Mining Relationships among Variables in Large Datasets from Complex Systems [MIRACLE]

Terry Dawson (University of Dundee)

and **Gary Polhill** (The James Hutton Institute)

with

Dawn Parker, Xiongbing Jin, Kirsten Robinson (University of Waterloo);
Tatiana Filatova, Alexey Voinov (University of Twente);
Michael Barton (Arizona State University)

The James Hutton Institute

Project overview

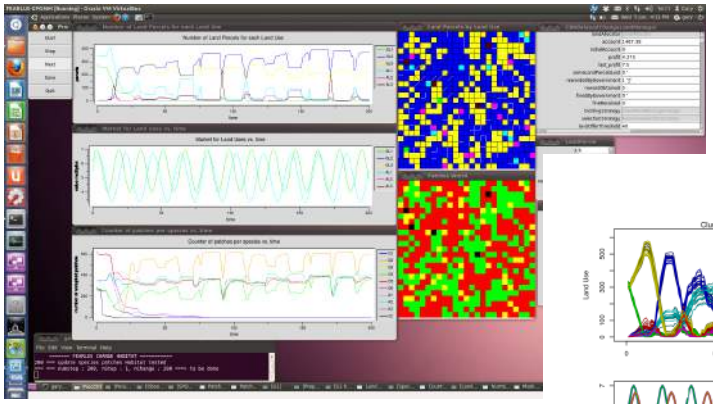
The James Hutton Institute

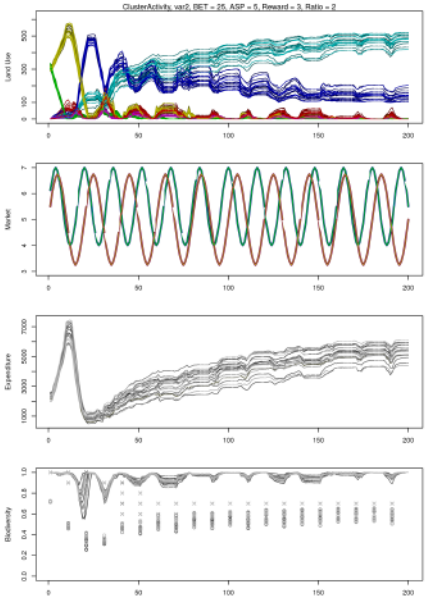
- Complex systems have statistical properties that challenge their analysis using traditional statistics:
    - Path dependence
    - Irreducibility
    - Nonlinearity
    - Leptokurtic distributions of phenomena
    - Heteroskedastic time series
- Models need to use large parameter sweeps, but data can be difficult to summarise

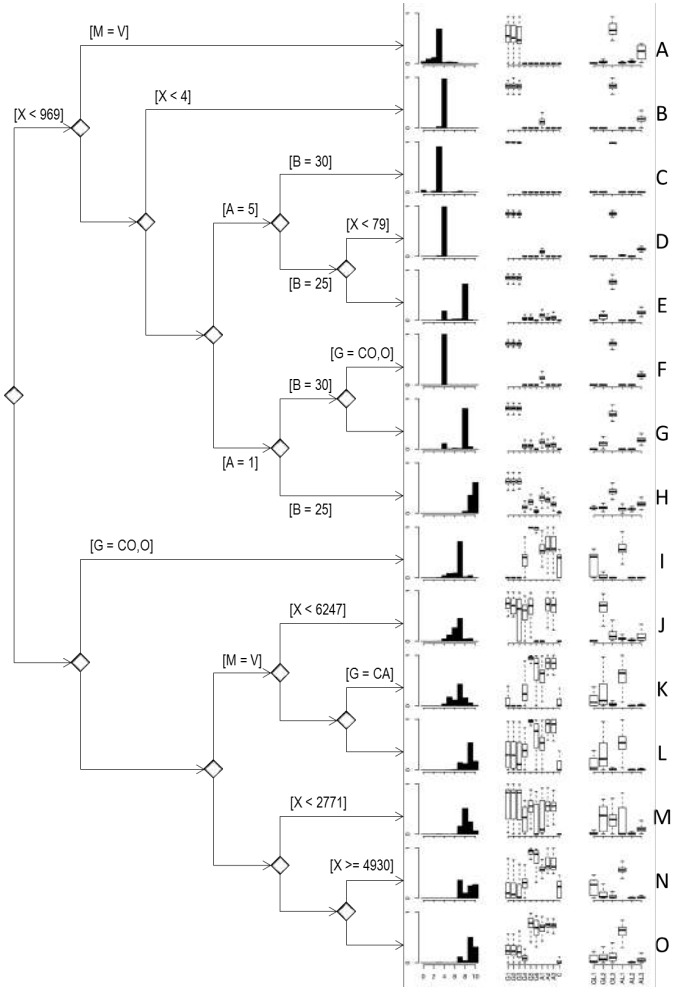
Simulation output data as research data

- Visualisation and analysis challenges

e.g. 20,000 runs from FEARLUS-SPOMM







Overview of project plan

The James Hutton Institute

- Agree simulation output metadata standard
- Create tools allowing analysis and visualisation of simulation output data
- Make these available on the web
- Use the tools to explore relationships among variables in our own models

Objectives

The James Hutton Institute

- Collect, extend and share analysis and visualisation methods
- Create web-based tools to facilitate this
- Conduct meta-analysis of our own projects and invite others to participate
- Try the techniques with empirical (non-model) datasets

Institutional context and objectives (1)

The James Hutton Institute

- University of Waterloo, Canada
    - Co-ordination
    - Development of web-based tools for analysis and visualisation
    - Running workshops
- Arizona State University, USA
    - Web-based interface development
    - Workflow management
    - Embedding in CoMSES Net

Institutional context and objectives (2)

- University of Twente, The Netherlands
    - Development of tools for analysis and visualisation of outputs from social simulations
    - Co-organisation of workshops
- University of Dundee and The James Hutton Institute, UK
    - Development of tools for discovering relationships among variables
    - Prototype agent-based data-mining algorithm

Deliverables

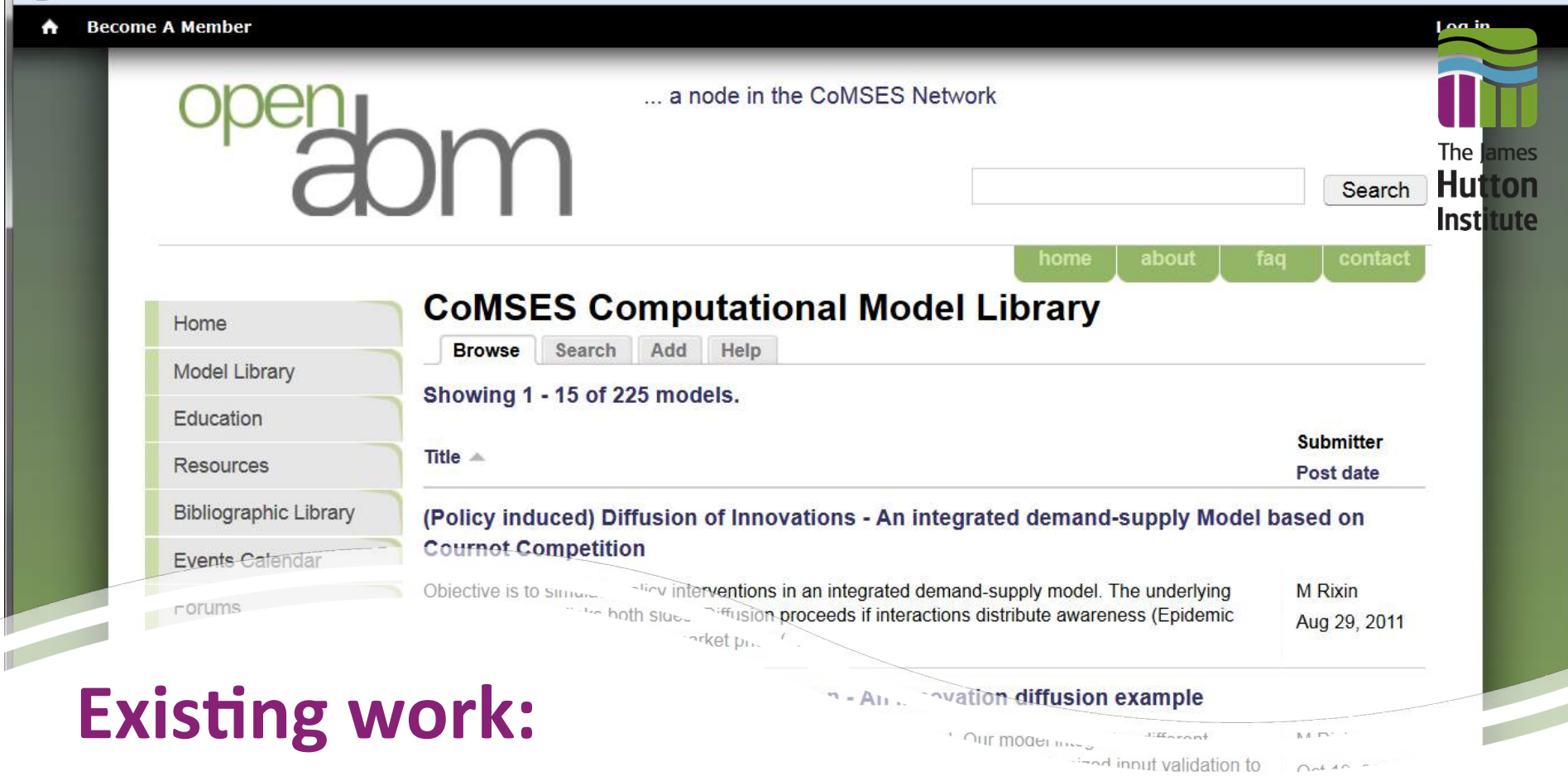
The James Hutton Institute

- Reports
    - Tools for visualising high-dimensional nonlinear data
    - Final report
- Software
    - Data analysis tool set for R
    - Metadata tools
    - Visualisation tools
    - Website

Use cases (agent-based models)

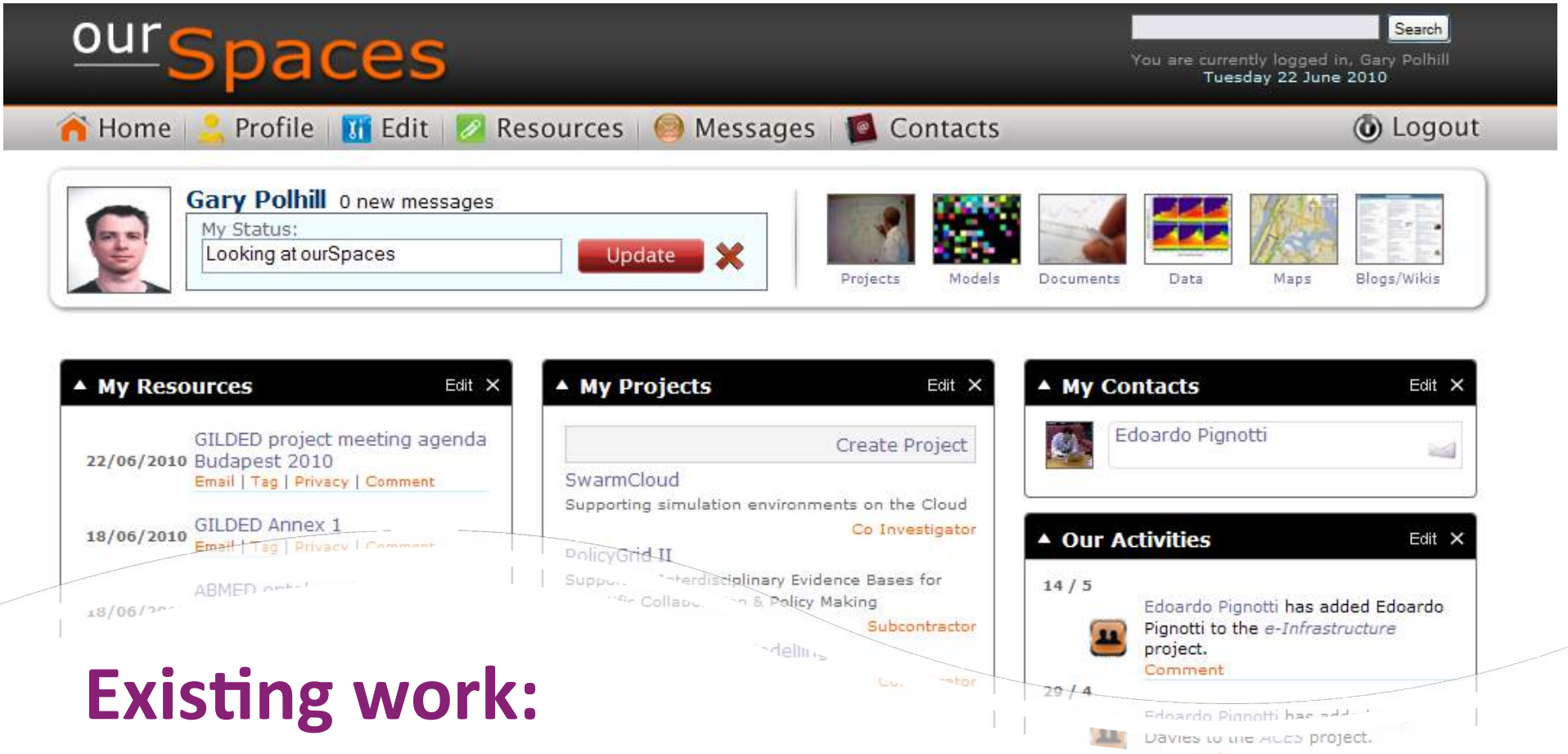
The James Hutton Institute

- Land management and biodiversity (UK)
- Energy demand (UK)
- Land use, transportation and landscape management (Canada)
- Housing market in hazard-prone areas (The Netherlands)



Existing work: CoMSES Network

- The Open ABM library allows users to upload their models
- Aim is to provide a digital archive of ABM artefacts (source code, configuration files, etc.)
- Implements DDI and Dublin Core

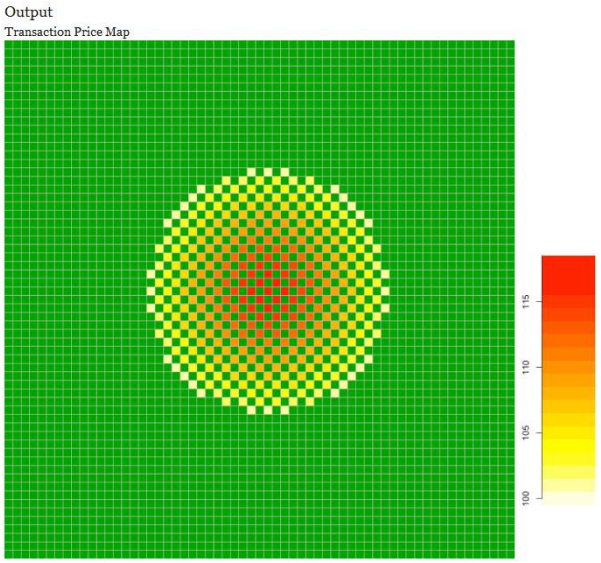


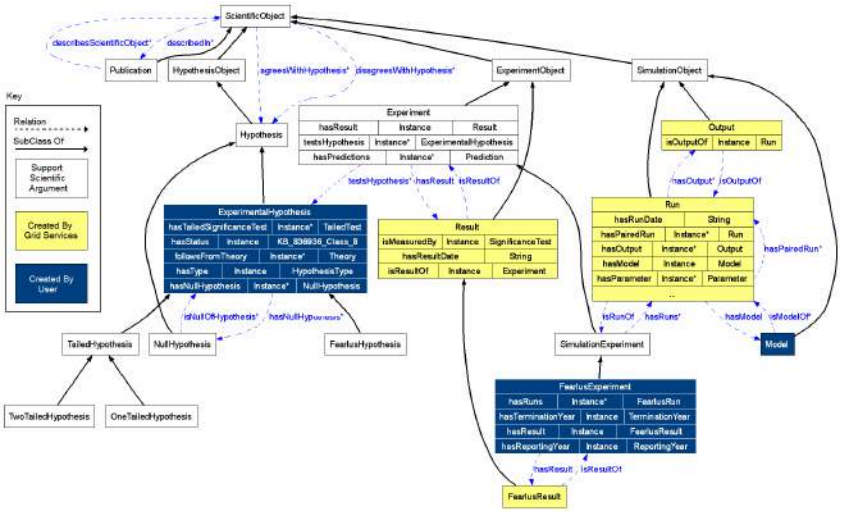
Existing work: ourSpaces VRE

- Prototype provenance-enabled Virtual Research Environment allowing upload of digital artefacts

Other existing work

- MyExperiment – scientific workflows
- FEARLUS-G – semantic grid enabled agent-based model experimentation tool
- SLUCE2 Land Market Model web-based visualisation tool





Communication

The James Hutton Institute

- Engage with relevant scholarly societies at conferences and with journal articles
    - European Social Simulation Association
    - International Environmental Modelling and Software Society
- Two international workshops are planned to engage with the research community
    - First is at iEMSs 2014 in June

"**Workshop G2**: Analyzing and Synthesizing Results from Complex Socio-ecosystem Models with High-dimensional Input, Parameter and Output Spaces"

Success criteria

The James Hutton Institute

- Agreed metadata standard for recording simulation output data
- Novel visualisation and analysis tools available to all via CoMSES website
- Interesting findings from our ABM output