Modeling counts in Trawling Data

Ritei Shibata and Yuki Sugaya (Keio University)
with Ross Darnell, Mick Haywood, Charis Burridge(CSIRO) and Hideyasu Shimadzu (GeoScience Australia)

2008-03-15~3/31: Visit of Charis Burridge to Keio in the occasion of Cherry Bud Workshop

www.csiro.au

A new project to foster Japanese-Australian discovery through Data Science for environmental and ecological management

Charis Burridge

Cherry Bud Workshop, Mar '08, Keio University, Japan

CSIRO

Trawling Data

Quantifying the Effects of Trawling on Seabed Fauna in the Northern Prawn Fishery

Mick Haywood (Principal Investigator), Burke Hill (Co-Investigator), Anthea Donovan, Wayne Rochester, Nick Ellis, Andrzej Welna, Scott Gordon, Sue Cheers, Karl Forcey, Ian McLeod, Chris Moeseneder, Greg Smith, Fiona Manson, Ted Wassenberg and Steve Thomas

CSIRO Marine and Atmospheric Research

Petra Kuhnert, Geoff Laslett, Charis Burridge and Sarah Thomas

CSIRO Mathematical and Information Sciences.

October 2005

CSIRO

Australian Government

Fisheries Research and Development Corporation

FRDC Project 2002/102

Effects of Trawling Subprogram: Quantifying the Effects of Trawling on Seabed Fauna in the Northern Prawn Fishery

2008-06-28~2008-07-06: Visit of Ross and Mick to Keio

慶應義塾大学データサイエンス研究室
水曜セミナー（特別講義）

Marine Ecological Data Analysis

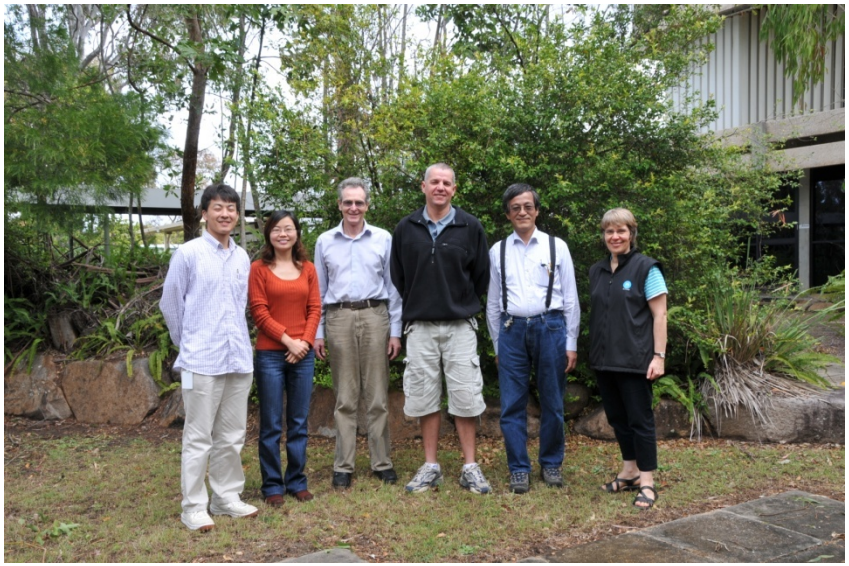
Quantifying the effects of trawling on seabed fauna in the Northern Prawn Fishery

講演者：Dr. Ross Darnell and Dr. Mick Haywood
(CSIRO Marine Laboratories, Australia,
Partly supported by Australia -Japan Foundation)

日時：2008－07－02　11:00 ~ 12:15
場所：慶應義塾大学理工学部矢上キャンパス 36 棟 205 号室

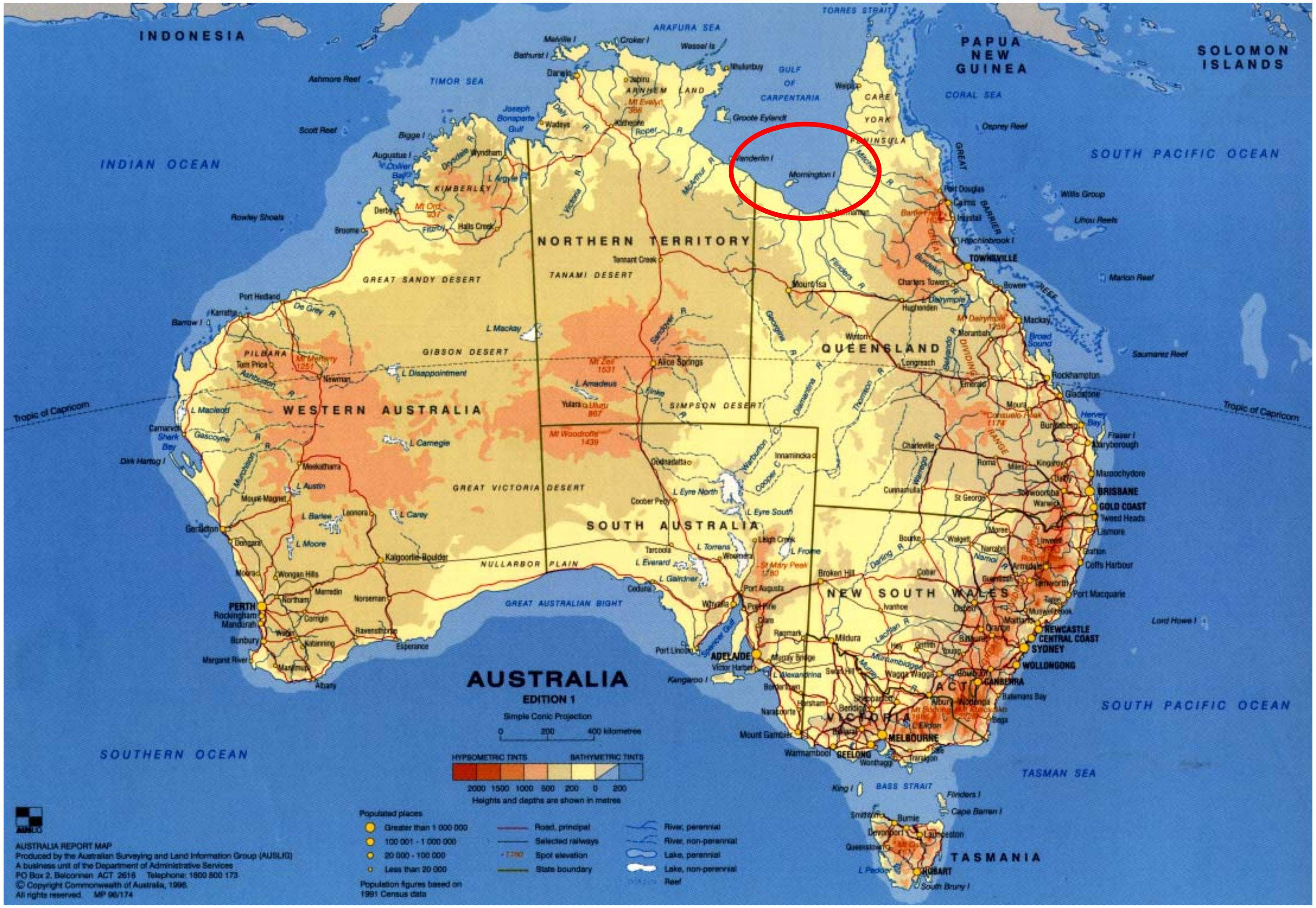
問い合わせ先
柴田里程
横浜市港北区日吉 3-14-1
慶應義塾大学理工学部数理科学科
Tel：045-566-1654

2008-07-30~2008-09-04: Visit of Ritei Shibata to Marine Labo., Clevelnd

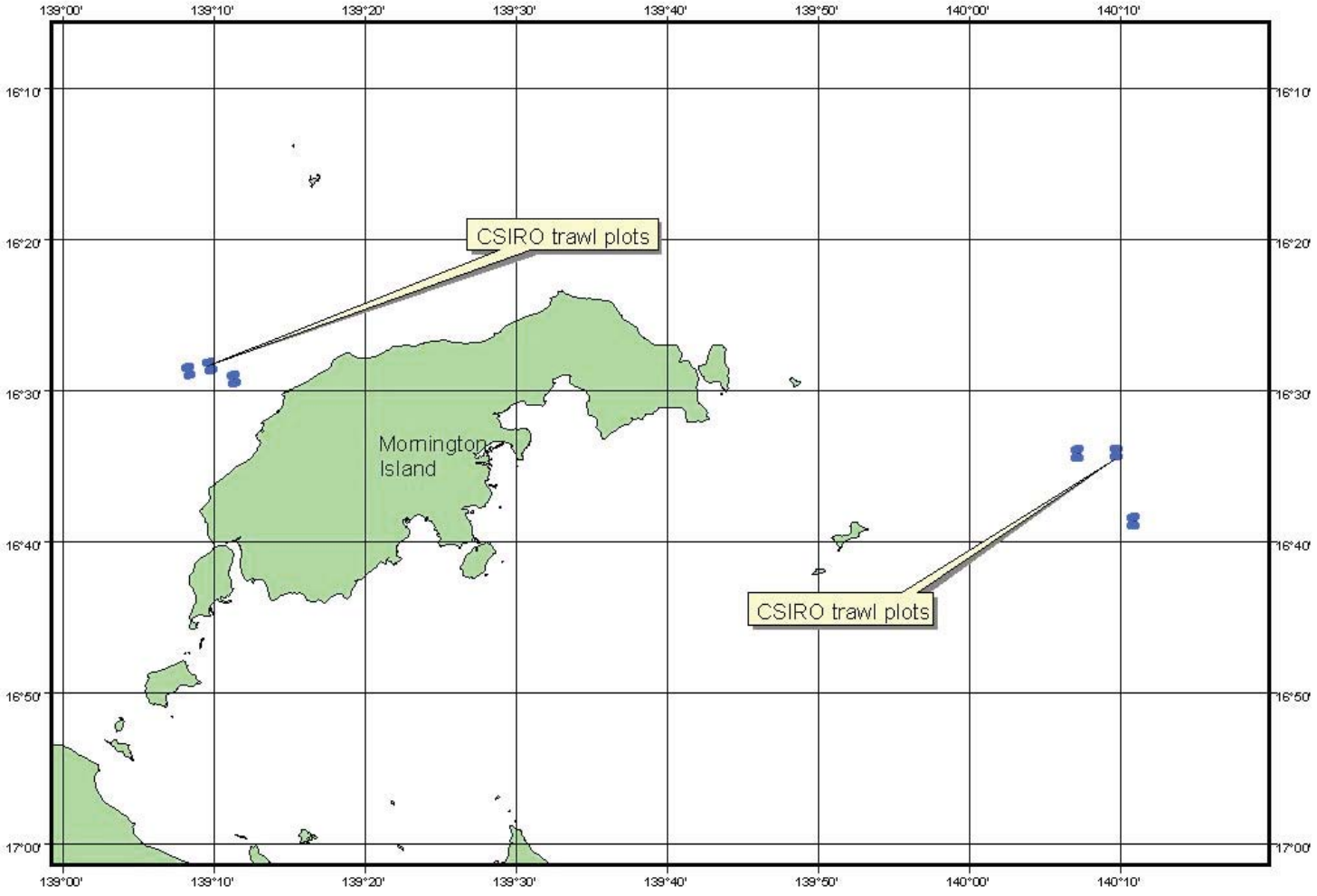


2008-09-14~2008-10-04: Visit of Yuki Sugaya to Marine Labo., Clevelnd

Northern Prawn Fishery(NPF)

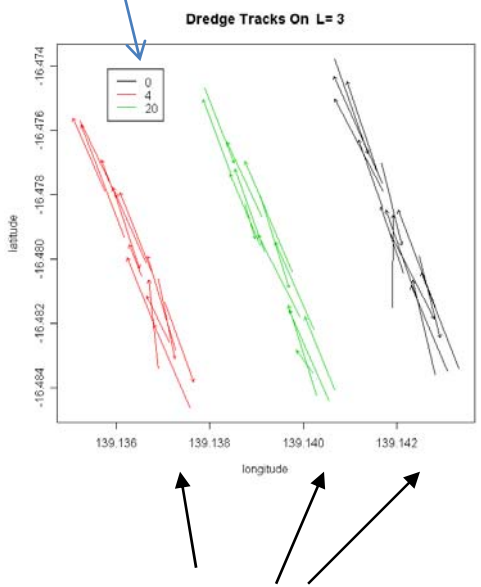


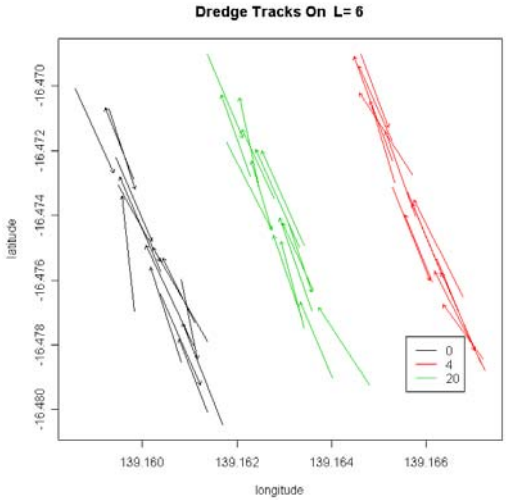
Plots for Experiment



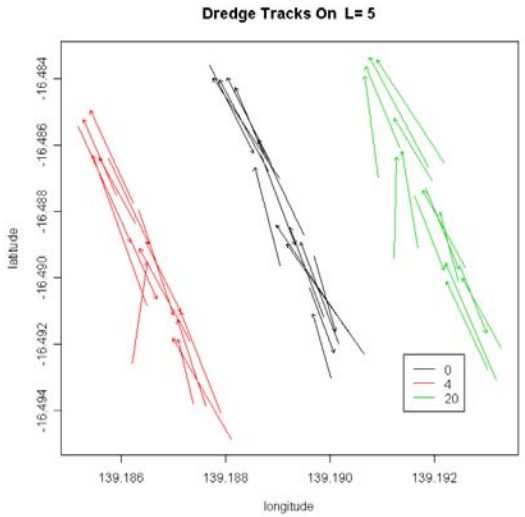
Experiment 1 (West Region)

treatments

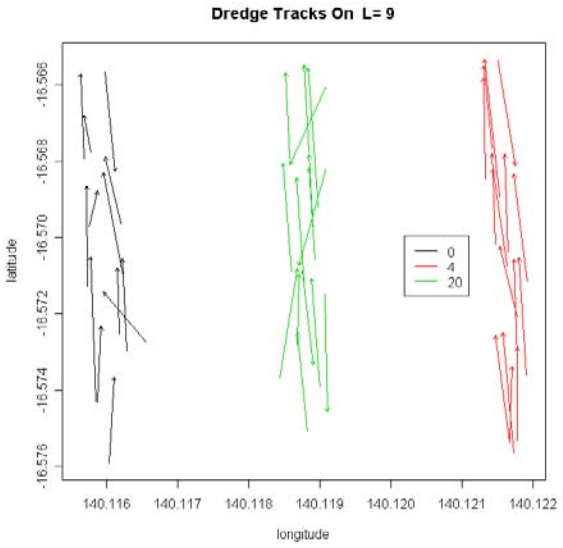


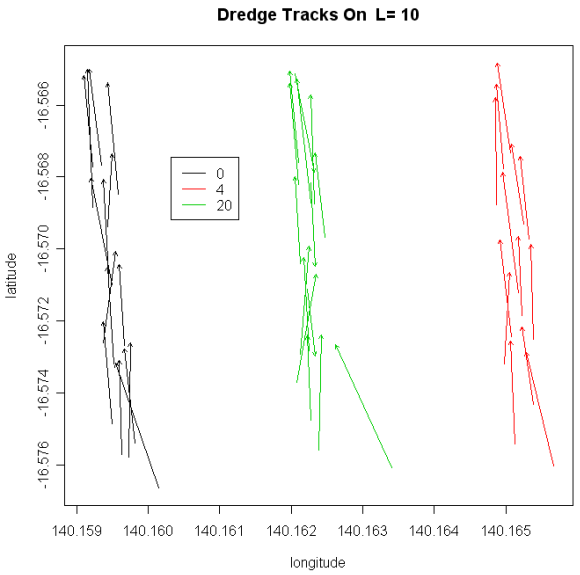


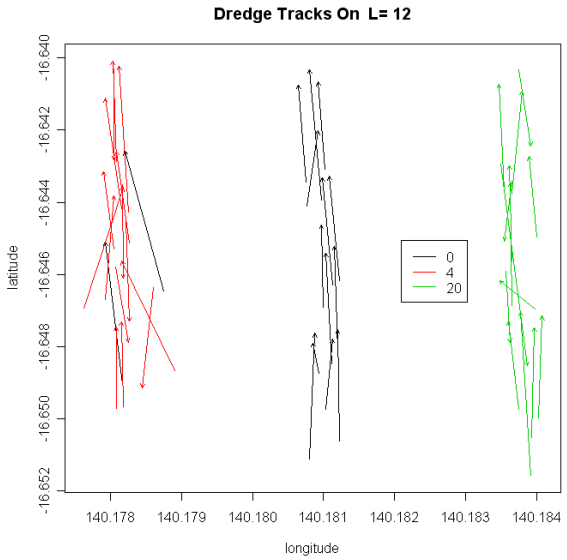
(Before After 6month 12month 18month) x 3



Experiment 1 (East Region)







Experiment 1 Data

```
> expt1[100:110,]
    OPERATION_NO     sdate  CRUISE   SPCODE        SCIENTIFIC_NAME        FAMILY       CLASS
100            1 02MAR2003 CC03/03 11208807  Veretillidae  OPNO 142  Veretillidae     Anthozoa
101            1 02MAR2003 CC03/03 11208808 Virgularia sp OPNO 034  Virgulariidae    Anthozoa
102            1 02MAR2003 CC03/03 11219902 Pteroides sp  OPNO 186  Pteroeididae     Anthozoa
103            1 02MAR2003 CC03/03 11224800 Virgularia sp OPNO 281  Virgulariidae  Pennatulacea
104            1 02MAR2003 CC03/03 11224801 Virgularia sp OPNO 398  Virgulariidae  Pennatulacea
105            1 02MAR2003 CC03/03 11224802 Virgularia sp OPNO 420  Virgulariidae  Pennatulacea
106            1 02MAR2003 CC03/03 11232800      Anemone OPNO 008    Actiniidae      Anthozoa
107            1 02MAR2003 CC03/03 11287001 Sphenopus marsupialis    Spenopidae      Anthozoa
108            1 02MAR2003 CC03/03 11288800 Hormathiidae OPNO 214   Hormathiidae    Anthozoa
109            1 02MAR2003 CC03/03 11288801   Zoanthidae OPNO 149    Zoanthidae      Anthozoa
110            1 02MAR2003 CC03/03 11288802     Anthozoa OPNO 002    Zoanthidae      Anthozoa

    operation_code REGION LOCATION TREATMENT time   nperha   gperha          COMMON_NAME
100       1:Before    EMN       12         4    B  0.00000  0.00000               Seapen
101       1:Before    EMN       12         4    B  0.00000  0.00000               Seapen
102       1:Before    EMN       12         4    B  0.00000  0.00000               Seapen
103       1:Before    EMN       12         4    B  0.00000  0.00000 Sea Pen Pennatulacea
104       1:Before    EMN       12         4    B  0.00000  0.00000 Sea Pen Pennatulacea
105       1:Before    EMN       12         4    B  0.00000  0.00000 Sea Pen Pennatulacea
106       1:Before    EMN       12         4    B  0.00000  0.00000             Anemone
107       1:Before    EMN       12         4    B 23.22880 39.25668 [a zoanthid anemone]
108       1:Before    EMN       12         4    B  0.00000  0.00000            Zooanthid
109       1:Before    EMN       12         4    B  0.00000  0.00000            Zooanthid
110       1:Before    EMN       12         4    B  0.00000  0.00000            Zooanthid
>
```

Structure of Experiment 1 Data

nperha : *Number per hectare*

and

gperha : *Gram per hectare*

on

$SPCODE \times LOCATION \times TREATMENT \times time$

SPCODE : *species code* (777); 10000813~99560007

LOCATION : *location of plot* : 3, 5, 6, 9 , 10, 12

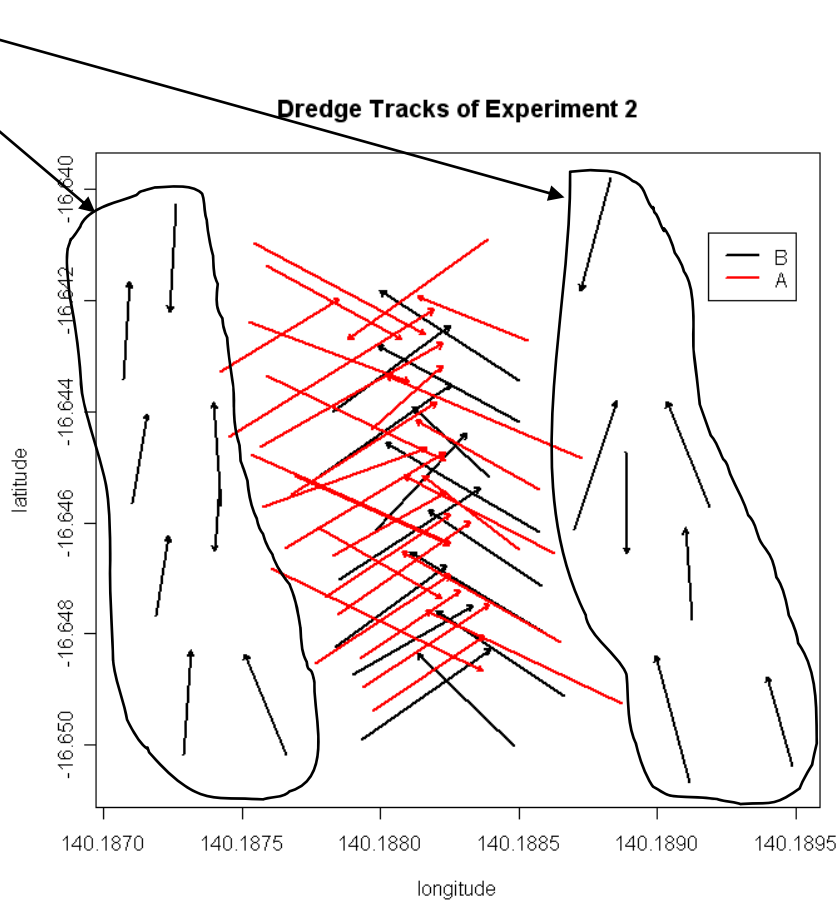
TREATMENT : *number of trawlings* : 0, 4, 20

time : *time of dredge* : B, A, 6m, 12m, 18m

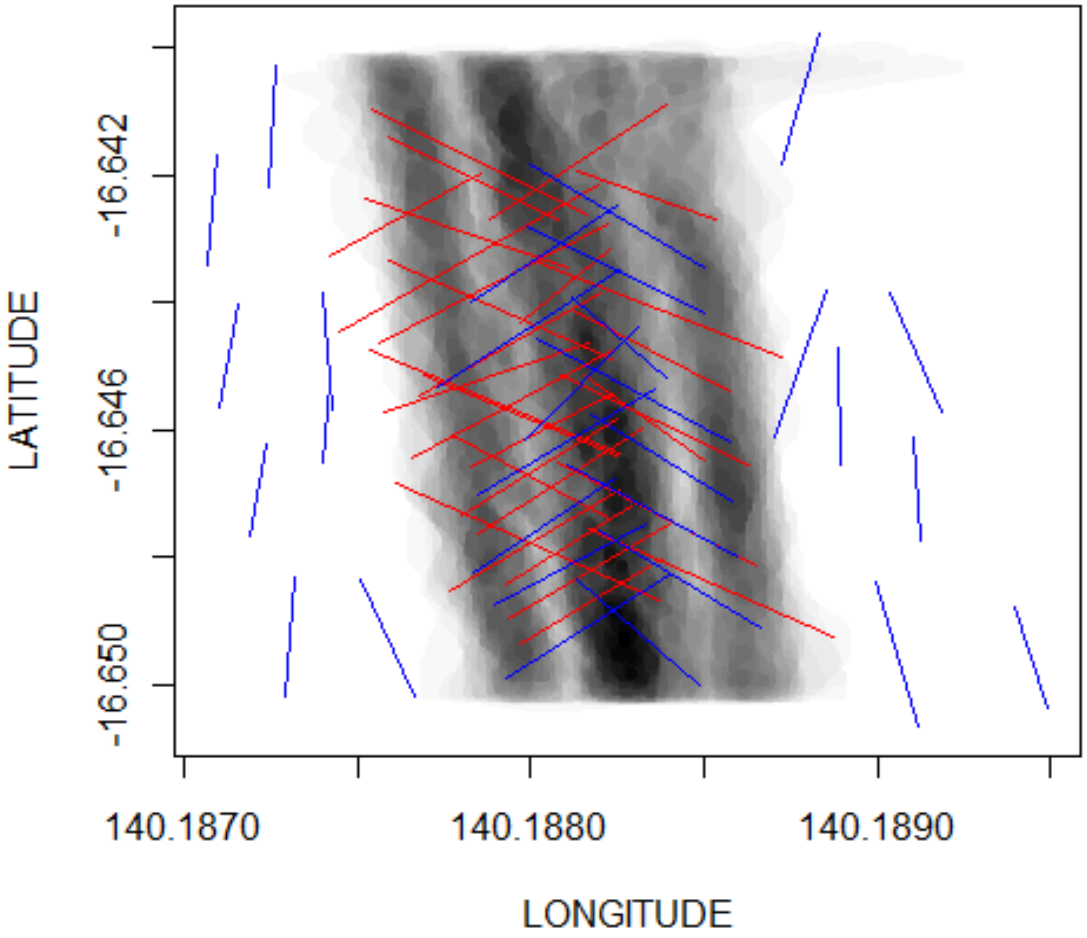
Experiment 2 (East Region, Location 13)

$$B = \begin{cases} B1 \\ B2 \end{cases} \quad (Before\,Trawling), \quad A\ (After\,Trawling)$$

40 dense trawlings in the middle



Trawling Intensity



Experiment 2 Data

	CLASS	COMMON_NAME	CRUISE	FAMILY	gperha	LOCATION	nperha	operation_code	OPERATION_NO
1	Anthozoa	[a zoanthid anemone]	CC05/02	Spenopidae	0	13	0	1: Before	443
2	Anthozoa	[a zoanthid anemone]	CC05/02	Spenopidae	0	13	0	1: Before	446
3	Anthozoa	[a zoanthid anemone]	CC05/02	Spenopidae	0	13	0	1: Before	447
4	Anthozoa	[a zoanthid anemone]	CC05/02	Spenopidae	0	13	0	1: Before	448
5	Anthozoa	[a zoanthid anemone]	CC05/02	Spenopidae	0	13	0	1: Before	449
6	Anthozoa	[a zoanthid anemone]	CC05/02	Spenopidae	0	13	0	1: Before	450
7	Anthozoa	[a zoanthid anemone]	CC05/02	Spenopidae	0	13	0	1: Before	451
8	Anthozoa	[a zoanthid anemone]	CC05/02	Spenopidae	0	13	0	1: Before	452
9	Anthozoa	[a zoanthid anemone]	CC05/02	Spenopidae	0	13	0	1: Before	453
10	Anthozoa	[a zoanthid anemone]	CC05/02	Spenopidae	0	13	0	1: Before	454

	SCIENTIFIC_NAME	SPCODE	TGROUP	START.LAT	START.LON	END.LAT	END.LON	time	START_EST
1	Sphenopus marsupialis	11287001	Anthozoa	-16.65039	140.1895	-16.64877	140.1894	B	24FEB2005:09:29:00
2	Sphenopus marsupialis	11287001	Anthozoa	-16.63978	140.1888	-16.64181	140.1887	B	24FEB2005:10:46:00
3	Sphenopus marsupialis	11287001	Anthozoa	-16.64612	140.1887	-16.64381	140.1889	B	24FEB2005:11:07:00
4	Sphenopus marsupialis	11287001	Anthozoa	-16.64472	140.1889	-16.64655	140.1889	B	24FEB2005:11:23:00
5	Sphenopus marsupialis	11287001	Anthozoa	-16.65068	140.1891	-16.64840	140.1890	B	24FEB2005:11:42:00
6	Sphenopus marsupialis	11287001	Anthozoa	-16.64027	140.1873	-16.64220	140.1872	B	24FEB2005:11:59:00
7	Sphenopus marsupialis	11287001	Anthozoa	-16.64570	140.1874	-16.64385	140.1874	B	24FEB2005:12:23:00
8	Sphenopus marsupialis	11287001	Anthozoa	-16.65019	140.1877	-16.64837	140.1875	B	24FEB2005:13:20:00
9	Sphenopus marsupialis	11287001	Anthozoa	-16.64526	140.1874	-16.64650	140.1874	B	24FEB2005:13:34:00
10	Sphenopus marsupialis	11287001	Anthozoa	-16.65018	140.1873	-16.64830	140.1873	B	24FEB2005:13:48:00

Structure of Experiment 2 Data

$nperha$: *Number per hectare*

and

$gperha$: *Gram per hectare*

on

$SPCODE \times time$

$SPCODE$: *species code* (777); 10000813~99560007

(*LOCATION* : *location of plot* : 13)

(*TREATMENT* : *number of trawlings* : 40)

$time$: *time of dredge* : B($B1$, $B2$), A

Project Details

- 2008-03~2008-07: Understanding the background of this survey
  Understanding the preliminary analysis
- 2008-08~2008-09: Data correction and enhancement
  Exploring the data
  Challenges for inflated zeros
  From normalised to un-normalised counts and biomass
- 2008-09~2008-10: Application of Thomas model
- 2008-11~2009-01: Theoretical Development of application of Thomas model
- 2009-02~2009-03: Thomas model for the number and Gamma model for the weight data of Experiment 2

Analysis in the report

- Descriptive statistics
- Depletion model, Burridge et al. (2003)

$$T_i = \sum_{j=1}^{i-1} c_j : \ Total \ \ amount \ of \ \ benthos \ caught \ \ prior \ to \ trawl \ i,$$

$$where \ c_j = q_j (X_0 - T_j), \ \ q_j \sim Beta(\alpha, \beta), \ \ and \ \ X_0 : \ initial \ biomass$$

- Epi-benthic sled process, Kuhnert

$$\log(y_{ijk}) = \alpha + \gamma d_{jk} + l_j + tl_{jk} + \varepsilon_{ijk}; \ \ at \ location \ \ j \ with \ treatment \ k,$$

$$l_j \sim N(0, \sigma_l^2), \ \ tl_{jk} \sim N(0, \sigma_{tl}^2), \ \varepsilon_{ijk} \sim N(0, \sigma^2)$$

Normalised or un-normalised

A measure of abundance

$$nperha:\ N_i / \alpha_i,\ \ i = 1, 2, ..., k$$

$$gperha:\ W_i / \alpha_i,\ \ i = 1, 2, ..., k$$

$$N_i : the\ number\ of\ \ catches\ a\ dredge$$

$$W_i : total\ biomass\ (gram)\ a\ dredge$$

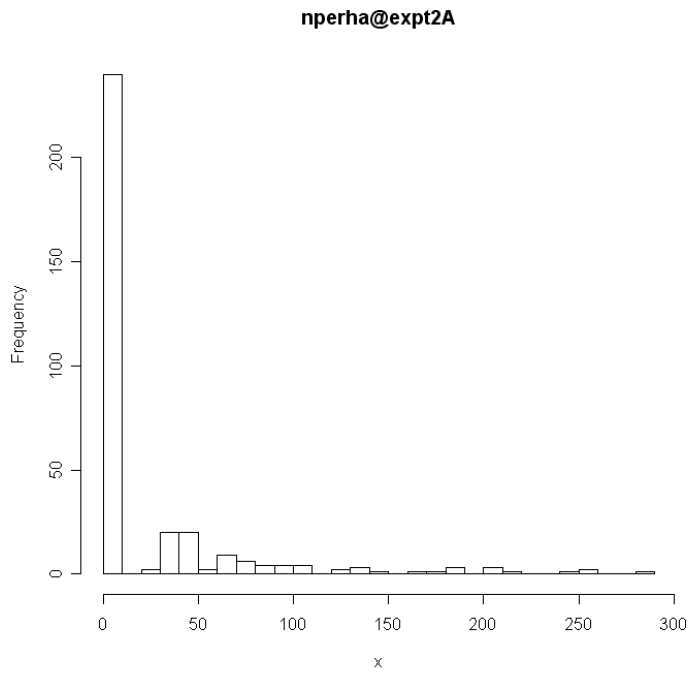
$$\alpha_i\ : swept\ area\ of\ \ the\ dredge$$

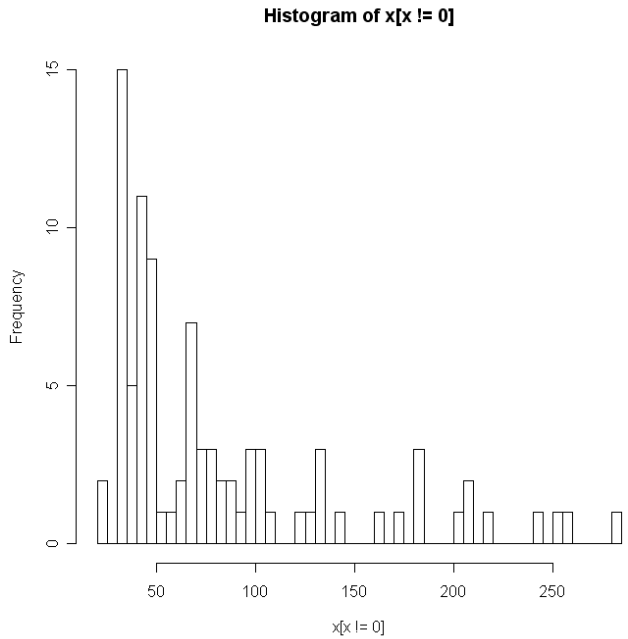
Ecology

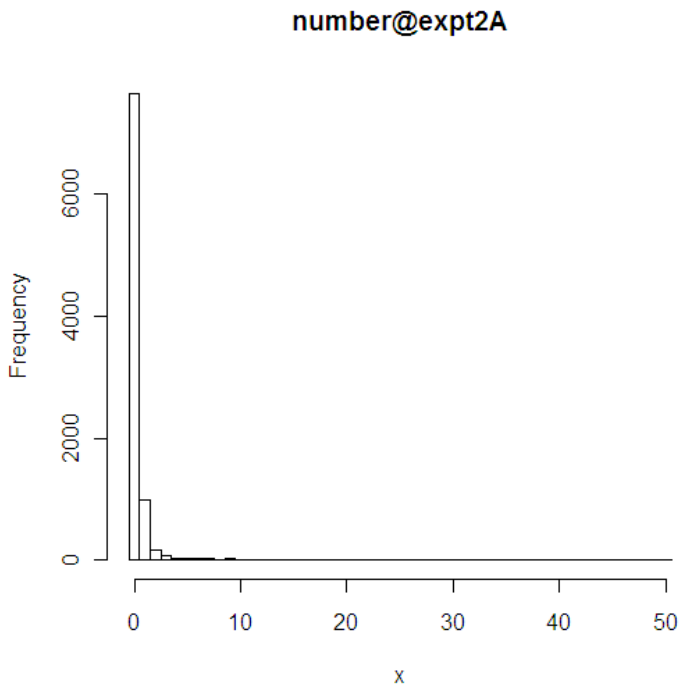
$$number:\ N_i,\ i = 1, 2, ..., k$$

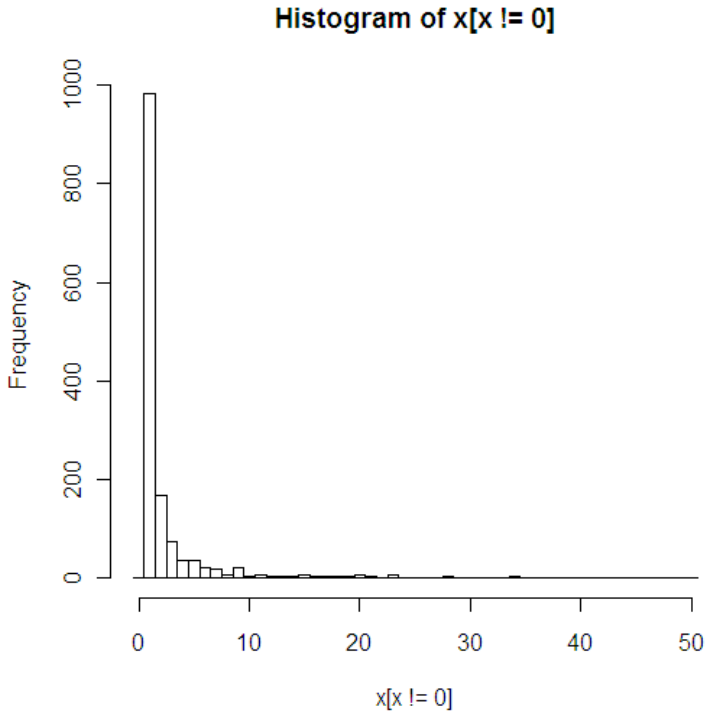
$$biomass\_g : W_i,\ i = 1, 2, ..., k$$

$$\alpha_i,\ i = 1, 2, ..., k\ :\ attribute\ of\ \ each\ dredge$$


nperha@expt2A
Frequency
0
50
100
150
200
0
50
100
150
200
250
300
x


Histogram of x[x != 0]
Frequency
0
5
10
15
50
100
150
200
250
x[x != 0]


number@expt2A
Frequency
0
2000
4000
6000
0
10
20
30
40
50
x


Histogram of x[x != 0]
Frequency
0
200
400
600
800
1000
0
10
20
30
40
50
x[x != 0]

Models for the Number of catches

Poisson

$$N_i \sim Po(\lambda\alpha_i),\ \ i = 1, 2, ..., k$$

Negative Binomial

$$N_i \sim NB_N(\alpha_i, p)$$

$$\Leftrightarrow$$

$$N_i \sim Po(\lambda),\ \ \lambda \sim Gamma(\alpha_i,\ (1-p)/p)$$

The meaning of λ *?*

Poisson – Stopped Sum (*Generalised Poisson*)

$$N_i = X_1 + X_2 + \cdots X_{Q_i},$$

$$X_1, X_2, \cdots \sim Po(\phi),$$

$$Q_i \sim Po(\alpha_i \lambda)$$

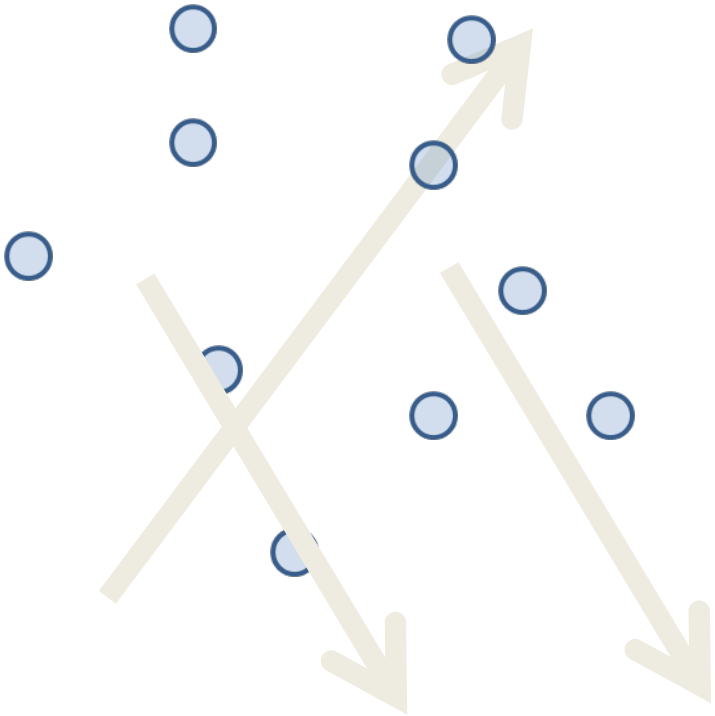
Thomas

$$N_i = X_1 + X_2 + \cdots X_{Q_i},$$

$$X_1 - 1 \geq 0, X_2 - 1 \geq 0, \cdots \sim Po(\phi),$$

$$Q_i \sim Po(\alpha_i \lambda)$$

Patch Model



$$N = X_1 + X_2 + \cdots + X_Q : \text{the number of catches a dredge}$$

$$X_1, X_2, \ldots : \text{the number of catches a patch}$$

$$Q : \text{the number of hit patches}$$

Poisson Stopped Sum Model

$$N_i = X_1 + X_2 + \cdots + X_{Q_i} : \text{the number of catches by dredge } i$$

$$X_1, X_2, \ldots : \text{the number of catches a patch} \sim Po(\phi)$$

$$Q_i : \text{the number of hits to patch} \sim Po(\alpha_i \lambda)$$

$$\phi : \text{intensity in each patch}$$

$$\lambda : \text{hit rate}$$

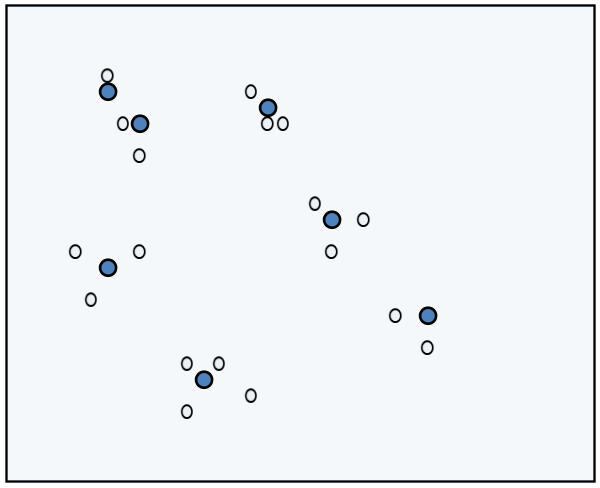
Poisson model is not a part of this model

Zeros are caused by two reasons:

$$X_1 = X_2 = \cdots = X_{Q_i} = 0 \text{ or } Q_i = 0$$

Thomas Model

- M. Thomas, *A generalization of Poisson's Binomial limit for use in ecology*, Biometrika 36, 18-25, 1949.
  - Abundance of a given species in a commonwealth of plants.
  - Tendency of the plants to cluster together



a patch: ancestor and its offsprings

$$N_i = (X_1 + 1) + (X_2 + 1) + \cdots + (X_{Q_i} + 1) : \textit{ the number of catches a dredge } i$$

$$X_1, X_2, \ldots : \textit{the number of catches a patch} \sim Po(\phi)$$

$$Q_i : \textit{the number of hit patches} \sim Po(\alpha_i \lambda)$$

Poisson is a special case

$$if\ \phi = 0\ \ then\ \ N_i \sim Po(\lambda \alpha_i)$$

Zeros are caused by a reason:

$$Q_i = 0$$

Explicit representation of probability

$$P(N_i = 0) = e^{-\alpha_i \lambda}$$

$$P(N_i = k) = \sum_{x=1}^{k} \frac{(\alpha_i \lambda)^x e^{-\alpha_i \lambda}}{x!} \frac{(\phi x)^{k-x} e^{-\phi x}}{(k-x)!} \quad for\ k = 1, 2, \ldots$$

Maximum Likelihood Estimate

Before Trawling in Experiment 2

Demospongiae (sedentary)　　普通海綿網

SPCODE	Thomas		SPCODE	Thomas	
	lambda	phi		lambda	phi
10000816	0.011542	0	10093800	0.126962	0
10000819	0	0	10098803	0.011542	0
10000820	0.011542	0	10100000	0.092336	0
10000824	0	0	10101801	0	0
10009801	0.131156	0.496031	10101802	0.011542	0
10009802	0.034626	0	10101803	0.311634	0
10009803	0.126962	0	10112800	0.011542	0
10020800	0.11542	0	10112803	0.049362	0.16911
10026801	0.184672	0	10112804	0.077755	0.039082
10029800	0.277552	0.205965	10112805	0.023084	0
10029801	0.034626	0	10114800	0	0
10066800	0	0	10119800	0.034626	0
10067013	0.023084	0	10119801	0.011542	0
10073800	0.034626	0	10119802	0	0
10086800	0.011542	0	10119804	0.011542	0
10090801	0.034626	0	99100804	0.069252	0

Gastropoda (limited movement)　腹足網

SPCODE	Thomas	
	lambda	phi
24040001	0.023084	0
24057800	0.308853	0.494821
24079002	0	0
24079800	0	0
24079801	0.011542	0
24125802	0.023084	0
24145800	0.034626	0
24145801	0.023084	0
24155800	0.034626	0
24155801	0.011542	0
24156003	0.011542	0
24165802	0.011542	0
24170800	0	0
24171001	0	0
24176800	0.023084	0
24191801	0	0
24200800	0.829764	2.547051
24200801	0.207756	0
24200802	0.034626	0
24200803	0.023084	0
24200804	0.011542	0
24200805	0.011542	0
24207003	0.011542	0
24207801	0	0
24215800	0.023084	0
24215802	0.011542	0
24215803	0.011542	0
24221800	0.011542	0
24420800	0	0

Echinoidea (limited movement)　ウニ綱

SPCODE	Thomas	
	lambda	phi
25241803	0.877193	0
25241804	0.631691	1.009876
25241809	0.011542	0
25266800	0.496888	0.440173
25290800	0	0
25304800	0.023084	0
25306800	0.218219	1.062781

Holothuroidea (very limited movement)　ナマコ綱

SPCODE	Thomas	
	lambda	phi
25400800	0	0
25400807	0.011542	0
25400810	0.023084	0
25400812	0	0
25400814	0.011542	0
25408007	0.011542	0
25408804	0.034626	0
25416029	0	0

Crustacea (mobile)

甲殻類

SPCODE	Thomas		SPCODE	Thomas	
	lambda	phi		lambda	phi
28711016	0.011542	0	28880804	0.023084	0
28711017	0	0	28880805	0.193332	0.25371
28711019	0.011542	0	28880812	0.108326	0.172038
28711044	0.080794	0	28880813	0.023577	0.468625
28711054	0.05771	0	28880815	0.011542	0
28765801	0.011542	0	28880816	0.034626	0
28765805	0.046168	0	28880817	0.023084	0
28821007	0.023084	0	28880818	0.063027	0.098761
28821015	0.011542	0	28880820	0.011542	0
28827004	0.310818	0.893847	28880824	0.011542	0
28827800	0.011542	0	28880827	0	0
28827802	0.011542	0	28895001	0.20922	0.268836
28827805	0.208388	1.824742	28895019	0.023084	0
28835804	0.08999	0.154326	28895800	0	0
28850800	0.011542	0	28895801	0.092336	0
28852800	0.069252	0	28895806	0.011542	0
28852803	0.011542	0	28895807	0.138268	0.586033
28870001	0.023084	0	28895809	0.023084	0
28870800	0.080427	0.148074	28911022	0.011542	0
28870802	0.147994	0.169845	28911026	0.011542	0
28876016	0	0	28911027	0.034626	0
28876018	0.023084	0	28911033	0	0
28876021	0.094927	0.459052	28911075	0.034626	0
28876022	0.147022	1.276661	28911081	0.011542	0
28876805	0.193287	2.16486	28911800	0.150862	0.606653
28876806	0	0	28920800	0.011542	0
28876813	0.034626	0	28920801	0.011542	0
28876816	0.02356	0.469674	28920802	0.011542	0
28880011	0.023084	0	28926006	0.011542	0
28880028	0.011542	0	28926015	0.079218	0.019893
28880038	0.023084	0	28926801	0.427054	0
28880111	0.049539	0.164936			

What to be done

- Diagnostics
  - non i.i.d.
  - PP plot ? $F_{\lambda\alpha_i,\phi}(N_i) \sim U(0,1)$
- Grouping of Species
- Effect of Location
- Effect of Trawling through parameters of Thomas distribution
- Application to Experiment 1 Data