BESS Summer School

Argo Quality Control Analysis

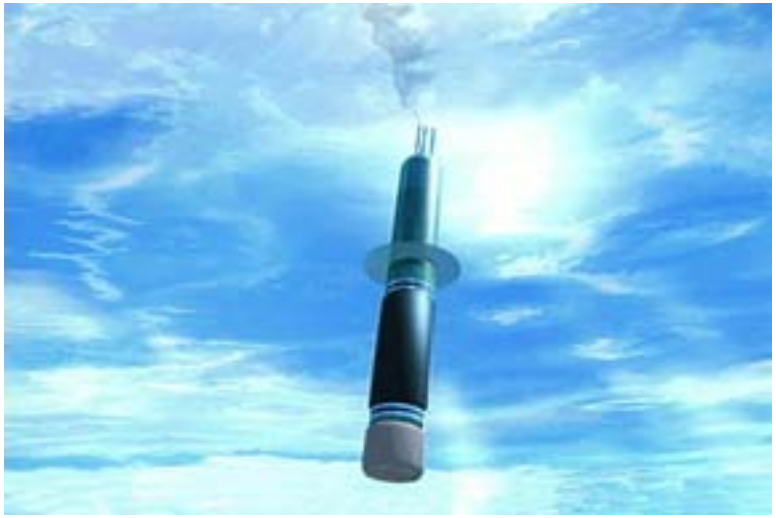
Supervisors: Robert Gurney, Francois Taiani

Analysis sub-group

- Alastair Gemmell
- Jennine Jonczyk
- Dusan Kostic
- Ulirich Weber

Visualisation sub-group

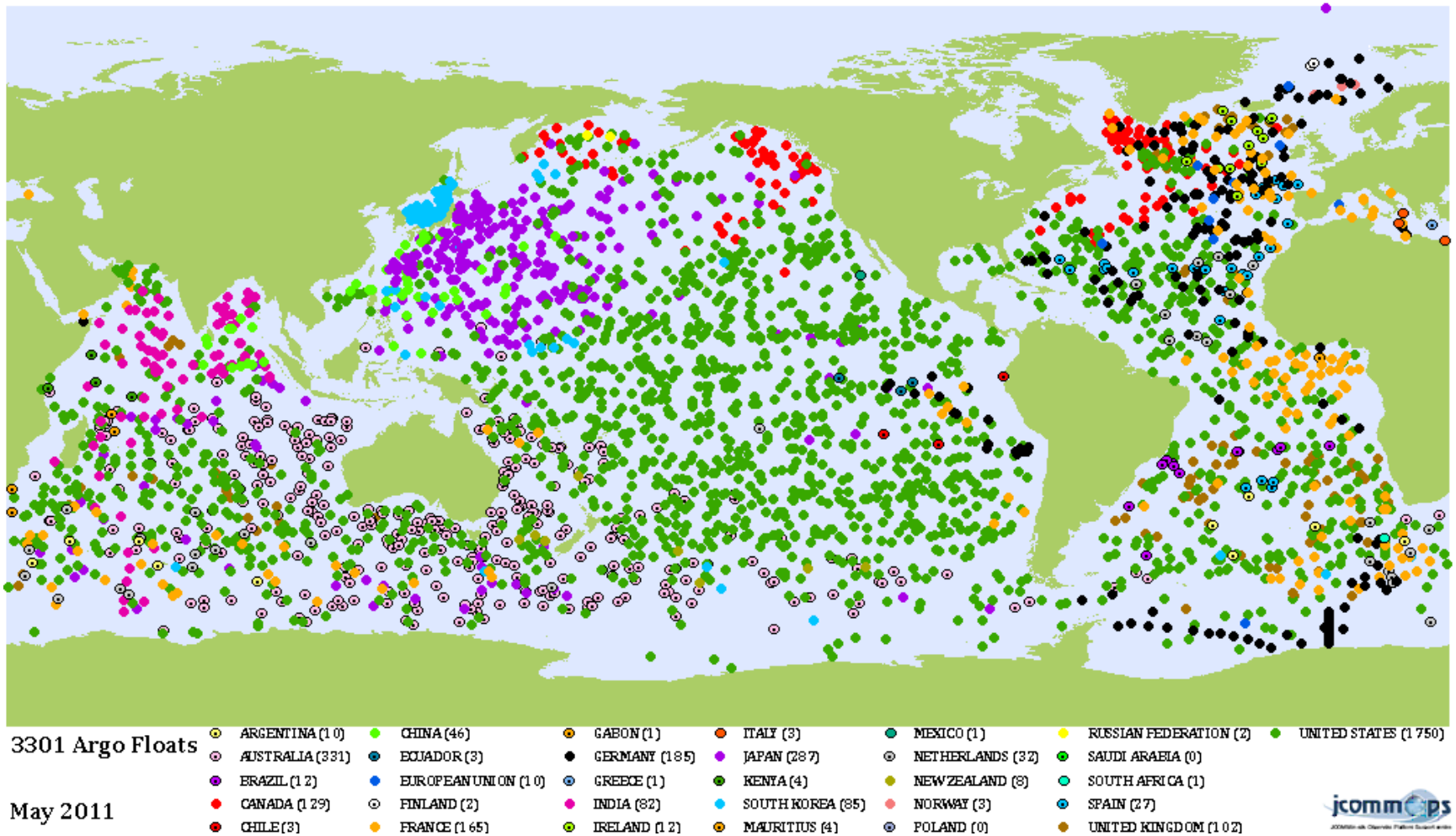
- Radoslaw Guzinski
- Fei Ma
- Eleanor Mackay
- Mark Wilkinson



What are Argo floats?

- Autonomous profiling floats. Phased in since ~2000
- Goal of 3000 floats was achieved by 2007
- 10 day cycle
  - ➔ Mostly float at (up to) 2000m depth
  - ➔ Every 10 days ascend, collecting data on Temperature and Salinity at frequent depth intervals
  - ➔ At surface transmit latest profile data via satellite to operational centres and Argo data centres
- Paradigm shift in sampling of T and S in the ocean
- Undergo real time QC ‘quick and dirty’
  - ➔ This is updated months later by the results of detailed DELAYED MODE QC. This is what we are looking at

Where are Argo floats?



The data for this project

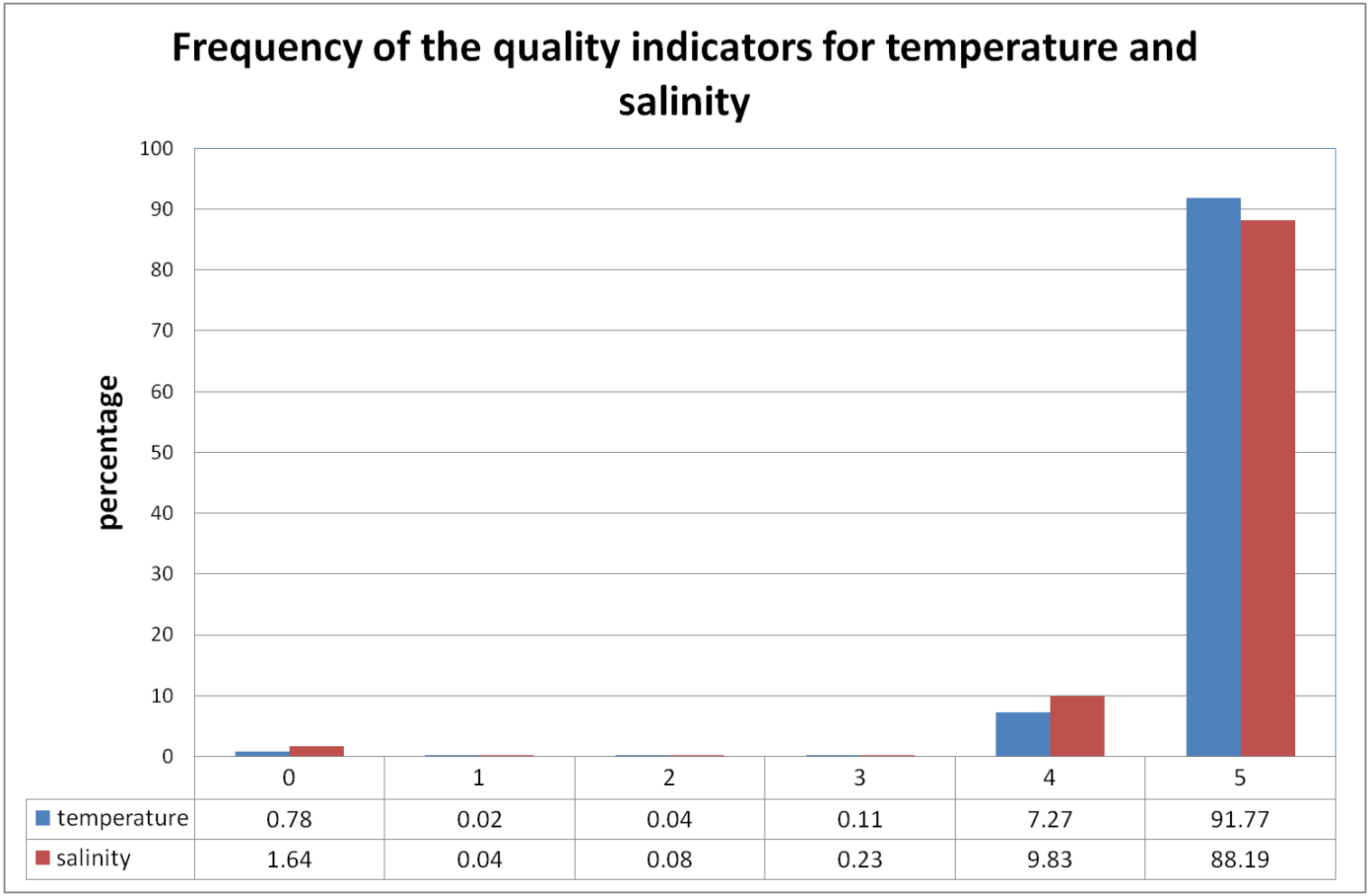
- **Argo Data**
  - Detailed delayed mode Quality Control.
  - Profiles are coded as containing percentages of **G**ood **D**ata:
    - Code **0**: GD = **0**
    - Code **1**: **0** < GD <= **25**
    - Code **2**: **25** < GD <= **50**
    - Code **3**: **50** < GD <= **75**
    - Code **4**: **75** < GD < **100**
    - Code **5**: GD =**100**

- **Accept/Reject decisions from operational centres**
  - Courtesy of Jim Cummings at US Navy via the GODAE project
  - Available for:
    - BMRC: Australia's **B**ureau of **M**eteorology **R**esearch **C**entre
    - FNMOC: US Navy's **F**leet **N**umerical **M**eteorology and **O**ceanography **C**enter
    - MEDS: Canada's **M**arine **E**nvironment **D**ata **S**ervice
    - UKMO: **UK**'s **M**et **O**ffice

Motivation

- Data assimilation extremely important for accuracy of forecasts
- Argo floats important source of these data
- Operational centres must QUICKLY decide to accept or reject recent Argo profiles
- They don’t have time to wait for detailed QC done by Argo project
- So…how accurately can they detect profiles which (upon detailed analysis) are later shown by the Argo project to contain bad data?

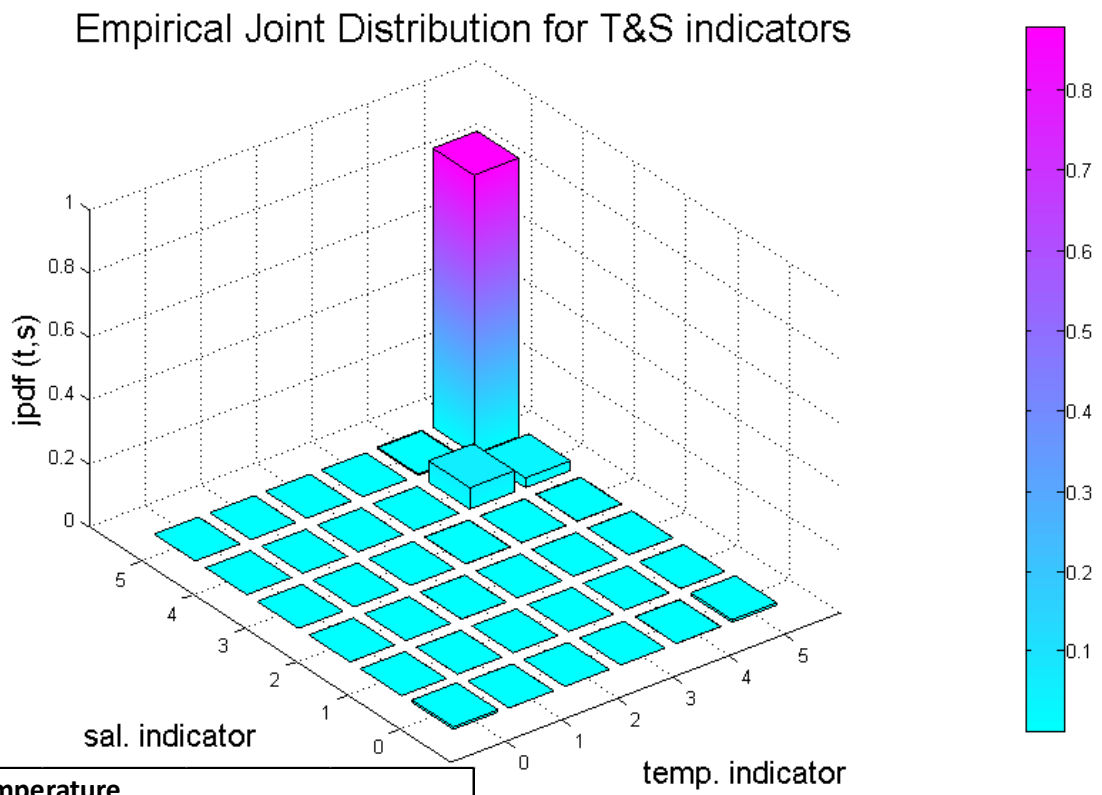
Distibution of quality indicators



	0	1	2	3	4	5
temperature	0.78	0.02	0.04	0.11	7.27	91.77
salinity	1.64	0.04	0.08	0.23	9.83	88.19

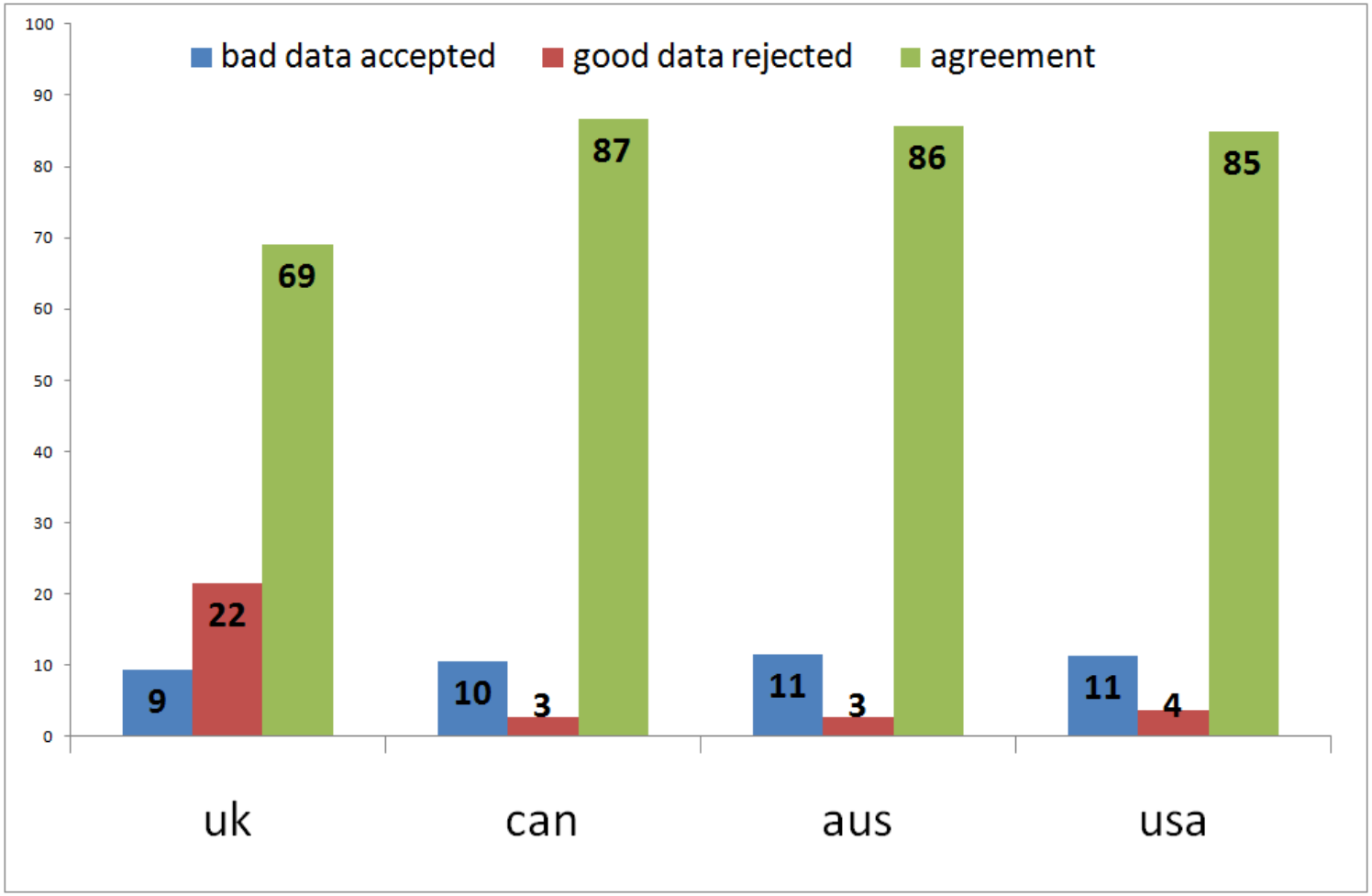
Indicators as dependent variables

		0	1
salinity	0	1155	4
	1	1	22
	2	8	0
	3	2	0
	4	9	0
	5	29	0
		No	**153350**

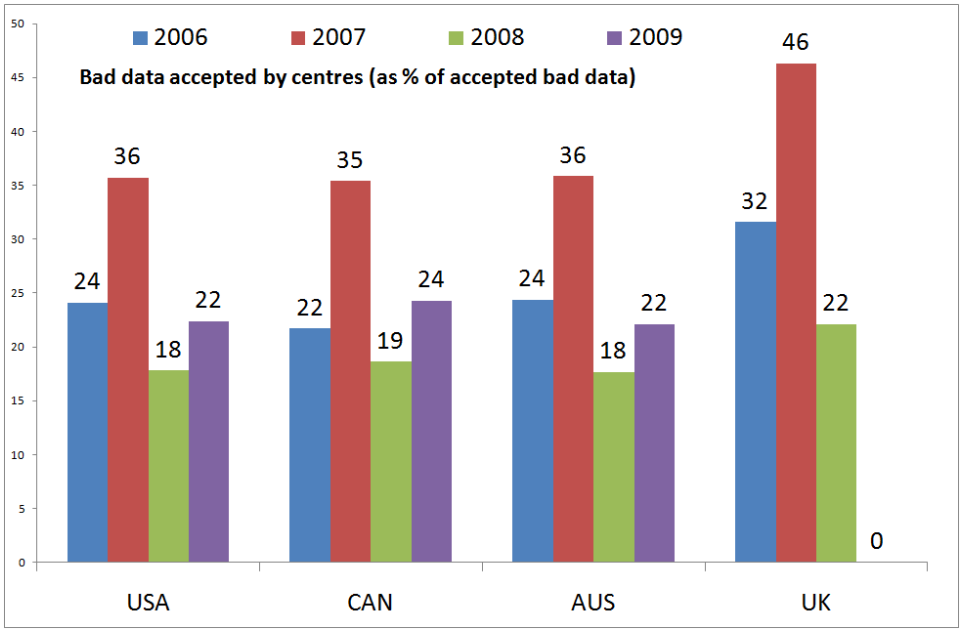


		temperature					
		0	1	2	3	4	5
salinity	0	7.53E-03	2.61E-05	2.61E-05	3.91E-05	8.35E-04	7.95E-03
	1	6.52E-06	1.43E-04	6.52E-06	0.00E+00	5.22E-05	1.76E-04
	2	5.22E-05	0.00E+00	3.65E-04	1.96E-05	5.22E-05	3.59E-04
	3	1.30E-05	0.00E+00	1.30E-05	9.72E-04	1.96E-04	1.06E-03
	4	5.87E-05	0.00E+00	6.52E-06	3.91E-05	6.71E-02	3.08E-02
	5	1.89E-04	0.00E+00	3.26E-05	3.91E-05	4.40E-03	8.77E-01
		sum	**1.0**				

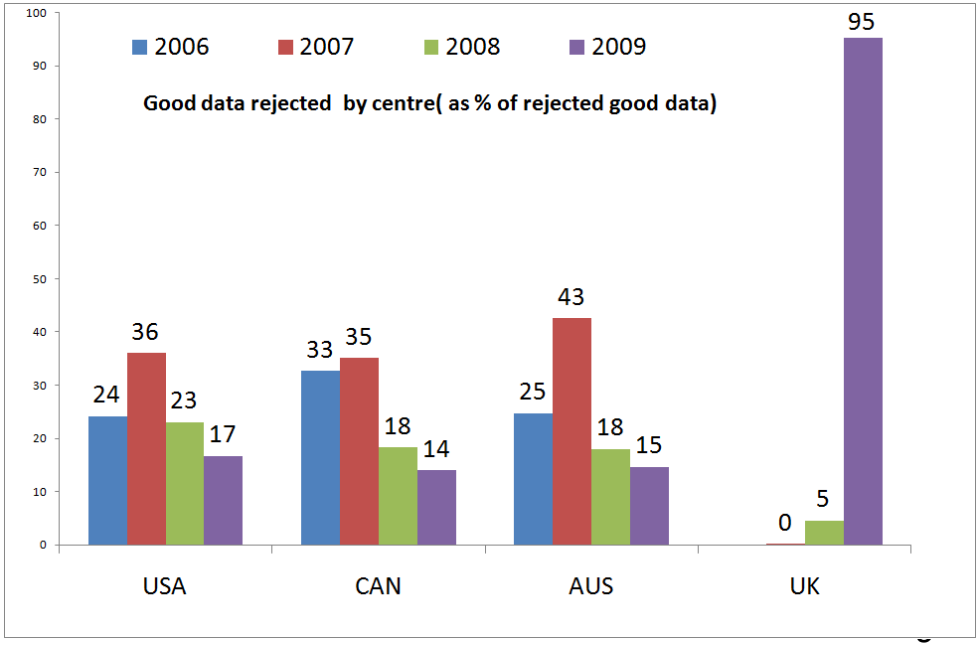
First Look.....


bad data accepted
good data rejected
agreement
100
90
80
70
60
50
40
30
20
10
0
9
22
69
10
3
87
11
3
86
11
4
85
uk
can
aus
usa

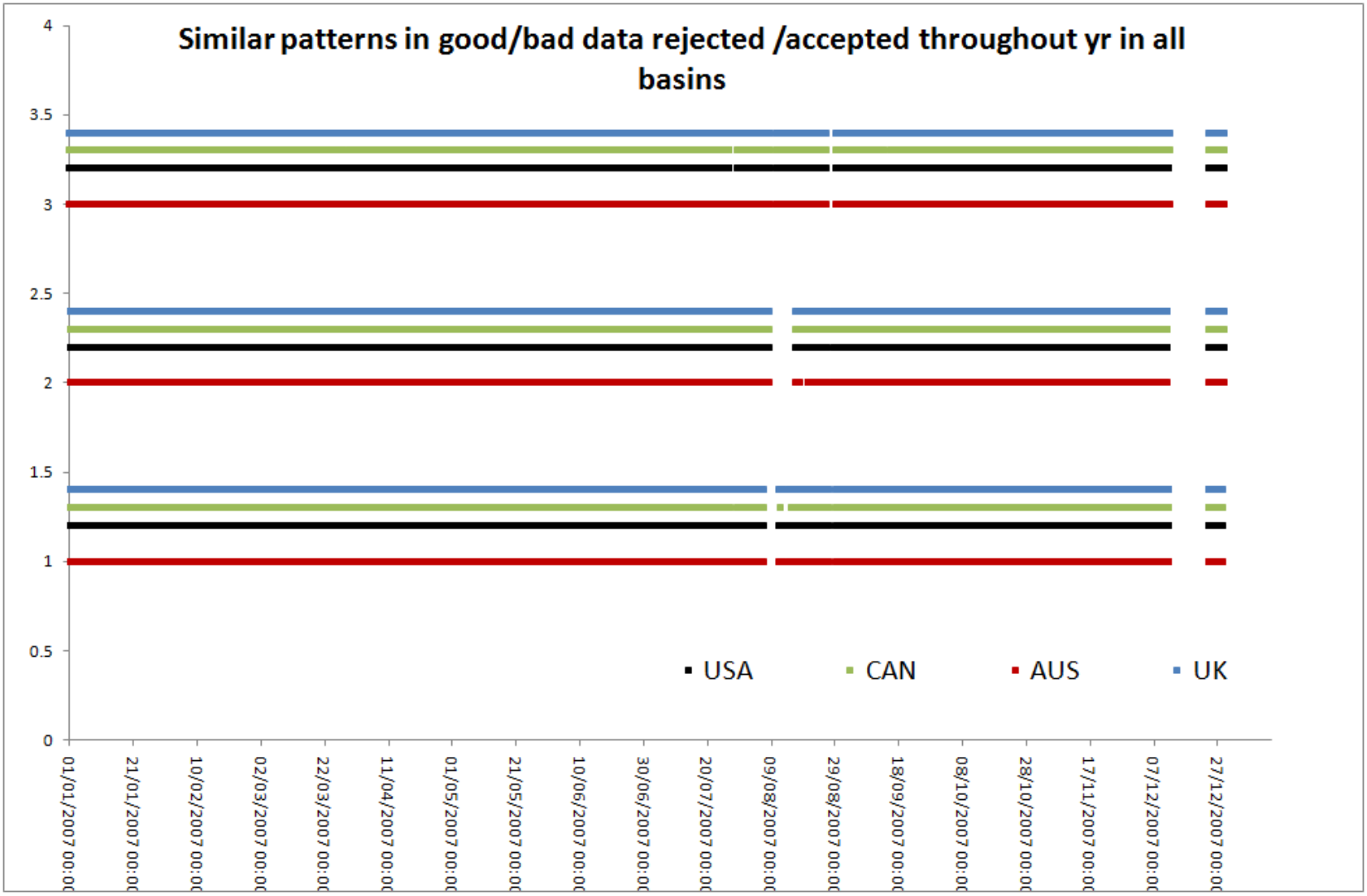
8



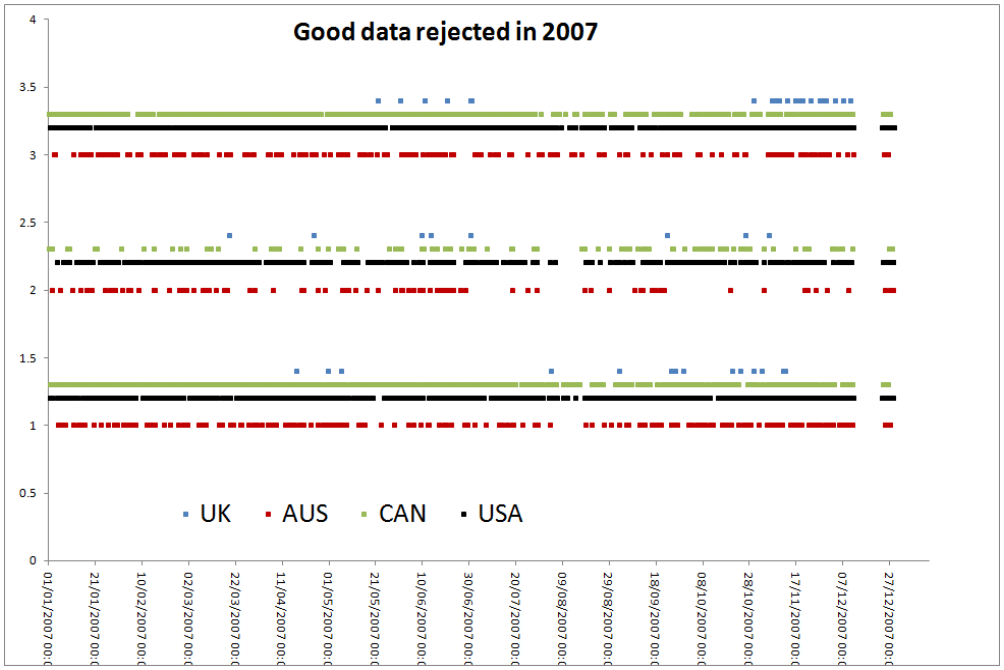
- Q: is it worse to accept bad data or reject good data?



Any particular times?



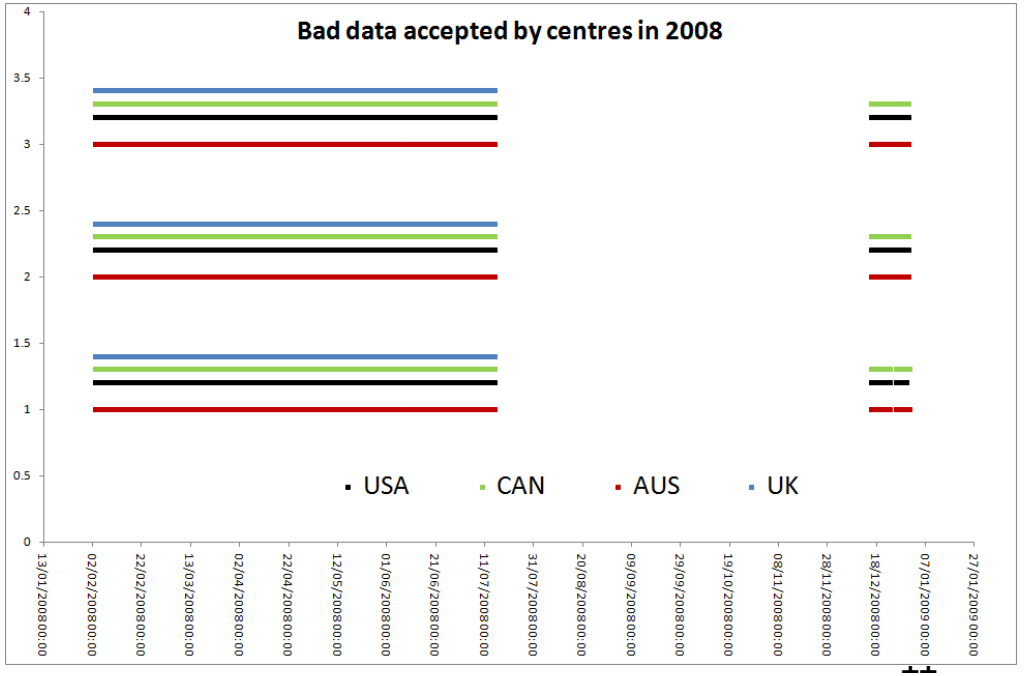
10



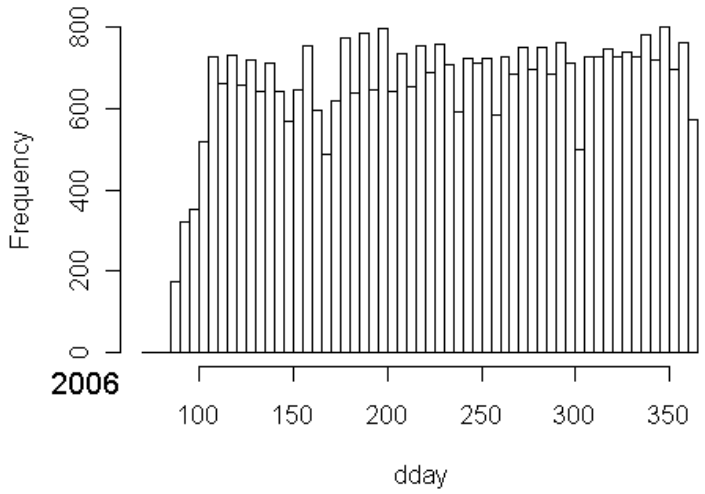
EXCEPT:

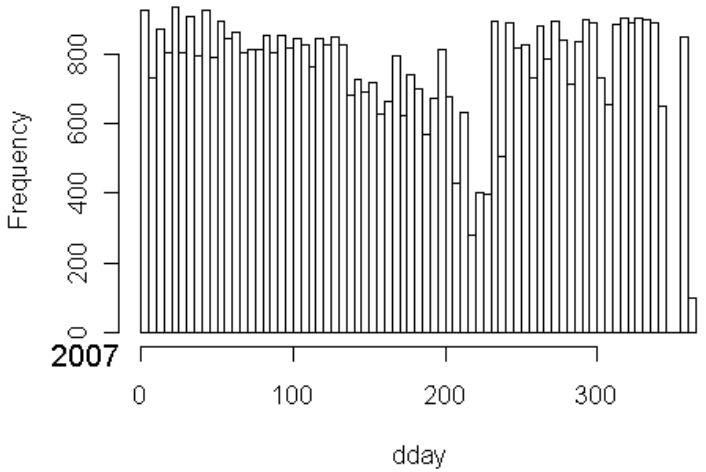
Good data rejected in 2007

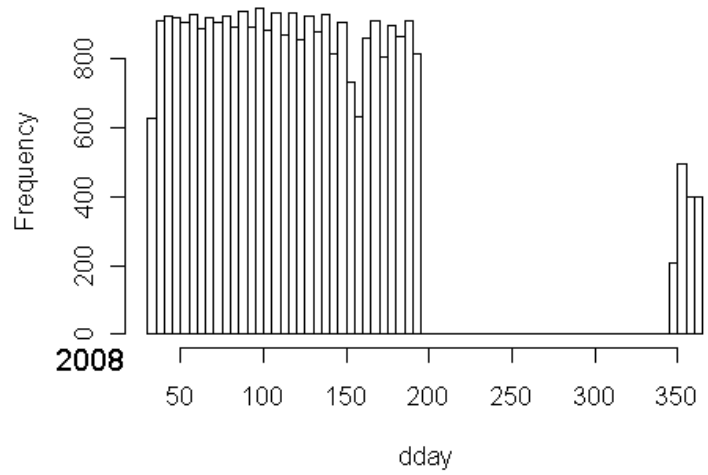
Bad data accepted in 2008

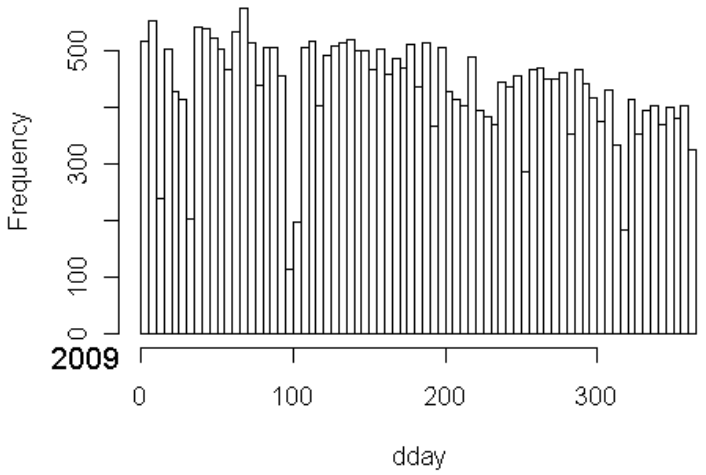


Data availability in days/year (bin=3)

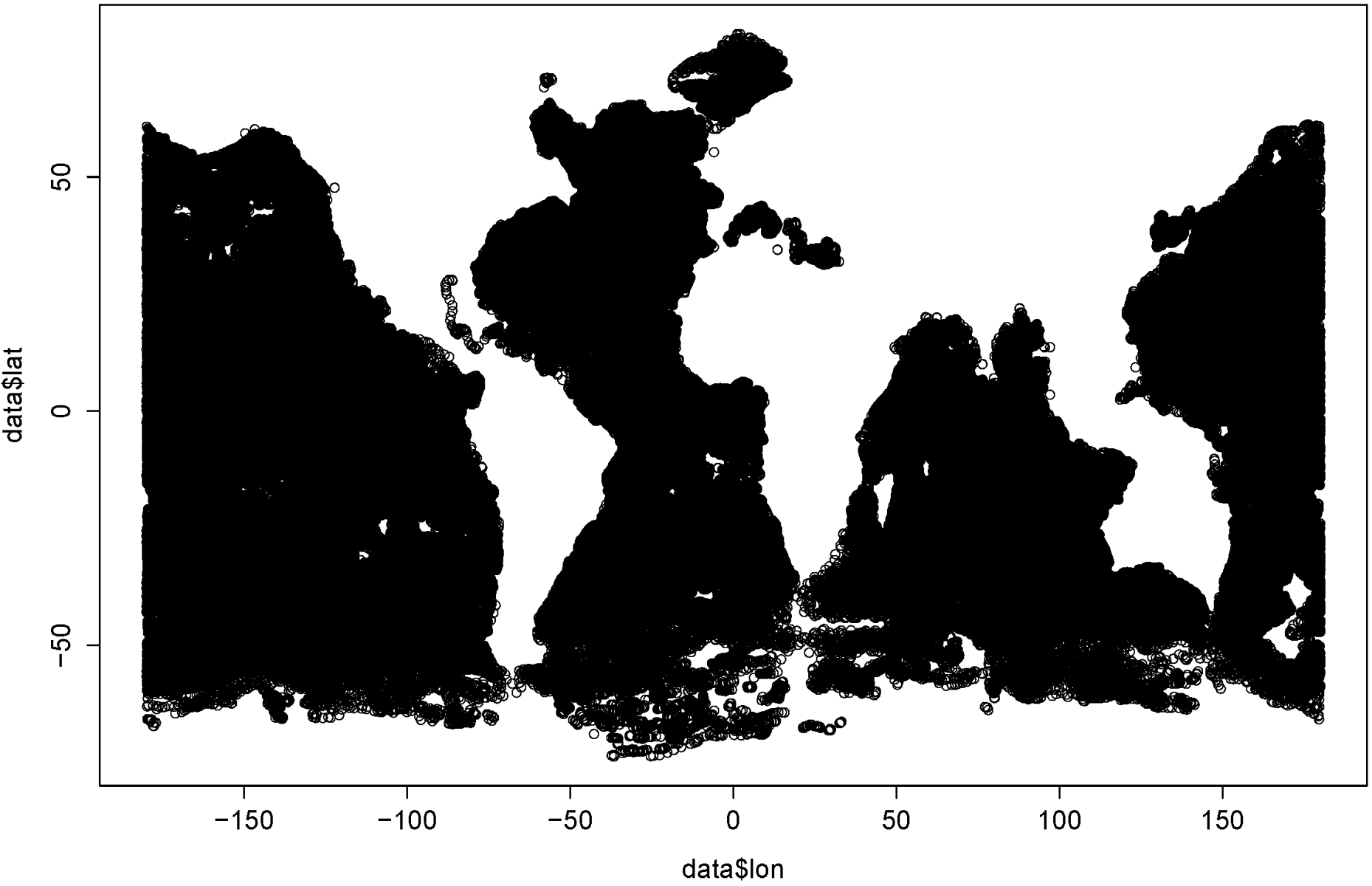




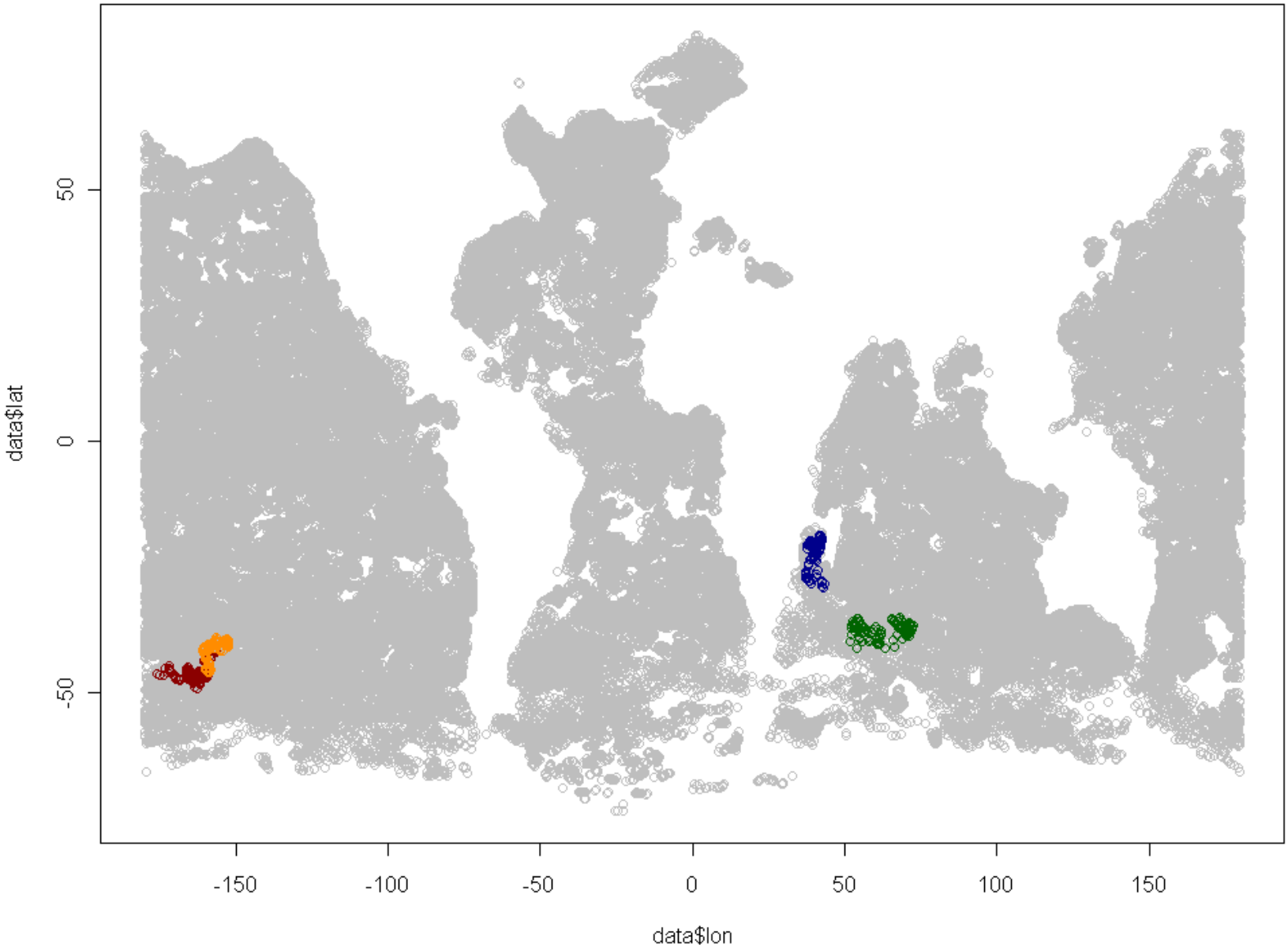




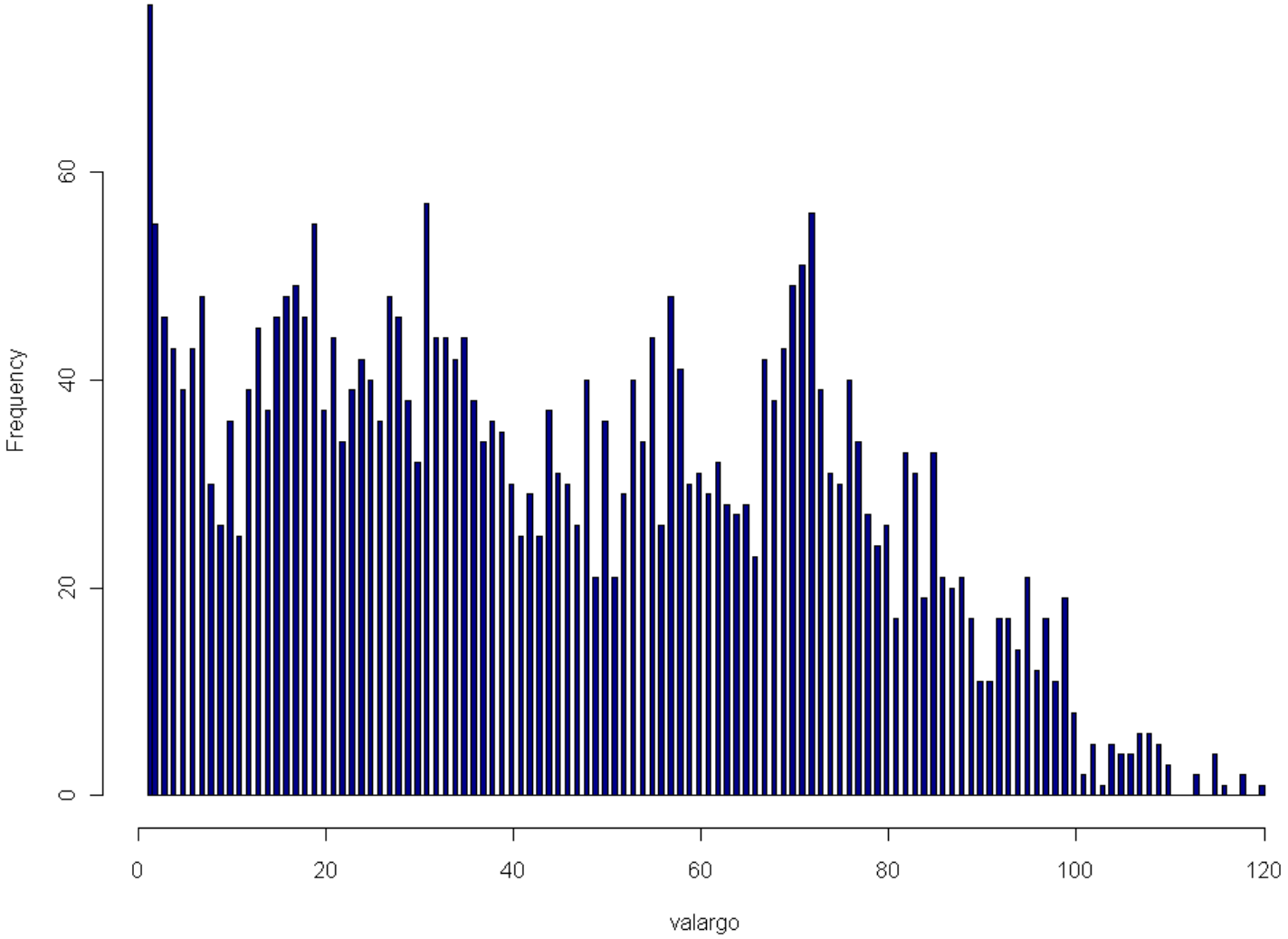
Spatial extend of QC dataset



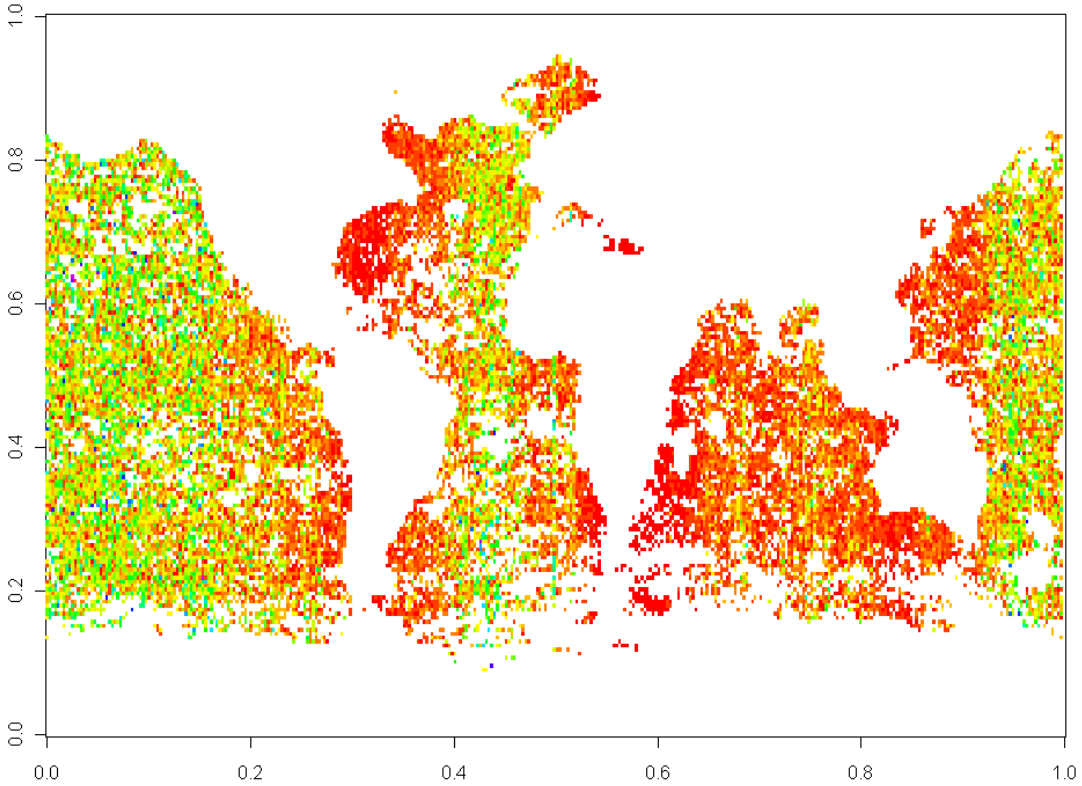
Spatial extend of all.valid.argo.float (-999 excluded) + coloured paths of sel.argo.float

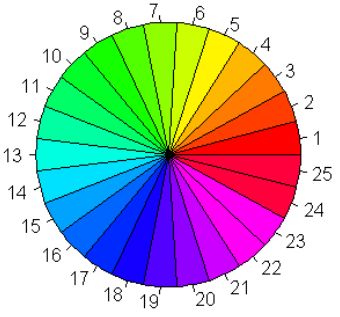


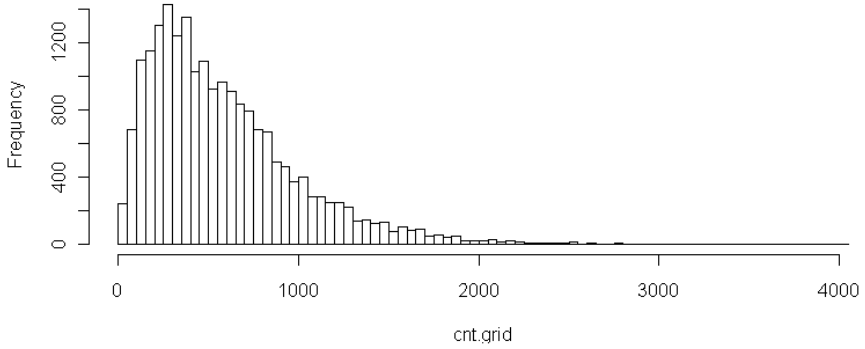
Histogram of Valid Argo.floats per meassurement



1 deg grid Cumulative Argo Counts



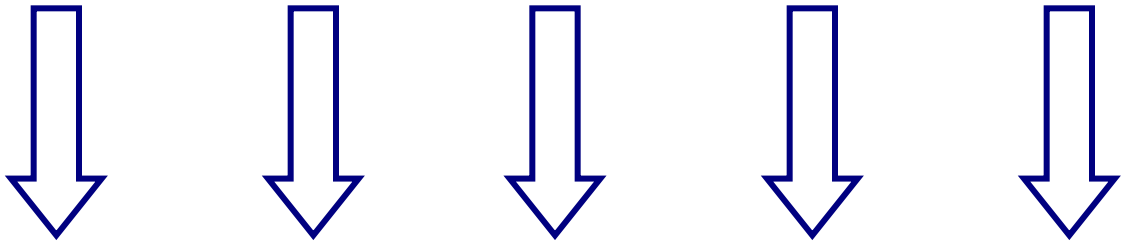




Clustering

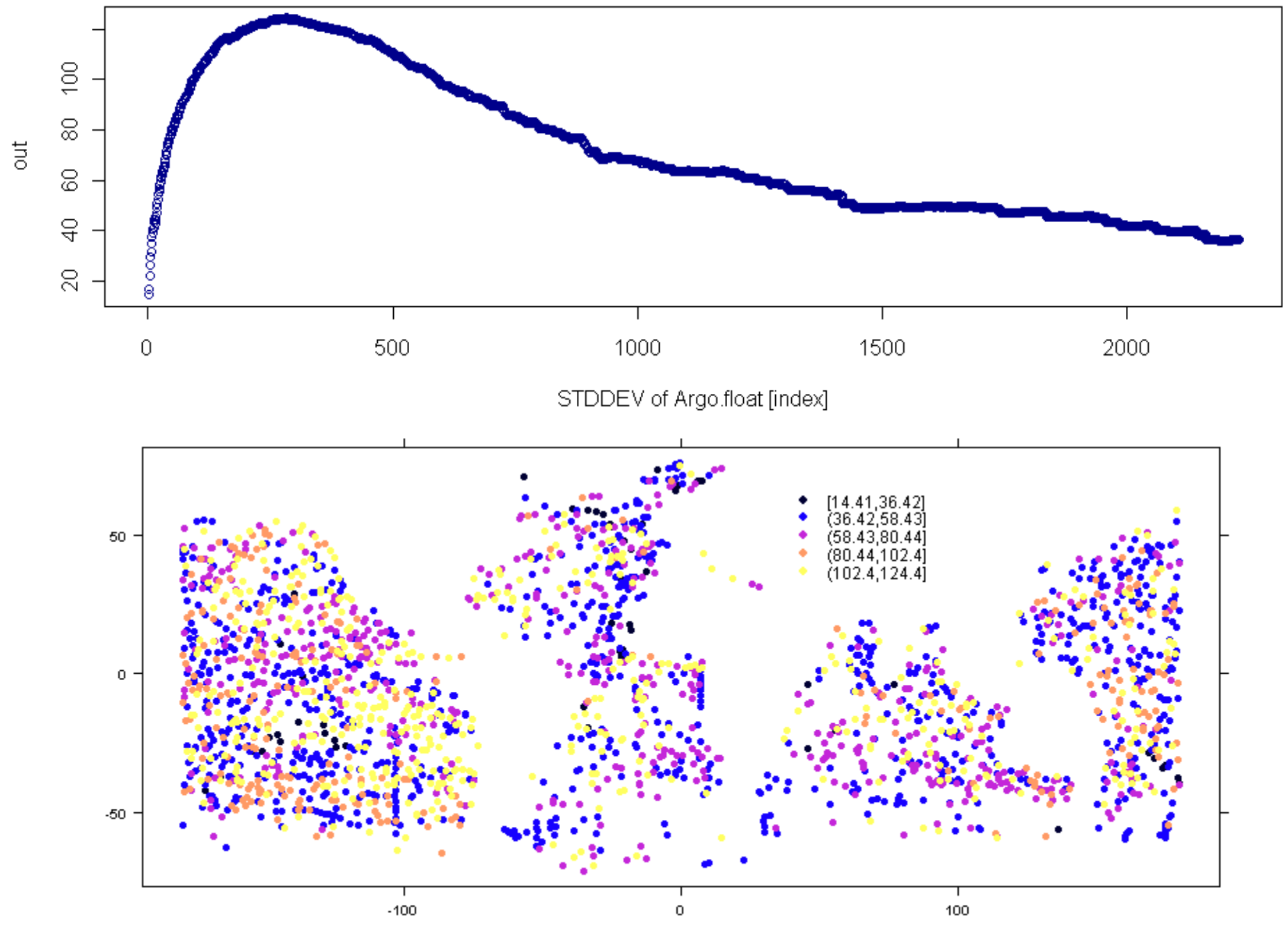
	Day 1	Day 2	…	…	Day 365
d_mode_qc	0	0	0	0	0
bmrc-qc	55	0	0	55	55
fnmoc_qc	0	0	0	0	0
meds_qc	55	50	0	55	0
ukmo_qc	0	0	0	0	0

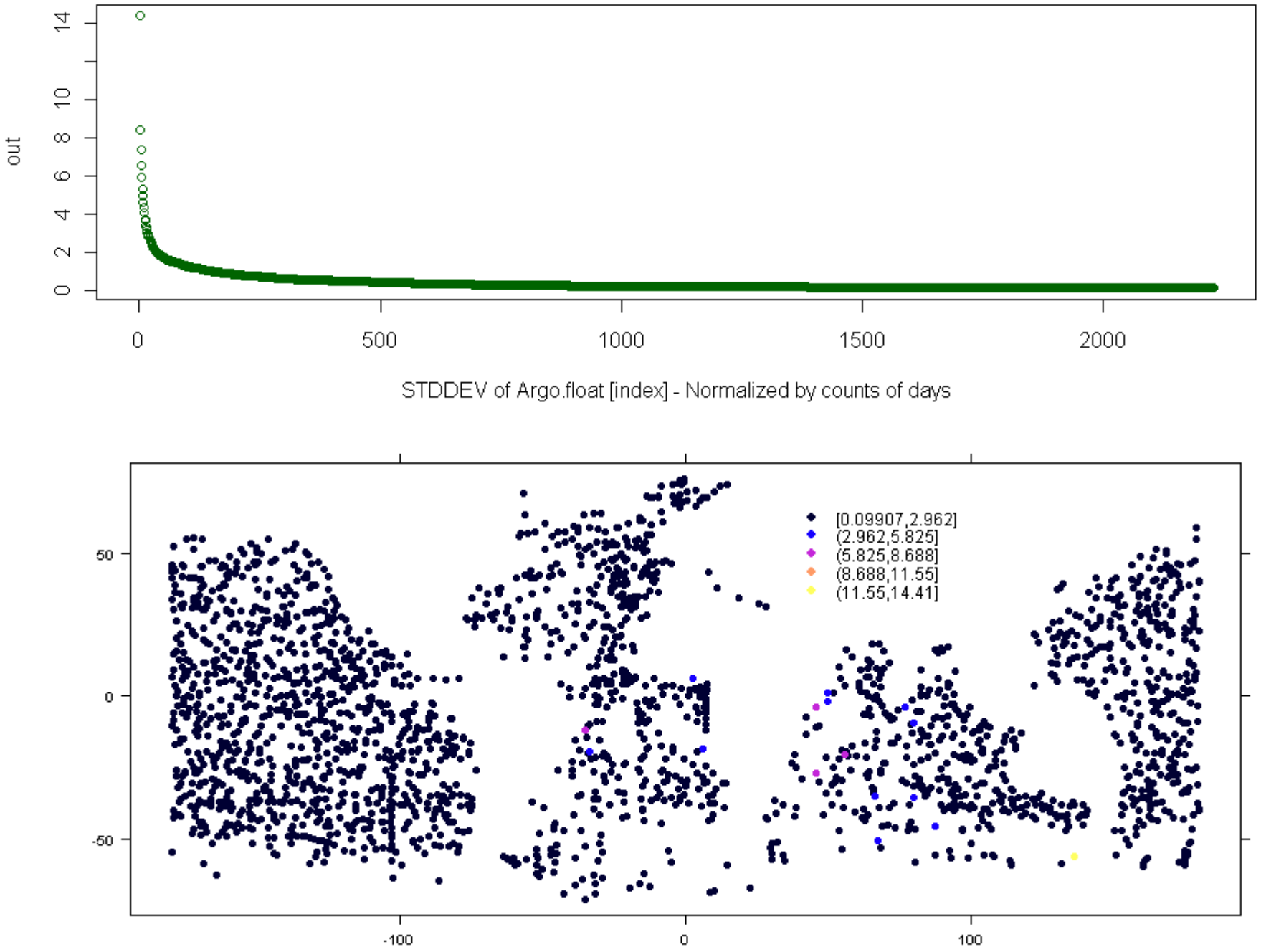
TOTAL, 2



	110	50	0	110	55

STDDEV(TOTAL, 2)

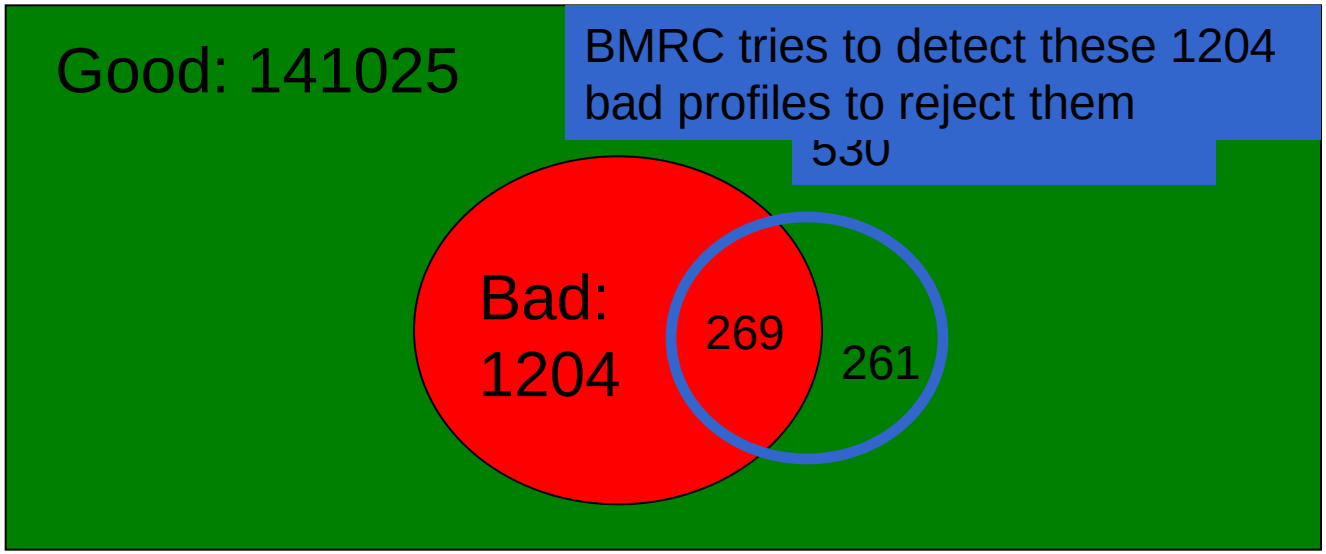

out
20
40
60
80
100
0
500
1000
1500
2000
STDDEV of Argo.float [index]
50
0
-50
-100
0
100
[14.41,36.42]
(36.42,58.43]
(58.43,80.44]
(80.44,102.4]
(102.4,124.4]


out
0
2
4
6
8
10
14
0
500
1000
1500
2000
STDDEV of Argo.float [index] - Normalized by counts of days
[0.09907,2.962]
(2.962,5.825]
(5.825,8.688]
(8.688,11.55]
(11.55,14.41]
50
0
-50
-100
0
100

Recall and Precision

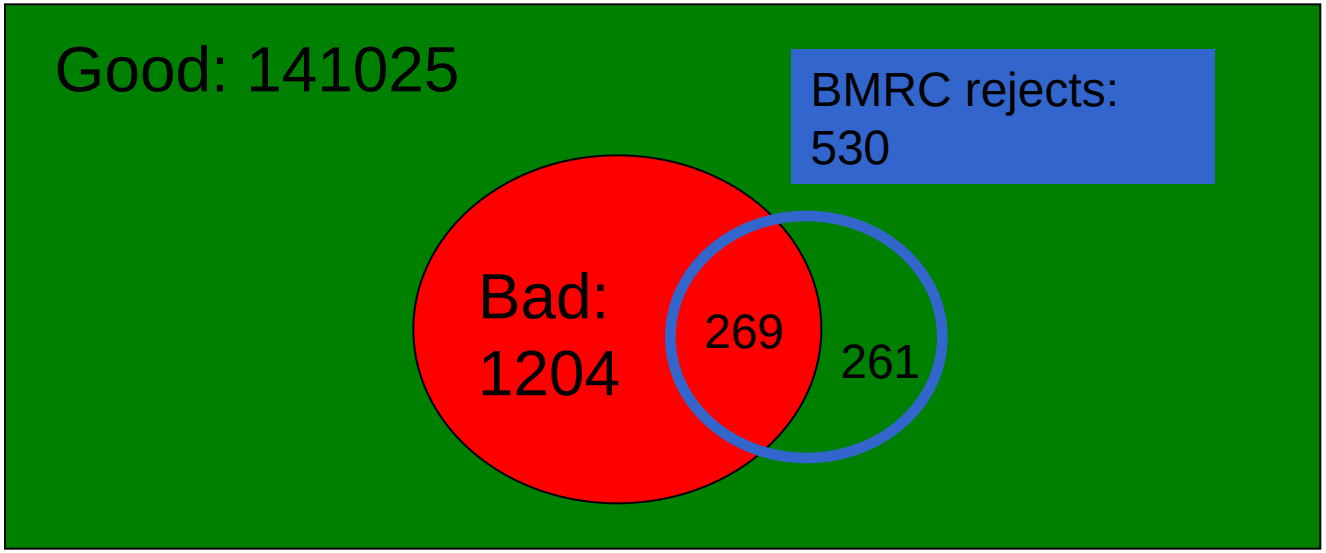
- Metrics to measure degree of success of the operational centres in detecting the bad data

Total: 142229



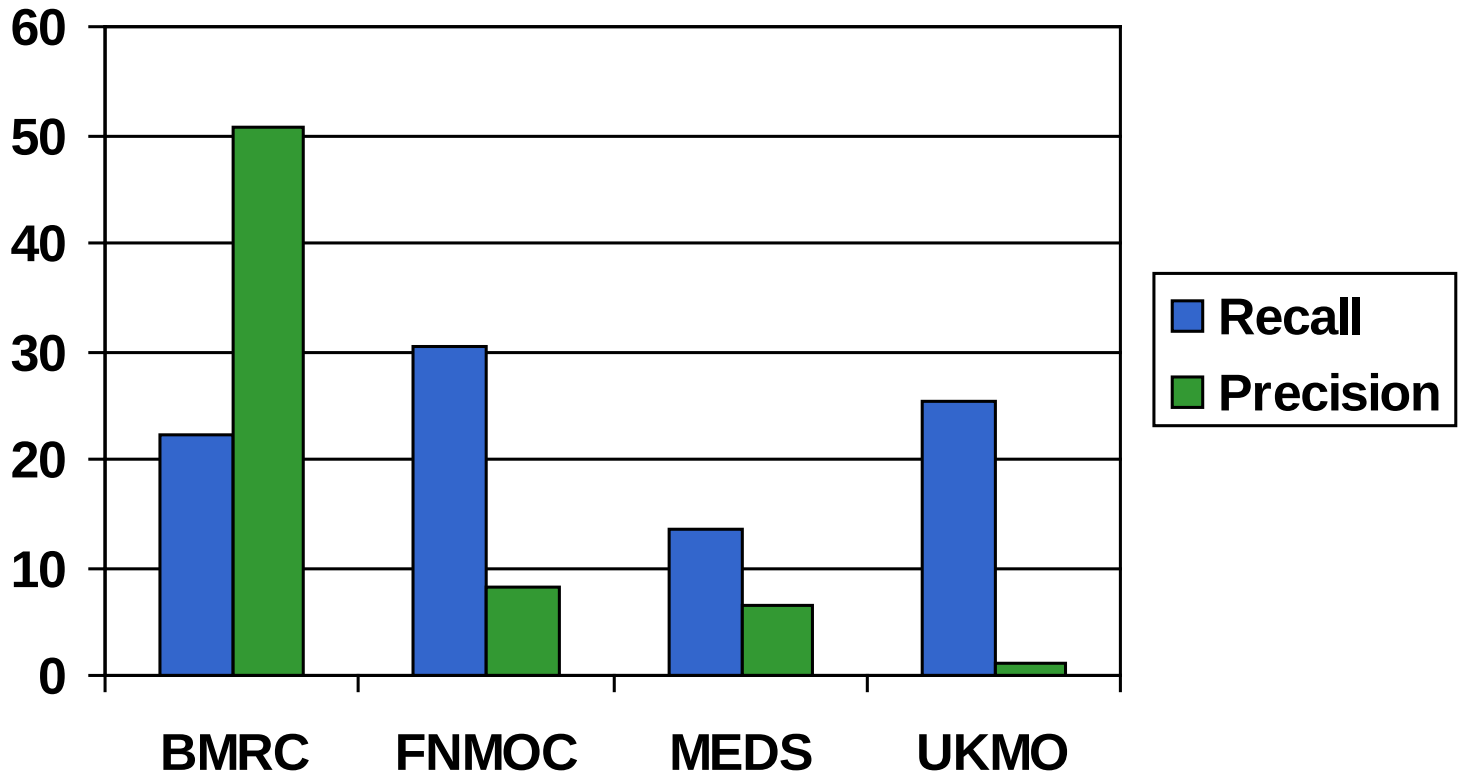
Recall and Precision

Total: 142229

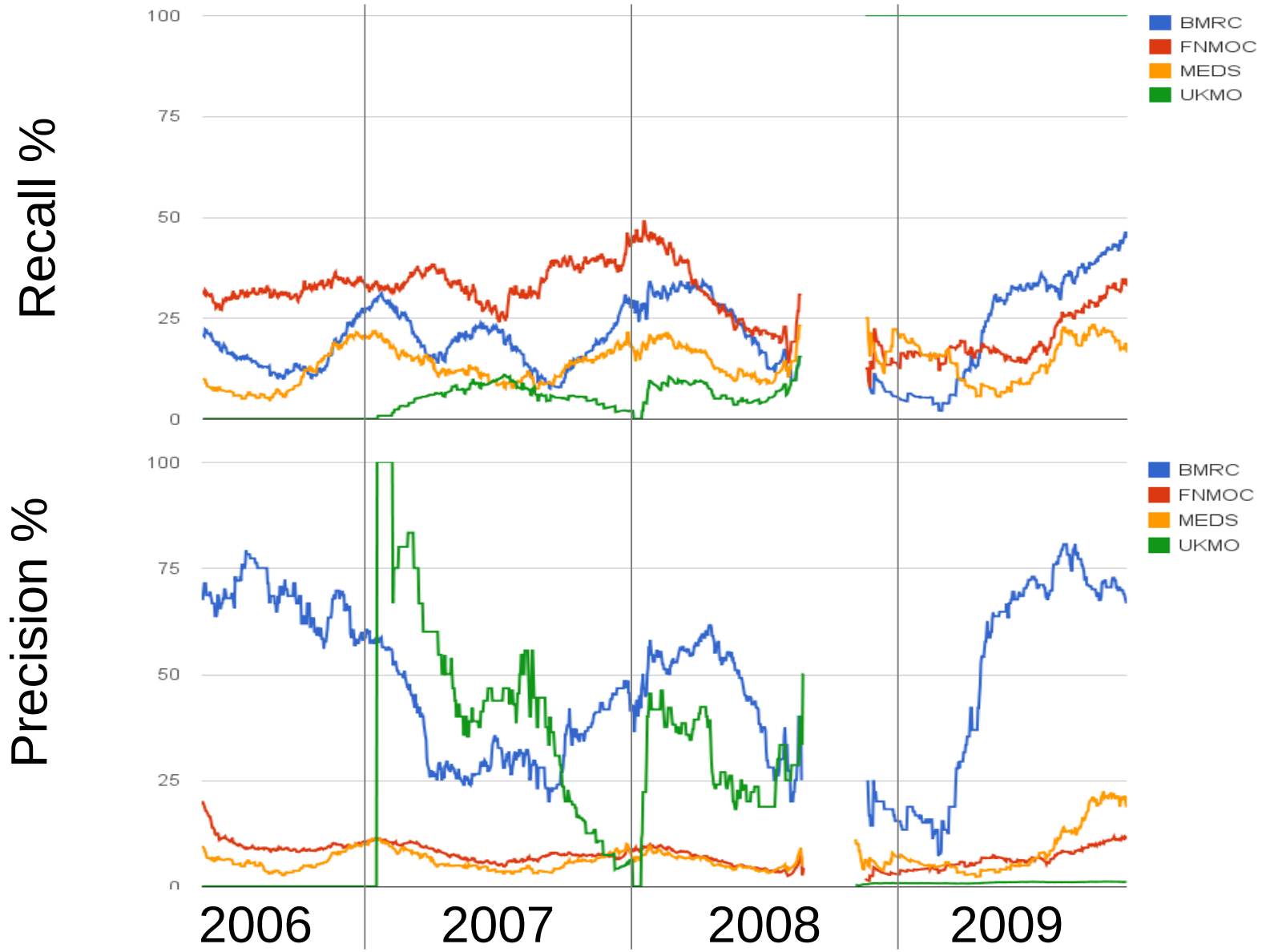


- Recall = what proportion of the **bad data** did they detect?
  - ➔ 269 / **1204** = 22.3%
- Precision = what proportion of their **rejects** were actually bad?
  - ➔ 269 / **530** = 50.8%

Recall and Precision by Centre



Recall and Precision over time



Google Earth – Salinity quality visualisation

- Salinity Quality code (QC)data was used
- Q.C. is ranked 0 to 5
- D_Mode QC was subtracted from data centre QC to produce a ranking system

Data centre wrongly rejected the data

Data centre QC matched the D_mode

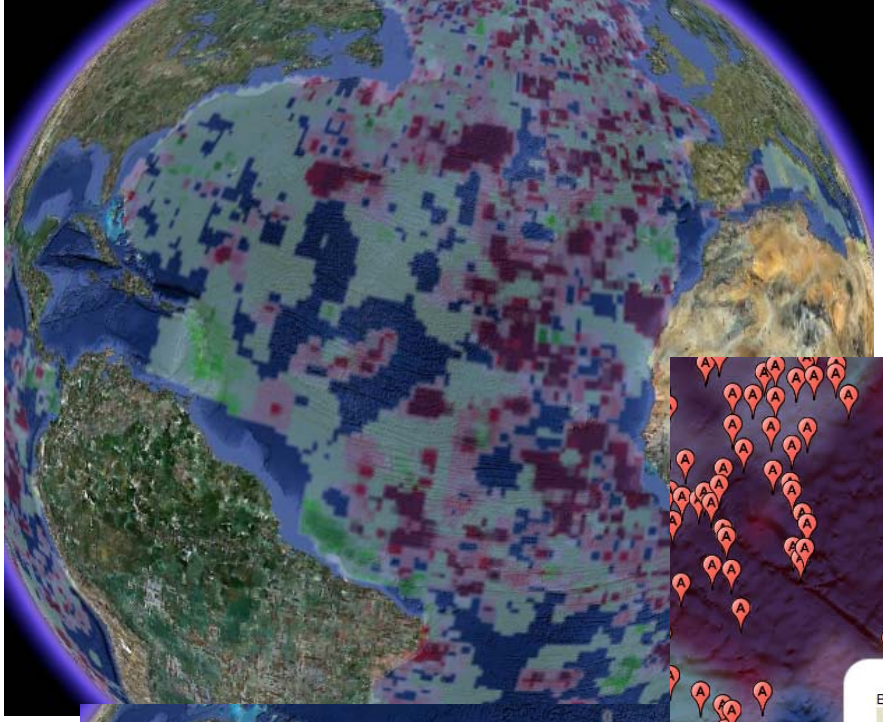
Data centre wrongly accepted the data

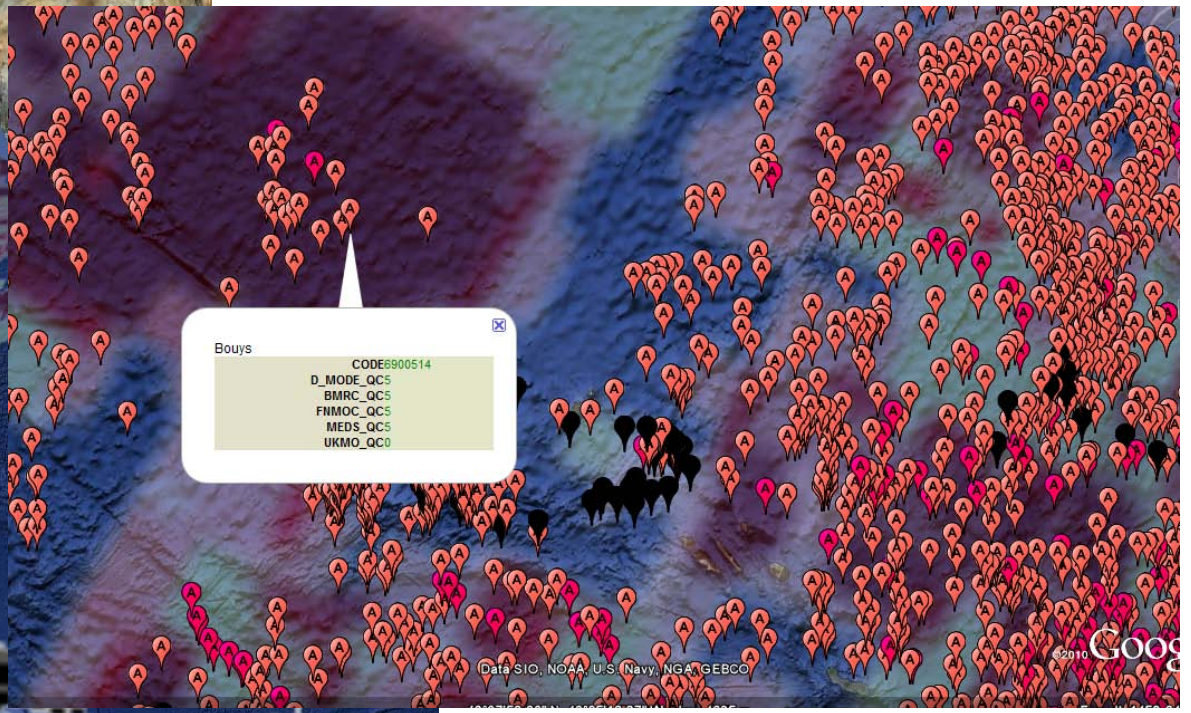
Google Earth – Salinity quality visualisation

- Raster maps were produced in ArcGIS and then exported into Google Earth. This file can easily be downloaded from the web:

http://dl.dropbox.com/u/29469014/Centres_sal.kmz

- **Inverse distance weighting** (**IDW**) is a method for multivariate interpolation




Bouys
CODE6900514
D_MODE_QC5
BMRC_QC5
FNMOC_QC5
MEDS_QC5
UKMO_QC0
Data SIO, NOAA, U.S. Navy, NGA, GEBCO

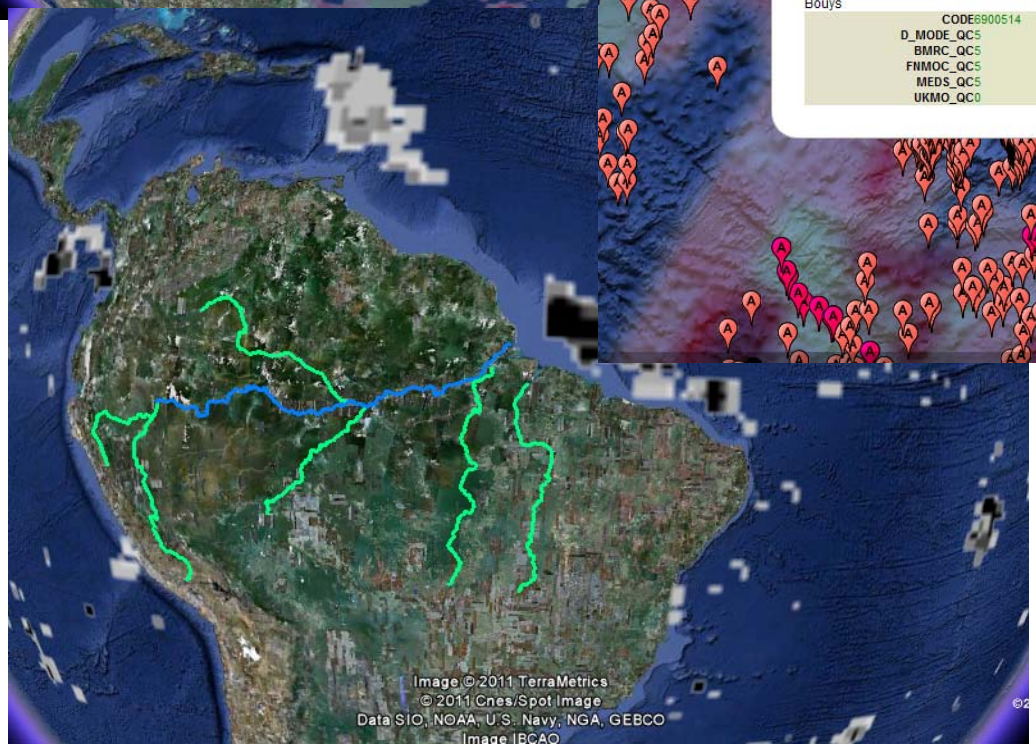


Image © 2011 TerraMetrics
© 2011 Cnes/Spot Image
Data SIO, NOAA, U.S. Navy, NGA, GEBCO
Image IBCAO

Algorithm Input / Output

- Two data sets - two values e.g:

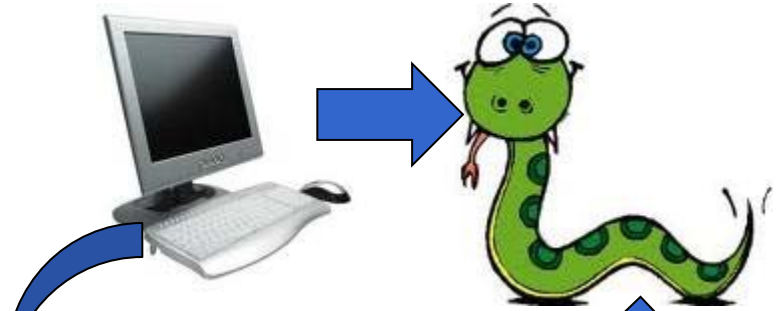
  select where d_mode_qc = good AND bmrc_qc = bad

- Other inputs:
  - ➔ Colour, presentation, temperature or salinity, comparison operator
- Output: KML file showing
  - ➔ Location of the selected floats on the globe
  - ➔ Percentage of floats that met the condition in an area

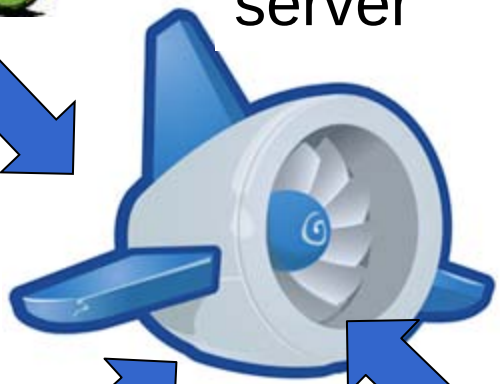
Internal Work

- Data kept in text file
- Open the file and:
  - ➔ Scan the file line by line
  - ➔ Evaluate the condition
  - ➔ IF pins:
    - Create KML pin
  - ➔ IF grid:
    - Bin the result
    - Compute statistic
    - Create Grid
  - ➔ Output KML file

Online visualisation



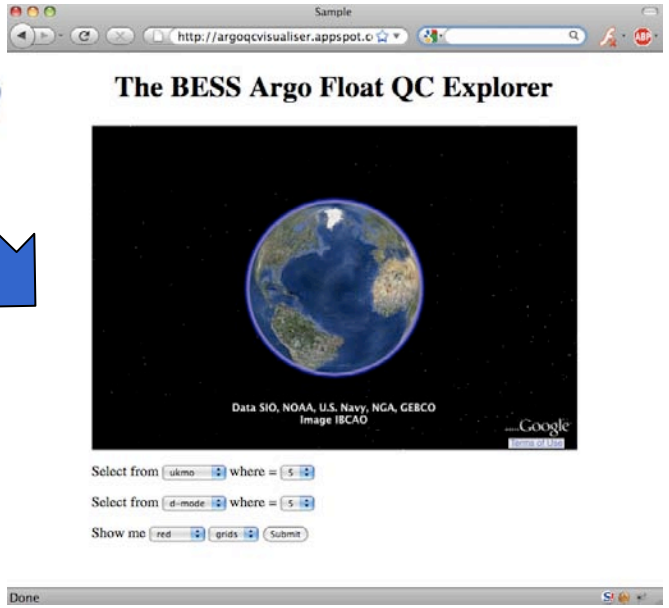
Google App Engine server



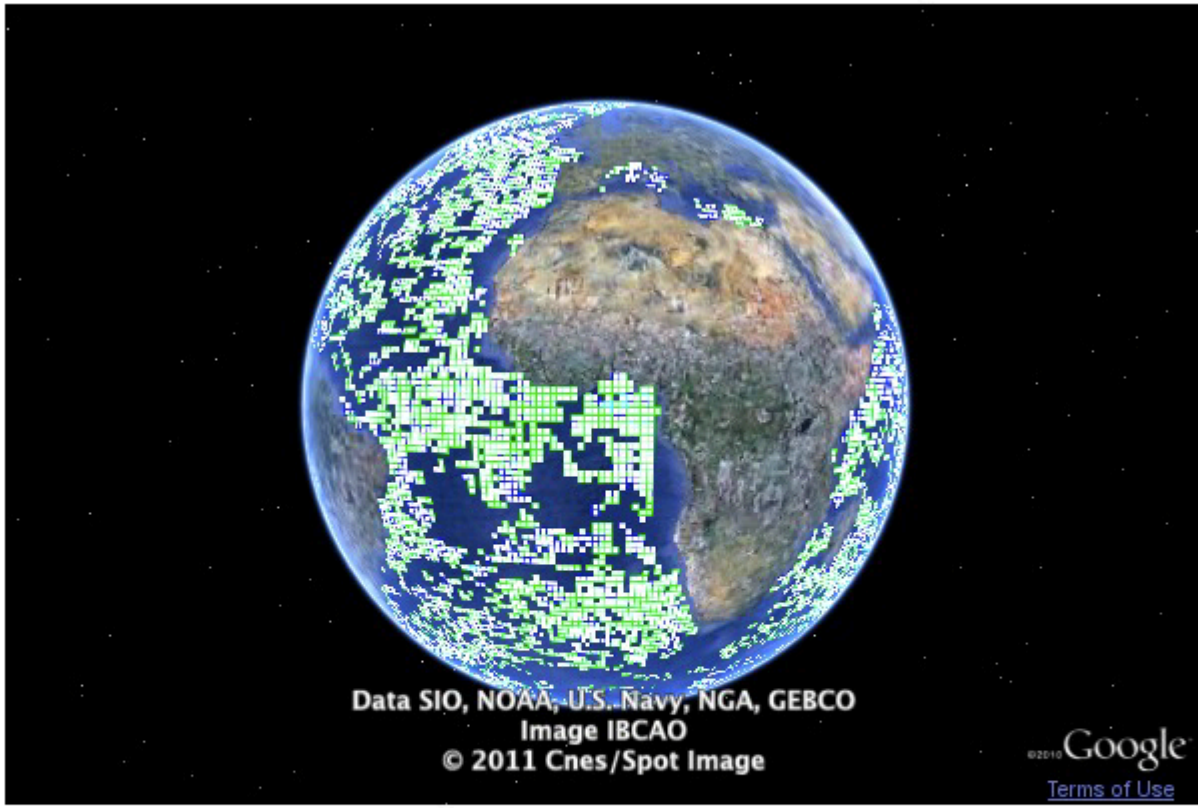
Argo Data

Data storage

User query



Query options



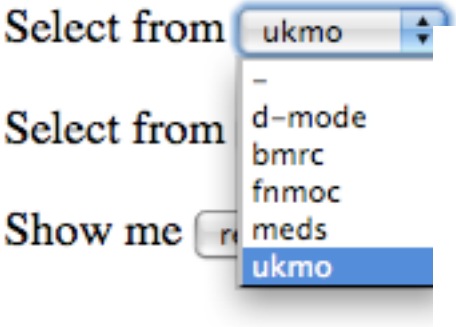
Please select the year:

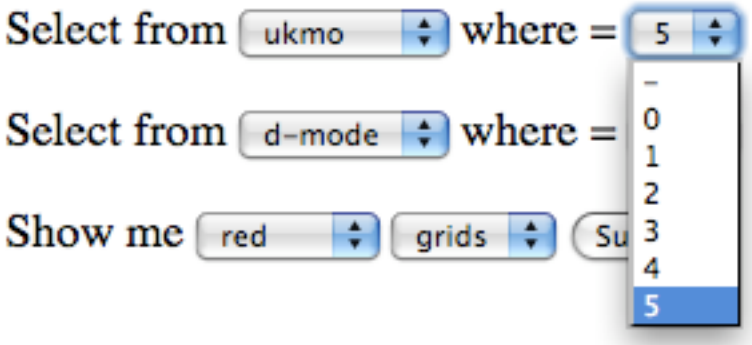
○ 2006 ○ 2007 ◉ 2008 ○ 2009

submit!

User Queries

1. Select data centre

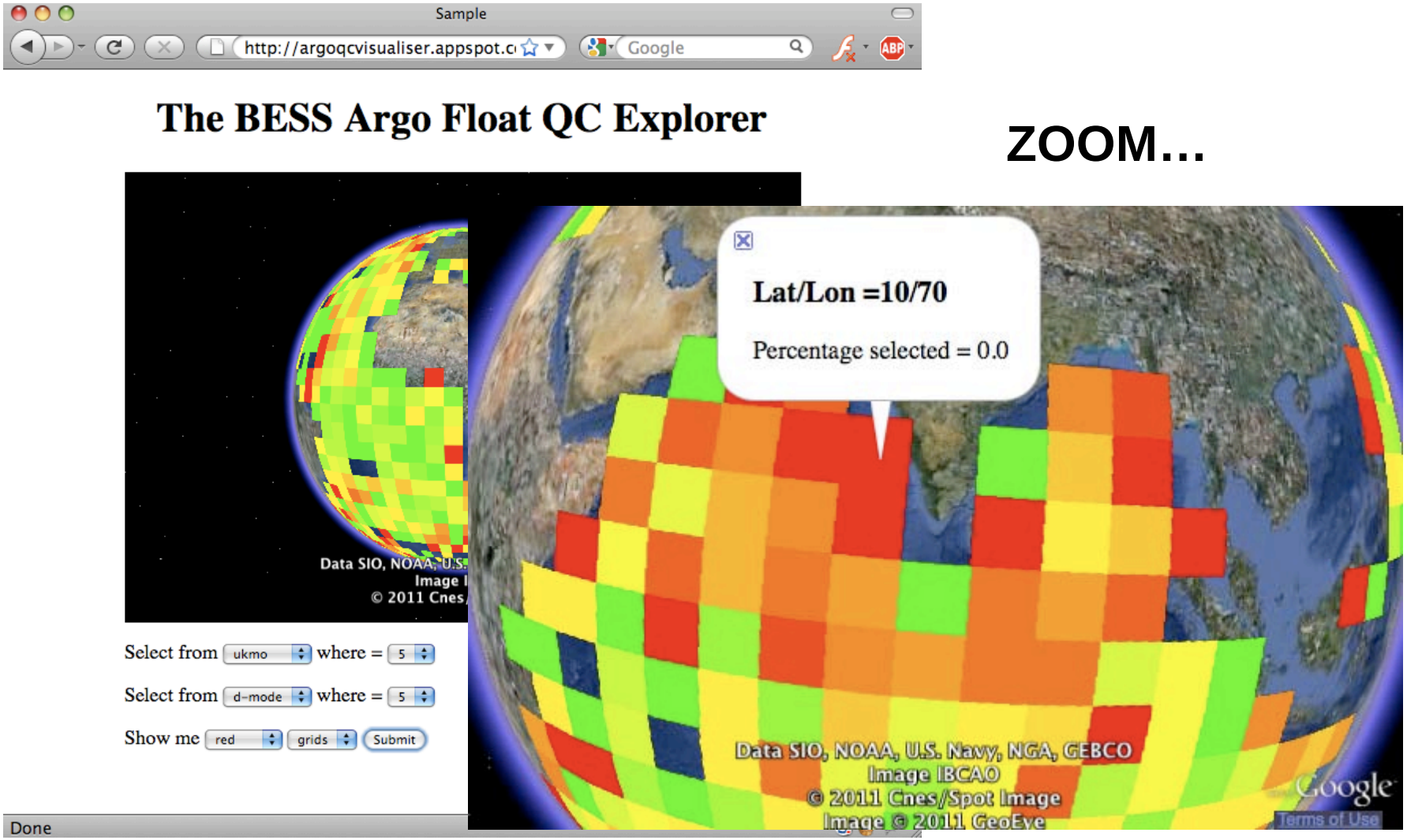




2. Select data qc value

3. Submit!

And the result…



http://argoqcvisualiser.appspot.com/

Summary 1

- All centres have relatively low recall rates for bad data
- Precision rates are much lower for MEDS and FNMOC suggesting fundamental differences in qc methodology compared to BMRC and UKMO
- Most accepted data is of good quality
- There is a high consistency between temperature and salinity qc values
- No particular time of year that causes more good or bad data

Summary 2

- ‘Spatial-temporal clustering’ may be a good tool for evaluating the consistency of argo floats in terms of data quality
- Visualisation tools such as Google Earth have the potential to provide a powerful way to visualise large spatial datasets
- Online application of visualisation tools can provide the ability for rapid and simple queries for a large number of potential users