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

Michael D. Gerst[a]
Mark E. Borsuk[a]

[a]Thayer School of Engineering, Dartmouth College, Hanover, USA

In this paper we propose and demonstrate a methodology for combining two major approaches in the nascent field of data mining-based global change scenario analysis: threshold-based[1] and cluster-based[2] scenario discovery.

Providing quantitative support for climate change policy is a challenging problem because doing so involves projecting linked social and technological systems over long time spans. Such systems, which are complex and adaptive, are difficult to model with reasonable scientific accuracy because they contain both irreducible (also known as aleatoric or statistical) and reducible (also known as epistemic or knowledge) uncertainties. For example, the likelihood that research and development (R&D) programs will reduce renewable energy costs to be competitive with energy produced from fossil fuels is considerably uncertain and fundamentally unknowable. Past results of R&D can be used to provide a guide to what is possible, but ultimately the uncertainty surrounding cost reductions is irreducible. Other uncertainties, such as how households or firms make decisions, are in theory reducible, but the state of our knowledge often still requires considering multiple hypotheses about real-world behavior.

Historically, construction of scenarios has proven valuable as a means for organizing and communicating the many uncertainties associated with climate policy support. A scenario can be thought of as a “coherent, internally consistent, and plausible description of a possible future state of the world”. By illuminating the span of possible futures, consideration of diverse scenarios has the potential to highlight the interaction of complex uncertainties that would otherwise be difficult to analyze.

Climate policy scenarios have mostly been produced by a sequential, piecewise process. First, subject-matter experts are convened to create storylines that qualitatively describe plausible, internally-consistent outcomes for irreducibly uncertain processes, such as future population change, economic growth, and technological progress. These storylines are then translated into quantitative projections that are thought to be representative of the storyline themes. Finally, the exogenous projections are used as inputs to formal models that produce key outputs such as energy technology market shares, greenhouse gas emissions, and atmospheric CO_2 concentration.

However, after over a decade of utilization, the modeling community began to appreciate that these sequential methods often hindered effective use of scenarios. Because storylines were drafted separately from model construction, it was often difficult for the models to completely engage with scenario themes. Furthermore, how to interpret the scenarios in a decision-making context was often unclear, as disagreement among modelers and practitioners has surrounded the issue of assigning probabilities to scenario outcomes.

A recent effort to overcome these issues has been the Representative Concentration Pathway (RCP) framework[3]. In contrast to SRES, RCP scenarios are first defined by outcomes instead of driving forces: four radiative forcing stabilization pathways ranging from ambitious climate stabilization at 2.6 W/m^2 forcing to a more baseline scenario of 8.5 W/m^2 forcing, which correspond, respectively, to atmospheric

[1] Bryant, B.P. and R.J. Lempert. 2010. *Technological Forecasting and Social Change* 77: 34-49.
[2] Gerst et al. 2013. *Environmental Modelling & Software* 44: 76-86.
[3] Moss, R.H., et al. 2010. *Nature* 463: 747-756.

greenhouse gas concentrations of about 430 and 1230 ppm CO_2-eq. in year 2100. Then, pathways are used in one of two ways: (i) as forcing inputs into complex climate system models or as (ii) targets for climate policy models.

Beginning scenario planning with policy targets defined by physical variables introduces new challenges and opportunities. On the positive side, modeling teams have more freedom to define social, economic, and technological scenario attributes. However, this new flexibility adds an additional layer of uncertainty to comparison of model results because storyline and model assumptions are now likely to be different. As a result, the scientific community has begun the task of defining a set of Shared Socioeconomic Pathways (SSPs) to serve as baselines for comparison[4]. A first step in that direction has been to compare existing scenarios, looking for consistent patterns of socio-economic drivers across differing emissions trajectories. Using scenarios from previous model comparison exercises, Van Vuuren et al.[5] found that much overlap existed in the range of socio-economic drivers for any given emission trajectory. This indicates that RCPs, or emissions trajectories, alone may not sufficiently identify individual socioeconomic scenarios. Resultantly, van Vurren et al. have proposed a matrix framework whereby RCP forcing targets define four matrix rows, and SSP drivers, such as mitigative and adaptive capacity, define matrix columns. How to fill in the matrix elements remains an open question. Among the many issues are how to ensure consistency among rows and columns and how to address co-variance among SSP drivers.

In a first attempt at addressing these questions, Rozenberg et al.[6] use 286 simulations of the IMACLIM-R model[7] and Bryant and Lempert's scenario discovery method[8] to generate self-consistent scenarios to populate the matrix. Scenario discovery operates in the opposite direction of the sequential approach. Probabilistic simulations from a quantitative model are generated first. Then, using non-parametric statistical methods, model outputs are grouped according to chosen metrics, and determinant driving forces for each group are identified. As we discuss in Gerst et al.[9], Bryant and Lempert's method, while clearly a step forward, requires selecting *a priori* performance thresholds in order to group model outputs. This introduces the possibility that interesting dynamics might be overlooked, as it is difficult to determine whether selected thresholds appropriately delineate multi-dimensional model output.

Our previous work demonstrated a more generalized version of scenario discovery in the absence of a policy target[10] and in the presence of a carbon tax to meet RCP4.5[11]. Both applications allow for multiple performance dimensions without the need for *a priori* threshold selection by first clustering model simulations based on cumulative emissions and cost. These identified candidate scenarios are then further refined using classification and regression tree analysis to identify common scenario drivers. While we see this as an improvement to scenario discovery methodology, Wang, et al.'s[12] exclusion of potentially relevant information implicit in *a priori* thresholds might hinder the usability of the defined scenarios. In this paper, we use Wang, et al.'s data to explore how inclusion of relevant thresholds might improve multi-dimensional scenario discovery. Specifically, we hypothesize that including the achievement of a mitigation target along with cumulative emissions and cost in the clustering step will enhance scenario definitions. In a sense, this proposed method merges aspects of Bryant and Lempert's and Gerst et al.'s approaches toward a more generalized template for scenario discovery methodology that will be widely applicable to forecasted and backcasted scenario exercises.

---

[4] Kriegler, E., et al. 2012. *Global Environmental Change* 22: 807-822.
[5] van Vuuren, D.P., et al. 2012. *Global Environmental Change* 22: 21-35.
[6] Rozenberg, J., et al. 2013. *Climatic Change* 10.1007/s10584-013-0904-3.
[7] Rozenberg, J. et al. 2010. *Climatic Change* 101: 663-668.
[8] Bryant, B.P. and R.J. Lempert. 2010. *Technological Forecasting and Social Change* 77: 34-49.
[9] Gerst et al. 2013. *Environmental Modelling & Software* 44: 76-86.
[10] Ibid.
[11] Wang, P., et al. 2013. In *Energy Policy Modeling in the 21st Century*. Ed H. Qudrat-Ullah. Springer: 251-269.
[12] Ibid.