Detecting change points in marine time series using state-space models

Matthew Spencer, Nova Mieszkowska, **Leonie Robinson**, Steve Simpson, Mike Burrows, Chris Frid, and others

`m.spencer@liverpool.ac.uk`

This work was supported by DEFRA, NERC, MAFF, the Scottish Government, the Crown Estate, the British Geological Survey, the University of Liverpool, and Natural England.

Regime shifts in marine communities

- “Low-frequency, high-amplitude changes in oceanic conditions that may propagate through several trophic levels and be especially pronounced in biological variables” [a]

[a] Collie et al. (2004), Prog. Oceanogr. 60:281-302

[b] Lees et al. (2006), Fish and Fisheries 7:104-127

Regime shifts in marine communities

- "Low-frequency, high-amplitude changes in oceanic conditions that may propagate through several trophic levels and be especially pronounced in biological variables" [a]
- "Sudden, high-amplitude, infrequent events, which are detectable in multiple aspects of the physical and biological components and on large spatial scales."[b]

---

[a] Collie et al. (2004), Prog. Oceanogr. 60:281-302

[b] Lees et al. (2006), Fish and Fisheries 7:104-127

Regime shifts in marine communities

- “Low-frequency, high-amplitude changes in oceanic conditions that may propagate through several trophic levels and be especially pronounced in biological variables” [a]
- “Sudden, high-amplitude, infrequent events, which are detectable in multiple aspects of the physical and biological components and on large spatial scales.”[b]
- Although it is widely believed that there are regime shifts in marine communities as a result of environmental change, there is actually little quantitative evidence.

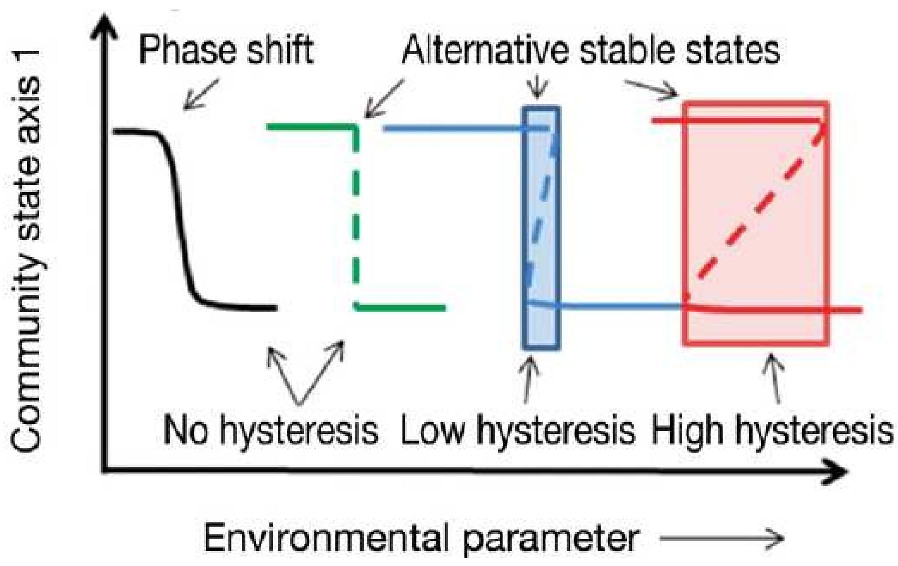
---

[a] Collie et al. (2004), Prog. Oceanogr. 60:281-302

[b] Lees et al. (2006), Fish and Fisheries 7:104-127

Regime shifts and alternative stable states

- The concept of regime shifts in marine communities is often associated with step-like changes between alternative stable states[a].



---

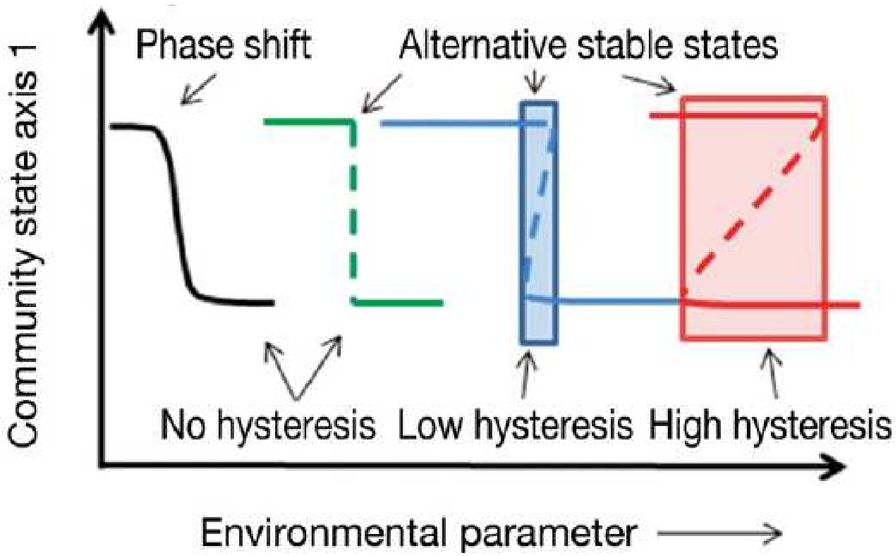
[a] Figure from Dudgeon et al. (2010) MEPS 413: 201-216

[b] Petraitis and Dudgeon (2004), JEMBE 300: 343-371

[c] Petraitis et al. (2009) Oecologia 161:139-148

Regime shifts and alternative stable states

- The concept of regime shifts in marine communities is often associated with step-like changes between alternative stable states[a].
- There is little experimental evidence for alternative stable states in marine communities[b], with the exception of mussel beds in the North-Eastern USA[c].

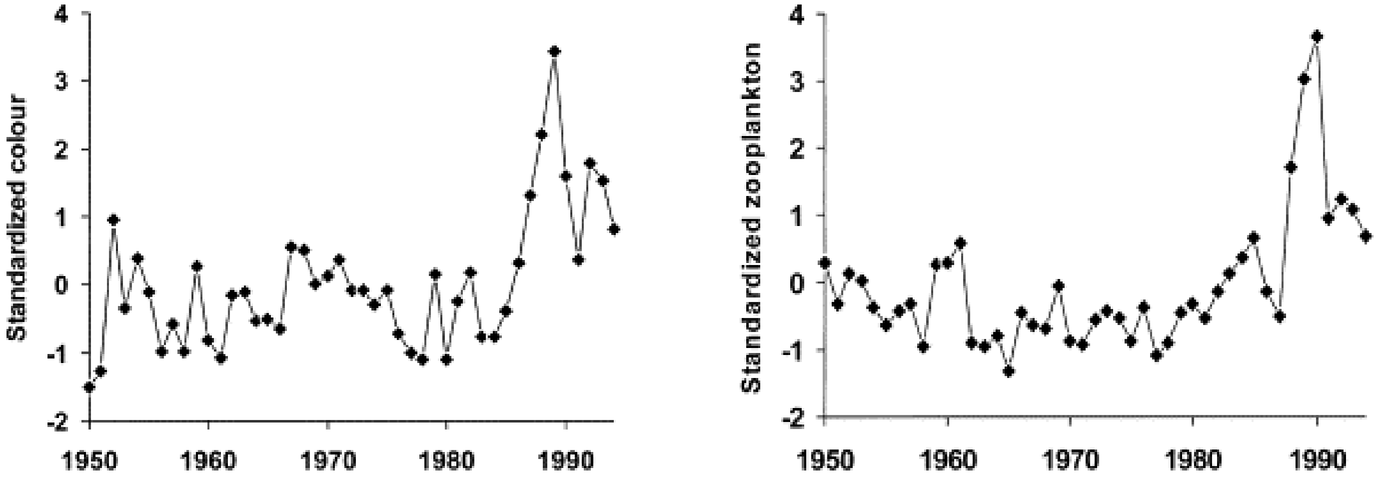


[a] Figure from Dudgeon et al. (2010) MEPS 413: 201-216

[b] Petraitis and Dudgeon (2004), JEMBE 300: 343-371

[c] Petraitis et al. (2009) Oecologia 161:139-148

Plankton in the North Sea: a proposed regime shift



The late 1980s and early 1990s appear to be unusual for both phytoplankton (left: CPR greenness index) and zooplankton (right: first principal component of CPR zooplankton counts) in the North Sea[a].

---

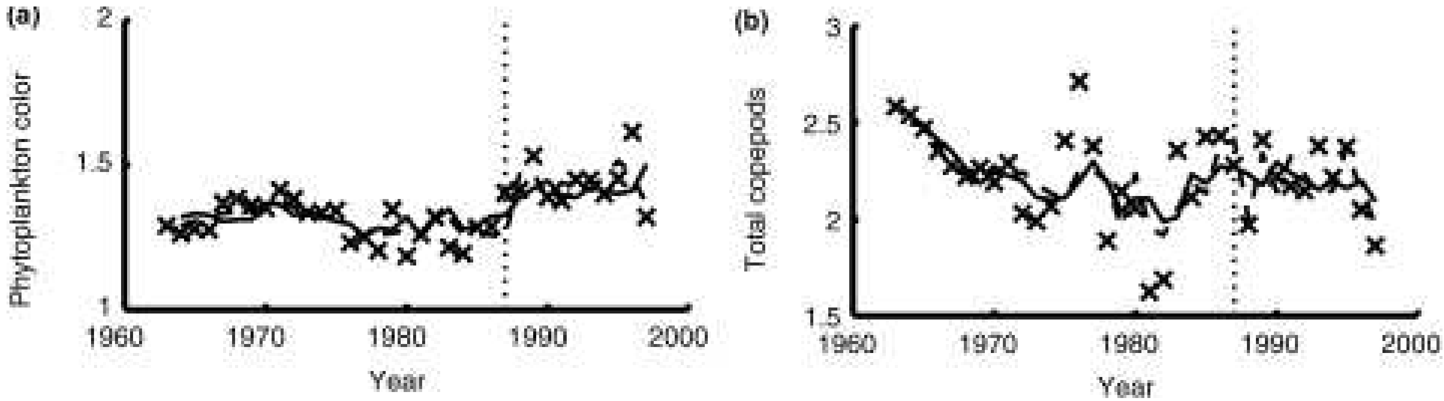
[a] Reid et al. (2001), Fisheries Research 50:163-171

A statistical test for the North Sea regime shift

- Solow and Smith[a] fitted models to a set of 5 North Sea time series, representing the dynamics of the system as a linearized vector autoregressive process, varying around either one equilibrium or two.

A statistical test for the North Sea regime shift

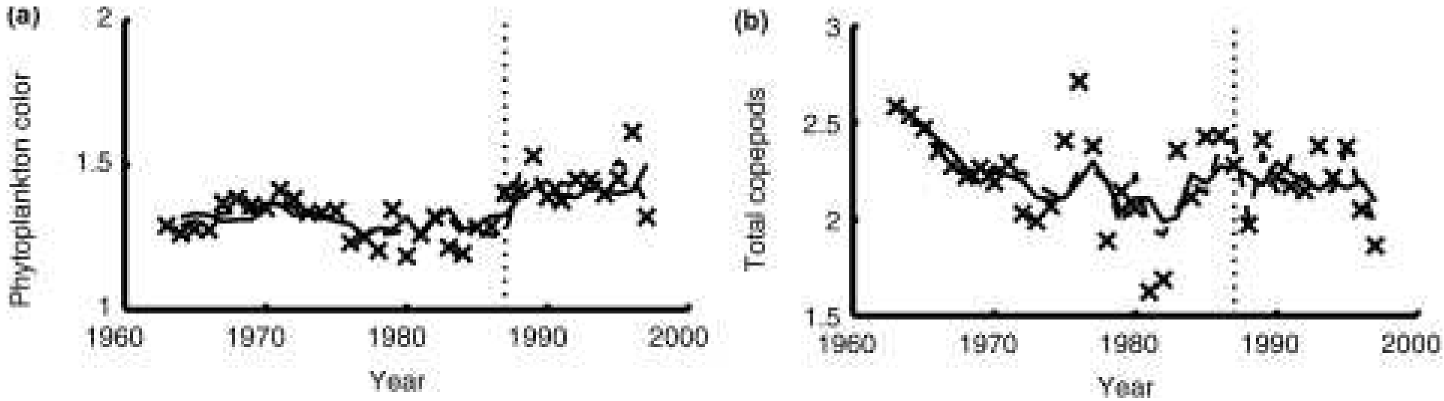
- Solow and Smith[a] fitted models to a set of 5 North Sea time series, representing the dynamics of the system as a linearized vector autoregressive process, varying around either one equilibrium or two.



(a) phytoplankton colour, (b) copepods.

A statistical test for the North Sea regime shift

- Solow and Smith[a] fitted models to a set of 5 North Sea time series, representing the dynamics of the system as a linearized vector autoregressive process, varying around either one equilibrium or two.



(a) phytoplankton colour, (b) copepods.

- Although the best change point is 1989, models with and without a change point are almost indistinguishable.

[a] Solow and Smith (2005), Fisheries Oceanography 14:236-240

Aims

- Assemble a large collection of biological time series from marine communities around the UK.

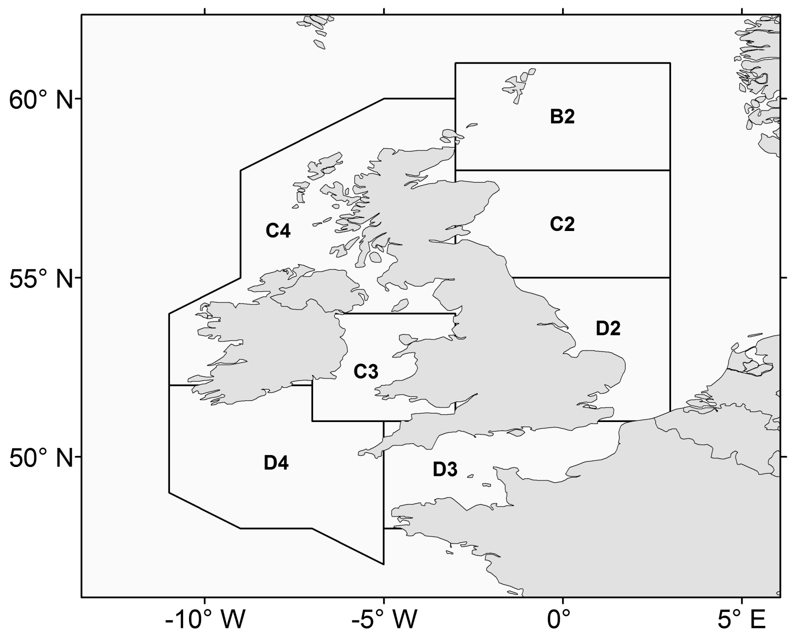
Aims

- Assemble a large collection of biological time series from marine communities around the UK.
- Develop statistical models that are flexible enough to capture the behaviour of these time series.

Aims

- Assemble a large collection of biological time series from marine communities around the UK.
- Develop statistical models that are flexible enough to capture the behaviour of these time series.
- Use these models to look for evidence of regime shifts.

UK marine regions



Region	Series	Years
B2	35	30
C2	65	28
D2	34	30
D3	68	19
D4	47	25
C3	37	19
C4	38	19

324 time series ranging from phytoplankton to seals, contributed by a large number of institutions. Quality control and data preparation took several months.

Principal components

- We want to study the community, so models of individual species won’t answer our questions.

Principal components

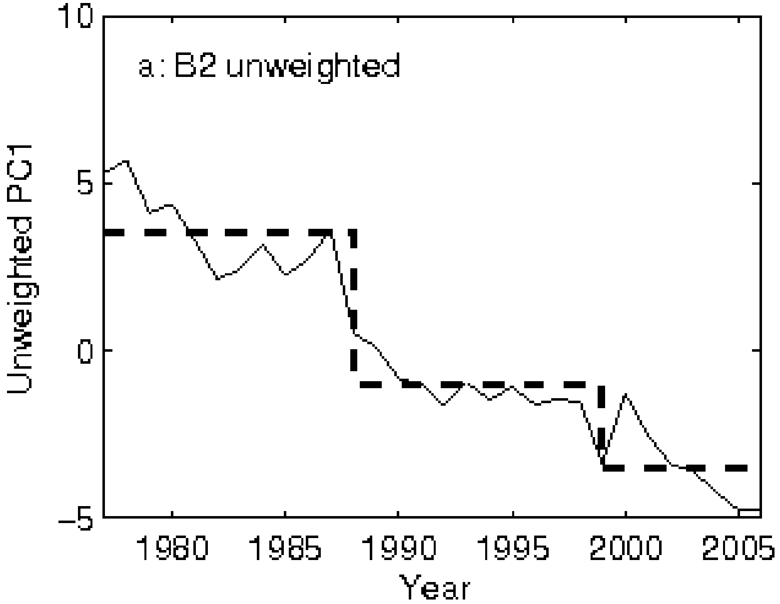
- We want to study the community, so models of individual species won’t answer our questions.
- It isn’t practical to build empirical multivariate models of interactions between large numbers of species.

Principal components

- We want to study the community, so models of individual species won't answer our questions.
- It isn't practical to build empirical multivariate models of interactions between large numbers of species.
- Summarizing the community-level dynamics in the form of the first principal component (PC1) is common practice in studies of regime shifts, and seems appropriate here.

Regime Shift Detection

A common approach known as Regime Shift Detection [a] assumes that the process is stationary except at possible shift points, and that we have independent and identically distributed observations drawn from a normal distribution within each regime.



Region B2 (Northern North Sea)

[a] Rodionov (2004), Geophysical Research Letters 31: L09204

Statistical issues

- Population growth: empirically, biological populations show long-term trends

Statistical issues

- Population growth: empirically, biological populations show long-term trends
- Autocorrelation: observations in successive years are not independent.

Statistical issues

- Population growth: empirically, biological populations show long-term trends
- Autocorrelation: observations in successive years are not independent.
- Measurement error: we cannot observe the sizes of marine populations exactly.

Statistical issues

- Population growth: empirically, biological populations show long-term trends
- Autocorrelation: observations in successive years are not independent.
- Measurement error: we cannot observe the sizes of marine populations exactly.
- Temporal variability: growth rates vary unpredictably from year to year.

Statistical issues

- Population growth: empirically, biological populations show long-term trends
- Autocorrelation: observations in successive years are not independent.
- Measurement error: we cannot observe the sizes of marine populations exactly.
- Temporal variability: growth rates vary unpredictably from year to year.
- Multiple testing: if we do not know where the possible change point is, we have to test each year.

Statistical issues

- Population growth: empirically, biological populations show long-term trends
- Autocorrelation: observations in successive years are not independent.
- Measurement error: we cannot observe the sizes of marine populations exactly.
- Temporal variability: growth rates vary unpredictably from year to year.
- Multiple testing: if we do not know where the possible change point is, we have to test each year.
- We have only a few tens of observations, so we need to work with the simplest plausible model.

Modelling the dynamics of the first principal component

- For a single population i, we could start with a stochastic exponential:

$$N_{i,t+1} = N_{i,t} R_{i,t}$$

  where $N_{i,t}$ is the size of population i at time t, and $R_{i,t}$ is the discrete-time growth rate at time t.

Modelling the dynamics of the first principal component

- For a single population i, we could start with a stochastic exponential:

$$N_{i,t+1} = N_{i,t} R_{i,t}$$

  where $N_{i,t}$ is the size of population i at time t, and $R_{i,t}$ is the discrete-time growth rate at time t.

- Log-transforming this gives us a first-order autoregressive model

$$\log N_{i,t+1} = \log N_{i,t} + \log R_{i,t}$$

Modelling the dynamics of the first principal component

- For a single population i, we could start with a stochastic exponential:

$$N_{i,t+1} = N_{i,t} R_{i,t}$$

  where $N_{i,t}$ is the size of population i at time t, and $R_{i,t}$ is the discrete-time growth rate at time t.

- Log-transforming this gives us a first-order autoregressive model

$$\log N_{i,t+1} = \log N_{i,t} + \log R_{i,t}$$

- The first principal component α_t of the logs of all the population sizes at time t is a linear combination of the individual logged populations, with coefficients that we treat as fixed. Thus

$$\alpha_{t+1} = \alpha_t + S_t$$

  where S_t is a linear combination of the log growth rates of all the populations.

Modelling the growth rate of PC1

- We assume that S_t has both deterministic and stochastic components:

$$S_t = \sum_{j=1}^{k} \beta_j x_{j,t} + \eta_t$$

Modelling the growth rate of PC1

- We assume that S_t has both deterministic and stochastic components:

$$S_t = \sum_{j=1}^{k} \beta_j x_{j,t} + \eta_t$$

- We have explanatory variables $x_{j,t}$ representing patterns of change, with unknown coefficients β_j.

Modelling the growth rate of PC1

- We assume that S_t has both deterministic and stochastic components:

$$S_t = \sum_{j=1}^{k} \beta_j x_{j,t} + \eta_t$$

- We have explanatory variables $x_{j,t}$ representing patterns of change, with unknown coefficients β_j.
- The stochastic component is $\eta_t \sim N(0, \sigma_\eta^2)$, where the amount of true process error σ_η^2 is unknown.

Modelling measurement error

- Measurement error is ubiquitous in ecological time series [a].

[a] Akçakaya et al. (2003) J. Anim. Ecol. 72:698-702

Modelling measurement error

- Measurement error is ubiquitous in ecological time series [a].
- We model the observation of PC1 at time t as

$$y_t = \alpha_t + \epsilon_t$$

[a] Akçakaya et al. (2003) J. Anim. Ecol. 72:698-702

Modelling measurement error

- Measurement error is ubiquitous in ecological time series [a].
- We model the observation of PC1 at time t as

$$y_t = \alpha_t + \epsilon_t$$

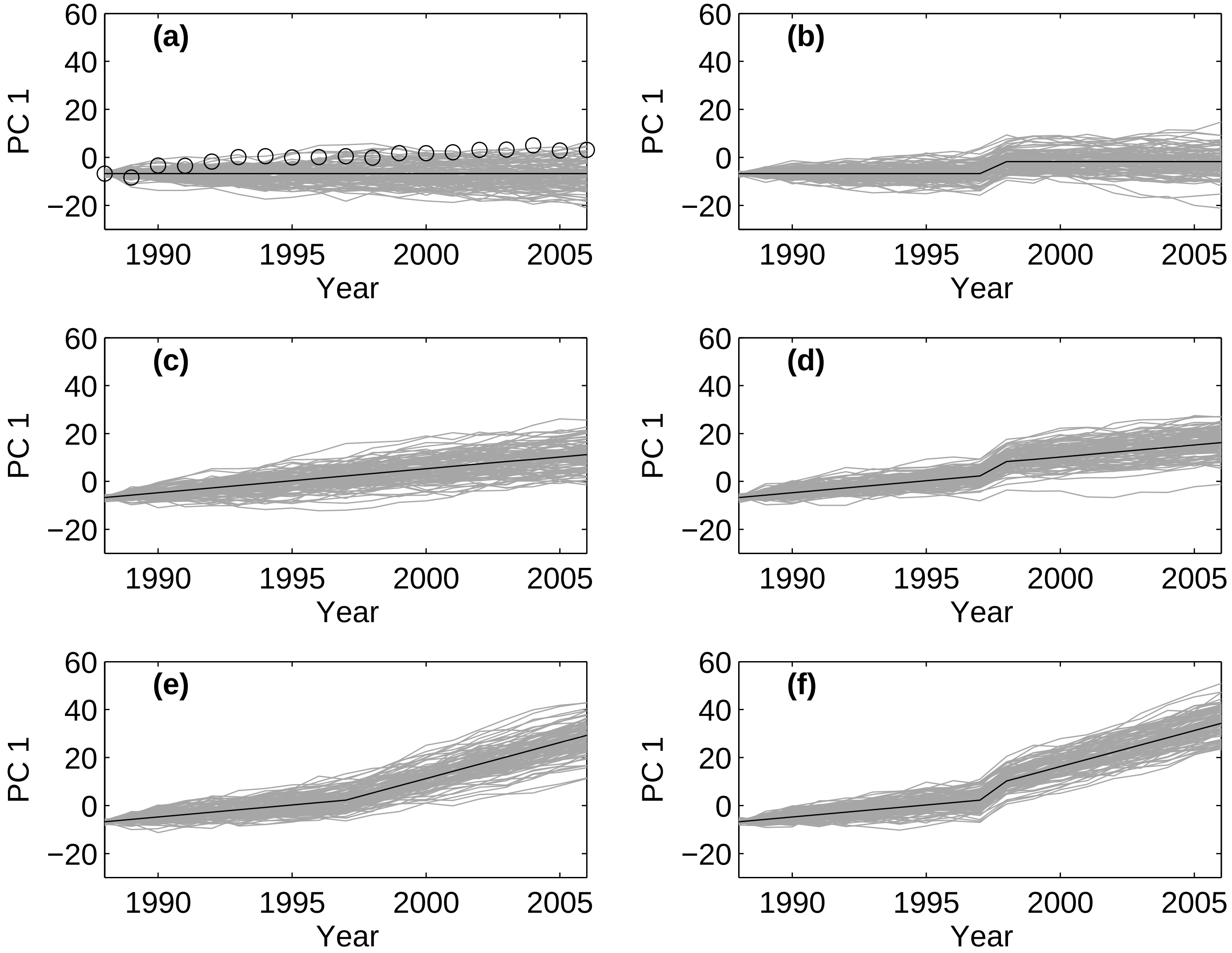
- The measurement error is

$$\epsilon_t \sim N(0, \sigma_\epsilon^2)$$

where σ_ϵ^2 is unknown

[a] Akçakaya et al. (2003) J. Anim. Ecol. 72:698-702

Models considered



Black line: deterministic skeleton. Grey lines: 100 simulated data sets. Black circles: real observations, region C3.

State space models

- Our model is a state-space model[a].

---

[a] Durbin and Koopman (2001), Time series analysis by state space methods

State space models

- Our model is a state-space model[a].
- We can estimate the parameters of this model by maximum likelihood, using a Kalman filter.

[a] Durbin and Koopman (2001), Time series analysis by state space methods

State space models

- Our model is a state-space model[a].
- We can estimate the parameters of this model by maximum likelihood, using a Kalman filter.
- We can compare pairs of models using likelihood ratio statistics.

---

[a] Durbin and Koopman (2001), Time series analysis by state space methods

State space models

- Our model is a state-space model[a].
- We can estimate the parameters of this model by maximum likelihood, using a Kalman filter.
- We can compare pairs of models using likelihood ratio statistics.
- For models with change points, we always select the best possible change point.

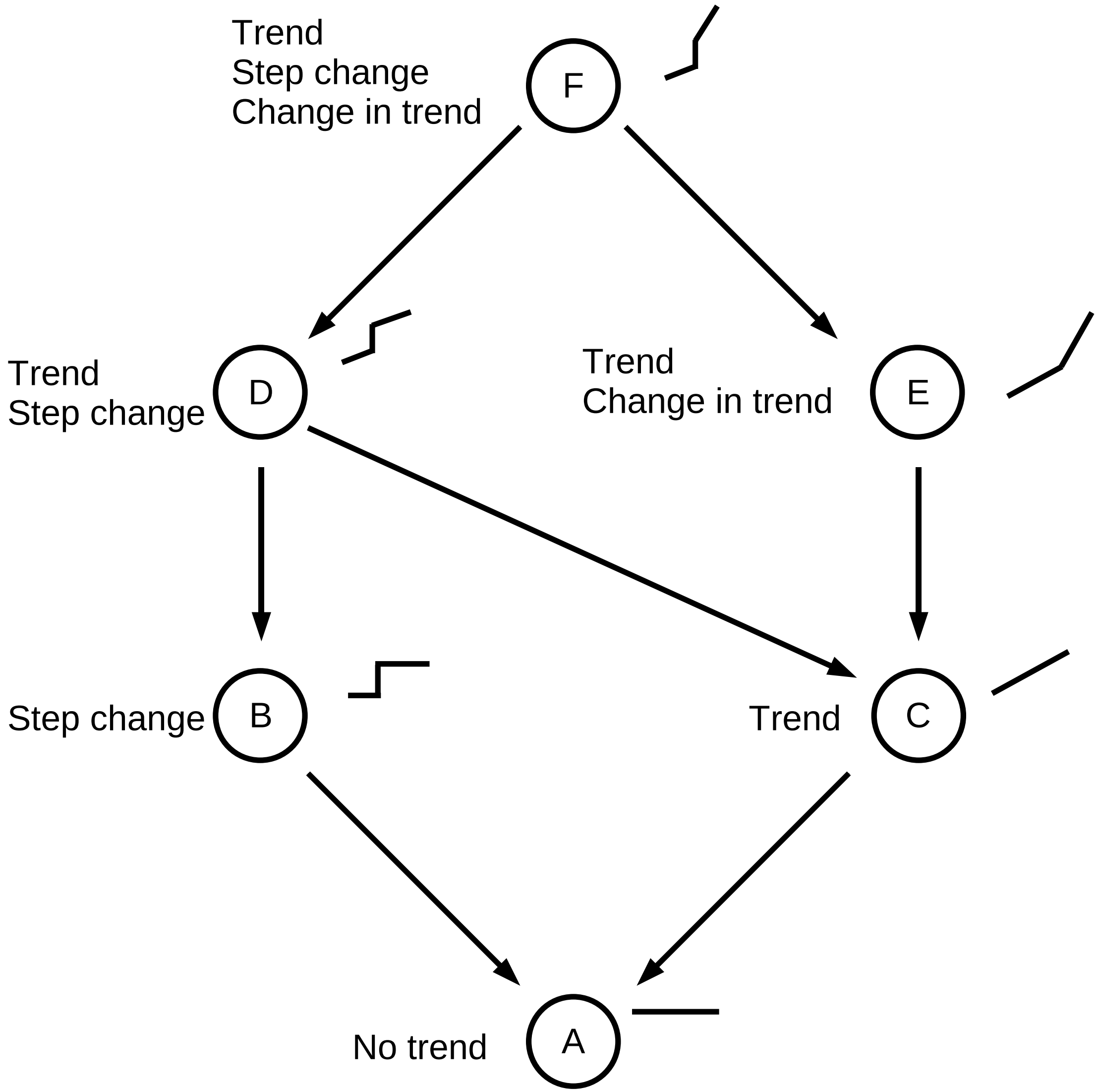
[a] Durbin and Koopman (2001), Time series analysis by state space methods

State space models

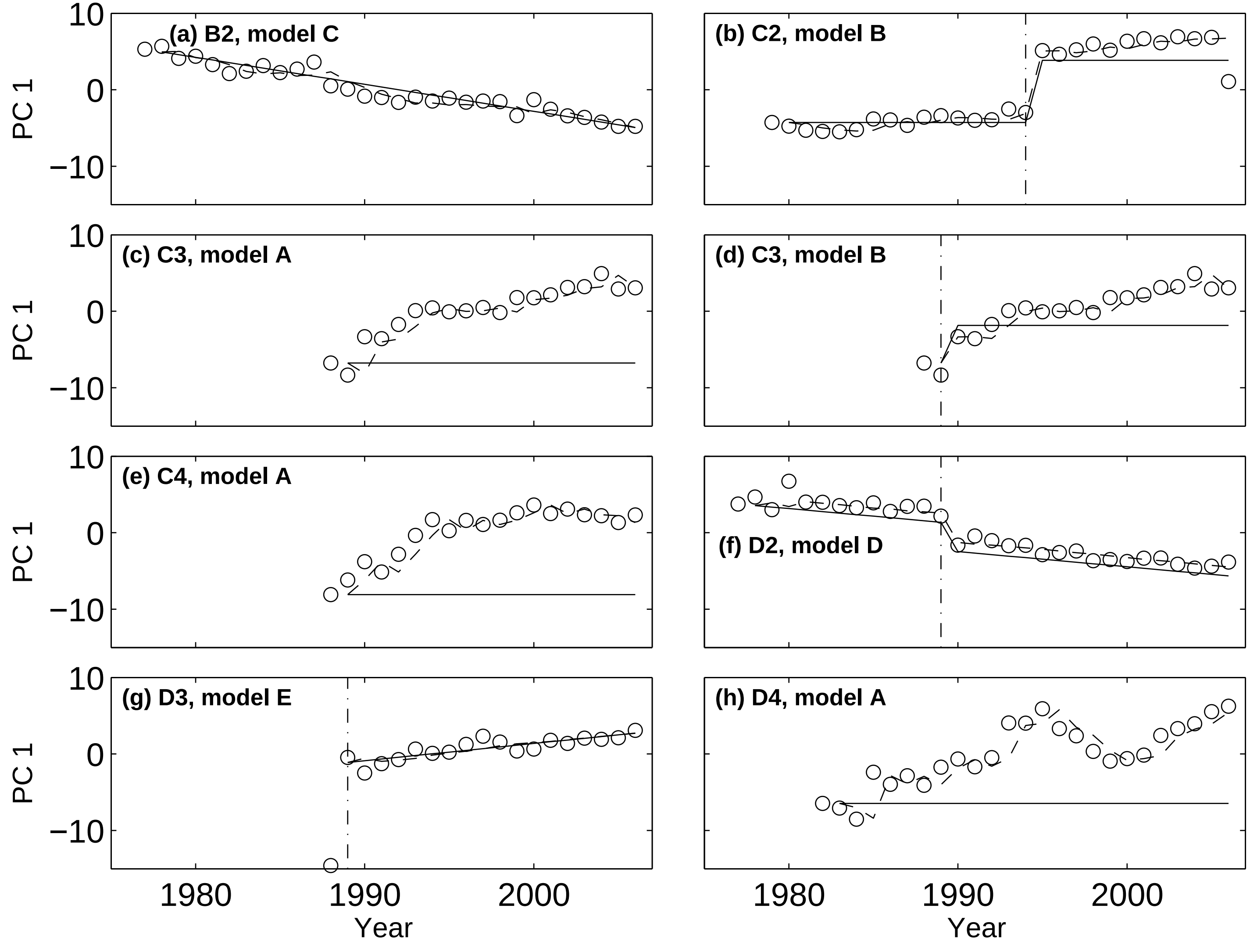
- Our model is a state-space model[a].
- We can estimate the parameters of this model by maximum likelihood, using a Kalman filter.
- We can compare pairs of models using likelihood ratio statistics.
- For models with change points, we always select the best possible change point.
- We deal with autocorrelation and multiple testing by parametric bootstrap: we estimate the null distribution of the likelihood ratio statistic by simulating under the simpler model.

[a] Durbin and Koopman (2001), Time series analysis by state space methods

Model comparisons



Fitted models for each region

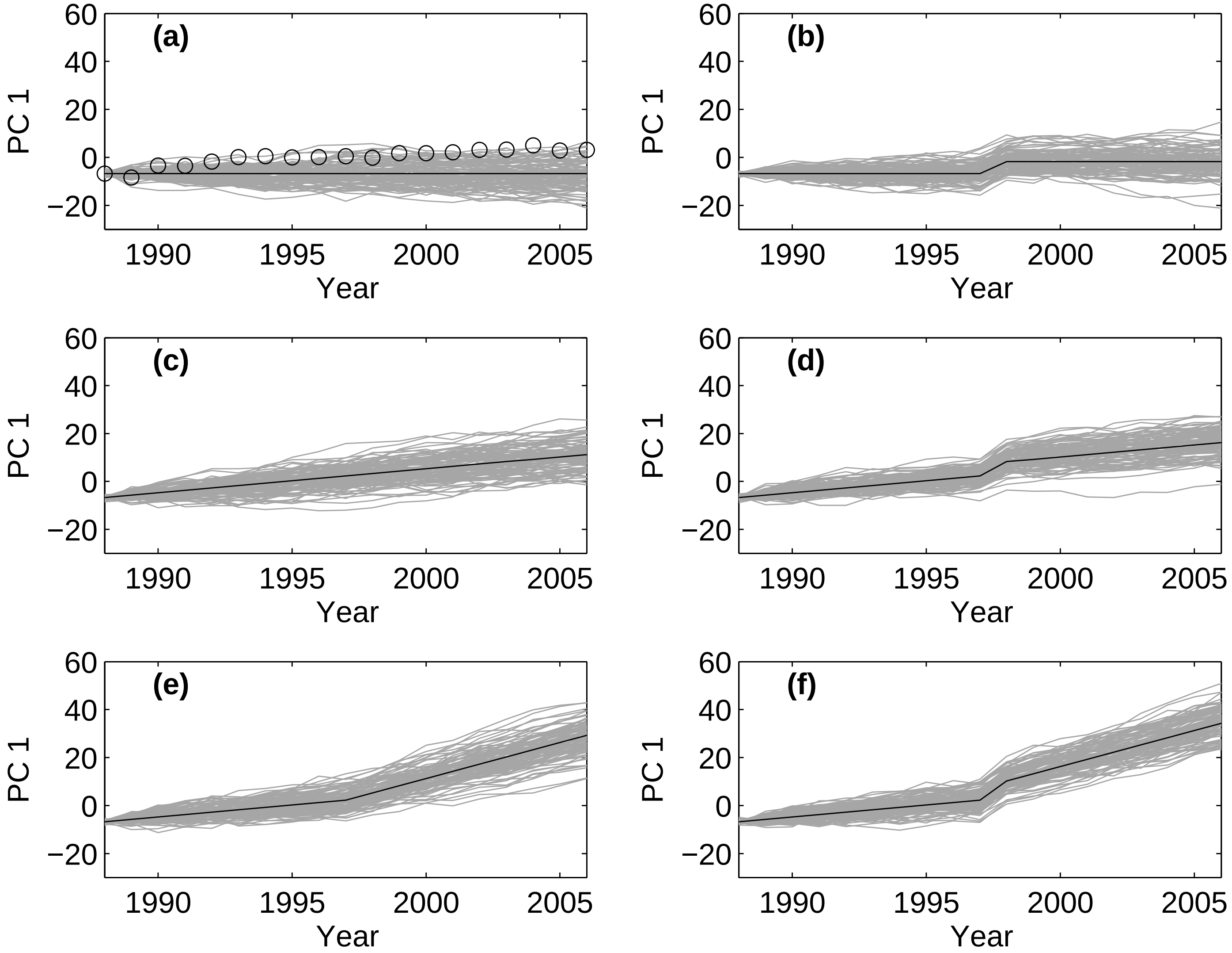


Model selection performance

In simulations where the true model is known, we usually select models that are too simple.

True	Selected					
	A	B	C	D	E	F
A	85	3	2	3	12	4
B	68	24	4	4	1	7
C	81	2	11	4	14	1
D	67	28	5	2	2	7
E	82	2	11	3	7	3
F	66	28	11	4	3	5

Possible patterns of change



Black line: deterministic skeleton. Grey lines: 100 simulated data sets. Black circles: real observations, region C3.

Sensitivity and specificity of step change identification

In simulation studies, when we select a model that includes a step change, the true model usually contained a step change.

True	Selected			
	No step	Step	Ambiguous	Sensitivity
No step	275	15	10	
Step	191	81	28	0.27
Specificity		0.84		

Conclusions

- There were convincing step changes in two out of seven regions. However, three out of seven regions showed trends, either as well as or instead of step changes.
- Models of marine community time series need to allow the varied patterns of change that we see in real data.
- With short time series, we will have low power to detect change points. However, the change points we do detect are likely to be reliable.

Images from MarLIN database, copyright Helmar Hinz, Sue Scott, and Robert Keech.