptness of the decrease.

4(a) Note that although the actual build-up of stress may be comparatively regular and uniform, our observations of time-delays and the unavoidable scatter will mean that any estimates of the level of fracture criticality and the time and magnitude of the impending earthquake will be subject to errors.

4(b) We try to account for this by using the shear wave splitting to estimate the range of time and magnitude in a SELL window. Currently, this is a subjective estimate which takes account of: the errors in estimates of: time and magnitude; the level of fracture criticality; and any variation in the rate of build-up of stress.

Justified by the arguments in these items, the following stress-forecast was issued on 1998 November 10. The e-mail message from the Department in Edinburgh to the Iceland Meteorological Office was the SELL window: ‘... an event could occur any time between now ($M \geq 5$) and the end of February ($M \geq 6$).’ (Paper 1). The forecast was successful as an $M = 5$ earthquake occurred three days later on November 13. Although, the pervasiveness of stress-changes (item 1, above) do not allow earthquake locations to be estimated from shear wave analysis alone, knowing that a larger earthquake is approaching allows other possible precursory behaviour to be realistically interpreted. Continuing small-scale seismic activity on the fault zone of the previous $M = 5.1$ earthquake in 1998 June, allowed the fault of the stress-forecast earthquake to be correctly predicted in Paper 1.

3 SUMMARY OF STATISTICAL IMPLICATIONS

The purely statistical implications of the above behaviour is that stress-forecasting is made from a series of increasing least-squares lines fitted to discontinuous segments of the data set. The start of each line is determined from a minimum in the nine-point moving averages. (Note that there is no particular significance to the nine-point moving average. It has no statistical relevance and is merely a convenient ‘summary’ of the variations in time-delays.) The delays increase until there is an appropriate larger earthquake. Typically the end is marked by a decrease in stress and a drop in the value of shear wave splitting. Because of the averaging, this drop is not necessarily reflected immediately by a drop in the nine-point moving average.

In the case of the forecast earthquake, there was a precursory drop in time-delays (and implied stress) before the main shock (Volti & Crampin 2003b), but the sequence of foreshocks and aftershocks induced rapid variations in time-delays which are irrelevant to the stress-forecasting process. (This precursory drop occurred before the stress-forecast, but we were not aware of this as the data were not yet available on the internet, and the corresponding time-delays had not been measured.) Thus the least-squares lines are fixed by the time-delay data and the occurrence of larger earthquakes. They are almost entirely constrained. The only judgement required is to estimate the position of the minimum of the nine-point moving average from which to start the least-squares line for the stress increase before an incipient earthquake. Typically the minimum is obvious and there is no choice of timing, but occasionally a decision between several options is required, however, such typically small differences in timing are easily accommodated by usually minor changes to the range of the SELL window.

In general, the time-delays following the larger earthquakes represent relaxation phenomena following the release of stress by the earthquake. They are irrelevant to stress-forecasting the next larger earthquake, apart from providing the minima in the nine-point moving average used to identify the start of the accumulation of stress.

In particular, the intervals of relaxation between the least-squares lines are not included in the stress-forecasting analysis of Paper 1. This means that in the 22 month period 1997 January–1998 October (Fig. 1), only approximately 58 per cent of the time were the data responding to stress changes before five impending earthquakes (four recognized with hindsight and one in real time). These changes are episodic. Consequently, the stress-forecast cannot be made with statistical analysis of a continuous data set. The forecast tries to assess stress accumulation before the earthquake. What happens after the earthquake has occurred is irrelevant to the stress-forecasting hypothesis as proposed.
A simple analogy for the mixing of two inappropriate data sets would be estimating the average height of boys from the average height of a mixed population of boys and girls. Whatever the outcome it cannot be a statistically meaningful average for the height for boys, as the girls could be pygmies or amazons.

4 ANALYSIS OF SEHER AND MAIN

4.1 Use of Bayesian information criteria

The bulk (∼80 per cent) of Seher & Main (2004) is an attempt to find the goodness of fit for competing statistical models by using the Bayesian information criteria developed by Leonard & Hsu (1999). BIC assesses the statistical significance of any forecast, and penalises complex models with which have larger numbers of parameters. Although BIC is valid for irregular sampling, it assumes a continuous data set where every point has equal weight. Techniques for statistical analysis of discontinuous segments of a continuous data set, as in Fig. 1, are not available and would require considerable effort to develop (Thomas Leonard, private communications, 1999, 2003). This work has not been done. Consequently, Seher & Main avoid the issue by analysing models with continuous data sets.

They try a variety of models, including: a constant (Poisson) model; a sinusoidal model; and various multilinear models, and reject all models as statistically inadequate. They also define a model includes the least-squares lines (1997 January–1998 October) in Fig. 1 but join the end of each line to the start of the following line to result in a continuous multilinear model. They claim that ‘We reproduced the model of increase in normalized time-delays suggested by Crampin et al. (1999)’. Adding 42 per cent of statistically irrelevant data to make it continuous is not meaningful. In particular, following an earthquake, it is observed that the time-delays immediately fall in value to approximately half the maximum value. They do not decrease linearly to the start of the next increase. Consequently, the results and conclusions of the ‘BIC analysis’ of Seher & Main are irrelevant and can make no valid evaluation either of the hypothesis or of the stress-forecast.

There are several other details in the BIC analysis that are worthy of comment.

(1) Seher & Main claim that there are a large number (17) degrees of freedom in our model, and use this parameter in their analysis. In fact, the model is almost wholly constrained and repeatable, but not completely based on the time-delay data set (as the end of a least-squares line is determined by a larger earthquake). The five most pronounced minima in the nine-point moving average for 1997 and 1998 in Fig. 1 are fixed by the data (apart from small ambiguities in timing which lead to small changes in the SELL window), and the stress-forecast follows from the least-squares line following the minima. The only degrees of freedom are the judgement of: (a) whether the analysis of the data is reliable and justifies a stress-forecast; and (b) the range, particularly the start, of the SELL window. In principle, assuming good data, a large part of this process could be automated but we currently have too few examples to be able to establish a pattern of behaviour and, in view of irregular data, no doubt a judgement will always be required.

(2) Note that in plots of such time-series, the ratio of vertical to horizontal axes is important for visual identification. If the horizontal axis is too short, as it tends to be in fig. 2 of Seher & Main, the data points are clustered and the increases are not visible.

(3) Seher & Main note that to be valid the BIC criteria of Leonard & Hsu (1999) require data sets of more than 46 points. The episodic least-squares lines indicating increasing stress in 1997 and 1998 in Fig. 1, each have less than 25 points. So that even for the discontinuous data sets analysis in terms of BIC is inappropriate.

(4) Seher & Main suggest that the ‘. . . scatter in the data is the main reason for the lack of definitive results.’ We claim at least one definitive result despite the scatter (the successful stress-forecast, and ten cases, including the four seen at BJN in 1997 and 1998, where the effects have been seen with hindsight), but agree that the scatter is a serious problem. The scatter has now been explained by high fluid-pressures on seismic-active fault-planes causing 90°-flips in shear wave polarizations (Crampin et al. 2002, 2004), but the scatter cannot be eliminated as we have insufficient knowledge of the rock mass and stress regime which varies after each source earthquake (Crampin et al. 2004). In contrast to the large scatter observed above small earthquakes, shear wave time-delays measured in reflection surveys, and vertical seismic profiles, in rocks away from seismic-active fault-planes in controlled-source show minimal scatter (Crampin 1999). We have recognized the need for controlled-source experiments data in the EC funded SMSTES Project. SMSTES measures traveltimes from a controlled-source between 500 m deep boreholes in Iceland. This is now indicating the extraordinary sensitivity of fluid-saturated microcrack rock to small tectonic (seismic) disturbances at 70 km (several hundred times the likely fault dimensions) from the energy equivalent of an $M = \sim 4$ earthquake (Crampin et al. 2003). This unique data set confirms that both the science and technology of stress-monitoring sites are valid for stress-forecasting impending large earthquakes.

(5) Seher & Main also claim that ‘The main source of scatter is probably the use of ray paths with different azimuths, thereby mixing transverse and azimuthal anisotropy.’ This is incorrect on two counts. The source of the scatter is now recognized, see previous item, and is not caused by ‘different azimuths’. Secondly, mixtures of different forms of anisotropy do not necessarily cause scatter in time-delays. Combinations of transverse and azimuthal anisotropy produce specific orthorhombic anisotropic symmetry which in many combinations cause comparatively smooth variations of properties (Wild & Crampin 1991). The only irregularities are the effects of the deviations of group velocities at shear wave singularities, and these may occur in all anisotropic symmetry systems.

4.2 Continuity of data

The major disagreement with Seher & Main is that they insist that the data set is continuous. Our analysis is based on the reasonable assumption that the behaviour after the event is not necessarily related to the accumulation of stress before the earthquake. Seher & Main are mislead by the series of five earthquakes in a two year time span. In an area less seismic than SW Iceland, there might be, say, one large earthquake that might be stress-forecast in ten years of data. By implication Seher & Main would analyse the whole ten years data. We suggest the analysis of ten years of data would clearly be irrelevant to trying to forecast a one-off event if there was an identifiable change immediately before the earthquake.

The continuous model that Seher & Main claim is our data set, where the ends of our least-squares lines are joined by straight lines to the starts of the next lines is particularly inappropriate. Our hypothesis is that following an earthquake the stress and time-delays drops abruptly (unless modified by stress release by aftershocks), and in particular, does not decrease linearly to the start of the next increase as in the Seher & Main model.

Seher & Main may interpret the data as continuous, but they demonstrate that the continuous data set does not have a useful
meaning with respect to earthquake occurrence. Whereas the discontinuous data sets, that we analyse, have stress-forecast times and magnitude of nine earthquakes with hindsight, and one in real time (Paper 1).

4.3 Misleading error bars

Seher & Main appear to misunderstand how magnitude is related to time-delays in the stress-forecast. Fig. 3 of Seher & Main inappropriately combines values of normalized time-delays (left-hand axis) with equivalent magnitudes (right-hand axis). It is the rate and duration of change of time-delays, not their value, which are related to earthquake magnitude (item 3b in Section 2, above). Consequently, the magnitude and time errorbars (the ‘large cross’: $3.9 \leq M \leq 6.7$ and $-42$ d to infinity) derived from time-delays, but interpreted in terms of magnitude, are both incorrect and misleading. The only correct elements of their fig. 3 are the time delays referred to the left-hand axis, and the successful time–magnitude forecast of Paper 1 (referring to the right-hand axis), which starts three days before the earthquake occurred and ends three months later. Since the ‘large cross’ errorbars are referred to both directly and indirectly in Seher & Main’s summary, discussion and conclusions, their interpretation is incorrect and misleading.

4.4 Gutenberg–Richter relationship

Fig. 4 of Seher & Main is complicated and it is not clear what it represents. Minimally defined in a two-line caption, it attempts to relate the incremental frequency magnitude relationship of Gutenberg–Richter with the time–magnitude relationship of the stress-forecast. Gutenberg–Richter is an overall statistical relationship between magnitude and cumulative occurrence, which all earthquake data are necessarily subject to. However, the duration is reset at each minima of the nine-point moving average. Hence, duration and incremental frequency are very different quantities and the relevance of combining them is not demonstrated. Incidentally, the Seher & Main equation for the time–magnitude relationship is not good. $T = -11.5 + 4.2 M$ describes the data (Fig. 2a, Volti & Crampin 2003b) much better than $T = -6.91 + 3.05 M$, and it is misleading to plot the equation for the (curved) section $2.5 \leq M \leq 5.1$ when it is only defined for the short section: $3.5 \leq M \leq 5.1$, which is almost linear in their fig. 4.

Seher & Main suggest that ‘The graph also shows the estimated waiting time that (sic) might be predicted from the background seismicity (inverse frequency) so that direct comparison can be made.’ If this were true it would quantify the relationship between the physical process as stress accumulates in the overall rock mass and the eventual earthquake magnitude. Although this would be extremely valuable if true, it is highly unlikely and has certainly not been demonstrated.

4.5 Time-delay magnitude relationship

Seher & Main try to draw conclusions from a relationship between shear wave time-delays and magnitudes plotted in their fig. 5. As stated above (item 3b, Section 2), the magnitude of the earthquake is not related to the values of time-delays. Consequently, their fig. 5 is spurious.

5 CONCLUSIONS

Theory and observations show that shear wave splitting can monitor the build-up of stress and assess when the crack geometry reaches levels of fracture-criticality and the earthquakes occur (Crampin & Zatspin 1997; Zatspin & Crampin 1997). Observations in Iceland (Volti & Crampin 2003a,b) show that the duration of stress build-up is directly proportional to, and the rate of increase is inversely proportional to, the size of the impending earthquake, in the limited magnitude range available.

Examining the two years 1997 and 1998 in Fig. 1. The first four least-squares lines specified as above were fitted with hindsight. It was recognized in 1998 October that the fifth line was so similar to the previous line before the $M = 5.1$ earthquake of 1998 June, that the stress-forecast was not difficult to make. Items 1–4, above, in Section 2, attempt to rationalize and categorize, we believe successfully, the effects in the first four least-squares lines that gave the successful stress-forecast of the fifth event.

As Leonard, one of the originators of the Bayesian information criteria (Leonard & Hsu 1999) on which Seher & Main base most of their analysis, considered that BIC is not valid for discontinuous data sets, this cast doubts on the validity of the analysis of Seher & Main. We suggest that the vast number of failed attempts to predict earthquakes by the purely statistical analyses of earthquake times, locations, magnitudes and other parameters, indicates that purely statistical analyses ignoring geophysics must be treated with great caution.

Figs 3–5 of Seher & Main appear to result from serious misunderstandings of the stress-forecast relationships. Consequently, in the context of the evaluating the stress-forecast, these diagrams and conclusions therefrom are meaningless. Statistical analysis of geophysical data has to take account of the geophysics. Thus forcing data to be continuous just to permit BIC analysis is invalid.

Finally, the fact that the first ever stress-forecast was successful, suggests that either there was a highly unlikely statistical coincidence, or there must be something significant in the data in Fig. 1. We suggest it would be more useful for statistics to demonstrate why the stress-forecast was successful, rather than attempt to show by inadequate techniques why what had actually occurred was impossible.

ACKNOWLEDGMENTS

This paper was partially supported by the European Commission SSMITES and PREPARED Projects, contract numbers EVR1-CT1999-40002 and EVG1-CT2002-00073, respectively. We thank Sergei Zatspin for valuable comments and discussions.

REFERENCES


