Analyzing complex social-ecological system simulations

Gary Polhill

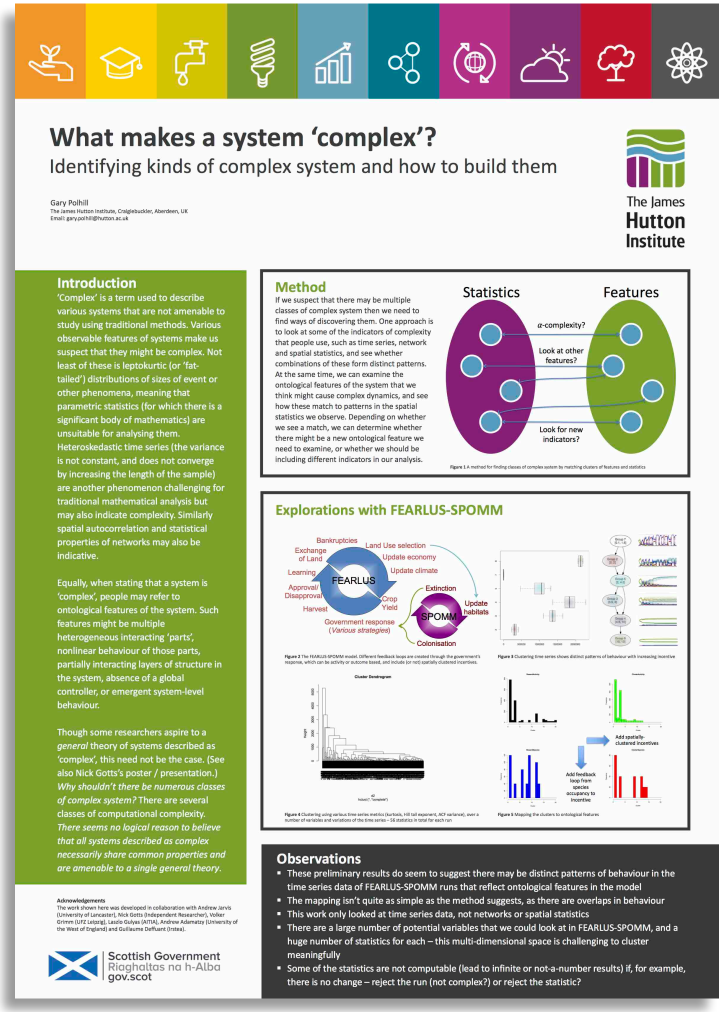
The James Hutton Institute

Outline

- Background and motivation
- Complexity of analyzing social simulations
- Indicators of complexity
- Towards a method

Background and motivation

- Various bids and projects
  - CAVES, MIRACLE
  - OCCAM, CSoS
- Is 'complexity' one thing…
- …or numerous quite distinct kinds of system…
  - …with characteristic patterns of behaviour



What makes a system 'complex'?

Identifying kinds of complex system and how to build them

Gary Polhill
The James Hutton Institute, Craigiebuckler, Aberdeen, UK
Email: gary.polhill@hutton.ac.uk

The James Hutton Institute

Introduction

'Complex' is a term used to describe various systems that are not amenable to study using traditional methods. Various observable features of systems make us suspect that they might be complex. Not least of these is leptokurtic (or 'fat-tailed') distributions of sizes of event or other phenomena, meaning that parametric statistics (for which there is a significant body of mathematics) are unsuitable for analysing them. Heteroskedastic time series (the variance is not constant, and does not converge by increasing the length of the sample) are another phenomenon challenging for traditional mathematical analysis but may also indicate complexity. Similarly spatial autocorrelation and statistical properties of networks may also be indicative.

Equally, when stating that a system is 'complex', people may refer to ontological features of the system. Such features might be multiple heterogeneous interacting 'parts', nonlinear behaviour of those parts, partially interacting layers of structure in the system, absence of a global controller, or emergent system-level behaviour.

Though some researchers aspire to a *general* theory of systems described as 'complex', this need not be the case. (See also Nick Gotts's poster / presentation.) *Why shouldn't there be numerous classes of complex system?* There are several classes of computational complexity. *There seems no logical reason to believe that all systems described as complex necessarily share common properties and are amenable to a single general theory.*

Method

If we suspect that there may be multiple classes of complex system then we need to find ways of discovering them. One approach is to look at some of the indicators of complexity that people use, such as time series, network and spatial statistics, and see whether combinations of these form distinct patterns. At the same time, we can examine the ontological features of the system that we think might cause complex dynamics, and see how these match to patterns in the spatial statistics we observe. Depending on whether we see a match, we can determine whether there might be a new ontological feature we need to examine, or whether we should be including different indicators in our analysis.

Statistics
Features
α-complexity?
Look at other features?
Look for new indicators?

Figure 1 A method for finding classes of complex system by matching clusters of features and statistics

Explorations with FEARLUS-SPOMM

Figure 4 Clustering using various time series metrics (kurtosis, Hill tail exponent, ACF variance), over a number of variables and variations of the time series – 56 statistics in total for each run

Figure 5 Mapping the clusters to ontological features

Observations

- These preliminary results do seem to suggest there may be distinct patterns of behaviour in the time series data of FEARLUS-SPOMM runs that reflect ontological features in the model
- The mapping isn't quite as simple as the method suggests, as there are overlaps in behaviour
- This work only looked at time series data, not networks or spatial statistics
- There are a large number of potential variables that we could look at in FEARLUS-SPOMM, and a huge number of statistics for each – this multi-dimensional space is challenging to cluster meaningfully
- Some of the statistics are not computable (lead to infinite or not-a-number results) if, for example, there is no change – reject the run (not complex?) or reject the statistic?

Acknowledgements
The work shown here was developed in collaboration with Andrew Jarvis (University of Lancaster), Nick Gotts (Independent Researcher), Volker Grimm (UFZ Leipzig), Laszlo Gulyas (AITIA), Andrew Adamatzy (University of the West of England) and Guillaume Deffuant (Irstea).

Scottish Government
Riaghaltas na h-Alba
gov.scot

Complexity of social simulations

- Characterising a system's behaviour is challenging
    - Individual data of various types
    - Spatial data
    - Network data
    - Time-series data
        - Emergent system-level properties
        - Individual, spatial and network
- Numerous methods and statistics for each variable
- 'Summary' of a system difficult

Methods (Lee et al. 2015)

- Issues of non-normal and multivariate data
- Significance testing
    - Critique that sample size can be arbitrarily increased until 'inconsequential' significance observed (see also White et al. 2013 Oikos)
        - Coefficient of variance $\sigma/\mu < E$ used by Lorscheid et al. (2012, CMOT)
- Stonedahl & Wilensky (2011, LNCS) use GAs to search for parameters that produce a particular emergent effect

Methods (Lee et al. 2015)

- Time series
    - Decomposition
        - trend, seasonal, cyclical (irregular periodicity) and random
    - Dynamic Time Warping for comparison (Keogh & Ratanamahatana, 2005, K&IS)
- Space
    - Lee et al. mention several without recommending any
        - e.g. Cohen's kappa, Moran's I
        - Pontius (2002, PE&RS) compares at multiple spatial resolutions

Indicators (Hetman & Magnuszewski 2008)

The James Hutton Institute

- Networks
  - Clustering coefficients
    - Various measures of how likely a friend of a friend is also your friend
  - Assortativity
    - ‘Degree homophily’ (you are more likely to interact with other people having a similar node degree to you)
  - Density
    - How many of the possible connections are made
      - Denser networks more resilient
      - Network diameter also a possible metric

Indicators (Hetman & Magnuszewski 2008)

- Space
  - Recommend Moran's I
    - -1: maximally uncorrelated; 0: no correlation; +1: maximally correlated
- Time
  - Kurtosis
    - -ve: thin-tailed; 0: normal; +ve: fat tailed
  - Power law: Hill tail exponent
  - Heteroskedasticity: Analysis of ACF
    - Not strictly automatable
      - Lag choice; inspection of ACF plot

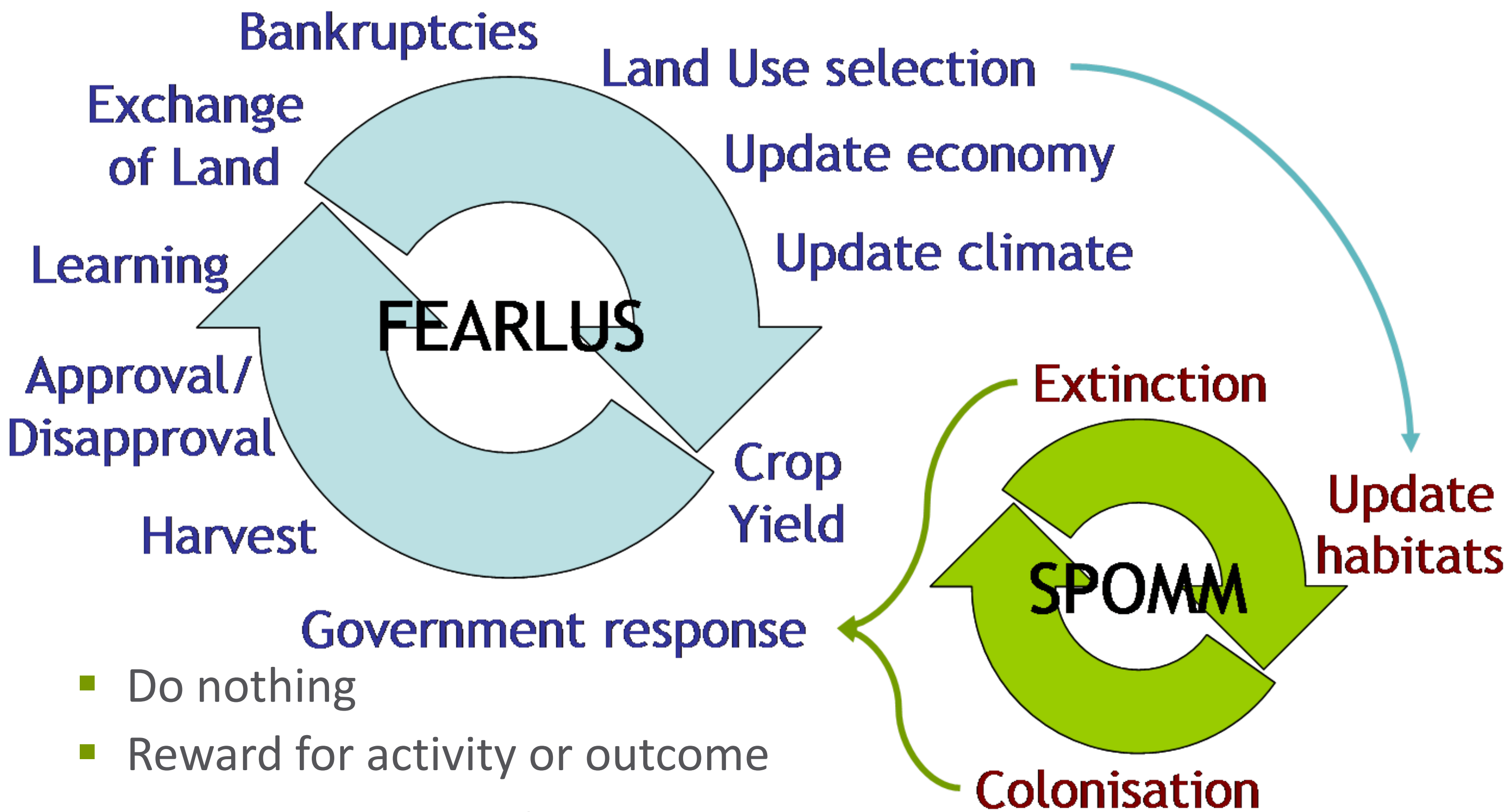
(Not) Indicators (Shalizi 2006)

- Kolmogorov complexity
    - But not computable and often (wrongly) interpreted as ‘gzip’ complexity
        - Still, there might be an indicator there
- Logical Depth
    - Ditto (but no gzip equivalent)
- Minimum description length + stochastic complexity
    - Feasible to compute
    - But still about the number of bits you need to encode some dynamics

Possible indicators (Shalizi 2006)

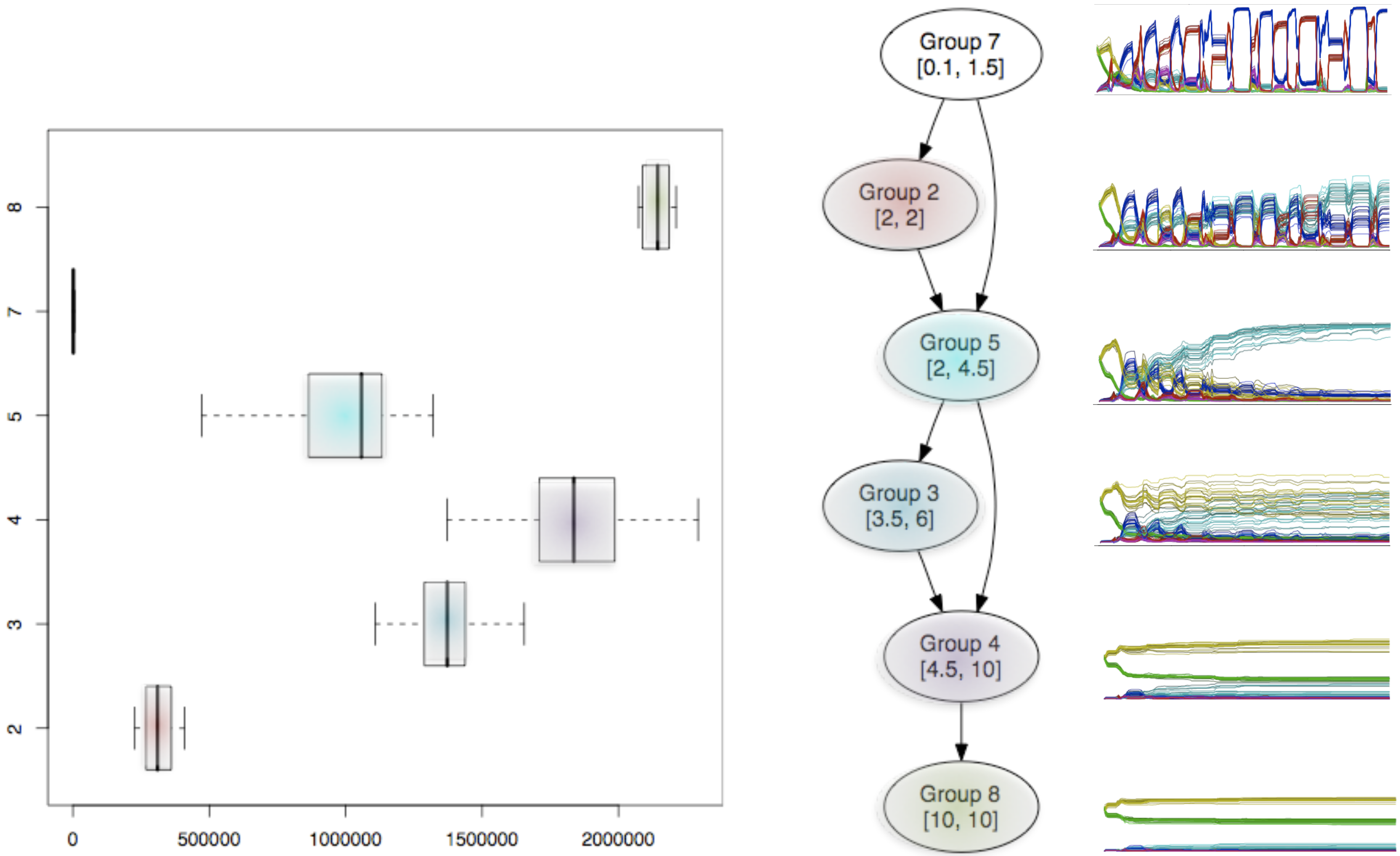
- Complexity as the *minimum* amount of information needed to encode a system state for optimal prediction (Grassberger)
- Grassberger-Crutchfield-Young complexity
  - 'Causal states': equivalence classes of historical states such that they have the same probabilities over the future states
  - Information content of these causal states as a complexity measure
- Power laws
  - Not necessarily complex
  - Things that look like power laws might be other kinds of fat-tailed distribution

FEARLUS SPOMM

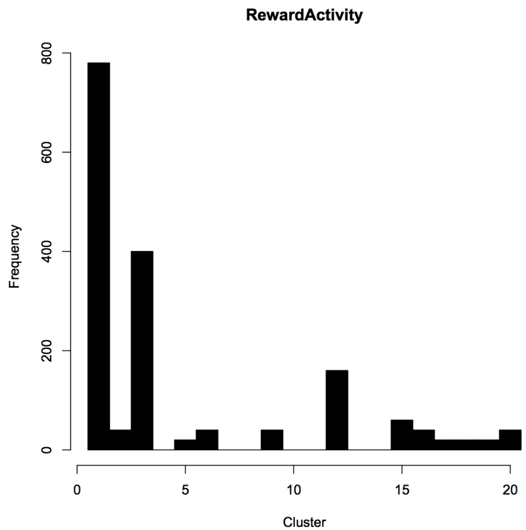


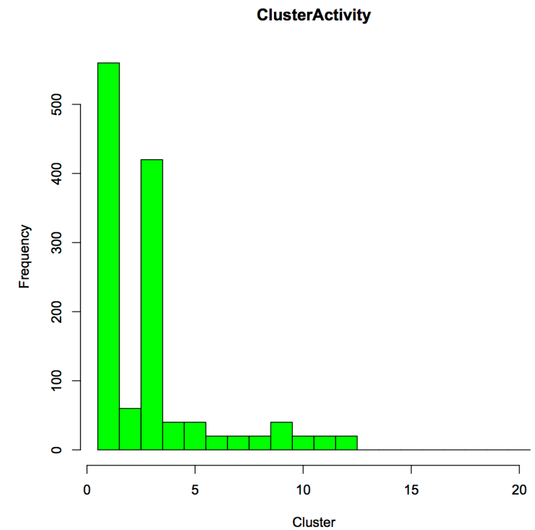
- Do nothing
- Reward for activity or outcome
  - Feedback from species
- Reward individually or ‘clustered’
  - Feedback from neighbours

Previous analysis (1)

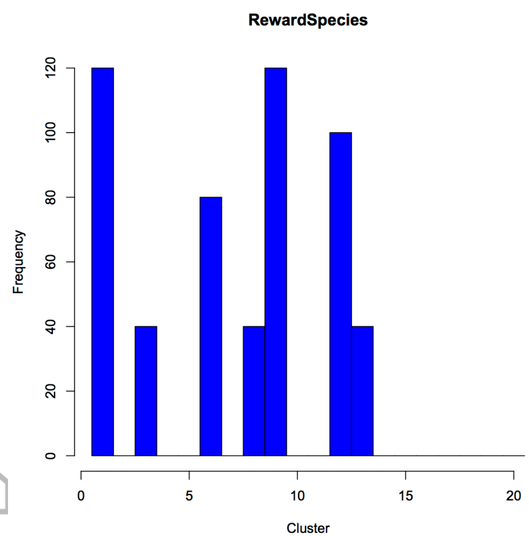


Previous analysis (2)

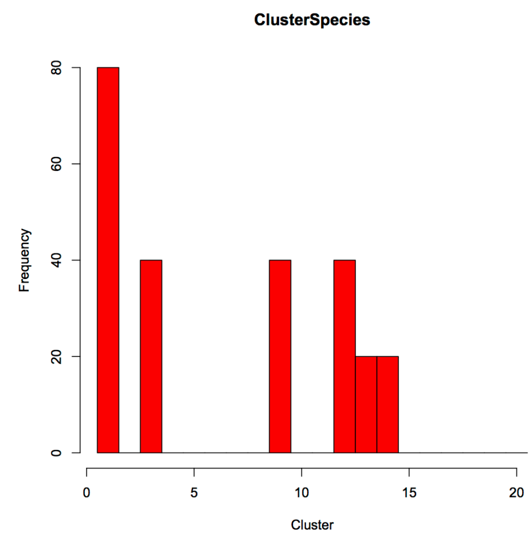




Add spatially-clustered incentives

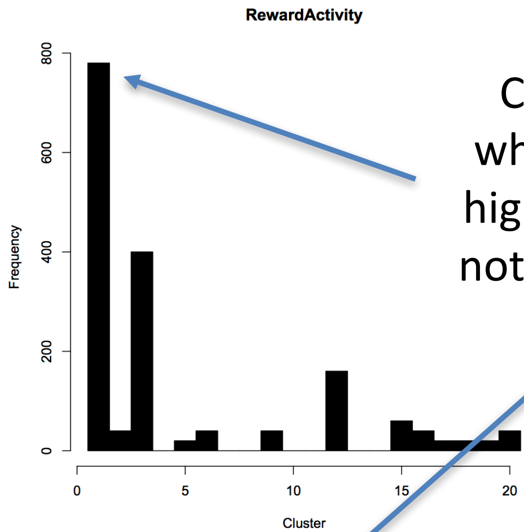


Add feedback loop from species

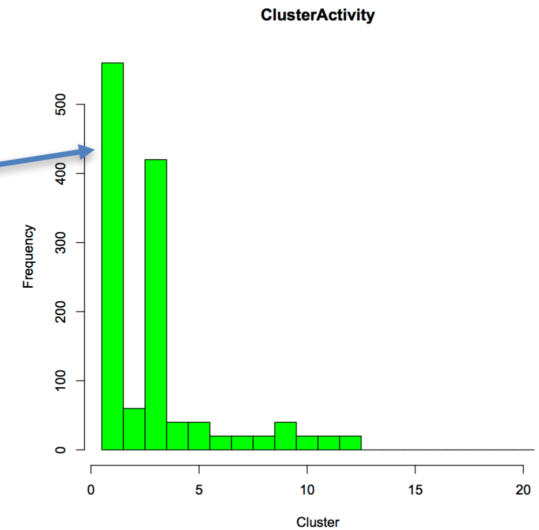


Previous analysis (2)

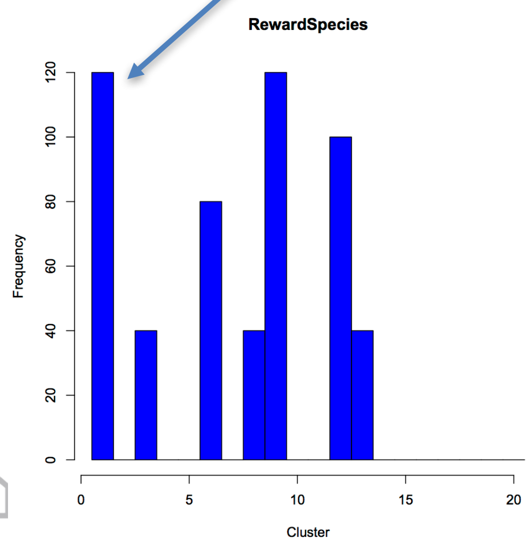
The James Hutton Institute



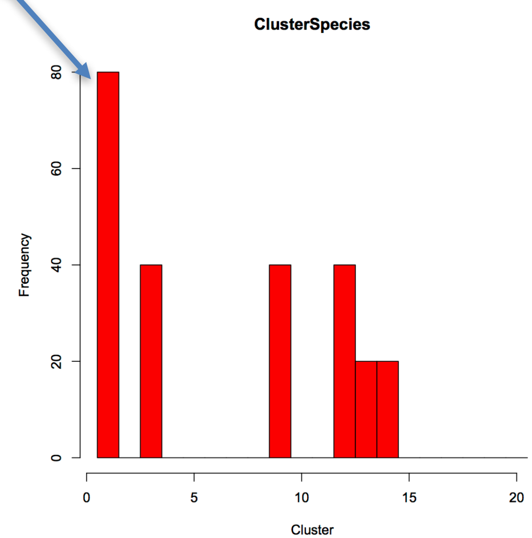
Commonality where incentive high enough that nothing changes?



Add spatially-clustered incentives



Add feedback loop from species

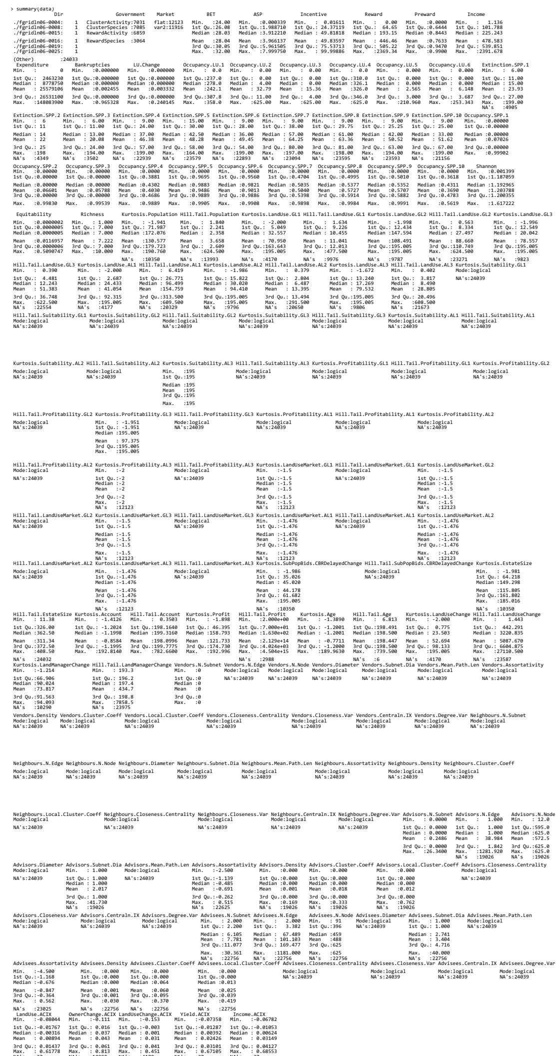


Experiment design

- Use Hetman-Magnuszewski indicators
- Monte Carlo sample in parameter space
    - Five switches and dials
        - Policy design
        - Market time series
        - Aspiration
        - Costs
        - Incentive
    - ~100k runs
- Analyse the results looking for patterns of behaviour

Results?

- 166 variables!
- Several are uncomputable for a significant proportion of runs
  - No network
    - No advice or exchange of land
  - No dynamics
    - No change happens because everyone is happy
  - 166 => 58
    - Remove vars with *all* NA
  - 58 => 17
    - Remove vars with >= 25% NA
- Analysis techniques don't always accept 'NA's



Workflow

- Dimensionality reduction and simplification
    - Principal components analysis (PCA)
    - Feature selection
    - Discretization
- Classification
    - Clustering
    - Other unsupervised machine learning
- Mapping to ontological features
    - Supervised machine learning
        - Decision trees?
        - Something Bayesian?

(No) conclusion

- Lots of indicators mean a significant task to analyse all the data
- Indicators need to have 'null' or NA data when not computable
    - Reject these runs? (Not complex?)
    - Use numbers representing 'normal'
- Data analysis techniques need no NAs
- Previous work does suggest there may be identifiable links between ontological properties in FEARLUS-SPOMM and patterns of dynamics

Acknowledgements / References

Paulina Hetman & Piotr Magnuszewski
(Wrocław University of Technology)
"CAVES Generalisation Framework", 2008

Ju-Sung Lee, Tatiana Filatova, Arika Ligmann-Zielinska et al.
(University of Twente and various other places)
"The complexities of agent-based modeling analysis", 2015, JASSS 18/4/4

Shalizi, C. R. (2006) Methods and techniques of complex systems science: An overview. In Deisboeck, T. S. and Yasha Kresh, J. (eds.) *Complex Systems Science in Biomedicine*. New York: Springer.

Thanks also to various collaborators on the OCCAM and CSoS bids.