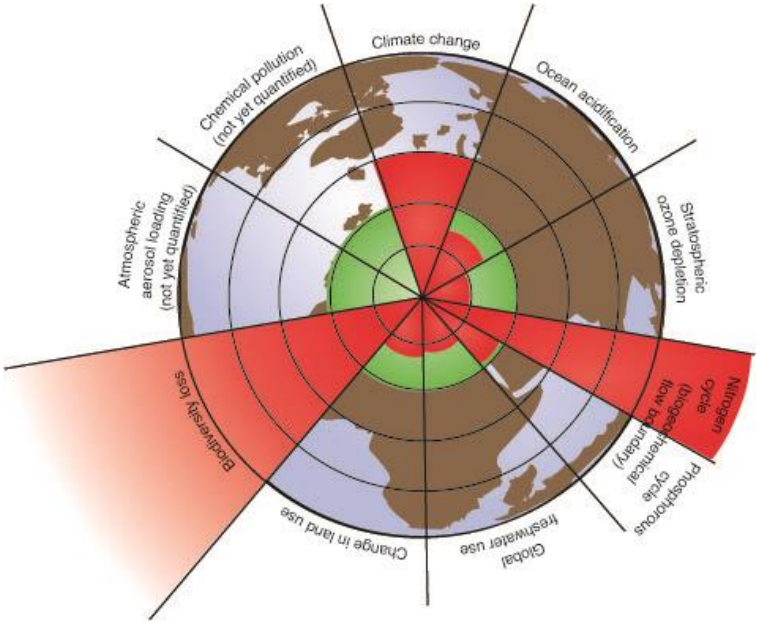
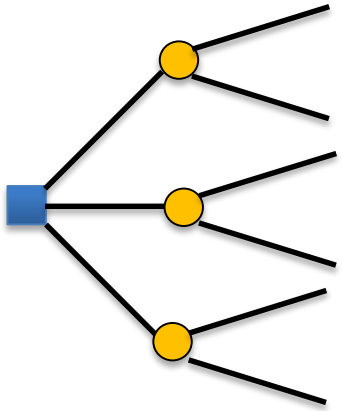
Combining threshold- and cluster-based scenario discovery methods to improve scenario interpretability and usability

Michael D. Gerst

THAYER SCHOOL OF ENGINEERING AT DARTMOUTH



The Present



Which future?

March 26, 2014
IQ-SCENE Workshop

Background

Collaborators/Funders

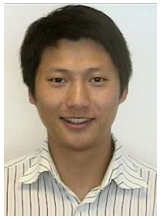
Dr. Richard Howarth
Ecological Economics
Environmental Studies
Dartmouth College

Dr. Mark Borsuk
Engineering
Dartmouth College

Dr. Giovanni Dosi
Economics
Sant'Anna School of Advanced Studies
(Italy)

Dr. Andrea Roventini
Economics
University of Verona
(Italy)

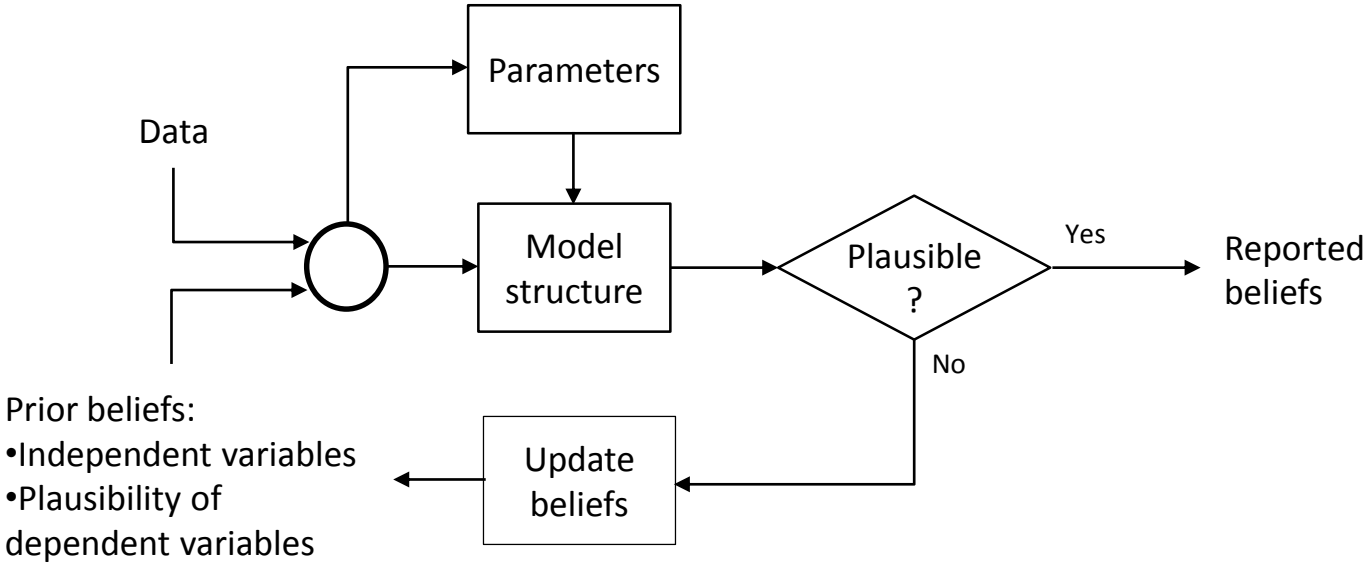
Dr. Guri Bang
Political Science
CICERO
(Norway)



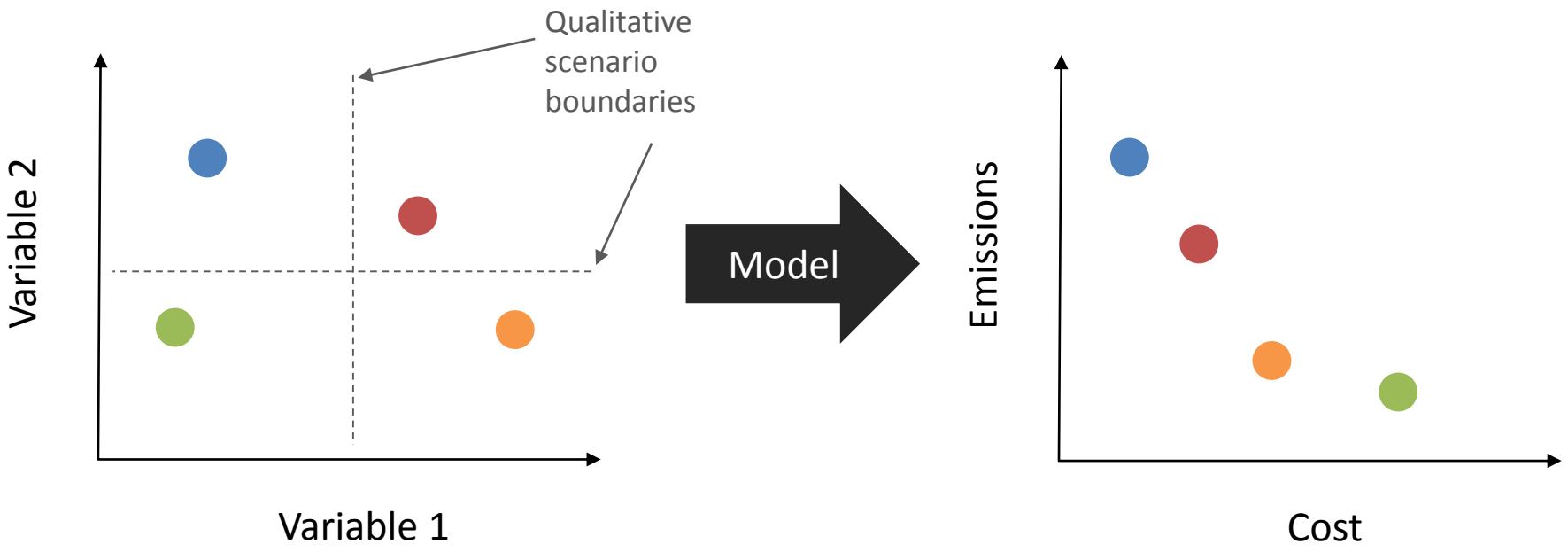
Peng Wang
(M.S., Dartmouth)
Bridgewater Assoc.

NSF DRMS Grant
SES-0962258

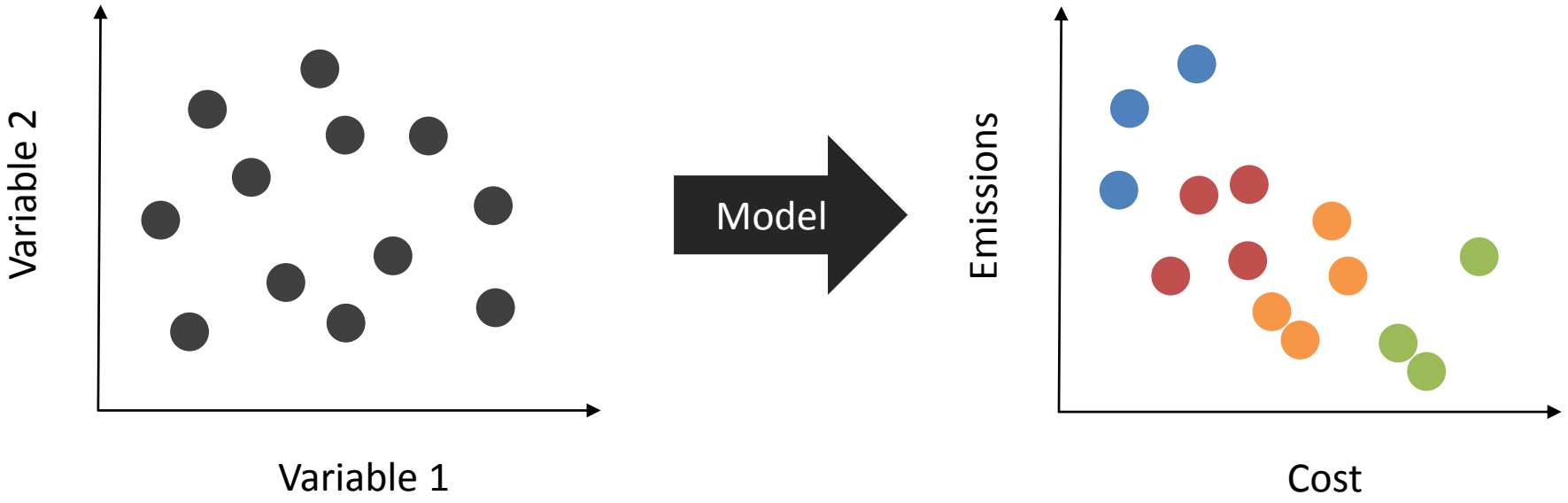
A Bayesian view of exploratory modeling



Story-and-simulation

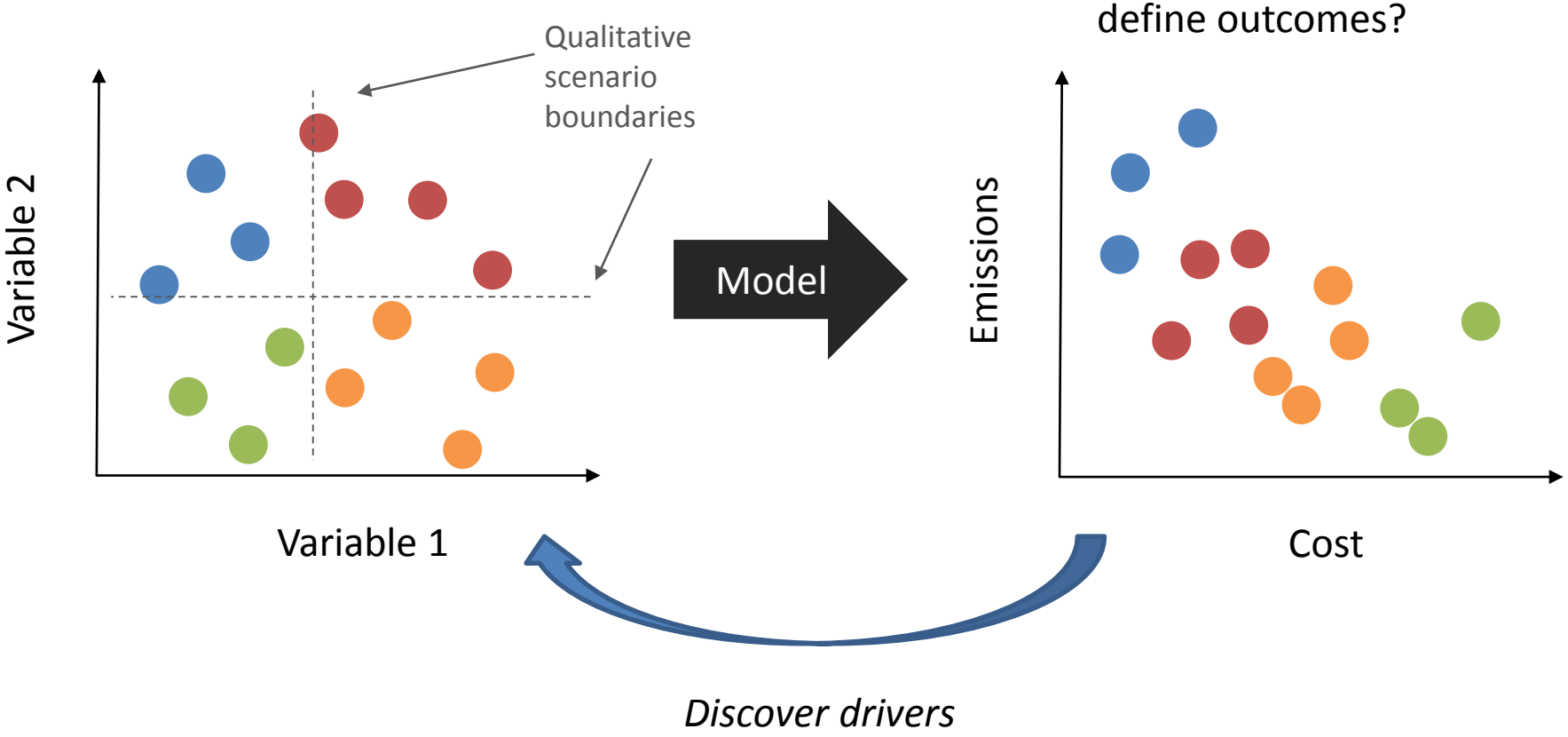


Scenario discovery

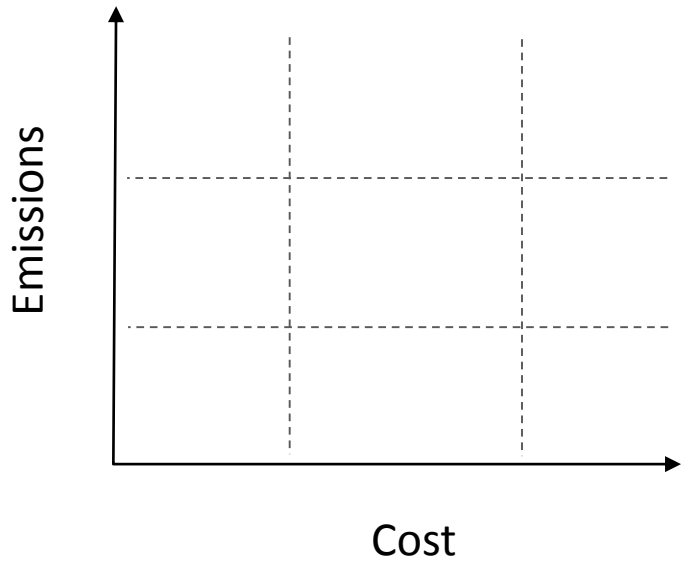


Scenario discovery

Key issue: If scenario outcomes are multi-dimensional, how to define outcomes?



Threshold-based Method



- Outcomes might not conform neatly to threshold boxes
- Thresholds may not be meaningful in a decision-making context
- Decision-makers might not be able to agree on thresholds

An Alternative Method: Clustering Outcome Data

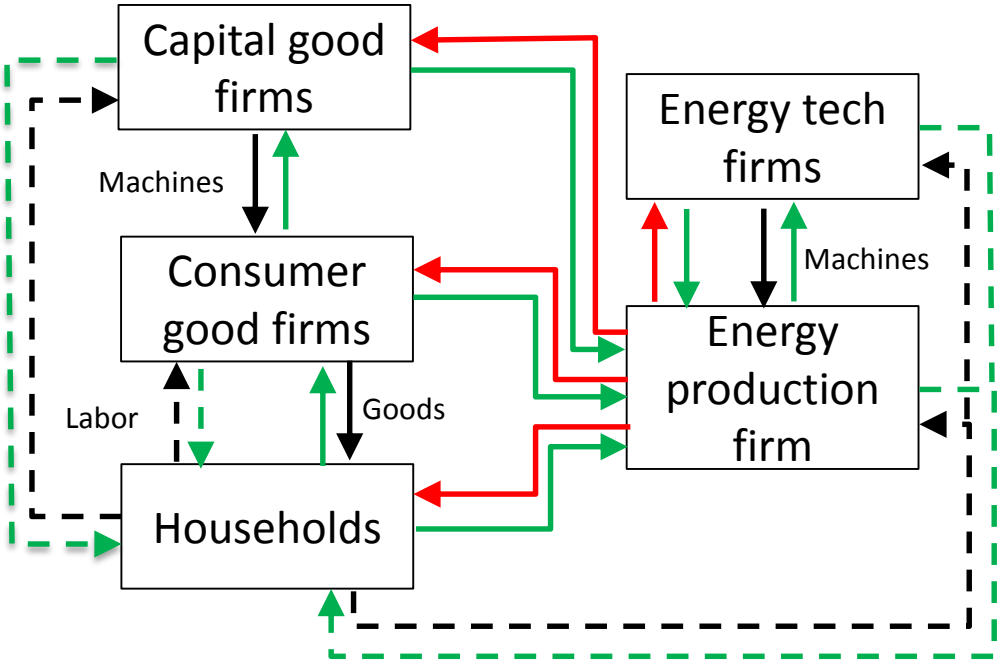
Simple example using an agent-based model

Documented in:

Gerst, MD, et al. 2013. *Environmental Modelling & Software* 44:62-75.

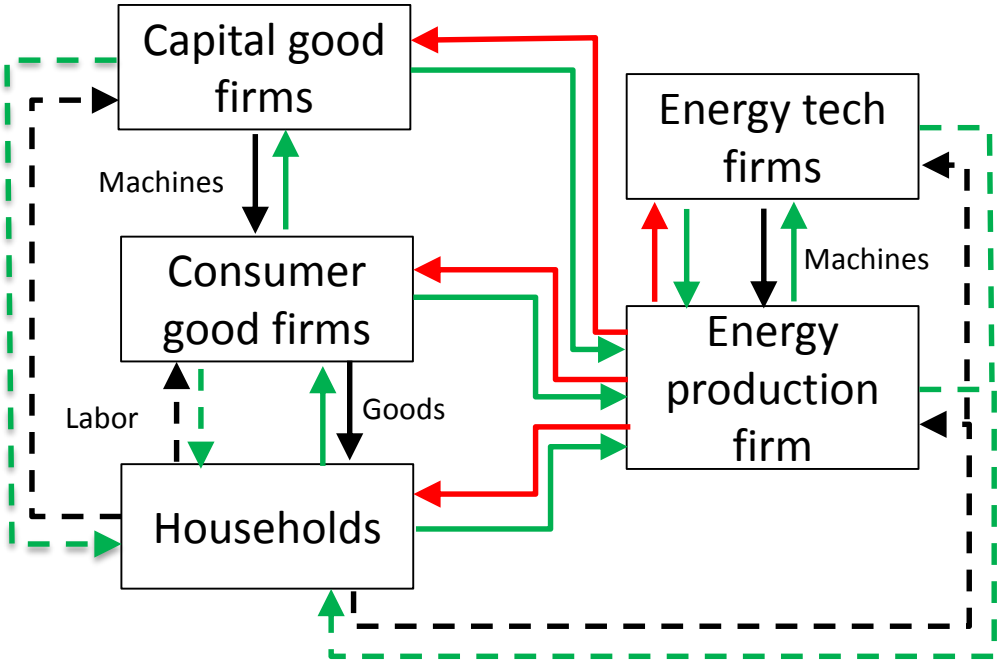
Gerst, MD, et al. 2013. *Environmental Modelling & Software* 44:76-86

1st generation domestic energy-economy model overview



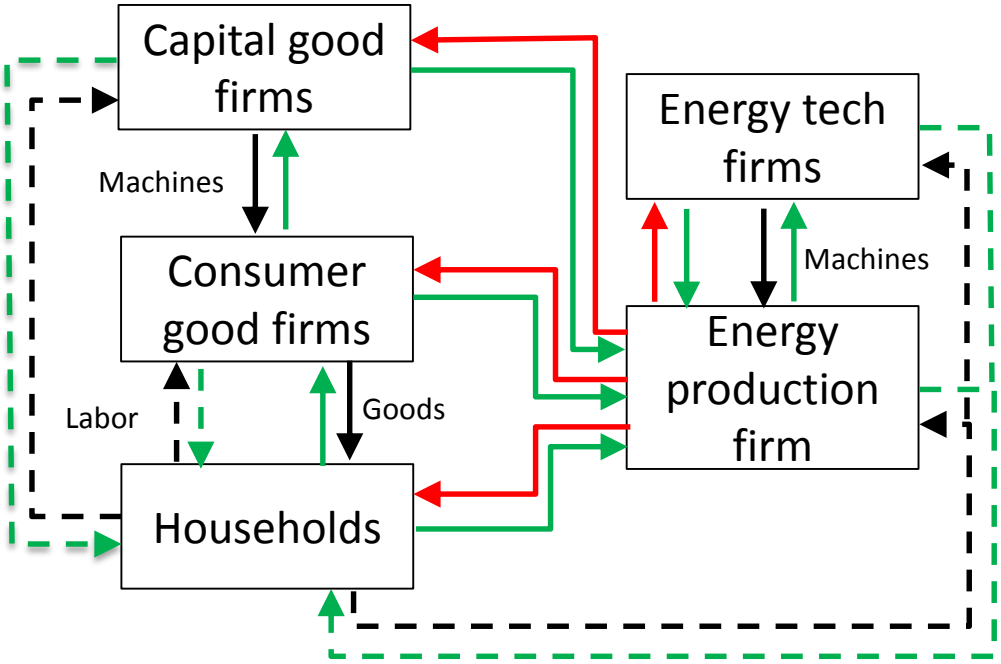
- Household heterogeneity: goods owned
- Consumer good firm heterogeneity: machine stock
- Capital good firm heterogeneity: machines produced
- R&D has random affects on labor productivity and energy efficiency

1st generation domestic energy-economy model overview



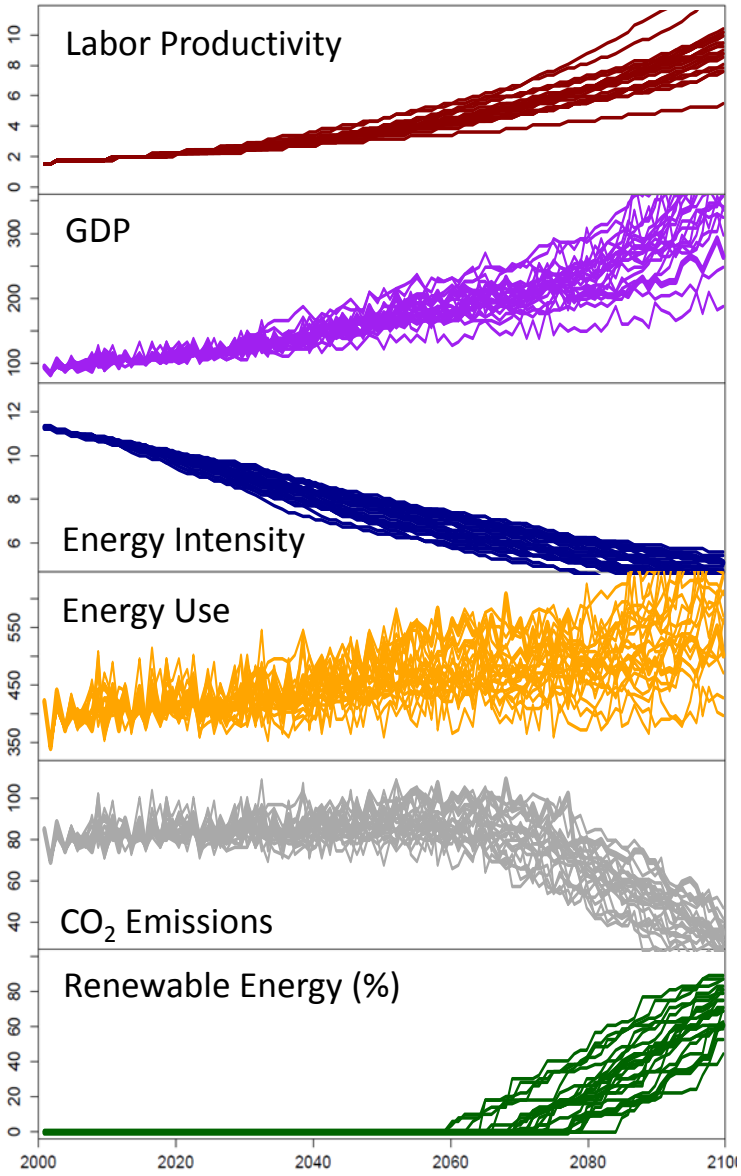
- Wage tracks average productivity changes
- Price changes pegged to changes in firm competitiveness
- Three types energy technologies: carbon-heavy, carbon-light, and carbon-free
- One type of energy produced
- Fuel production exogenous

1st generation domestic energy-economy model overview



- Effects of R&D success calibrated to historical US GDP growth and residential energy use
- Starting conditions match US economic accounts and energy use data

Simulation of carbon tax w/ recycling to carbon-free R&D



Successful R&D leads to uncertain productivity growth

Which translates to increasing GDP

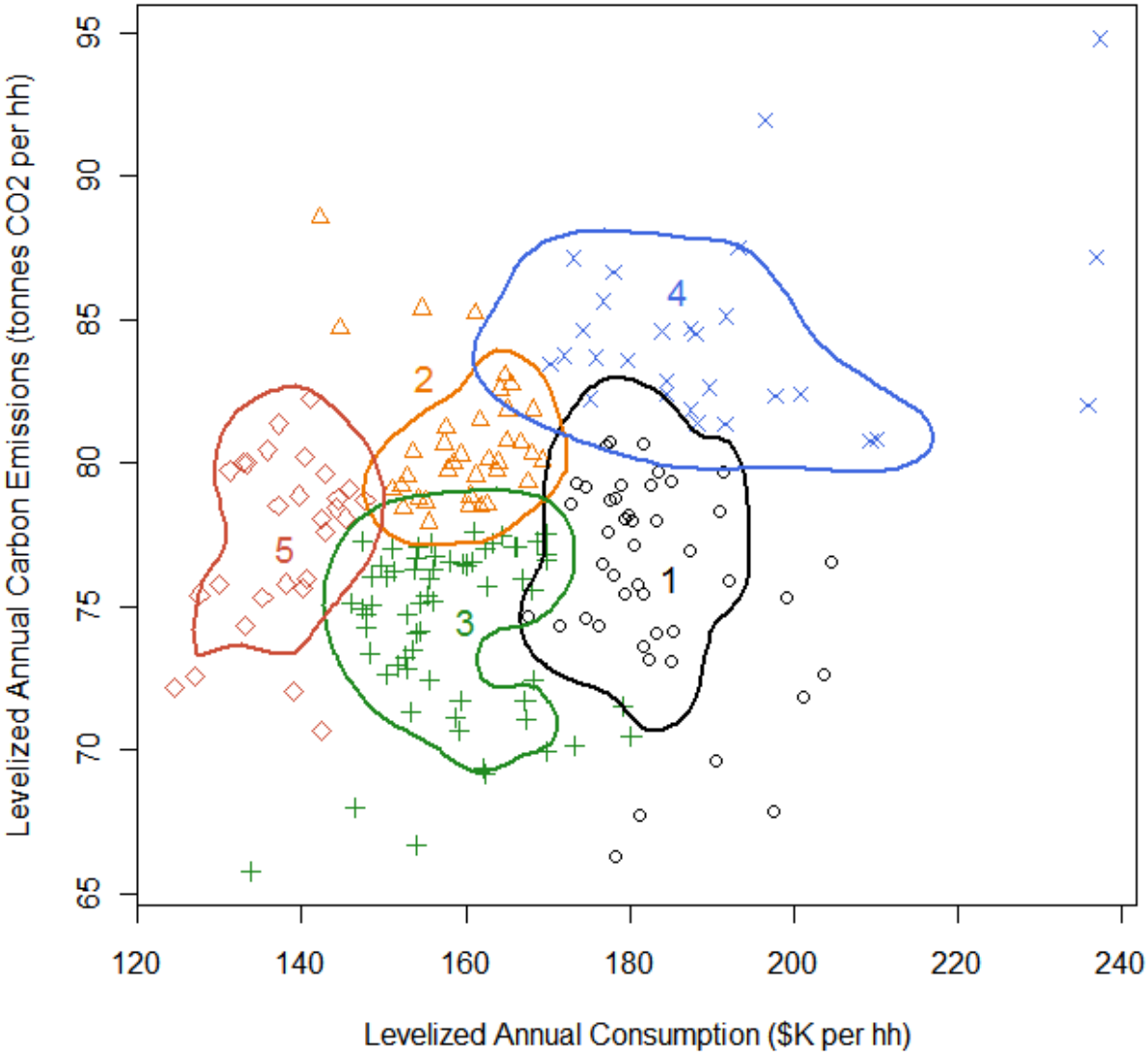
Increasing GDP combined with improving energy efficiency leads to decreasing energy intensity

Which leads to a wide range of energy use

CO_2 emissions decrease with

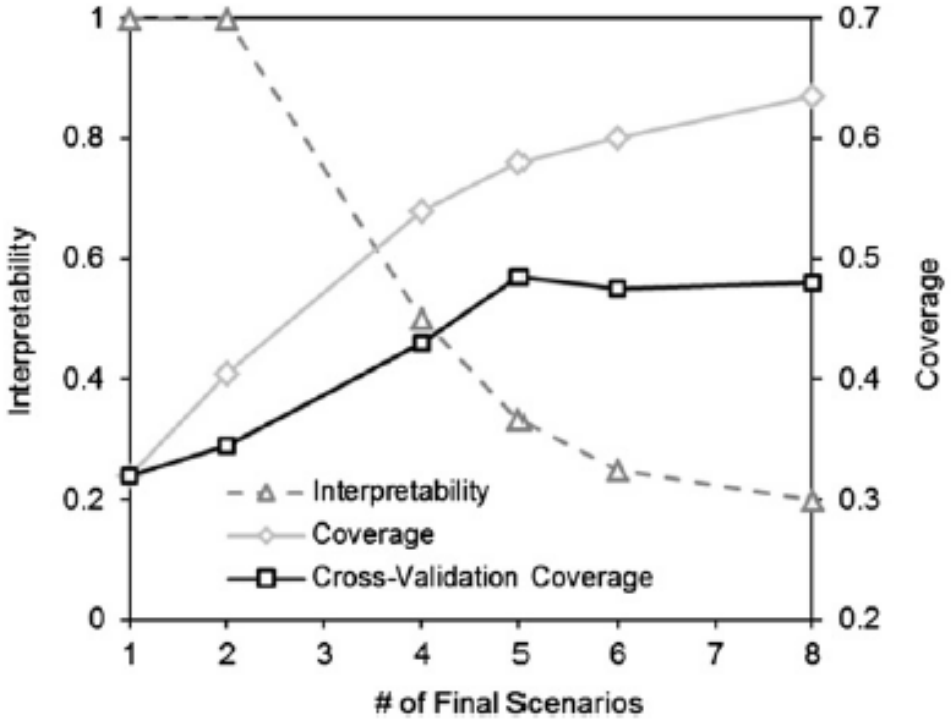
Penetration of carbon-free technology

Multi-Dimensional Scenario Discovery

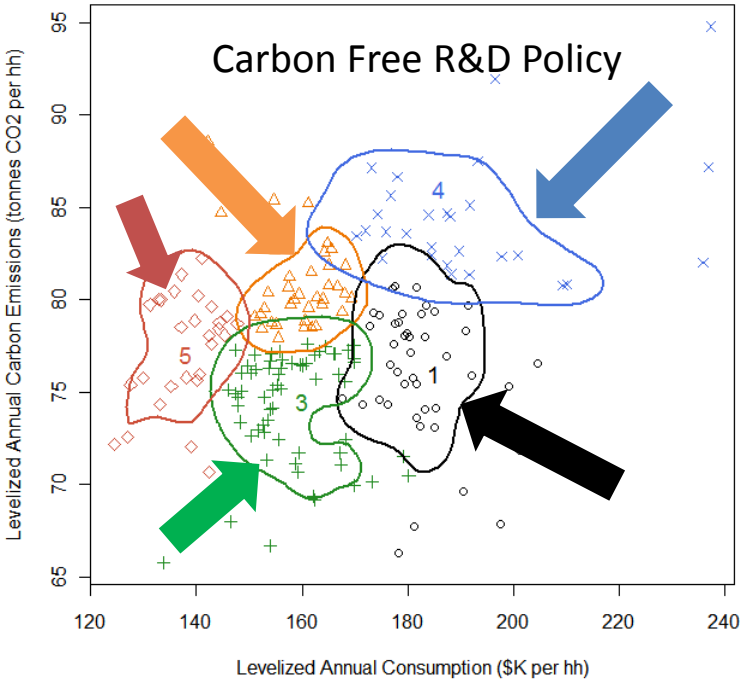


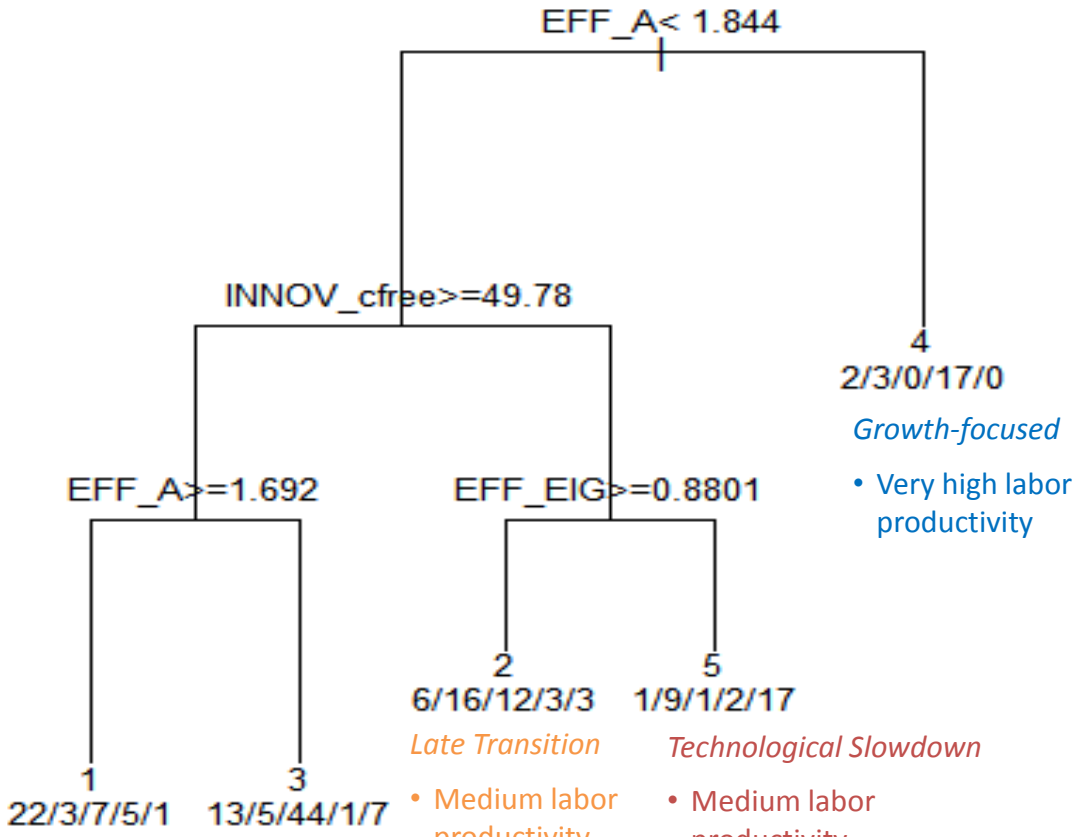
1. Choose output variables of interest (consumption and carbon emissions)
2. Use hierarchical clustering to see if natural clusters emerge from data
3. Use classification tree to see which of 9 stochastic variables are important for determining scenarios

Pruning classification tree



Classification Tree





Green High-Growth

- Medium labor productivity
- High carbon-free innovation

Green Low-Growth

- Very low labor productivity
- High carbon-free innovation

Late Transition

- Medium labor productivity
- Low carbon-free innovation
- High energy efficiency of good use

Technological Slowdown

- Medium labor productivity
- Low improvements in energy efficiency and carbon-free costs

Growth-focused

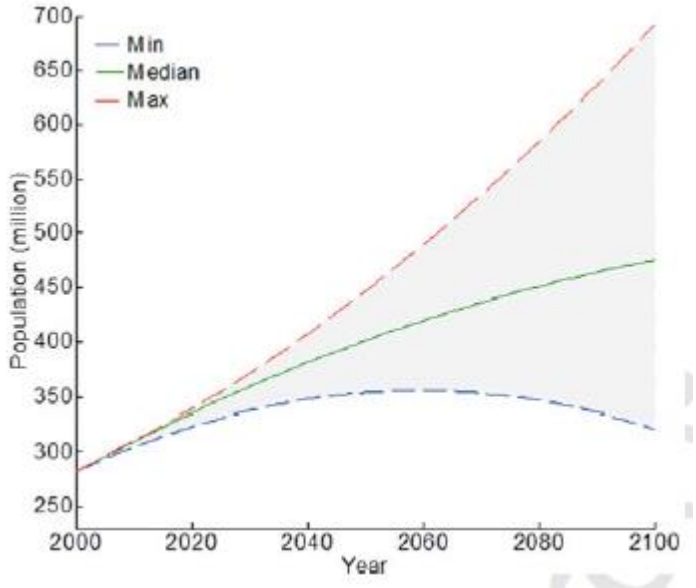
- Very high labor productivity

Combining cluster and thresholds

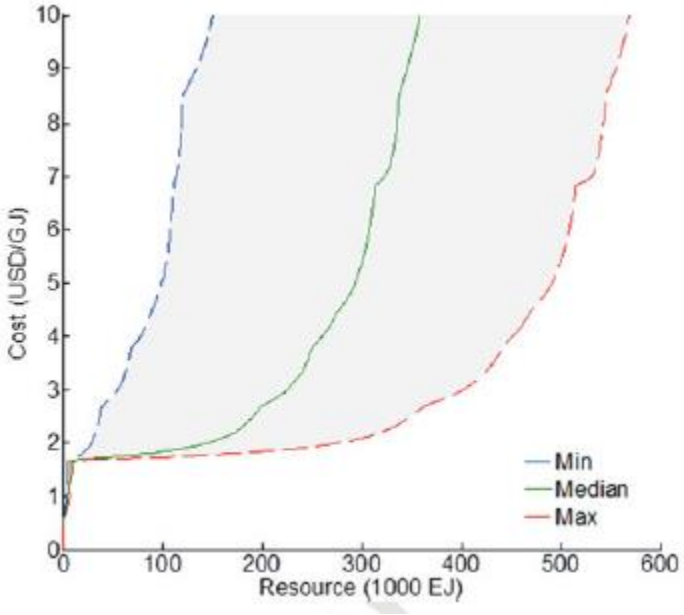
Another simple example … leading to more questions than answers

Take previous model and add in …

Probabilistic population growth for the US



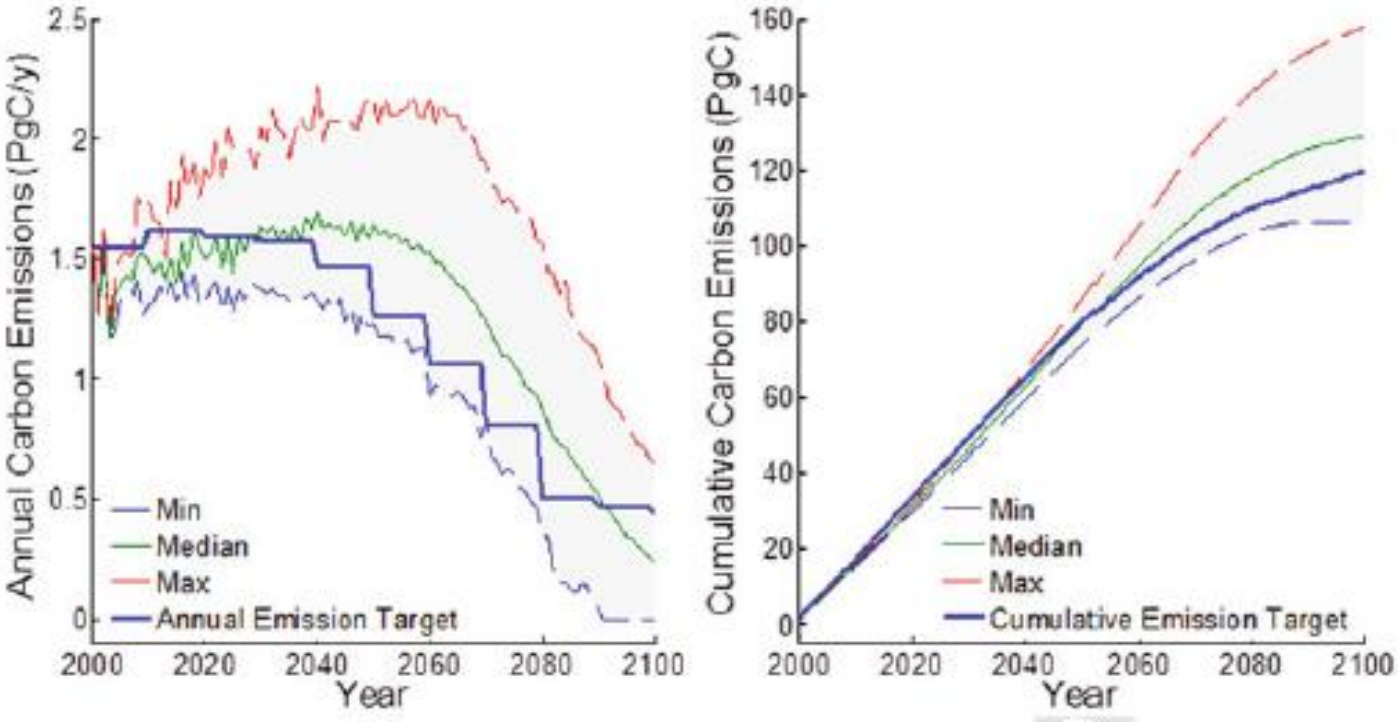
Probabilistic fossil fuel costs and resources



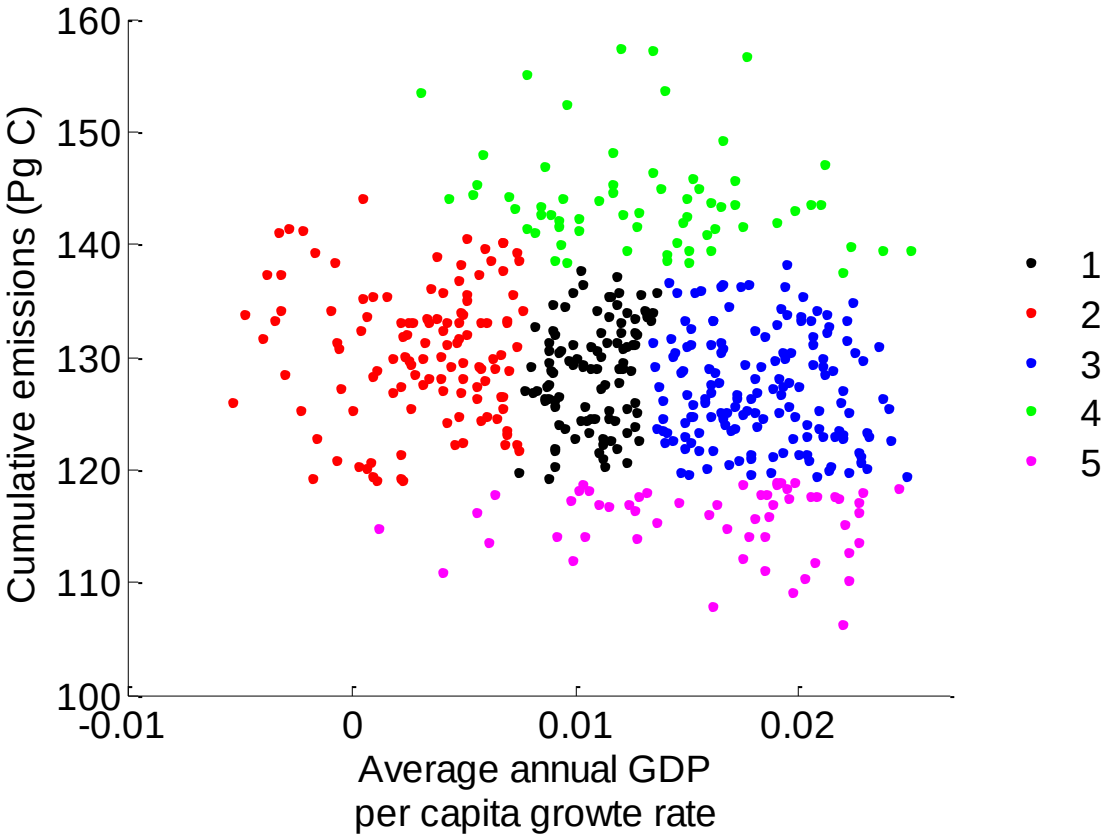
And an *adaptive* carbon tax to meet RCP4.5

Documented in: Wang, P et al. 2014. in Energy Modeling in the 21st Century. ed. H. Qudrat-Ullah. Springer.

Emission trajectories

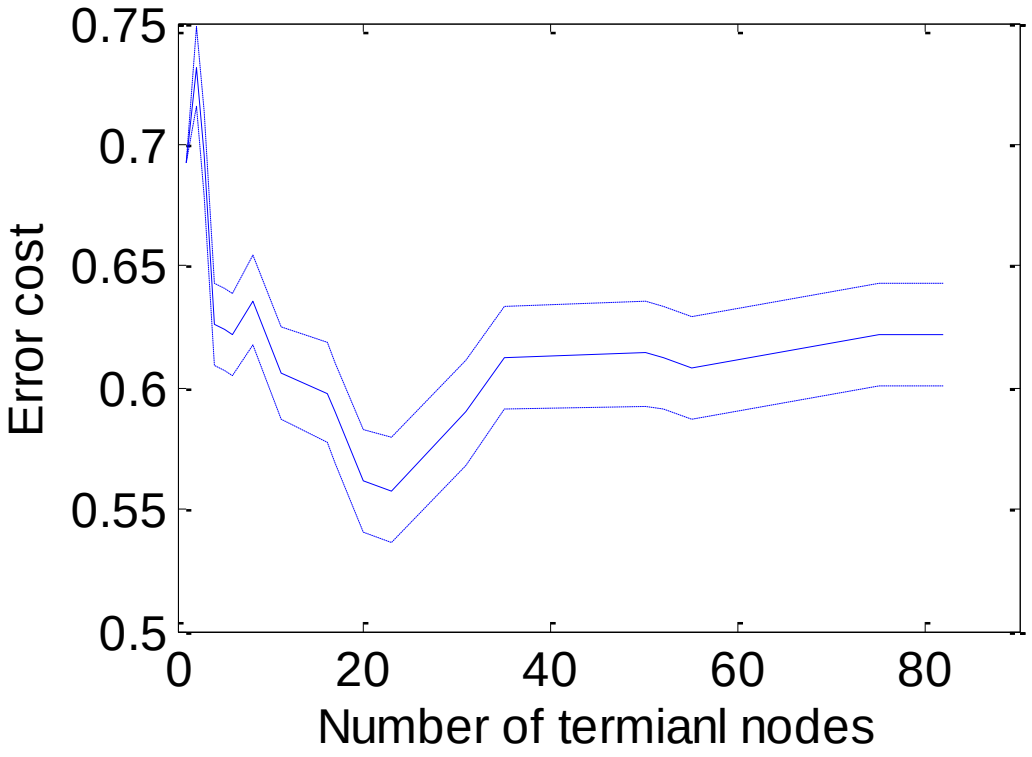


Resulting clustering



- Use meeting or not meeting target as a response variable
- Now have three response variables as inputs into cluster analysis

Classification tree error costs



- Large number of terminal nodes (~20) implies large number of final scenarios
- Also, difficulties encountered in forming well-defined scenarios

Final thoughts

1. Consider more outcome variables (e.g., energy inflation index as a proxy for cost)
2. How to consider time-dependent nature of predictors and response variables?
3. Reduction of number of predictors?