Luigi Spezia

Biomathematics & Statistics Scotland

Aberdeen

NEW INSIGHTS INTO LONG-TERM RIVER CHEMISTRY TIME SERIES FROM PERIODIC MULTIVARIATE NORMAL HIDDEN MARKOV MODELS

joint work with Martyn Futter and Mark Brewer

Second International Workshop on Catchment Management in the Rio Cuiaba Basin

18th – 20th August 2009, Macaulay Institute, Aberdeen

Overview

NEW INSIGHTS INTO LONG-TERM RIVER CHEMISTRY TIME SERIES FROM PERIODIC MULTIVARIATE NORMAL HIDDEN MARKOV MODELS:

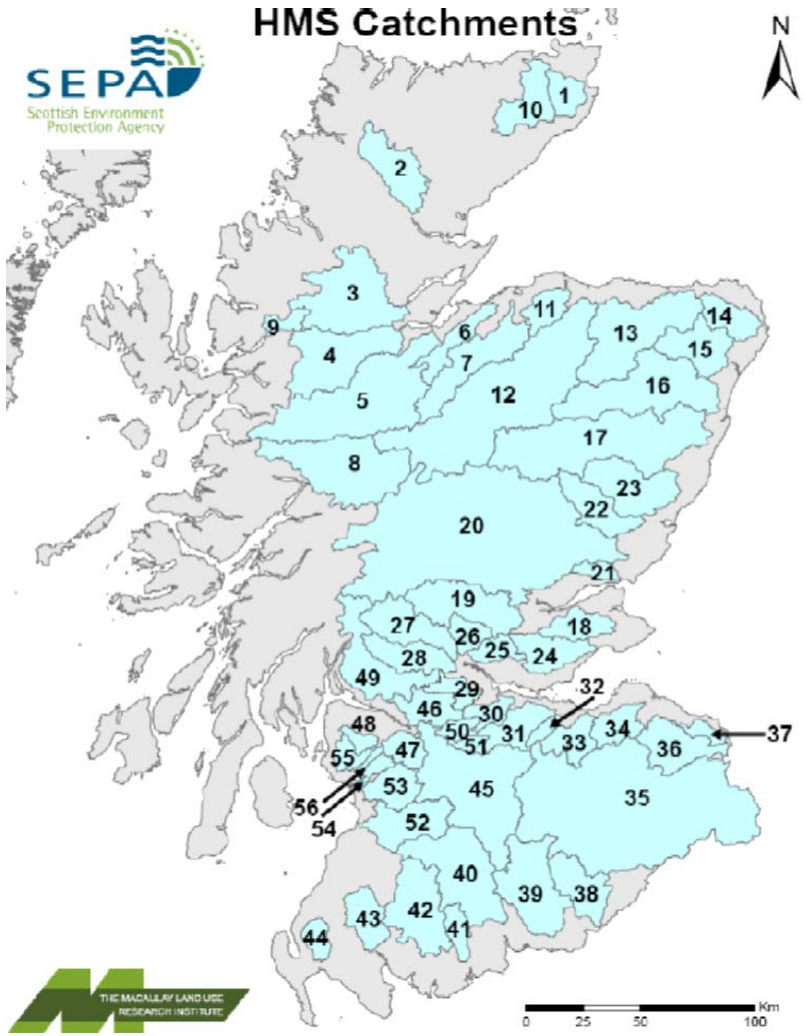
1) Data

2) Classification problem

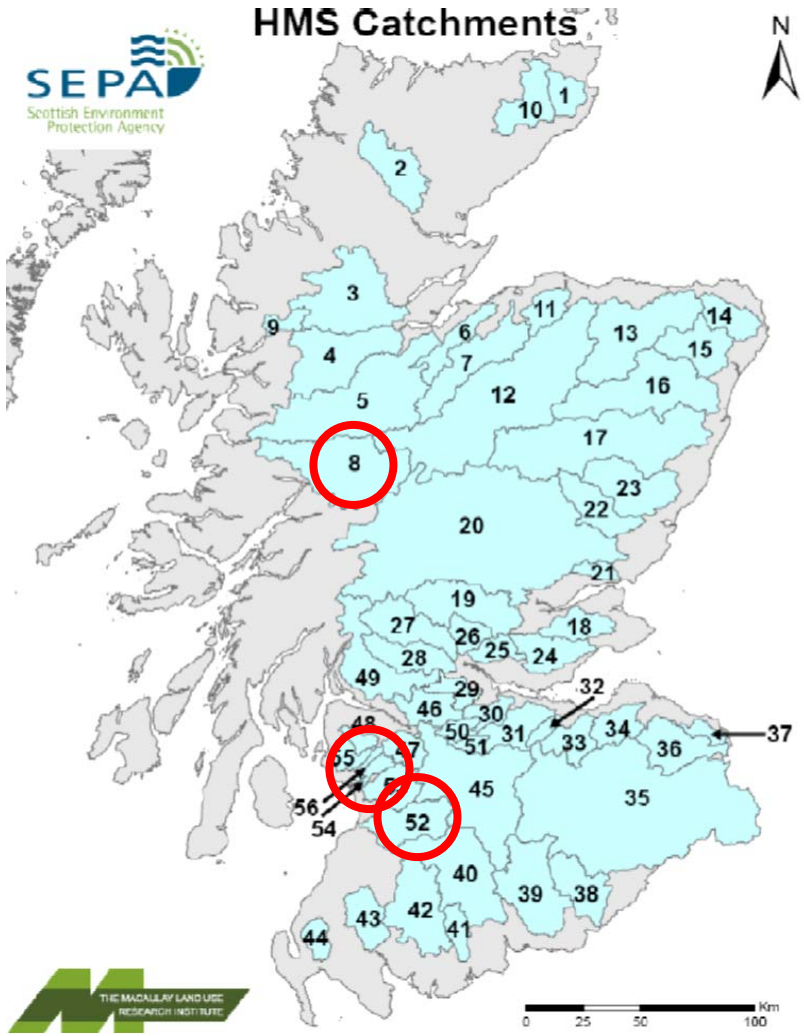
3) Model

4) Results

Data


HMS Catchments
SEPA
Scottish Environment Protection Agency
N
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
THE MACAULAY LAND USE RESEARCH INSTITUTE
0
25
50
100
Km

Data

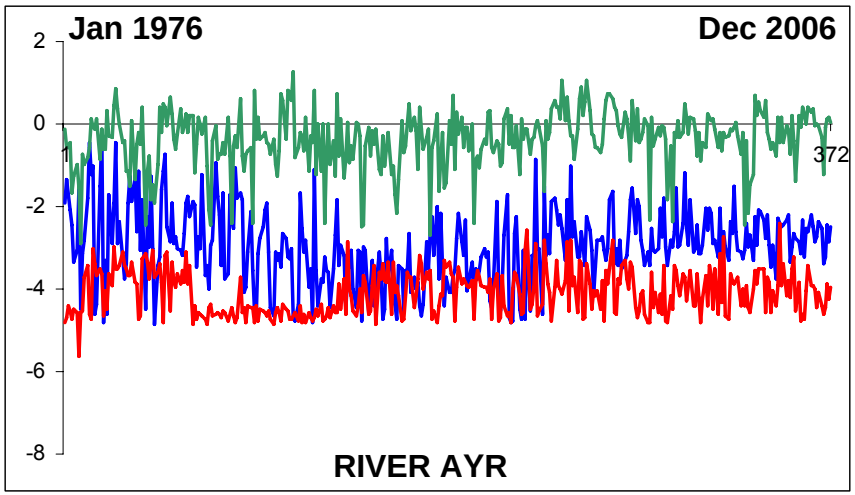


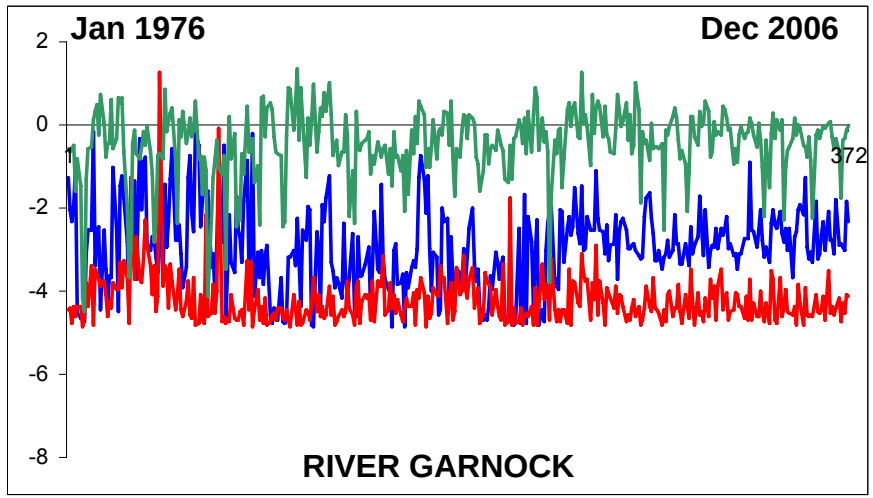
8 River Lochy

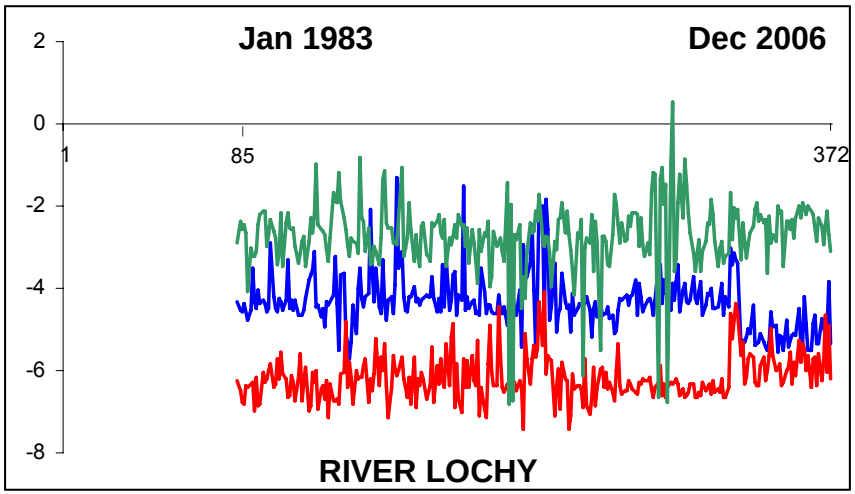
56 River Garnock

52 River Ayr

Data

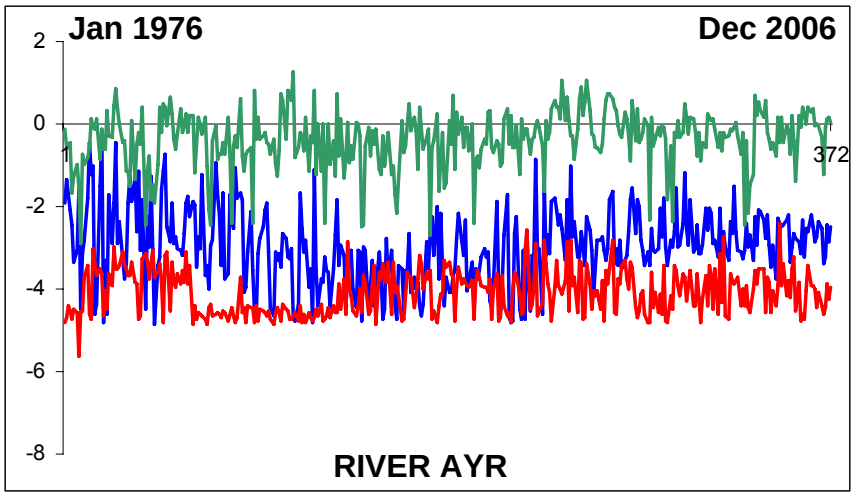


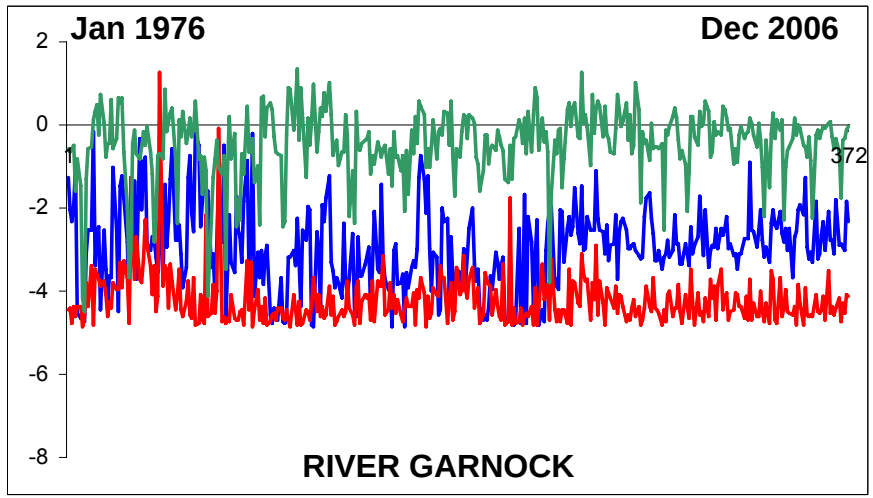


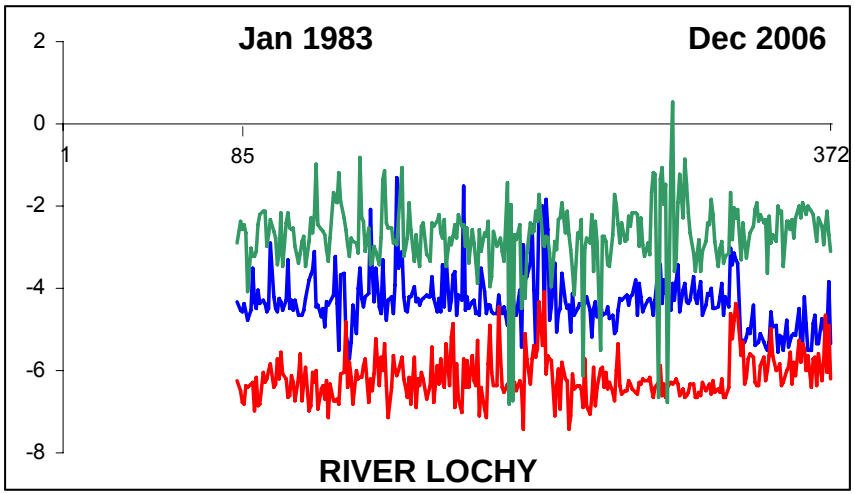


In-concentrations of NH4 (-), NO2 (-), and NO3 (-) in Rivers Ayr, Garnock, and Lochy

Data







weak dependency
non-linearity
non-normality
monthly periodicity
missing values

Data

missing values	River Ayr	NH4: 30/372 (8%)
		NO3: 31/372 (8%)
		NO2: 30/372 (8%)
	River Garnock	NH4: 43/372 (12%)
		NO3: 43/372 (12%)
		NO2: 43/372 (12%)
	River Lochy	NH4: 12/288 (4%)
		NO3: 11/288 (4%)
		NO2: 12/288 (4%)

Time series analysis

(i) the description of the main features of the series

(ii) the modelling of the sequence of the data

(iii) the prediction of the future values

(iv) the control of the process to take actions to adjust it

(v) the classification of the observations in a small number of homogeneous groups

(vi) the occurence of change-points within the dynamics of the series

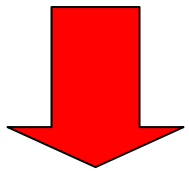
(vii) the modelling of the heterogeneity in the data through the switching among different means and covariance matrices

Time series analysis

(v) the classification of the observations in a small number of homogeneous groups

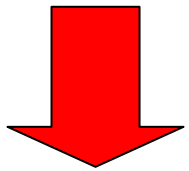
(vi) the occurence of change-points within the dynamics of the series

(vii) the modelling of the heterogeneity in the data through the switching among different means and covariance matrices



Hidden Markov Models

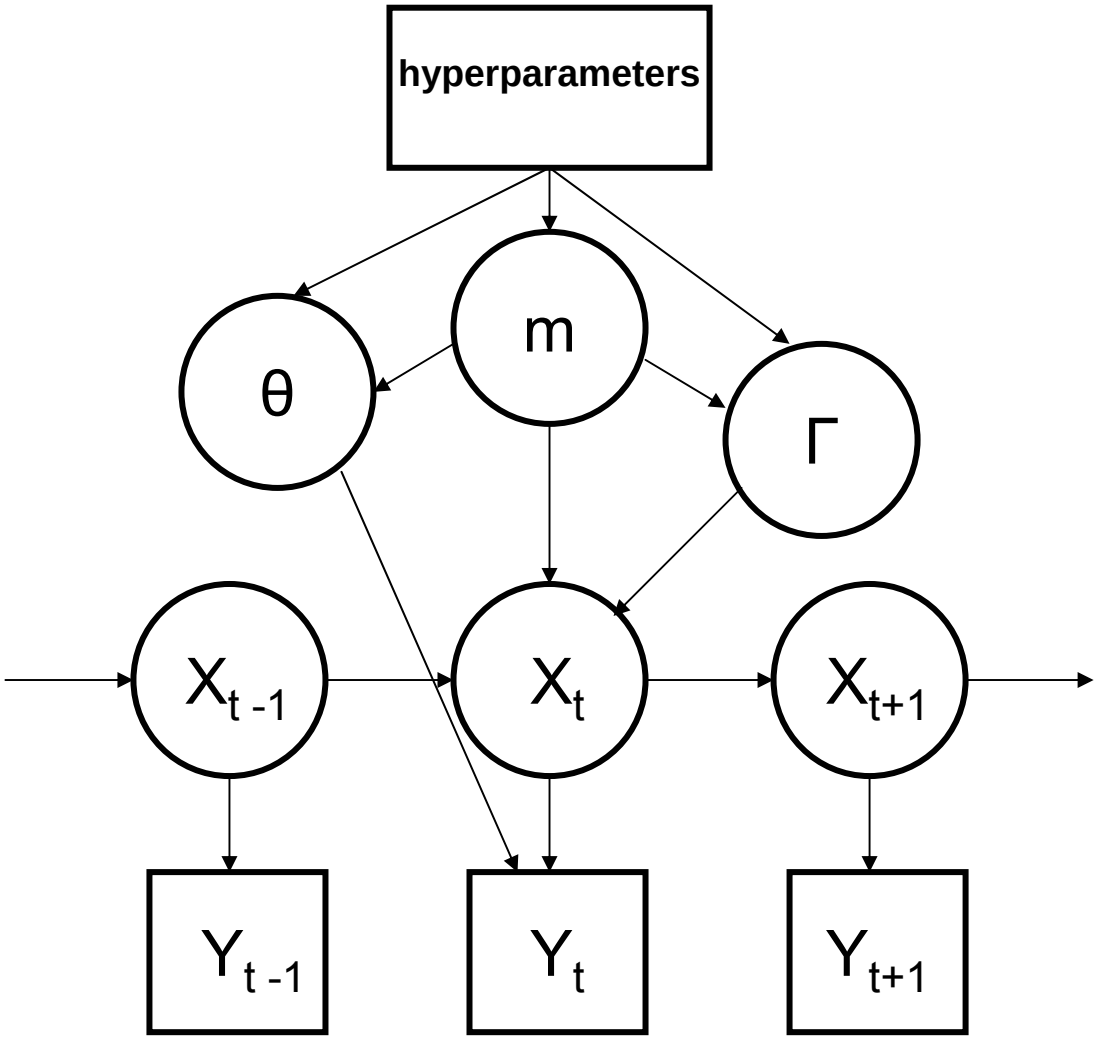
Hidden Markov Models



* the number of groups is not fixed a priori

* * no threshold is defined

Hidden Markov Models



Periodic Multivariate Normal Hidden Markov Models

$$(Y_t \mid X_t=i) = \mu_i + \beta_t + E_t \qquad i \in \{1,\ldots,m\};\quad t = 1,\ldots,T$$

$$(E_t \mid X_t=i) \sim \mathcal{N}_{(p)}(0;\ \Sigma_i)$$

$$(Y_t \mid X_t=i) \sim \mathcal{N}_{(p)}(\mu_i + \beta_t;\ \Sigma_i)$$

$$\mu_i = (\mu_{i,1},\ \ldots,\ \mu_{i,p})$$

$$\beta_t = (\beta_{t,1},\ \ldots,\ \beta_{t,p})$$

$$\beta_{t;i} = \sum_{s=1}^{q} \left(\beta_{1;s;j} \cos\left(\frac{\pi s t}{q}\right) + \beta_{2;s;j} \sin\left(\frac{\pi s t}{q}\right) \right) \qquad 2q = 12$$

$$\Sigma_i = [\sigma_{i,j,k}]_{(p\times p)} \qquad i \in \{1,\ldots,m\};\quad j,k = 1,\ldots,p$$

Bayesian Inference

reversible jump Markov chain Monte Carlo algorithms

2 steps:

- m
  - [Ayr $\Rightarrow$ 4]
  - [Garnock $\Rightarrow$ 5]
  - [Lochy $\Rightarrow$ 5]
- $\theta = (\mu, \Sigma, \beta)$
- Γ
- missing values

ex post: hidden states

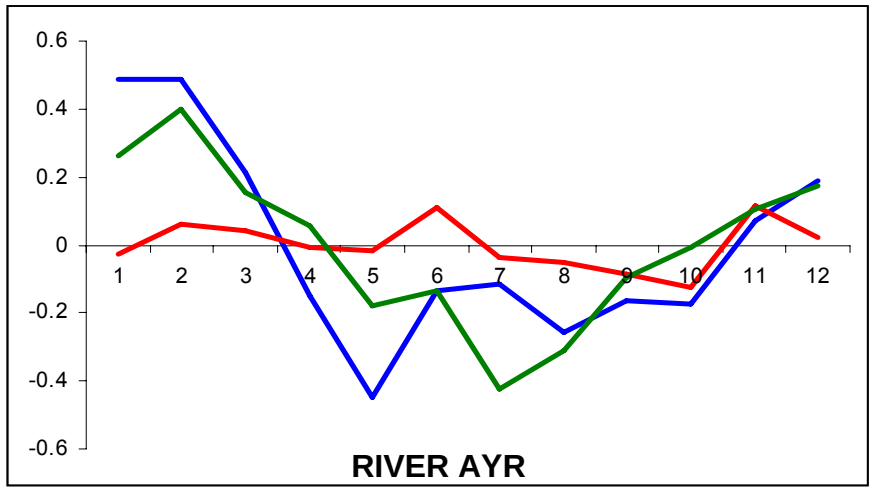
Results

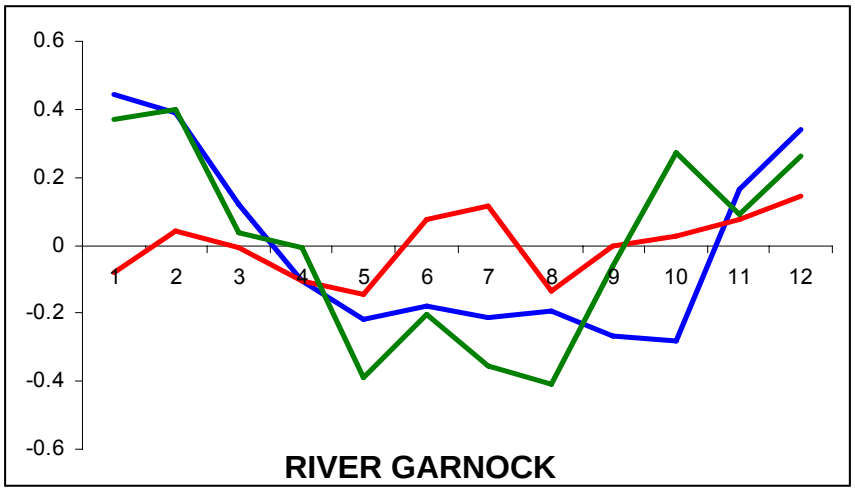
monthly periodicity

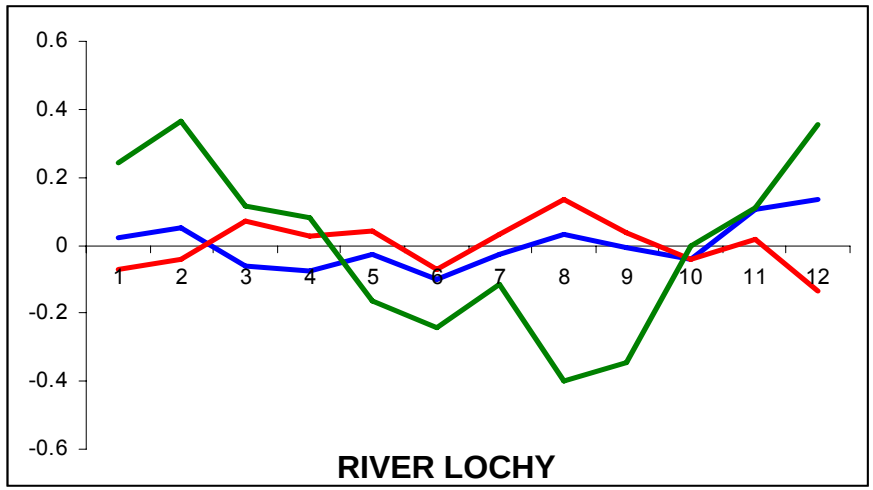
NH4 (-)

NO2 (-)

NO3 (-)





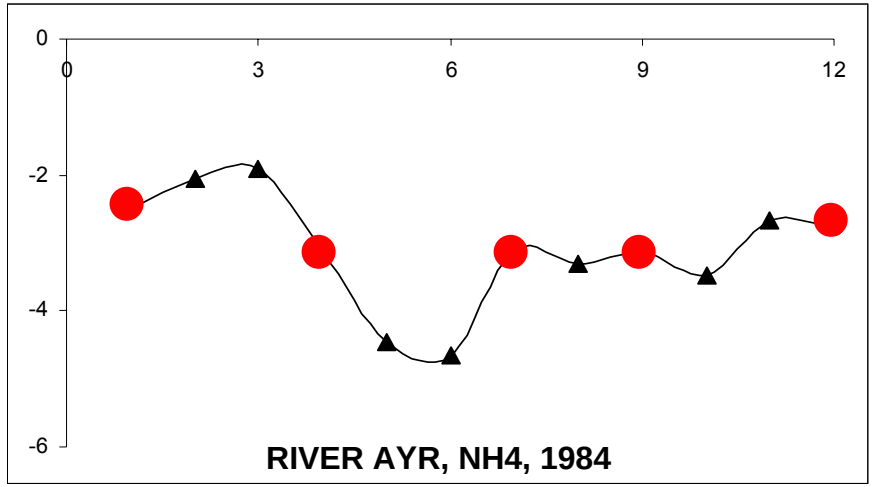


Results

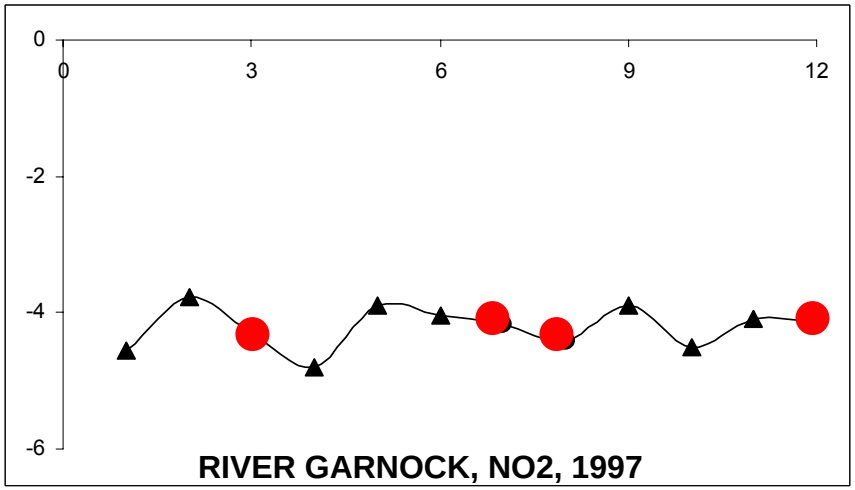
missing values

▲ = actual

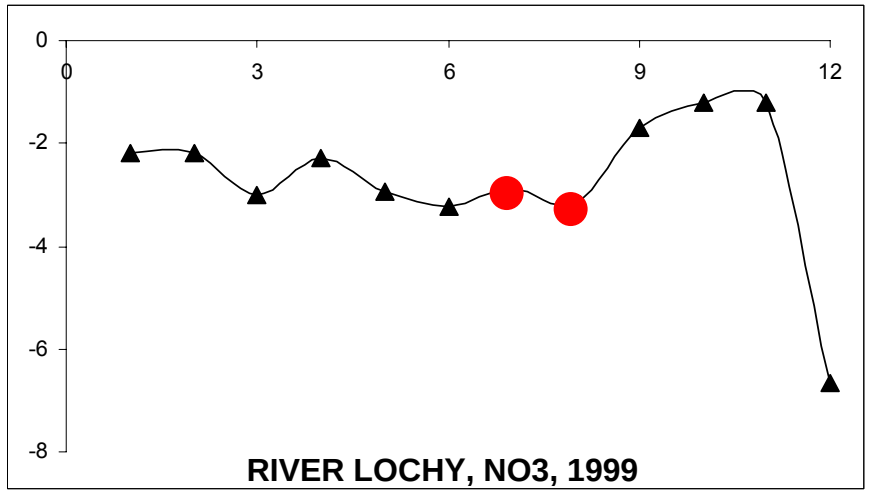
● = fitted



RIVER AYR, NH4, 1984



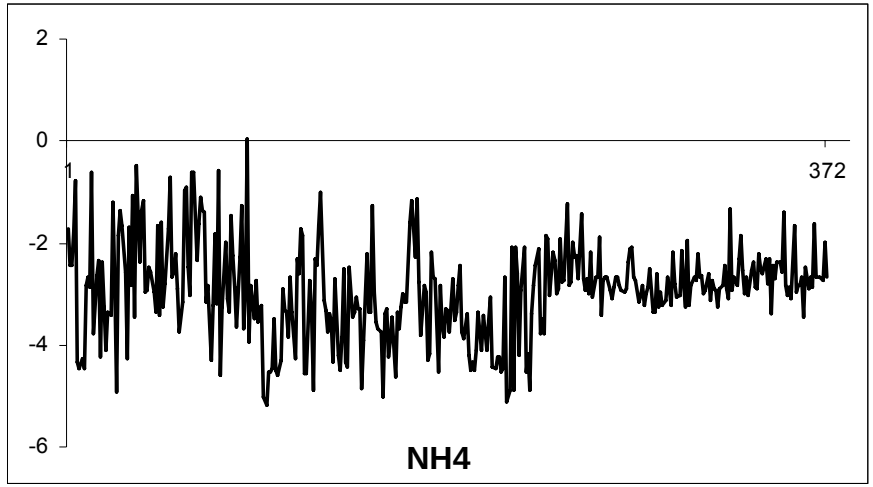
RIVER GARNOCK, NO2, 1997

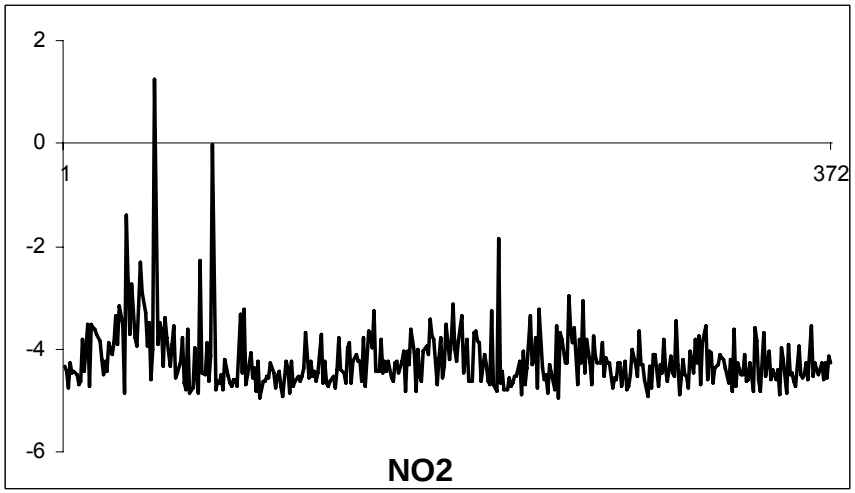


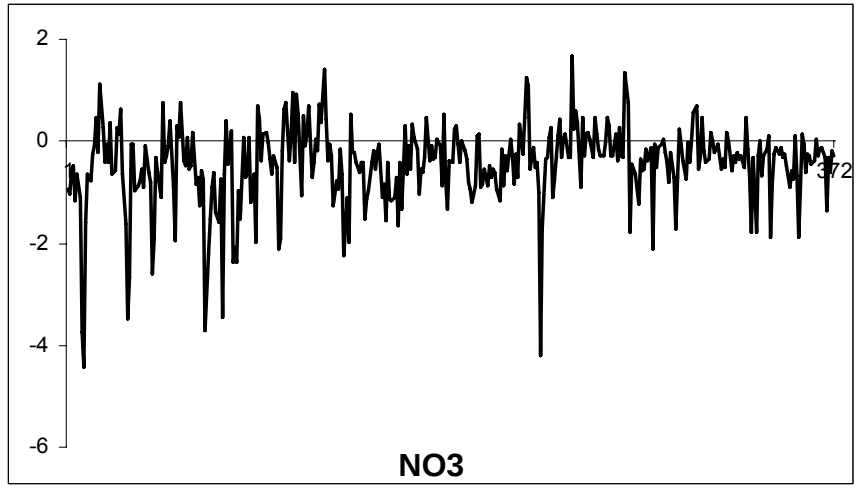
RIVER LOCHY, NO3, 1999

Results

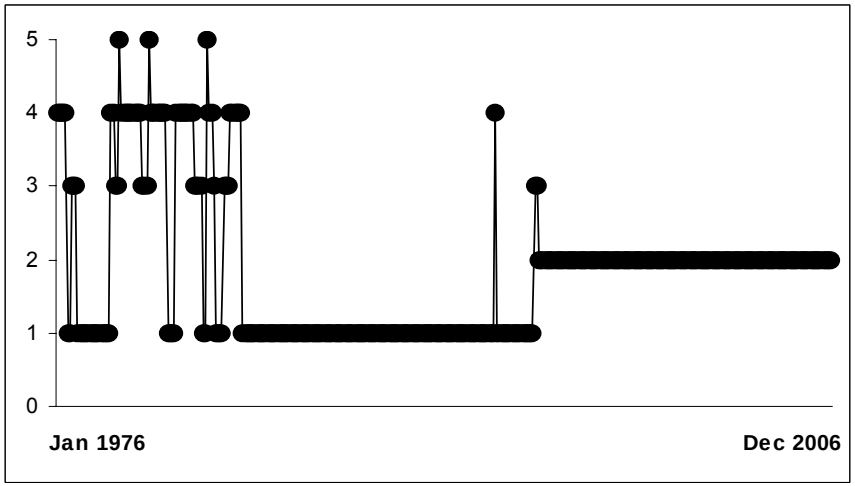
River Garnock: deseasonalized observations and hidden states (y_t - β_t)

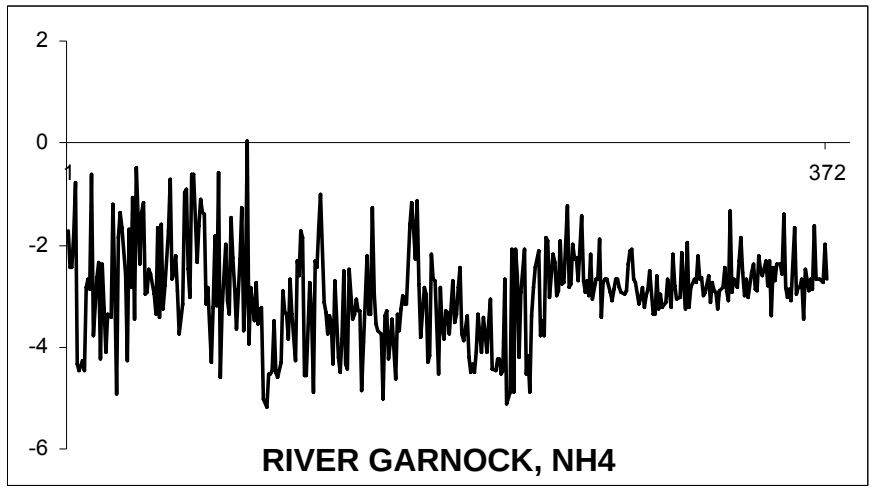


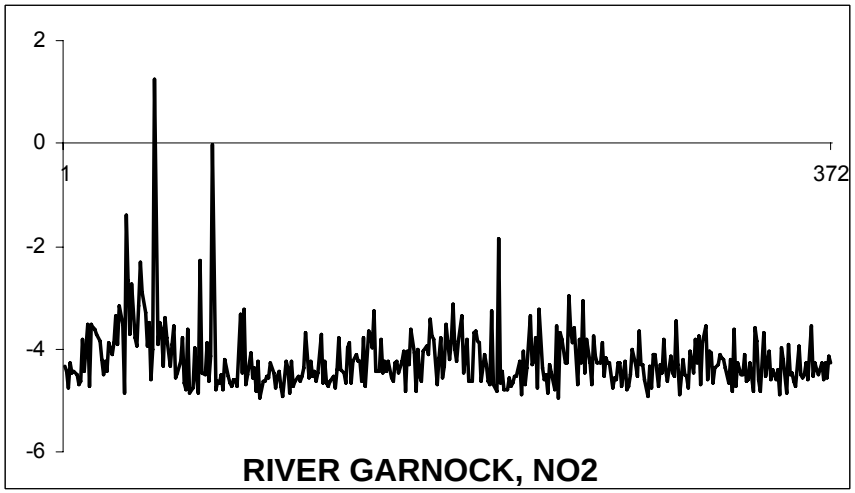


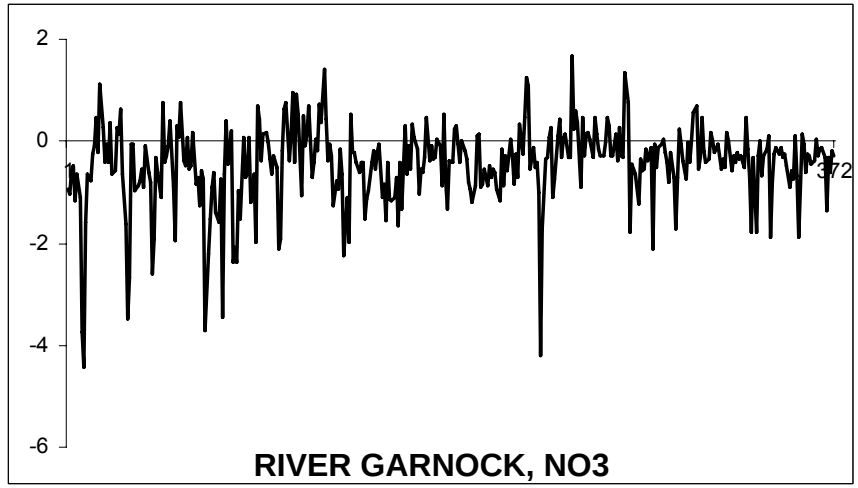


Results

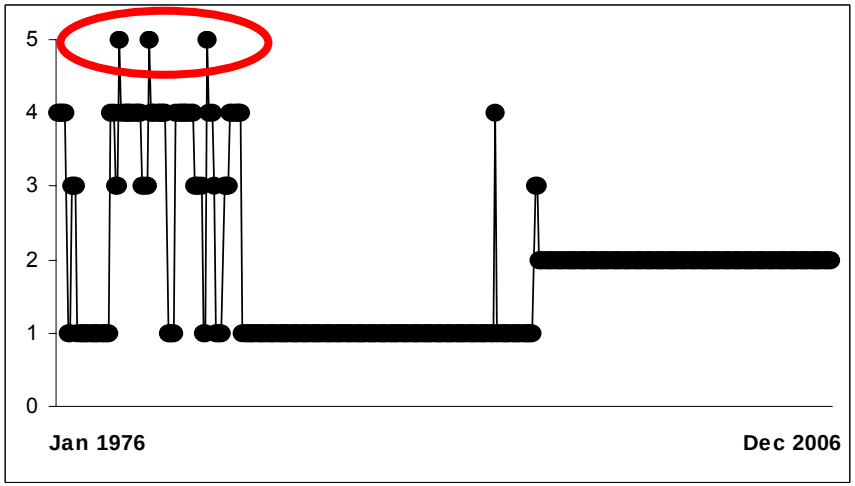

5
4
3
2
1
0
Jan 1976
Dec 2006

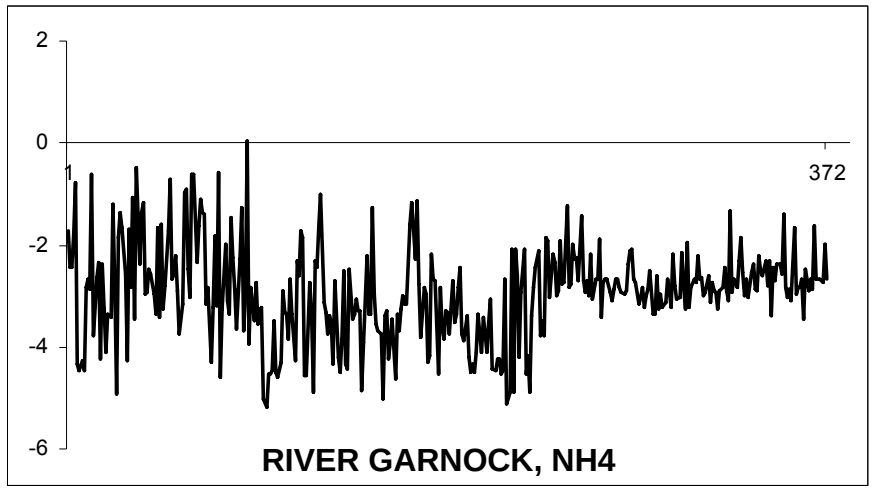

2
0
-2
-4
-6
1
372
RIVER GARNOCK, NH4

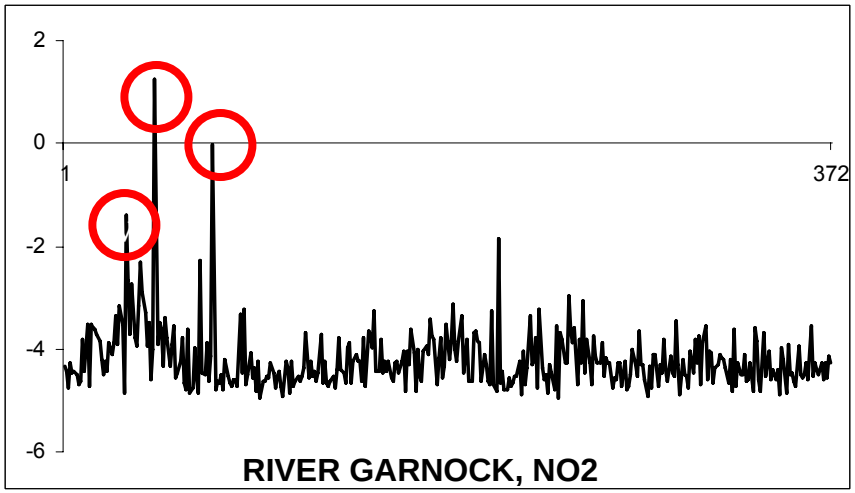

2
0
-2
-4
-6
1
372
RIVER GARNOCK, NO2

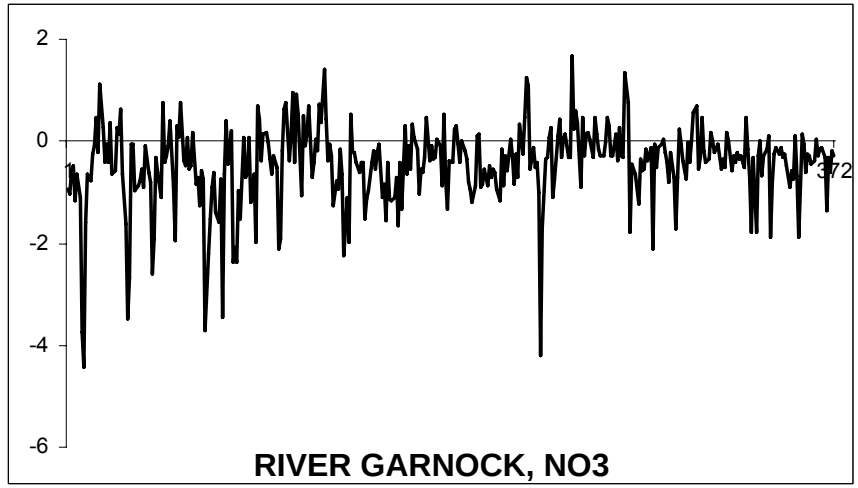

2
0
-2
-4
-6
1
372
RIVER GARNOCK, NO3

Results

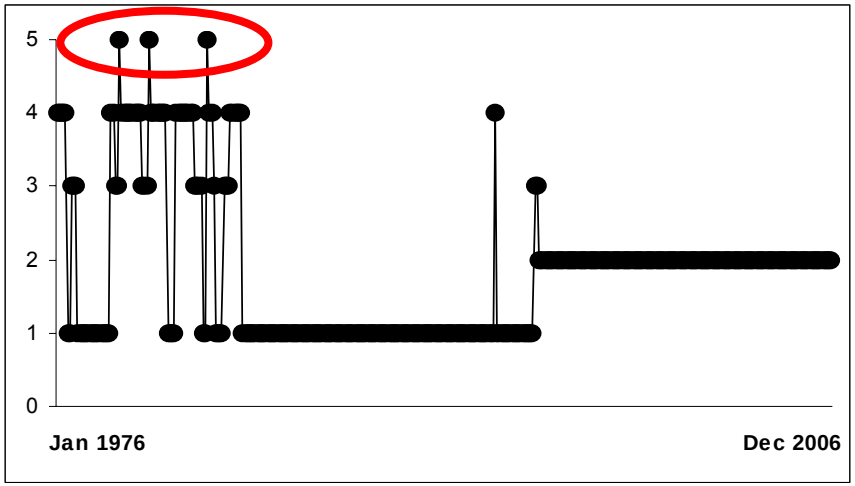

5
4
3
2
1
0
Jan 1976
Dec 2006

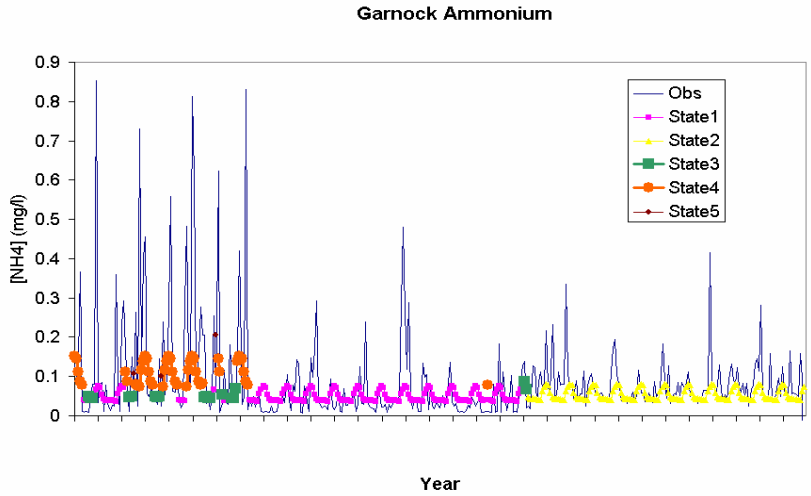

2
0
-2
-4
-6
1
372
RIVER GARNOCK, NH4

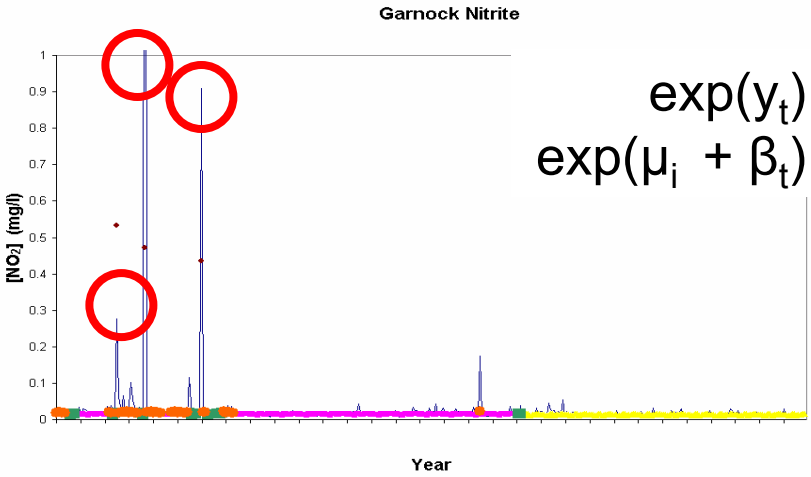

2
0
-2
-4
-6
1
372
RIVER GARNOCK, NO2


2
0
-2
-4
-6
1
372
RIVER GARNOCK, NO3

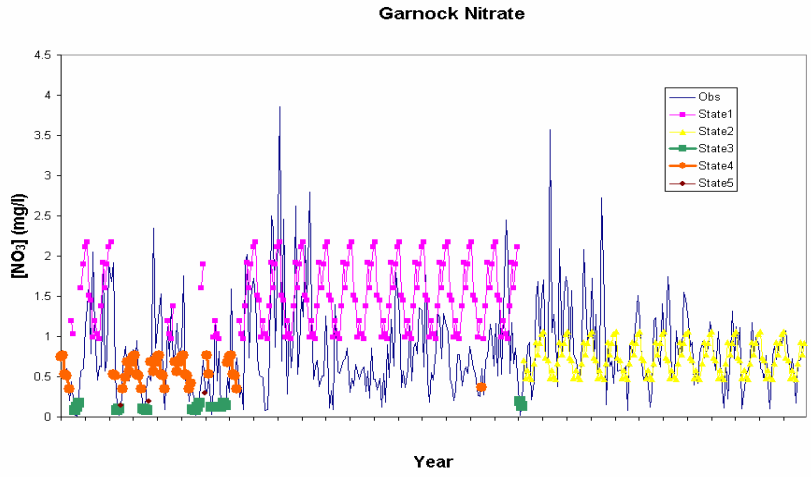
Results



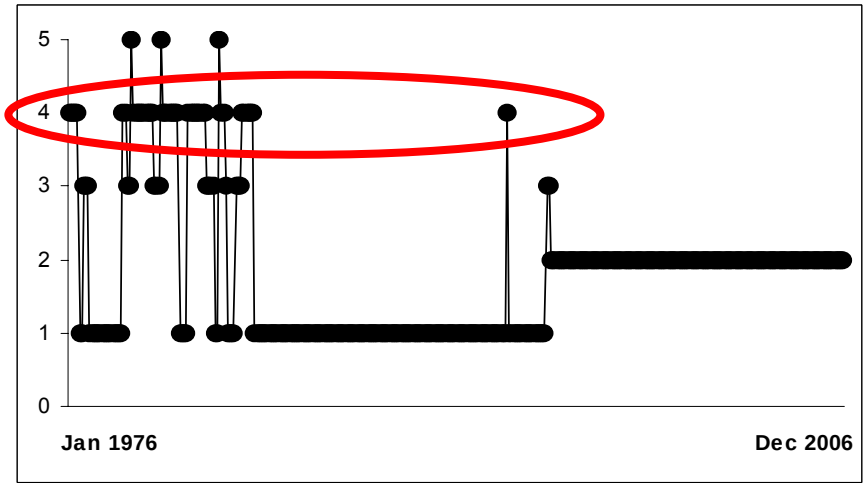


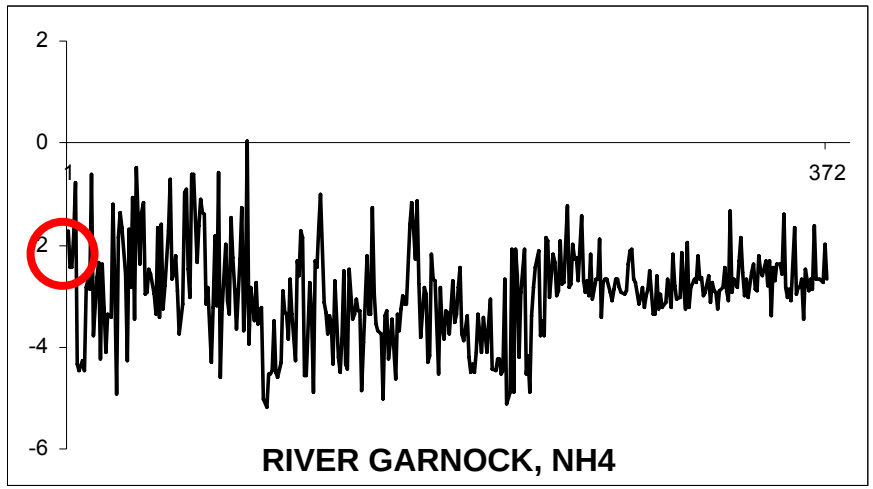


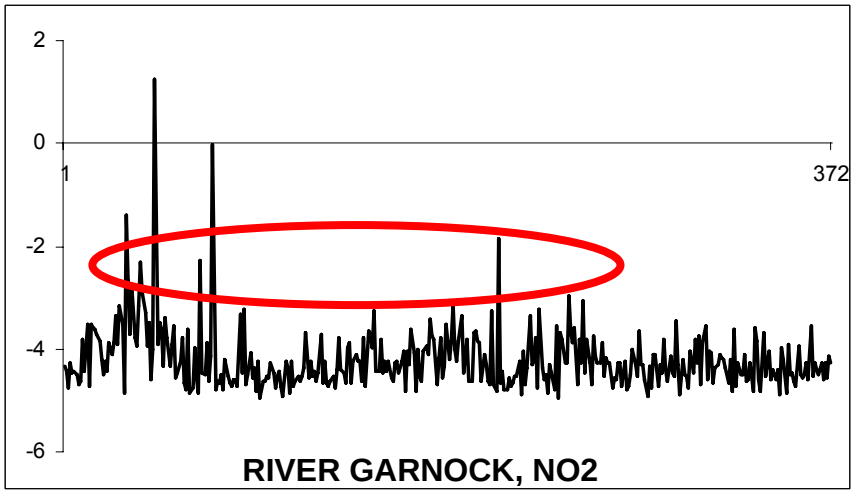
$$\exp(y_t)$$
$$\exp(\mu_i + \beta_t)$$

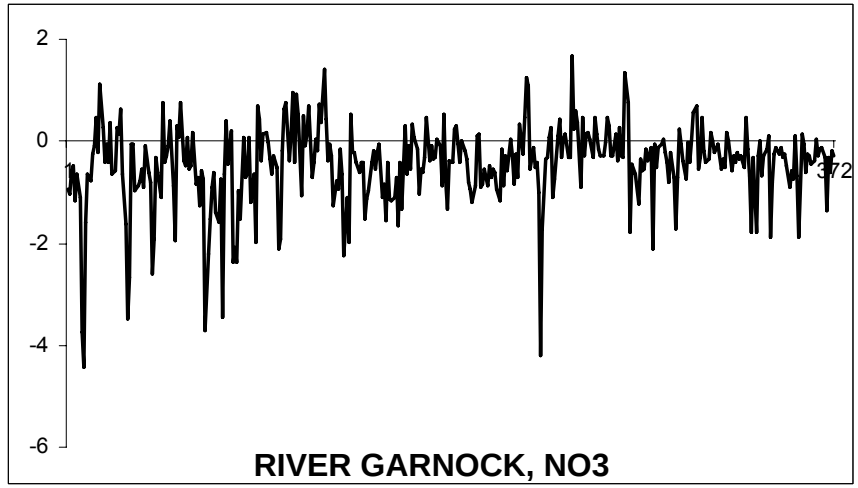


Results

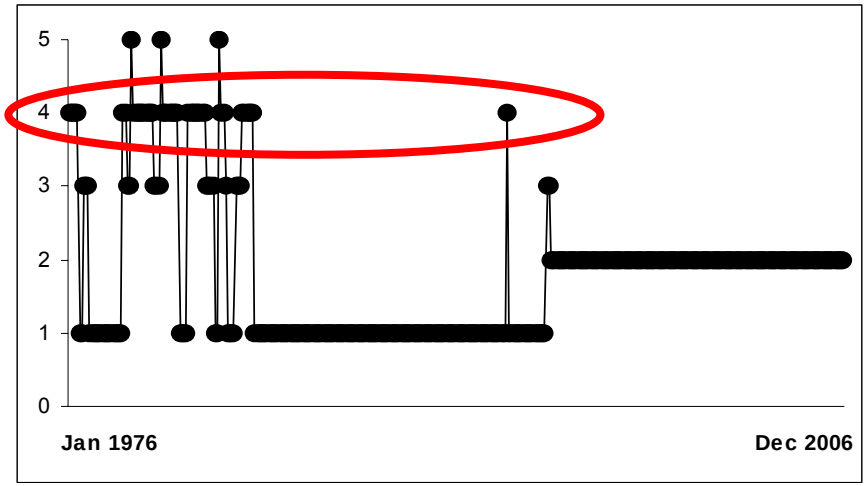

5
4
3
2
1
0
Jan 1976
Dec 2006

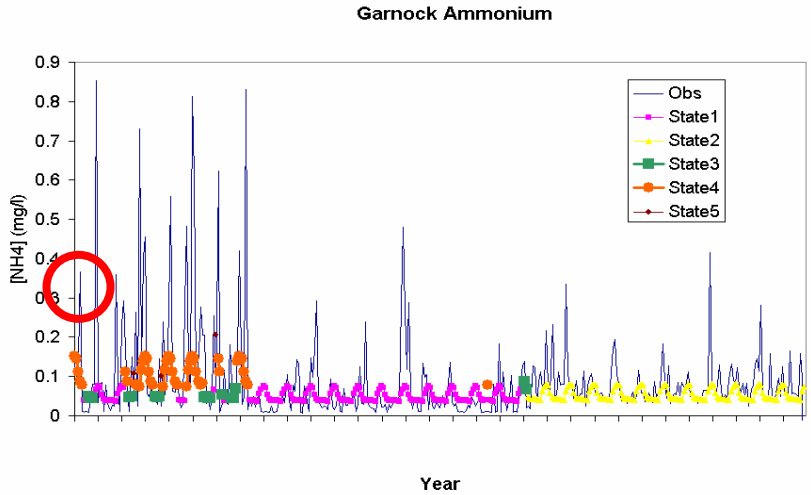

2
0
-4
-6
1
372
RIVER GARNOCK, NH4

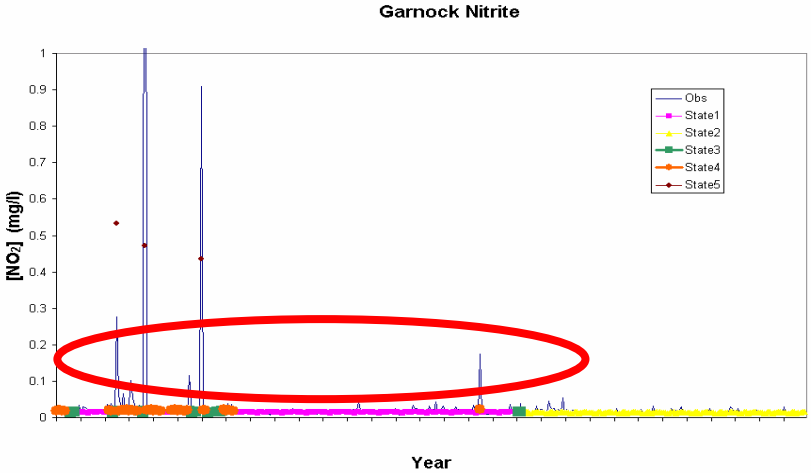

2
0
-2
-4
-6
1
372
RIVER GARNOCK, NO2

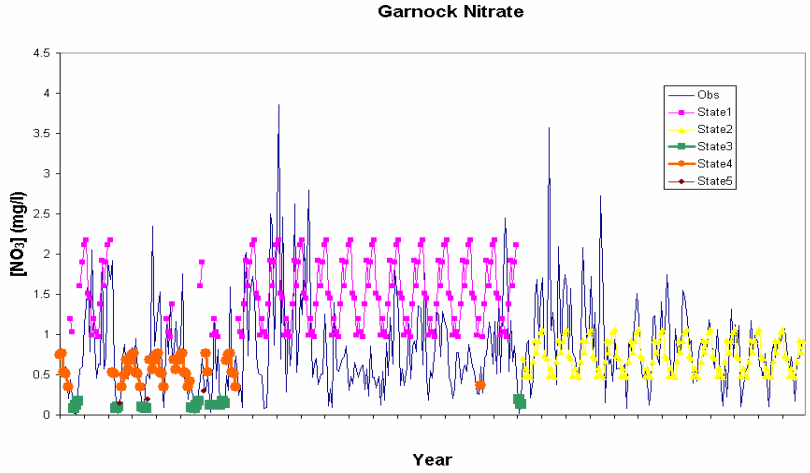

2
0
-2
-4
-6
372
RIVER GARNOCK, NO3

Results

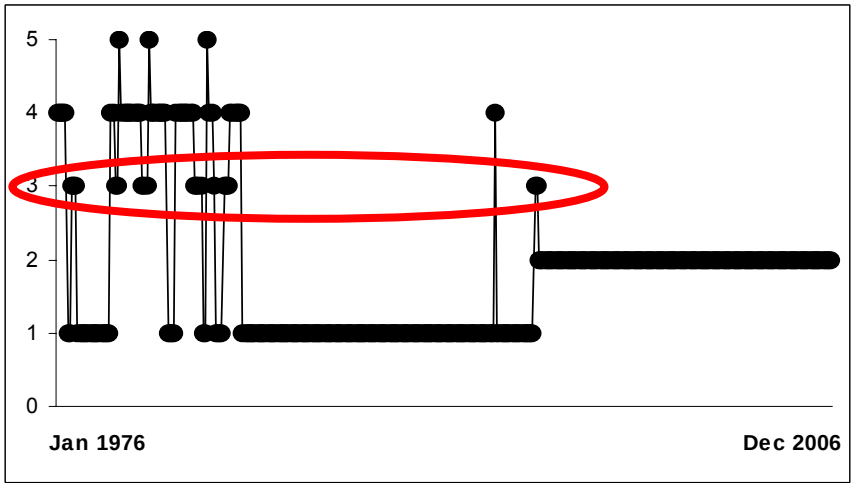

5
4
3
2
1
0
Jan 1976
Dec 2006

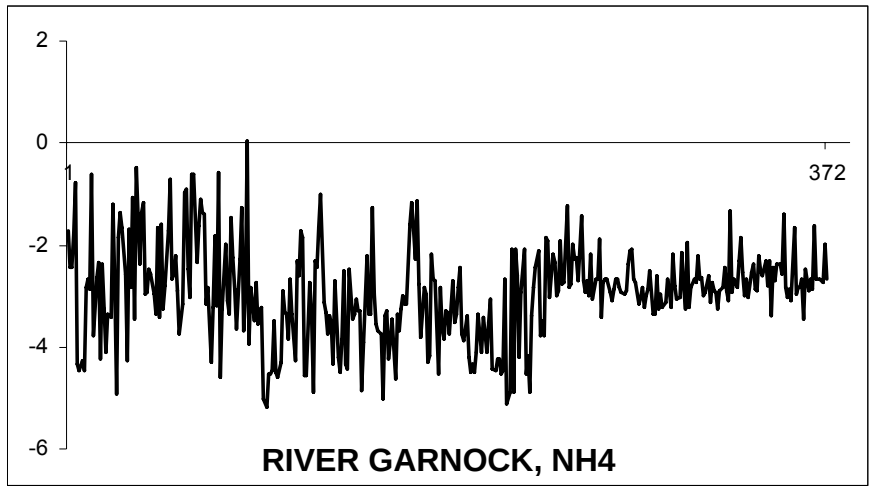

Garnock Ammonium
[NH4] (mg/l)
Year
Obs
State1
State2
State3
State4
State5

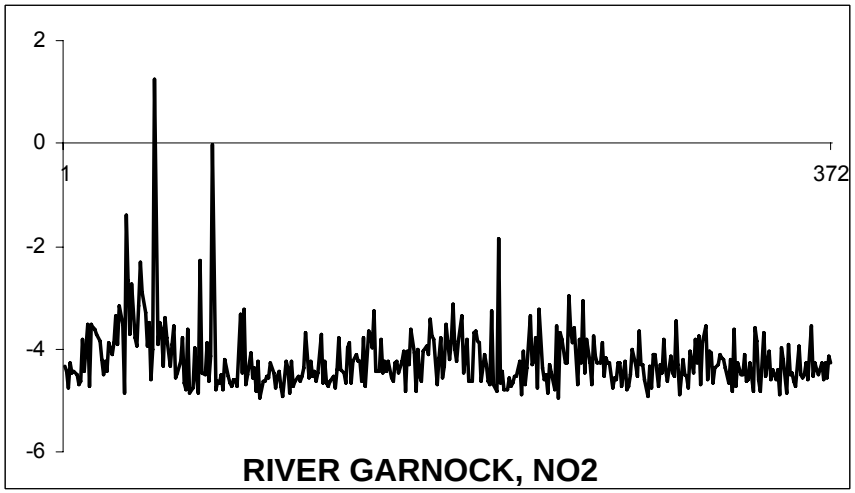

Garnock Nitrite
[NO2] (mg/l)
Year
Obs
State1
State2
State3
State4
State5

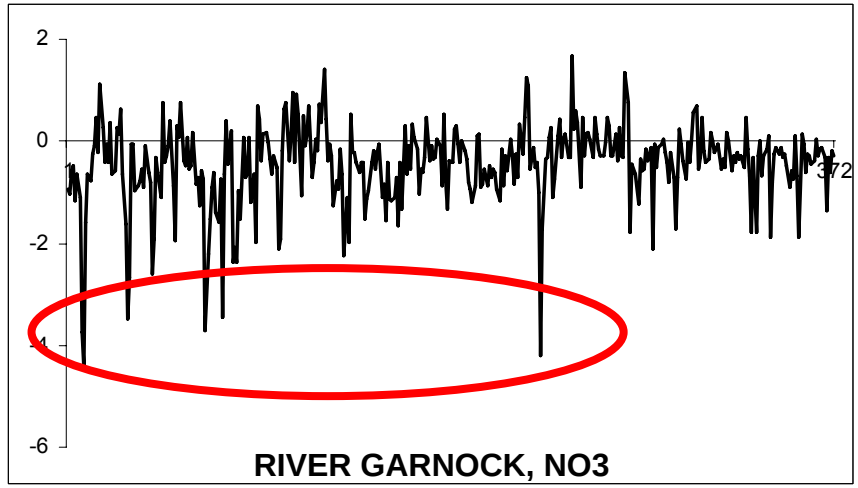

Garnock Nitrate
[NO3] (mg/l)
Year
Obs
State1
State2
State3
State4
State5

Results

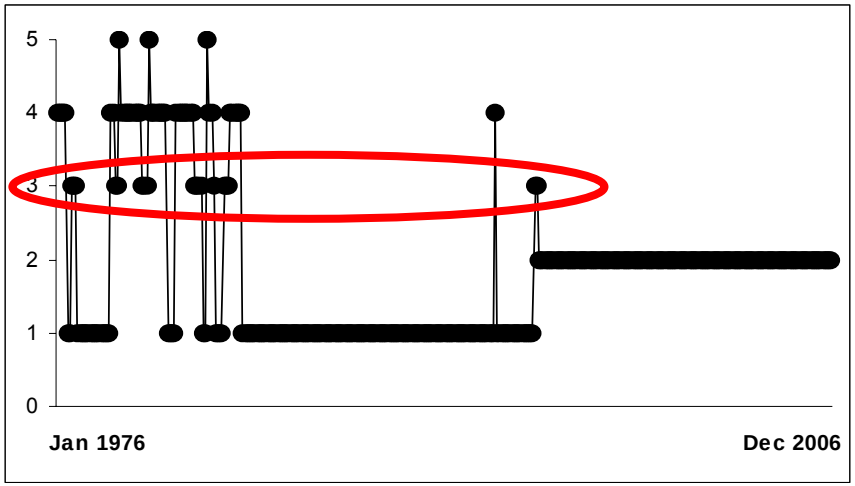

5
4
3
2
1
0
Jan 1976
Dec 2006

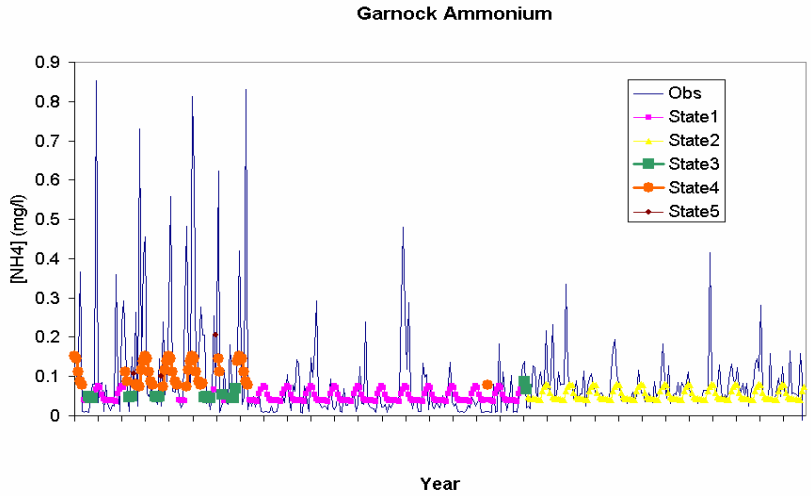

2
0
-2
-4
-6
1
372
RIVER GARNOCK, NH4

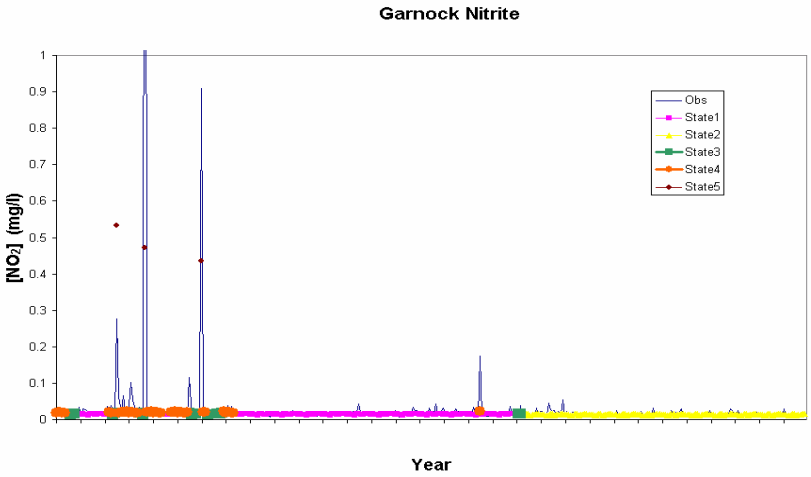

2
0
-2
-4
-6
1
372
RIVER GARNOCK, NO2

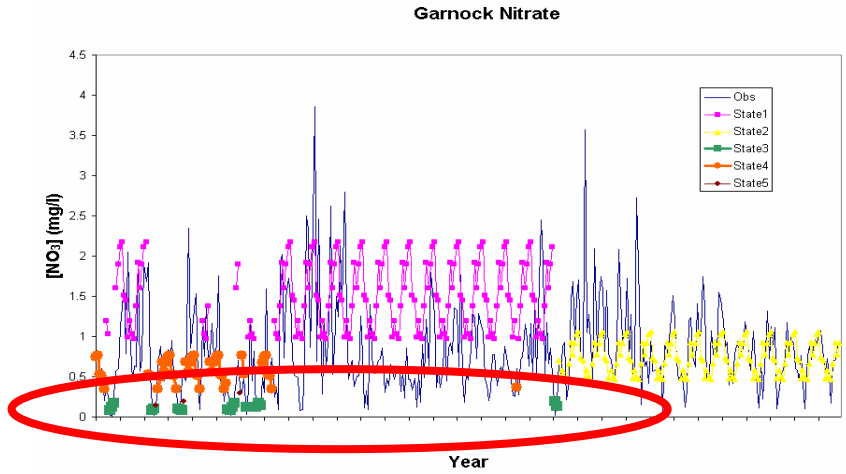

2
0
-2
-6
372
RIVER GARNOCK, NO3

Results

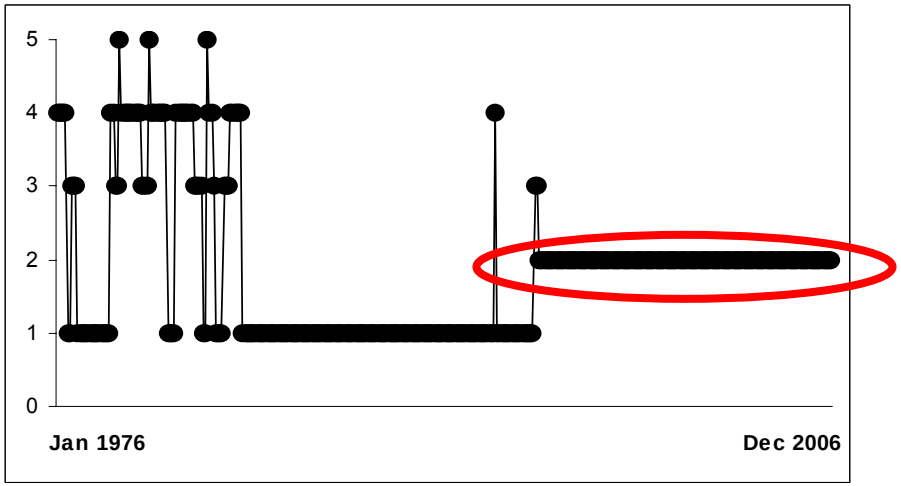

5
4
3
2
1
0
Jan 1976
Dec 2006

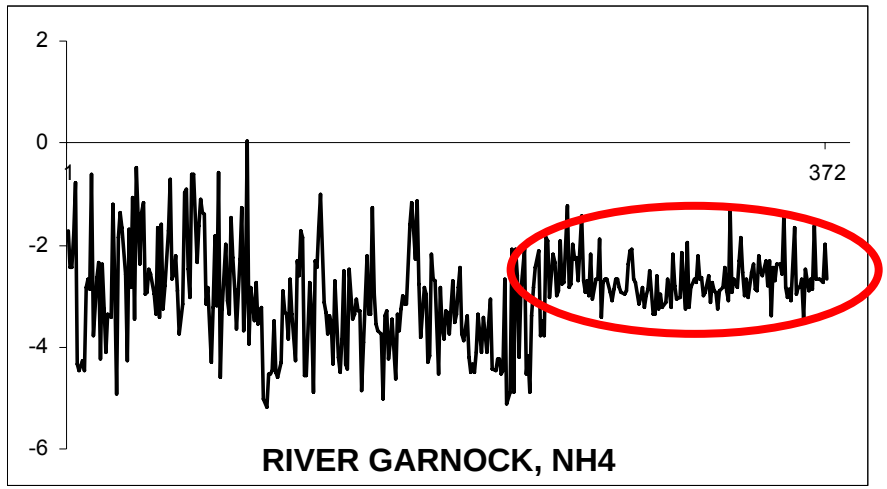

Garnock Ammonium
[NH4] (mg/l)
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
Obs
State1
State2
State3
State4
State5
Year

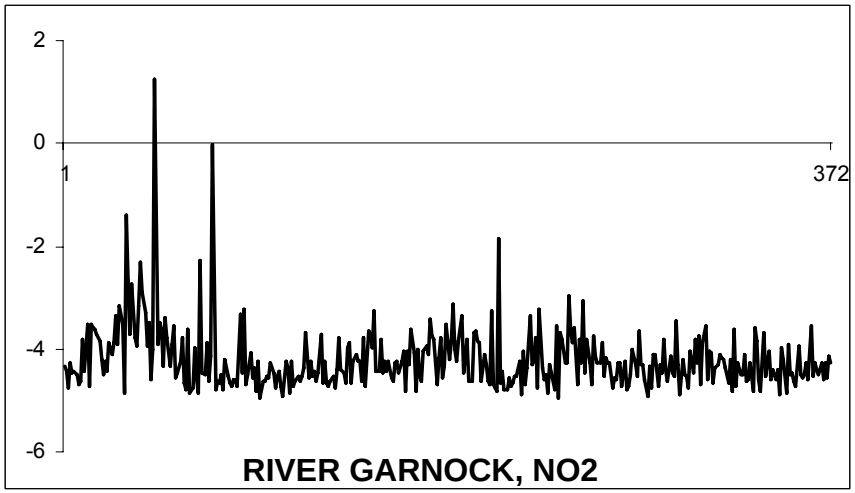

Garnock Nitrite
[NO2] (mg/l)
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
Obs
State1
State2
State3
State4
State5
Year

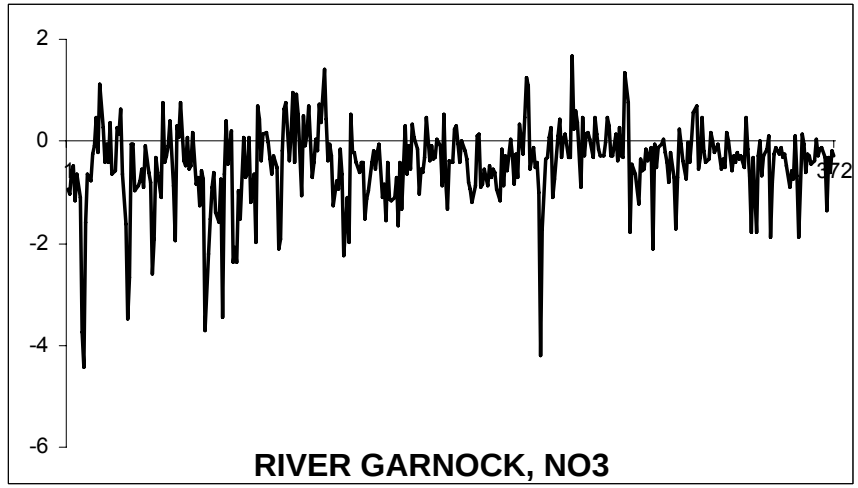

Garnock Nitrate
[NO3] (mg/l)
4.5
4
3.5
3
2.5
2
1.5
1
0.5
0
Obs
State1
State2
State3
State4
State5
Year

Results

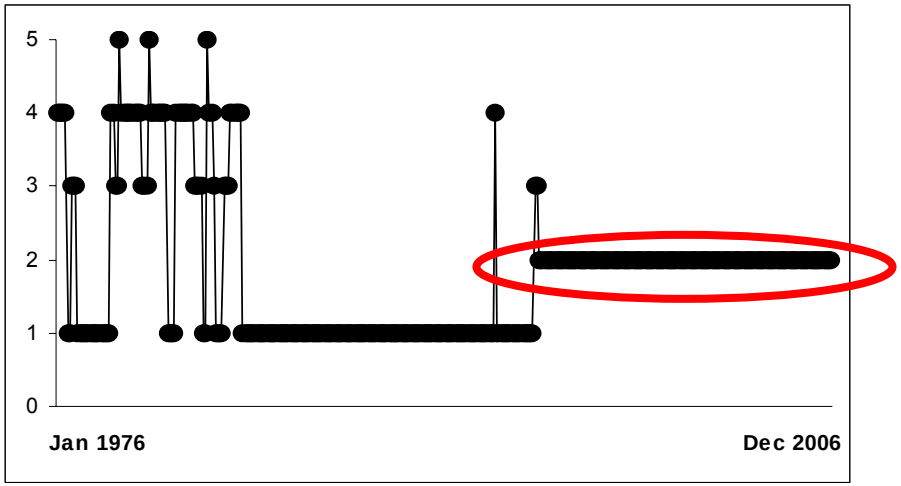

5
4
3
2
1
0
Jan 1976
Dec 2006

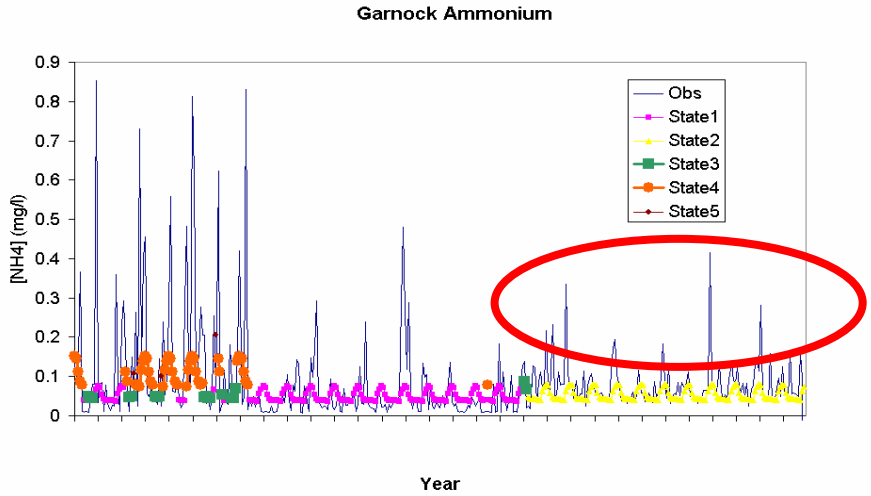

2
0
-2
-4
-6
1
372
RIVER GARNOCK, NH4

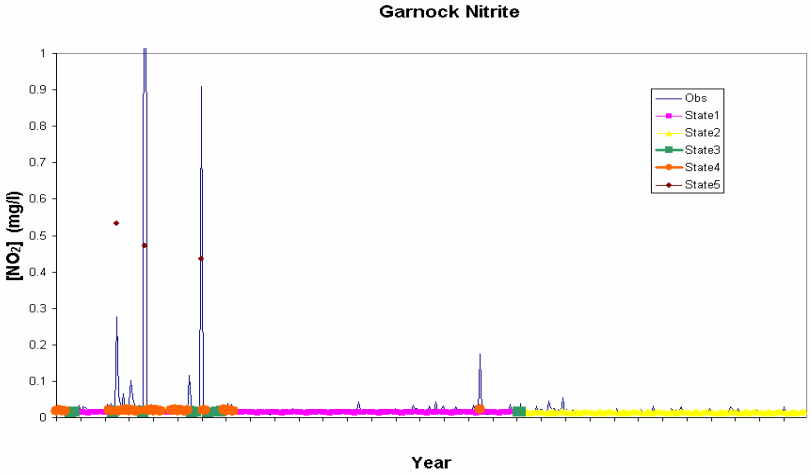

2
0
-2
-4
-6
1
372
RIVER GARNOCK, NO2

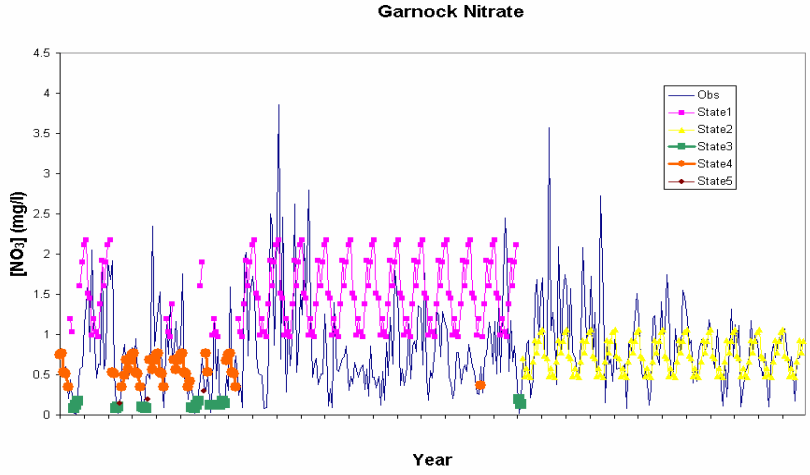

2
0
-2
-4
-6
1
372
RIVER GARNOCK, NO3

Results

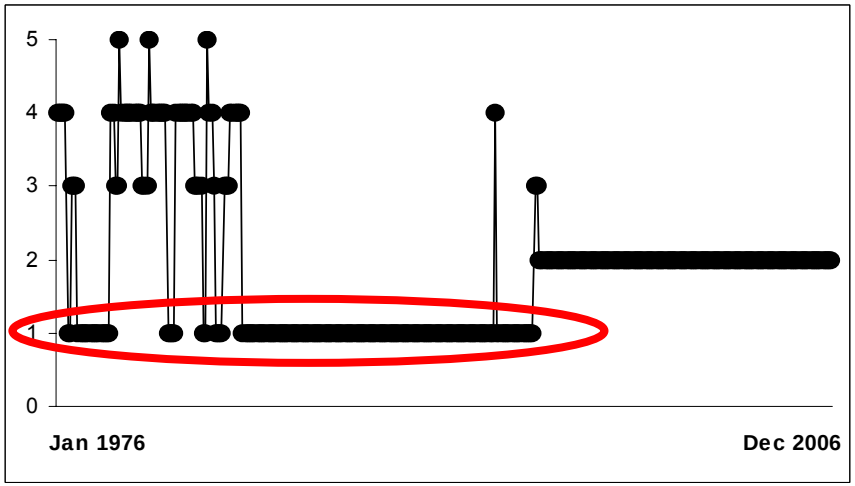

5
4
3
2
1
0
Jan 1976
Dec 2006

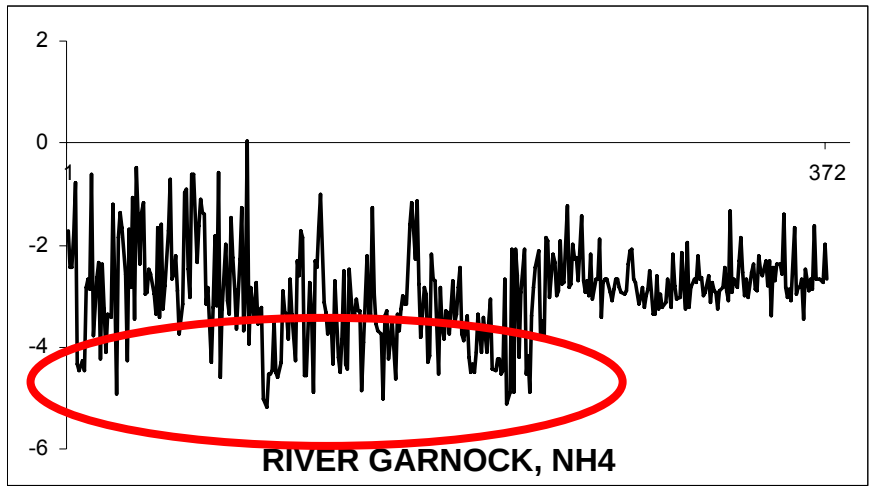

Garnock Ammonium
[NH4] (mg/l)
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
Obs
State1
State2
State3
State4
State5
Year

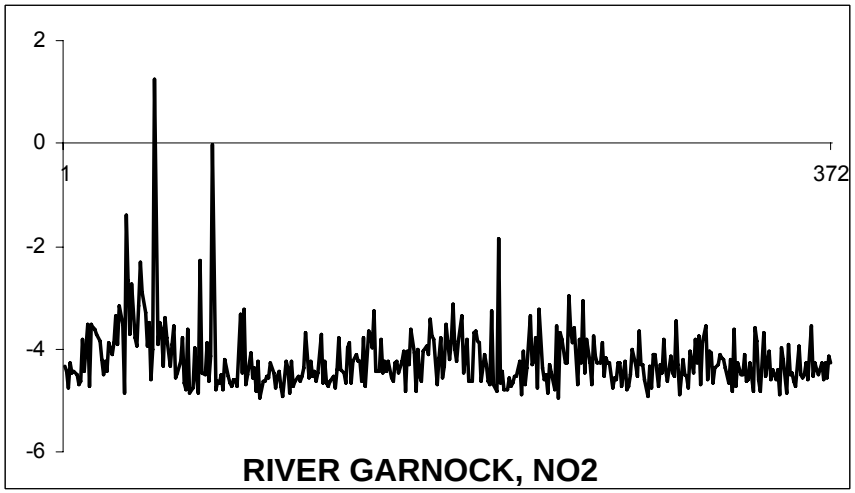

Garnock Nitrite
[NO2] (mg/l)
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
Obs
State1
State2
State3
State4
State5
Year

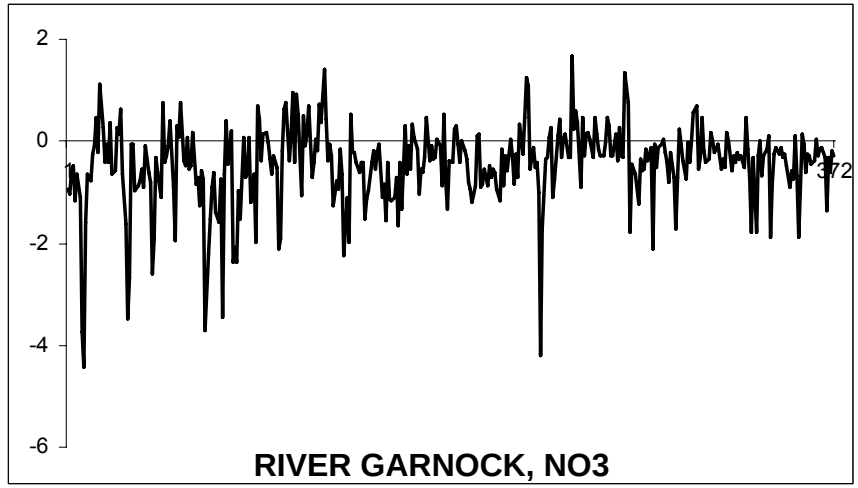

Garnock Nitrate
[NO3] (mg/l)
4.5
4
3.5
3
2.5
2
1.5
1
0.5
0
Obs
State1
State2
State3
State4
State5
Year

Results

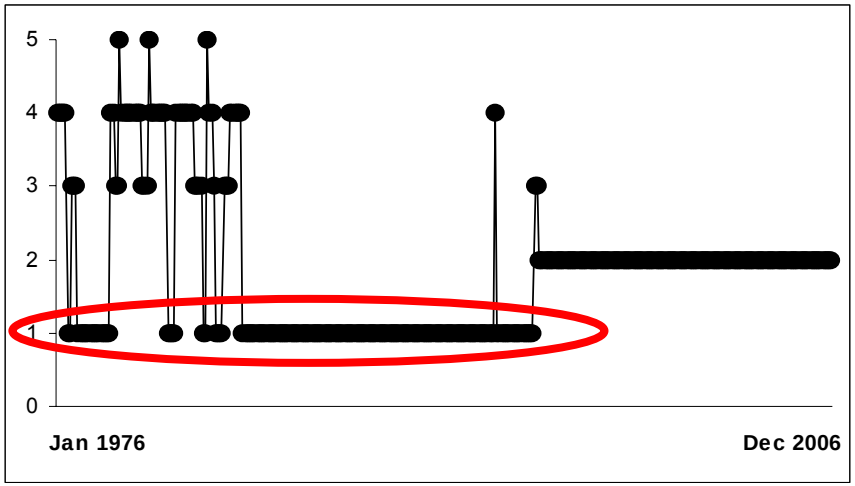

5
4
3
2
1
0
Jan 1976
Dec 2006

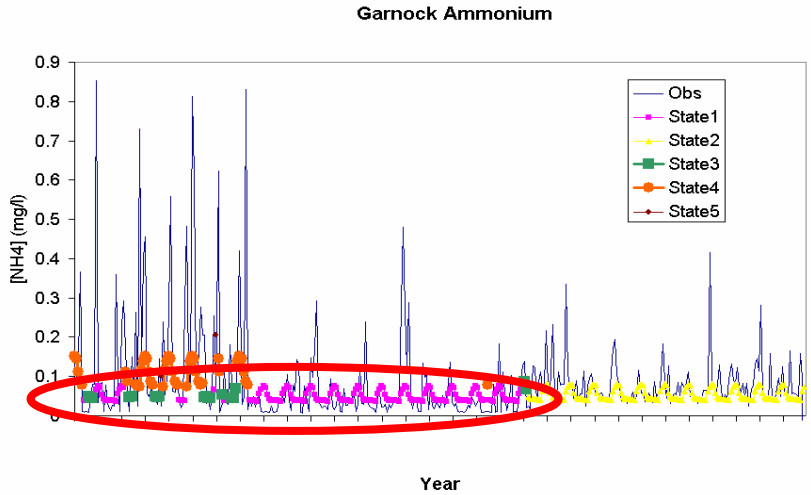

2
0
-2
-4
-6
1
372
RIVER GARNOCK, NH4

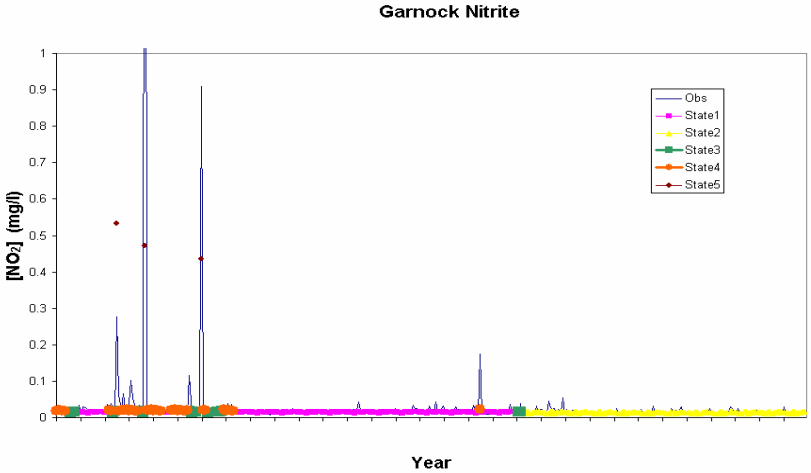

2
0
-2
-4
-6
1
372
RIVER GARNOCK, NO2

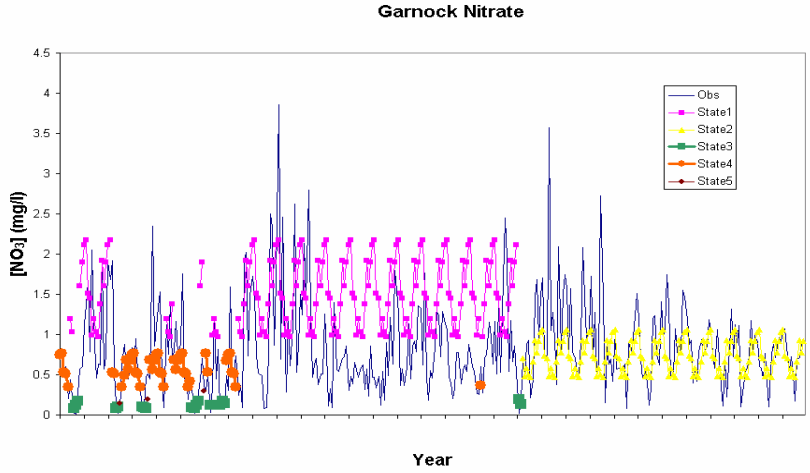

2
0
-2
-4
-6
1
372
RIVER GARNOCK, NO3

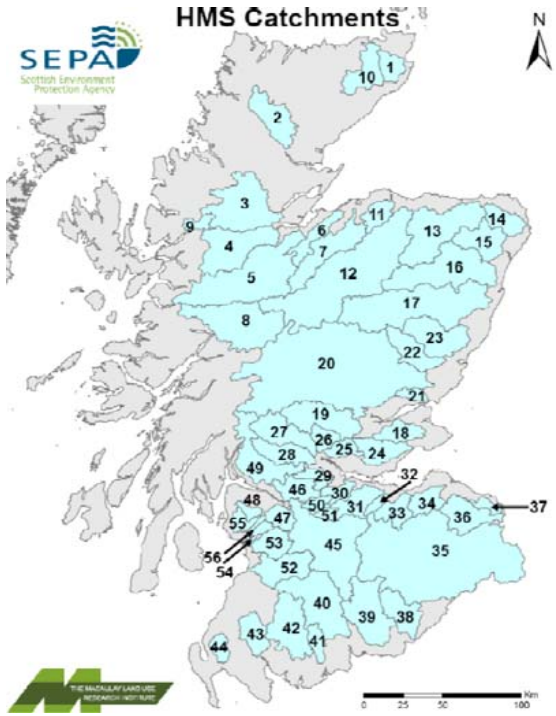
Results


5
4
3
2
1
0
Jan 1976
Dec 2006


Garnock Ammonium
[NH4] (mg/l)
Year
Obs
State1
State2
State3
State4
State5


Garnock Nitrite
[NO2] (mg/l)
Year
Obs
State1
State2
State3
State4
State5


Garnock Nitrate
[NO3] (mg/l)
Year
Obs
State1
State2
State3
State4
State5

Current and Future Work

spatio-temporal analysis
of the 56 rivers through
Hidden Markov Random Fields