Microbiology and food safety in Limfjorden mussels from 1996 to 2013

Kadri Kuusemäe, Paula Canal and Jens Kjerulf Petersen

Danish Shellfish Centre, DTU Aqua

Aim of the report

The aim of this analysis is to examine possible trends in results for food safety sampling in relation to microorganisms. Study focuses mainly on the distribution of *E. Coli* showing the distribution in Limfjorden over the years and season. The distribution is shown by measurement points and risk maps.

Methods

Work done in excel:

Data was received from Fødevarestyrelsen as excel spreadsheets that had send the data to Foreningen MuslingeErhvervet. The spreadsheets were examined for inconsistencies and requirements for simple manipulations:

- GPS coordinates were converted into decimal degrees when needed. Unneeded and repetitive columns were deleted. Only columns containing important information to map manipulations and statistical analysis were left (Zone nr, Mussel farm nr, Ship nr, date of sample taken, Species, Lab name, salmonella and *E. Coli*)
- Salmonella and *E. Coli* were converted to positive (1) and negative (0) values using IF function. Salmonella was considered negative when the sample wasn't analysed (it) or tested negative (ip). Sample was considered positive if marked p or pv. An unknown abbreviation "s" was found and this sample was left out.
- An *E. Coli* the sample was considered positive if the value was over 230. If three samples were collected in the same place, the point/sample was considered positive if one of the samples was positive.
- A combined column was made out of mussel farm numbers and indication of samples taken in fishery. To ease the statistical analysis the data was reduced to binary data using the IF function in excel. The mussel farms were given the number 1, fishery samples were given the number 0 and if the origin was unknown the samples were given N/A.
- Extra cells were made for month and year, for keeping better track of data in MapInfo and to check seasonal variations. After manipulations the exact date of the sample was discarded and for both statistical and map manipulations the year and month were used to look at seasonality and variations in between years.

- Some renaming was done to ensure that the statistical program R would treat the data set correctly. Å was replaced with AA and Ø with OO, also blåmuslinger was changed to blåmusling. Points that didn't have coordinates were ignored.

Manipulations in MapInfo:

Tables were made based on the data manipulations done in excel. The projection of the tables was chosen to be longitude/latitude. Columns were specified according to the data they were holding. As shown in Figure 1. Tables included all the necessary data needed for analysing *E. Coli* and salmonella abundance in Limfjorden. Zone number was placed as integer to not allow any decimal places, all data that consisted of words or numbers and a letter were introduced as characters, the analysis date was set ad date, rest was introduced as float, which allows as many decimal points as needed.

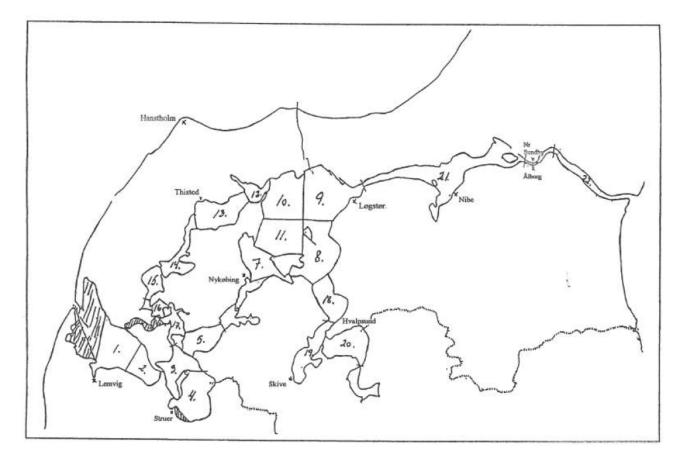**Figure 1**   **Build-up of the tables made based on excel data.**

Data from excel was copy-pasted into the newly made tables. Tables were made for each year separately based on the same model table. These base tables also included the exact date of the taken sample, exact *E. Coli* measurements. These tables (for each year) were combined into a main table containing all the data from all the years. Points were created to visualize the distribution of samples. To limit the study area a SQL query was made using two tables: the newly made table and the fishery area table a new table was made containing only samples from Limfjorden.

Year	Number of samples taken in Limfjorden	Year	Number of samples taken in Limfjorden
1996	60	2006	336
1997	29	2007	211
1998	2	2008	467
2000	37	2009	334
2001	40	2010	255
2002	74	2011	359
2003	85	2012	357
2004	101	2013	109
2005	208	Total number of samples taken over the years	3064

Tables were made for all the separate years and one combined table containing all the years (made by appending rows from all the tables into a new table). In the separate year layers, positive *E. Coli* results were highlighted with red. The combined table was copied to excel to further manipulate the table. As explained before, the exact date of samples was discarded, the measurements of *E. Coli* was removed and only the binary response was left, also a new combined column called type was made. This new table was copied back to a new MapInfo table. This step was done to bring map manipulations closer to statistical analysis.

The next two tables were compared to confirm the point's locations. In 2004 the fishery numbering system was changed. Due to that some fishery zones had multiple zone numbers within them. The old and new fishery areas can be seen in Figure 2 and Figure 3.
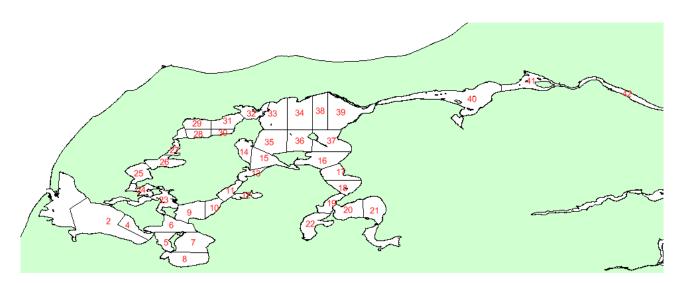
Figure 2   **Fishery zones until 2004**

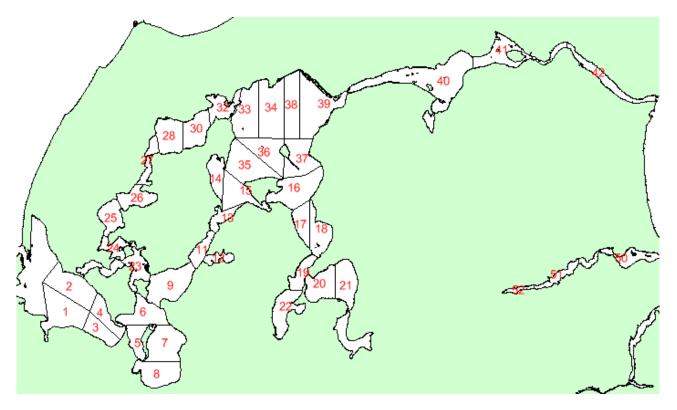**Figure 3**   **Fishery zones after 2004**

Figure 4   Map of fishery zones after 2011

Point manipulations were done after comparing the fishery zones before and after the name change. When the points didn't fit either new or old numbering the year and neighbouring zones were compared and the points were moved to the correct zone, if the location was unproven the points were deleted. These can be seen in Figure 4 and Table 2.

Table 2   Deleted points or the changed zone number after comparing coordinates with mussel farm location table.

Zone nr	Type	Ship nr	y coordinate	x coordinate	Month	Year	Species	Lab	Salmonella	E. Coli	Proven location
19	1	N/A	56.8702	9.14067	Nov	2010	Blåmusling	Aalborg	0	1	
35	0	SK925	56.8727	8.63845	August	2011	Blåmusling	Aalborg	0	0	
35	0	SK925	56.8727	8.63845	August	2011	Blåmusling	Aalborg	0	0	
35	0	SK925	56.8727	8.63845	August	2011	Blåmusling	Aalborg	0	0	
35	0	T229	56.8727	8.63845	July	2011	Blåmusling	Aalborg	0	0	
35	0	N/A	56.8727	8.63845	July	2011	Blåmusling	Aalborg	0	0	
13	1	N/A	56.7683	8.49167	July	2009	Blåmusling	Aalborg	0	0	
9	0	T194	56.6767	9.155	Aug	2008	Blåmusling	Aalborg	0	0	
36	1	N/A	56.6741	9.125	May	2013	Blåmusling	Eurofins	0	0	zone 19
19	0	L919	56.8167	8.93333	Sep	1996	Blåmusling	Højmark	0	0	
20	0	L919	56.8167	8.93333	Sep	1996	Blåmusling	Højmark	0	0	
19	0	L919	56.8	8.91667	Nov	1996	Blåmusling	Højmark	0	0	
13	0	L919	56.8167	8.93333	Sep	1996	Blåmusling	Højmark	0	0	
8	0	L919	56.8	8.91667	Nov	1996	Blåmusling	Højmark	0	0	

| 8	0	L919	56.8167	8.93333	Sep	1996	Blåmusling	Højmark	0	0	
9	1	N/A	56.83	8.85433	Aug	2010	Blåmusling	Aalborg	0	0	
13	1	N/A	56.6975	8.85417	July	2009	Blåmusling	Aalborg	0	0	
13	1	N/A	56.6975	8.85417	Aug	2010	Blåmusling	Aalborg	0	0	
13	1	N/A	56.6975	8.85417	Sep	2010	Blåmusling	Aalborg	0	0	
13	1	N/A	56.6975	8.85417	May	2012	Østers	Eurofins	0	0	zone 12
13	1	N/A	56.6975	8.85417	June	2012	Blåmusling	Eurofins	0	0	
13	1	N/A	56.6975	8.85417	Aug	2012	Blåmusling	Eurofins	0	0	
13	1	N/A	56.6975	8.85417	Sep	2012	Blåmusling	Eurofins	0	0	
9	0	T194	56.4881	8.69053	March	2011	Blåmusling	Eurofins	0	0	
9	1	N/A	56.5461	8.64108	Feb	2008	Blåmusling	Aalborg	0	0	
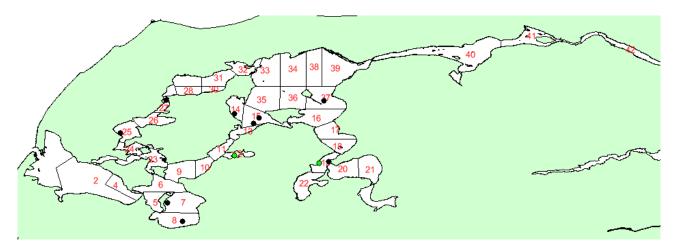
Figure 5 **Locations of deleted points shown with black dots and green dots show the points with changed zone number. Red numbers indicate the current zone numbers.**

After confirming the locations of the points a new table was prepared and queries were made to find all the points in each zone one by one. The results of the queries were copied to excel and the zone numbers were manually changed to the current zone numbers and copied back to the new table.

When the new master table was completed queries were made to make all the necessary tables for data analyse. Tables were made for each year, months and mussel farms. Based on these maps, risk maps were made.

Risk maps were made with vertical mapper using natural neighbouring interpolation. For this map manipulation the separate year tables were used that contained both negative and positive values of *E. Coli* results. In some cases dummy points were created in between two positive values when a land mass was seen in between them to improve the map. A combined risk map based on all the tests over the years were done using the built in thematic map creator due to vertical mapper not creating natural neighbouring interpolation with bigger point data table than around 2600 data points. With salmonella only an overall risk map was created using the built in thematic map function.

All risk maps were adjusted for their colour ranges. For maps done in vertical mapper: Instead the given colour distribution all values based on the natural neighbouring calculations that were under 0.5 (0 to 0.49) were considered as zero based on the common rounding method and given the same colour, all above 0.5,

it included were considered positive. For the two maps done with the built in thematic map function, similar manipulation was done, but in this case the colour ranges were limited to 3: one for 0%, 50% and 100%.

A map of relative *E. Coli* positive samples to total number of samples taken in each fishery zone was done by modifying the fishery table by adding a centroid x and y coordinate to make a query to get information about the total number of samples with in each fishery zone, the number of positive E. Coli results within the zone and a centroid location for each zone. Based on the query, a table was made and points created. When needed points were moved to the centre of the fishery manually, because the centroid position within MapInfo depends on the shape of the area and can be located outside of the area of interest. Relative number of positives was calculated by dividing the total number of positive results with total number of samples taken in the fishery zone. That was done for each zone by using update column function. Based on the table an interpolation map was created to give visual illustrations.

Statistical analysis:

For statistical analysis the combining table from MapInfo containing only the points in Limfjorden was copy-pasted to a new Excel file.

Statistical analysis was done with two programs. Simpler analysis was done in excel and the main significance test was conducted in R.

Histograms of the data covering all the years were made by two methods. First method included query made in MapInfo asking for the count (frequency) of wanted variable and grouping it by *E. Coli* and the variable itself, results were copied into Excel and made into bar graphs after reordering the data. Second method used the Data analysis function in Excel that made histograms automatically. Excel was also used to calculate simple percentages to illustrate the dataset given.

Before analysing data in R, the data had to be saved in a single sheet Excel file as Macintosh CSV file, which is easy for R to read. In R logistical linear regression analysis was done to analyse binary response data.

Data was read into R with the command:

mussel1 <- read.csv("new_e.coli_stat3.csv", header=TRUE, sep=";", na.strings="N/A", dec=".", strip.white=TRUE)

Different specifications were given to introduce data correctly and later checked with the function *head.* AS R interprets all numerical data as numbers special conditions were made so R would read Zone numbers, Years and Type as categories. General linear model was conducted:

EColi1 <- glm(EColi ~ Zonenr+year+month+species+Lab+salmonella+type, data=mussel1, family="binomial")

summary (EColi1)

anova(EColi1, test="Chi")

Based on the linear model an ANOVA was made to see more complete picture of the different categories effects on *E. Coli* response.

Due to data not being totally independent, a second approach was taken by using mixed effects model. For that, an extra R package was downloaded called lme4, which contains better facilities for this type of models than the built in version in R. As the response variable is in binary data *glmer* function was used. Before model was implemented data piling was needed. For the model to run and give an estimate it needs a certain amount of data in each grid cell. Species, Lab (analysing laboratory) and salmonella were discarded from the model due to having too little data points in different grid cells and with that interrupting model from working. In case of zones, some of the zones were combined into 1 with Years. To improve power of the analysis, two zones (zone 41 and 42) were left out of the statistical analysis due to too few measurements and their position far away from rest of the fishery zones.

```
, ,  = 2010-2011

        April Aug Dec Feb Jan July June March May Nov Oct Sep
2           6   0   3   2   4    0    0     7   0   6   6   6
4           2   0   1   0   0    0    0     3   0   2   2   2
5           5   2   2   1   1    1    1     4   2   3   6   2
6           4   0   3   4   4    0    6     1   6   5   4   0
7           4   0   2   5   0    0    2     4   4   1   0   0
8           4   0   6   0   0    0    2     3   5   4   0   0
9-10        1   4   1   1   0    4    3     0   6   1   2   4
11          0   0   0   0   0    0    0     0   0   0   0   0
12          0   5   0   0   0    3    1     0   2   2   0   4
13          0   1   0   2   0    2    1     2   2   2   2   5
14          1  11   0   0   0    7   11     2   3   4   4   7
15          3   0   3   3   0    1    2     3   5   4   6   2
16          3   0   0   0   0    0    3     2   2   0   0   0
17-18       2   0   2   0   1    1    4     1   3   4   7   3
19          6   2   2   0   0    3    5     2   5   8   2   2
20          1   0   0   0   0    0    1     1   2   3   2   0
21          0   0   0   0   0    0    0     0   2   0   0   0
22          8   0   1   0   0    0    6     3  10   6   5   2
25          0   0   1   0   0    0    0     0   0   1   0   0
26          6   2   3   0   0    3    8     0  10   6   6   6
27          6   3   3   0   1    4    3     8   5   5   0   0
28-29       0   0   0   0   0    0    3     0   1   0   0   0
30-31       1   2   2   0   0    0    1     1   2   3   2   2
32          0   0   0   0   0    0    0     0   1   0   0   0
33          4   0   2   0   0    0    0     4   2   6   1   0
34          0   0   0   0   0    0    0     0   0   0   0   0
35-36       7   0   5   4   0    1    5     5   5  11   1   0
37          0   0   0   0   0    0    0     0   0   0   0   0
38          0   0   0   0   0    0    0     0   0   0   0   0
39          0   0   0   0   0    0    0     0   0   1   1   0
```

Figure 6   Lack of data points in one of the year groups in respect to Zone numbers and Months.

	1996-2002	2003-2005	2006-2007	2008-2009	2010-2011	2012-2013
2	16	62	54	36	40	9
4	0	15	17	10	12	3
5	2	0	0	125	30	27
6	1	18	3	3	37	12
7	3	16	0	1	22	12
8	0	4	0	0	24	13
9-10	21	22	25	123	27	19
11	2	0	0	4	0	13
12	0	5	5	26	17	21
13	2	4	6	21	19	22
14	0	16	17	9	50	13
15	12	16	21	4	32	49
16	9	10	12	24	10	4
17-18	12	29	42	85	28	10
19	1	1	62	129	37	30
20	16	6	31	13	10	3
21	2	2	29	5	2	1
22	22	8	33	18	41	12
25	27	2	38	21	2	9
26	12	9	24	17	50	11
27	3	32	29	23	38	18
28-29	1	4	1	0	4	60
30-31	6	9	38	33	16	20
32	0	5	0	9	1	0
33	16	9	1	16	19	10
34	4	26	2	14	0	4
35-36	15	15	4	0	44	53
37	21	25	25	10	0	6
38	9	18	7	12	0	0
39	0	6	16	9	2	2

Figure 7   **Illustration of combined zone numbers and the year groups made to run mixed effects model.**

EColi2 <- glmer(EColi ~ ZONE+month+type+(1|YEAR), data=mussel1, family="binomial")

Model build up to see the effects of Zone, month and type on E. Coli and where the year is introduced as a random effect, meaning year doesn't have a fixed effect on the presence and absence of *E.Coli* in mussels.

Multiple variations of base models were conducted by removing on variable at the time, the model fit was accessed by looking at the AIC value and the models were compared with each other one by one to get significance values for each parameter, due to mixed effects model not directly providing a P value.
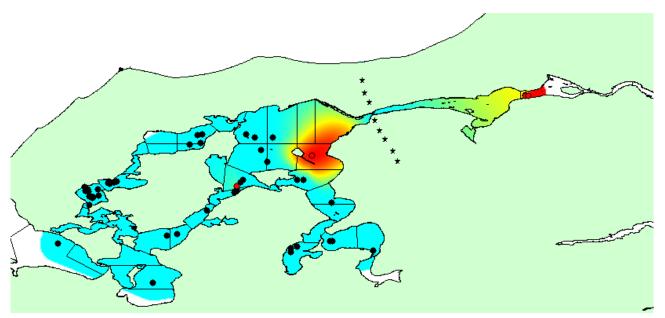
Results

MapInfo

Figure 8 Combined layer map showing the measurement points in 1996 in black and red, where red represents the positive E. Coli results. Underneath is a risk map based on the values where red represents positive result (1) and light blue negative value (0). Black stars are dummy points used to manipulate the map for better results.
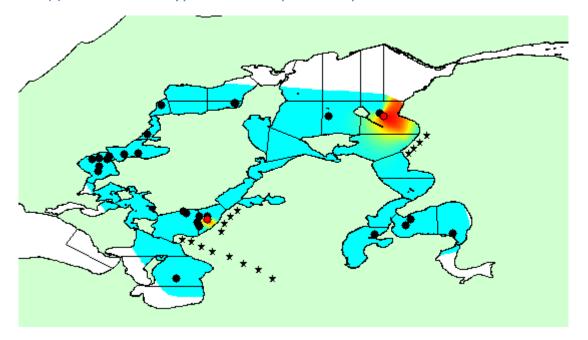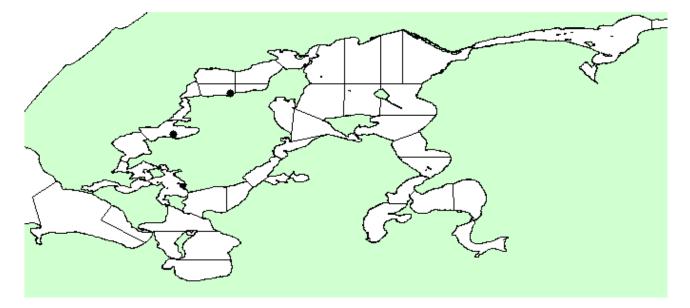

Figure 9 Combined map from 1997 with actual measurements and a risk map. Black stands for negative *E. Coli* and red for positive, stars are once again dummy points to manipulate the graph, colour code is as described in Figure 7.

Figure 10   **Map of 1998 with two samples taken in Limfjorden that are marked with black dots.**

In 1998 only two samples were provided and both of them were negative, so no interpolation map was made. No data was provided for 1999.
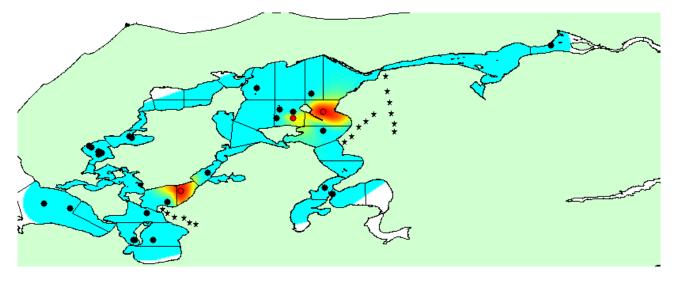
Figure 11   **Map of 2000, black dots *E. Coli* negative and red dots positive, stars represent dummy values. Underneath risk map with colours as described in Figure 7.**
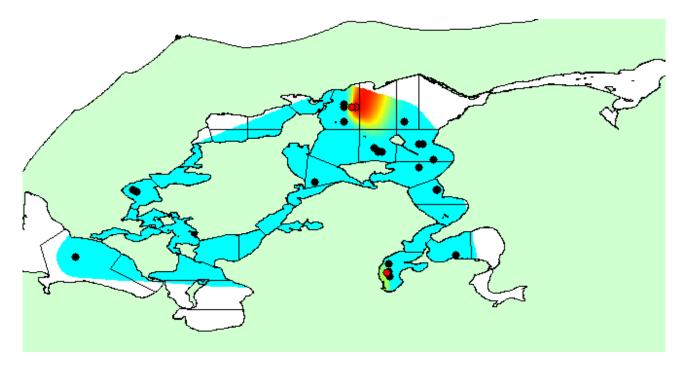
Figure 12   **Part of Limfjorden, mussel analyse results from 2001, black-** *E. Coli* **negative and red- positive. Risk map included.**
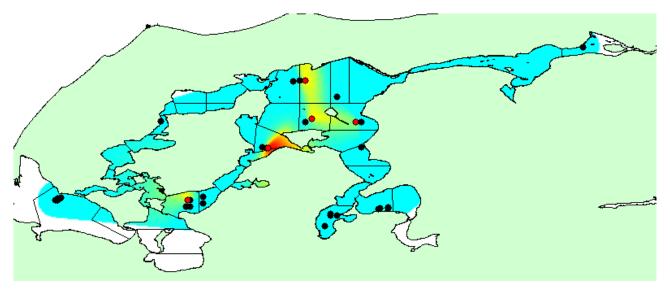
Figure 13   **Distribution of samples taken in 2002; black dots stand for negative result and red for positive. Risk map included.**
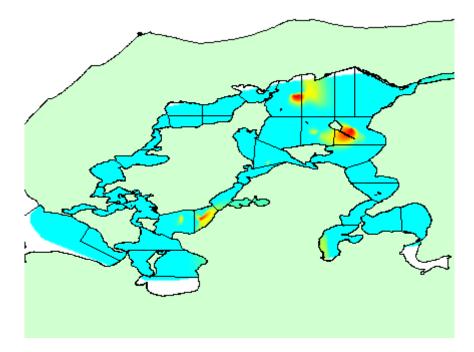
Figure 14   Risk map for combined data of samples that were taken from 1996 to 2002.
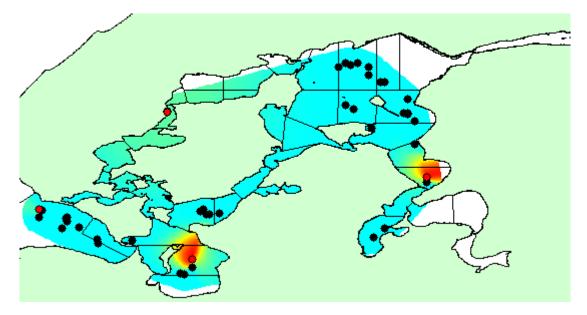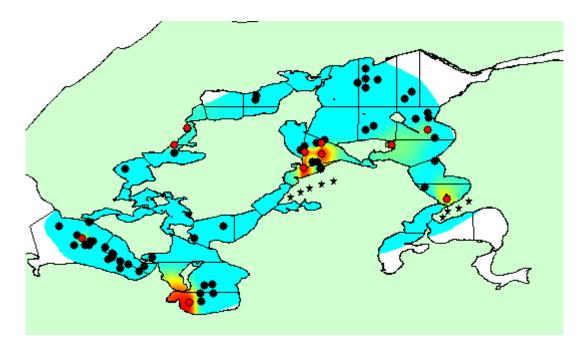
Figure 15   Map of measurement distribution and *E. Coli* risk in 2003 with a risk map. Red dots are for positive and black for negative test results.

Figure 16 **Map of 2004 for *E. Coli* risk and distribution. Red- positive, black- negative response.**
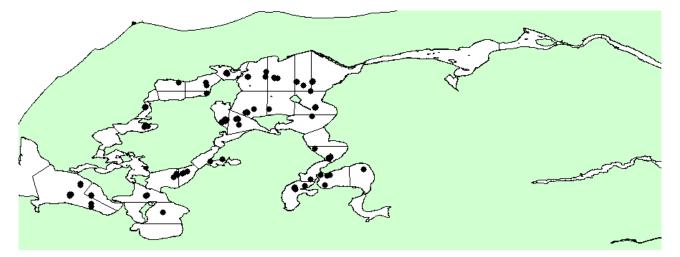
Figure 17 **Distribution of measurements in 2005**

No positive values were found for either tested measures (*E. Coli* and salmonella).
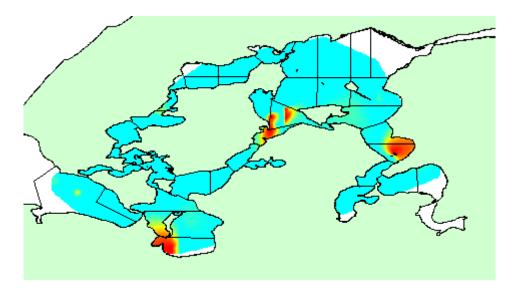
Figure 18   **Risk map for combined data from 2003 to 2005.**
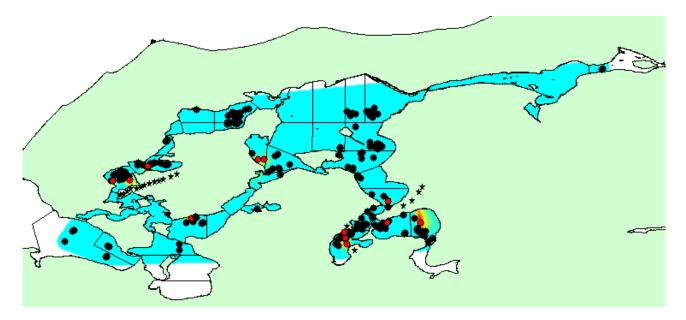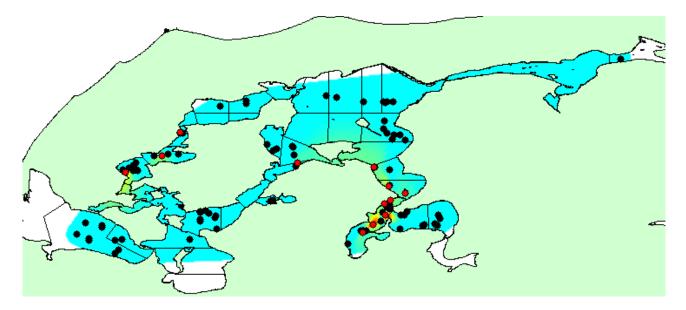
**Figure 19**   **Distribution and a risk map of 2006 for *E. Coli.* Black dot- negative, red- positive and star is a dummy value.**

Figure 20 Distribution and a risk map of 2007 for *E. Coli.* Black dot- negative, red- positive.
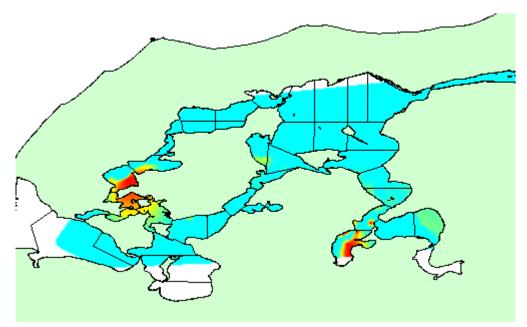
Figure 21 Risk map for combined years from 2006 to 2007.
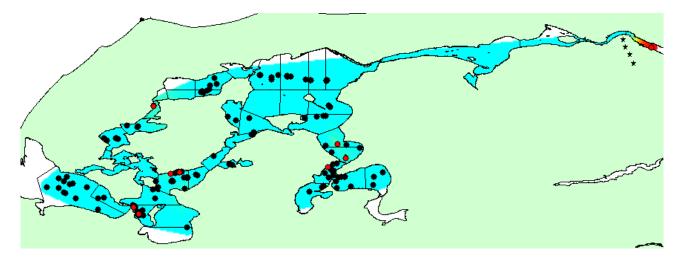
Figure 22 Distribution and a risk map of 2008 for *E. Coli.* Distribution map has been manipulated to get more realistic interpolation map. Black- negative, red- positive and black star- dummy variables.
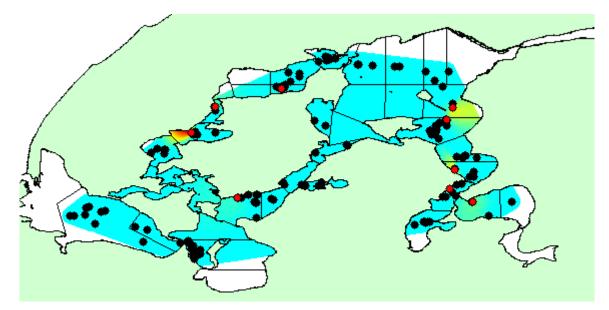
Figure 23 Distribution and a risk map of 2009 for *E. Coli.* Black dot- negative, red- positive.
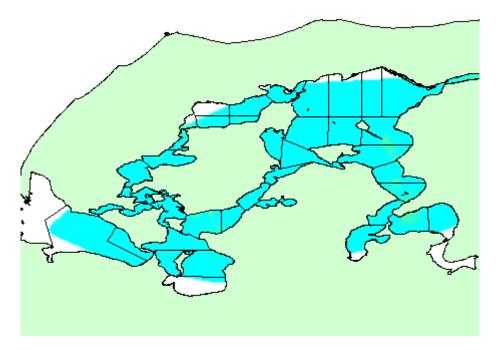
Figure 24   Risk map for combined years from 2008 to 2009.
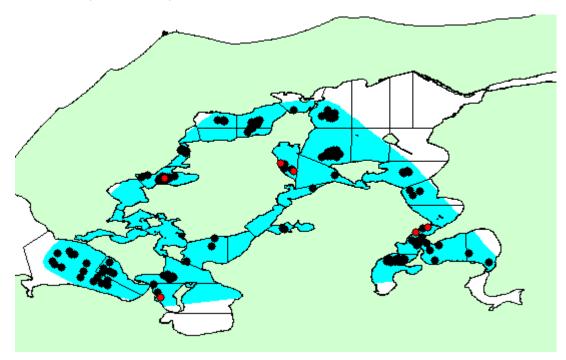
Figure 25   Distribution and a risk map of 2010 for *E. Coli*. Black dot- negative, red- positive.
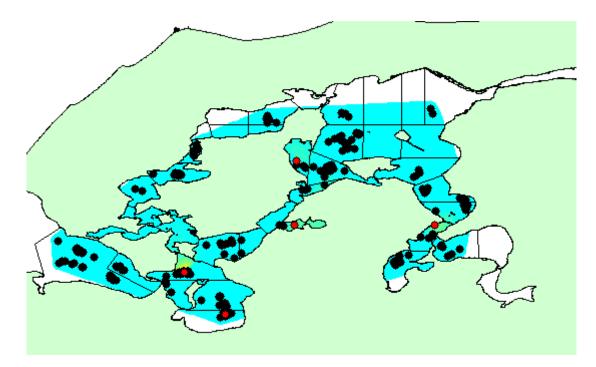
Figure 26　　　**Distribution and a risk map of 2011 for *E. Coli*.　Black dot- negative, red- positive.**
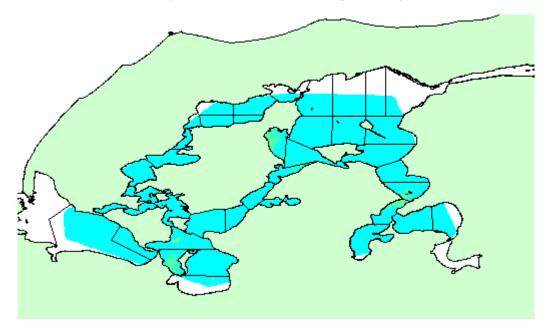
Figure 27　　　**Combined risk map for years 2010 to 2011.**
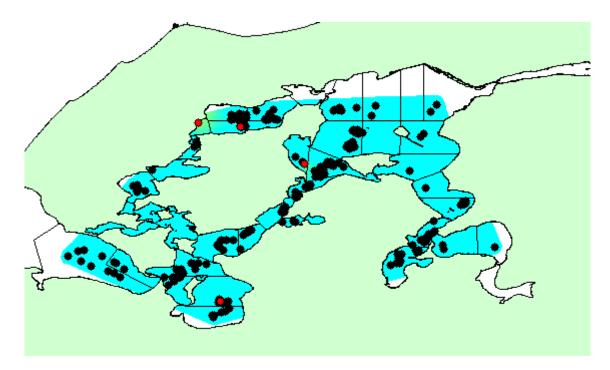
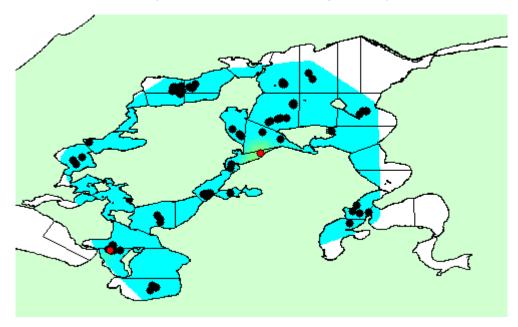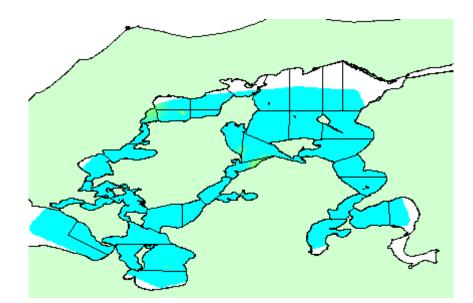**Figure 28**   Distribution and a risk map of 2012 for *E. Coli*. Black dot- negative, red- positive.

Figure 29   Distribution and a risk map of 2013 for *E. Coli*. Blue-green dot- negative, pink- positive.

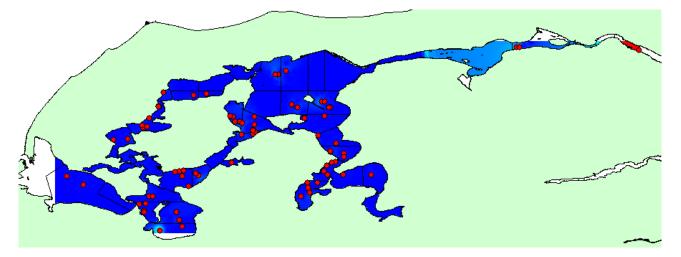**Figure 30**　　　　**Risk map for combined years from 2012 to 2013.**

Figure 31　　　　**Distribution and risk map over all the years. Dots on the maps show places where *E. Coli* has been found positive. Risk map colour code: Dark blue 0, light blue 0.5 and red 1.**
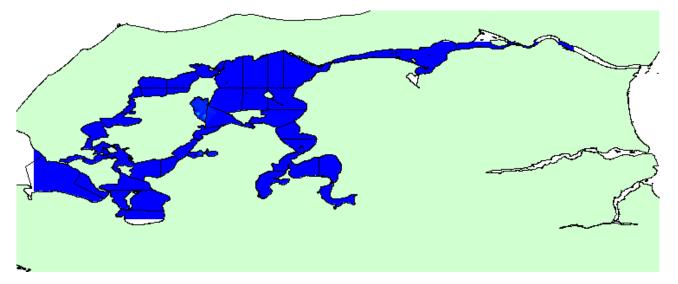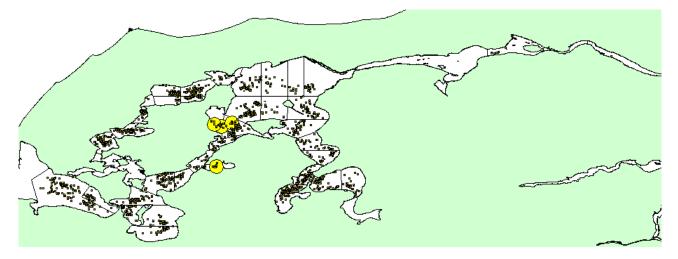
Figure 32   **Risk map for salmonella for all over the years analysed.**

Only six positive values for Salmonella were found in total over the years so light blue dots are hard to notice in Figure 31. The results are better seen in the graduated map in Figure 32.
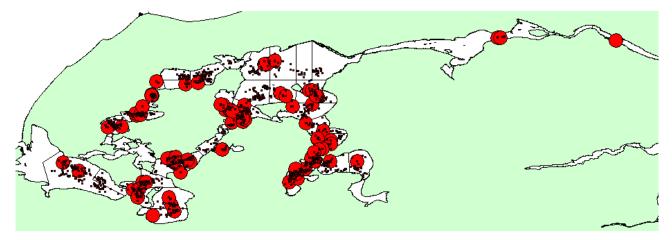

Figure 33   **Graduated map of salmonella findings from 1996 to 2013. Small dots represent negative values and big dots positive values.**

Figure 34   **Graduated map of *E. Coli* findings from 1996 to 2013. Small dots represent negative values and big dots positive values.**

For seasonality purposes similar maps were made only this time grouped by month instead of the year.
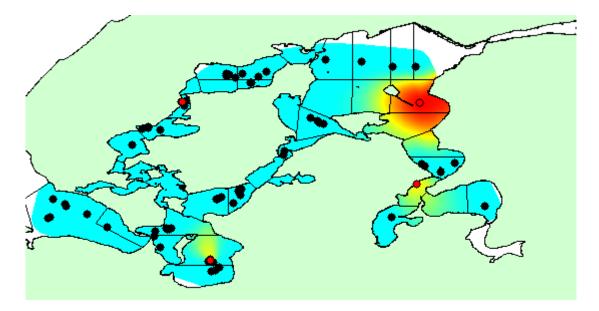
Figure 35   Combination of risk map and measured data for *E. Coli* in January including all analysed years. Black – negative and red dot positive measurements.
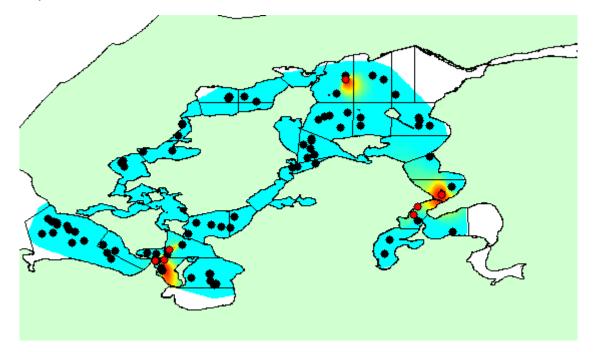

Figure 36   Combination of risk map and measured data for *E. Coli* in February including all analysed years. Black – negative and red dot positive measurements.
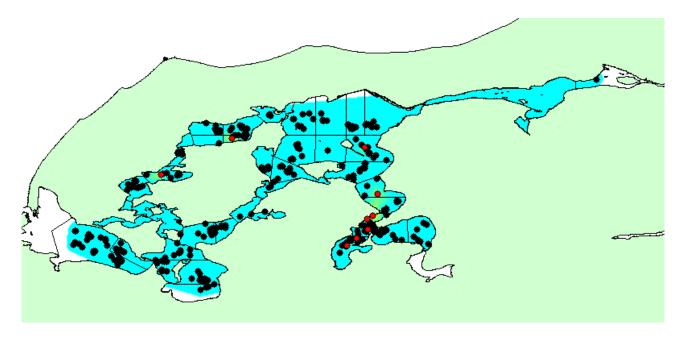
Figure 37 Combination of risk map and measured data for *E. Coli* in March including all analysed years. Black – negative and red dot positive measurements.

Figure 38 Combination of risk map and measured data for *E. Coli* in April including all analysed years. Black – negative and red dot positive measurements.

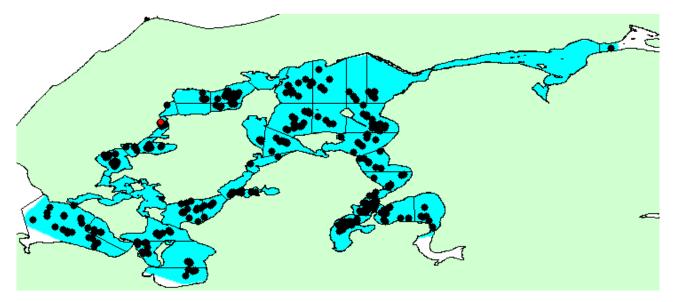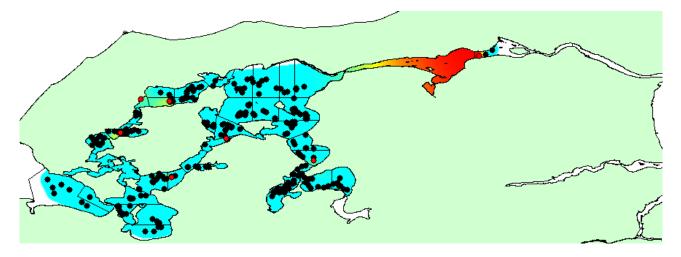**Figure 39** **Combination of risk map and measured data for *E. Coli* in May including all analysed years. Black – negative and red dot positive measurements.**
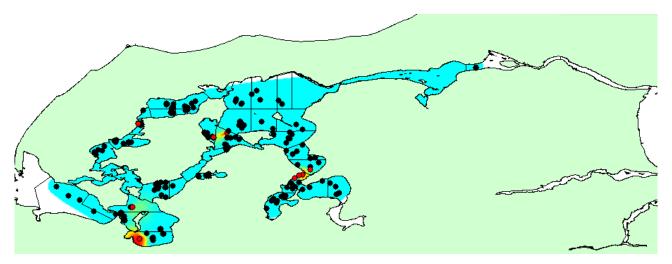
Figure 40 ***E. Coli* risk map in June including all analysed years with real measurements. Black – negative and red dot positive measurements.**
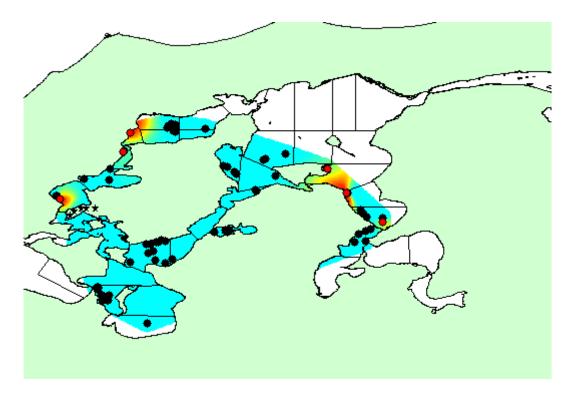
Figure 41 *E. Coli* risk map in July including all analysed years with real measurements. Black – negative and red dot positive measurements, stars dummy values.
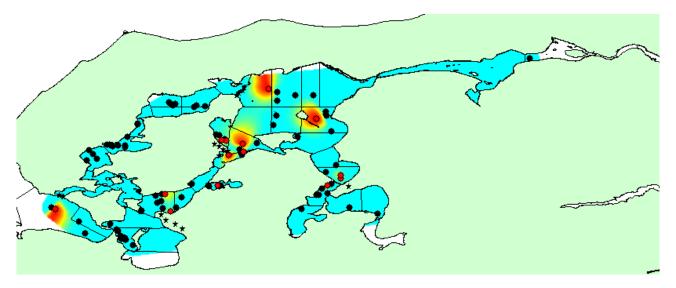
Figure 42 *E. Coli* risk map in August including all analysed years with real measurements. Black – negative and red dot positive measurements, dummy values represented with stars.
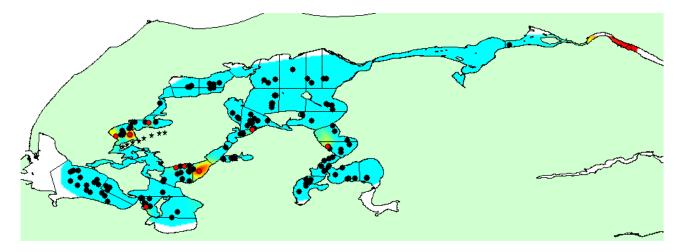
Figure 43 *E. Coli* risk map in September including all analysed years with real measurements. Black – negative and red dot positive measurements, stars represent dummy values.
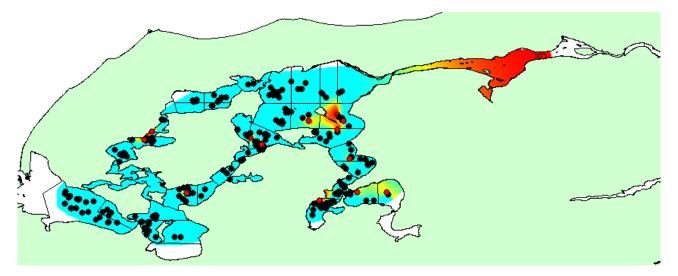

Figure 44 *E. Coli* risk map in October including all analysed years with real measurements. Black – negative and red dot positive measurements.
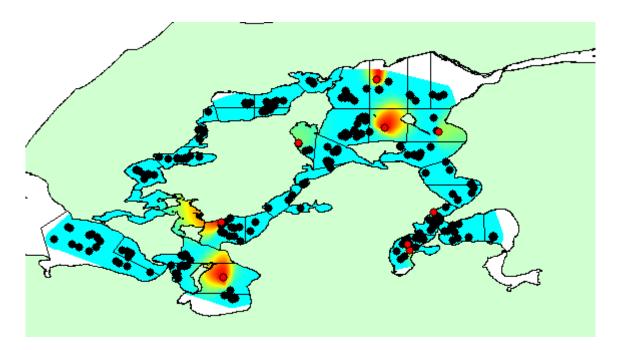
Figure 45 *E. Coli* **risk map in November including all analysed years with real measurements. Black – negative and red dot positive measurements.**
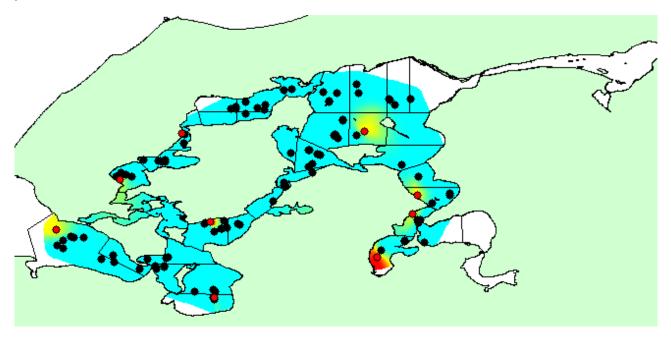
Figure 46 *E. Coli* **risk map in December including all analysed years with real measurements. Black – negative and red dot positive measurements.**
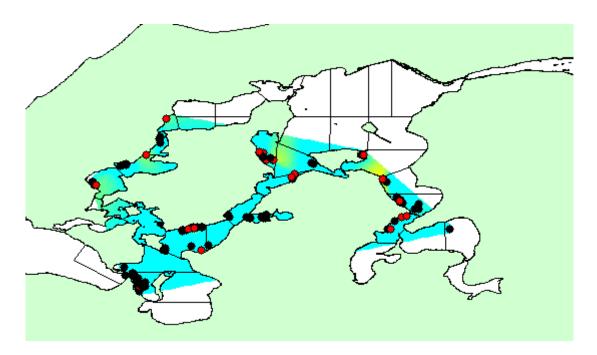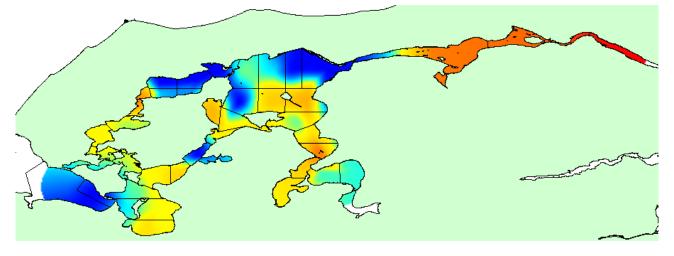
Figure 47 Mussel farm map covering results from all the years analysed. Black dots represent data points where *E. Coli* measurements were negative and red dots places where the results were positive.

	Value	Percentile
(blue)	0,0000	2,1978
(cyan)	0,0216	25,0048
(yellow)	0,0376	50,0129
(orange)	0,0911	75,0000
(red)	1,0000	100,0000

Figure 48 Relative number of positive samples per total number of samples taken in each fishery and legend for the colours.

Statistical analysis

Histograms were made to illustrate the frequency in between *E. Coli* if grouped by different categories. The data set used for this included all the data points available from all the years in Limfjorden.

Depending on the year and source, the coordinates for the studied points varied on the number of digits recorded. This variation could mean that zeroes were left out for simplification, or that a varying degree of precision took place over the years. The implications of these variations could be that a maximum of 20% of the data points could be in a false location. The distance from the correct position could vary between 20 m to almost 4 km depending on the coordinate precision shown later in Table 13.
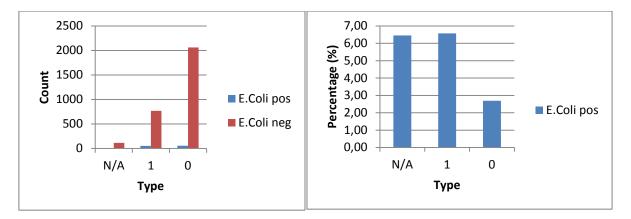
Figure 49   *E. Coli* frequency based on where the sample was taken: N/A- unknown, 1- mussel farm, 0- fishery Graph on the right shows the percentage of positive results with in the different groups.

Table 3   Values for *E. Coli* frequency grouped by type, (groups explained in Figure 48). Percentages of positive results calculated within the group and out of total measurements.

Type	E. Coli pos Count	E. Coli neg Count	Percentage of positive results within the group	Total Type	Percentage of samples collected for each group
N/A	8	116	6.45	124	4.05
1	54	768	6.57	822	26.83
0	57	2061	2.69	2118	69.13

More than half of the measurements were taken from the fishery that also had the lowest percentage of discovered positive results. Both the mussel farms and unknown sampling places had a similar percentage of positive results within the group.

Table 4   Distribution of *E. Coli* in different type (explained in Figure 48) per years.

Type	Year	Ecoli pos Count	Ecoli neg Count
N/A	1997		1
	2000	3	34
	2002		1
	2003	5	80
1	2004	6	14
	2005		3
	2006	5	37
	2007	13	68
	2008	12	237
	2009	4	161
	2010	5	74
	2011	4	85
	2012	4	67
	2013	1	22
0	1996	5	55
	1997	2	26
	1998		2
	2001	3	37
	2002	6	67
	2004	8	73
	2005		205
	2006	8	286
	2007	7	123
	2008	4	214
	2009	7	162
	2010	1	175
	2011	3	267
	2012	2	284
	2013	1	85
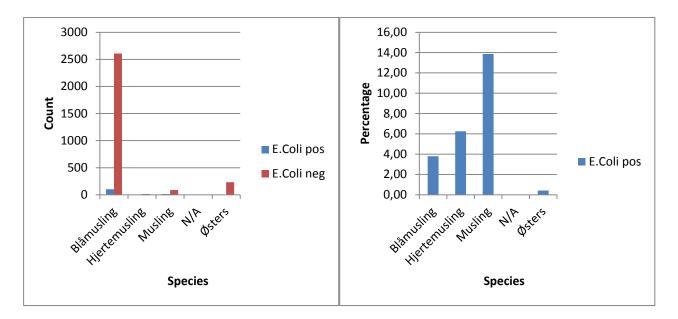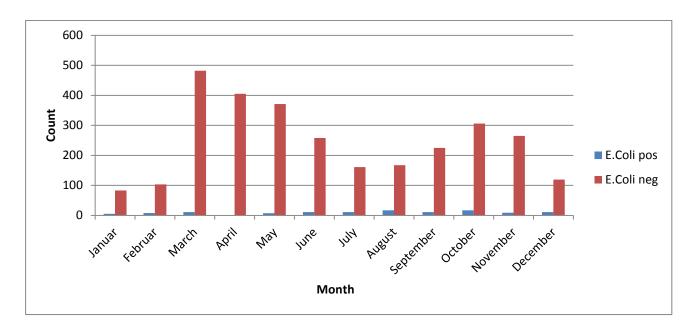
Figure 50 *E. Coli* frequency histogram grouped by species from all the years analysed. Graph on the right shows the percentage of positive results with in the different groups.

Table 5 Values for *E. Coli* frequency grouped by species. Percentages of positive results calculated within the group and out of total measurements.

Species	E. Coli pos Count	E. Coli neg Count	Percentage of positive results within the group	Total Species	Percentage of samples collected for each group
Blåmusling	103	2609	3.80	2712	88.51
Hjertemusling	1	15	6.25	16	0.52
Musling	14	87	13.86	101	3.30
N/A		2	0.00	2	0.07
Østers	1	232	0.43	233	7.60

Close to 90% of all the mussels collected were blue mussels. The E. Coli findings in that group were fairly low compared to the heart mussels (clams). Musling had a high percentage of positive results, but they are most likely part of other groups.
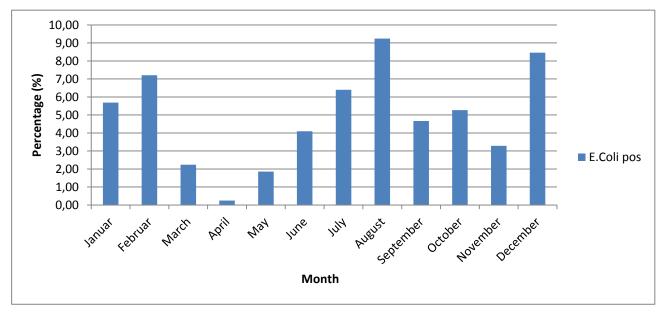
Figure 51 *E. Coli* **frequency histogram grouped by month from all the years analysed. Graph on the below shows the percentage of positive results with in the different groups.**

Table 6 **Values for** *E. Coli* **frequency grouped by month. Percentages of positive results calculated within the group and out of total measurements.**

Month	E. Coli pos	E. Coli neg	Percentage of positive results within the group	Total	Percentage of samples collected for each group
	Count	Count		Month	
January	5	83	5.68	88	2.87
February	8	103	7.21	111	3.62
March	11	482	2.23	493	16.09
April	1	405	0.25	406	13.25
May	7	371	1.85	378	12.34

June	11	258	4.09	269	8.78
July	11	161	6.40	172	5.61
August	17	167	9.24	184	6.01
September	11	225	4.66	236	7.70
October	17	306	5.26	323	10.54
November	9	265	3.28	274	8.94
December	11	119	8.46	130	4.24

There is no month where there has been a lot more sampling compared to other months. The most fishing over the years has been done in March, closely followed by April, May and October. The highest *E. Coli* counts are in August and October, the lowest in April. April has the biggest difference in between the amount fished and positive samples found.
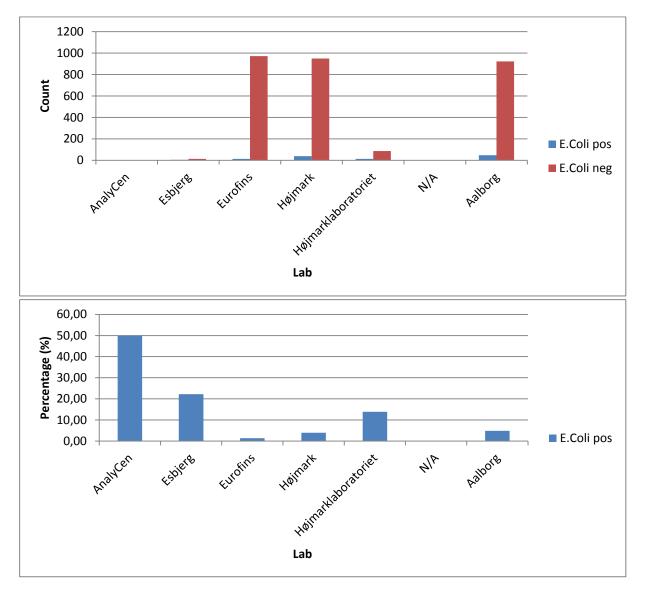
Figure 52 *E. Coli* **frequency histogram grouped by lab the analysis was made, from all the years analysed. Graph on the below shows the percentage of positive results with in the different groups.**

Table 7   Values for *E. Coli* frequency grouped by lab. Percentages of positive results calculated within the group and out of total measurements.

Lab	E. Coli pos Count	E. Coli neg Count	Percentage of positive results within the group	Total Lab	Percentage of samples collected for each group
AnalyCen	1	1	50.00	2	0.07
Esbjerg	4	14	22.22	18	0.59
Eurofins	14	971	1.42	985	32.15
Højmark	39	948	3.95	987	32.21
Højmarklaboratoriet	14	87	13.86	101	3.30
N/A		2	0.00	2	0.07
Aalborg	47	922	4.85	969	31.63

The main labs where the samples were analysed were Eurofins, Højmark and Aalborg lab, they all had low percentage of positive results, while labs that got less samples to analyse have bigger positive result production.
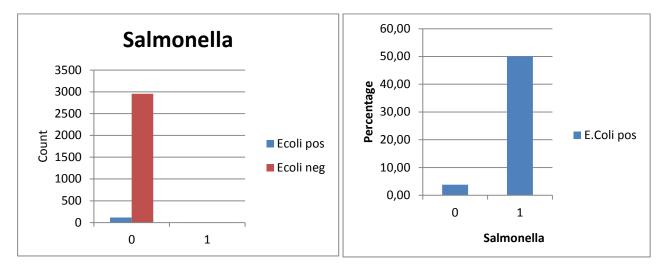
Figure 53   Histogram with *E. Coli* frequency in relation to salmonella in combination of results from all the years where data was available. Graph on the right shows the percentage of positive results with in the different groups.

Table 8   Values for *E. Coli* frequency grouped by salmonella response: positive (1) or negative (0). Percentages of positive results calculated within the group and out of total measurements.
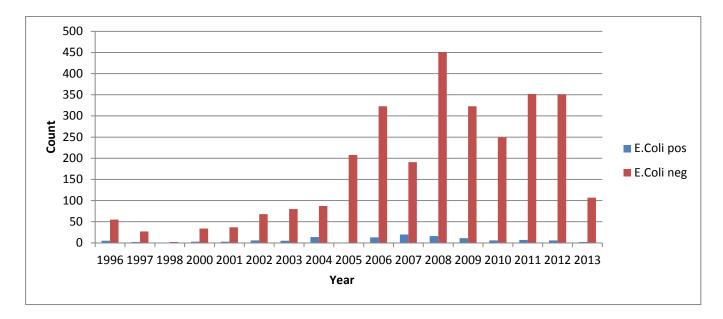
Salmonella	E. Coli pos Count	E. Coli neg Count	Percentage of positive results within the group	Total Salmonella	Percentage of samples collected for each group
0	116	2942	3.79	3058	99.80
1	3	3	50.00	6	0.20

Close to all samples analysed were negative for salmonella. Total of 6 samples were identified as positive. Distribution of salmonella positive samples can be seen in Figure 31.

Table 9 summary tables of salmonella positive samples. Type : N/A- unknown, 1- mussel farm, 0- fishery, *E. Coli*: negative 0, positive 1.

Zone no	Type	Ship	Month	Year	Species	Lab	Salmonella	E Coli
15	0	Limfjorden	Dec	2007	Blåmusling	Aalborg	1	0
14	1	N/A	June	2012	Blåmusling	Eurofins	1	1
14	1	N/A	June	2006	Blåmusling	Højmark	1	1
14	1	N/A	June	2006	Blåmusling	Højmark	1	1
14	1	N/A	March	2011	Blåmusling	Aalborg	1	0
12	1	N/A	Sep	2009	Blåmusling	Aalborg	1	0

All of the salmonella results have been found in the blue mussels and in mussel farm samples. Half of them were also positive for E. coli at the same time and they were found in June.
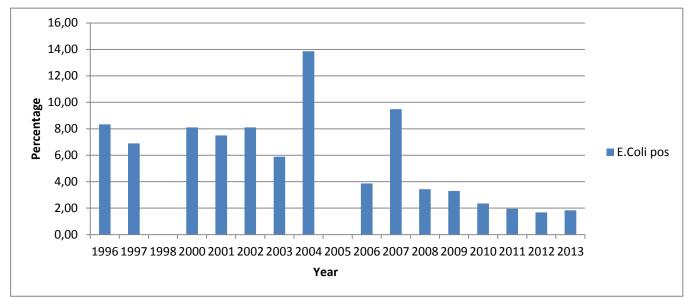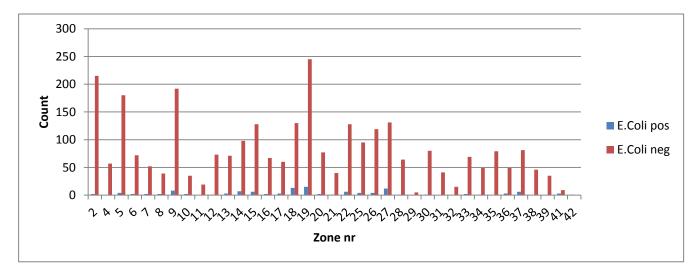
Figure 54 Histogram with *E. Coli* frequency in relation to year in combination of results from all the years where data was available. Graph on the lower shows the percentage of positive results with in the different groups.

Table 10 Values for *E. Coli* frequency grouped by year. Percentages of positive results calculated within the group and out of total measurements.

Year	E. Coli pos Count	E. Coli neg Count	Percentage of positive results within the group	Total Year	Percentage of samples collected for each group
1996	5	55	8.33	60	1.96
1997	2	27	6.90	29	0.95
1998		2	0.00	2	0.07
2000	3	34	8.11	37	1.21
2001	3	37	7.50	40	1.31

2002	6	68	8.11	74	2.42
2003	5	80	5.88	85	2.77
2004	14	87	13.86	101	3.30
2005		208	0.00	208	6.79
2006	13	323	3.87	336	10.97
2007	20	191	9.48	211	6.89
2008	16	451	3.43	467	15.24
2009	11	323	3.29	334	10.90
2010	6	249	2.35	255	8.32
2011	7	352	1.95	359	11.72
2012	6	351	1.68	357	11.65
2013	2	107	1.83	109	3.56

Over the year the amount of data stored has grown. Over the resent years the positive sample number has gone down.
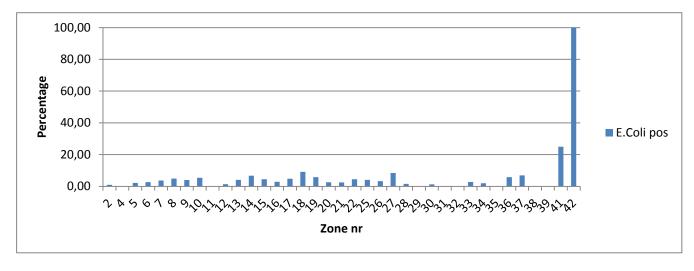

Figure 55 Histogram with *E. Coli* frequency in relation to year in combination of results from all the years where data was available. Graph on the lower shows the percentage of positive results with in the different groups.

Table 11 **Values for *E. Coli* frequency grouped by zone number. Percentages of positive results calculated within the group and out of total measurements.**

Zone nr	E. Coli pos Count	E. Coli neg Count	Percentage of positive results within the group	Total Zone nr	Percentage of samples collected for each group
2	2	215	0.92	217	7.08
4		57	0.00	57	1.86
5	4	180	2.17	184	6.01
6	2	72	2.70	74	2.42
7	2	52	3.70	54	1.76
8	2	39	4.88	41	1.34
9	8	192	4.00	200	6.53
10	2	35	5.41	37	1.21
11		19	0.00	19	0.62
12	1	73	1.35	74	2.42
13	3	71	4.05	74	2.42
14	7	98	6.67	105	3.43
15	6	128	4.48	134	4.37
16	2	67	2.90	69	2.25
17	3	60	4.76	63	2.06
18	13	130	9.09	143	4.67
19	15	245	5.77	260	8.49
20	2	77	2.53	79	2.58
21	1	40	2.44	41	1.34
22	6	128	4.48	134	4.37
25	4	95	4.04	99	3.23
26	4	119	3.25	123	4.01
27	12	131	8.39	143	4.67
28	1	64	1.54	65	2.12
29		5	0.00	5	0.16
30	1	80	1.23	81	2.64
31		41	0.00	41	1.34
32		15	0.00	15	0.49
33	2	69	2.82	71	2.32
34	1	49	2.00	50	1.63
35		79	0.00	79	2.58
36	3	49	5.77	52	1.70
37	6	81	6.90	87	2.84
38		46	0.00	46	1.50
39		35	0.00	35	1.14

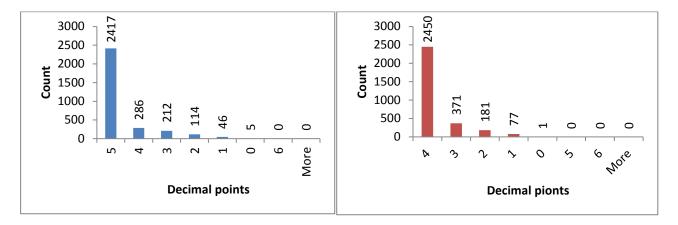| 41	3	9	25.00	12	0.39
42	1		100.00	1	0.03

Figure 56 **Histogram for analysing the number of decimal places in coordinates. Blue colour represents decimal place frequency for x coordinate and red for y coordinate.**

Table 12 **Frequency for number of decimal places in x and y coordinates with percentages of total amount of adequate coordinates.**

For x coordinate			For Y coordinate		
Bin	Frequency	%	Bin	Frequency	%
5	2417	78.5	4	2450	79.5
4	286	9.3	3	371	12.0
3	212	6.9	2	181	5.9
2	114	3.7	1	77	2.5
1	46	1.5	0	1	0.0
0	5	0.2	5	0	0.0
6	0	0.0	6	0	0.0
More	0	0	More	0	0

Table 13 **Example of coordinate precision based on Y coordinate on the location of the sample site.**

Nr of Decimals	Decimal degrees	Northing	Location change in m
4	56.8342	6298932.26	0
3	56.834	6298909.998	22.26246476
2	56.83	6298464.749	467.5116092
1	56.8	6295125.389	3806.871036
0	56	6206081.789	92850.47108

Little over 20 % of both coordinates were weak in their precision. As seen in Table 13 the amount of decimal points greatly affects the location of the point. That can also be seen in MapInfo looking at the points and the Zone number, some of them do not line up.

```
Model: binomial, link: logit

Response: EColi

Terms added sequentially (first to last)


        Df Deviance Resid. Df Resid. Dev  Pr(>Chi)
NULL                     2927     918.22
ZONE   29   50.512      2898     867.71 0.0079498 **
month  11   57.387      2887     810.32 2.824e-08 ***
type    1   11.992      2886     798.33 0.0005342 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
Advarselsbesked:
glm.fit: fitted probabilities numerically 0 or 1 occurred
```

Figure 57   Linear model without any random effects implemented.

```
Data: mussel1
Models:
EColi2: EColi ~ ZONE + (1 + 1 | YEAR)
EColi4: EColi ~ ZONE + type + (1 + 1 | YEAR)
       Df    AIC    BIC  logLik deviance  Chisq Chi Df Pr(>Chisq)
EColi2 31 976.77 1163.5 -457.39   914.77
EColi4 32 899.88 1091.3 -417.94   835.88 78.894      1  < 2.2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Figure 58   Comparison of two mixed models using ANOVA to show overall significance value for type.

There seems to be a difference in between where the samples are coming from (fishery or mussel farms).

```
Data: mussel1
Models:
EColi5: EColi ~ type + (1 + 1 | YEAR)
EColi4: EColi ~ ZONE + type + (1 + 1 | YEAR)
       Df    AIC     BIC  logLik deviance  Chisq Chi Df Pr(>Chisq)
EColi5  3 879.08  897.02 -436.54   873.08
EColi4 32 899.88 1091.30 -417.94   835.88 37.199     29     0.1412
```

Figure 59   Comparison of two mixed models using ANOVA to show overall significance value for Zone number.

No overall significance was found but in Figure 60 it can be seen that some zones are more significant, the significance values correlate well with the relative E. Coli positive graph (Figure 47).

```
Data: mussel1
Models:
EColi4: EColi ~ ZONE + type + (1 + 1 | YEAR)
EColi7: EColi ~ ZONE + month + type + (1 + 1 | YEAR)
        Df    AIC    BIC  logLik deviance  Chisq Chi Df Pr(>Chisq)
EColi4 32 899.88 1091.3 -417.94   835.88
EColi7 43 870.71 1127.9 -392.36   784.71 51.166     11  3.859e-07 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Figure 60   Comparison of two mixed effects models using ANOVA to show overall significance to month.

```
Generalized linear mixed model fit by maximum likelihood ['glmerMod']
 Family: binomial ( logit )
Formula: EColi ~ ZONE + month + type + (1 + 1 | YEAR)
   Data: mussel1

      AIC       BIC    logLik  deviance
 870.7128 1127.9420 -392.3564  784.7128

Random effects:
 Groups Name        Variance Std.Dev.
 YEAR   (Intercept) 0.3336   0.5776
Number of obs: 2928, groups: YEAR, 6
```

```
Fixed effects:
              Estimate  Std. Error  z value  Pr(>|z|)
(Intercept)   -22.5107   1386.4676   -0.016   0.98705
ZONE4         -14.9087   3498.6235   -0.004   0.99660
ZONE5           0.9792      1.1731    0.835   0.40392
ZONE6           1.8552      1.2672    1.464   0.14318
ZONE7           1.6407      1.4437    1.136   0.25575
ZONE8           2.7672      1.2726    2.174   0.02967 *
ZONE9-10        1.7098      1.0902    1.568   0.11681
ZONE11        -14.3863   5187.9504   -0.003   0.99779
ZONE12          0.4520      1.4652    0.308   0.75773
ZONE13          1.5848      1.1977    1.323   0.18579
ZONE14          1.9557      1.1283    1.733   0.08302 .
ZONE15          1.9470      1.1069    1.759   0.07857 .
ZONE16          1.9121      1.2620    1.515   0.12974
ZONE17-18       2.1284      1.0734    1.983   0.04738 *
ZONE19          2.0246      1.0787    1.877   0.06053 .
ZONE20          1.7704      1.2480    1.419   0.15602
ZONE21          1.6984      1.4434    1.177   0.23931
ZONE22          2.4931      1.1012    2.264   0.02358 *
ZONE25          1.7826      1.1511    1.549   0.12148
ZONE26          1.8970      1.1458    1.656   0.09781 .
ZONE27          2.4654      1.0744    2.295   0.02175 *
ZONE28-29       1.4817      1.4588    1.016   0.30977
ZONE30-31       0.8628      1.4320    0.603   0.54684
ZONE32        -14.7724   6410.4289   -0.002   0.99816
ZONE33          2.0828      1.2524    1.663   0.09629 .
ZONE34          1.5546      1.4443    1.076   0.28176
ZONE35-36       1.6719      1.2514    1.336   0.18155
ZONE37          2.5756      1.1196    2.300   0.02142 *
ZONE38        -15.0038   4002.3044   -0.004   0.99701
ZONE39        -14.5517   4093.7133   -0.004   0.99716
monthAug       17.7431   1386.4672    0.013   0.98979
monthDec       18.0246   1386.4672    0.013   0.98963
monthFeb       18.0622   1386.4672    0.013   0.98961
monthJan       17.8651   1386.4673    0.013   0.98972
monthJuly      17.2769   1386.4672    0.012   0.99006
monthJune      17.0301   1386.4672    0.012   0.99020
monthMarch     16.7055   1386.4672    0.012   0.99039
monthMay       16.1144   1386.4673    0.012   0.99073
monthNov       17.0971   1386.4672    0.012   0.99016
monthOct       17.4075   1386.4672    0.013   0.98998
monthSep       17.2308   1386.4672    0.012   0.99008
type1           1.0868      0.3163    3.436   0.00059 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Figure 61   **Breakdown of the main mixed effects model used for analysis.**

For the model fit AIC number was accessed, which stayed relatively similar throughout all the different simulations made. In quite a few cases the standard error is very high, this comes from very low number of data points in the grids, for improvement of the data further data piling would be necessary, for example

grouping months and some extra zones. If ignoring this source of error the model seems to work for fishery zones and is comparable to map manipulations. Lacks power in respect seasonality.

Brief summary

- The amount of data collected and stored before 2005 is low.
- Most of the mussels are collected from fisheries, while positive E. Coli samples found in them were lower compared to samples collected in the mussel farms. Mussel farms seem have higher probability to get positive results, most likely because mussel farmers were unaware of the contamination that can occur during sampling. Numbers of positive samples in mussel farms have decreased in the latter part of the period.
- No distinct zones with higher probability of positive E. coli results were found.
- No seasonality of positive E. coli results were found due the lack of evenly distributed and continuously taken samples. The standard error of the estimates was too big to believe the significance result achieved with statistical analysis.
- Salmonella positive results were in very a low numbers and no tendencies were found.