Estimating fish concentration and location using trawl data with proportional errors

Sue Paul G A Vignaux

May 28, 1999

Abstract

This extends work reported on by Paul and Vignaux (1999). A number of randomly placed 'trawls' are placed over a concentration of fish. Each trawl makes a catch of fish that depends on the overlap of the trawl with the fish concentration but has some random error. A Bayesian analysis shows how to estimate the location and density of the fish concentration from observations of the catches. This reports on a short series of experiments where the variance of the assumed errors are assumed to be proportional to the overlap of a trawl onto a school and where there is some non-zero density of fish outside the school.

This is $Revision : 1.4$

Contents

1 Introduction

A number of randomly placed 'trawls' are placed over a concentration of fish. Each trawl makes a catch of fish that depends on the overlap of the trawl with the fish concentration but has some random error. The Bayesian analysis reported in Paul and Vignaux (1999) shows how to estimate the location and density of the fish concentration from observations of the catches. Here we study a modification of that model where the variance of the assumed errors are proportional to the overlap. At the same time we extend the model to include the case where the density is non-zero outside the school. this is done to remove the case of zero assumed error in catches.

2 The Model

With the model of constant variance we can write the likelihood of the set of data, $\{d_i\}$, as

$$\prod_{i=1}^{n} P(d_i|x,R) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(d_i - fO_i)^2}{2\sigma^2}} \tag{1}$$

$$= \left(\frac{1}{2\pi\sigma^2}\right)^{\frac{n}{2}} e^{-\sum_{i=1}^{n} \frac{(d_i - fO_i)^2}{2\sigma^2}} \tag{2}$$

where $d_i = O_i f + \epsilon_i$. O_i denotes the overlap of the i^{th} trawl into the school and f is the linear fish density. In the previous model the ϵ_i were distributed as $N(0, \sigma^2)$, implying that $d_i - O_i f$ was also distributed as $N(0, \sigma^2)$. The σ was both known and and constant for all d_i.

We now consider a case where σ varies with each d_i, so that $d_i - O_i f$ is distributed as $N(0, \sigma_i^2)$. Specifically, the σ_i are proportional to the overlap (and hence the theoretical catch).

We can write $\sigma_i \propto O_i f$ or similarly $\sigma_i = gO_i f$ where g is the proportionality constant. Rearranging the above, we can write

$$g = \frac{\sigma_i}{O_i f} \tag{3}$$

Previously we had typically used a fixed σ of 0.1. Corresponding to this level of error we would use $g \sim 0.05$.

We can now rewrite the likelihood as

$$P(\{d_i\}|x,R) \propto \left(\frac{1}{2\pi g O_i f^2}\right)^{\frac{n}{2}} e^{-\sum_{i=1}^{n} \frac{(d_i - fO_i)^2}{2gO_i f^2}} \tag{4}$$

and using Bayes' rule we can write

$$P(x, R|\{d_i\}) \propto \left(\frac{1}{2\pi g O_i f^2}\right)^{\frac{n}{2}} e^{-\sum_{i=1}^{n} \frac{(d_i - f O_i)^2}{2 g O_i f^2}} \tag{5}$$

Note that the posterior distribution is a function of $O_i f$ and it becomes becomes undefined if $O_i f = 0$. Thus if there is no overlap the above density is not defined. If we introduced a small background density, however, to ensure that we never got a zero overlap, then this problem would be eliminated. We could set the background density $b \sim 1 \times 10^{-10}$.

3 One dimensional case

For one school of fish in the one-dimensional case, the model is as before. Figure 1 shows the situation. The model for the school of fish is as before.

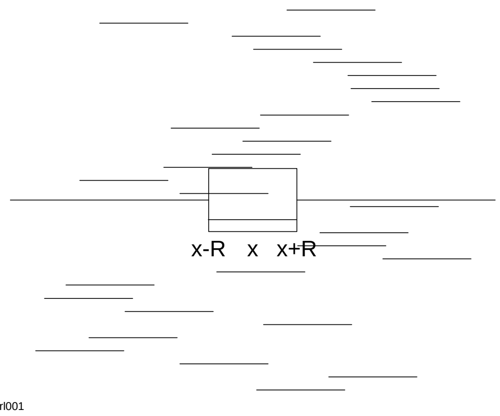

Figure 1: One dimensional case. The horizontal dotted lines surrounding the school represent the trawls. Here all trawls are horizontal. In other models they can be in random directions. The *theoretical catch* of a given trawl depends on its overlap with the school.The school is centred at x and has a width of $2R$.

All calculations were carried out using S-PLUS (Becker et al. (1988)) on Sun SPARC workstations.

3.1 Contour Plots

Given a set of horizontal trawl data, including the location and catches of the trawls we next calculate the posterior probability distribution as a function of a range of putative values of the school location. These can be plotted in the form of a contour plot. A typical plot is shown in Figures 2 and 3 which also compare constant variance results (left side) with proportional variance results (right side). Random trawls with uniformly distributed midpoints and uniform lengths with a mean of 2 were used. In all cases a small background density was present to eliminate the special cases where the error variance would be zero.

The figures show how accurate the location is. The peak is in each case represented by the small dot near the junction of the horizontal and vertical lines in the centre of the figures. They cross at the true location of the school. Note also that these figures are in terms of log-probability.

The strange linear shape of the contours in the proportional case have not satisfactorily been explained. Each line certainly corresponds to the case when one particular trawl drops off the end of the school. It varies with both x and R because the edge of the school varies linearly with both of these.

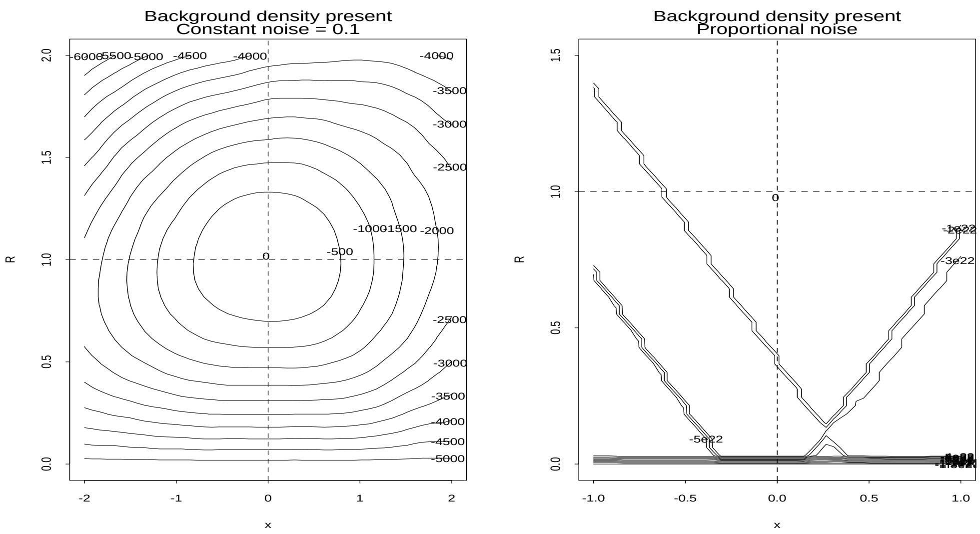

Figure 2: One-dimensional case. Log posterior probability contours for $-1 < x < 1$ and $0 < R < 1.5$ after 50 trawls were generated around the school. The figure on the left shows the results when constant variance errors were used; that on the right shows the results when proportional errors with $g = 0.05$ were used. The peak location in the right-hand figure is the small dot close to the true x, R point.

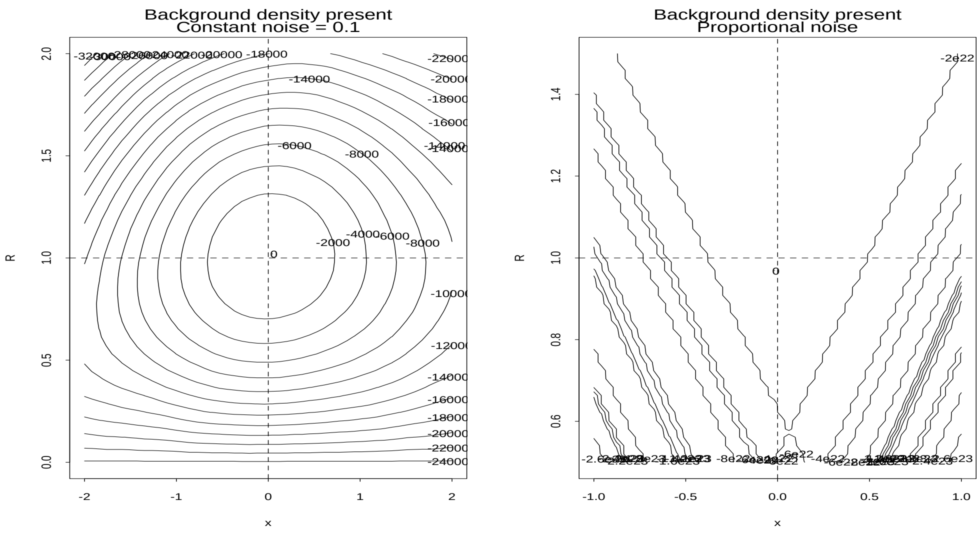

Figure 3: One-dimensional case. Log posterior probability contours for $-1 < x < 1$ and $0 < R < 1.5$ after 300 trawls were generated around the school. The figure on the left shows the results when constant variance errors were used; that on the right shows the results when proportional errors with $g = 0.05$ were used. The peak in the right-hand figure is the small dot close to the true x, R point.

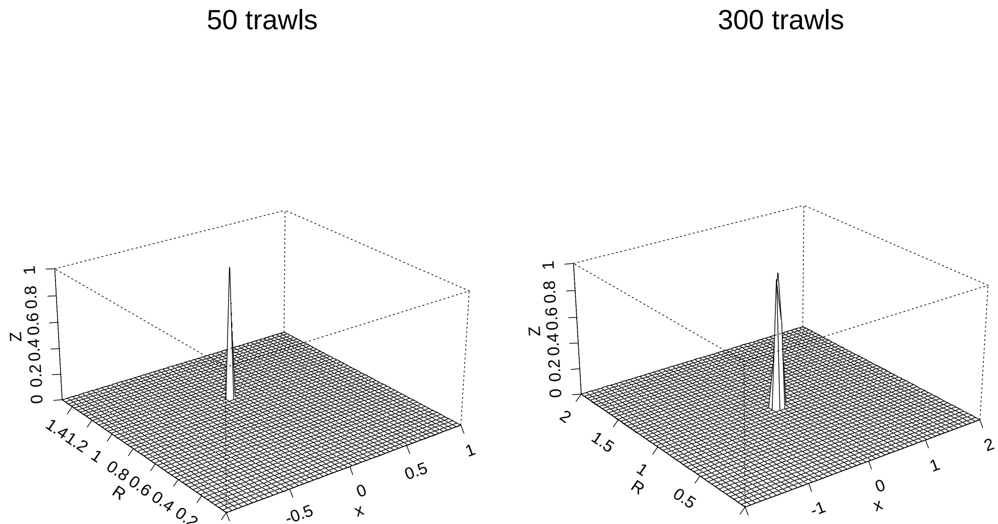

Figure 4: One-dimensional case. Perspective plots of the posterior probability using proportional errors. $-1 < x < 1$ and $0 < R < 1.5$ and 50 and 300 trawls were generated around the school. The sharpness of the peaks results in difficulties in getting the location numerically.

3.2 Nonlinear Maximisation and Error Analysis

To estimate the parameters x and R numerically, we used the *Splus* function *nlmin* as in previous studies. We would like to see the effect of the number of trawls in a sample on the accuracy of the estimates. Table 1 and Figure 5 show the results.

Table 1: One-dimensional case. Estimates of x and R as the number of trawls change.

Number of trawls	x	R
50	3.40×10^{-3}	0.9891
100	2.10×10^{-3}	1.0002
200	8.00×10^{-4}	1.0004
400	-1.25×10^{-5}	1.0000
800	-1.24×10^{-3}	0.9993
1600	2.02×10^{-6}	1.0000
3200	-3.30×10^{-4}	0.9990
6400	0.00	1.0000

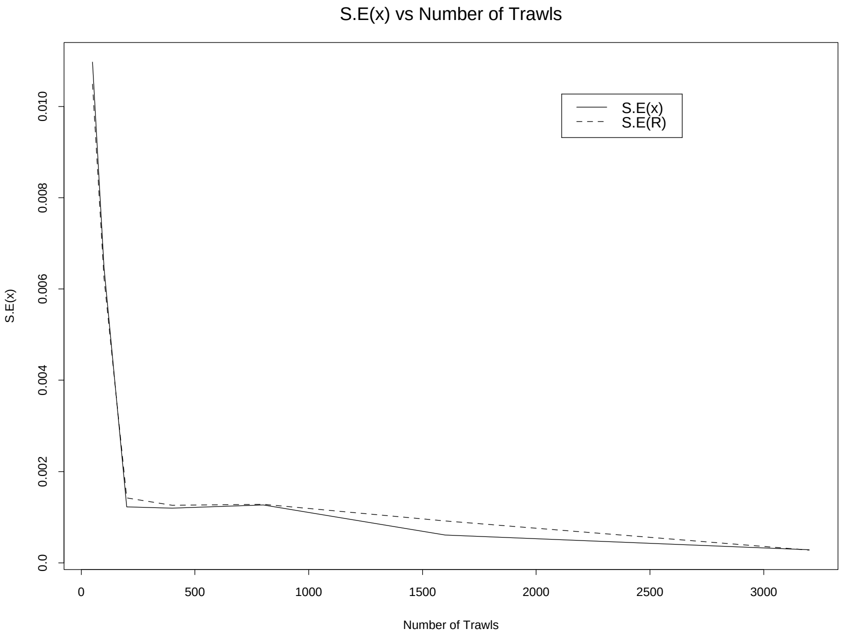

Figure 5: One-dimensional case. The accuracy of the estimates of x and R increase with the number of trawls.

4 Two dimensional rectangular school

A few experiments were attempted using proportional errors with the the two-dimensional model shown in Figure 6. The model for the rectangular school is exactly as developed in Paul and Vignaux (1999) but the errors are now assumed to be proportional to the overlap, as in equation 5.

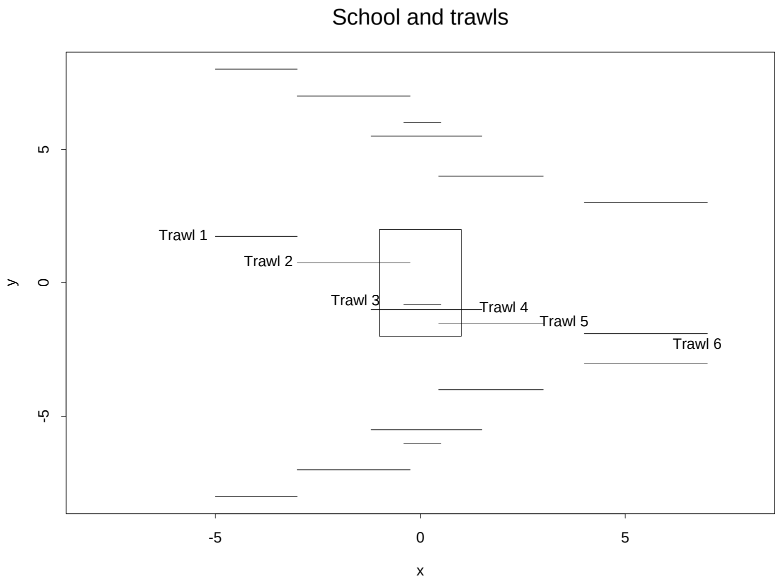

Figure 6: Two dimensional case with a rectangular school. The school is centred at x and has a horizontal width of $2R_x$ and a vertical width of $2R_y$. The horizontal lines represent the trawls and are set out to show the different sorts of overlaps that can occur.

4.1 Contour Plots

Contour plots of the posterior log-probability calculations are shown in Figures 7, and 8. In these experiments we were varying the inside and outside densities but we did not attempt to estimate them.

The different behaviour of the contours as R_x and R_y (particularly the former) get near 0 is of interest. Of course, neither can go negative and it may be useful to see if a Jeffrey's prior would be more sensible for these parameters.

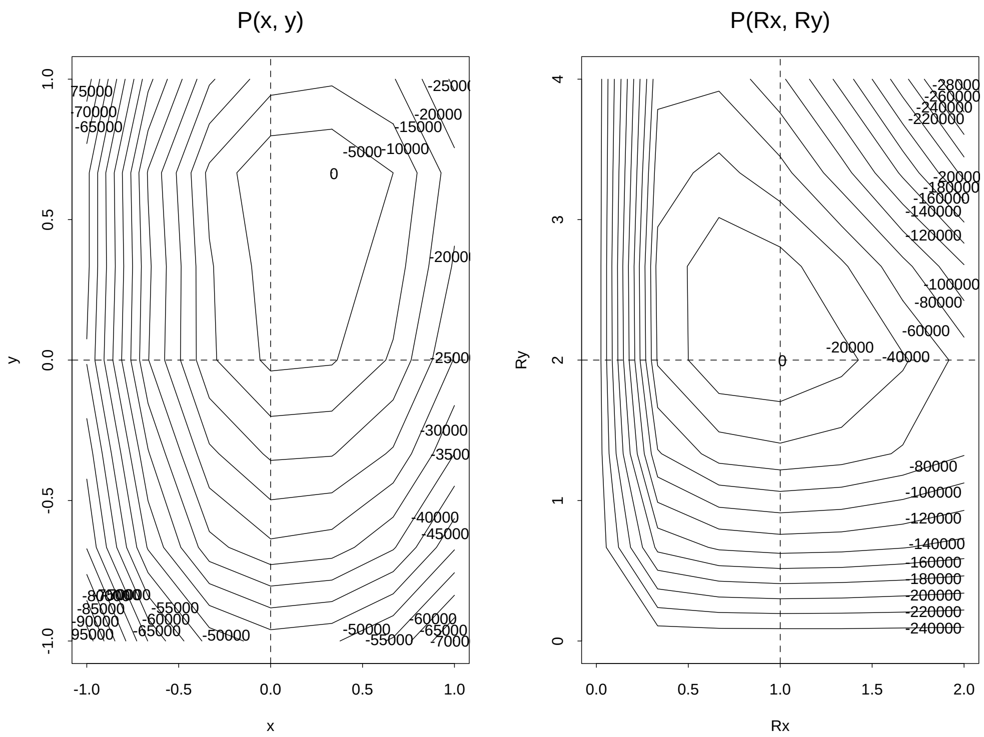

Figure 7: Two dimensional case with a rectangular school. 100 trawls, a background density of 0.8, and a school density (f) of 3 was used. $R_x = 1$ and $R_y = 2$. The estimates of the location of the centre of the school, x and y are not very accurate in the y direction. This may be because the trawls are all horizontal. R_x and R_y are located quite accurately.

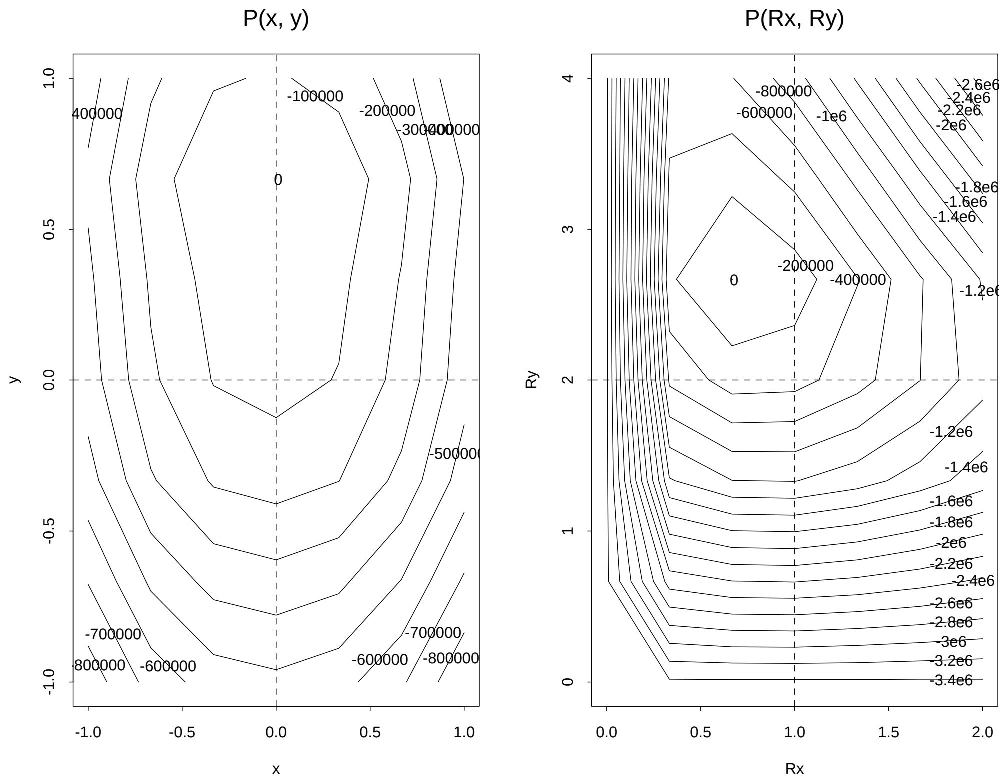

Figure 8: Two dimensional case with a rectangular school. 300 trawls, a background density of 0.5, and a school density (f) of 3 was used. $R_x = 1$ and $R_y = 2$. The estimate for x is close to its "true" value but the y estimate is not accurate. The estimates of R_x and R_y are also wrong. More worrying is the different behaviour of the these as they get near 0.

5 Conclusion

Experiments in using proportional errors have produced some useful results. In particular, the accuracy of the size of the school has been good. However not all variables have been tested and further investigation into the interaction between the non-school density, b, and the proportional errors should be looked at.

We found numerical difficulties when we tried to used *nlmin* in the more complicated two-dimensional case.

Further work should look at:

1. Varying the relative magnitudes of the school and non-school density, f and b, and the error variance coefficient, g.

2. Further investigate numerical methods for finding the maximum. This may require rewriting the code in C rather than S-PLUS.

3. Try a Jeffrey's prior for the dimension parameters, R_x and R_y.

6 Acknowledgement

We acknowledge the support given by a research grant from IGNS.

References

Becker, R. A., Chambers, J. M., and Wilks, A. R. (1988). *The New S Language*. Wadsworth & Brooks, Pacific Grove, CA.

Paul, S. and Vignaux, G. A. (1999). Estimating fish concentration and location using trawl data. Technical Report 99a, MCS, Victoria University, Wellington, NZ.