Environmental Modelling & Software 80 (2016) 336–353

Contents lists available at ScienceDirect

Environmental Modelling & Software

journal homepage: www.elsevier.com/locate/envsoft

The diversity of socio-economic pathways and CO_2 emissions scenarios: Insights from the investigation of a scenarios database

Céline Guivarch [a, b, *], Julie Rozenberg [a, c], Vanessa Schweizer [d, e]

[a] Cired, Nogent-sur-Marne, France
[b] Ecole des Ponts ParisTech, Champs-sur-Marne, France
[c] Chief Economist Office for Sustainable Development, World Bank, Washington DC, USA
[d] National Center for Atmospheric Research (NCAR), Boulder, CO, USA
[e] University of Waterloo, Department of Knowledge Integration, Waterloo, ON, Canada

A R T I C L E I N F O

Article history:
Received 26 June 2015
Received in revised form
2 March 2016
Accepted 7 March 2016
Available online 21 March 2016

Keywords:
Scenario discovery
Diversity
Socioeconomic pathways
Emissions scenarios
Database

A B S T R A C T

The new scenario framework developed by the climate change research community rests on the fundamental logic that a diversity of socio-economic pathways can lead to the same radiative forcing, and therefore that a given level of radiative forcing can have very different socio-economic impacts. We propose a methodology that implements a “scenario discovery” cluster analysis and systematically identifies diverse groups of scenarios that share common outcomes among a database of socio-economic scenarios. We demonstrate the methodology with two examples using the Shared Socio-economic Pathways framework. We find that high emissions scenarios can be associated with either high or low per capita GDP growth, and that high productivity growth and catch-up are not necessarily associated with high per capita GDP and high emissions.

Data availability

The database of scenarios outcome analyzed in this article, together with associated scenario drivers is available upon request.

1. Introduction

The scientific community has been developing a new generation of scenarios for climate change research and assessment (Moss et al., 2010) to replace the SRES scenarios (Nakicenovic et al., 2000) that have been widely used over the past decade.

One of the main innovations in this new generation of scenarios is relaxing the coupling between socioeconomic scenarios and emissions scenarios. Whereas previous generations of scenarios followed a sequential process starting from projecting different socioeconomic futures and then estimating the corresponding emissions, concentrations, radiative forcing and the ensuing climate change, the new generation initiated a parallel process (O'Neill and Schweizer, 2011; Ebi et al., 2014). This process starts with the selection of radiative forcing pathways (van Vuuren et al., 2011), from which investigations are proceeding in parallel: climate modellers simulate the climate change resulting from these radiative forcing pathways (Taylor et al., 2012) while scientists from the mitigation and adaptation communities develop socio-economic scenarios that could lead to these pathways.

The fundamental logic of this new architecture is the idea that each radiative forcing pathway is not associated with a unique socioeconomic scenario, and instead can result from different combinations of economic, technological, demographic, policy and institutional futures. This was a fundamentally important finding of the SRES scenarios, that alternative combinations of driving forces (e.g. population and economic growth) could lead to similar levels and structures of energy and land use.

This new architecture has therefore the potential to overcome some of the main difficulty in scenario studies: “The more detail that one adds to the story line of a scenario, the more probable it will appear to most people, and the greater the difficulty they likely will have in imagining other, equally or more likely, ways in which the same outcome could be reached.” (Morgan and Keith, 2008). The new scenarios architecture offers the potential to help to identify a range of different technological, socioeconomic and policy futures that could lead to a particular concentration pathway. But exploring and managing the plethora of possible future trends in socio-economic development is also a great challenge for the developers of scenarios.

* Corresponding author. Cired, Nogent-sur-Marne, France.
E-mail address: guivarch@centre-cired.fr (C. Guivarch).

http://dx.doi.org/10.1016/j.envsoft.2016.03.006
1364-8152/

To address this challenge, it was decided to focus first on a small set of socio-economic pathways contrasted along two axes, the challenges for mitigation and the challenges for adaptation, in the Shared Socio-economic Pathways (SSP) framework (Nakicenovic et al., 2014; O'Neill et al., 2014). The Story and Simulation approach (Alcamo, 2008; O'Neill et al., 2015) was used to develop five storylines (O'Neill et al., 2015) and quantify five respective marker scenarios. To exploit the full potential of the new scenarios architecture, it was suggested to establish and utilize large databases of scenarios (Ebi et al., 2014) so as to allow scenario users to select a set of well-informed, self-consistent scenarios customized for their particular application. Indeed, the "Shared Socio-economic Pathways are canonical, but the canon is not exclusive", as O'Neill et al. (2015) conclude, and for some uses of scenarios, it may be necessary to explore the diversity of socio-economic drivers that lead to specific outcomes.

Here we propose, with two example applications, a methodology that contributes to this longer-term research agenda by giving a quantified and systematic way to examine the diversity of socio-economic scenarios leading to similar outcomes. As noted by O'Neill et al. (2015), a novel feature of the SSP framework is that the relevant uncertainties of scenarios are associated primarily with alternative *outcomes* or *results* rather than the scenario *drivers* or *inputs* leading to outcomes. Nevertheless alternative drivers should be considered in order to explore differences in outcomes. The methodology featured in this paper builds on Rozenberg et al. (2014), who constructed a large set of scenarios by varying scenario drivers systematically to explore the space of possible outcomes relevant to the SSP framework. This process generated hundreds of scenarios that comprise what we call a "scenarios database". These scenarios were then classified into different scenario archetypes systematically rather than through developing only a few contrasting cases. Schweizer and O'Neill (2014) similarly assess a large number of scenarios by systematically varying the states of qualitative scenario elements. However only Rozenberg et al. (2014) apply this technique to the inputs of an integrated assessment model and run the model a large number of times to construct a database of integrated assessment outputs. Here, we investigate the scenario classifications further and select subsets of scenarios of interest, by their outcomes. To do this, we adapt a "scenario discovery" cluster analysis to uncover the diversity of driver combinations leading to the outcomes of interest.

The rest of this article is structured as follows. Section 2 presents the methods used. Section 3 describes two applications of this methodology. Because the development of the methodology was inspired by the SSP framework, the first application explores the diversity of scenarios classified as a particular type of SSP. The second application is not strictly linked to the SSP framework, as it explores more broadly the diversity of scenarios with high cumulative CO_2 emissions. Section 4 discusses the results and methods and concludes.

2. Methods

The methodology proceeds in three steps. First we identify *a priori* the main driving forces (or main uncertainties) affecting the future outcomes of the system, such as population growth or fossil fuel reserves in the case of the human-climate system. Second we translate these driving forces into parameters for a model of the system studied, and we combine these parameters to build a large number of model runs. Note that the purpose of doing this is to systematically explore the implications of different combinations of drivers. Third we select a subset of scenario outcomes of interest, e.g. scenarios leading to high greenhouse gases (GHG) emissions. Because a diversity of drivers could have led to these outcomes, we iterate a "scenario discovery" cluster analysis to identify the diversity of combinations of drivers that lead to the selected subset of scenario outcomes. Section 2.1 describes the iteration of the scenario discovery cluster analysis. Section 2.2 presents the model we used and the driver or parameter sets that are systematically varied to build a database of scenarios.

2.1. Scenario discovery to uncover the diversity of scenarios

"Scenario discovery" cluster analysis provides a computer-assisted method of scenario development that applies statistical algorithms to databases of simulation model results to characterize the combinations of uncertain input parameter values (or "drivers") most predictive of specified classes of results (Lempert et al., 2006). In other terms, the "scenario discovery" analysis is a systematic manner to find which combinations of the model input parameters lead to specific results of interest, i.e. cases where a given output variable is above or below a defined threshold or, more generally, cases where output variables are in specified areas of the results space. To describe the subset of scenarios of interest, i.e. the cases where results are in specified areas of the results space, we define a binary indicator associated with scenarios. It takes the value 1 if scenarios belong to the subset of scenarios of interest and the value 0 otherwise.

We use the version of the PRIM (Patient Rule Induction Method) for the cluster analysis by Bryant and Lempert (2010). PRIM identifies several combinations of drivers, called "boxes", and their coverage and density. Coverage is the fraction of scenarios consistent with the outcome indicators of interest, i.e. for which the indicator is equal to 1, that are also in the box (i.e. the combination of drivers) identified. Coverage corresponds to "recall" in the classification and information retrieval literature. Density is the fraction of all scenarios in the box that are also in the subset of scenarios of interest. For example a density of 100% means that all scenarios in the box are in the subset of scenarios of interest; a density below 50% means that more than half of the scenarios in the box are, in fact, outside the subset of scenarios of interest. Density is analogous to "precision" or "positive predictive value" in the classification and information retrieval literature. Since these two measures are generally in tension with one another, PRIM provides the user a set of options representing different trade-offs among density and coverage.

PRIM cluster analysis has been applied to a number of cases: pollution-control strategies (Lempert et al., 2006), long-term water planning (Groves and Lempert, 2007), the efficacy of a proposed renewable energy standard (Bryant and Lempert, 2010), low-carbon energy technologies (McJeon et al., 2011), near term policy choices for greenhouse gas transformation pathways (Isley et al., 2015), vulnerabilities of the port of Rotterdam (Halim et al., 2015), or the design of effective policies for stimulating bio-methane production in the Netherlands (Eker and van Daalen, 2015). These applications explicitly tie scenario discovery to policy design by identifying scenarios that give rise to vulnerabilities for a proposed policy (i.e. that cause the policy to fail). In recent years, scenario discovery has been used independently from a specific policy design. For example, Gerst et al. (2013) use scenario discovery to "discover" plausible energy and economic futures, Kwakkel et al. (2013) to study the future of copper, and Rozenberg et al. (2014) to explore the space of challenges to mitigation and challenges to adaptation in the SSP framework. Other clustering methods also exist, in particular CART (Breiman et al., 1984), and have been applied to "discover" combinations of drivers leading to specific scenario outcomes (for example Gerst et al., 2013 use CART). For a comparison between PRIM and CART cluster analysis, one may refer to Lempert et al. (2008), Hadka et al. (2015) and Kwakkel and Jaxa-Rozen (2016). However, all of those applications of cluster analysis were focusing on identifying the main combination of drivers