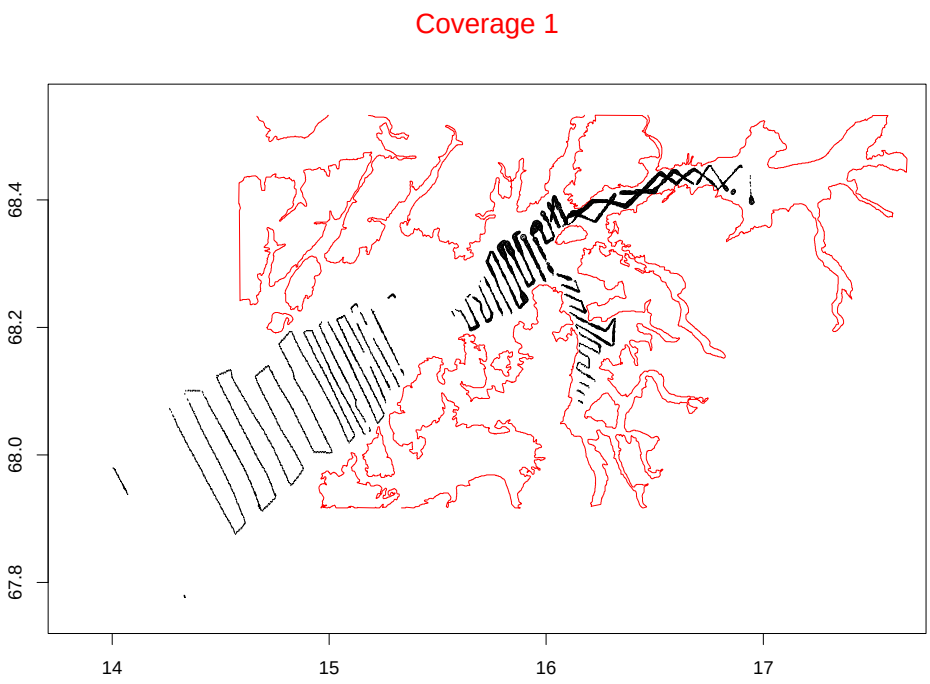
# Initial analysis of data from Norwegian Spring Spawning Herring Survey in Vestfjord in November 2001




**SAMBA/20/02**

Anders Løland
Gudmund Høst

July 2002

Norsk Regnesentral
ANVENDT DATAFORSKNING

**NR-notat**/NR Note

**Tittel**/Title: Initial analysis of data from Norwegian Spring Spawning Herring Survey in Vestfjord in November 2001

**Dato**/Date: July
**År**/Year: 2002
**Notat nr.**/Note no.: SAMBA/20/02

**Forfatter**/Author: Anders Løland
Gudmund Høst

**Sammendrag**/Abstract: This note sums up the initial analysis of the acoustic measurements of Norwegian Spring Spawning Herring (NSSH) in Vestfjord, Norway, November 2001. The data are corrected for shadow effect, avoidance effect and target strength effect and then integrated over the water column. One characteristic of the data is the presence of many 0 measurements. Generally, the variograms are increasing with distance, while there seems to be no trend for the variograms with respect to time.

**Emneord**/Keywords: Space-time data, empirical variograms, Norwegian Spring Spawning Herring

**Målgruppe**/Target group: Researchers in the project

**Tilgjengelighet**/Availability: Open

**Prosjekt**/Project: Absolute abundance estimation of fish

**Prosjektnr.**/Project no.: 220140

**Satsningsfelt**/Research field: Marine Resources

**Antall sider**/No. of pages: 33

Norsk Regnesentral/Norwegian Computing Center
Gaustadalléen 23, Postboks 114 Blindern, 0314 Oslo, Norway
Telefon 22 85 25 00, telefax 22 69 76 60

# Contents



# List of Tables



# List of Figures

# 1 Introduction

This note is meant to sum up the initial analysis of the acoustic measurements of Norwegian Spring Spawning Herring (NSSH) in Vestfjord, Norway, November 2001.

# 2 Data

The data are adjusted for shadow effect, avoidance effect and target strength effect (Zhao and Ona, 2002; Ona, 2001) and then integrated over the water column, so the locations are 2-dimensional, the time is 1-dimensional and the measurements are also 1-dimensional. We have measurements $Z(\boldsymbol{s}_i, t_i)$ of a spatial random field $Z(\boldsymbol{s}, t)$ at geographical locations $\boldsymbol{s}_i$ and times $t_i$, $i = 1, \ldots, n$.

The fjord regions of interest are given in Figure 1. In Table 1 we have summed up the start and stop times and locations for the transects from the different areas in the fjord system. A Tysfjord survey takes roughly 7 hours, Vestfjord survey roughly 40 hours and Ofotfjord survey roughly 10 hours.

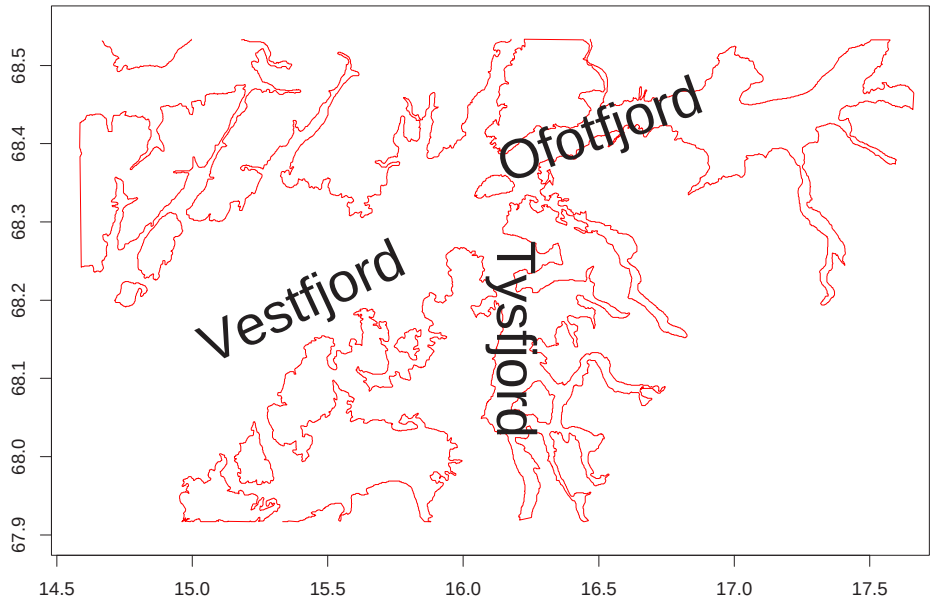



Figure 1: The fjord system.

In Figures 2-13 we have plotted the different fjord transects and measurements. The axes are Longitude and Latitude *decimal* degrees. In the measurement plots, the abundance of herring in each measurement location is proportional to the area of the circle. We see that there are large areas where the measured abundance of herring

| Area | Start date & time | Start location | Stop date & time | Stop location |
|---|---|---|---|---|
| Tysfjord 1 | 21.11 2:56 | N68:17<br>E16:3.2 | 21.11 8:32 | N68:3.7<br>E16:6.2 |
| Vestfjord 1 | 21.11 10:13 | N68:22<br>E16:1.3 | 23.11 4:33 | N67:48.6<br>E14:17.5 |
| Ofotfjord 1 | 25.11 9:27 | N68:22<br>E16:5.0 | 25.11 20:15 | N68:23.1<br>E16:9.8 |
| Tysfjord 2 | 26.11 14:21 | N68:15.8<br>E15:52.2 | 26.11 23:14 | N68:6.3<br>E16:9.5 |
| Vestfjord 2 | 27.11 22:23 | N68:24.3<br>E16:2 | 29.11 11:42 | N68:4.5<br>E14:16.1 |
| Ofotfjord 2 | 25.11 21:55 | N68:22.3<br>E16:10.8 | 26.11 8:07 | N68:21.6<br>E16:2.5 |

Table 1: Data start and stop locations with times.

equals zero, especially in Figures 2 (Vestfjord 1), 10 (Ofotfjord 1) and 12 (Ofotfjord 2). For each transect, we have produced histograms of the measurements. Since there are so many zero measurements, we have put the zeros in a single bar on the left of the other measurements.

In Figure 14 we have combined the fjord regions into one plot. There seems to have been a movement of herring from the 1st to the 2nd coverage from the centre of the fjord west into Vestfjord and south into Tysfjord. In Figure 15 we have made a map of zero (black) and non-zero (red) measurements. The zeros seem to be concentrated in certain areas and not spread out evenly, as we would expect if the distribution of the zeros had a very short correlation range.

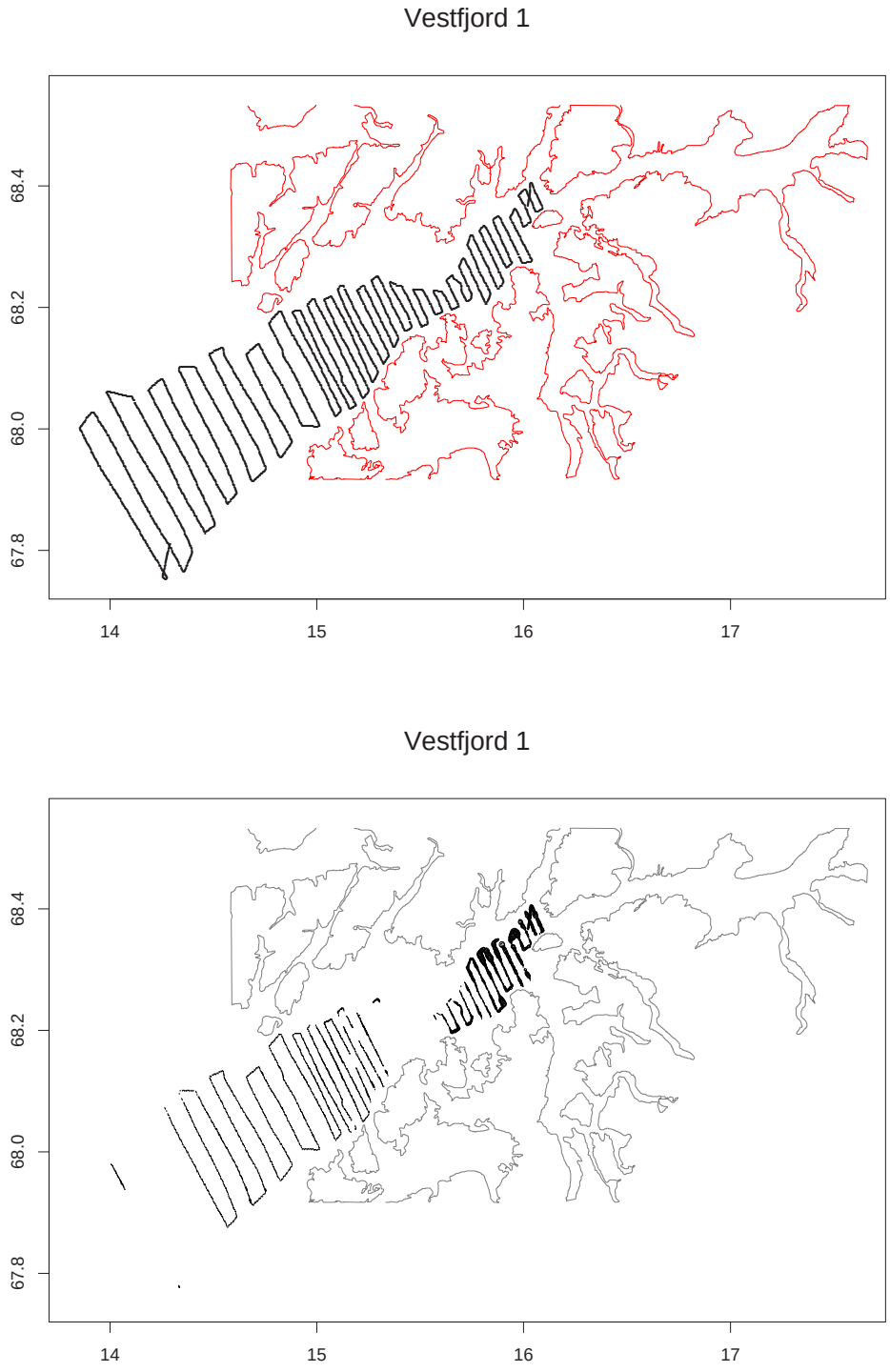



Figure 2: Vestfjord 1 transect (upper) and measurements (lower).

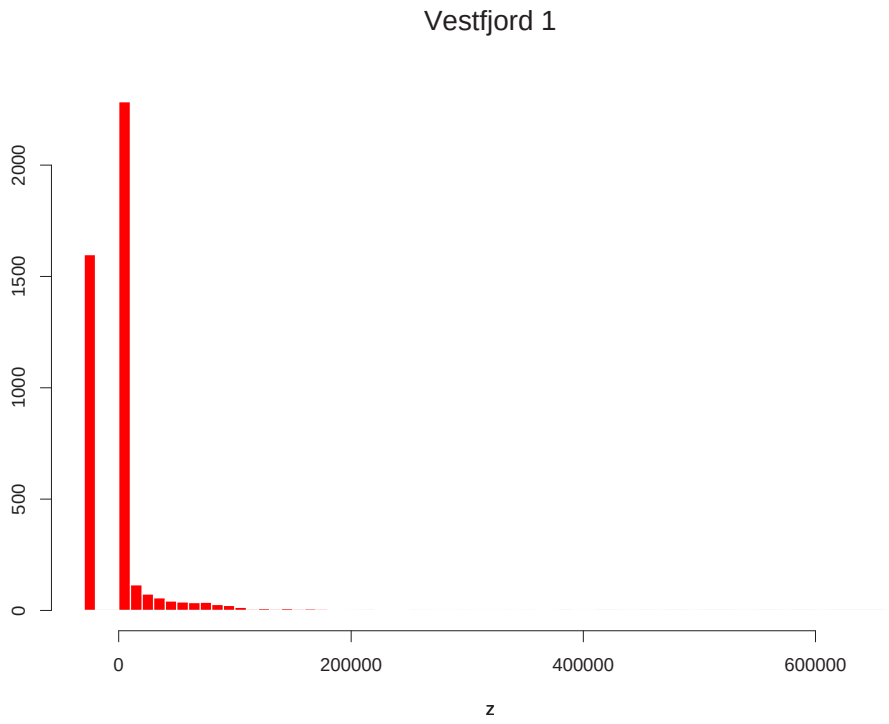



Figure 3: Histogram of measurements on original scale. Vestfjord 1. The single bar on the left represents the (many) zeros.

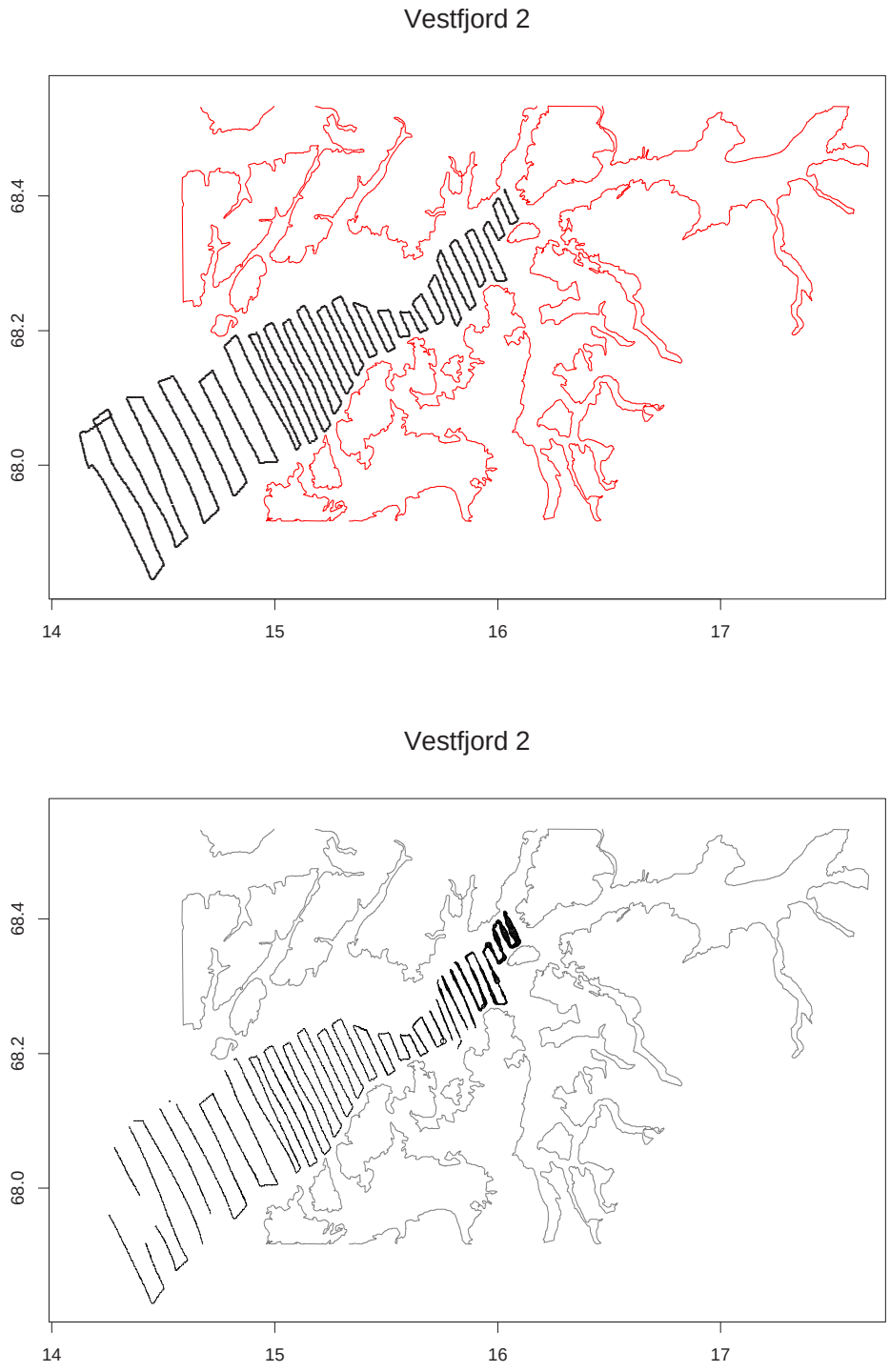



Figure 4: Vestfjord 2 transect (upper) and measurements (lower).

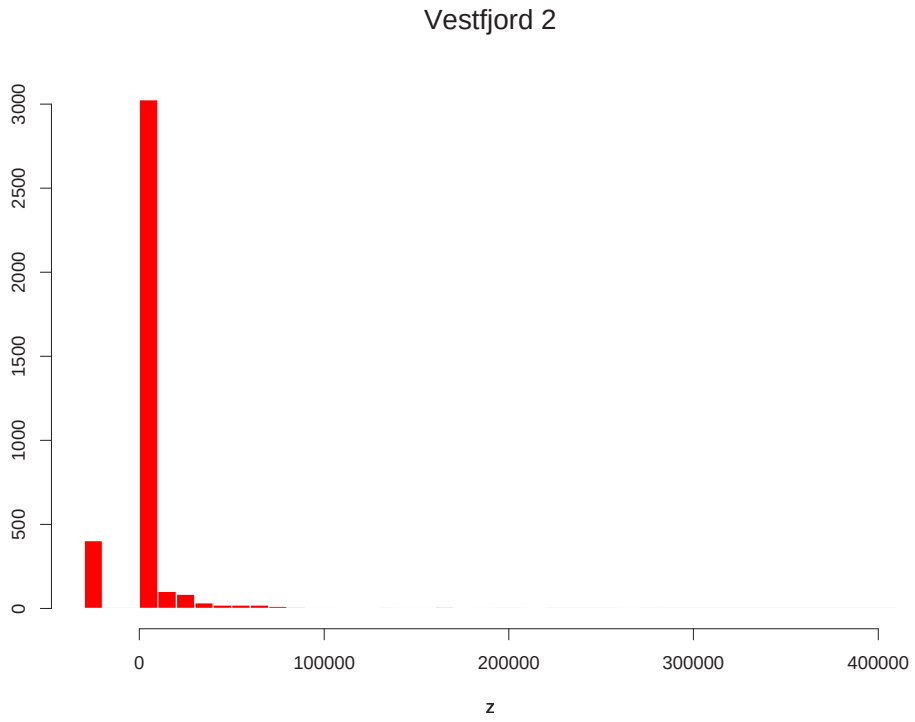



Figure 5: Histogram of measurements on original scale. Vestfjord 2. The single bar on the left represents the (many) zeros.

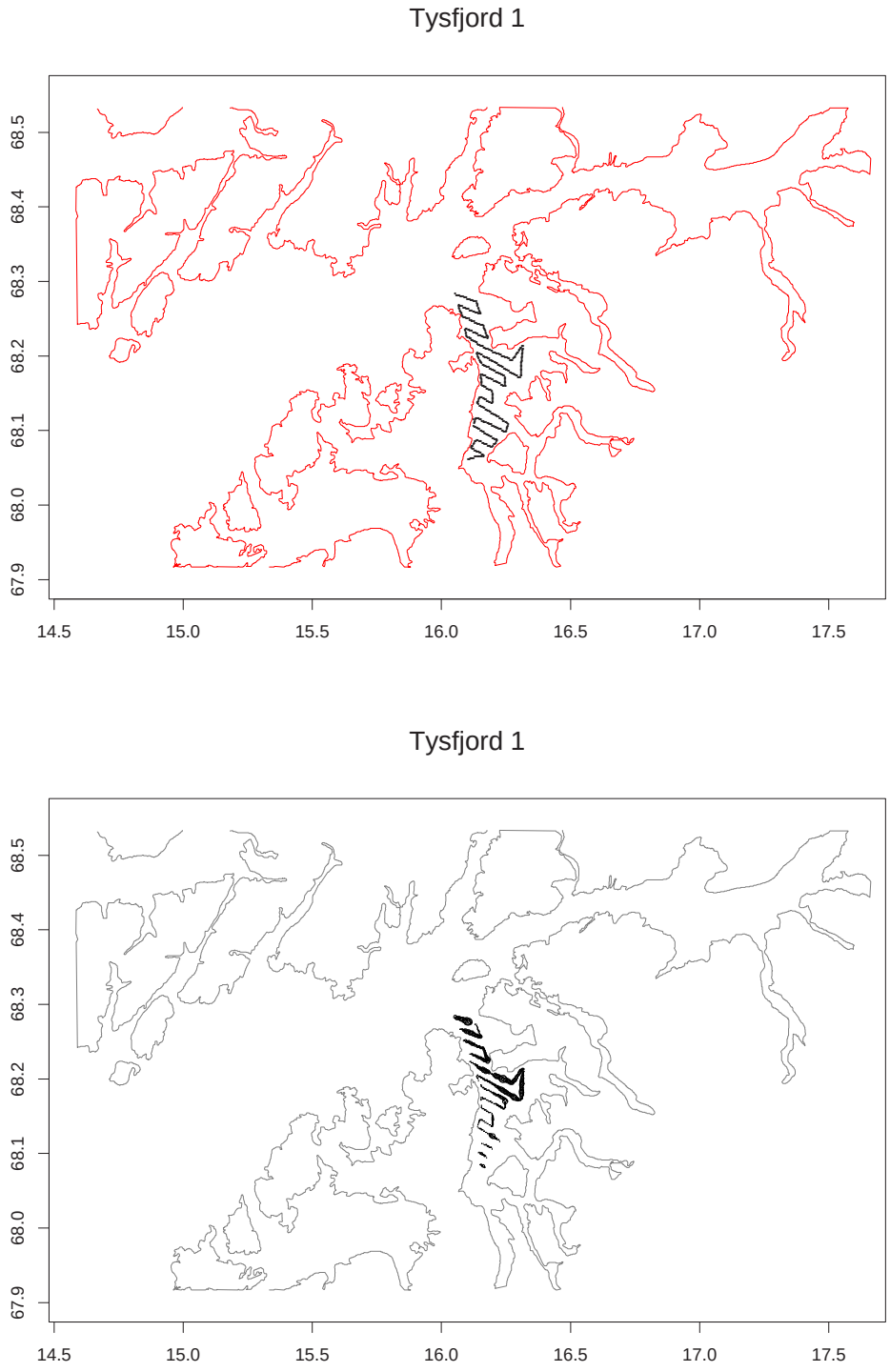



Figure 6: Tysfjord 1 transect (upper) and measurements (lower).

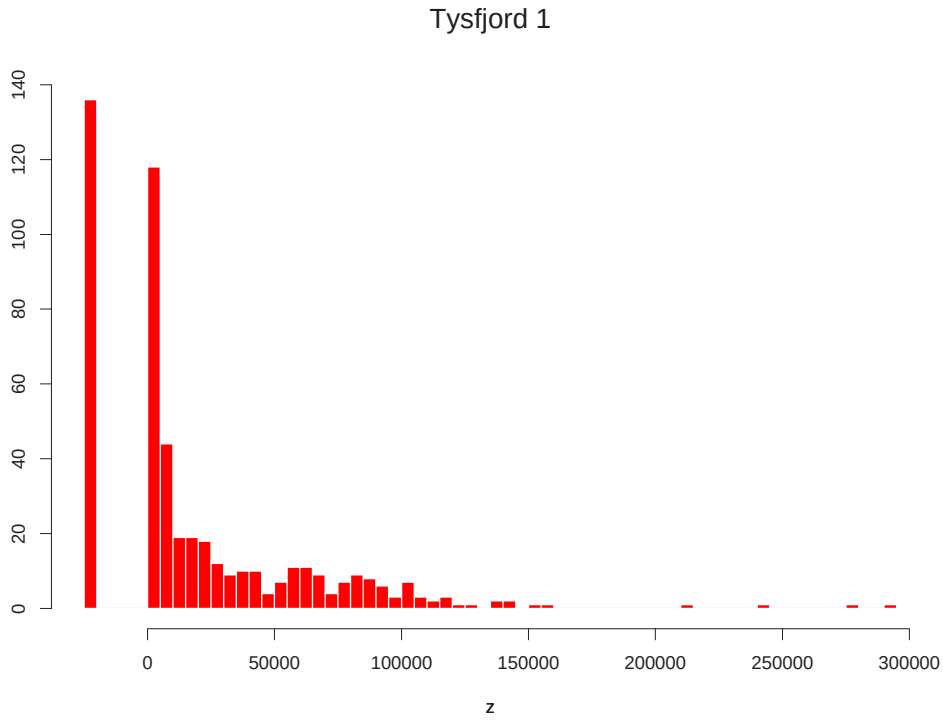



Figure 7: Histogram of measurements on original scale. Tysfjord 1. The single bar on the left represents the (many) zeros.

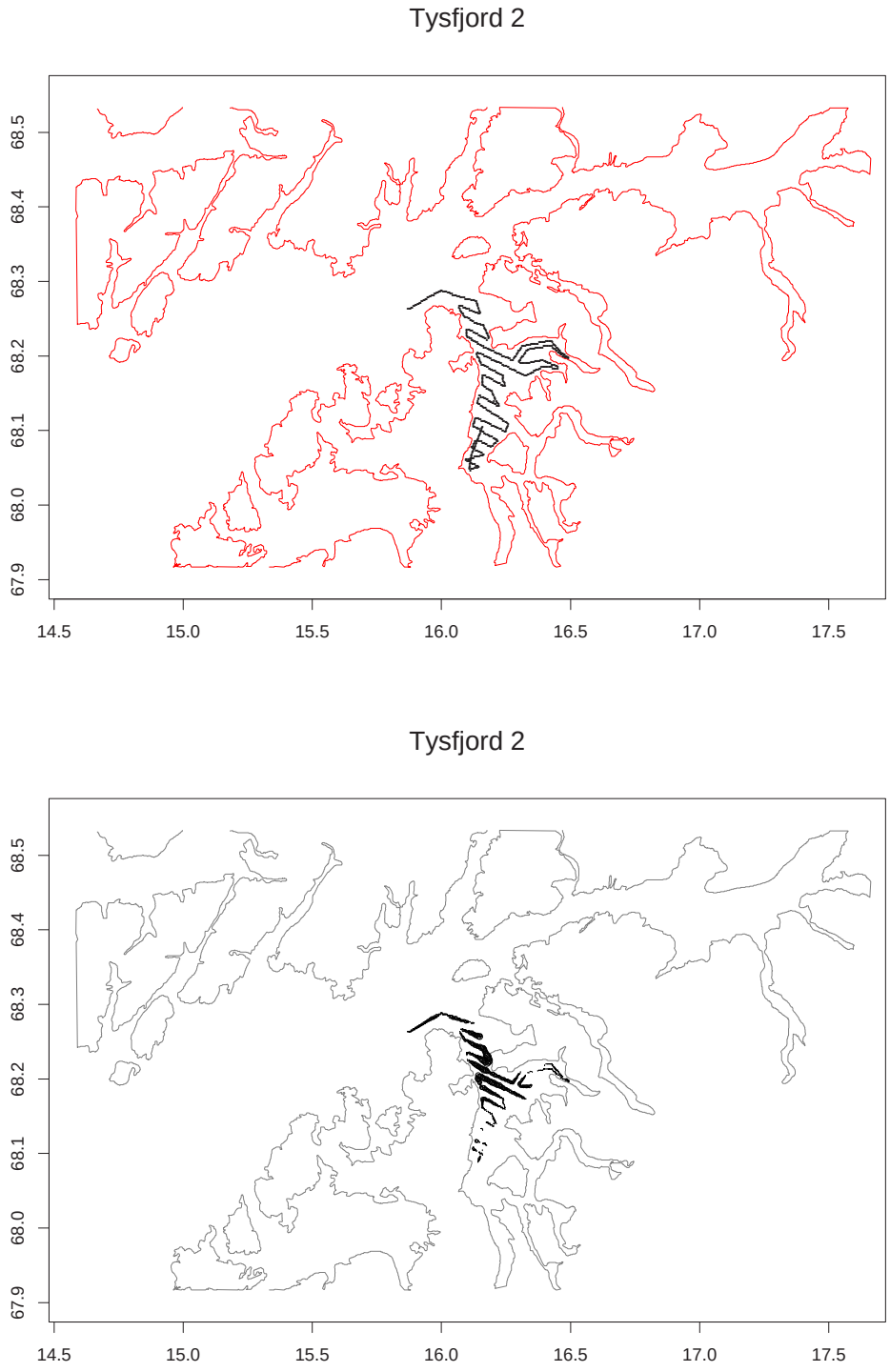



Figure 8: Tysfjord 2 transect (upper) and measurements (lower).

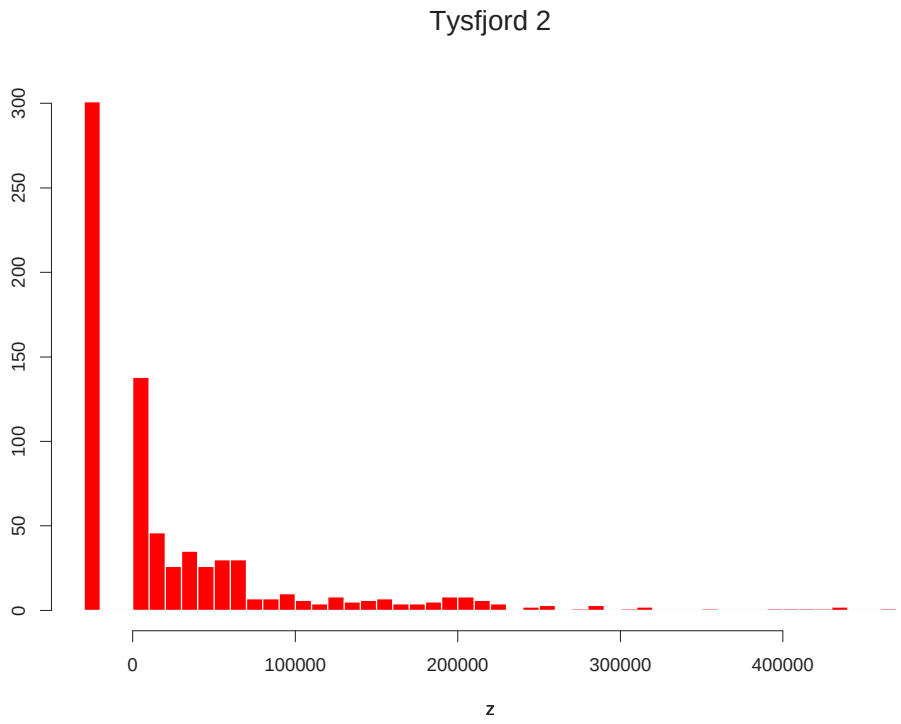



Figure 9: Histogram of measurements on original scale. Tysfjord 2. The single bar on the left represents the (many) zeros.

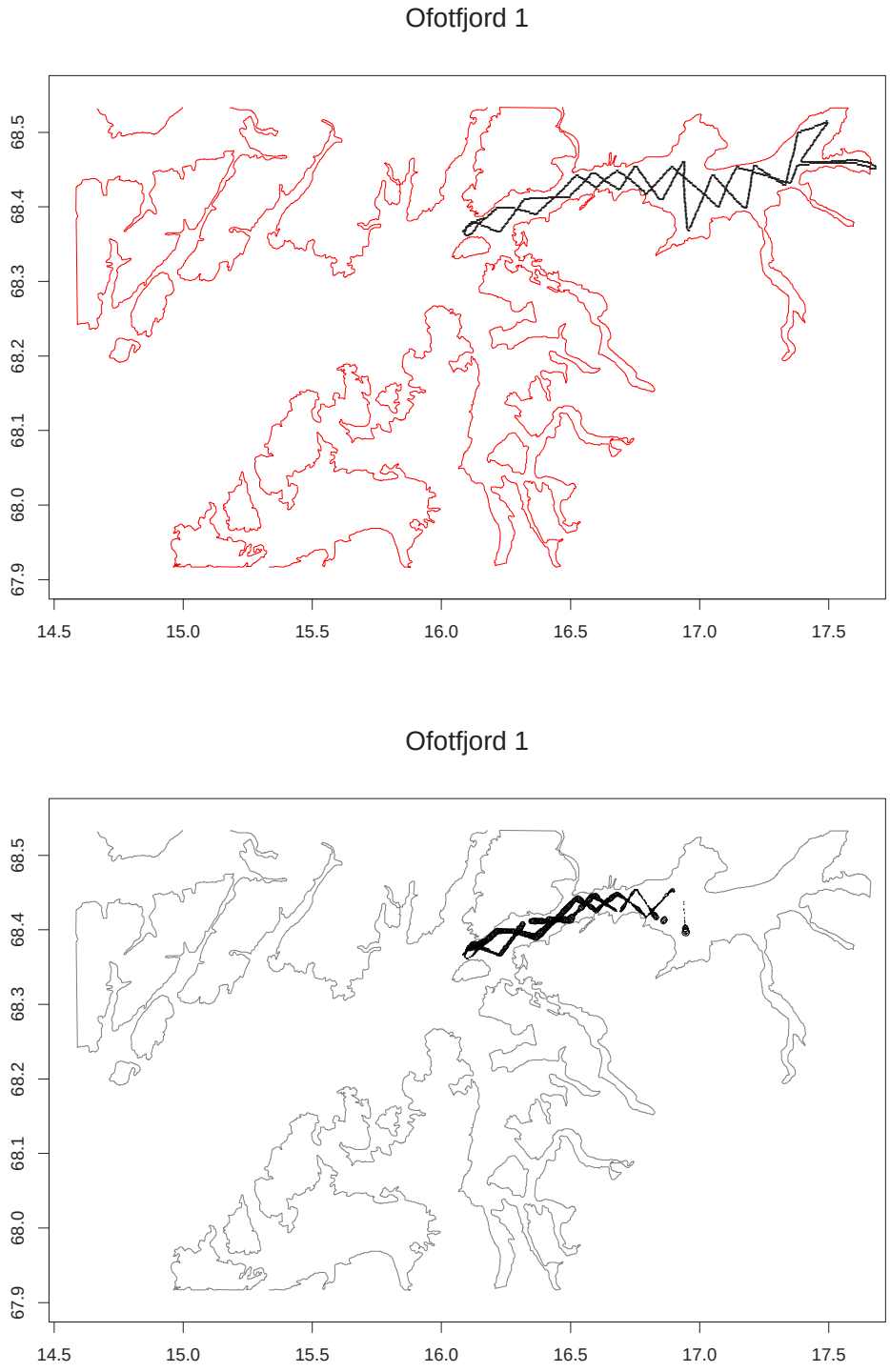



Figure 10: Ofotfjord 1 transect (upper) and measurements (lower).

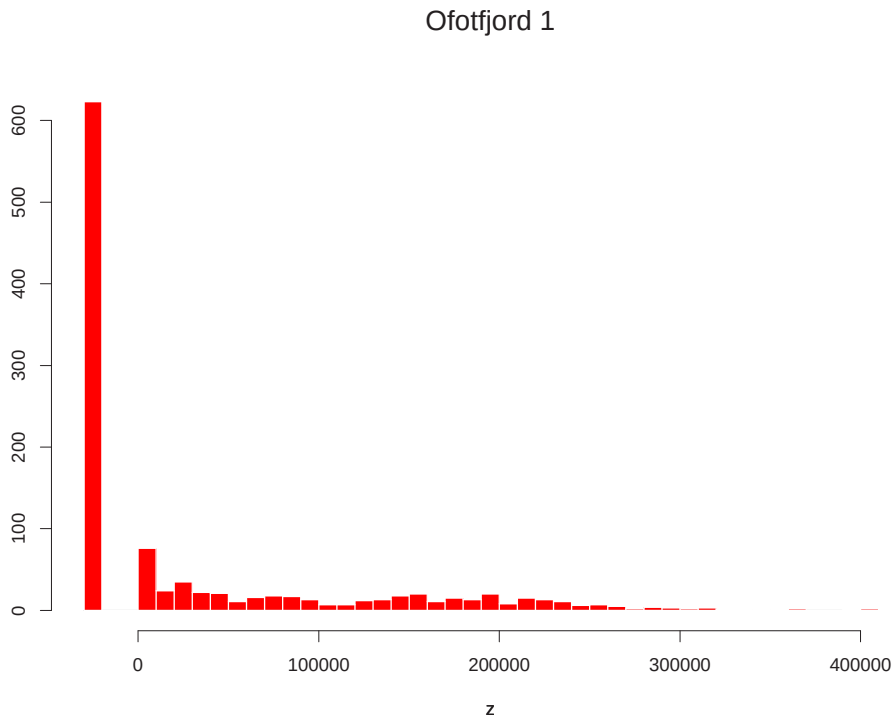



Figure 11: Histogram of measurements on original scale. Ofotfjord 1. The single bar on the left represents the (many) zeros.

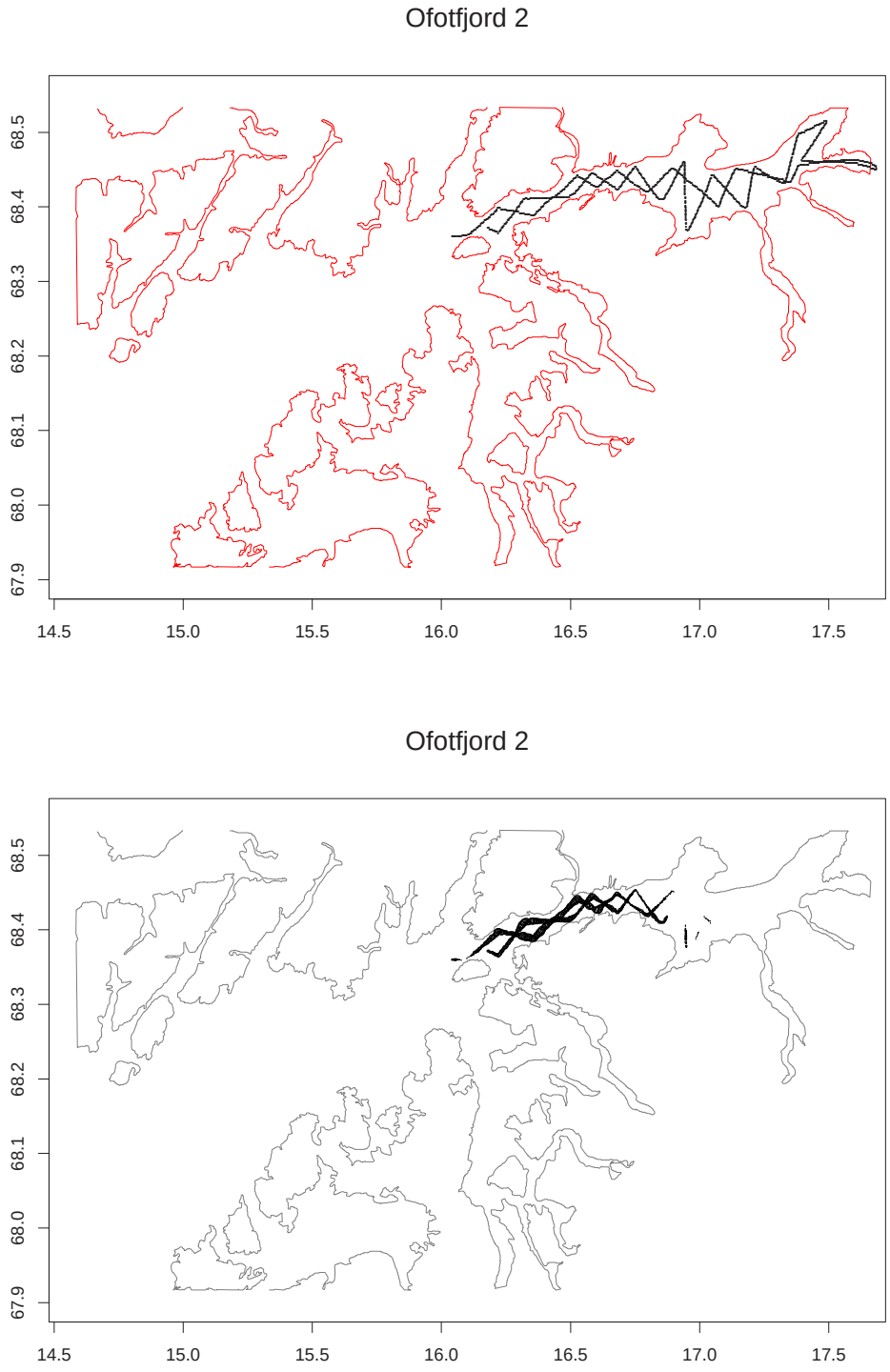



Figure 12: Ofotfjord 2 transect (upper) and measurements (lower).

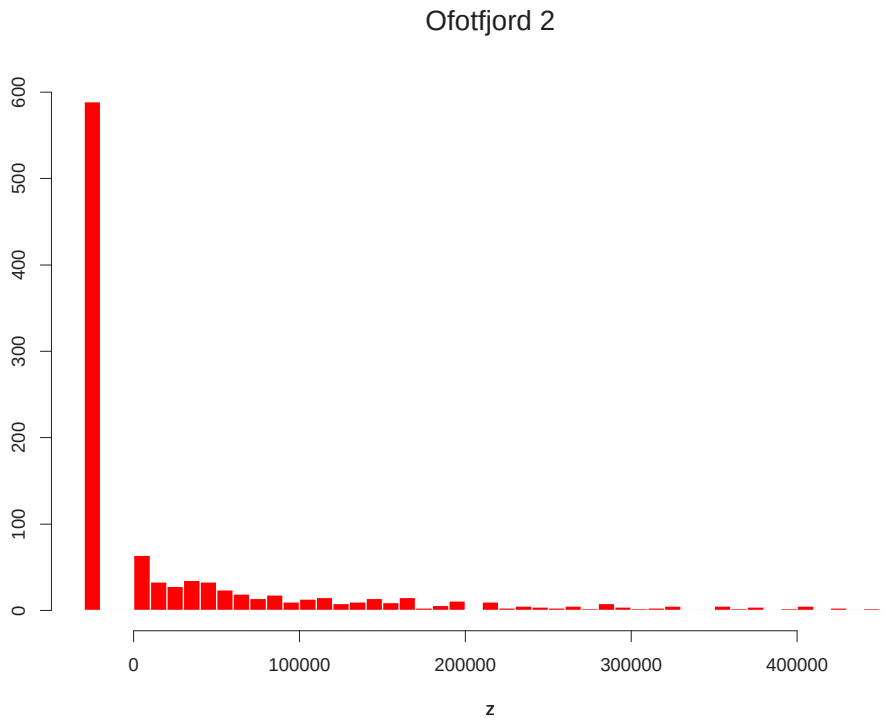



Figure 13: Histogram of measurements on original scale. Ofotfjord 2. The single bar on the left represents the (many) zeros.

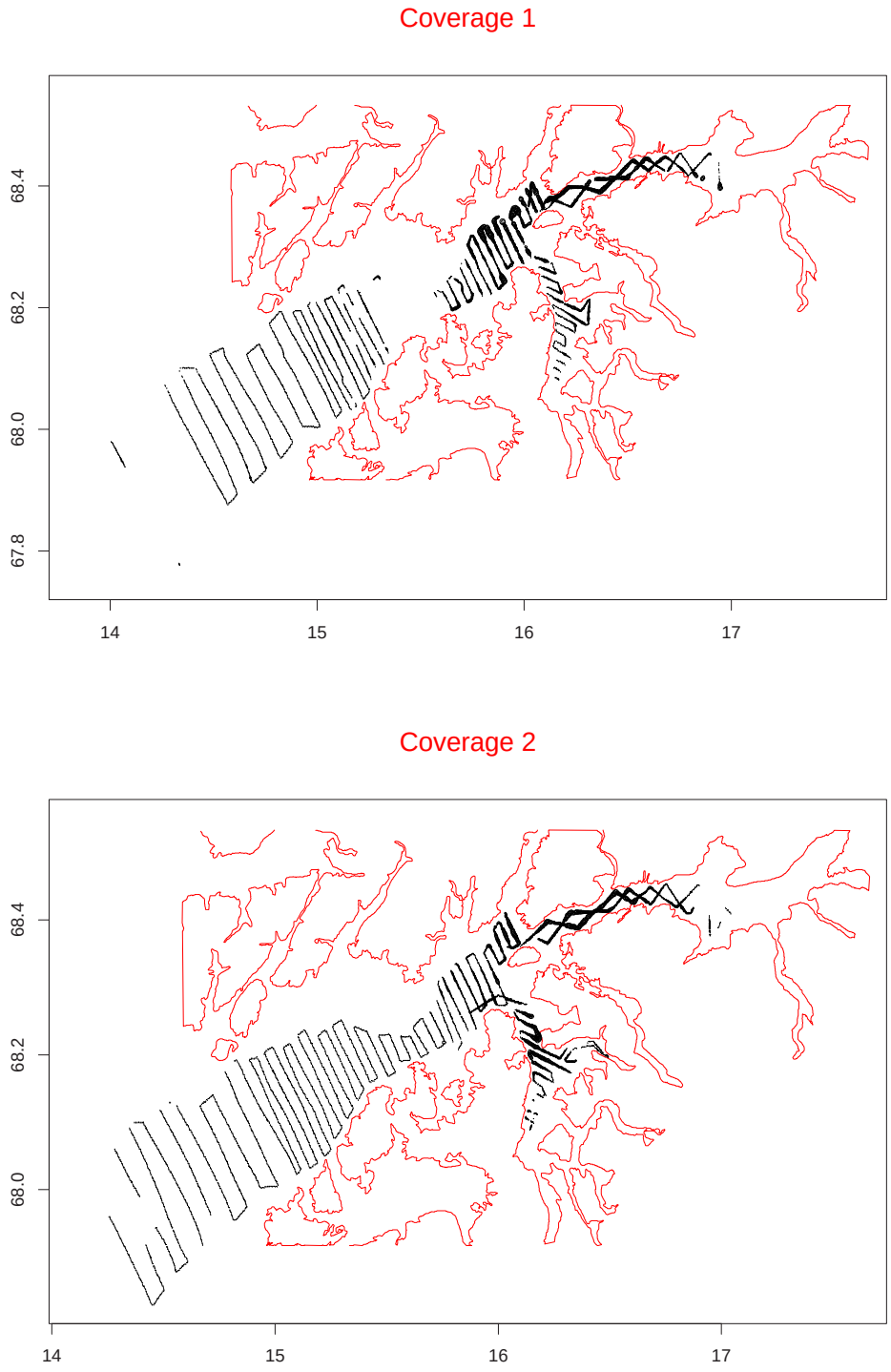



Figure 14: Total coverage.

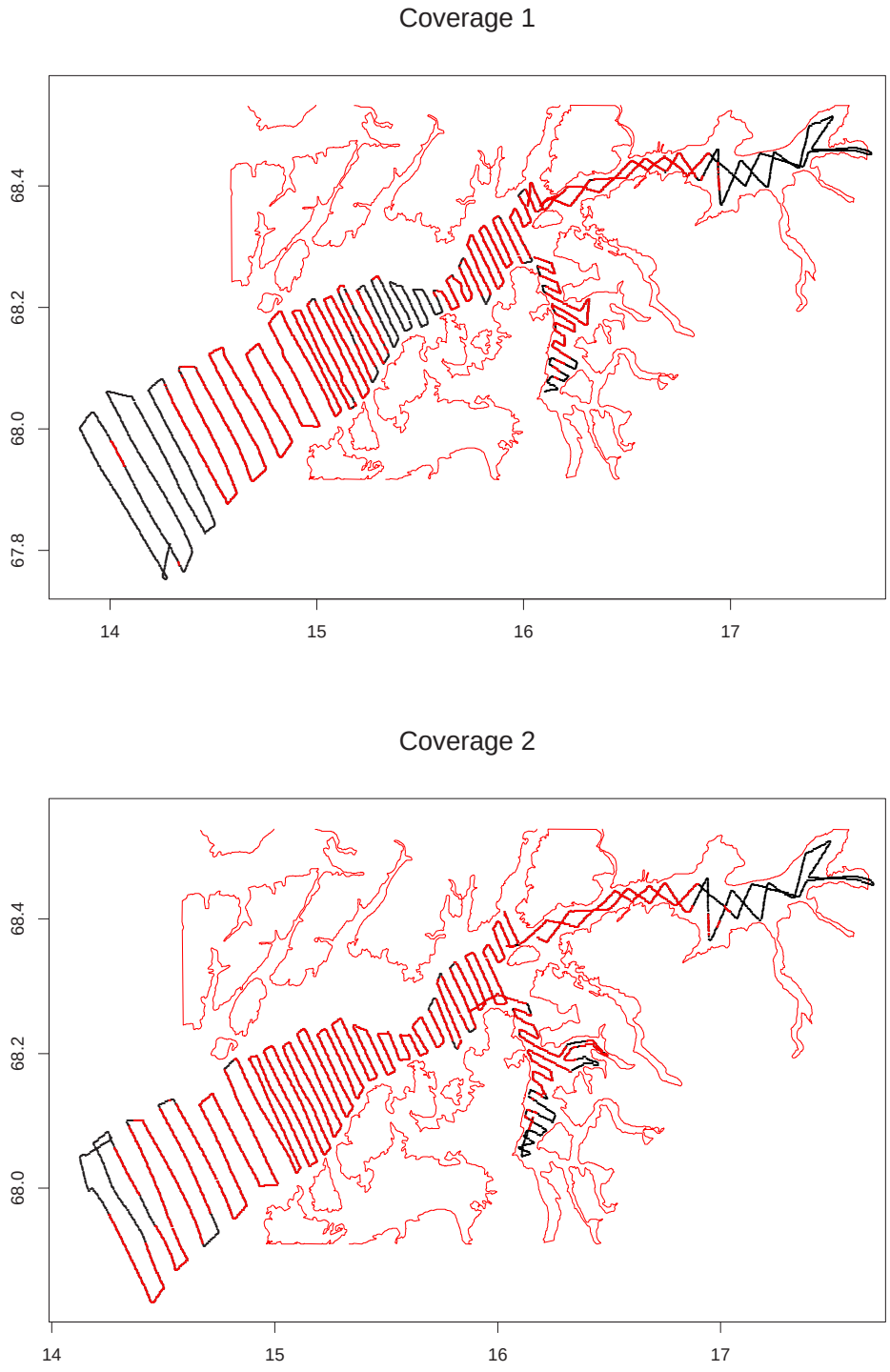



Figure 15: Map of zero (black) and non-zero (red) observations.

# 3 Empirical Variograms

In Figures 16-23 we have plotted classical variogram estimates (Cressie, 1993).

$$2\hat{\gamma}(\boldsymbol{h}_s, h_t) = \frac{1}{|N(\boldsymbol{h}_s, h_t)|} \sum_{N(\boldsymbol{h}_s, h_t)} \big(Z(\boldsymbol{s}_i, t_i) - Z(\boldsymbol{s}_j, t_j)\big)^2.$$

Here, $N(\boldsymbol{h}_s, h_t) = \{(\boldsymbol{s}_i, \boldsymbol{s}_j) : \boldsymbol{s}_i - \boldsymbol{s}_j = \boldsymbol{h}_s; (t_i, t_j) : t_i - t_j = h_s; i, j = 1, \ldots, n\}$ and $|N(\boldsymbol{h}_s, h_t)|$ is the number of distinct pairs in $N(\boldsymbol{h}_s, h_t)$. The empirical variograms are computed using both distance and time lags, and we thus get a 3-dimensional plot of the empirical variogram (with the 3 dimensions being distance lag, time lag and empirical variogram values). In the variogram plots the bins with no data points are excluded (and an image plot is given to show exactly where the observations are missing).

Since the survey ship moves from one part of the fjord to another part of the fjord in a predestined pattern, we have either a good resolution on the distance scale (like in Vestfjord 1 and 2) or a good resolution on the time scale (like in Ofotfjord), but not a simultaneously good resolution of both distance and time. The survey ship can not, for instance, be at two locations 10 Nm (Nautical miles) apart in less than 1 hour when it is cruising at 10 knots (1 Nm = 1852 m and 1 knot = 1 Nm/hour). Due to this fact the (time and distance) scale varies for the different parts of the fjord. For Tysfjord and Ofotfjord especially the fjord is rather narrow, so the distances beyond a certain range (e.g. 5 Nm) are *along* the fjord.

In Figure 16 and 17 we have plotted empirical variograms for roughly the same area in Vestfjord. We see that in neither of them we have coverage of the same location several hours apart. Holding the time fixed, we see that the empirical variogram for Vestfjord 2 is much steeper than the one for Vestfjord 1 and the other way around; holding the distance fixed. This may be due to the spatial trend in the data from the selected region of Vestfjord 2. In Figure 18 we have combined the data from Figures 16 and 17. We see, however, that the time gap is very large (roughly 6 days).

Figures 19 and 20 show similar analysis for the upper (or North-East) part of Vestfjord. The results are roughly similar, but again we see that the missing combinations of time and distance observations and spatial trends in the data influence on the results.

Figure 21 shows plotted empirical variogram for the data from Ofotfjord 1 & 2. Those two surveys are only separated by a two hours break (see Table 1), so they are more or less one survey. The sampling design here is to cruise back and forth on a rather small area, and thus we get good coverage of the time domain. We have used data located west of the 17th Longitude, since the other measurements were all from sites with zero measurements (as we see from Figures 10, 12 and 15). The variogram is increasing with distance , while it is more or less constant with time.

Variogram for the Tysfjord 1 & 2 data (Figure 22) is presented in Figure 23. Since we want to avoid to use of non-water distances: distances between two sites *not* traversed entirely over water. We therefore divided the dataset into a Northern and Southern dataset. For the variogram plot, we added 1000 hours to the times in the Northern dataset to avoid the use of non-water distances. Figure 23 shows the variogram using the data from Tysfjord. The maximum time difference being used is only

4 hours. There is no simultaneous coverage of for instance 3 hours and 2 Nm and 1 hour and 8 Nm. The variogram is increasing with increasing distance, while it is more or less constant with increasing time.

Finally, we have plotted a variogram using data from all the other plots in Figure 24. The general trend is again an increasing variogram with distance, but a more diffuse (if any?) trend for the time.

By combining the results from the good coverage of the time domain in Ofotfjord and the good coverage of the distance domain in Vestfjord, we think we can approximate/estimate the empirical variogram function rather good. Beyond the range of the data (about 20 hours and 15 Nm), we have to say that the data are uncorrelated. And we do not have measurements to estimate diurnal effects.

## 3.1 Empirical Covariograms

In Figures 25 and 26 we have plotted classical covariogram estimates (Cressie, 1993).

$$\hat{C}(\boldsymbol{h}_s, h_t) = \frac{1}{|N(\boldsymbol{h}_s, h_t)|} \sum_{N(\boldsymbol{h}_s, h_t)} \big(Z(\boldsymbol{s}_i) - \bar{Z}\big)\big(Z(\boldsymbol{s}_j) - \bar{Z}\big),$$

where $\bar{Z} = \sum_{i=1}^{n} Z(\boldsymbol{s}_i)/n$, $N(\boldsymbol{h}_s, h_t) = \{(\boldsymbol{s}_i, \boldsymbol{s}_j) : \boldsymbol{s}_i - \boldsymbol{s}_j = \boldsymbol{h}_s; (t_i, t_j) : t_i - t_j = h_s; i, j = 1, \ldots, n\}$ and $|N(\boldsymbol{h}_s, h_t)|$ is the number of distinct pairs in $N(\boldsymbol{h}_s, h_t)$.

In Figure 25 we have plotted covariogram for Ofotfjord 1 & 2 (the same data as in Figure 21) and in Figure 26 we have plotted covariogram for all the data (the same data as in Figure 24). Generally we see that in the time domain, the covariance is more or less constant, while it is decreasing with increasing distance. This might be what we expect: If the fish moves around in the fjord very slowly, maybe we would have discovered that it has moved in a week or so, but not in 10-15 hours. The falling covariance with distance might be either due to a spatial trend of some sort or that the random field is non-stationary.

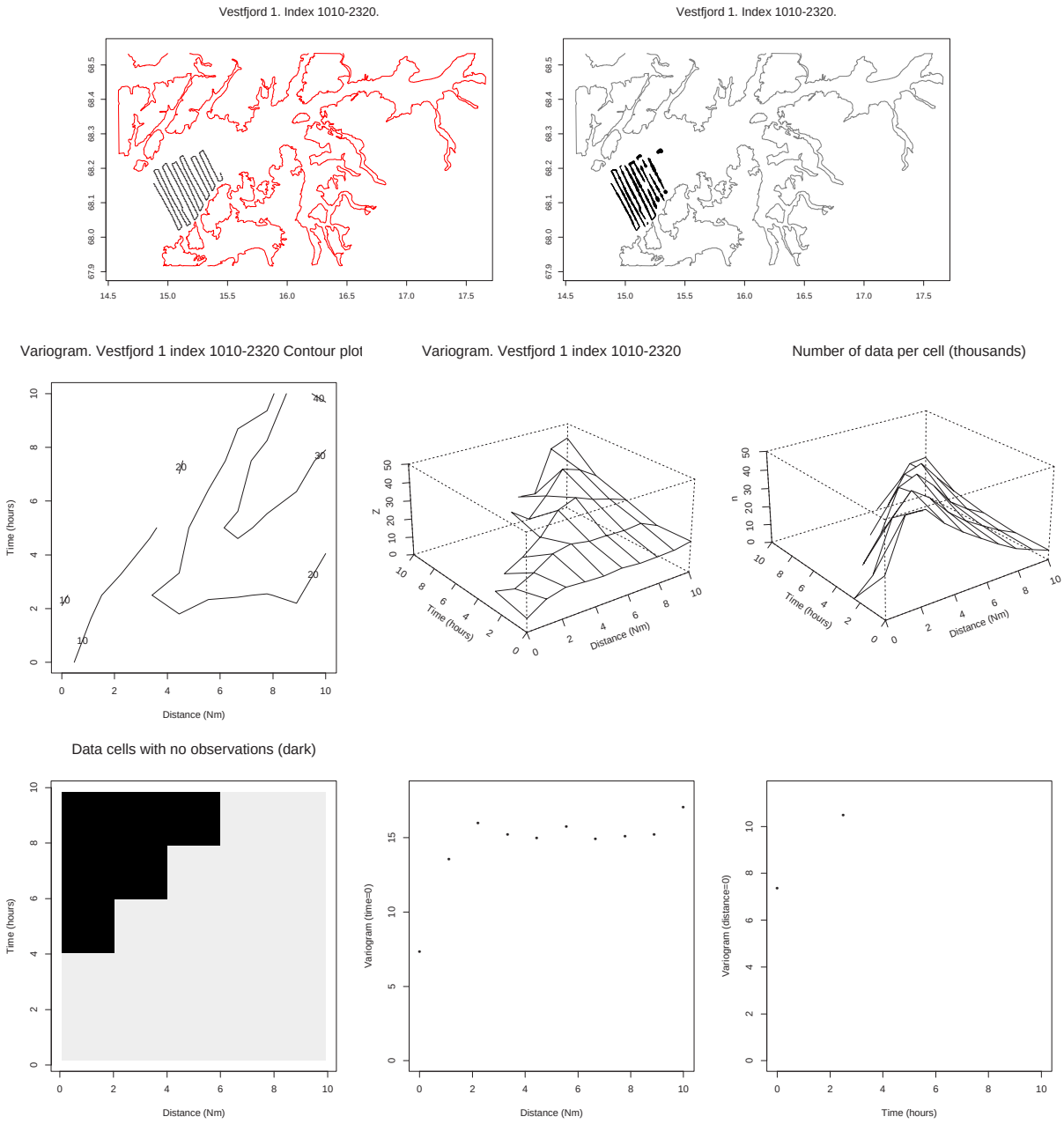



Figure 16: Vestfjord 1. Index 1010-2320. Transect (upper left), data (upper right), variogram (middle left and middle centre), number of data points used in the variogram (middle right), data points used in variogram with no observations (lower left), variogram with time=0 (lower middle) and variogram with distance=0 (lower right).

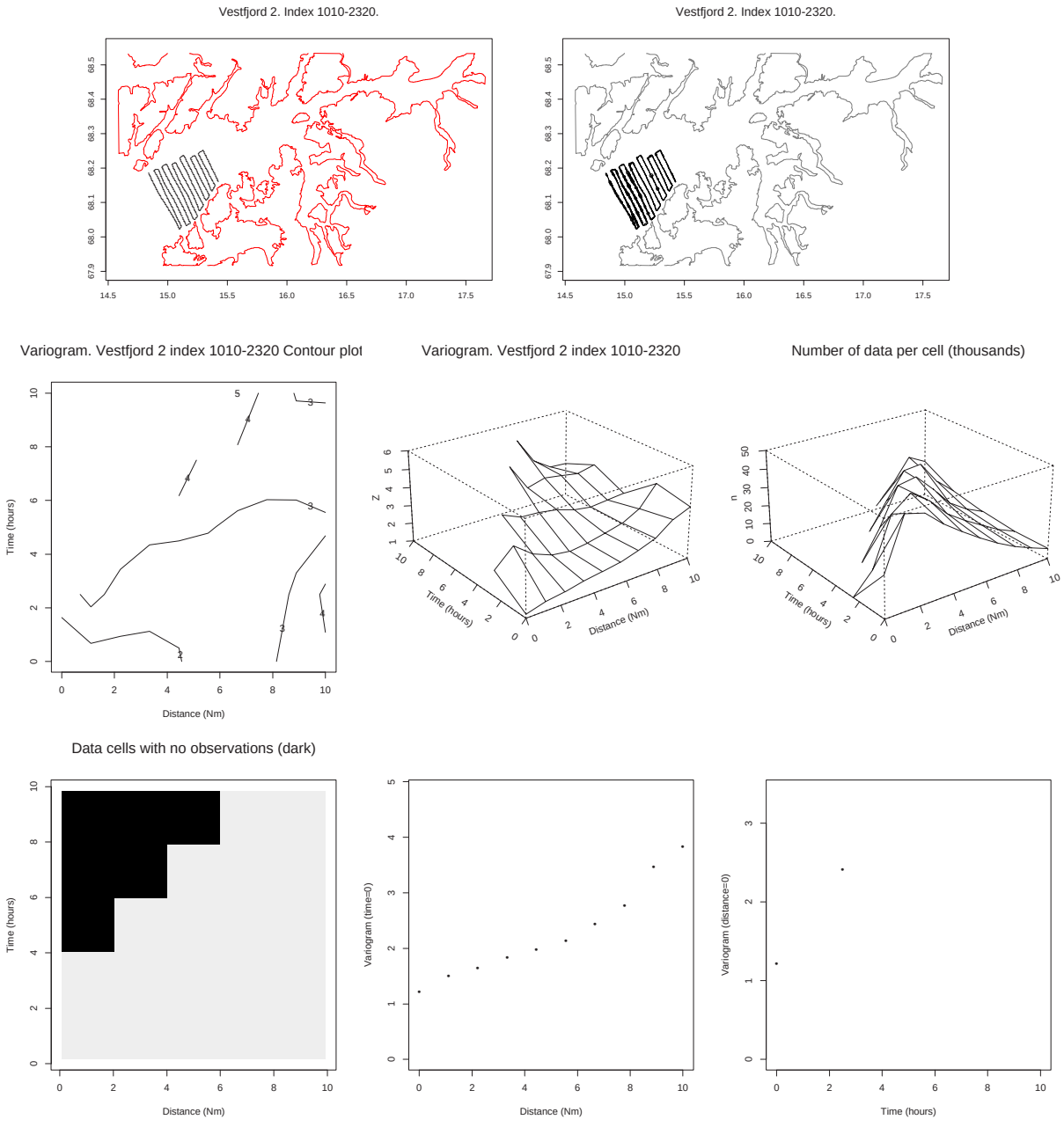



Figure 17: Vestfjord 2. Index 1010-2320. Transect (upper left), data (upper right), variogram (middle left and middle centre), number of data points used in the variogram (middle right), data points used in variogram with no observations (lower left), variogram with time=0 (lower middle) and variogram with distance=0 (lower right).

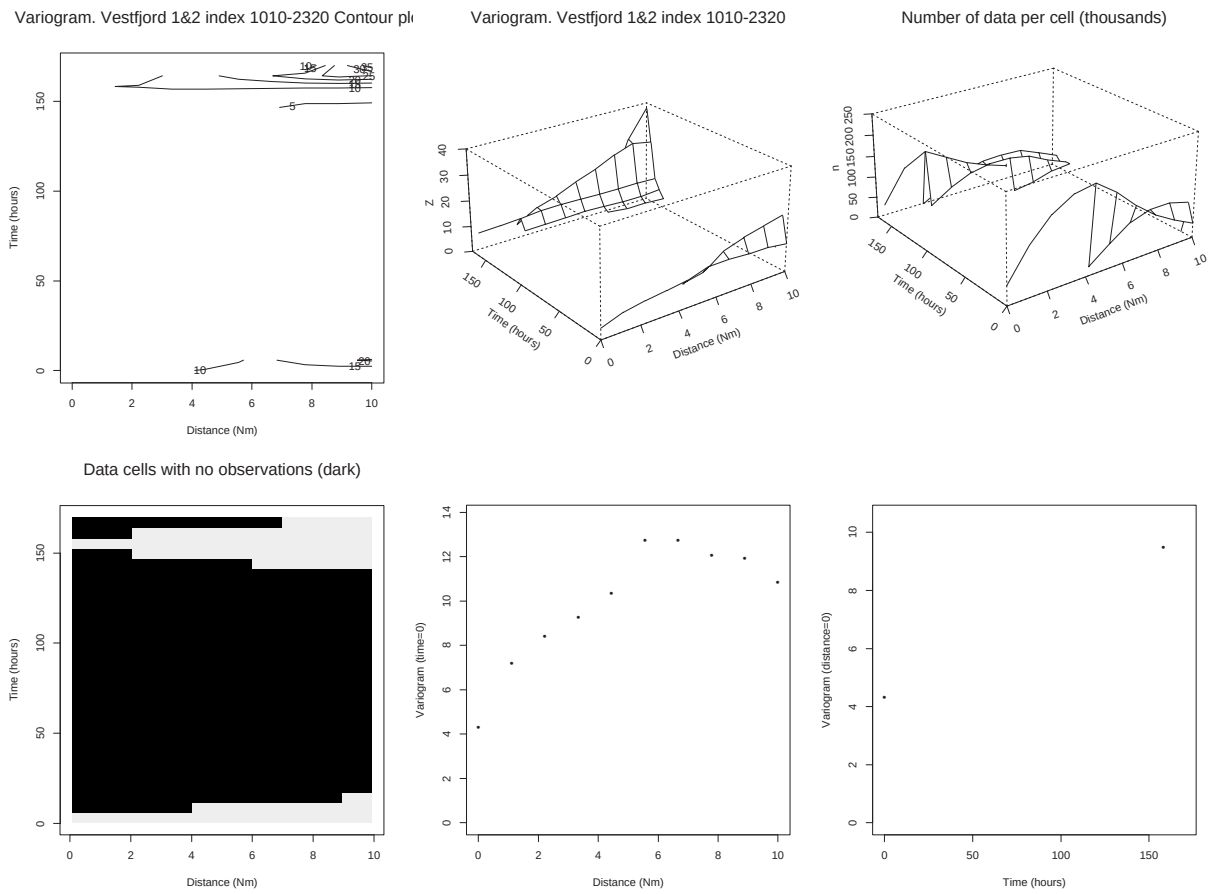



Figure 18: Vestfjord 1 & 2 (a combination of the data in Figures 16 and 17). Index 1010-2320. Variogram (middle left and middle centre), number of data points used in the variogram (middle right), data points used in variogram with no observations (lower left), variogram with time=0 (lower middle) and variogram with distance=0 (lower right).

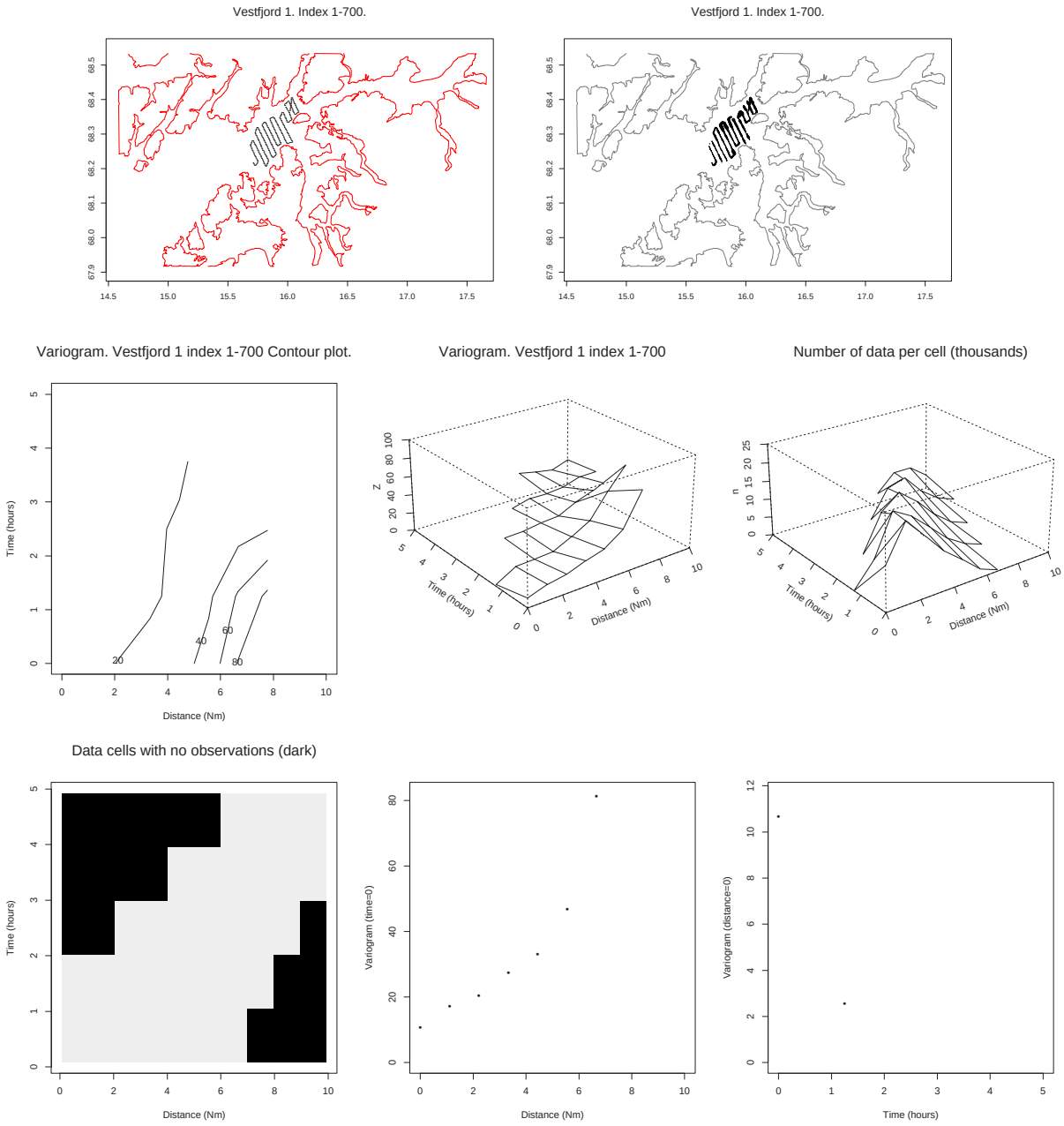



Figure 19: Vestfjord 1. Index 1-700. Transect (upper left), data (upper right), variogram (middle left and middle centre), number of data points used in the variogram (middle right), data points used in variogram with no observations (lower left), variogram with time=0 (lower middle) and variogram with distance=0 (lower right).

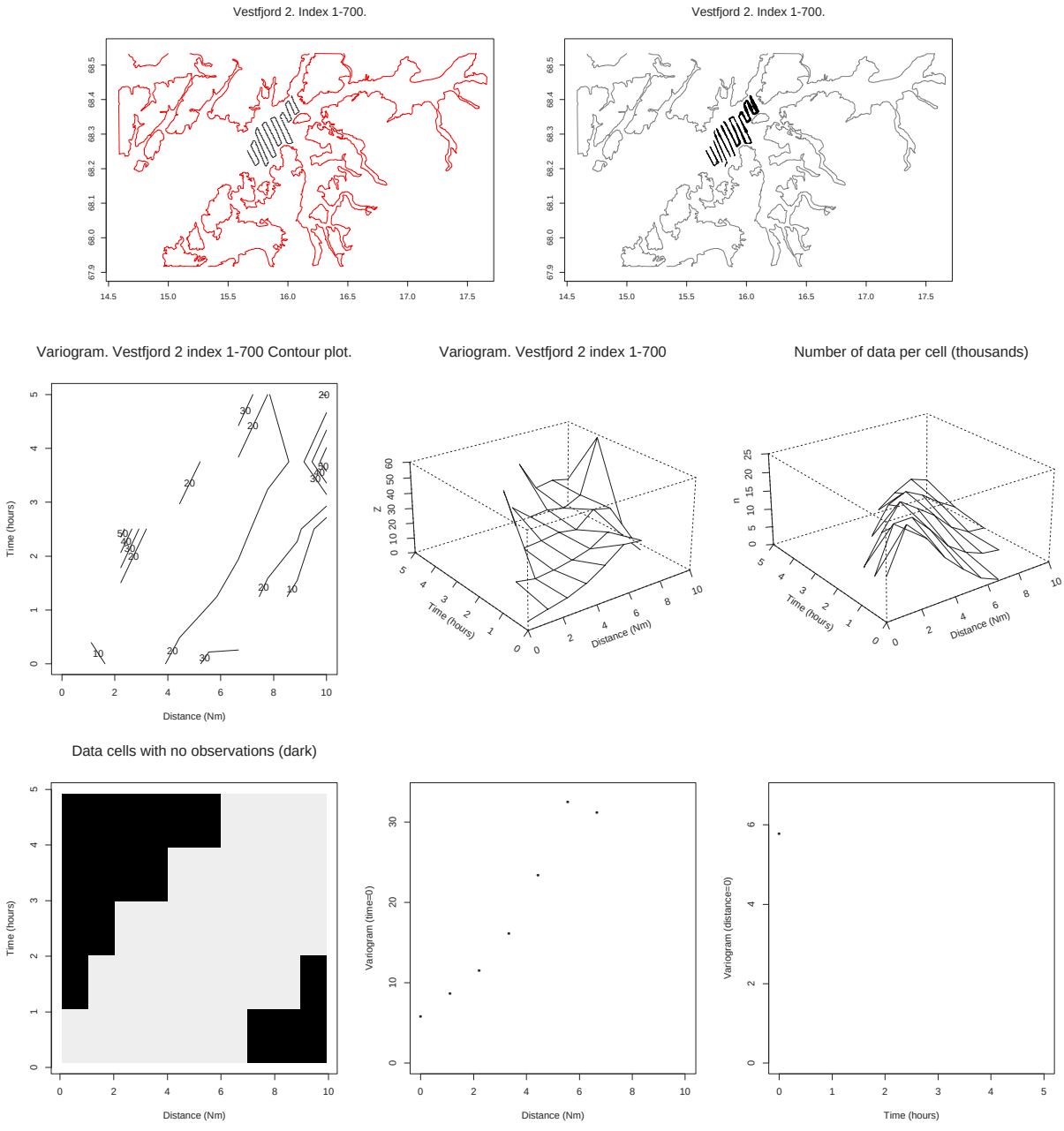



Figure 20: Vestfjord 2. Index 1-700. Transect (upper left), data (upper right), variogram (middle left and middle centre), number of data points used in the variogram (middle right), data points used in variogram with no observations (lower left), variogram with time=0 (lower middle) and variogram with distance=0 (lower right).

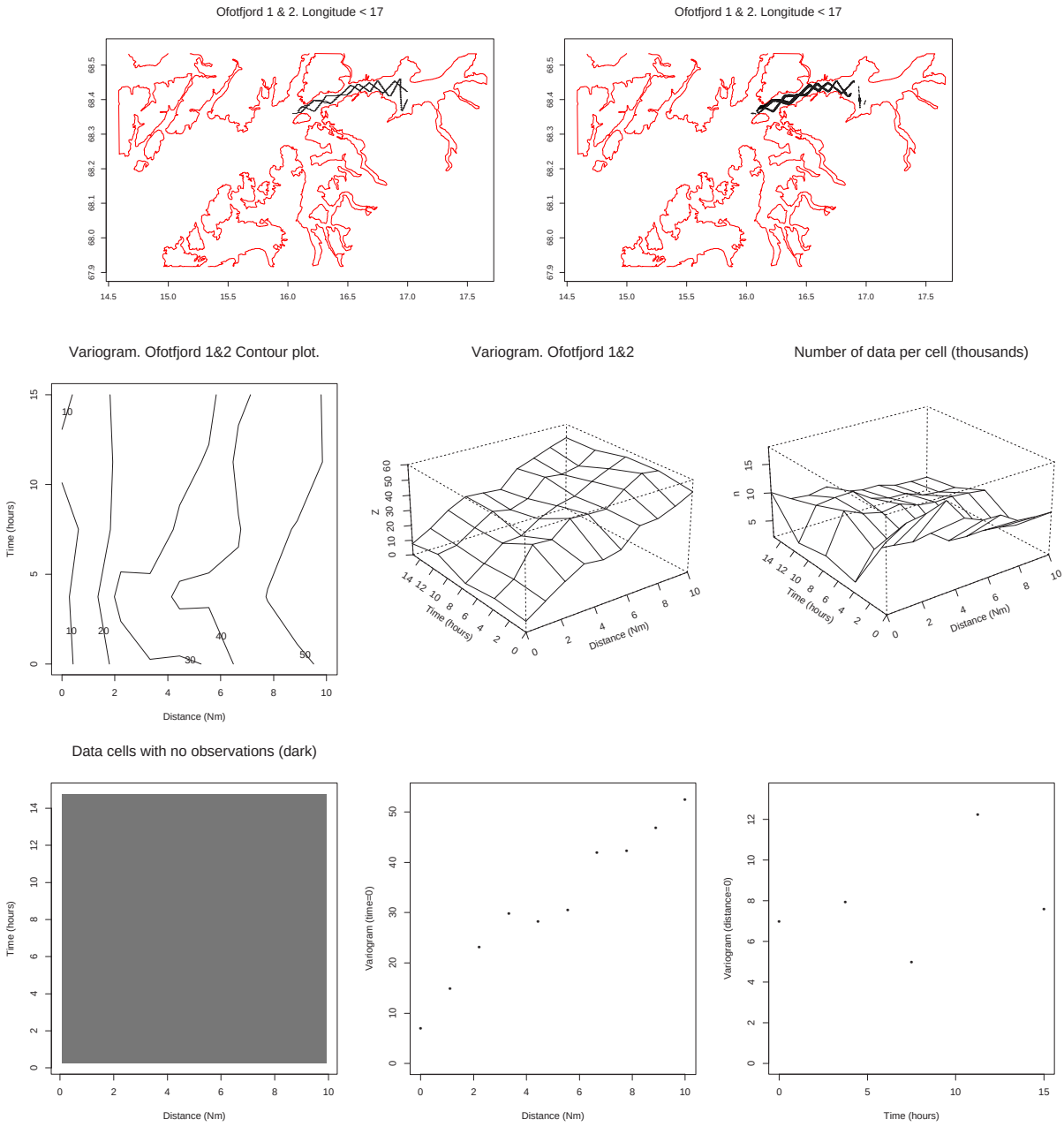



Figure 21: Ofotfjord 1 & 2. Transect (upper left), data (upper right), variogram (middle left and middle centre), number of data points used in the variogram (middle right), data points used in variogram with no observations (lower left), variogram with time=0 (lower middle) and variogram with distance=0 (lower right).

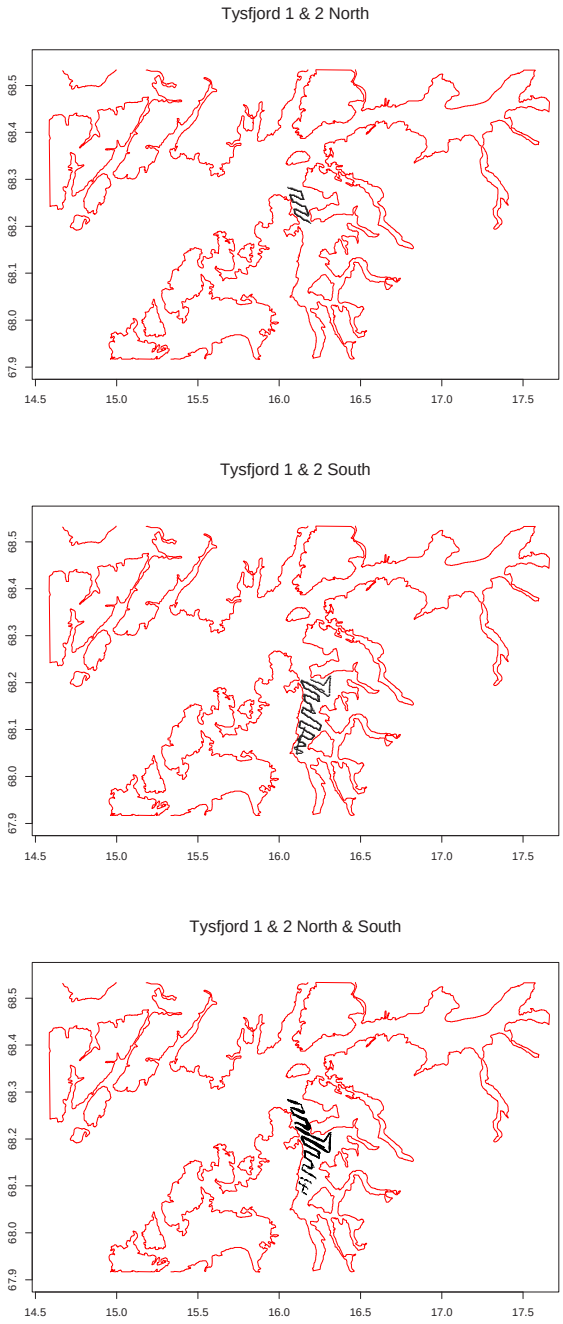



Figure 22: Tysfjord 1 & 2. Transect (upper and middle) and data (lower).

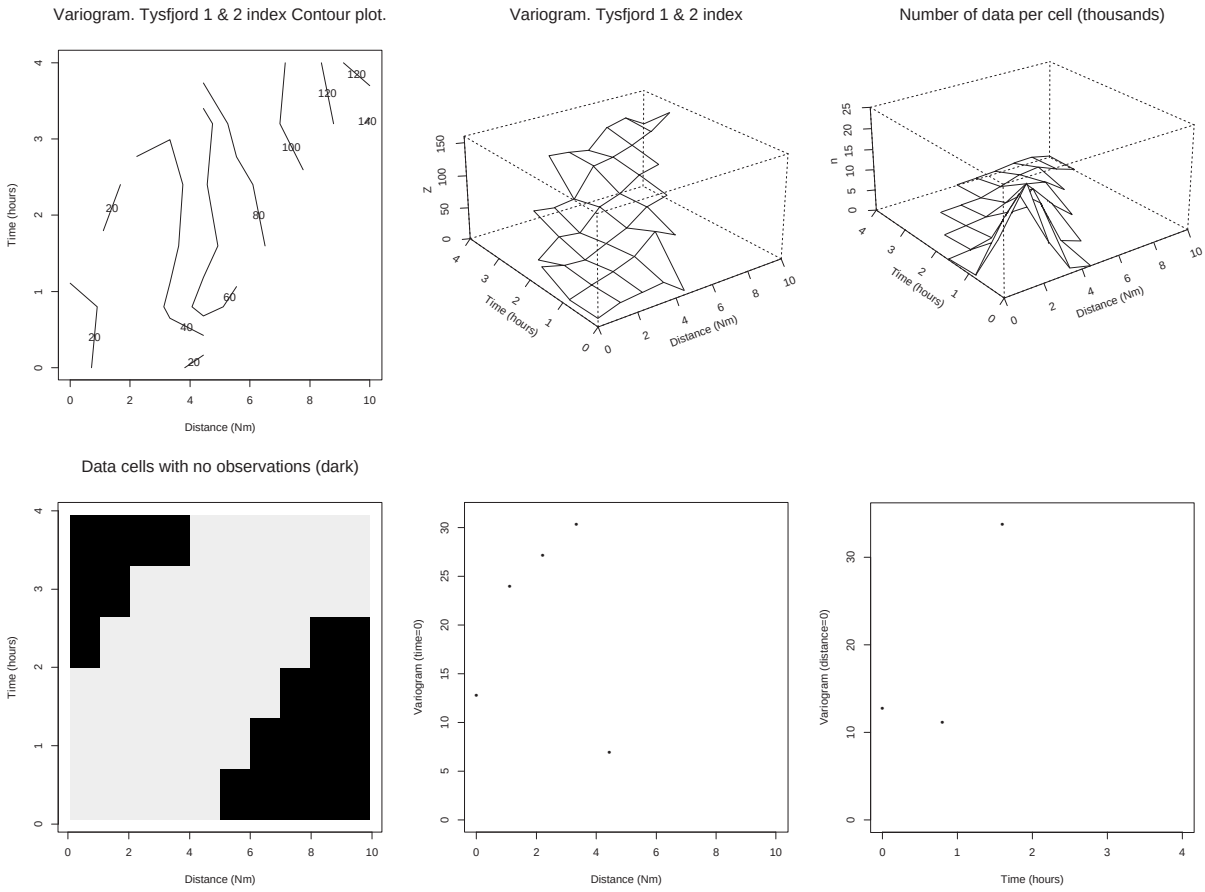



Figure 23: Tysfjord 1 & 2. Data from Figure 22. Variogram (upper left and upper centre), number of data points used in the variogram (upper right), data points used in variogram with no observations (lower left), variogram with time=0 (lower middle) and variogram with distance=0 (lower right).

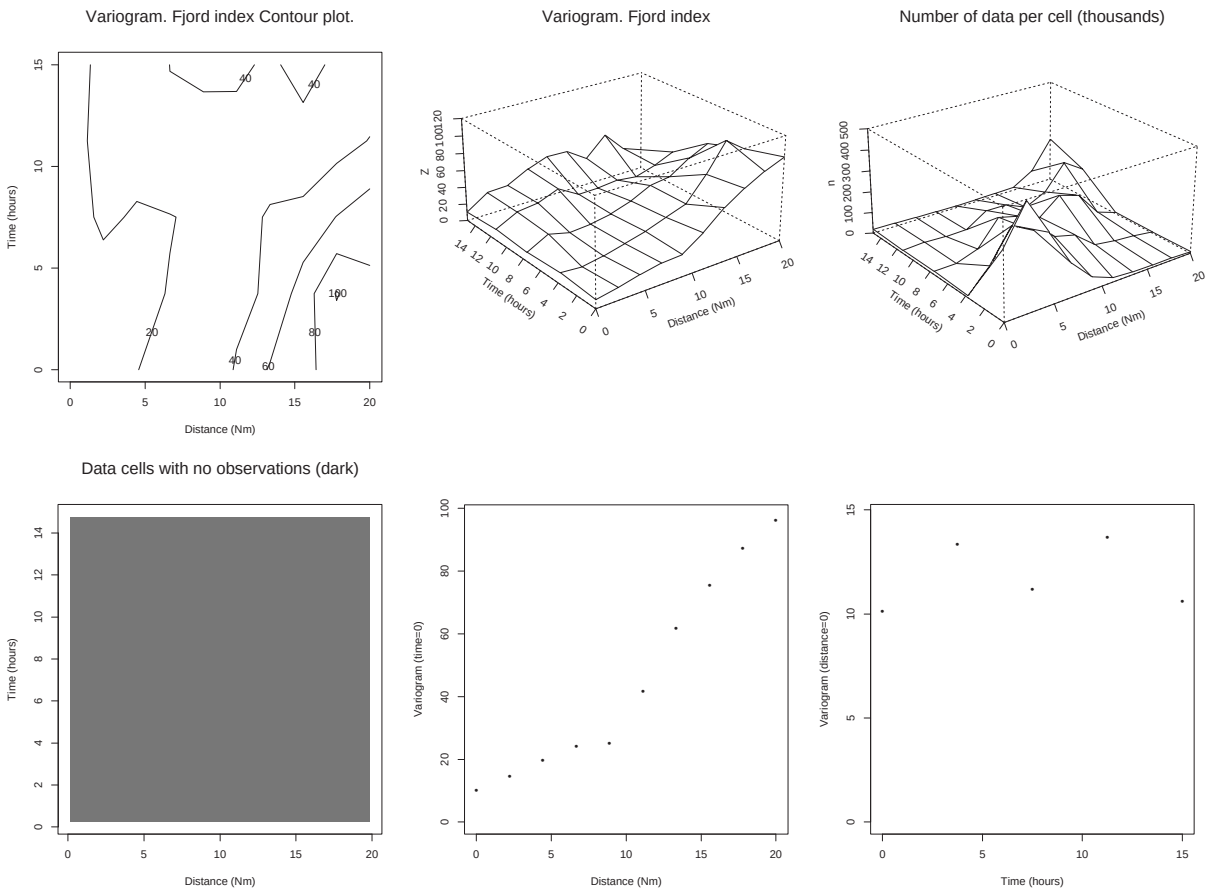



Figure 24: Data used in all the other plots. Variogram (upper left and upper centre), number of data points used in the variogram (upper right), data points used in variogram with no observations (lower left), variogram with time=0 (lower middle) and variogram with distance=0 (lower right).

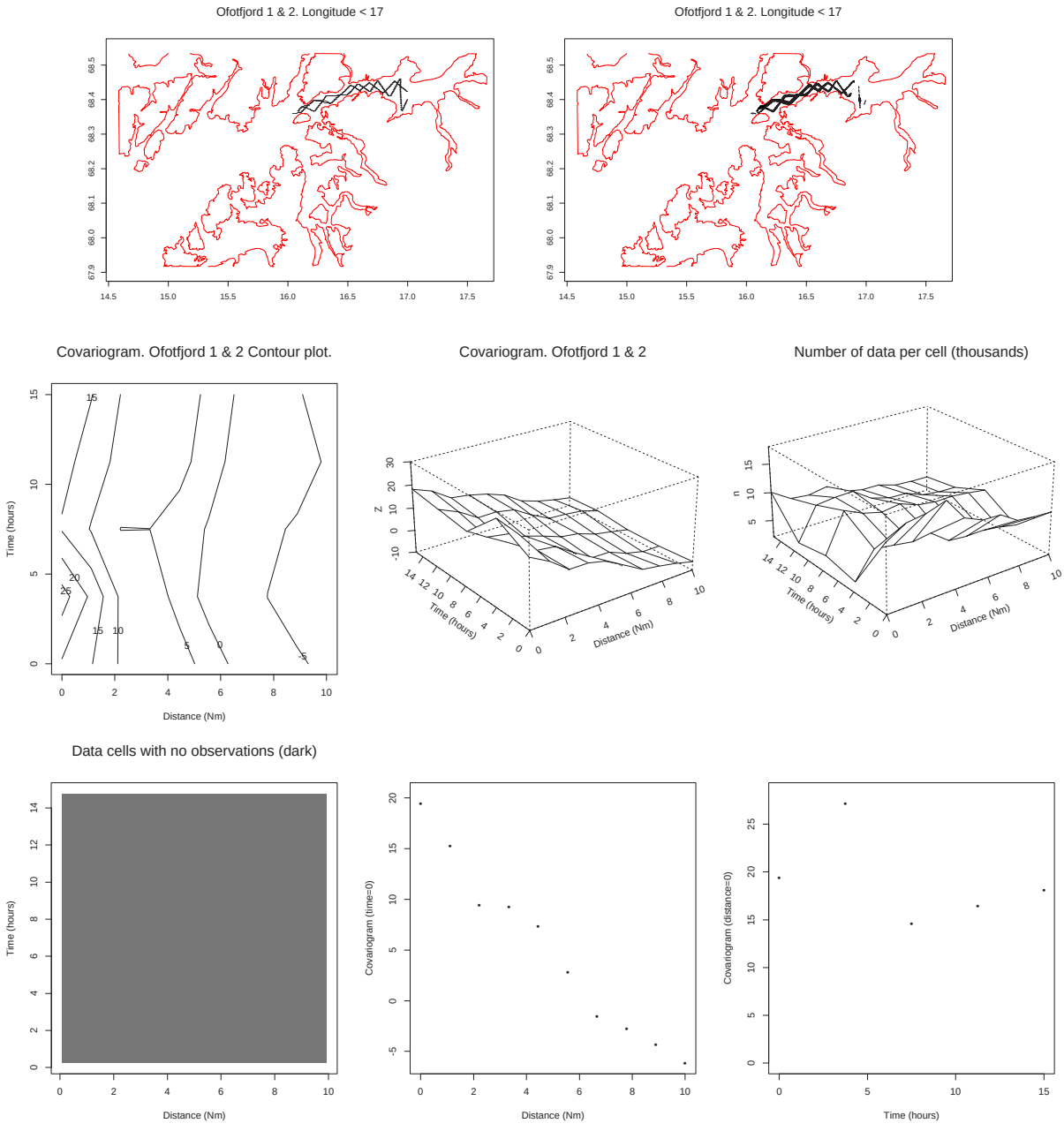



Figure 25: Ofotfjord 1 & 2. Transect (upper left), data (upper right), covariogram (middle left and middle centre), number of data points used in the covariogram (middle right), data points used in covariogram with no observations (lower left), covariogram with time=0 (lower middle) and covariogram with distance=0 (lower right).

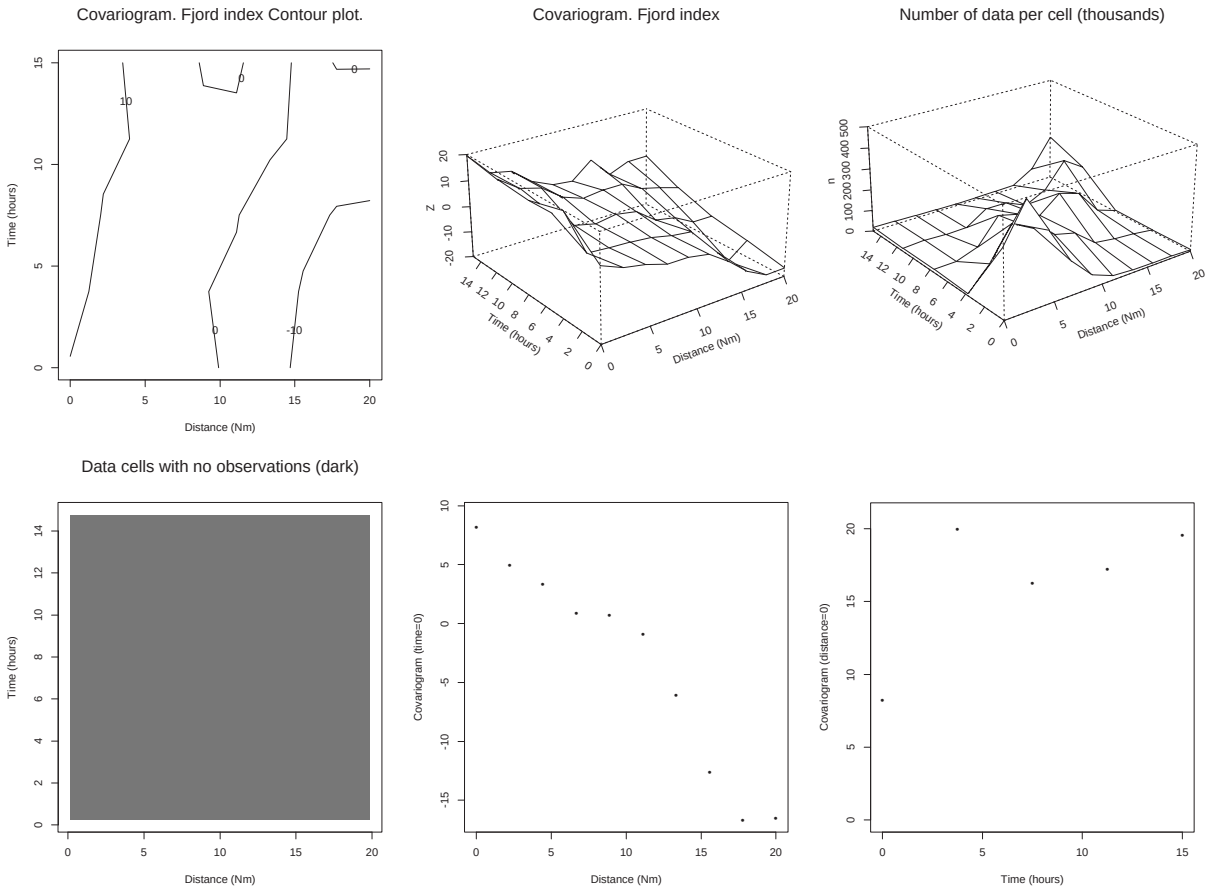



Figure 26: Data used in all the other plots. Covariogram (upper left and upper centre), number of data points used in the covariogram (upper right), data points used in covariogram with no observations (lower left), covariogram with time=0 (lower middle) and covariogram with distance=0 (lower right).

# 4 Future work

We have several problems to work with:

1. In the observations where we measure 0 fish special care has to be taken to model the abundance. Maybe we can model the probability of measuring 0 in a particular location/time? Then, given a measurement greater than zero, model the amount.

   How do we deal with migration between the areas?

2. What covariance or kernel function are we going to choose for space-time?

3. Distance metric. In the Figures we have assumed that the geographic distance is appropriate, but we assume the water distance is a more appropriate measure of distance. (Water distance is the shortest path between those two sites that may be traversed entirely over water.)

   We may assume that data from sites far apart are uncorrelated. For sites close to each other the water distance equals the geographic distance. So the problem is where the distances between locations are of medium length, e.g. 5 Nm.

# References

Cressie, N. A. C. (1993). *Statistics for Spatial Data*. Wiley-Interscience, revised edition.

Ona, E. (2001). About avoidance effect and target strength effect. Personal comment.

Zhao, X. and Ona, E. (2002). Models for estimation of the $\sigma_e/\sigma$ ratio of fish: Compensation for the shadowing effect in dense fish aggregations. *Unpublished*. Yellow Sea Fisheries Research Institute, 106 Nanjing Road, Qingdao 266071, P. R. China.