Detecting causality with Convergent Cross Mapping

August Birk, Clotilde Armand Prætorius, Frederikke Rasmussen & Sune Halkjær

09/03/2023

UNIVERSITY OF COPENHAGEN

RESEARCH ARTICLE

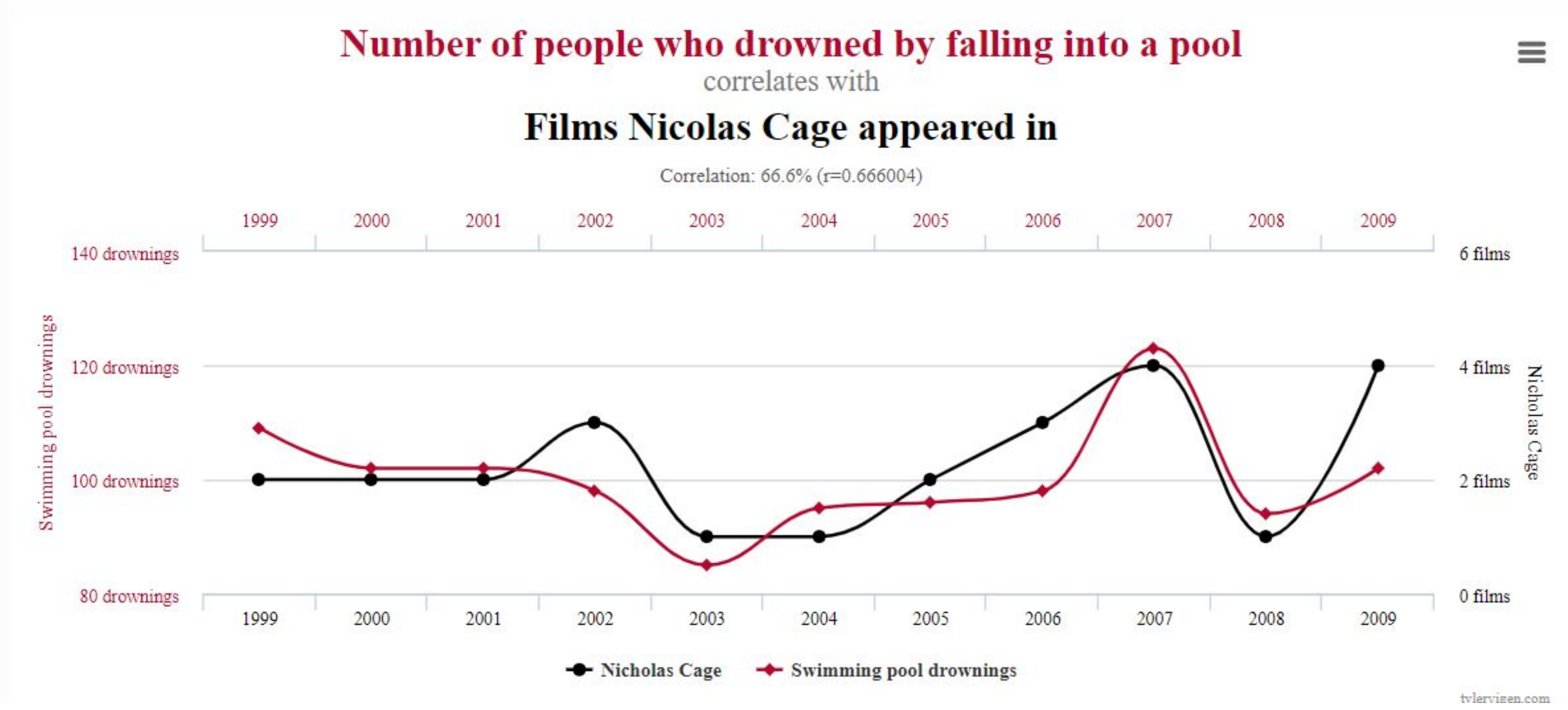
Detecting Causality in Complex Ecosystems

George Sugihara,[1]* Robert May,[2] Hao Ye,[1] Chih-hao Hsieh,[3]* Ethan Deyle,[1]
Michael Fogarty,[4] Stephan Munch[5]

Identifying causal networks is important for effective policy and management recommendations on climate, epidemiology, financial regulation, and much else. We introduce a method, based on nonlinear state space reconstruction, that can distinguish causality from correlation. It extends to nonseparable weakly connected dynamic systems (cases not covered by the current Granger causality paradigm). The approach is illustrated both by simple models (where, in contrast to the real world, we know the underlying equations/relations and so can check the validity of our method) and by application to real ecological systems, including the controversial sardine-anchovy-temperature problem.

Causality and correlation

- Correlation is not necessary for causality neither does it directly imply causality



Causality and correlation

- Correlation is not necessary for causality neither does it directly imply causality

- Detecting causality is important for understanding complex systems

- The article presents two methods for detecting causality
  - Granger causality (GC)
  - convergent cross mapping (CCM)

Granger Causality (GC)

- X “Granger causes” Y if the predictability of Y declines when X is removed
- Good for systems that are
  - purely stochastic
  - separable
  - linear
  - strongly coupled
- Missing a method that
  - addresses nonseparable systems
  - identifies weakly coupled variables
  - distinguishes interactions among variables from the effect of shared driving variables

CCM

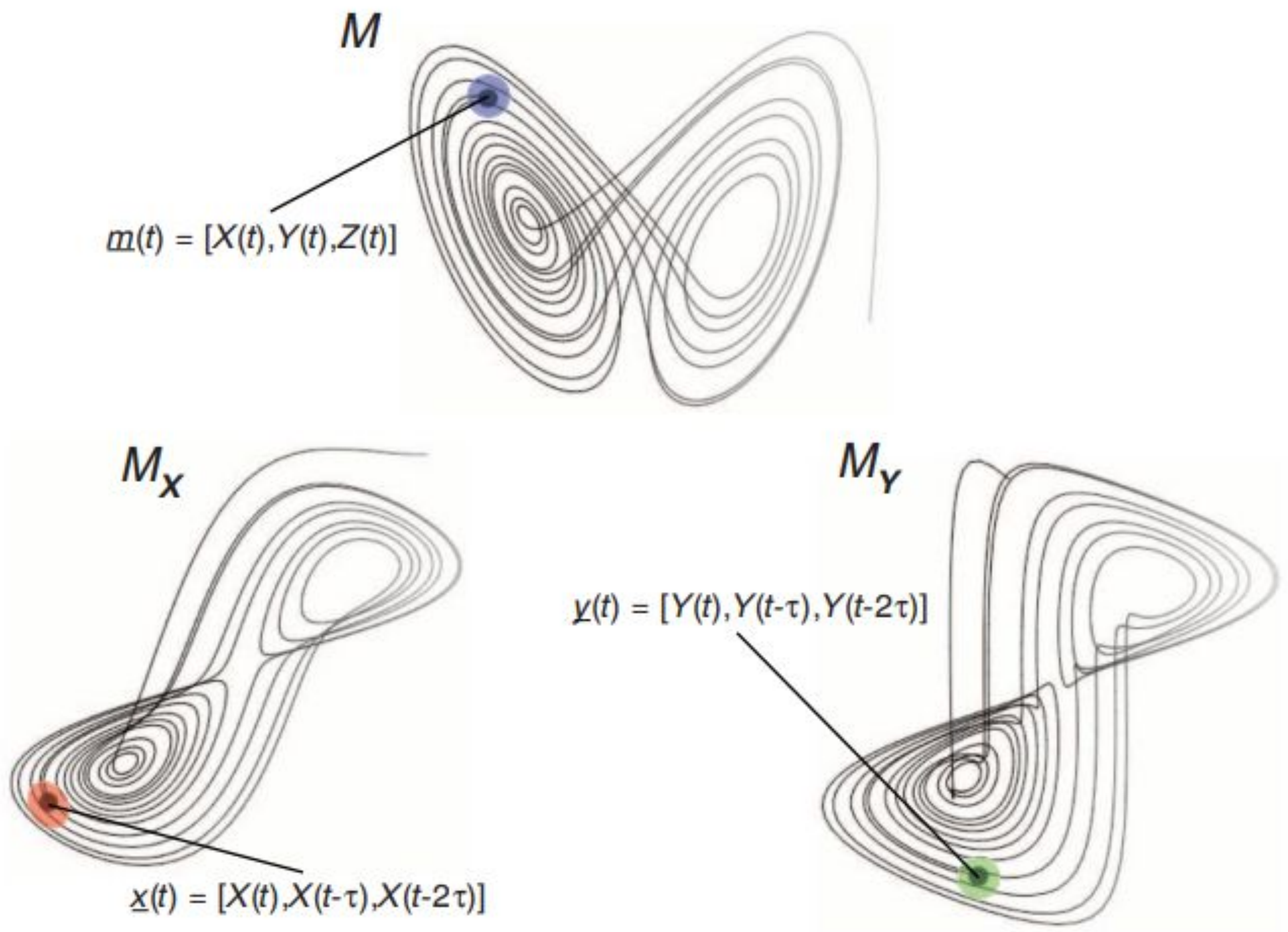
- Extends to nonseparable weakly connected dynamic systems. Aimed at identifying causation in ecological time series (not purely stochastic)
- Tests for causation by measuring the extent to which the historical record of Y values can reliably estimate states of X
  - X causes Y, we can predict X from Y
- Cross mapping
  - Given points on one manifold, M_y look for corresponding points on other manifold M_x
  - Shadow manifolds
- Convergence
  - Predictability of Y, from X should increase with increasing data

CCM

- Extends to nonseparable weakly connected dynamic systems. Aimed at identifying causation in ecological time series (not purely stochastic)
- Tests for causation by measuring the extent to which the historical record of Y values can reliably estimate states of X
  - X causes Y, we can predict X from Y
- Cross mapping
  - Given points on one manifold, M_y look for corresponding points on other manifold M_x
  - Shadow manifolds
- Convergence
  - Predictability of Y, from X should increase with increasing data

CCM

- Extends to nonseparable weakly connected dynamic systems. Aimed at identifying causation in ecological time series (not purely stochastic)
- Tests for causation by measuring the extent to which the historical record of Y values can reliably estimate states of X
  - X causes Y, we can predict X from Y
- Cross mapping
  - Given points on one manifold, M_y look for corresponding points on other manifold M_x
  - Shadow manifolds

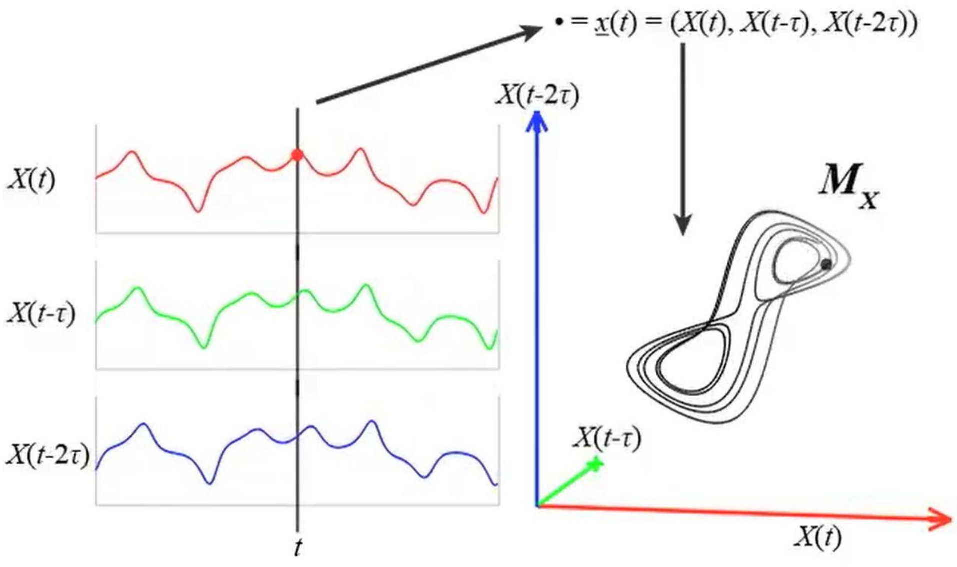

M
m(t) = [X(t),Y(t),Z(t)]
MX
x(t) = [X(t),X(t-τ),X(t-2τ)]
MY
y(t) = [Y(t),Y(t-τ),Y(t-2τ)]

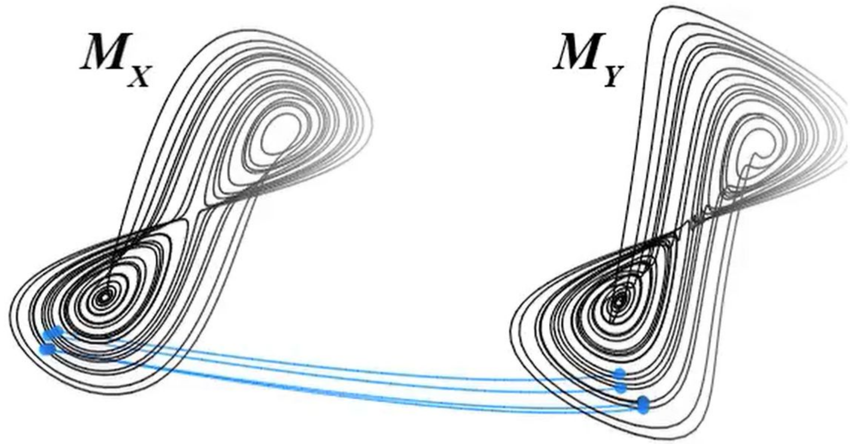
CCM

- Extends to nonseparable weakly connected dynamic systems. Aimed at identifying causation in ecological time series (not purely stochastic)
- Tests for causation by measuring the extent to which the historical record of Y values can reliably estimate states of X
  - X causes Y, we can predict X from Y
- Cross mapping
  - Given points on one manifold, M_y look for corresponding points on other manifold M_x
  - Shadow manifolds
- Convergence
  - Predictability of Y, from X should increase with increasing data

CCM algorithm

1. Create shadow manifold from X and Y
2. Locate lagged data in the shadow manifold
3. Find the nearest neighbours
4. Construct the predictions of Y from X
5. X & Y are coupled if corresponding points are densely clustered
6. Correlation between predicted Y and true Y are calculated





Results

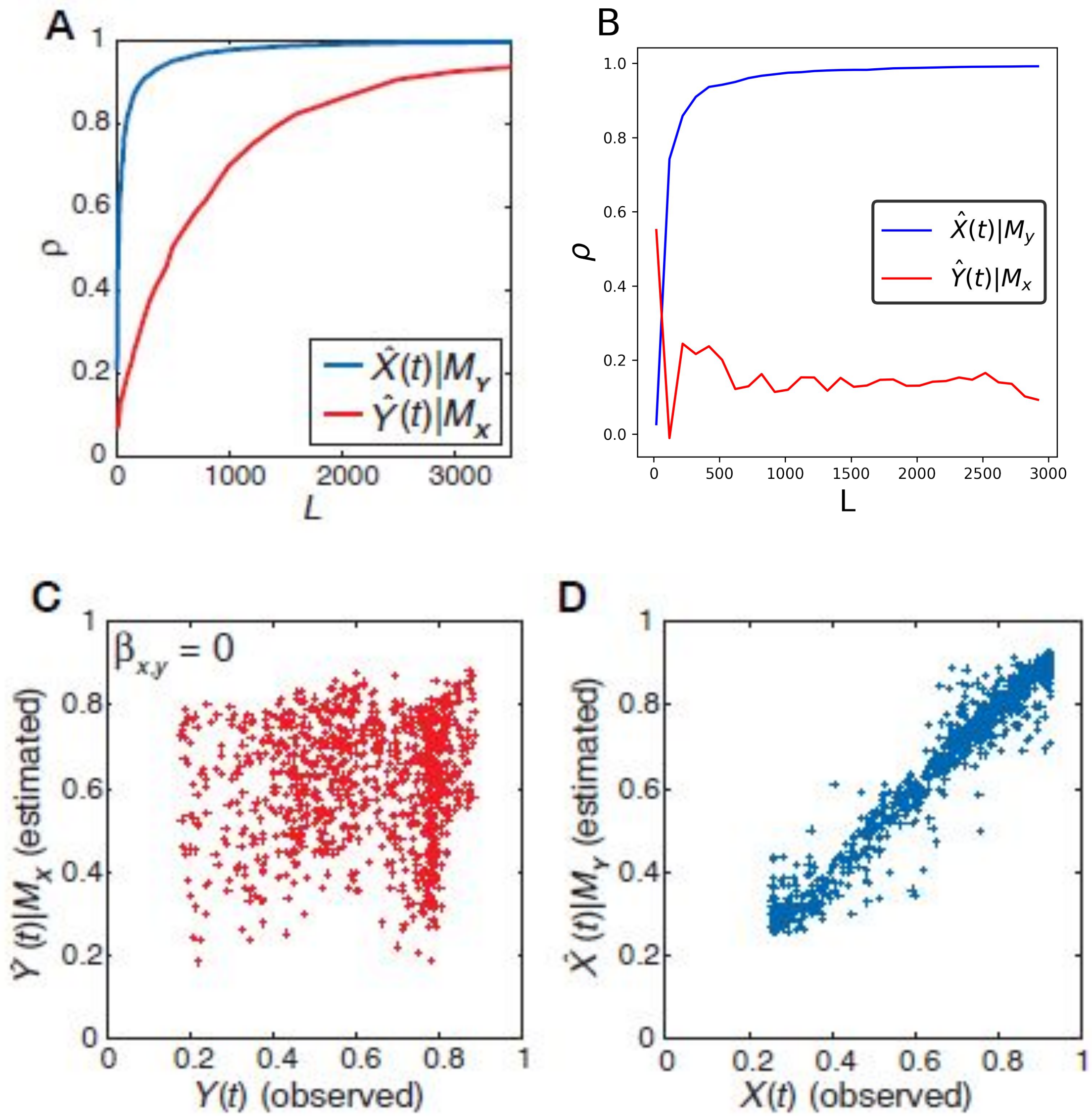
Simple models

- Bidirectional

$$X(t+1) = X(t)[r_x - r_x X(t) - \beta_{x,y} Y(t)]$$
$$Y(t+1) = Y(t)[r_y - r_y Y(t) - \beta_{y,x} X(t)]$$

  - For $\beta_{x,y}$= 0.02 and $\beta_{y,x}$=0.1

- Unidirectional
  - $\beta_{x,y}$= 0
  - “Synchrony” problem

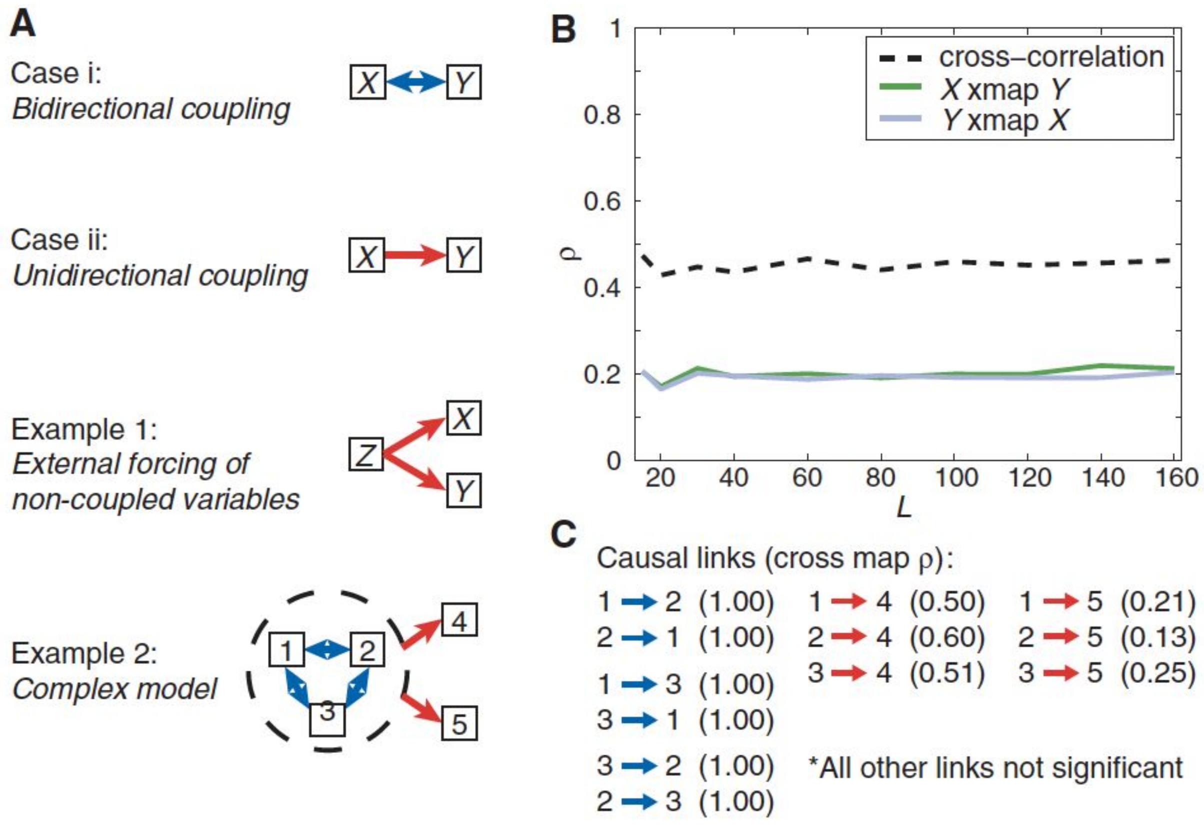


Results

More complex systems

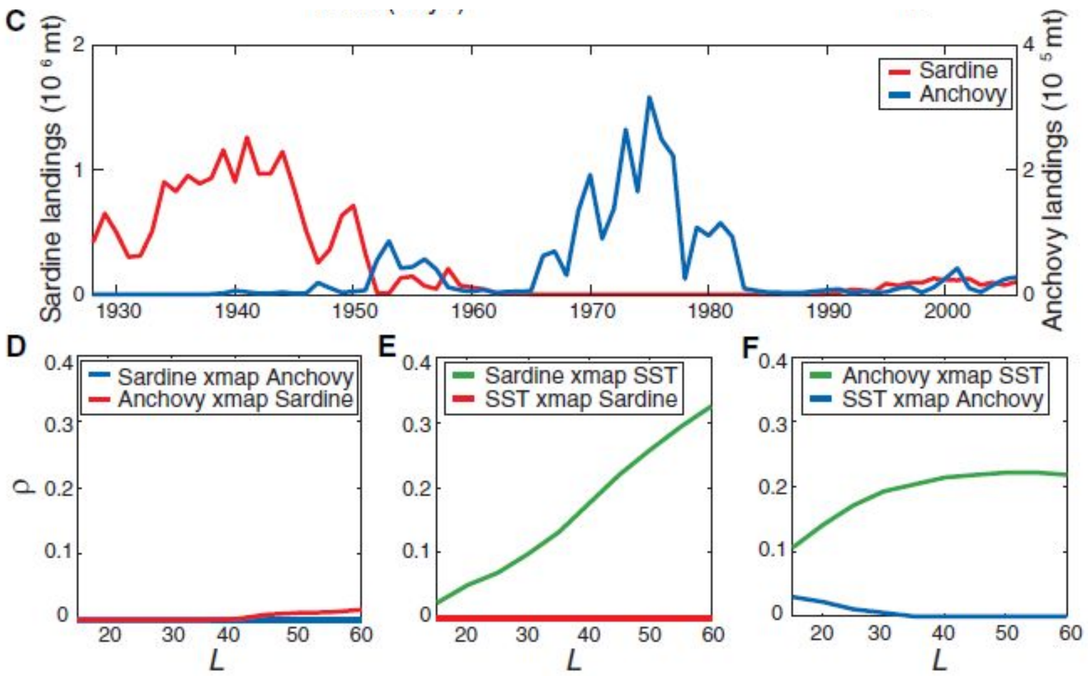
- External forcing of non-coupled variables

- 5-species system



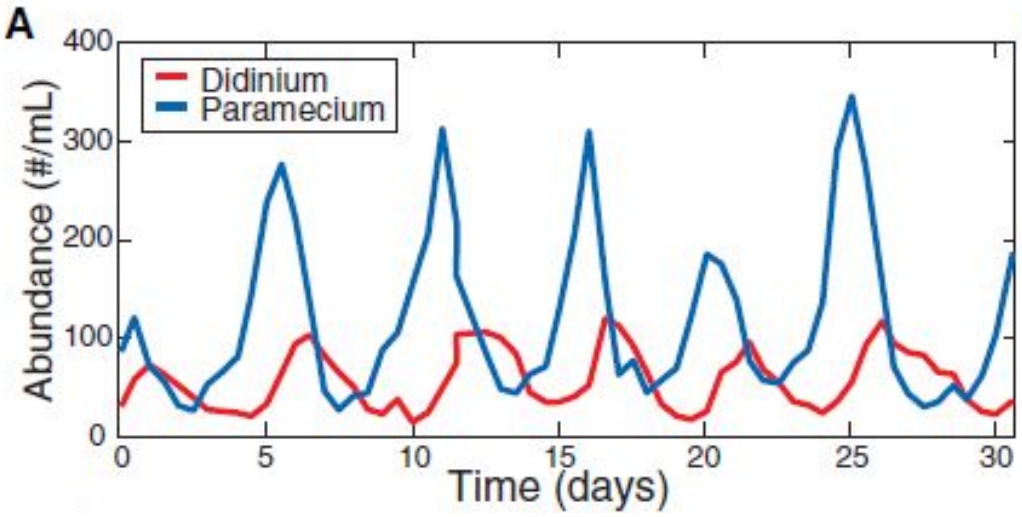
Results

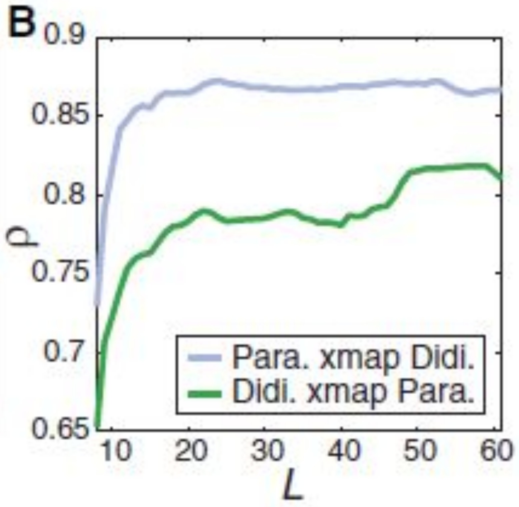
- Sardine-anchovy-temperature problem
- Two competing theories:
  - Competition for resources
  - Driven by external factor (e.g. temperature)
- CCM suggests:
  - No causation between the fishes
  - None of them affect SST
  - SST affecting both asymmetrically and unidirectional



Results

- The Paramecium-didinium-relation
- A “classic” in Predator-Prey-models
- Shows a bidirectional coupling with CCM
  - Heavier influence by predator than prey (henceforth still asymmetrical)





Conclusion

- Important to understand connection and distinction between correlation and causality
- GC and CCM are methods that aim to detect causality
- The article shows that CCM is a good alternative to GC in cases where GC does not apply
- CCM is tested and shown to perform well and as expected
  - Predicts couplings
  - Distinguishes interactions from shared driving variables