[redacted]

VICTORIA UNIVERSITY
OF WELLINGTON

PO Box 600, Wellington, New Zealand

INSTITUTE OF STATISTICS
AND
OPERATIONS RESEARCH

Statistical Modelling of Felt Intensity Data for Three New Zealand Earthquakes

A Report to the New Zealand Earthquake Commission 93/141

by David Vere-Jones
Jiayan Pang
Euan G. Smith

Statistical Modelling of Felt Intensity Data for Three New Zealand Earthquakes

by David Vere-Jones [1]
Jiayan Pang [2]
Euan G. Smith [3]

Abstract

The aims of this study are to develop objective statistical methods for fitting attenuation curves and smooth isoseismals of simple form to felt intensity data. The main analyses were carried out with data from three major New Zealand earthquakes: Buller (1929), Inangahua (1968) and Edgecumbe (1987). Both non-parametric and parametric approaches were used, making use of subroutines available within the statistical language S-PLUS.

The non-parametric methods produced useful results for the attenuation curves, less so for the smoothed isoseismals. A simple elliptical model, representing expected intensity as a function of distance and angle relative to the centre of shaking, was then fitted to the felt intensity data using two approaches. It was fitted first directly by minimising the sum of squares between observed and expected intensities, and secondly indirectly by using the expected intensity to generate a family of probability distributions for the observed intensity, and fitting these probabilities by maximum likelihood.

Even with such simplified forms, the three events gave rise to significantly different parametric values. However the attenuation curves (corrected for different relative attenuations along the major and minor arcs of the ellipse) were very similar to those estimated non-parametrically, and close also to a common power law decay with distance (geometric spreading), with index somewhat below 2. All three fits also suggested an elliptical pattern of isoseismals, with the major axis oriented approximately NNE/SSW. However the ellipticity term was clearly statistically significant only for the Inangahua earthquake.

Preliminary comparisons of the histograms of observed and expected numbers of felt intensities of different values within a given isoseismal band suggested that more consistent isoseismal patterns could be obtained by correcting the biases due to under-reporting of small intensities and absence of reports from regions with low population densities.

[1] Institute of Statistics and Operations Research, Victoria University of Wellington
[2] Department of Geology, Victoria University of Wellington
[3] Institute of Geophysics, Victoria University of Wellington

1 Introduction

Making the best use of felt intensity observations is a challenging methodological problem. They contain unique information about both short-range and long-range patterns of shaking, about attenuation and azimuthal dependence, and about the local effects of particular geological structures. At the same time the extreme unevenness in the spatial distribution of reported observations, coupled with the low reliability of individual reported values, puts unusually severe obstacles in the path of routine statistical analyses. The total absence of intensity observations from the sea, which in different cases may be near or far from the centre of shaking; the effects of dense clusters of observations from local centres of population; and of regions of sparse or even non-existent observations from wilderness areas, are characteristic causes of difficulty.

As a consequence of these difficulties, the drawing of isoseismal maps remains an area where objective statistical methods commonly yield place to the use of expert knowledge and experience. Nevertheless there have been an number of important attempts to develop more objective procedures. There have been several previous studies of MM intensity attenuation in New Zealand. Smith (1978a) developed an empirical attenuation model and used it to estimate seismic hazard in New Zealand. Smith's study allowed for the ellipticity observed in a number of the isoseismal maps, but used as a starting point the interpreted isoseismals rather than the original intensity observations. Such an approach has the drawback that any subjective biases or inconsistencies between interpreters are built into the data from the outset. Other studies, such as Dowrick (1991, 1992), have used standard functional form to fit attenuation against distance, but without allowing for ellipticity. Some overseas studies (see for example Ortolani et al (1987)) have developed quite elaborate models using orthogonal series in radial and angular components for fitting intensity data, but face the considerable danger that the resultant curves reflect accidental groupings of intensity observations rather than any underlying physical pattern of attenuation behaviour. Particular mention may be made of the paper by Brillinger and Preisler (1985) which was one of the first to illustrate the possibilities of using modern non-parametric methods to intensity data. Very recently Kozuch (1995) has extended the method of the simple functional form to include systematic site effects.

This paper represents a further attempt to develop more effective statistical approaches to modelling felt intensity data, taking advantage of the wider range of techniques available within modern statistical languages such as S-PLUS, but without developing special-purpose software. The main aims of the paper are:

1. To develop more objective approaches to estimating attenuation factors and large scale isoseismal patterns from felt intensity data.
2. To use these approaches to better understand the similarities and differences between the felt intensity patterns for different New Zealand earthquakes, and between New Zealand and overseas earthquakes.

In pursuit of these aims a standardised data base was set up comprising felt intensity data from three major New Zealand earthquakes, as well as a number of overseas earthquakes, the latter serving as a reference and comparison.

These data sets were then modelled in several different ways. First the non-parametric regression methods available in S-PLUS, in particular the Loess routines, were used to obtain smoothed attenuation curves and simplified isoseismal patterns. Then a simple model was developed for the expected intensity at a given distance from and orientation about the centre of shaking, and fitted by the S-PLUS non-linear least squares procedure. Finally, the observed intensities were modelled by a parametric family of discrete distributions, and fitted by the general likelihood maximisation routine from S-PLUS.

We shall make the following broad distinction between parametric and non-parametric procedures – a distinction which is becoming increasingly blurred as techniques emerge, such as thinplate smoothing splines, which incorporate some features of both. In a parametric modelling situation, the main features of the data are built into the model structure itself, leaving only a few parameters to be fitted from the data. In a non-parametric situation, however, the numbers of free parameters may be very large, so that very different types of data may be accommodated, and the obvious problems of overfitting are controlled by the choice of smoothing parameters, such as a bin-width or bandwidth, which may be left largely to subjective criteria, or handled more objectively by model selection procedures such as cross-validation, dividing the data into training and evaluation sets, etc.

The general strategy we shall employ is to use non-parametric methods as an intermediate stage between the raw data and simplified elliptical or circular parametric models for the isoseismal patterns. The non-parametric methods are more responsive to local fluctuations and can be helpful therefore in showing up the existence of departures from the simplified model patterns, and in roughly indicating their character. The simplified parametric models are useful particularly in comparing and combining information for groups of events, and for developing common models which can be used with a reasonable degree of confidence for predicting attenuation effects and isoseismal patterns throughout New Zealand.

One particular difficulty which arises in developing objective methods for drawing isoseismals is the lack of clarity over precisely what an isoseismal contour is. It would seem that the concept of Isoseismal has never been precisely or consistently defined – presumably because of the difficulties caused by the great spread of distances across which an individual MM Intensity might be observed. According to Richter (Richter 1958, p17):

> Isoseismals ... were at first mapped as curves connecting localities where equal intensity was observed in a given quake. They are now more commonly mapped as boundaries between regions of successive intensity ratings, such as IV and V.

In fact, neither definition has ever been, or could ever be, exactly achieved. The breadth of the zone containing any Intensity makes the earlier concept difficult to implement consistently, while attempts at implementing Richter's latter-day usage, of Isoseismals enclosing all observations, will invariably result in some observations of an Intensity lying beyond that Intensity's isoseismal.

Consistency can be achieved with a distributional definition, i.e. defining Isoseismals to be lines enclosing (on average) a fixed percentage of the observations of that Intensity. The percentage could be 50% – corresponding approximately to the original use of Isoseimal, or something much higher, such as 80%, to correspond with the most common current practice. In this paper we treat it as the contour line for a particular intensity value in a model for expected intensity. If the model fits well, and effects of bias are allowed for, this should correspond approximately to a line passing through the middle of the band of observed intensities of that value. This seems more convenient, for an objective definition, than defining the isoseismal contour as a line encompassing some arbitrary proportion (75%, 90%) of the observations of a given intensity value. However, should such a model be required, it could readily be developed from the procedures outlined in §4 – 5.

Details of the data sets used are given in §2, while §3, 4, 5 outline respectively the applications of the non-parametric, least squares, applications of the non-parametric, least squares and maximum likelihood procedures. §6, contains a brief discussion of the fit of the models, in terms of the relation of the observations to the fitted isoseismal contours, while §7 summarises the main findings and indicates some directions for further study.

2 Observational Data Base

Three well-observed New Zealand earthquakes (Edgecumbe (1987), Inangahua (1968) and Buller (1929) were selected as the basis of the study. These were supplemented by felt data from a group of historical Italian earthquakes, supplied by courtesy of The Italian Commission for Nuclear and Alternative Energy Resources (ENEA), and a group of earthquakes from Sichuan (China) supplied by courtesy of the Sichuan branch of the State Seismological Bureau, China.

Data for the three New Zealand earthquakes made use of the revised values prepared recently by the New Zealand Seismological Observatory (Downes (1995)) – see Figures 1 - 3. These maps initially used different projections, i.e. a rectangular projection for the near epicentral area (within 70 km) of the Edgecumbe earthquake, and a Lambert conformal conic projection for the large-scale Buller and Inangahua data. These were all converted to standard latitude-longitude coordinates by a two-stage process, first digitising the locations using the Sierra software package from the VUW Institute of Geophysics and then using an approximate interpolation procedure to convert from the map coordinates to latitudes and longitudes. The final format used for all data sets was latitude and longitude in decimal fractions, felt intensity in integers on the Modified Mercalli scale. (It should be noted that the versions of the MM scale used in Italy, China and New Zealand are each adapted to local conditions and so may not be exactly comparable.)

Summary information concerning the data sets used in the study is shown in Table 1. The epicentre is unsuitable to use to define the centre of shaking, as it indicates the location of initial movement and typically lies towards the edge of the region of maximum shaking. To allow for this effect, the centre of shaking for each of the earthquakes was defined as the mean of the latitude and longitudes of all intensity estimates at the maximum recorded intensity level.

The events from outside New Zealand have not been used in the present study, but are intended to serve as the basis for future comparative studies.

3 Non-parametric estimates of intensity and attenuation

The use of modern non-parametric methods to fit attenuation data goes back at least to the study by Brillinger et al (1985) of the Joyner-Boore data on peak accelerations (Joyner and Boore (1981)). Then concern was not only to smooth the data but also to seek modified functional forms for the attenuation law. Here we shall be concerned with non-parametric methods simply as a means of flexibly and objectively smoothing the data. We have applied the non-parametric procedures in three ways: first to fit smoothed attenuation laws (intensity v. distance from centre of shaking); second to develop isoseismal lines as contours on a smoothed map of intensity as a function of latitude and longitude; and third to present plots of the residuals from fitting the simplified model described in the next two sections. The first two of these applications are described in the present section; the third will be described briefly in §5.

3.1 Attenuation of intensity with distance

Plots of the intensity I_j against radial distance from the centre of shaking, are shown in Figures 4 - 6. The radial distances have been converted to kilometres for ease of comparison and interpretation.

S-PLUS contains several procedures for non-parametric regression with one response variable (intensity) and one explanatory variable (radial distance). The featured method, "Loess", is a local regression procedure which fits a polynomial of given order to the data in the neighbourhood of the point at which the estimate is required, and returns as the estimate for that point the value there of the fitted polynomial. Adjustable parameters include "span" (the length of the neighbourhood over which the local regression is fitted) and "degree", the degree of the fitted polynomial. Special adjustments are used near the ends of the range of the explanatory variable; we refer to Cleveland et al [1991] for further details.

Other methods include a kernel smoothing method, in which the estimated value at a given point in the curve is a weighted average of the observed values in the neighbourhood of the point, and a smoothing spline which in this 1-dimensional context fits cubic splines with knots at the data points.

Plots 4 - 6 show the results of using each of these methods (Loess, kernel, spline) on each of the New Zealand data sets. Only a rough adjustment of the smoothing parameters has been made, and it is clear for example that the choice of band-width for the kernel method undersmooths the Buller and Inangahua data, but rather oversmooths the Edgecumbe data. Nevertheless the three methods give reassuringly similar results. The behaviour for large distances is not well controlled, as a result both of the paucity of small intensity observations and the limitations of the method near the ends of the variable range.

Other features to note are:

1. the extreme scatter of distances over which a given intensity value may be observed, and
2. the differences in behaviour of these sets of curves for high intensities.

In particular, the absence of high intensities for the magnitude 7.8 1928 Buller earthquake is to be attributed largely to the lack of observations close to the epicentre, which was located in a largely uninhabited wilderness region. The attenuation curves are clearly affected by this feature, particularly when compared with the smoother magnitude 7 Inangahua earthquake, the epicentre of which was close to several population centres. The scatter of distances for middle values of the intensity is accentuated by the high ellipticity of the isoseismal contours: similar values would thus be expected at different distances along the major and minor axes of the ellipse. A method of correcting for this effect will be discussed in §4, where the non-parametric fits are compared to fits from the parametric model.

3.2 Non-parametric fitting of isoseismal lines

As a general rule the amount of data needed to obtain satisfactory results from non-parametric methods increases rapidly with dimension. This is partly due to the "curse of dimensions" – many more data points are needed to obtain a given density in a 2- or 3-dimensional region than in a 1-dimensional interval of similar linear dimension. Although all three methods described in the previous subsection can be extended to 2 and higher dimensions, only "Loess" is immediately available in the S-PLUS release used for these studies. The smoothed intensity estimates from this routine are evaluated on a grid of spatial locations, forming a matrix of values which is then fed into a contouring routine to give families of smoothed contours.

The results of applying the "Loess" routine to the Edgecumbe data are shown for different values of the smoothing parameter in Fig. 7. The extensions of the contours into the sea, which are not based on actual values, are meaningless and should be disregarded. The contour lines are reassuringly stable for values of "span" between 0.4 and 0.6. The fits to the Buller and Inangahua data (Figs. 8 and 9) seem less satisfactory. For values of the smoothing parameter "span" in the same range (0.4 - 0.6) the contour lines are reasonably stable, but somewhat biased, the observed high intensities near the epicentres towards the West Coast of the South Island appearing to be seriously underestimated. This seems to be a consequence of the large scale of the overall map relative to the scale of the high intensity contours – in effect the high values occur in a small cluster and are averaged out by the non-parametric smoothing. Reducing the value of the smoothing parameter reduces the bias, but at the cost of producing irregular and unstable contours. The effect may be exaggerated for the Buller and Inangahua earthquakes because the high values occur near the Western boundary of the region, causing an apparent East/West trend, which may have been enhanced by the local linear regressions selected for the plots (degree = 1; although it was also present for other values of "degree").

While these problems could doubtless be improved by more careful selection and adjustment of the parameters, their main cause is the irregular pattern of the data point locations, which makes local smoothing a very tricky operation. Such difficulties are one of the main reasons why we have returned to the use of simple parametric models, involving global estimation of the parameters. Our impression at this stage is that local smoothing methods may be best left to the graphical display of residuals from fitting a simple parametric model.

4 Parametric modelling using least squares

The S-PLUS routines include a non-linear least squares programme (nls) which can be applied to a wide range of modelling situations. Although the use of least squares as the fitting criterion is unlikely to be optimal in the present circumstances (this would require Gaussian data rather than the integer-valued intensity data here available) the fitting algorithms are robust, easy to understand, and can provide useful results even in non-optimal situations. In our context the least squares procedures are useful as a prelude to fitting a more complete model for the probability distribution of the intensities at different locations around the centre of shaking.

The model used assumes

- elliptical contours about the centre of shaking
- attenuation which includes terms both for geometrical spreading and for non-elastic energy dissipation.

These assumptions lead to estimating the intensity I at a point with polar coordinates (r, θ) about the centre of shaking by the model

$$I = I_0 - a\log(1 + R/\sigma) - bR + E \tag{1}$$

where

I is the estimated intensity at a point $P(r, \theta)$ with polar coordinates (r, θ) about the centre of shaking;

I_0 is the maximum intensity ("predicted" to occur at the centre of shaking);

$R = r\sqrt{1 - e^2\cos^2(\theta - \epsilon)}$ is the "elliptical distance" of P from the centre, i.e. its distance reduced to the minor axis of the ellipse on which it lies;

e is the eccentricity of the ellipse;

ϵ is the orientation of the major axis of the ellipse relative to the parallel of latitude through the centre of shaking;

a is the index of the power law for the attenuation due to geometrical spreading ($a = 2$ if the wave spreads uniformly in 3 dimensions);

b is the coefficient of the attenuation due to anelastic energy absorption (friction etc.);

σ is a scale factor which effectively controls the size of the highest MM isoseismals;

E is the error term needed to produce integer values of I.

Attempts to fit the model in this form led to a difficulty with the term bR. As can be seen from Figures 4 - 6, many of the data sets show unreliable values at large distances from the source. For the Edgecumbe earthquake, for example, the only recorded observations 60 km from the epicentre are for MM 6 and 7, and there are no recorded observations of MM 4 or 5. Fitting the full model (1) then leads to negative estimates of b, which are clearly un-physical and result in nonsensical attenuation contours for large distances. It was concluded that the data quality did not permit meaningful fitting of the full model and so the term bR was dropped in all subsequent calculations, which were therefore based on the simplified model:

$$I = I_0 - a \log(1 + r/\sigma) + E \tag{2}$$

In this form the model was fitted to all three New Zealand data sets, with results set out in Tables 2 and 3 and illustrated graphically in the LH parts of Figures 10 - 12.

Table 2 shows the parameter values, residual sums of squares, and Table 3 shows the correlation matrices from fitting model (2) to each of the three New Zealand earthquakes. There are some substantial numerical differences between the three sets of parameters, notably in the value of the scale parameter σ.

Assessment of these differences is complicated by the spatial clumping of the data and the near-collinearity in the parametrisation chosen for the model (2). This is best understood by rewriting (2) in the form:

$$I = I_0 + a \log \sigma - a \log(\sigma + R) + E$$

For moderate to large R, the last term is largely independent of σ; for this range of R at least, $\log \sigma$ behaves simply as a contribution to the constant term, so that an increase in I_0 is largely offset by a decrease in σ and vice versa. Similar comments apply to the relation between I_0 and a. These correlations lead to a loss of stability in the least squares minimisation and to lack of precision in the estimates.

While these defects could be remedied to some extent by a more careful equivalent parameterisation, a fundamental source of difficulty lies in the irregularity of the spatial distribution of the data and the crudeness and variability of individual intensity measurements.

The upshot of these difficulties is that relatively minor or accidental features of the locations of intensity measurements for a given earthquake can produce quite large changes in the least squares fit. This calls into question the physical significance of the differences in parameter values between the earthquakes. For example, the parameter a is controlled by the properties of the earth and should be similar for earthquakes of similar size at similar locations, such as the Buller and Inangahua events. Similarly one might expect the form of the isoseismals well away from the source to be controlled by geological factors and thus be common for events in the same part of the country. The parameter σ should reflect the size of the seismic source, and so should vary in a systematic way with magnitude or seismic moment.

To examine such effects various simplified versions of the model were run and compared with the results from (2) itself. In particular, the second part of Table 2 shows the results of fitting the model with attenuation factor $a = 2$, the value it should have for pure geometrical spreading, and $\sigma = 10$. The statistical significance of the differences in fit can be assessed by comparing the increase in the residual sum of squares with the square of the residual standard error. For the Buller earthquake, for example, fixing $a = 2$ leads to an increase in the sum of squares of approximately 10 where the squared residual standard error is approximately $(0.71)^2 = 0.5$, representing roughly a χ^2 value of 20 on 1 degree of freedom or a t-value of about 4.5 on 114 d.f. While the effect for Buller is clearly significant, there is virtually no reduction in the quality of fit for the Edgecumbe event. Inangahua is intermediate between these.

The Buller event is a special case because of the absence of high intensity observations in the epicentral region. The model (2) tends therefore to choose a combination of parameter values which gives a broad region of relatively low intensities near the epicentre, i.e. an anomalously large value of σ combined with a relatively high value of a and a low value of I_0. Constraining both a and σ leads to a more physically plausible value of I_0 (11.2) for a magnitude 7.8 event and suggests that the initial fit was strongly affected by the spatial distribution of intensity observations.

The Inangahua earthquake is associated with the fullest set of felt intensity observations for any New Zealand earthquake. The freely fitted parameters, shown in the first section of Table 2, include a low value of a (1.63), a small σ, and a high value of $I_0 = 10.7$, combining to give a rather sharp peak around the epicentre and a slower decay at longer distances. Constraining a to successively higher values, as in the third and second parts of Table 2, results in a steady increase in the residual sum of squares, and causes σ to increase to around 9 or 10 km, and I_0 to decrease to around 10.1. The impression here is the reverse to that for the Buller earthquake: high intensities around the centre are well-represented, even exaggerated, resulting in what may be an unduly high value of I_0 in the model without constraints. However, the preference for lower a and σ values is consistent throughout the table, and is unlikely to be due entirely to accidental features of the pattern of observations. The values are lower than suggested for some overseas attenuation models (e.g. Donovan(1973)) and may be characteristic of New Zealand earthquakes.

Further insight can be gained from combining the results from all three earthquakes, using a common value of a, but allowing the other parameters to be chosen freely, and separately, for each earthquake. Figure 13 shows a plot of the total sum of squares, from all three events, against the common value of a. The minimum occurs around $a = 1.9$. Details of the other parameters for this value of a are shown in the third section of Table 2. This compromise value gives a reasonable fit for all three data sets, and in each case is associated with physically plausible estimates.

These analyses suggest that for New Zealand earthquakes the geometrical spreading factor a is indeed somewhat below 2, but that for individual events the general pattern can be somewhat obscured by accidental features such as the location or absence of population centres close to the centre of shaking.

We next examine the question of ellipticity. All three Figures show that the preferred contours are elliptical in character with the major axis oriented approximately NE/SW. The degree of ellipticity is measured by the eccentricity e, which is very significantly different from zero for the Inangahua earthquake, marginally significant for the Buller earthquake, and not significant for the Edgecumbe earthquake. This impression was confirmed by repeating the analyses with e and ϵ both set to zero, to give circular contours. The decrease in the residual sum of squares is very highly significant for Inangahua, marginally significant if at all for the other two. Overall, however, the generally elliptical pattern seems well established and the expected similar patterns between the Buller and Inangahua earthquakes is observed.

Finally, we use the non-parametric procedures ("Loess" with degree = 1 and span = 0.3) to model the residuals from fitting the ellipse model with $a = 1.9$. The resulting contours are shown in the RH parts of Figures 10 - 12. The Buller model fits well around the epicentre but underestimates the observed intensities in the Taranaki region and overestimates them in the far North. The Edgecumbe model appears to underestimate the observed intensities along the coast, to the NW and E of the epicentre. The Inangahua model tends to underestimate the intensities over the North Island and to overestimate them in the South of the South Island. The smoothed discrepancies are not large, but there is some suggestion that events in the North of the South Island are felt relatively more strongly in the North Island than at comparable distances in the South Island. Of course the ellipticity built into the model implies that the shaking falls off more rapidly along the minor axis of the ellipse than along its major axis.

A different type of comparison is shown in Figure 14 - 16, where the "elliptical distances" R_i are plotted against observed intensities. Here the model values (crosses) can be directly compared with those obtained by non-parametric smoothing. The model case and the smoothed data case fit closely except at the two ends, where the model gives a more plausible fall-off for large distances, while both cases are subject to some uncertainty at small distances. The substantial agreement along the main part of the curve is reassuring. Note also how the use of elliptical distances has compressed the spread of the data, particularly for the Inangahua earthquake.

5 Parametric Modelling of Felt Intensity Probabilities

Since the felt intensities take only integer values, a better approach in principle is to model them through a parameterised family of discrete probability distributions rather than as a mean plus (normal) error. Moreover, as we shall see later, this more complete approach includes the possibility of allowing in the modelling both for under-reporting of low intensities and the variable population densities.

In this section we shall examine models for the intensity distribution of the multinomial logit form:

$$p_k(\underset{\sim}{x}) = D^{-1} \exp\{\psi_k(\underset{\sim}{x})\} \quad k = 2, 3, \ldots, 11 \tag{3}$$

where $\underset{\sim}{x}$ is the location of an observation point in the region under study; $p_k(\underset{\sim}{x})$ is the probability that a felt intensity (MM) of value k is reported at $\underset{\sim}{x}$, given that a report is made from $\underset{\sim}{x}$;

$$D = \sum_j \exp\{\psi_j(\underset{\sim}{x})\}$$

is a normalising constant; and the functions $\psi_k(\underset{\sim}{x})$ depend on the expected intensity of shaking at $\underset{\sim}{x}$, $I(\underset{\sim}{x})$ through the formula

$$\psi_k(\underset{\sim}{x}) = -b(k - I(\underset{\sim}{x}))^2 - C_k. \quad (k = 1, 2, \ldots, 11) \tag{4}$$

Note the restriction to integer values of k in the range [2, 11]. Here b is a parameter to be determined, $I(\underset{\sim}{x})$ varies with the angle θ and radial distance r of $\underset{\sim}{x}$ from the centre of shaking through the ellipse formula (2), and the constants C_k can be treated as given constants or further parameters to be determined.

Depending on the value of b, the distribution $\{p_k(\underset{\sim}{x})\}$ is more or less tightly distributed around the value $I(\underset{\sim}{x})$. If the constants C_i are all set to zero, the distribution is symmetric about

$I(\underset{\sim}{x})$ whenever $I(\underset{\sim}{x})$ itself takes an integer value. Otherwise the coefficients C_i are intended to model the extent of bias due to under-reporting of low intensities. Figure 17 shows two typical forms for the distribution when $I(\underset{\sim}{x}) = 5$, the first when $\underset{\sim}{C} = \{C_i\} = \underset{\sim}{0}$, and the second when $\underset{\sim}{C} = (5, 3, 1, 1, 0, 0, 0, 0, 0, 0)$, which might represent a more realistic situation in practice.

These model distributions for $I(\underset{\sim}{x})$ can be compared with histograms of the numbers of actual observed intensities falling within two isoseismal contours. More precisely, the distribution for $I(\underset{\sim}{x}) = k$ can be compared with the histogram of observed values of I within the two elliptical contours corresponding to $I(\underset{\sim}{x}) = k \pm \frac{1}{2}$. Histograms for the three New Zealand events are shown in Figures 18 - 20 (with $\underset{\sim}{C} = \underset{\sim}{0}$) and 21 - 23 (with $\underset{\sim}{C} = (4, 3, 2, 1, 0, 0, 0, 0, 0, 0)$).

The model parameters can be fitted by the method of maximum likelihood. The likelihood here takes the form:

$$L = \prod_{n=1}^{N} p_{I_n}(\underset{\sim}{x}_n) \tag{5}$$

where $\underset{\sim}{x}_n$ is the location and I_n the intensity observed at that location. The maximisation was done using the ms ( ) routine in S-PLUS, as described in Bates and Chambers (1991). Here the value of the log-likelihood plays a role analagous to minus the residual sum of squares in the least squares routine ns( ); the improvement in fit due to adding an additional parameter is gauged by the increase in the log likelihood, or "deviance".

The values of the parameters I_0, a, e, ϵ and σ obtained by this approach agree very closely with those obtained by the least squares procedures, at least when the vector of constants $\underset{\sim}{C} = \underset{\sim}{0}$. We shall not report these results in detail, but consider rather the effects of varying the constants within this vector. The isoseismals are shown in Figures 18 - 20 and 21 - 23 respectively. These figures also show the proportions of observed values within the elliptical annuli centred on integer values of $I(\underset{\sim}{x})$, and attenuation plots of log R against expected intensity. In all cases a was fixed at the value 1.9, and b (equation (4)) was set equal to 1.

The two sets of figures show little difference at a visual level. However there are some differences in the parameters. Increasing the bias substantially improves the fit for the Edgecumbe data, but not for the Buller or Inangahua data. However even in these last two cases, incorporation of the bias term improves the physical plausibility of the fit (in terms of I_0 and σ), and produces a closer resemblance between the histograms for the actual observations within an isoseismal band, and those predicted (as in Figure 17) by the model.

Clearly more work is needed to systematically study the effects of this bias term. However, there is a limit to the value of what can be done until it is possible to allow also for the probability that no observation at all will be recorded from a given location. This depends on two factors: location relative to the centre of shaking, with the probability of no observation increasing with distance; and population density, with the probability of no observation increasing as the population density approaches zero. Such an exercise is beyond the scope of the present study, but may be critical to obtaining a consistent picture of attenuation behaviour, as free as possible of the accidental effects of the location of the epicentre relative to centres of population.

6 Relation of observations to isoseismal contours

Implicit in our treatment is the assumption that the contour for intensity 7 (say) links those points where the expected intensity is 7 (see Introduction). This should run somewhere through the middle of the observed values with intensity 7.

A comparison with the observed intensity values reveal some systematic departures from this expectation. Figure 24 shows the proportion of observations of a given intensity lying inside the corresponding fitted isoseismal contour. The data are listed in Tables 4 and 5. It will be seen that this proportion increases from low values (0 - 30%) at the high intensities to high values (80 - 100%) at the low intensities. This is inconsistent with the concept of "isoseismal" however defined and is therefore unsatisfactory. We conclude that the effect is due to the fitted model peaking too sharply in the vicinity of the epicentre, and decaying too slowly at large distances from the epicentre. The latter effect is confounded with the increasing probability of no observation being recorded. Both effects are indicative of short-comings of the model structure, or the fitting procedures. Several points may be made:

1. The present model structure peaks very sharply near $r = 0$, even when the scale parameter σ is relatively large. Minor variations, such as a smooth peak, rather than a cusp, at $r = 0$ might help to overcome this point.

2. There are relatively few high intensity observations, and they may not carry sufficient weight in the regression procedures to force the fitted curve into matching the close obervations well. This could be critical in application to risk assessment, where the higher intensities dominate the contribution to loss.

3. The observations are fitted to the logarithmic distance scale which may produce some further distortions.

4. The fact that the estimated intensities are consistently too high at large distance suggests that the absorption term bR in (1) may have been discarded too hastily. In particular, improved modelling of the intensities at small distances may allow an absorption term to be retained in (1) without creating the problem referred to in §4.

5. The actual observations represent a compounding of the probability of observing a given intensity of shaking with the probability of there being an observer to make the observation. The actual distribution of observations is strongly affected by the relation between locations of population centres and the locations of regions with a given expected level of shaking. Thus the discrepancies cannot be finally resolved until the distribution of populations is taken into account in the model.

7 Conclusions

The main conclusions from this study may be summarised as follows:

1. Both non-parametric and parametric approaches can be used to obtain smoothed or simplified pictures of the isoseismal patterns around a major earthquake. Adequate procedures for both approaches exist within the standard S-PLUS software, although a greater choice of options for 2-dimensional smoothing would be desirable.

2. The most promising combination of these approaches for extracting useful information from the felt intensity data, is first to fit a simple parametric model, with only a small number of parameters, and follow that by non-parametric smoothing of the residuals from that model.

3. The simple elliptical form (2) for the attenuation, with just a single parameter to control geometric spreading, represents a reasonable compromise between simplicity and flexibility

as a baseline parametric model. However, some refinement of the parameterisation could help to improve the stability of the model-fitting.

4. The further step of directly modelling the probability distribution of intensities at a given location, as a function of the expected intensity predicted by the elliptical form (2), allows the observed distribution of intensities within an isoseismal band to be compared with the predicted distribution. It offers the possibility of taking into account the bias due to under-reporting of low intensities, but otherwise produces similar results to (3). More work is needed to realize the potential of this approach.

5. There is a need to clarify the definition of an isoseismal contour. If, as in the present study, it is taken as a contour where the expected intensity has the stated value, then there is evidence that with the present model the fitted isoseismals are too tight for high intensities and too large for low intensities. Refinements of the underlying model and the fitting procedures may be needed to overcome these inconsistencies.

6. The results of applying these methods to the Buller, Inangahua and Edgecumbe earthquakes suggest that even with a simplified model, accidental features of the spatial pattern of observations can affect the parameter values and the appearance of the fitted isoseismals. At the present stage it seems possible that a common form for the isoseismals, with the degree of the geometrical spreading below 2, and a NNE/SSW oriented elliptical shape, would adequately account for most if not all of the main features of the three events analysed. The residual features are likely to combine geological or tectonic factors with the effects of accidental groupings of felt intensity observations around centres of population, or the absence of such observations from areas of sea or wilderness.

Acknowledgement This work forms part of contract with the New Zealand Earthquake Commission. Their support is gratefully acknowledged. The assistance of IGNS staff in making available the New Zealand data, and the assistance and advice at various stages of staff in the VUW Institute of Geophysics and of the VUW Institute of Statistics and Operations Research, is also acknowledged with gratitude.

List of Figures

Figure 1	Felt intensity observations for Buller earthquake
Figure 2	Felt intensity observations for Inangahua earthquake
Figure 3	Felt intensity observations for Edgecumbe earthquake
Figure 4	Non-parametric attenuation curves for the Buller earthquake
Figure 5	Non-parametric attenuation curves for the Inangahua earthquake
Figure 6	Non-parametric attenuation curves for the Edgecumbe earthquake
Figure 7	"Loess" fitted attenuation and Isoseismals for the Buller data
Figure 8	"Loess" fitted attenuation and Isoseismals for the Inangahua data
Figure 9	"Loess" fitted attenuation and Isoseismals for the Edgecumbe data
Figure 10	Least squares fit of the ellipse model to Buller data, with residuals smoothed by "Loess"
Figure 11	Least squares fit of the ellipse model to Inangahua data, with residuals smoothed by "Loess"
Figure 12	Least squares fit of the ellipse model to Edgecumbe data, with residuals smoothed by "Loess"
Figure 13	Total residual sum of squares against attenuation parameter a, for the three combined New Zealand earthquakes
Figure 14	Attenuation versus elliptical distance: non-parametric and model curves for the Buller earthquake
Figure 15	Attenuation versus elliptical distance: non-parametric and model curves for the Inangahua earthquake
Figure 16	Attenuation versus elliptical distance: non-parametric and model curves for the Edgecumbe earthquake
Figure 17	A comparison of model probability distributions with and without the bias component
Figure 18	Histograms of the distribution of observed intensities within epicentral bands, attenuation model and isoseismals, without bias term: Buller earthquake
Figure 19	Histograms of the distribution of observed intensities within epicentral bands, attenuation model and isoseismals, without bias term: Inangahua earthquake
Figure 20	Histograms of the distribution of observed intensities within epicentral bands, attenuation model and isoseismals, without bias term: Edgecumbe earthquake
Figure 21	Histograms of the distribution of observed intensities within epicentral bands, attenuation model and isoseismals, incorporating bias term: Buller earthquake
Figure 22	Histograms of the distribution of observed intensities within epicentral bands, attenuation model and isoseismals, incorporating bias term: Inangahua earthquake
Figure 23	Histograms of the distribution of observed intensities within epicentral bands, attenuation model and isoseismals, incorporating bias term: Edgecumbe earthquake
Figure 24	Proportion of MMI observations within each Isoseismal band, modelled with bias term

References

Bates DM and JM Chambers (1991) *Non-linear Models* (Ch. 10 of Chambers JM and TJ Hastie (ed.) *Statistical Models in S*)

Brillinger DR and Preisler HK (1985) *Further Analysis of the Joyner-Boore Attenuation Data*, Bull. Seismol. Soc. America **75** 611-614.

Cleveland WS, E Grosse and WM Shyn (1991) *Local Regression Models* (Ch. 8 of Chambers JM and TJ Hastie (ed.) *Statistical Models in S*)

Donovan NC (1973) *Earthquake hazards for buildings* (in *Building Practices for Disaster Mitigation*, US Dept. of Commerce Report NBS BSS 46.

Downes G (1995) *An isoseismal atlas of New Zealand earthquakes* Bulletin Inst. Geolog. and Nuc. Sciences (in preparation)

Dowrick DJ (1991) *A revision of attenuation relationships for modified Mercalli intensity in New Zealand* Bull. NZ Nat. Soc. Earthq. Engy. **24** 210 - 224.

Dowrick DJ (1992) *Attenuation of Modified Mercalli intensity in New Zealand earthquakes* Engy. Struct. Dyn. **21** 181 - 196.

Härdle W (1990) *Smoothing Techniques with Implementation in S* Berlin (Springer)

Joyner WB and DM Boore (1981) *Peak horizontal acceleration and velocities from strong motion records including records from the 1979 Imperial Valley, California, Earthquake* Bull. Seismol. Soc. A. **71** 2011 - 2038

Kozuch MJ (1995) *Prediction of earthquake isoseismal patterns* PhD Thesis, Univ. of Colorado, Boulder.

Ortolani F and Gianmini R (1987) *Elaborazione delle Leggi di alternaziona per gli eventi sismici risentiti in Roma.* Rapporto N. 2 of *Dati Storici e Modello Probabilistico della sismicita in Roma* by Gianmini R, A Giuffu, F Ortolani, PE Pinto and D Veneziano.

Richter CF (1958) *Elementary Seismology* WH Freeman, San Francisco.

Smith WD (1978a) *Spatial distribution of felt intensities for New Zealand earthquakes* NZ J. Geol. Geophy. **21** 293 - 311

Smith WD (1978b) *Earthquake risk in New Zealand: statistical estimates.* NZ J. Geol. Geophys. **21** 313 - 327.

Table 1. Number of Felt Intensity Observations for Each Collected Data Set

Earthquake	Date	Epicentre	M	Depth	No.
Edgecumbe	2 Mar., 1987	(-37.92,176.76)	6.6	10	170
Inangahua	23 May, 1968	(-41.76,171.96)	7	15	245
Buller	16 June, 1929	(-41.7,172.2)	7.8	20	119
Diexi, China	25 Aug., 1933	(32,103.7)	7.5		136
Baranello, Italy	26 July ,1805	(41.53,14.51)			130
Campania, Italy	5 June, 1688	(41.2,14.75)			154
Irpinia, Italy	29 Nov., 1732	(41.13,15.1)			121
Irpinia, Italy	23 July, 1930	(41.07,15.35)			289
Irpinia, Italy	21 Aug., 1962	(41.23,14.93)			196
Napoletano, Italy	5 Dec., 1456	(41.52,14.52)			183
Vulture, Italy	14 Aug., 1851	(41.00,15.67)			51
Irpinia-Lucani, Italy	23 Nov., 1980	(40.8,15.37)			1129

Table 2. Summary of Trials with Different Parameter Combination

Parameters	Name	Estimated parameters a	e	ε	Io	σ	Residual sum of squares	Residual standard error
Parameters all free (a)	Buller	3.61	0.75	0.74	9.01	115.16	56.1	0.71 (114)*
	Inangahua	1.63	0.86	0.98	10.72	3.93	163	0.82 (240)
	Edgecumbe	1.35	0.71	1.25	9.33	2.88	201	1.10 (165)
a=2, σ=10 (b)	Buller		0.78	0.70	11.20		66.1	0.76 (116)
	Inangahua		0.88	0.97	10.13		172	0.84 (242)
	Edgecumbe		0.58	1.56	8.88		202	1.10 (167)
a=1.9 (c)	Buller		0.67	0.57	10.14	17.97	64	0.75 (115)
	Inangahua		0.87	0.98	10.12	8.92	168	0.83 (241)
	Edgecumbe		0.59	1.53	8.95	8.56	202	1.10 (166)

* Degree of freedom

Table 3. Correlations of Estimated Parameters

	Name	Correlation of parameter estimates	a	e	ε	Io
Parameters all free (a)	Buller	e	0.19			
		ε	0.36	0.18		
		Io	-0.74	-0.03	-0.19	
		σ	0.96	-0.02	0.28	-0.86
	Inangahua	e	0.17			
		ε	-0.09	0.02		
		Io	-0.48	-0.004	-0.03	
		σ	0.76	-0.06	-0.004	-0.92
	Edgecumbe	e	-0.39			
		ε	0.11	0.23		
		Io	-0.63	-0.05	-0.22	
		σ	0.93	-0.27	0.204	-0.86

Table 4. Percentage of Observations in Each Isoseismal Band

	% of total obervations	Isoseismal band	I=2	I=3	I=4	I=5	I=6	I=7	I=8	I=9	I=10	I=11
				%	of	obs.	within	band				
	0.101	[3,4)	0.167	0.25	0.5	0.083						
	0.286	[4,5)	0.029	0.265	0.176	0.471	0.059					
	0.294	[5,6)			0.057	0.457	0.457	0.029				
Buller	0.118	[6,7)					0.429	0.571				
	0.134	[7,8)					0.063	0.188	0.562	0.188		
	0.042	[8,9)							0.8	0..2		
	0.025	[9,10)							0.333	0.667		
	0.016	[2, 3)			0.5	0.5						
	0.139	[3, 4)		0.206	0.618	0.176						
	0.314	[4, 5)		0.078	0.429	0.468	0.026					
Inangahua	0.151	[5, 6)			0.19	0.568	0.216	0.027				
	0.139	[6, 7)			0.029	0.206	0.176	0.471	0.118			
	0.18	[7, 8)				0.045	0.068	0.318	0.432	0.113	0.023	
	0.037	[8, 9)					0.111	0.111	0.333	0.444		
	0.020	[9, 10)								0.2	0.8	
	0.004	[10,11)										1.0
	0.018	[4,5)					0.333	0.667				
	0.247	[5,6)			0.167	0.381	0.262	0.191				
	0.241	[6,7)				0.146	0.390	0.244	0.195	0.024		
Edgecumbe	0.329	[7,8)				0.143	0.089	0.304	0.214	0.25		
	0.165	[8,9)					0.036	0.179	0.107	0.607	0.071	

Table 5. Observations and Percentage of Each Isoseismal Band

		[2,3)	[3,4)	[4,5)	[5,6)	[6,7)	[7,8)	[8,9)	[9,10)	[10,11)
				Isoseismal	band					
	1.		12	34	35	14	16	5	3	
Buller	2.		0.101	0.286	0.294	0.118	0.134	0.042	0.025	
	1.	4	34	77	37	34	44	9	5	1
Inangahua	2.	0.0	0.538	0.516	0.284	0.3	0.438	0.115	0.1	0.0
	1.			3	42	41	56	28		
Edgecumbe	2.			0.0	0.533	0.471	0.405	0.13		

1. Total number of observations with value of I in the isoseismal band
2. Percentage of the observations of value I lying within the isoseismal band centred on the same value of I.

Fig. 1 Felt intensity observations for Buller earthquake

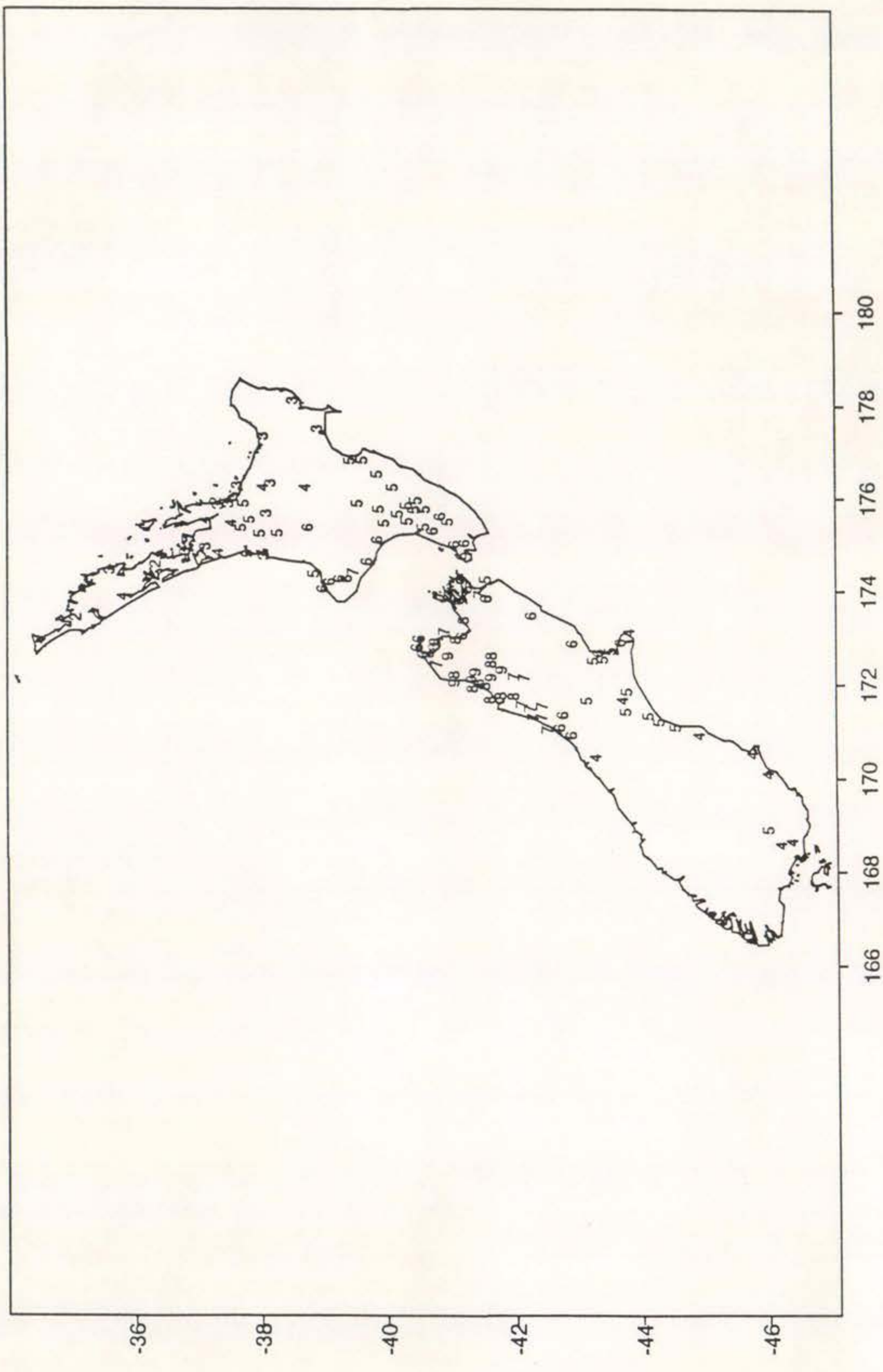


Fig. 2 Felt intensity observations for Inangahua earthquake

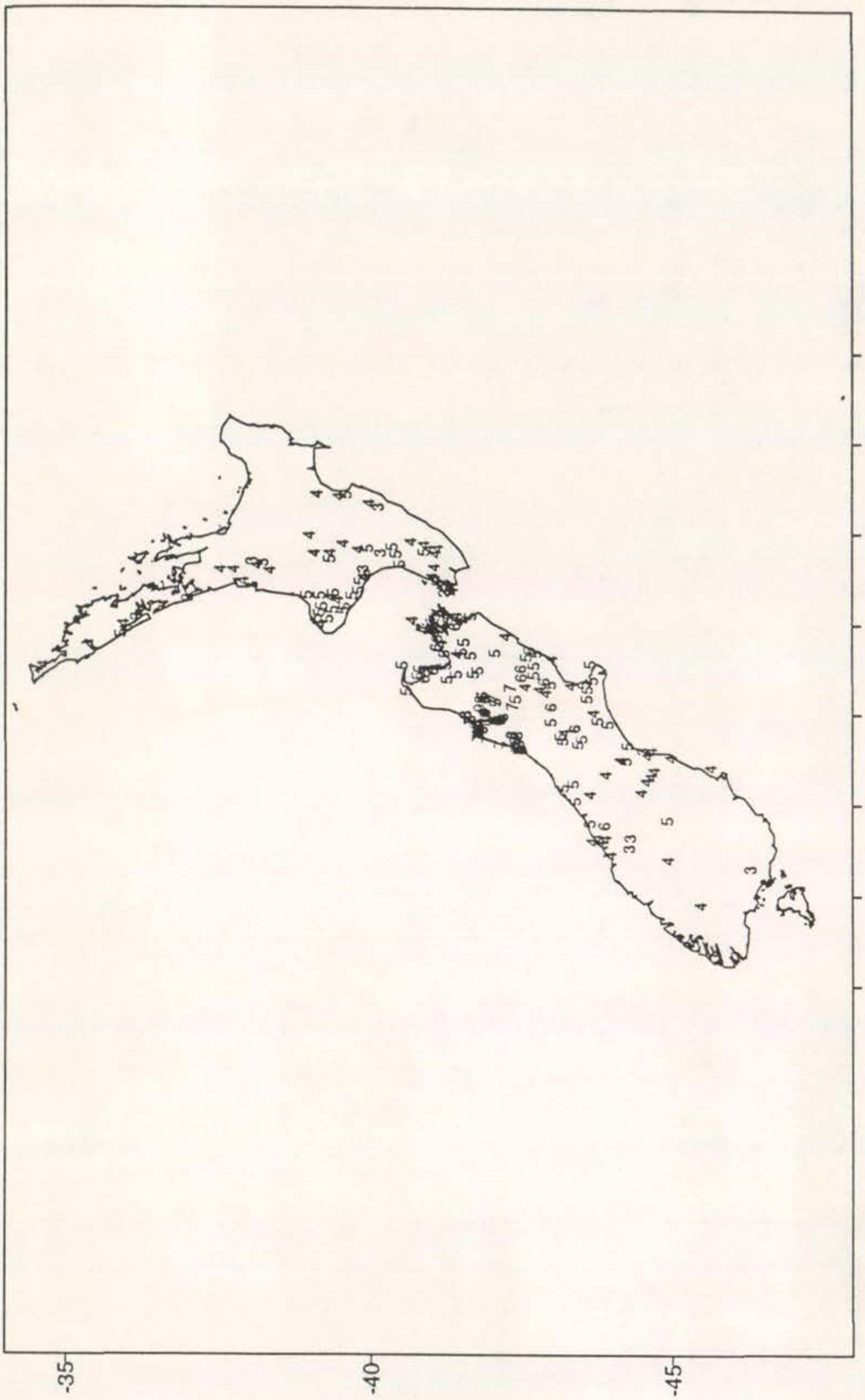
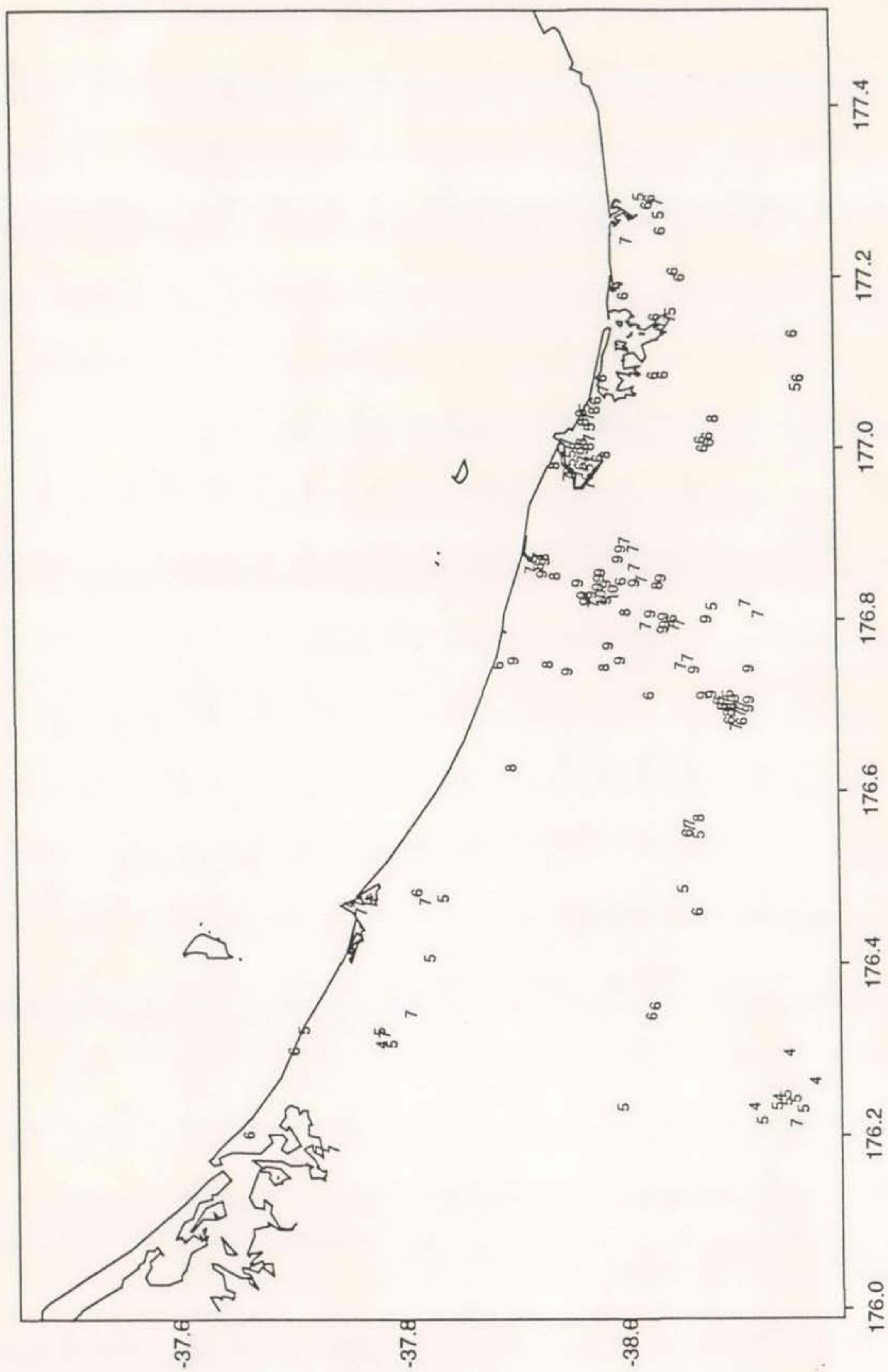


Fig. 3 Felt intensity observations for Edgecumbe earthquake



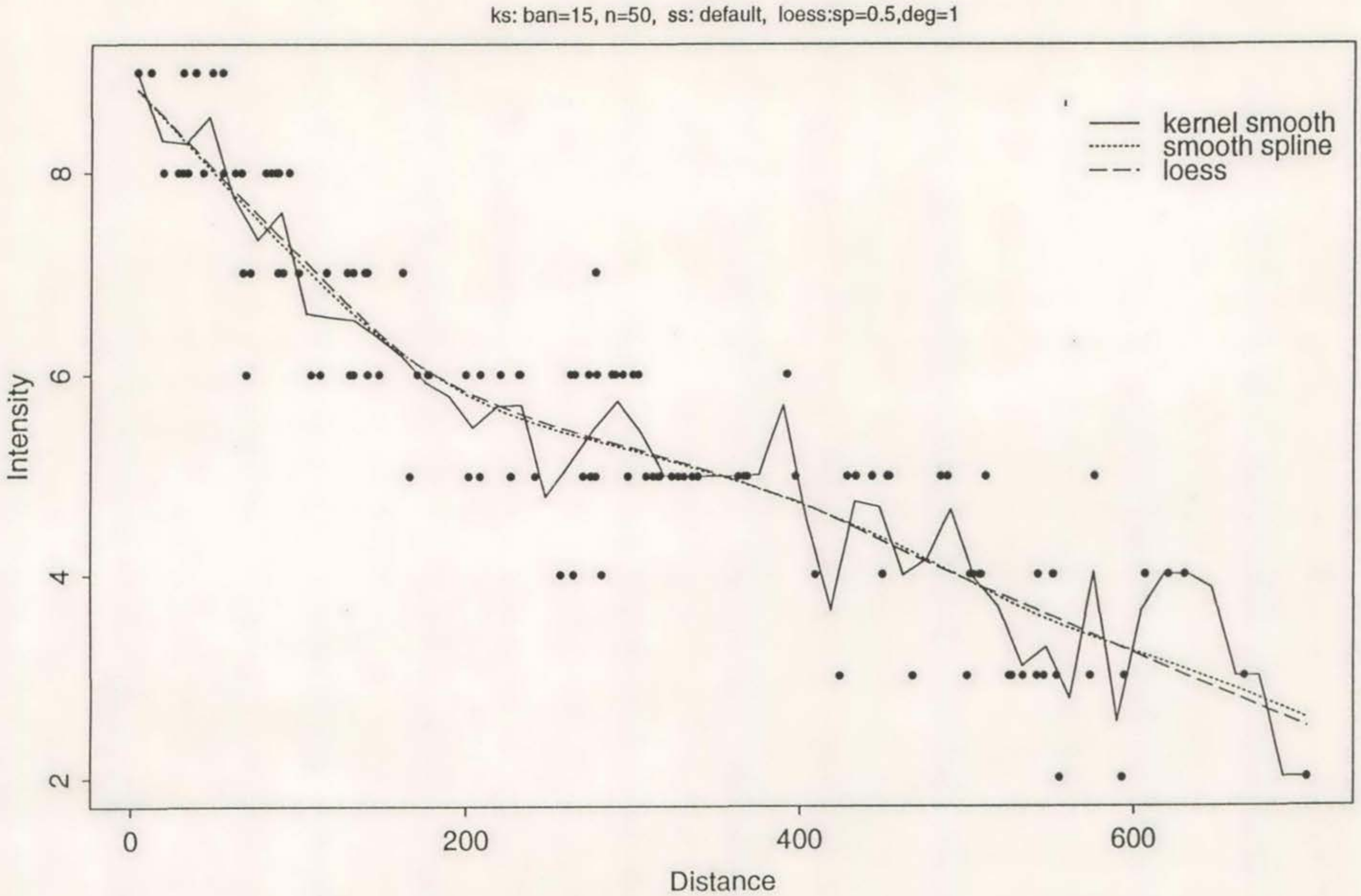


Fig. 4 Non-parametric attenuation curves for the Buller earthquake

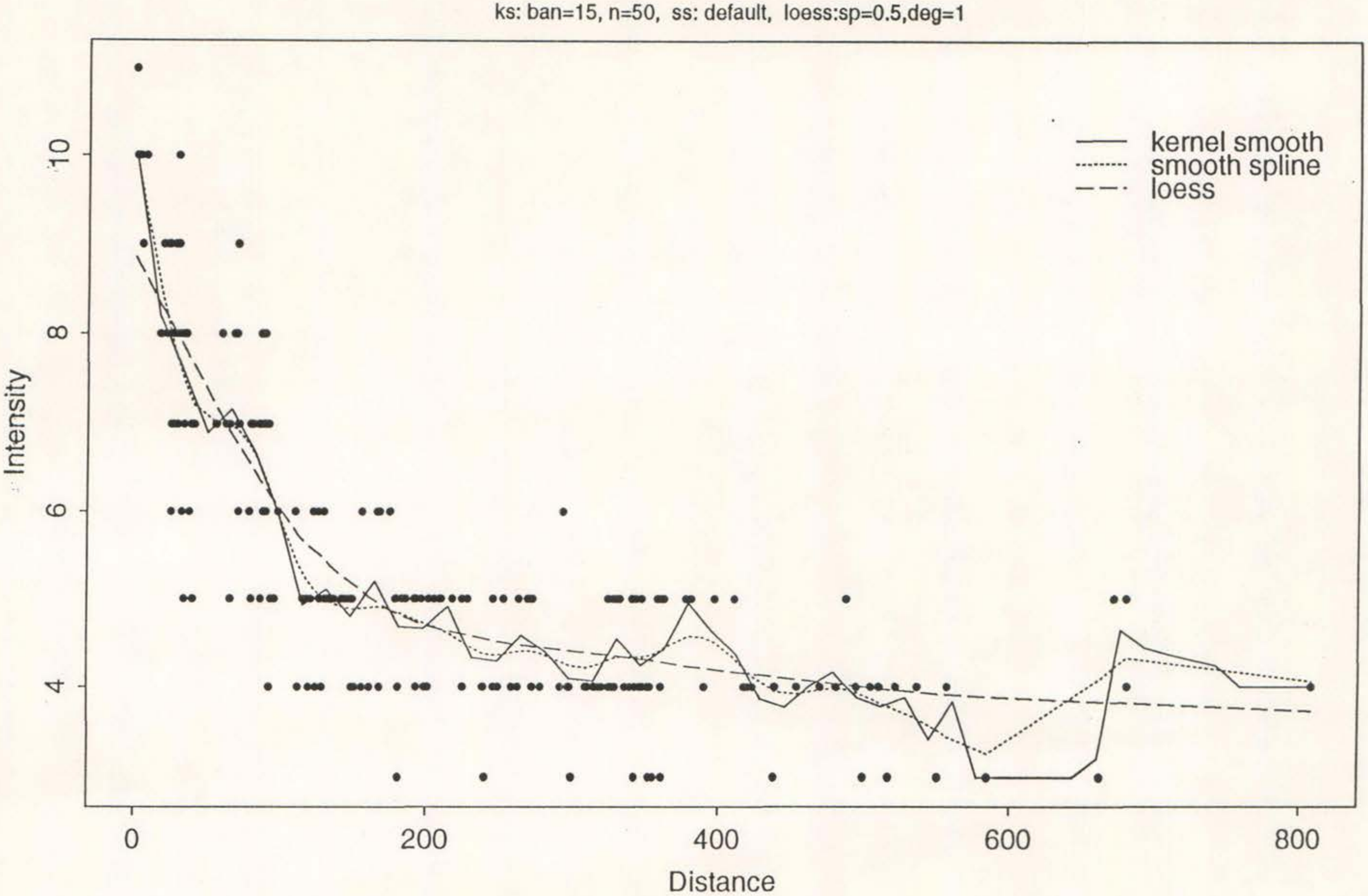


Fig. 5 Non-parametric attenuation curves for the Inangahua earthquake

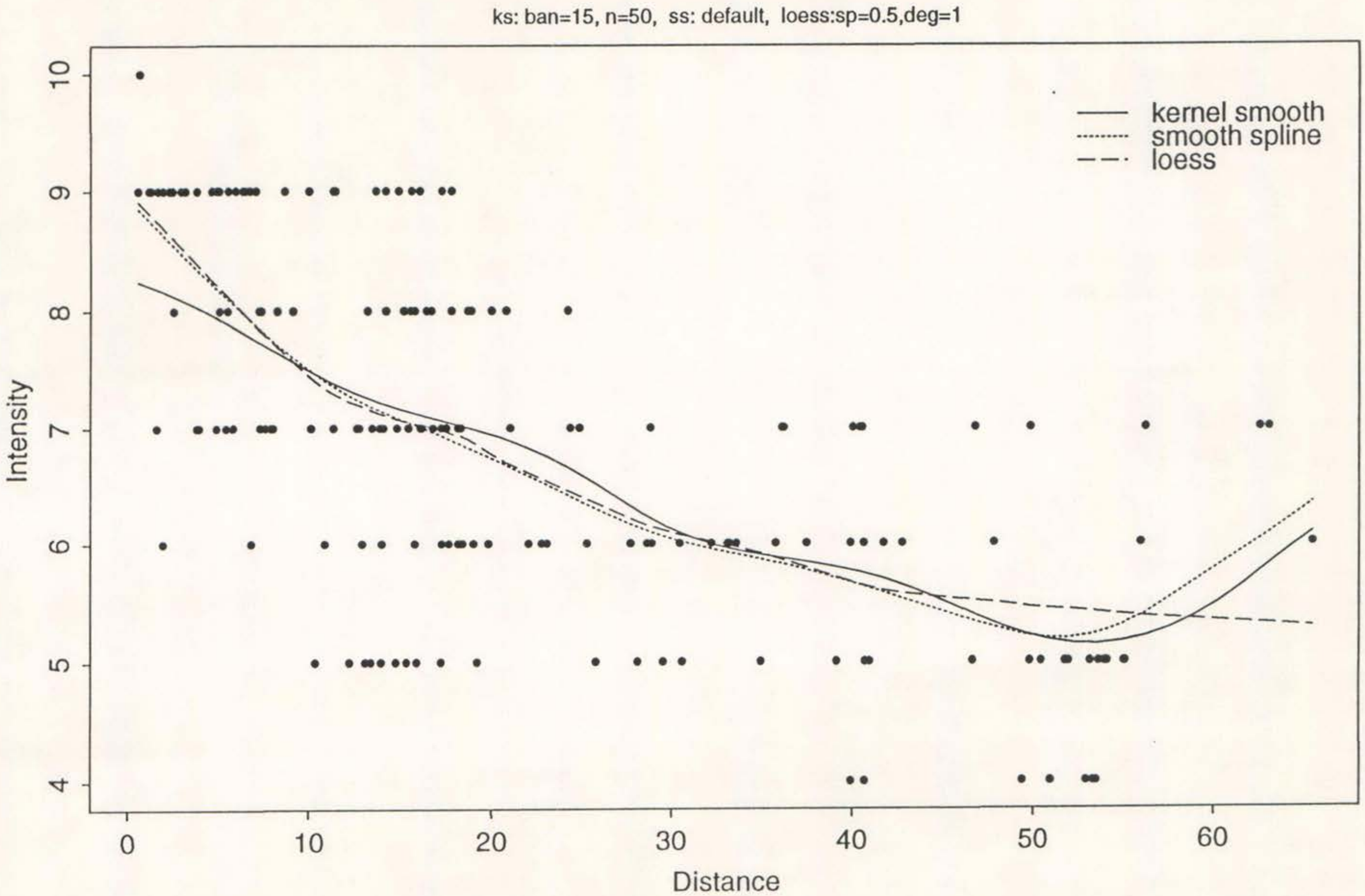
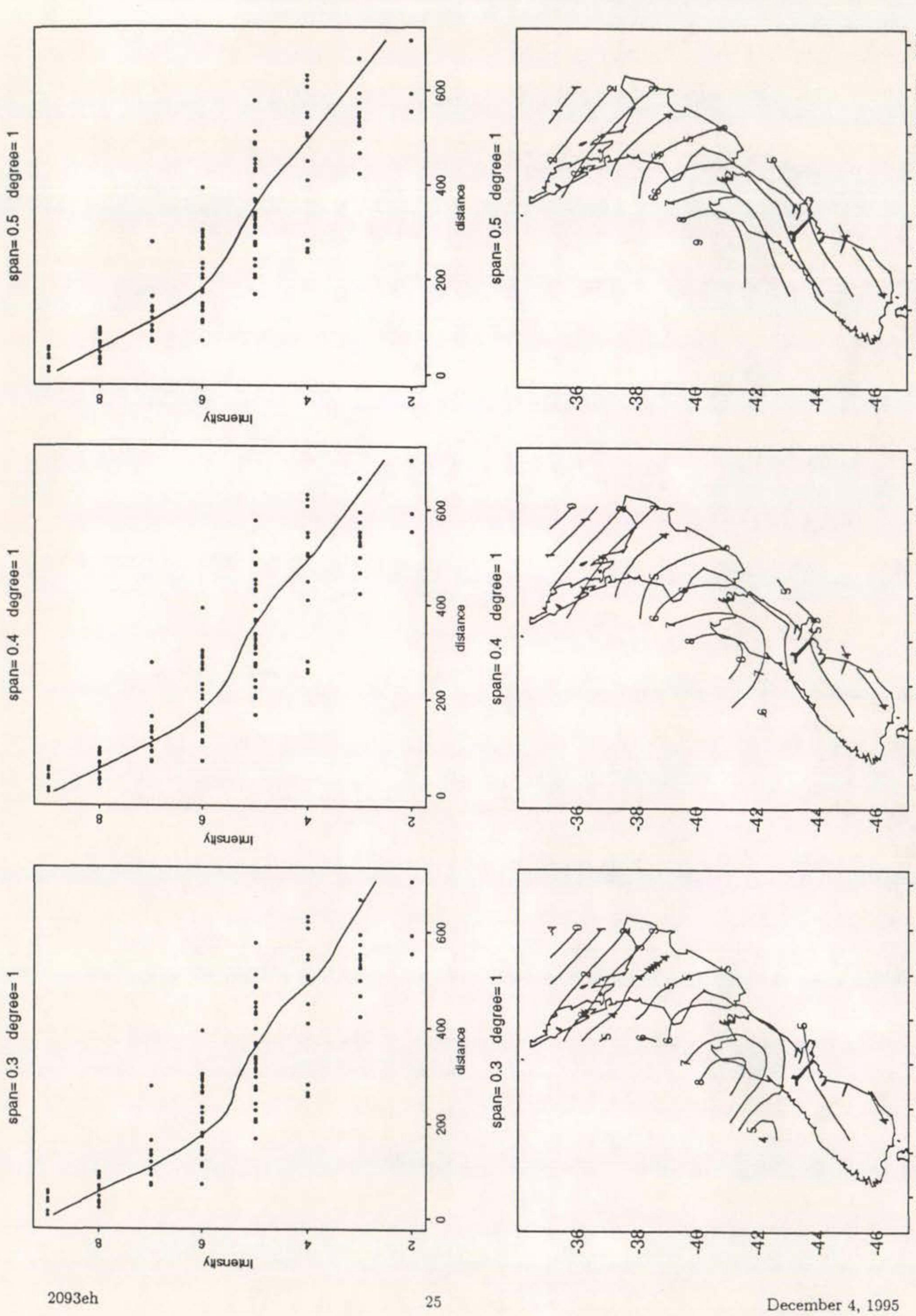


Fig. 6 Non-parametric attenuation curves for the Edgecumbe earthquake

Fig. 7 "Loess" fitted attenuation and Isoseismals for the Buller data



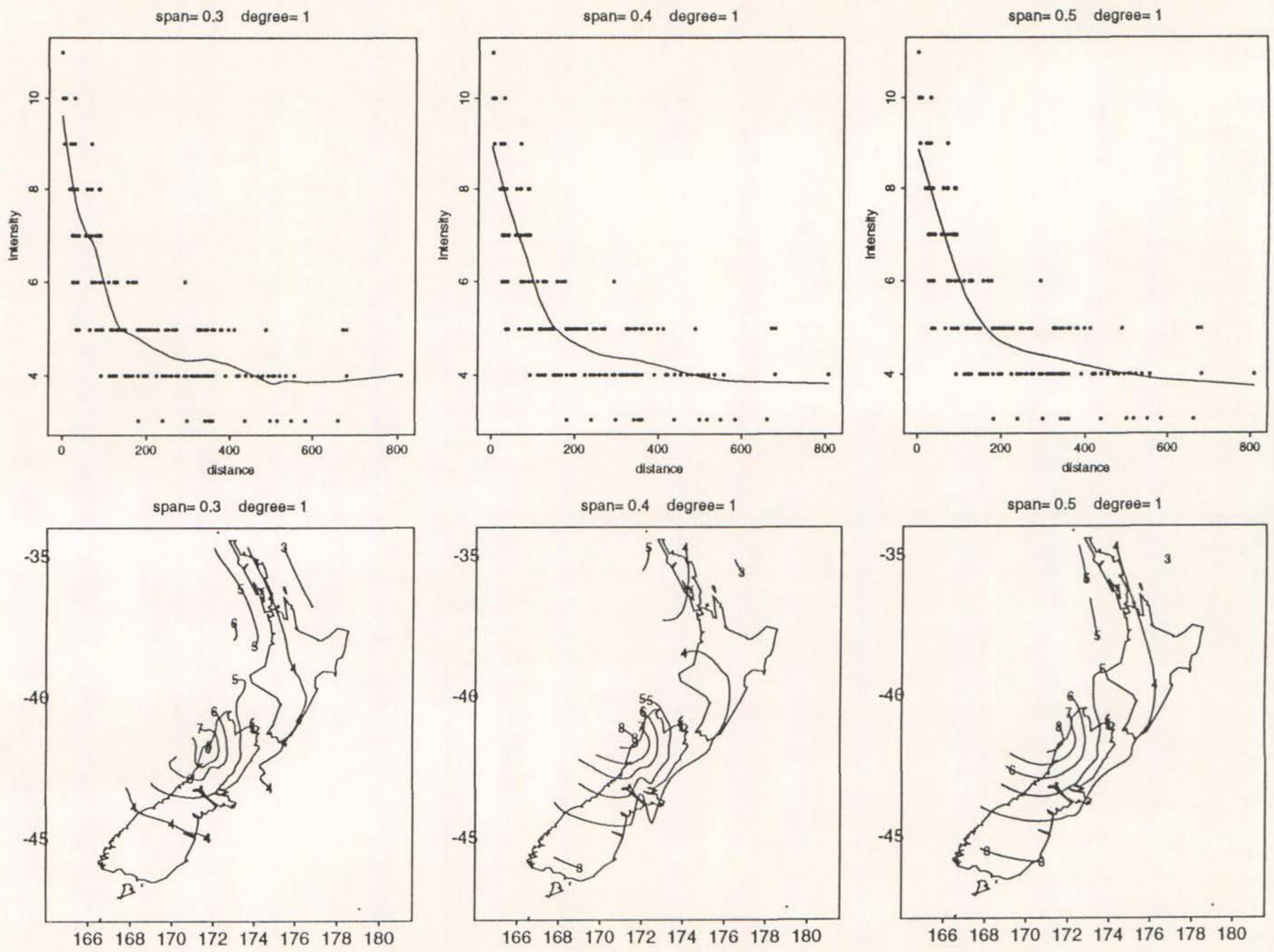


Fig. 8 "Loess" fitted attenuation and Isoseismals for the Inangahua data

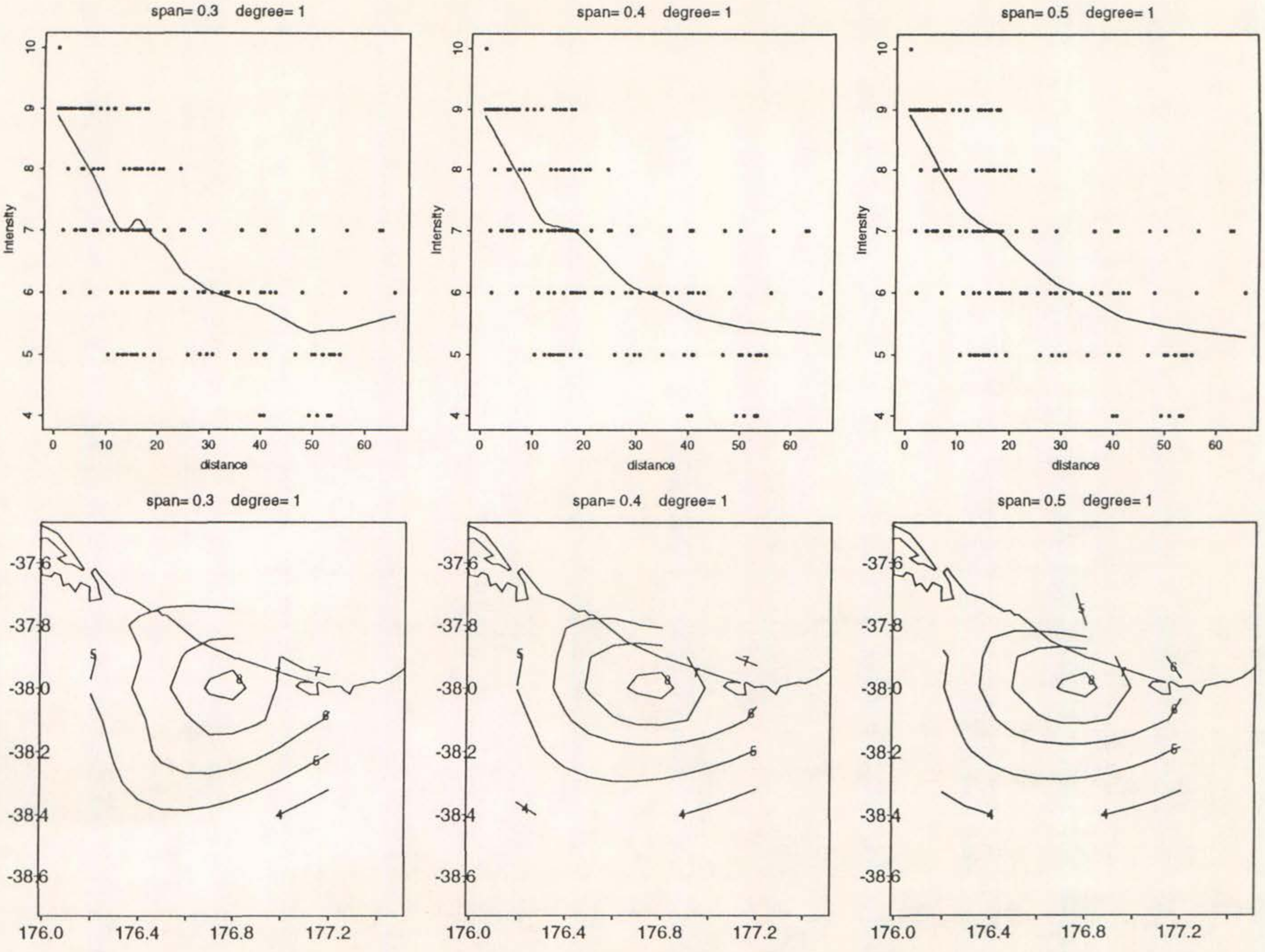


Fig. 9 "Loess" fitted attenuation and Isoseismals for the Edgecumbe data

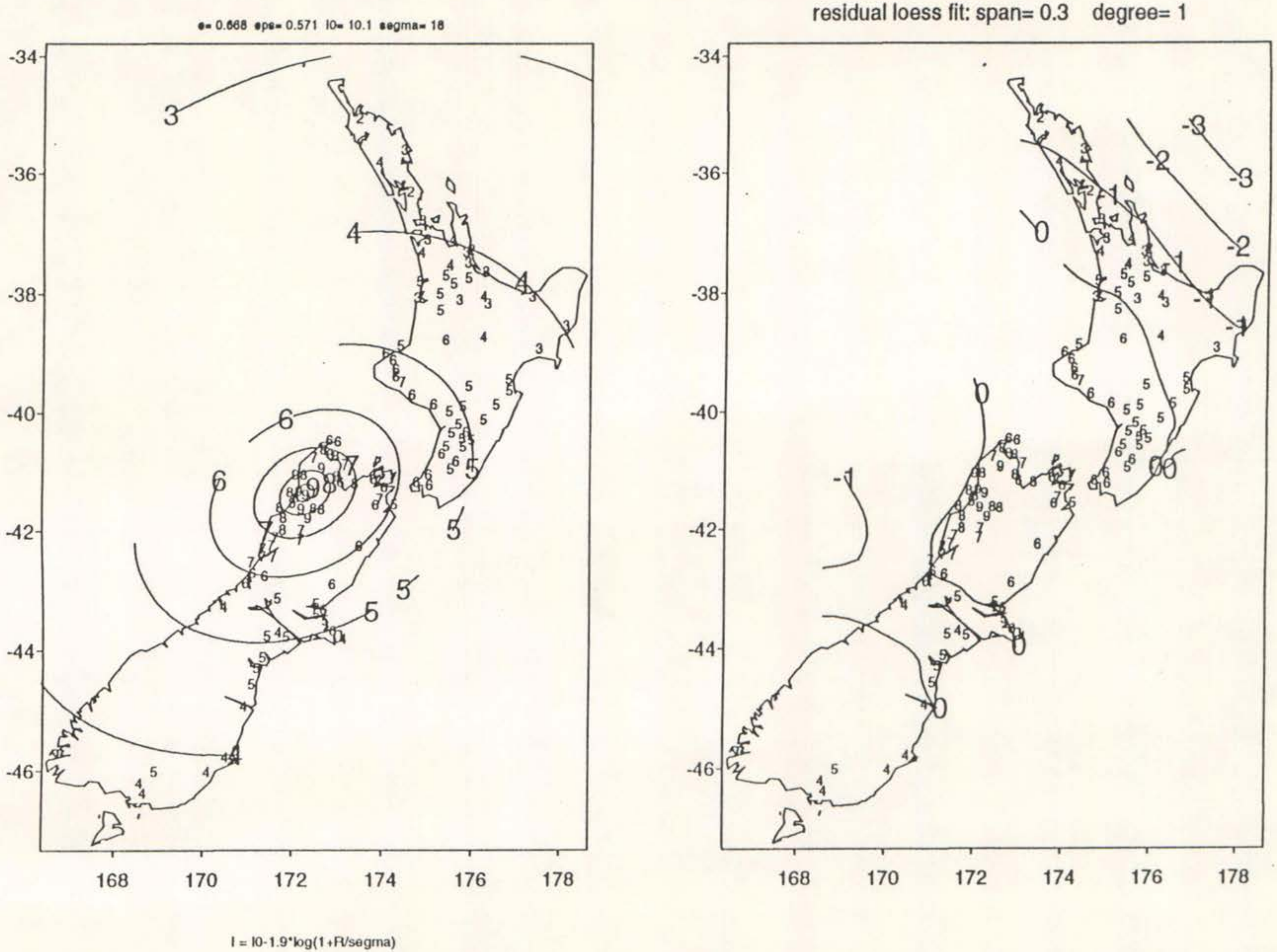


Fig. 10 Least squares fit of the ellipse model to Buller data, with residuals smoothed by "Loess"

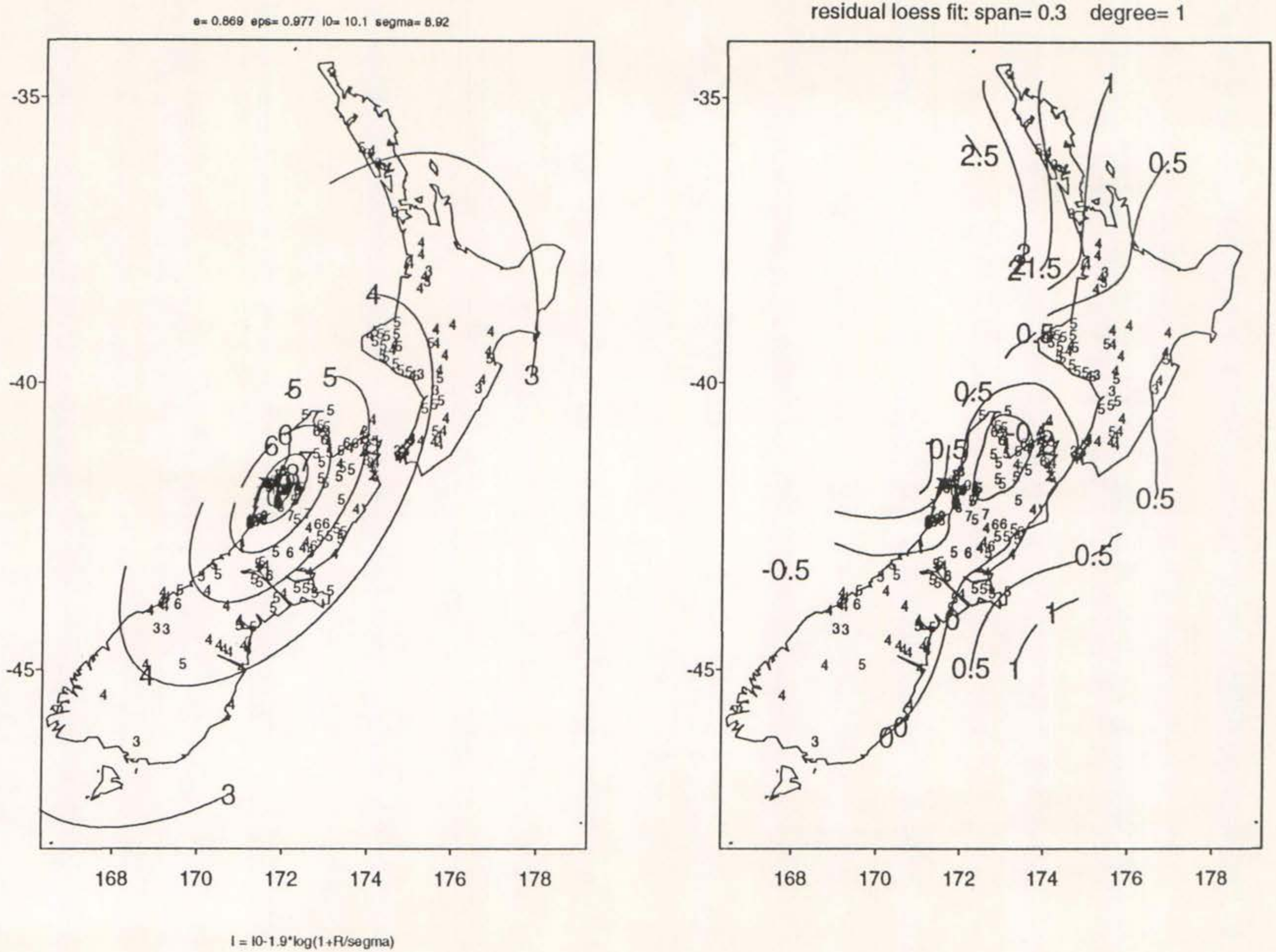


Fig. 11 Least squares fit of the ellipse model to Inangahua data, with residuals smoothed by "Loess"

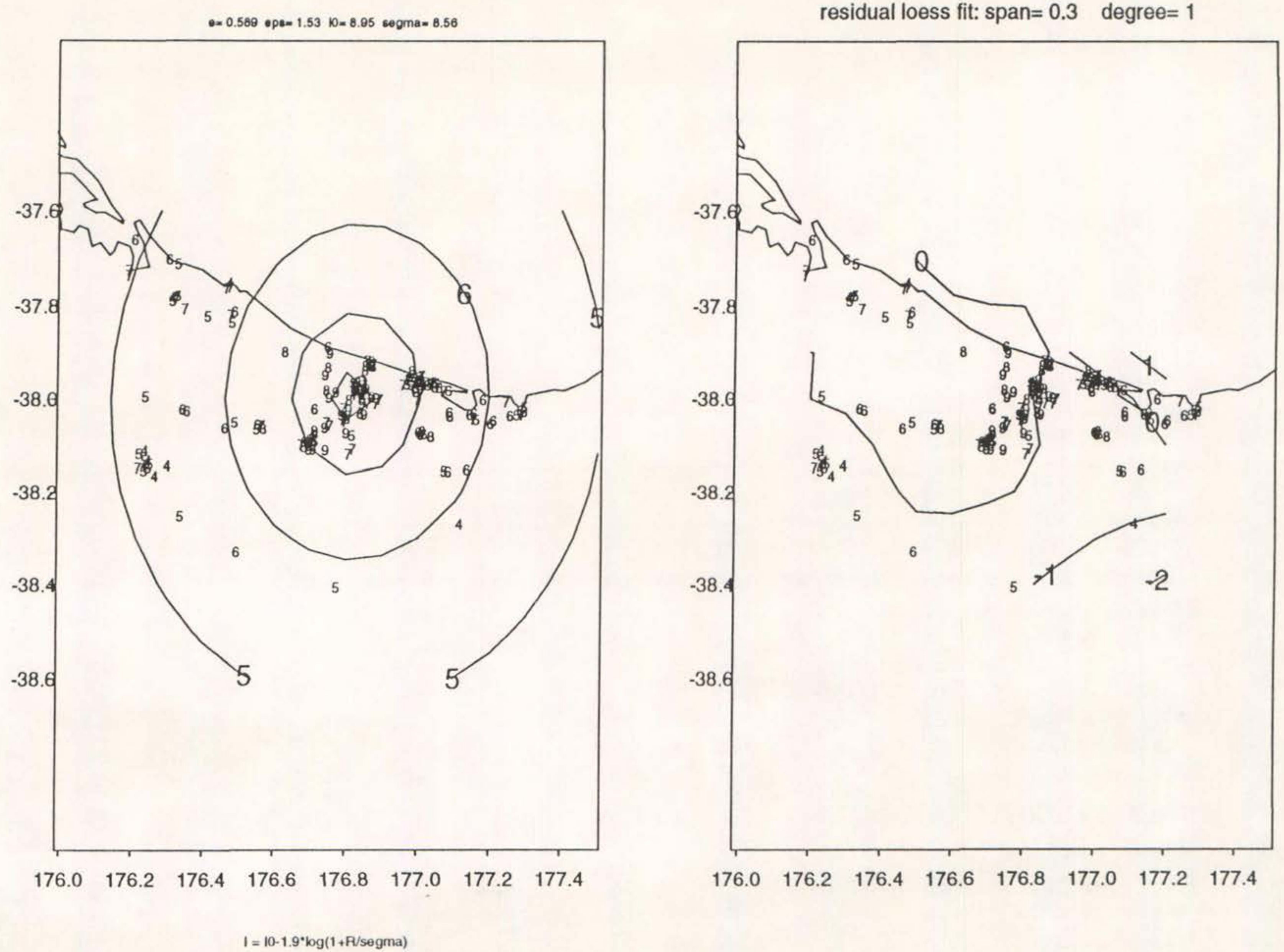
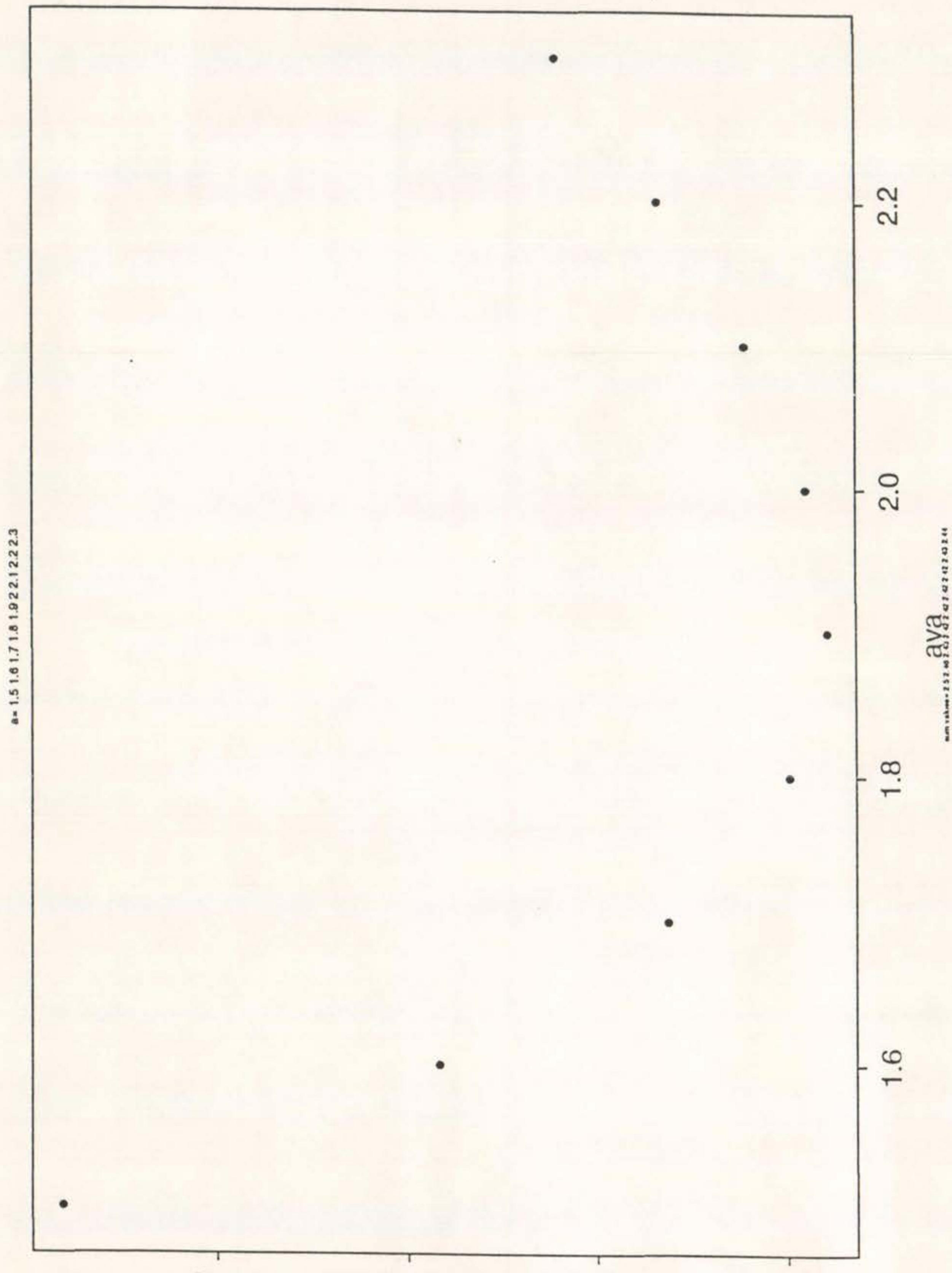


Fig. 12 Least squares fit of the ellipse model to Edgecumbe data, with residuals smoothed by "Loess"

Fig. 13 Total residual sum of squares against attenuation parameter a, for the three combined New Zealand earthquakes



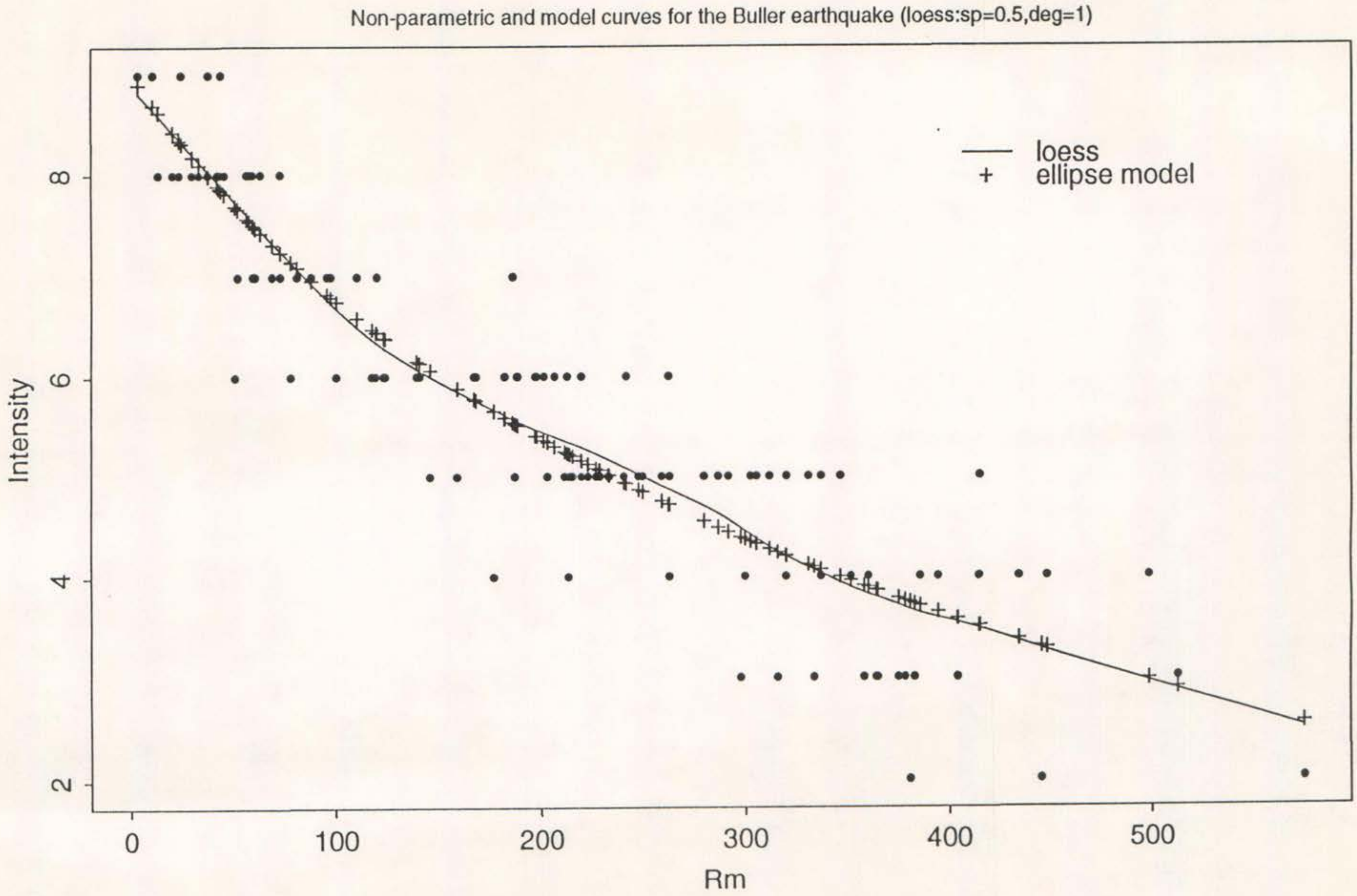


Fig. 14 Attenuation versus elliptical distance: non-parametric and model curves for the Buller earthquake

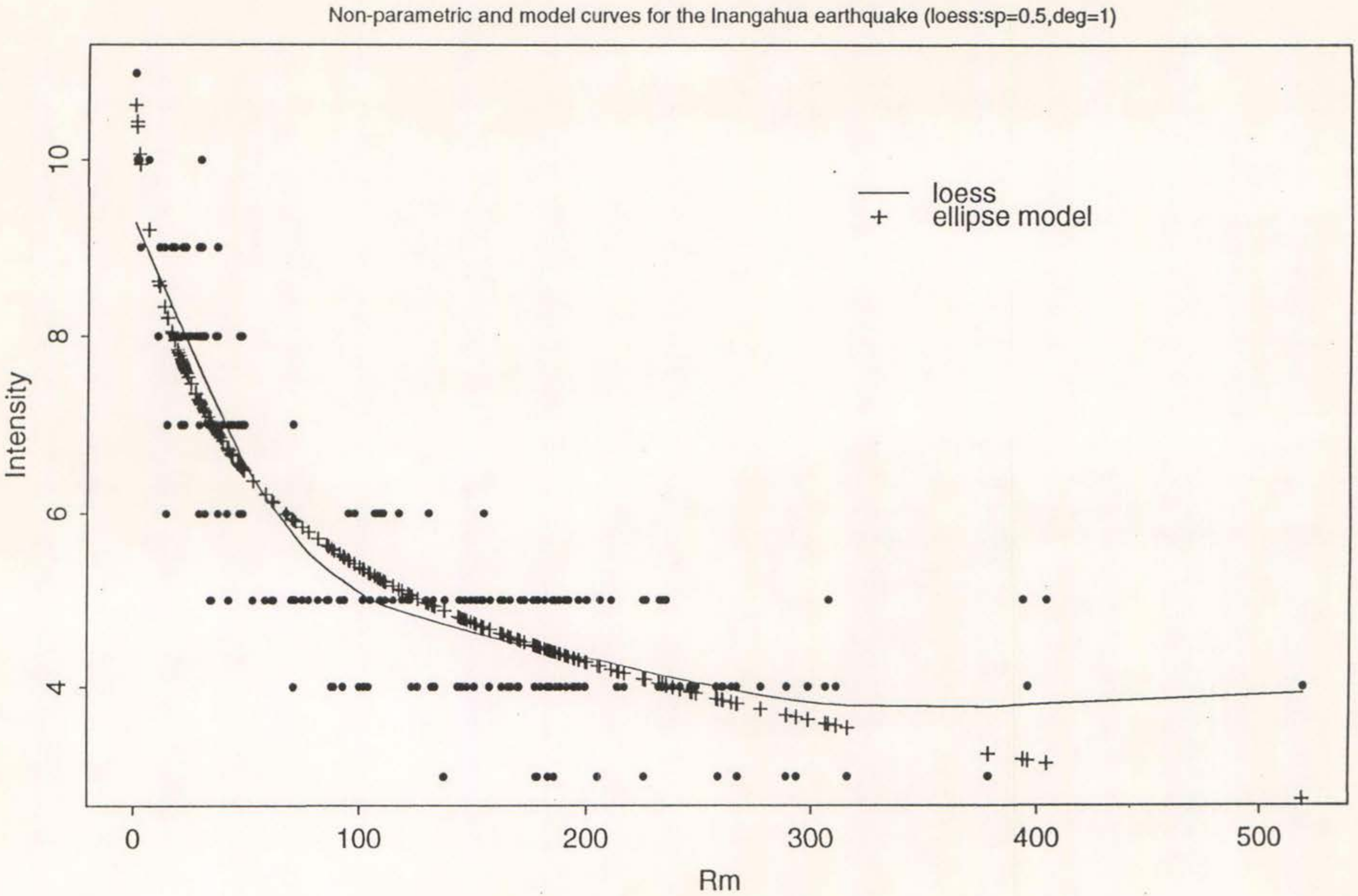


Fig. 15 Attenuation versus elliptical distance: non-parametric and model curves for the Inangahua earthquake

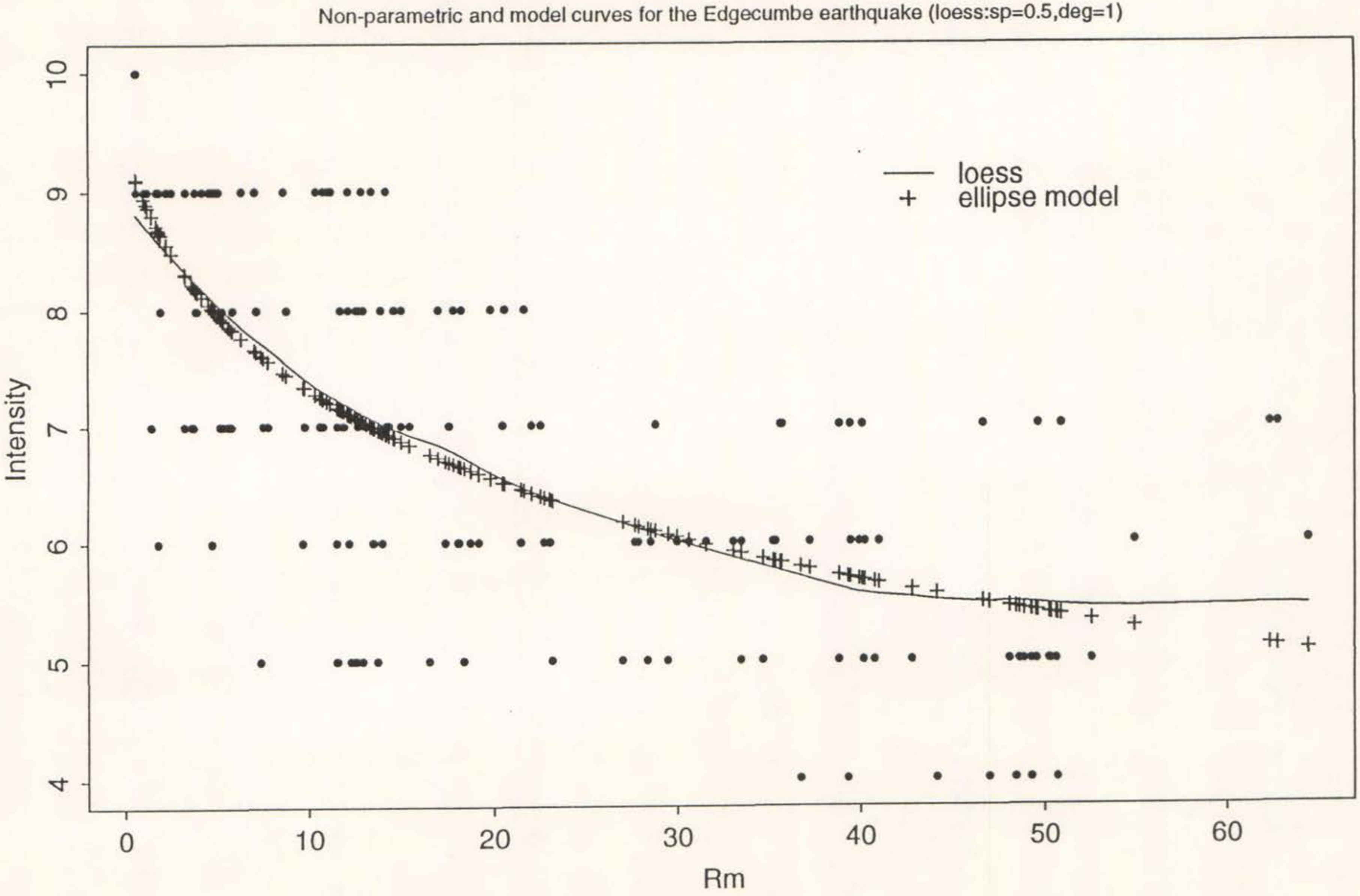


Fig. 16 Attenuation versus elliptical distance: non-parametric and model curves for the Edgecumbe earthquake

Fig. 17 A comparison of model probability distributions with and without the bias component

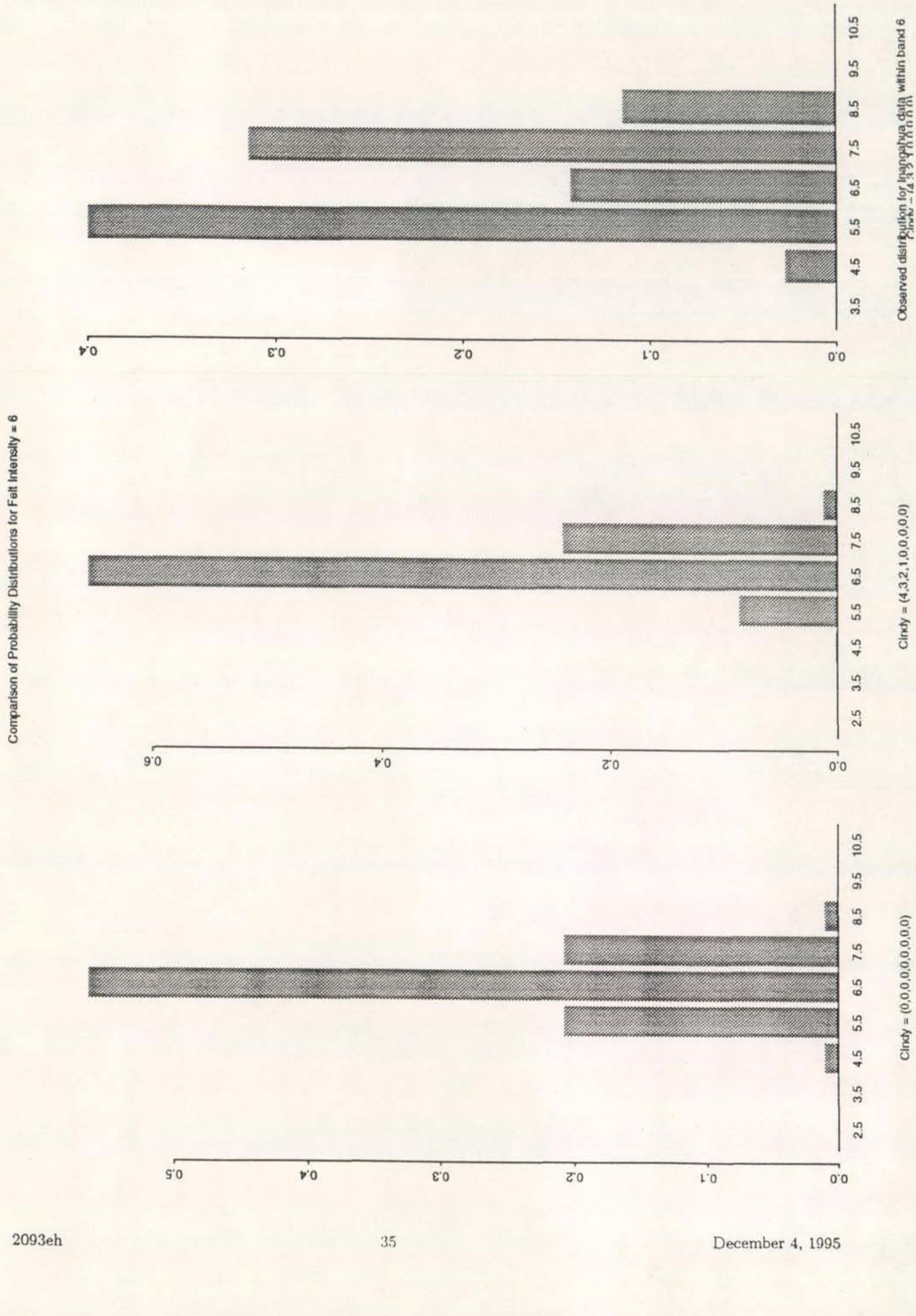


Fig. 18 Histograms of the distribution of observed intensities within epicentral bands, attenuation model and isoseismals, without bias term: Buller earthquake

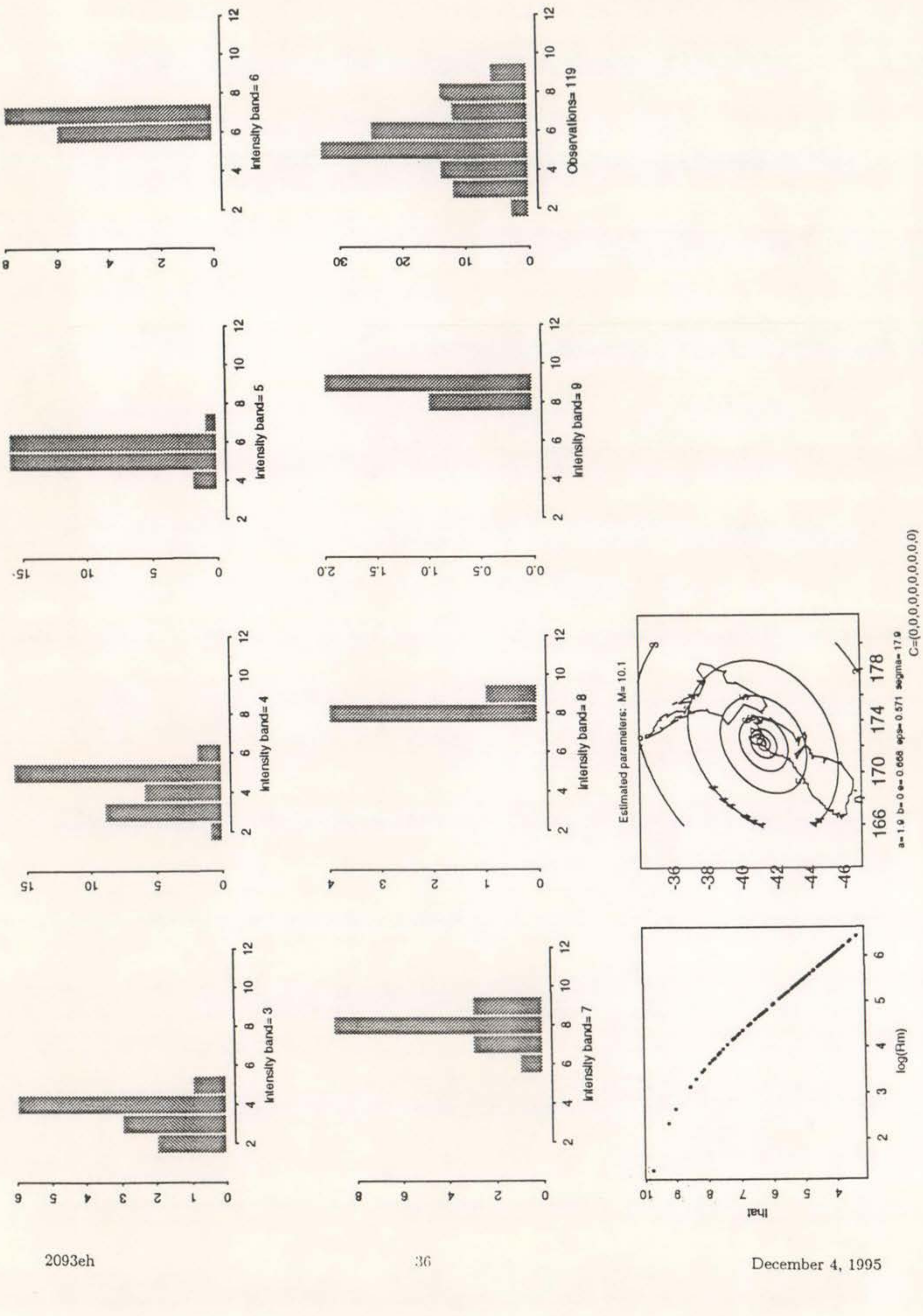


Fig. 19 Histograms of the distribution of observed intensities within epicentral bands, attenuation model and isoseismals, without bias term: Inangahua earthquake

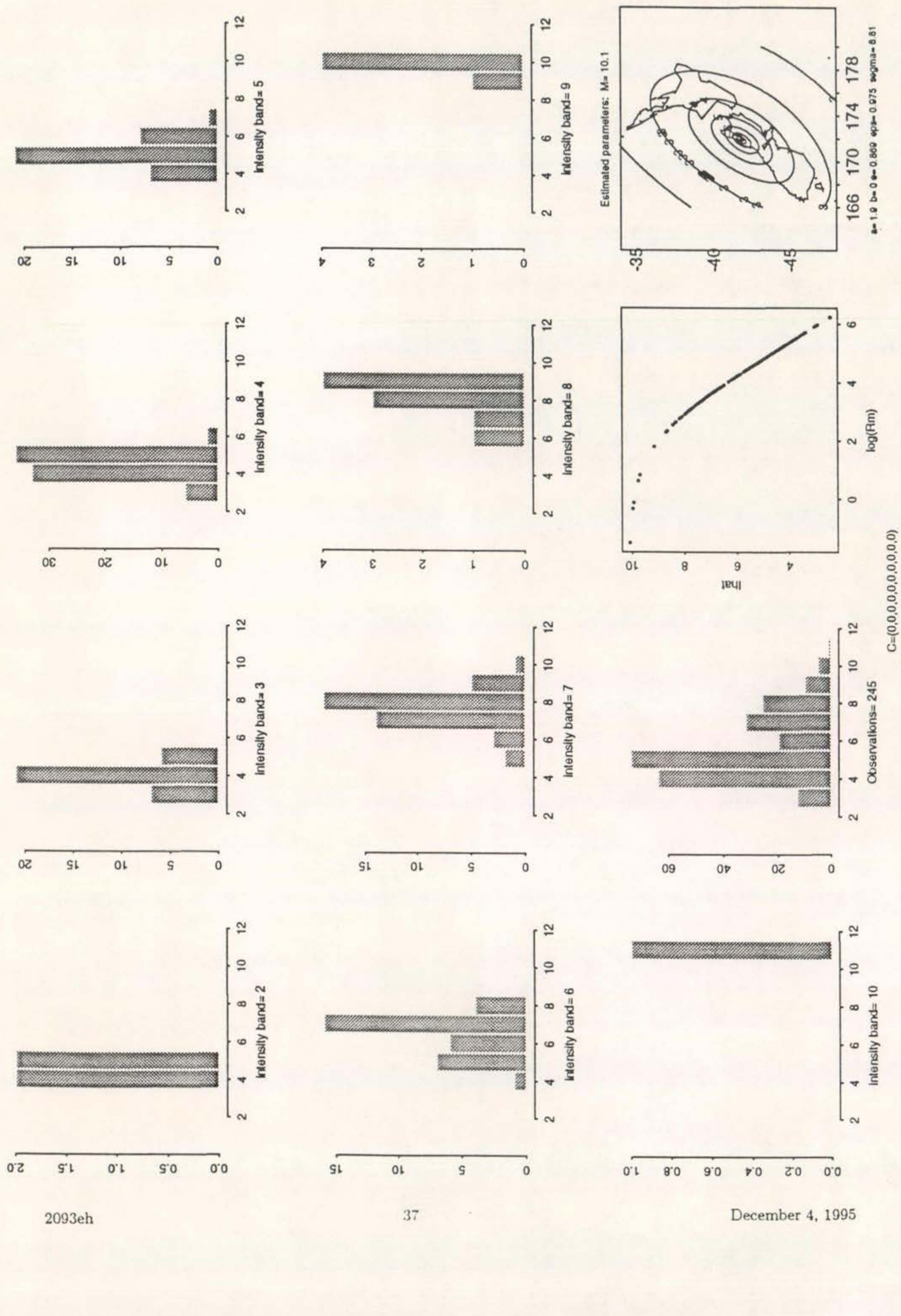


Fig. 20 Histograms of the distribution of observed intensities within epicentral bands, attenuation model and isoseismals, without bias term: Edgecumbe earthquake

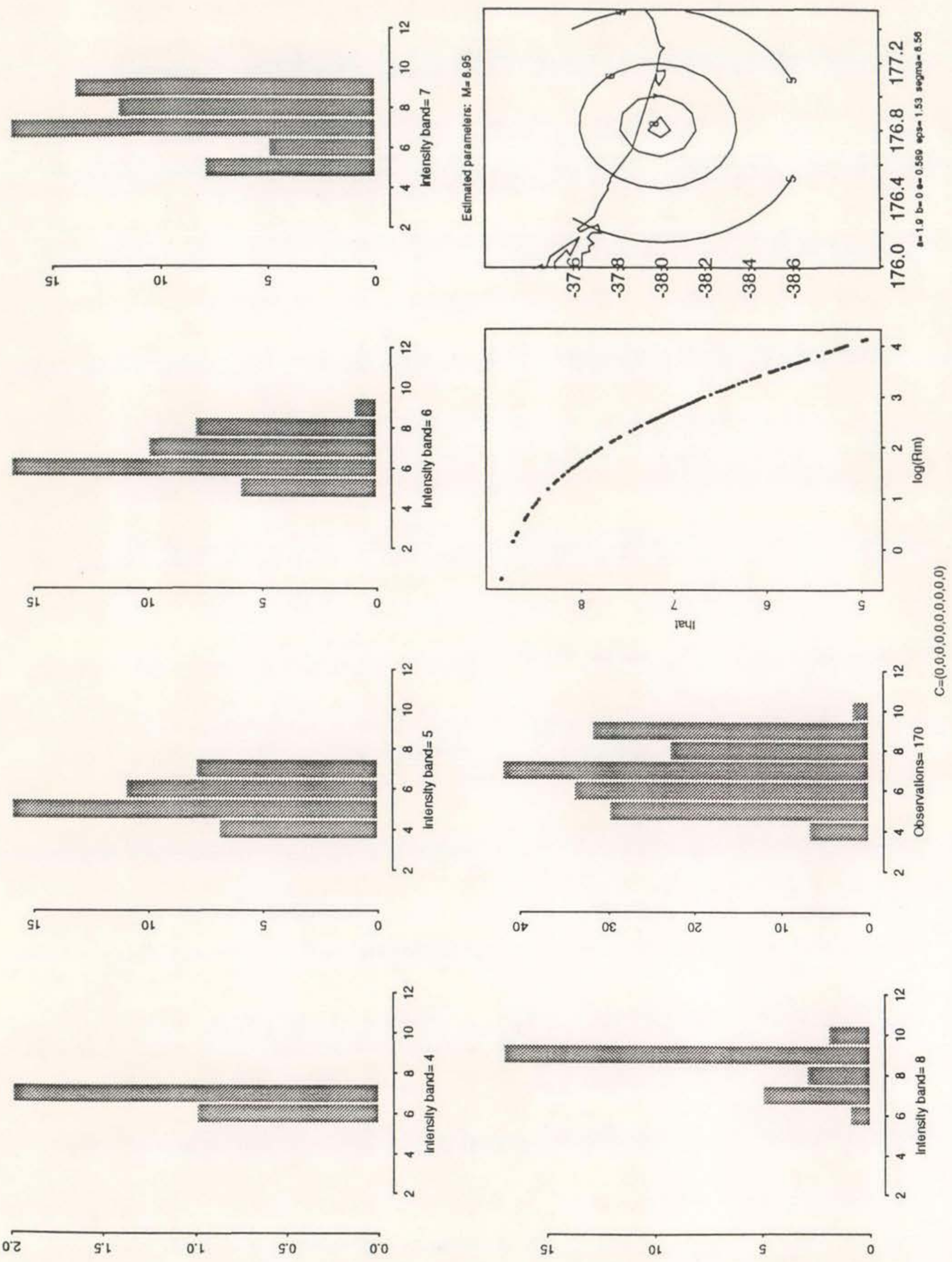


Fig. 21 Histograms of the distribution of observed intensities within epicentral bands, attenuation model and isoseismals, incorporating bias term: Buller earthquake

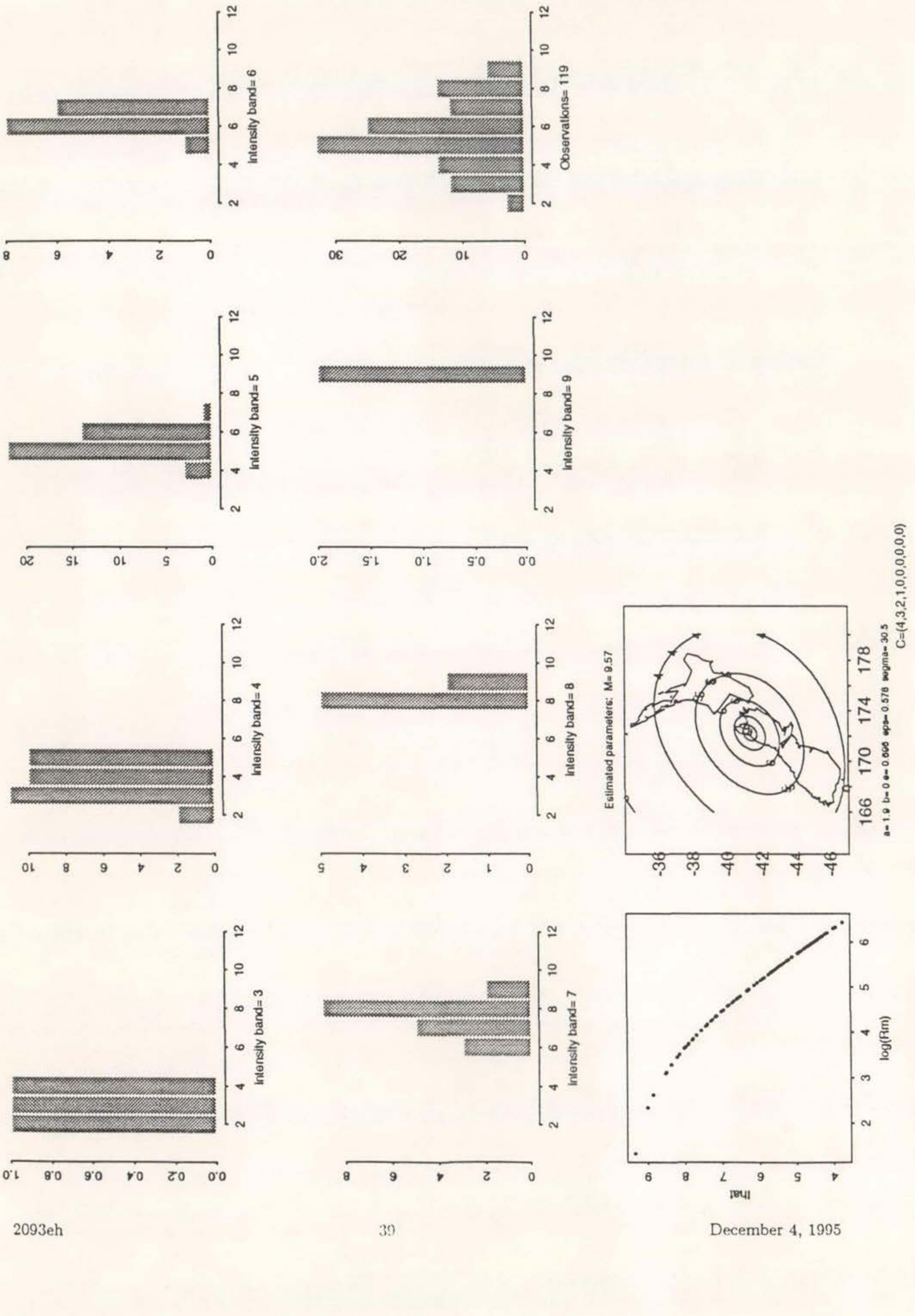


Fig. 22 Histograms of the distribution of observed intensities within epicentral bands, attenuation model and isoseismals, incorporating bias term: Inangahua earthquake

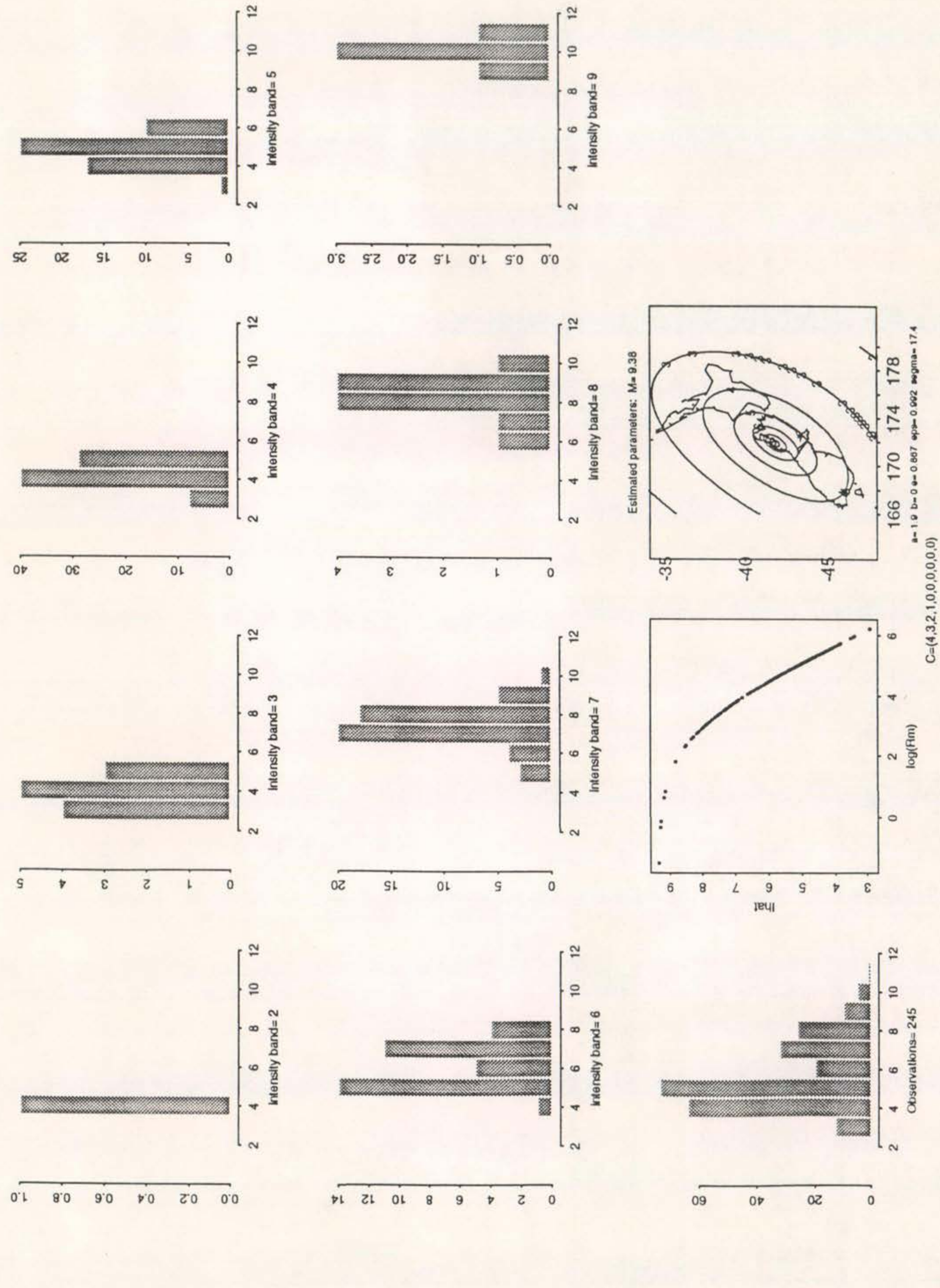


Fig. 23 Histograms of the distribution of observed intensities within epicentral bands, attenuation model and isoseismals, incorporating bias term: Edgecumbe earthquake

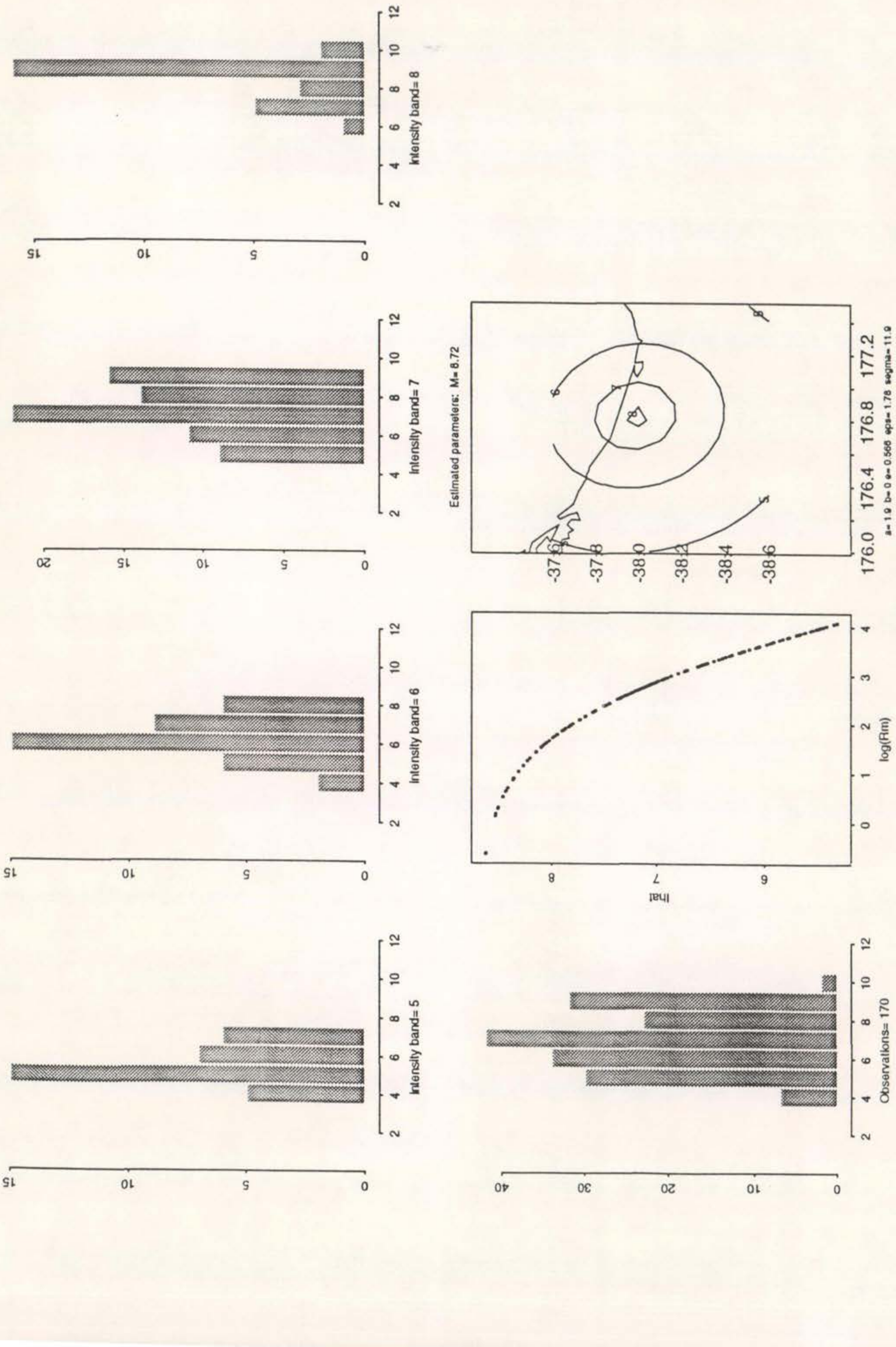


Fig. 24 Proportion of MMI observations within each Isoseismal band, modelled with bias term