

Exploring Energy and Economic Futures Using Agent-Based Modeling and Scenario Discovery

P. Wang, M. D. Gerst and M. E. Borsuk

Introduction

Providing quantitative support for climate change policy is a challenging problem because doing so involves representing 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 of 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 of 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’ (McCarthy et al. 2001). 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 (Groves and Lempert 2007).

Climate policy scenarios have mostly been produced by a sequential, piecewise process. Subject-matter experts are convened to create storylines that qualitatively describe plausible, internally consistent outcomes for irreducibly uncertain processes,

M. E. Borsuk (✉) · P. Wang · M. D. Gerst
Dartmouth College, Thayer School of Engineering, 14 Engineering Drive,
Hanover, NH 03755, USA
e-mail: mark.borsuk@dartmouth.edu

P. Wang
e-mail: peng.wang@dartmouth.edu

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 CO₂ concentration.

The most well-known application of the sequential scenario process to climate policy has been the Special Report on Emissions Scenarios (SRES; Nakicenovic and Swart 2000). It adopted the scenario axis method adopted by Schwartz (1991), which uses quadrants of a two-dimensional space to define four scenarios. In SRES, the axes are defined by degree of globalization and degree of sustainable development. Following the sequential process, the quadrants were used to sketch four storylines and quantify four sets of projected exogenous variables, which were used as model inputs for many climate policy studies. However, after more than a decade of utilization, the modeling community began to indicate that the scenario axis and sequential methods often hindered effective use of scenarios (Moss et al. 2010; Parson et al. 2007). 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 surrounded the issue of assigning probabilities to scenario outcomes.

A recent effort to overcome these issues has been the Representative Concentration Pathway framework (RCP; Moss et al. 2010). 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/m² forcing to a more baseline scenario of 8.5 W/m² forcing, which correspond, respectively, to atmospheric greenhouse gas concentrations of about 430 and 1,230 ppm CO₂-eq. in the year 2100. Then, pathways are used in one of the two ways: (i) as forcing inputs into complex climate system models or (ii) as 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 the 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 a baseline for comparison (Kriegler et al. 2012). The first step in that direction has been to compare existing scenarios, looking for consistent patterns of socioeconomic drivers across differing emissions scenarios. Using scenarios from EMF-22 (energy modeling forum; Clarke et al. 2009), AR4 (fourth assessment report; Fisher et al. 2007; Nakicenovic et al. 2006), and the RCPs (Moss et al. 2010), van Vuuren et al. (2012) found that much overlap existed in the range of socioeconomic drivers for any given emission trajectory. This indicates that RCPs, or emissions trajectories, alone may not sufficiently identify individual socioeconomic scenarios. Resultantly, van Vuuren et al. (2012) have proposed a matrix framework whereby RCP forcing targets define four matrix rows, and SSP drivers, such as mitigative and adaptive capacity, define

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 an initial attempt at addressing these questions, Rozenberg et al. (Rozenberg et al. 2012) use 286 simulations of the IMACLIM-R model (Rozenberg et al. 2010) and Bryant and Lempert's (2010) scenario discovery method to generate self-consistent scenarios to populate the matrix. Scenario discovery operates in the opposite direction of sequential approach. First, probabilistic simulations from a quantitative model are generated. Then, using nonparametric statistical methods, model outputs are grouped according to chosen metrics and determinant driving forces for each group are identified. As discussed in Gerst et al. (2013a), 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 multidimensional model output.

Our previous work (Gerst et al. 2013a) demonstrated a more generalized version of scenario discovery that allows for multiple performance dimensions without the need for a priori threshold selection. In the current contribution, we further demonstrate the utility of this approach by using an enhanced version of the agent-based ENGAGE model (Gerst et al. 2013b) to identify socioeconomic pathways for the 4.5 W/m² RCP. While ENGAGE remains a relatively simple model, we believe the results demonstrate how the combination of agent-based modeling and scenario discovery might be used to 'fill in' the matrix framework relating to RCPs and SSPs.

Method

ENGAGE is an agent-based, energy–economy model that is patterned after the family of evolutionary economic models recently developed by Dosi et al. (2010), (Fig. 1). This model consists of four types of agents—households, consumption goods firms, capital goods firms, and government—and one resource, labor. It is particularly well suited as a starting point for investigating the technological and economic aspects of climate policy because technological change is modeled as the driving force of economic growth and is represented as being both stochastic and endogenous. In our previous work (Gerst et al. 2013b), we expanded the original model to include energy as a resource, which involved adding firms that produce energy technologies and a single form of energy used by households and firms. In this section, we provide a brief description of the model and detail a new functionality added for the current study, including: (i) probabilistic population growth, (ii) probabilistic fuel costs, and (iii) endogenous climate policy. We encourage readers to refer to Gerst et al. (2013b) for details on the structure, parameterization, and motivation for ENGAGE.

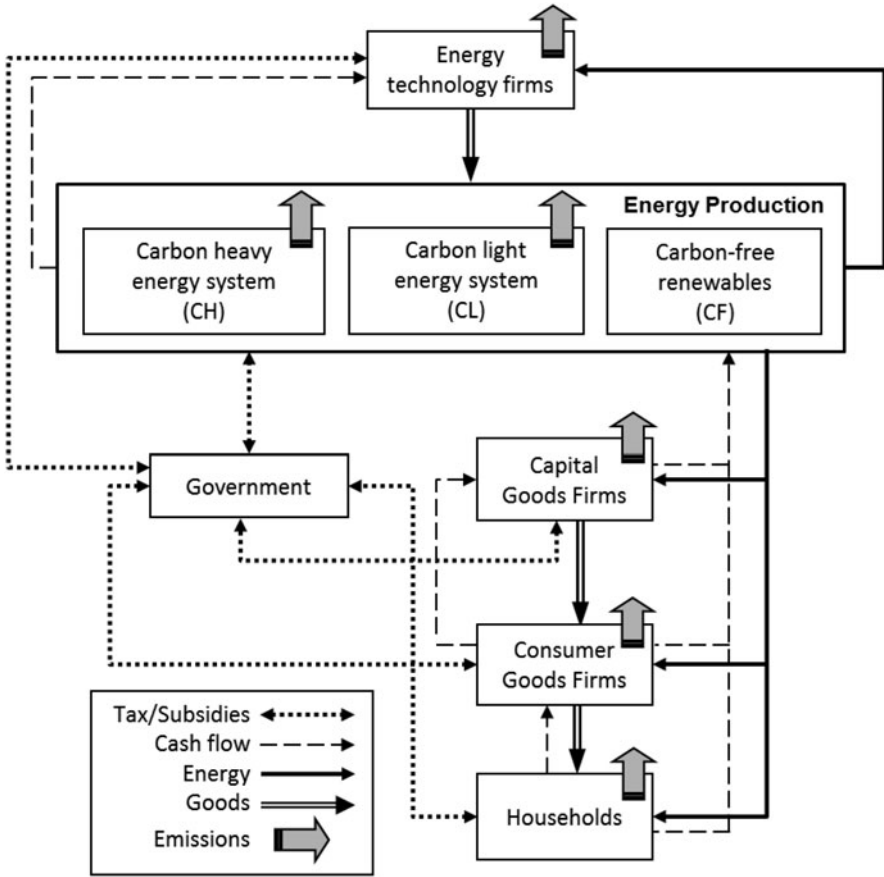


Fig. 1 Model schematic with boxes showing the various classes of agents and arrows indicating their interactions

Households

In our simplified economy, households supply labor to firms and spend all earned income on purchasing new generic consumption goods, which we call ‘thneeds,’ and energy to use existing thneeds. We do not explicitly model the labor market: wages (w) earned by households track closely with economy-wide changes in labor productivity and unemployed households receive an income subsidy provided by the government.

In the model version described by Gerst et al. (2013b), which represented the US economy, it was assumed that the number of households remained constant over time—a major simplifying assumption. In the current study, we relax this assumption

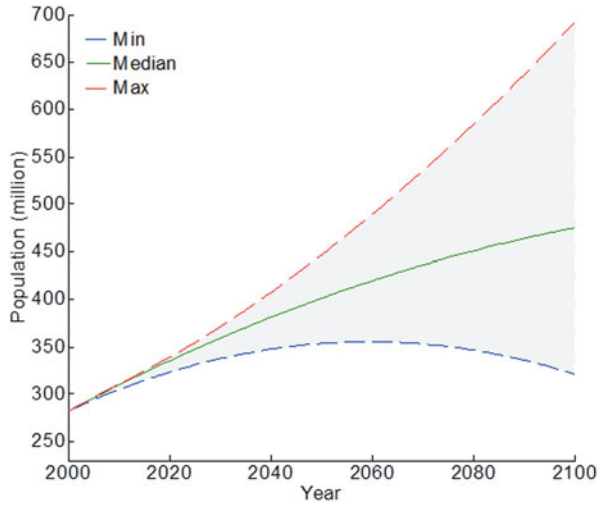


Fig. 2 Summary of probabilistic population projections used in the model. Lines indicate minimum (blue), median (green), and maximum (red) trajectories of 500 simulations

through a representation of the US population change that is fit to the probabilistic projections of Raftery et al. (2012). Specifically, population (P) is represented by a quadratic function

$$P[t] = a_1(t - 2000)^2 + a_2(t - 2000) + P_{2000}, \tag{1}$$

where population at $t = 2000$ (P_{2000}) is 282.5 million and the coefficients a_1 and a_2 are linked to the uncertain population in 2100 (P_{2100}) by

$$a_1 = 1.10 \cdot 10^{-7} (P_{2100}^2) - 1.21 \cdot 10^{-5} (P_{2100}) - 2.82 \cdot 10^{-2}, \tag{2}$$

$$a_2 = -1.10 \cdot 10^{-5} (P_{2100}^2) + 1.12 \cdot 10^{-2} (P_{2100}) - 3.88 \cdot 10^{-3}. \tag{3}$$

We represent P_{2100} by a log-normally distributed variable with arithmetic mean = 481.2 million and s.d. = 56.8 million (Fig. 2).

Capital Goods Sector

In our model, innovation activity is centered in the capital goods sector. Capital goods firms hire labor and use energy to produce machines, which are purchased by consumer goods firms for the purpose of producing consumer goods. Machines have five properties related to labor and energy use: (i) thneed production labor productivity (thneeds produced per worker), (ii) thneed production energy intensity (MWh per good), (iii) machine production labor productivity (machines produced

per worker), (iv) machine production energy intensity (MWh per machine), and (v) thneed use energy intensity (kWh per good).

Capital goods firms reinvest a fraction of their past sales in innovation and imitation activities, which have uncertain outcomes with regard to labor and energy intensity improvements. If a firm successfully innovates or imitates, it then compares the new machine to its currently produced machine and chooses the one having the lowest lifecycle cost. Lifecycle cost is composed of the sum of three terms: (i) the machine price and production capacity annualized by the annual interest on debt and expected machine lifetime; (ii) the cost of using the machine to produce goods; and (iii) the discounted cost of using the good. Machines are priced according a mark-up over operating costs that is homogenous across firms.

Importantly, the market for machines is defined by imperfect information. We model this by limiting the number of consumer goods firms to which capital goods firms may advertise. If a capital goods firm cannot find customers, then we assume it is subsequently replaced by a new firm.

Consumer Goods Sector

Consumer goods firms use their stock of purchased machines to produce thneeds. Each firm plans its desired level of thneed production according to expected demand, desired inventory, and the actual inventory. To meet increasing demand or to replace end-of-life machines, firms use lifecycle cost to compare the desirability of advertised machines. Firms may also replace machines before their end-of-life, but must consider the sunk cost of replacing a machine with a remaining useful life.

Like capital goods firms, consumer goods firms set prices using a mark-up over operating costs. However, the mark-up varies from firm to firm and is dependent on the firm's market share. Market shares evolve as a function of firm competitiveness relative to average sector competitiveness weighted by market share, where individual firm competitiveness is a function of price and cost of use.

Machine purchases and thneed production may be funded internally or through borrowing. Firms, however, have a limit to their debt to sales ratio. A consumer goods firm with a near-zero market share and negative liquid assets or an unfilled demand ceases operations and is replaced by a new firm.

Energy Sector

In our model, energy is represented by a generic form and is produced by a single-energy production firm. The firm meets overall energy demand by maintaining a stock of three 'stylized' energy technologies: carbon-heavy, carbon-light, and carbon-free.

New additions are made to the stock to replace end-of-life technologies or to meet increasing demand. The choice of which technology to purchase is made by a leveled, cost-decision rule:

$$\begin{aligned}
 c_{E,disc}[k, t] = & \frac{p_T[k, t]}{8760 \cdot u_T[k]} + \frac{w[t]}{AET[k, t] \cdot 10^6} \cdot \sum_{t'=1}^{\eta_T[k]} (1 + r_E)^{-t'} \\
 & + \frac{EFP[k, t]}{10^6} \cdot \sum_{t'=1}^{\eta_T[k]} c_F^*[t + t'] \cdot (1 + r_E)^{-t'} \\
 & + \frac{\sigma[k]}{10^3} \cdot \sum_{t'=1}^{\eta_T[k]} tax_C[t] \cdot (1 + r_E)^{-t'}.
 \end{aligned} \tag{4}$$

The leveled cost comprises the sum of four terms: (i) the price of the energy technology (p_T) accounting for the capacity factor (u_T); (ii) discounted labor costs calculated from the prevailing annual wage (w , \$ per worker), labor productivity of energy production (AET , GWh per worker); (iii) discounted fuel costs calculated from the heat rate (EFP , BTU per kWh) and forecasted fuel cost (c_F^* ; \$ per 10^6 BTU), and (iv) discounted carbon emissions costs calculated from an emission factor (σ , tonnes CO_2 per MWh) and carbon tax rate (tax_C , \$ per tonne CO_2). All discounting calculations are based on an annual discount rate (r_E), Energy technologies are manufactured by three separate firms. We assume that carbon-heavy is a mature technology, and thus its costs remain constant. Carbon-light and carbon-free technologies undergo uncertain learning-by-searching and learning-by-doing, which act to reduce technology capital costs. Learning-by-searching is a function of cumulative research and development effort and learning-by-doing is dependent on cumulative built capacity.

We repeat the simplifying assumption in Gerst et al. (2013b) that the carbon-heavy and carbon-light technologies use the same global fossil-fuel resource stock with a cost-supply curve based on the aggregation of coal, oil, and natural gas resources. Here, however, we adopt probabilistic cost-supply curves based on the method of Mercure and Salas (2012). In this setup, the supply of a particular energy resource available at a given cost is represented by the cumulative distribution function

$$N(c) = A \cdot e^{\left(-\frac{B}{c-C_0}\right)}, \tag{5}$$

where A represents the total energy supply potential for that resource, B represents the scaling of costs (e.g., due to inflation), and C_0 represents fuel extraction cost changes (e.g., due to learning-by-doing).

Parameters B and C_0 can be calculated using values for A and any two points on the cost-supply curve (C_1, Q_1) and (C_2, Q_2) from the following expressions:

$$C_0 = \frac{C_2 \ln \frac{Q_2}{A} - C_1 \ln \frac{Q_1}{A}}{\ln \frac{Q_2}{A} - \ln \frac{Q_1}{A}}, \tag{6}$$

Table 1 Fuel cost–supply curve parameters

Resource	Total technical potential, A (1000 EJ)	Cost at 1 % of technical potential, $C_{0.01}$ (USD/GJ)	Cost at 95 % of technical potential, $C_{0.95}$ (USD/GJ)
Crude oil	Tri(7, 11, 11)	1.7	Tri(5.5, 6.8, 8.2)
Oil shale	Tri(0, 27, 56)	6.8	Tri(6.8, 8.5, 10.2)
Oil sands	Tri(1, 29, 31)	8.5	Tri(13.7, 17.1, 20.5)
Conventional gas	Tri(7, 12, 16)	0.5	Tri(4.56, 5.7, 6.8)
Shale gas	Tri(0, 29, 47)	3.8	Tri(6.9, 8.6, 10.3)
Tight gas	Tri(0, 6, 12)	2.6	Tri(6.1, 7.6, 9.1)
Methane gas	Tri(0, 32, 32)	4.4	Tri(6.9, 8.6, 10.3)
Hard coal	Tri(24, 220, 419)	1.7	Tri(2.7, 3.3, 4)
Soft coal	Tri(5, 37, 75)	2.7	Tri(5.3, 6.7, 8)

$$B = -(C_1 - C_0) \ln \frac{Q_1}{A}. \quad (7)$$

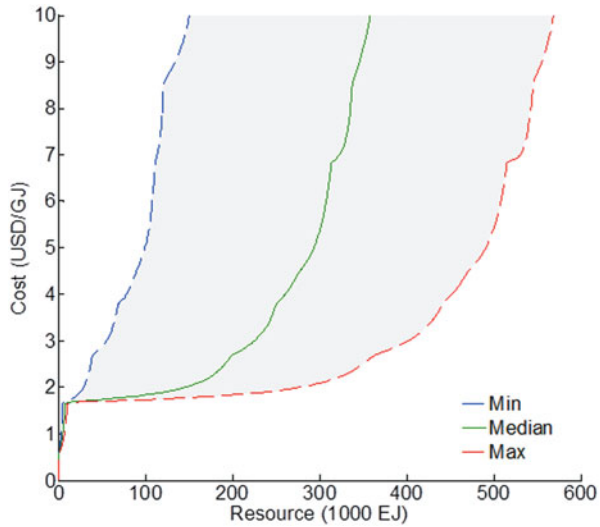
We adopt the values for A and costs at the 1st and 95th percentiles of A ($C_{0.01}$, $Q_{0.01}$) and ($C_{0.95}$, $Q_{0.95}$) provided by Mercure and Salas (2012). To represent uncertainty, for each model simulation we draw a value for A from a triangle distribution with mode, lower value, and upper value equal to the most probable, lower bound, and upper bound values on technical potential given by Mercure and Salas (2012). We also draw a value for $C_{0.95}$ representing the uncertainty in cost reduction due to technological innovation in extraction from a triangular distribution with mode at the value given by Mercure and Salas, lower bound at 80 % of the modal value, and upper bound at 120 % of the modal value. These randomly selected values are then used together with the given values of $C_{0.01}$, $Q_{0.01}$, and $Q_{0.95}$ to calculate values for C_0 and B from Eqs. (6) and (7) and all parameters are held constant over time for each simulation.

Distributions for the nine primary fossil-based energy resources are summarized in Table 1. For any given simulation, the nine cost–supply curves are assumed to be independent and are therefore aggregated by summing across all resources for each cost value (Fig. 3).

Government Agent

In the original DFR model, the government has the ability to collect a tax on other agents and use the revenue for a variety of purposes (e.g., to subsidize R&D by firms). Gerst et al. (2013b) use this modeling capability to assess the impact of a carbon tax on energy technology, energy use, carbon emissions, and economic growth. They use an exogenously-specified, increasing carbon tax and compare the impacts of three different revenue recycling schemes: (a) returning revenues to households in the form of a tax rebate, (b) using revenues to subsidize innovation by capital goods firms, and (c) investing revenues in renewable technology R&D.

Fig. 3 Summary of probabilistic fossil-based energy cost–supply curves used in the model. Lines indicate minimum (*blue*), median (*green*), and maximum (*red*) cost curves based on 500 simulations



Gerst et al. (2013b) found that, on its own, the carbon tax does not provide enough of a price signal to markedly alter the energy technology mix in the model: the carbon-light energy technology achieves significant market share only about 5 years earlier in schemes (a) and (b) than in a no-tax reference specification. Only when the carbon tax revenue is used to subsidize renewable energy technology R&D (scheme c) does the energy system transition away from carbon-emitting technologies within the next century. As mentioned earlier, however, the model of Gerst et al. (2013b) assumes a stable population, fixed fuel cost curve, and exogenous carbon tax schedule. All of these limitations can be expected to have a significant effect on results, both in terms of most likely outcomes and estimates of uncertainty.

Endogenous Policy Experiment

To simulate endogenous policy formation, we assume that in the year 2000, nations agree to emissions pathways that will lead to a stabilization of climate forcing of 4.5 W/m^2 by 2100. The necessary annual emissions commitments are given in Table 2 and are consistent with RCP4.5, as calculated by GCAM (global change assessment model; Thomson et al. 2011).

We assume that to meet its commitments the US government adopts a carbon tax with initial value of US\$ 25 per tonne CO_2 , increasing at a nominal rate of 5% per year. The effectiveness of the tax is monitored every 10 years by comparing actual cumulative carbon emissions against cumulative emissions commitments resulting from Table 2. If actual cumulative emissions are at or below the target, then the carbon tax growth rate remains the same. If cumulative emissions are above the target, then the annual carbon tax growth rate is adjusted upward by 0.5%. All carbon tax revenue

Table 2 Annual emissions commitments in PgC (petagrams of carbon) per year for the USA and the rest of the world (ROW)

Regions	Year										
	2000	2010	2020	2030	2040	2050	2060	2070	2080	2090	2100
USA	1.55	1.62	1.59	1.58	1.47	1.26	1.06	0.81	0.50	0.47	0.44
ROW	7.26	8.50	9.77	10.88	11.30	11.03	9.40	7.12	4.18	4.19	4.20

is used to subsidize renewable technology R&D, consistent with the most effective policy considered by Gerst et al. (2013b).

Our interest in the policy experiment, as described, is to determine the extent to which we can identify the socioeconomic and technological factors (the columns of the van Vuuren matrix) that lead to a specific RCP (the rows of the matrix framework). We accomplish this by generating a large number of stochastic model simulations to which we apply the multidimensional scenario discovery method described by Gerst et al. (2013a).

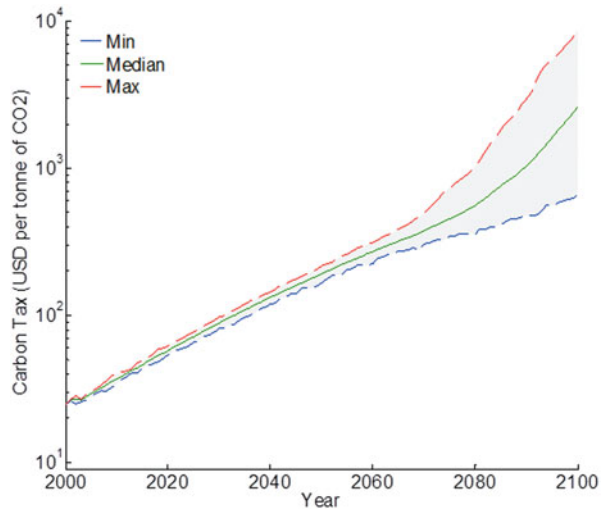
Model Simulation

Our model was calibrated to the U.S. historical rates of growth for GDP (gross domestic product) per household (1.7 % per year) and residential energy use per household (0.7 % per year) by adjusting the distributions representing stochasticity of labor productivity and energy efficiency improvement by the capital goods sector. For the purposes of calibration, we simulated the period 1820–2000, assuming a historical energy price increase of 1.0 % per year and constant average economy-wide labor and energy unit costs. This assumption was necessary to ensure that modeled technological improvements kept pace with increases in wages and energy price. Other model parameters mostly adopted the values of Dosi et al. (2010), as reported by Gerst et al. (2013b).

For our policy simulation, starting conditions were specified by selecting the simulated year 2000 state from the final calibration that most closely matched the actual investment fraction of GDP and household fraction of total energy use observed in 2000. Wages, energy price, and other parameters were then scaled to match the observed year 2000 values. This procedure preserved the agent heterogeneity generated in the calibration exercise, while allowing initial conditions to accord with overall macro variables observed for the year 2000.

For computational tractability, our model of the US economy is scaled to be represented by 50 capital goods firms, 200 consumer goods firms, and 250,000 households in the year 2000. The number of households then scales proportionally with population change, as represented by Eqs. (1–3). In the current version of the model, the number of firms remains constant, although production and labor demands can change with population.

Fig. 4 Modeled inflation-adjusted carbon tax. Lines indicate minimum (blue), median (green), and maximum (red) values of 500 simulations



To represent the range of possible model outcomes, 500 simulations were used to generate all figures and statistics. These simulations represent stochastic realizations of the model's dynamics emerging from the same set of initial conditions and model parameter values. Stochasticity arises from the uncertain technological development process. Random draws are taken each year from the distributions characterizing innovation and imitation success of firms seeking to improve labor productivity and energy efficiency. Similarly, energy technology firms reduce the cost of manufacturing low-carbon and carbon-free energy technologies through a two-factor learning curve characterized by stochastic rates of learning-by-searching and learning-by-doing effects.

Results

The price signal introduced by a growing carbon tax (Fig. 4) potentially acts through two channels to influence technological change and carbon emissions: (i) the machine purchasing decisions of consumer good firms (and therefore the incentive structure of capital goods firms) and (ii) the capital budgeting decisions of energy producers. As already revealed by Gerst et al. (2013b), the carbon tax on its own does not lead to substantial improvements in energy efficiency of produced machines or consumer goods beyond what would otherwise be achieved under a no-carbon tax scenario. Thus, even with an inflation-adjusted tax level of over US\$ 100 per tonne CO₂, model results indicate that the USA is unlikely to achieve the annual emissions commitments of Table 2 by mid-century (Fig. 5).

On the energy supply side, however, the effects of the carbon tax can be substantial—not necessarily because of the price signal, which is small compared

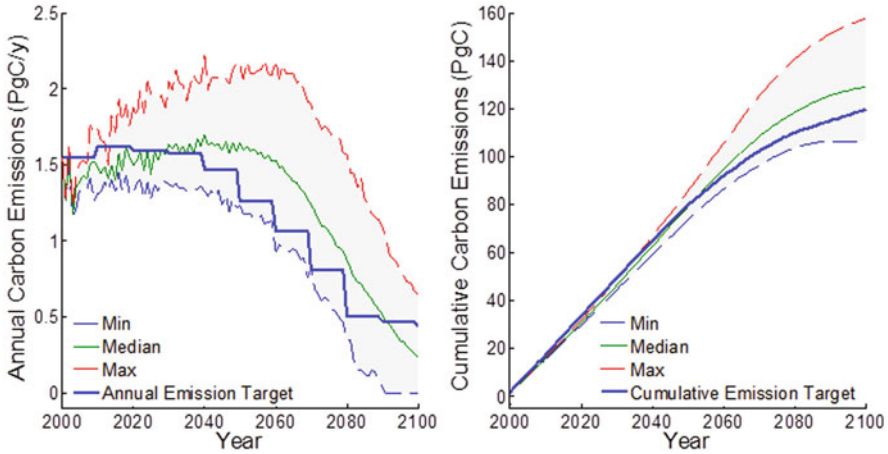


Fig. 5 Predicted annual and cumulative carbon emission. *Lines* indicate minimum (*blue*), median (*green*), and maximum (*red*) predicted trajectory from 500 simulations. Bold lines represent target emissions

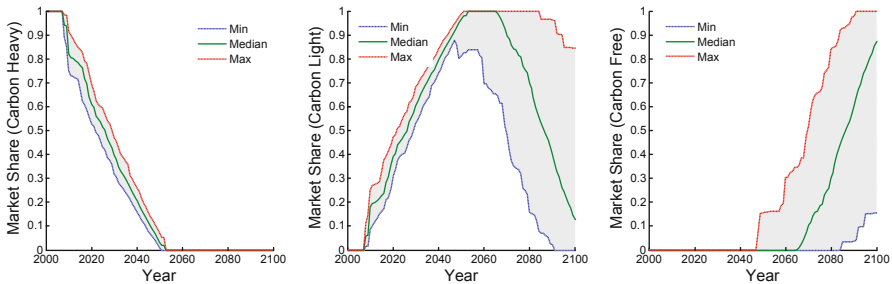


Fig. 6 Predicted market share for each energy technology. *Lines* indicate minimum (*blue*), median (*green*), and maximum (*red*) of 500 simulations

to the possible rise in future fuel costs (Fig. 3), but because of the dramatic influence of the subsidization of energy technology R&D that a carbon tax enables. By mid-century, carbon-free renewable energy begins to achieve significant market penetration in most simulations (Fig. 6). There is substantial uncertainty in the breakthrough year, due to the inherent stochasticity of technology improvement, giving rise to large uncertainty in predicted emissions in mid-century (see Fig. 5). This uncertainty is exacerbated by uncertainty in the fuel cost and population growth curves. However, once carbon-free sources take hold as a major contributor to the national energy mix, the economy becomes essentially uncoupled from fuel costs, resulting in the potential for dramatic economic growth by the end of the twenty-first century (Fig. 7a).

Due to technological improvement of energy technology, capital goods, and consumer goods, the economy-wide energy use per dollar GDP is predicted to decrease substantially over time (Fig. 7). Our model predicted decrease in energy intensity,

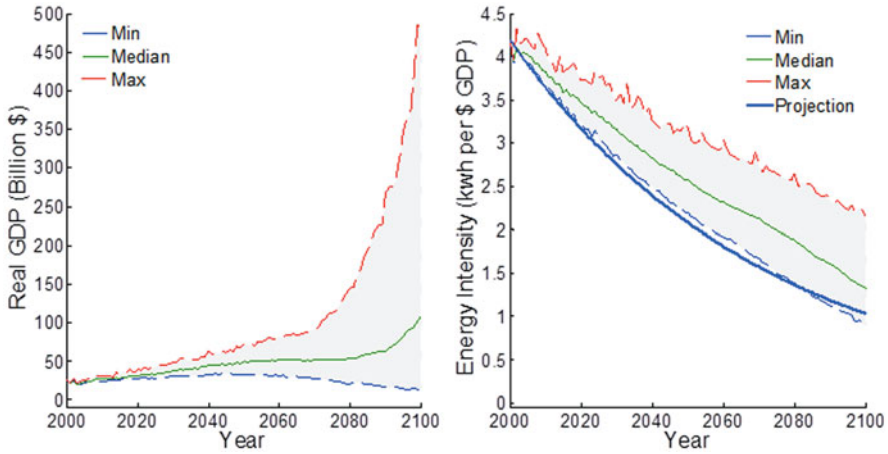


Fig. 7 Predicted trajectories for real GDP and energy intensity. *Thin lines* indicate minimum (*blue*), median (*green*), and maximum (*red*) trajectories of 500 simulations. *Bold line* indicates the projected change in energy intensity at the historical average rate of 1.39 % per year

however, is less than the average historical annual decline of 1.39 % from 1949–2009 (projected as the bold line in Fig. 7). The high historical decline in energy intensity is known to be due, at least in part, to broad structural changes that have occurred over the past 60 years, such as shifts from a manufacturing to a service-oriented economy, and changes in the international trade balance (Sue Wing 2008). These trends may or may not continue over the next century, but in any case, they are not currently represented in the model.

Scenario Discovery

Description of Method

We employ the method for multidimensional scenario discovery described by Gerst et al. (2013a). Each simulation is first represented by the values of two or more selected outcome variables. The full set of simulations is then subject to a hierarchical clustering algorithm to identify statistically similar groups according to these selected outcomes. Finally, these clusters, or ‘candidate scenarios,’ are subject to a classification analysis to identify the stochastic model inputs that serve as key scenario drivers. The results of this classification are then taken to represent the final ‘discovered’ scenarios. The notion of multidimensional similarity is what distinguishes our cluster-based technique from threshold-based methods (Bryant and Lempert 2010) or full-factorial “quadrant-based” scenario definitions.

We implemented our hierarchical cluster analysis using available functions in MATLAB. Distances between points were calculated using Euclidean distance and

clustering-employed Ward's method. For our application, we chose to cluster according to two dimensions: average GDP per capita growth rate (excluding climate damages) and cumulative carbon emissions, both for the period 2000–2100. These two outcome variables capture the key tradeoff of the climate policy: weighing the potential economic impacts of abatement versus the potential for climate impacts.

For classification analysis, we used the *ClassificationTree.fit* function of MATLAB. Classification trees represent dichotomous splits of independent variables that yield the strongest associations with a categorical dependent variable. In our context, independent predictors consisted of the nine constructed variables characterizing stochastic technological development used by Gerst et al. (2013b), as well as two additional probabilistic parameters used in the model extensions described in the present contribution: the US population in 2100 (P_{2100}) and the total energy supply potential across all fuel types (A_{tot}). The groups of model simulations (i.e., candidate scenarios) identified by the cluster analysis served as the dependent variable. The independent variables that best predict candidate scenario membership are then interpreted as the key driving forces, and the combination of conditions on these variables is then taken to define the final 'discovered' scenarios. To maintain an easily interpretable tree, we set the minimum number of simulations for splitting each node to 160 and the minimum number of simulations for each final branch to 45.

Scenario Results

As already shown in Figs. 5 and 7, there is large variation in carbon emissions and GDP growth under our simulated policy setting. This makes the results especially conducive to scenario discovery. Hierarchical cluster results (not shown) indicate that the model simulations, as represented by the two selected outcome variables, naturally divide into four clusters. A bivariate scatterplot of the cumulative carbon emissions and average GDP per capita growth rate for the four clusters (Fig. 8) indicates that this number represents a range of reasonably distinct groupings. The fact that these groupings do not conform neatly to quadrants of the two-dimensional space suggests that the use of empirical cluster analysis holds some value over threshold-based methods.

As the next step to scenario discovery, the classification tree (Fig. 9) indicates that the four clusters defined in the two-dimensional space of carbon emissions and GDP growth can also be reasonably distinguished by four partitions over three stochastic model variables. The three variables selected empirically as strong predictors are: (i) the population size in 2100 (P_{2100}), (ii) the relative efficacy of R&D with respect to labor productivity to produce consumer goods (EFF_A), and (iii) the relative efficacy of carbon free energy technology experience (EXPER_{cf}). The other eight variables in the candidate set of predictors appear not to be strong drivers of policy performance.

We take the partitioning defined by the classification tree in Fig. 9 to be our final set of four 'discovered' scenarios. Although defined with respect to only three variables, these scenarios represent a complete partitioning of the 500 model simulations in the

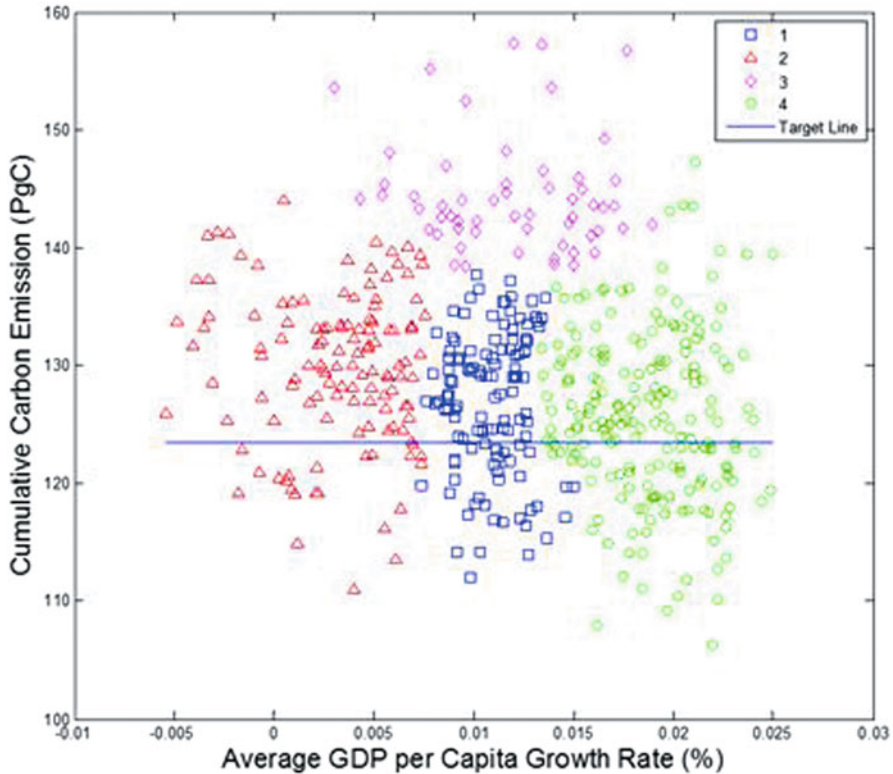


Fig. 8 Scatterplots of cumulative carbon emissions and average GDP per capita growth rate. *Points* represent the 500 stochastic simulation results. *Symbols* represent groupings identified by the cluster analysis and serve as candidate scenarios. The *horizontal line* indicates the cumulative emissions target for 2100

multidimensional space of all stochastic model variables and outcomes. The defining characteristics of these scenarios can be best viewed as a set of boxplots comparing the range of conditions experienced under each scenario (Fig. 10).

Scenario 1 is characterized by low levels of carbon emissions and moderate GDP per capita growth, associated with low to moderate levels of population growth and labor productivity improvement, but high efficiency in converting experience with carbon-free technology into emissions reductions (i.e., learning-by-doing). Scenario 2 on the other hand, has moderate emissions and very low GDP growth, associated with poor efficiency of learning-by-doing. Scenario 3 has the highest emissions levels and moderate-to-high levels of GDP per capita growth, associated primarily with very high population growth (greater than about 546 million by 2100). Finally, scenario 4 might be considered the most successful overall for achieving the lowest emissions and highest GDP growth. These results from low population growth are combined with high efficiency in converting R&D funding into improvements in labor productivity.

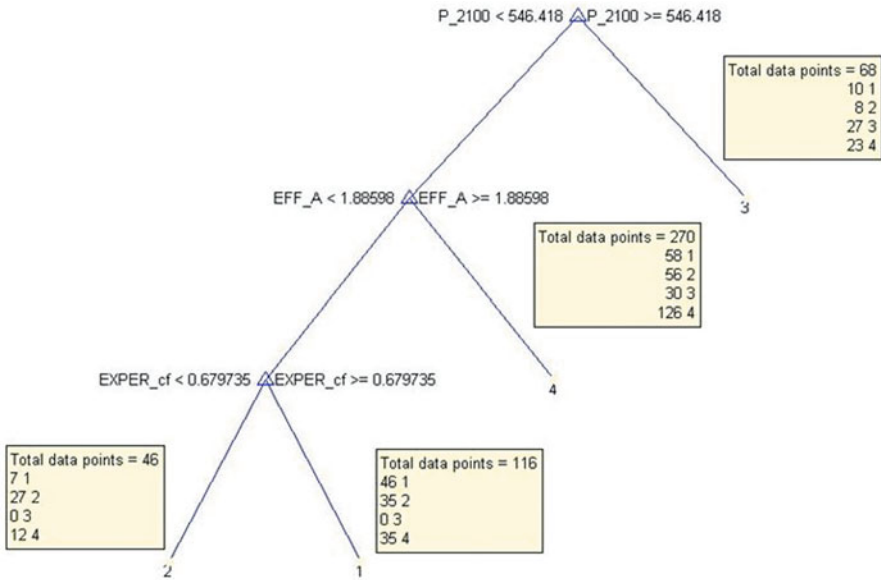


Fig. 9 Classification tree indicating the optimal partitioning of stochastic model variables for predicting candidate scenarios resulting from the cluster analysis. At each split, observations less than the indicated value proceed to the *left branch* and observations greater proceed to the *right*. Each split is conditional on the result of the splits above it in the tree. The bottom branches are labeled with the predicted cluster membership. *Boxes* indicate the total number of simulations that meet all the specified conditions leading to the corresponding branch, as well as the actual categorical membership frequencies among these simulations

Discussion

We demonstrate how the process of scenario discovery as applied to results of ENGAGE, a stochastic, dynamic agent-based model, might be used to generate socioeconomic scenarios relevant to a given emissions target, or RCP. For a carbon tax policy designed to meet the 4.5 W/m² RCP, population growth, improvement in labor productivity, and efficiency of learning-by-doing regarding carbon-free energy technology are revealed to be the key factors driving policy success. In particular, a low population growth and a high ability to convert experience in carbon-free energy technology into further cost reductions seem to be jointly, a key to meeting emissions targets with minimal negative economic impact. This implies that these features should form the key elements of the storylines underlying socioeconomic scenarios associated with the 4.5 W/m² RCP if they are to provide a meaningful exploration of policy efficacy. Such scenarios, which pair varying levels of population and economic growth with differing degrees of innovation in the energy sector, are consistent with those generated using more conceptual methods in the climate scenario literature (Moss et al. 2010; Parson et al. 2007; van Vuuren et al. 2012). However, by being derived from the results of a quantitative model, our specification is intrinsically consistent with practicable modeling assumptions and parameterizations.

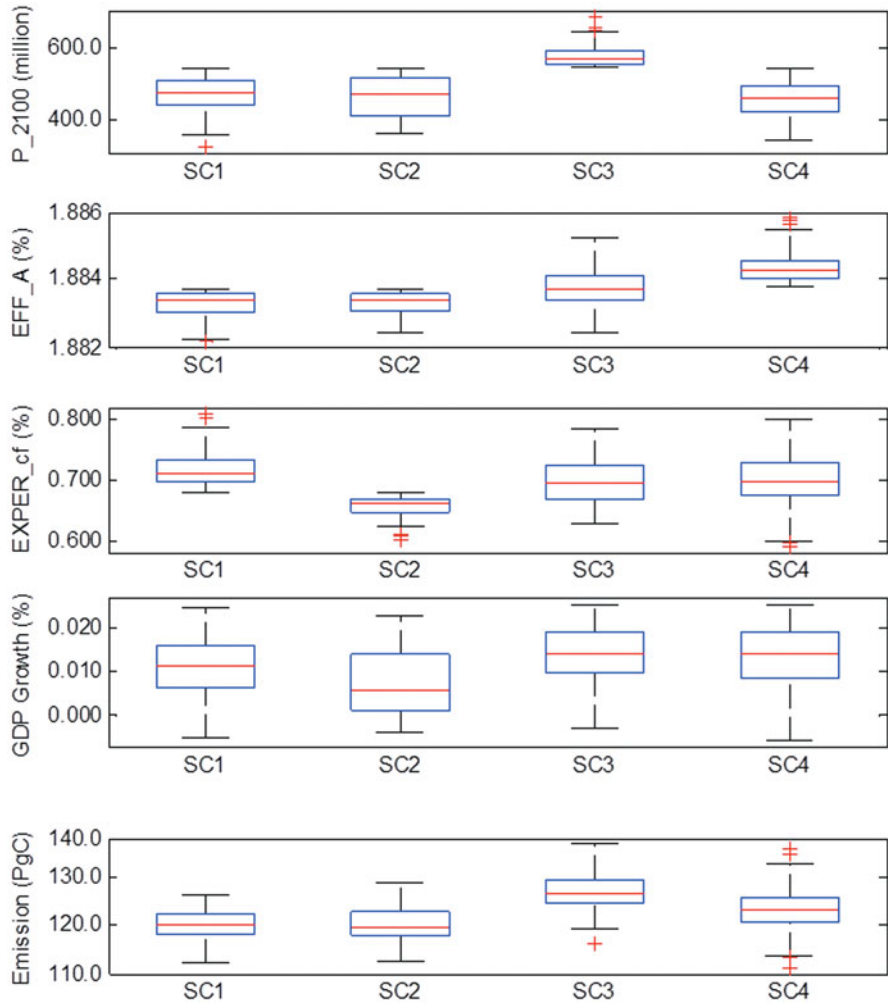


Fig. 10 Boxplots summarizing conditions associated with the four final discovered scenarios. *Boxes* indicate the middle 50% of the simulation values (interquartile range, IQR) for each scenario, *central lines* indicate median values, *vertical whiskers* extend out to the furthest simulation value within 1.5*IQR of the boxes, and *crosses* indicate further outlying values. Scenario numbering corresponds to Figs. 8 and 9 and variables are defined as described in text. All variables are reported on an annual basis, except for P2100 which is the population in the year 2100 and emissions which is cumulative from 2000 to 2100

While in the current contribution, we have overcome some of the key limitations of earlier versions of ENGAGE by allowing for a growing population and uncertain fuel price, there are still a number of simplifying assumptions that we believe are too great to allow direct application of our current results to real-world policy questions. For example, the current simplicity of the energy sector may overlook opportunities

for technology innovation and adoption. In particular, we only represent one energy production firm, and it is assumed to utilize the full lifetime of its energy technologies. Thus, it will not prematurely scrap any of its existing stock when improved carbon-light or carbon-free technology becomes available. Also, cost is currently the only factor in the model determining new technology adoption, precluding early adoption to meet moral obligation or public relations objectives. These factors add a significant lag to the achievement of carbon emissions reductions in the model.

Finally, the decision rules of households and firms in our current model are currently homogenous and simplified. For example, firms cannot focus R&D effort toward specific machine attributes or make decisions to hedge against anticipated energy price increases. Similarly, households have homogenous preferences for three needs that do not represent the true diversity of personal values and beliefs. We are currently working to alleviate these limitations by defining a suite of decision rules that households use to select goods that meet both their individual and social needs.

We recognize that further progress is necessary for ENGAGE to provide useful support for climate policy evaluation and formulation. Nevertheless, we believe that our proposed combination of stochastic, agent-based modeling and multidimensional scenario discovery can contribute to the ongoing climate scenario development effort by complementing traditional approaches. Furthermore, multidimensional scenario discovery may be used with any model that has the capability to generate probabilistic output. Other areas of energy and climate policy that exhibit considerable uncertainty and disagreement over metrics such as impacts, adaptation, vulnerability assessments, and regional infrastructure planning, could benefit from this approach.

References

- Bryant, B.P., Lempert, R.J.: Thinking inside the box: a participatory, computer-assisted approach to scenario discovery. *Technol. Forecast. Soc. Chang.* **77**(1), 34–49 (2010)
- Clarke, L., Edmonds, J., Krey, V., Richels, R., Rose, S., Tavoni, M.: International climate policy architectures: overview of the EMF 22 International Scenarios. *Energy Econ.* **31**, S64–S81 (2009)
- Dosi, G., Fagiolo, G., Roventini, A.: Schumpeter meeting Keynes: a policy-friendly model of endogenous growth and business cycles. *J. Econ. Dyn. Control.* **34**(9), 1748–1767 (2010)
- Fisher, B., Nakicenovic, N., Alfsen, K., Corfee Morlot, J., de la Chesnaye, F., Hourcade, J.-C., Jiang, K., Kainuma, M., La Rovere, E., Matysek, A., Rana, A., Riahi, K., Richels, R., Rose, S., van Vuuren, D.P., Warren, R.: Issues related to mitigation in the long-term context. In: Metz, B., Davidson, O., Bosch, P., Dave, R., Meyer, L. (eds.) *Climate Change 2007—Mitigation*. Cambridge University Press, Cambridge (2007)
- Gerst, M.D., Wang, P., Borsuk, M.E.: Discovering plausible energy and economic futures under global change using multidimensional scenario discovery. *Environ. Modell. Softw.* **44**, 76–86 (2013a)
- Gerst, M.D., Wang, P., Roventini, A., Fagiolo, G., Dosi, G., Howarth, R.B., Borsuk, M.E.: Agent-based modeling of climate policy: an introduction to the ENGAGE multi-level model framework. *Environ. Modell. Softw.* **44**, 62–75 (2013b)
- Groves, D.G., Lempert, R.J.: A new analytic method for finding policy-relevant scenarios. *Global Environ. Chang.* **17**(1), 73–85 (2007)

- Kriegler, E., O'Neill, B.C., Hallegatte, S., Kram, T., Lempert, R.J., Moss, R.H., Wilbanks, T.: The need for and use of socio-economic scenarios for climate change analysis: a new approach based on shared socio-economic pathways. *Global. Environ. Chang.* **22**(4), 807–822 (2012)
- McCarthy, J., Canziani, O., Leary, N., Dokken, D., White, K. (eds.): *Climate Change 2001: Impacts, Adaptations, and Vulnerability*. Cambridge University Press, Cambridge (2001)
- Mercure, J.-F., Salas, P.: An assessment of global energy resource economic potentials. *Energy.* **46**(1), 322–336 (2012)
- Moss, R.H., Edmonds, J.A., Hibbard, K.A., Manning, M.R., Rose, S.K., van Vuuren, D.P., Carter, T.R., Emori, S., Kainuma, M., Kram, T.: The next generation of scenarios for climate change research and assessment. *Nature.* **463**(7282), 747–756 (2010)
- Nakicenovic, N., Kolp, P., Riahi, K., Kainuma, M., Hanaoka, T.: Assessment of emissions scenarios revisited. *Environ. Econ. Policy. Stud.* **7**(3), 137–173 (2006)
- Nakicenovic, N., Swart, R. (eds.): *Special Report on Emissions Scenarios*. Cambridge University Press, Cambridge (2000)
- Parson, E., Burkett, V., Fisher-Vanden, K., Keith, D., Mearns, L., Pitcher, H., Rosenzweig, C., Webster, M.: *Global Change Scenarios: Their Development and Use*. U.S. Climate Change Science Program and the Subcommittee on Global Change Research. Department of Energy, Office of Biological & Environmental Research, Washington, D.C. (2007)
- Raftery, A.E., Li, N., Sevcikova, H., Gerland, P., Heilig, G.K.: Bayesian probabilistic population projections for all countries. *P. Natl. Acad. Sci.* **109**(35), 13915–13921 (2012)
- Rozenberg, J., Hallegatte, S., Vogt-Schilb, A., Sassi, O., Guivarch, C., Waisman, H., Hourcade, J.-C.: Climate policies as a hedge against the uncertainty on future oil supply. *Clim. chang.* **101**(3), 663–668 (2010)
- Rozenberg, J., Guivarch, C., Lempert, R.J., Hallegatte, S.: Building SSPs for climate policy analysis: a scenario elicitation methodology to map the space of possible future challenges to mitigation and adaptation. FEEEM Working Paper No 522012 (2012)
- Schwartz, P.: *The Art of the Long View: Planning for the Future in an Uncertain World*. Currency Doubleday, New York (1991)
- Sue Wing, I.: Explaining the declining energy intensity of the US economy. *Resour. Energ. Econ.* **30**(1), 21–49 (2008)
- Thomson, A.M., Calvin, K.V., Smith, S.J., Kyle G.P., Volke, A., Patel, P., Delgado-Arias, S., Bond-Lamberty, B., Wise, M.A., Clarke, L.E.: RCP4. 5: a pathway for stabilization of radiative forcing by 2100. *Clim. chang.* **109**(1), 77–94 (2011)
- van Vuuren, D.P., Riahi, K., Moss, R.H., Edmonds, J.A., Thomson, A., Nakicenovic, N., Kram, T., Berkhout, R., Swart, R., Jenetos, A., Rose, S.K., Arnell, N.: A proposal for a new scenario framework to support research and assessment in different climate research communities. *Global. Environ. Chang.* **22**, 21–35 (2012)