Modelling Uncertainty

Developing and Using Simulation Models for Exploring the Consequences of Deep Uncertainty in Complex Problems

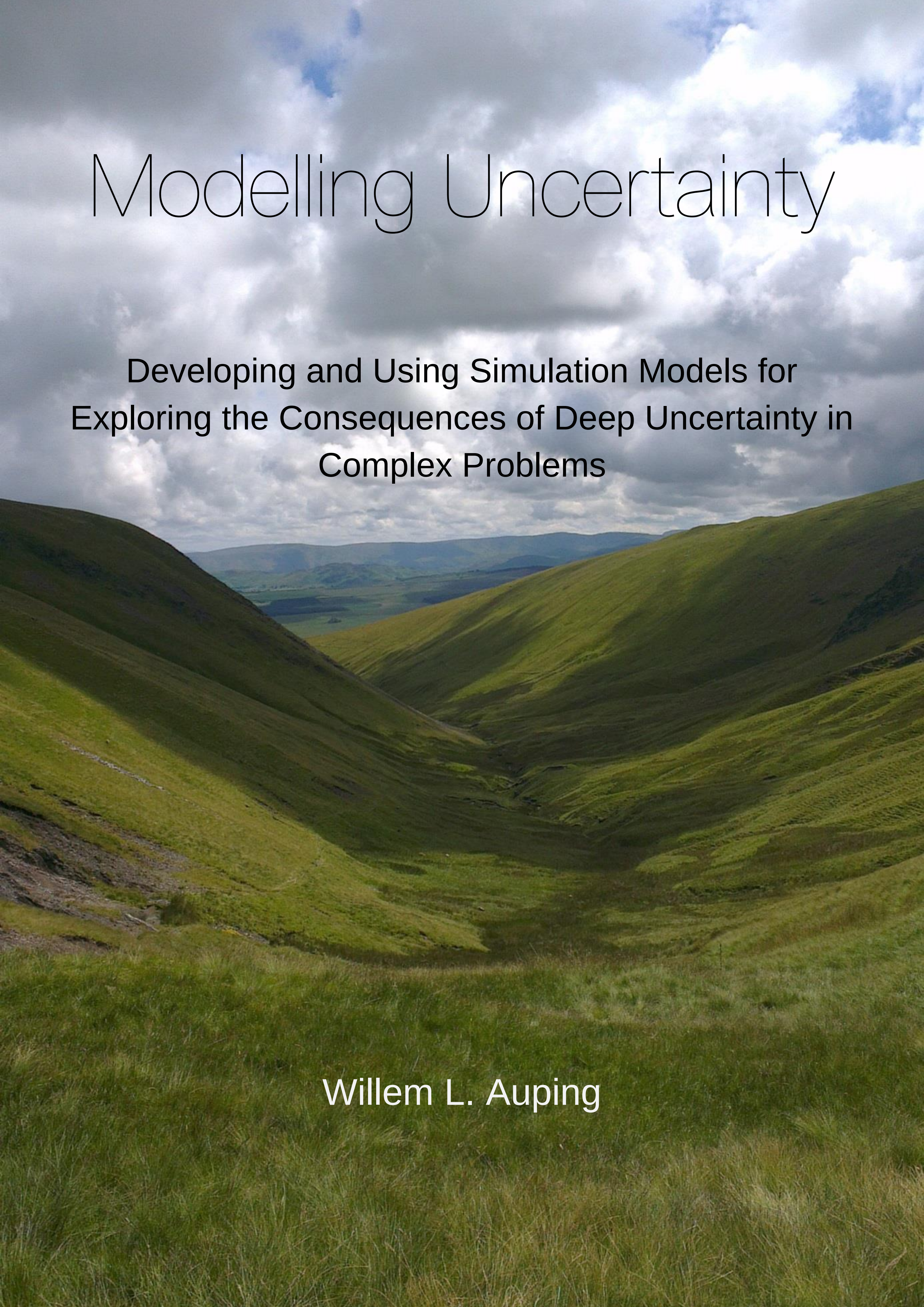
Auping, Willem

DOI
10.4233/uuid:0e0da51a-e2c9-4aa0-80cc-d930b685fc53

Publication date
2018

Document Version
Final published version

Citation (APA)
Auping, W. (2018). *Modelling Uncertainty: Developing and Using Simulation Models for Exploring the Consequences of Deep Uncertainty in Complex Problems*. [Dissertation (TU Delft), Delft University of Technology]. https://doi.org/10.4233/uuid:0e0da51a-e2c9-4aa0-80cc-d930b685fc53



Modelling Uncertainty

Developing and Using Simulation Models for Exploring the Consequences of Deep Uncertainty in Complex Problems

Willem L. Auping

Modelling Uncertainty

Developing and Using Simulation Models for Exploring the Consequences of Deep Uncertainty in Complex Problems

Willem Lucas AUPING

Delft University of Technology, 2018

Modelling Uncertainty. Developing and Using Simulation Models for Exploring the Consequences of Deep Uncertainty in Complex Problems

Cover design

Willem L. Auping, Dieneke Jongepier, in cooperation with GVO drukkers & vormgevers

Printed by

GVO drukkers & vormgevers

ISBN

978-94-6332-444-1

Modelling Uncertainty

Developing and Using Simulation Models for Exploring the Consequences of Deep Uncertainty in Complex Problems

PROEFSCHRIFT

ter verkrijging van de graad van doctor
aan de Technische Universiteit Delft,
op gezag van de Rector Magnificus Prof.dr.ir. T.H.J.J. van der Hagen,
voorzitter van het College voor Promoties,
in het openbaar te verdedigen op
13 december 2018 om 15:00

door

Willem Lucas AUPING
Ingenieur in Systems Engineering, Policy Analysis and Management,
Technische Universiteit Delft, Nederland,
geboren te Rotterdam, Nederland.

Dit proefschrift is goedgekeurd door de promotoren.

Samenstelling promotiecommissie bestaat uit:

Rector Magnificus	Voorzitter
Prof.dr.ir. W.A.H. Thissen	Technische Universiteit Delft, promotor
Dr. E. Pruyt	Technische Universiteit Delft, copromotor

Onafhankelijke leden:

Prof.dr.ir. P.M. Herder	Technische Universiteit Delft
Prof.dr. A. Petersen	University College London
Prof.dr. D.P. van Vuuren	Universiteit Utrecht / PBL
Prof. S.W. Popper, PhD.	Pardee Rand Graduate School
J. Morecroft, BSc., MSc, PhD.	London Business School
Prof.dr. B.A. van de Walle	Technische Universiteit Delft (reservelid)

Keywords: Policy Analysis, Deep Uncertainty, Complexity, Grand challenges, Exploratory Modelling & Analysis, System Dynamics, Scenario Discovery, Robust Decision Making

Contents

Summary

Simulation models are increasingly used for exploring the consequences of deep uncertainty in complex societal issues. The complexity of societal grand challenges, often characterised by the interrelatedness of different elements in the systems underlying these challenges, often renders mental simulation impossible, necessitating the use of simulation models to assist human reasoning. In addition, these grand challenges are typically also subject to deep uncertainty, making it, for example, impossible to come to a shared understanding of parts of the system and exogenous inputs to it, or even a shared problem definition. Deep uncertainty is defined by Lempert, Popper, and Bankes (2003) as conditions "where analysts do not know, or the parties to a decision cannot agree on, (1) the appropriate conceptual models that describe the relationships among the key driving forces that will shape the long-term future, (2) the probability distributions used to represent uncertainty about key variables and parameters in the mathematical representations of these conceptual models, and/or (3) how to value the desirability of alternative outcomes".

Under deep uncertainty, simulation models can be used to explore the consequences of different combinations of assumptions about uncertain factors or attributes of the problem situation and the underlying system. This type of simulation model use was introduced in 1993 as Exploratory Modelling and Analysis (EMA). In more recent years, this approach has become a major underpinning of the Decision Making under Deep Uncertainty (DMDU) field.

The treatment of deep uncertainty in much DMDU research can be improved, however. In most DMDU research to date, pre-existing models are used. These models were generally developed for 'consolidative' use: the modellers tried to unify existing knowledge to come a single, 'best' model. While most modellers will agree that these models are not perfect representations of reality, and often agree that they as such cannot be validated in the strict sense of the word, these modellers and their models do not acknowledge deep uncertainty. The use of consolidative models is arguably problematic if one agrees that the issue at hand is characterized by deep uncertainty. Therefore, models are needed that are explicitly developed for 'exploratory' use: models that explicitly incorporate deep uncertainty potentially relevant for the research question or questions at hand. However, little experience and guidance exists regarding development and use of specifically exploratory models.

In this dissertation, a first attempt is made to identify, and provide guidance for, the critical choices made during the development and use of exploratory models. I do this on the basis of four case studies published as five separate papers. The first and second paper concern the future availability of copper. The first paper, "Dealing with

Multiple Models in System Dynamics”, investigates under which conditions three different models with different perspectives on the same problem generate either similar or different behaviour. The second paper, “Dynamic scenario discovery under deep uncertainty: The future of copper”, investigates how to apply Scenario Discovery to time series data. Scenario Discovery is a method to identify scenarios from a large number of computer runs by the use of computerised learning algorithms, and – before this paper – exclusively focussed on end values of runs. The third paper, “The geopolitical impact of the shale revolution”, investigates how to apply exploratory modelling in case of two models with different scopes. This was demonstrated by exploring the potential consequences of the US’ shale boom on state stability of states heavily dependent on oil and gas exports. The fourth paper, “Societal Ageing in the Netherlands”, investigates how differences in problem perception can be taken into account through exploratory modelling. This paper applies this to the impact of societal ageing in the Netherlands on the affordability of ageing-related collective spending and the desirability of ageing policies in the eyes of Dutch citizens. Finally, the fifth paper, “Simulating endogenous dynamics of intervention-capacity development”, investigates a way of accounting for policy implementation uncertainty. This paper uses the intervention capacity development during the 2014 outbreak of the Ebola virus in Liberia as a case study. As most research underlying these papers was performed for clients, there is no strong connection between the different papers apart from the use of the same modelling paradigm (System Dynamics) and the same uncertainty methodology (EMA).

Methods

By reflecting critically on the case studies, I have derived a comprehensive overview of issues a modeller encounters when developing and using exploratory models. These issues can be structured into four broad categories. The first category concerns the different ways in which uncertainty can be acknowledged during the development of exploratory models. The second category focuses on the difficulties that arise during the use of specific tools and methods for exploratory analysis, and analysis of exploratory simulation results. The third category is about the costs of exploratory modelling compared to consolidative approaches. The fourth, and final, category, relates to the communication of EMA results to clients and stakeholders.

Exploratory model development

I found that deep uncertainty has an impact on every phase of the model development cycle. In this dissertation, I define a model as an internally connected set of equations, which is not necessarily parametrised. The model development cycle can be conceived to consist of 5 phases: problem articulation, model conceptualisation, model formulation, model evaluation, and policy testing.

The problem articulation phase aims at articulating the central problem which needs to be researched, using that problem formulation for selecting which elements need to be modelled endogenously (i.e., scoping or boundary selection), and what the time horizon of the model is. Here, uncertainty about the problem formulation (e.g., in the case of wicked problems or societal messes) may lead to multiple scopes: multiple models may need to be developed to accommodate the different ideas about problem and system. Next to this, the scope may be made wider than in traditional approaches if this allows multiple explanations of the same problem. The scope may be made narrower, if it is necessary to test the model's response to different well-established input scenarios (e.g., climate scenarios) or other input scenarios that may be used for testing system resilience.

In the conceptualisation phase, the modeller tries to identify main relations between key variables, which often builds on mental models of stakeholders and experts. Uncertainty in this phase may be reflected by identifying the most important structural uncertainties that need to be included in models. If in the problem articulation phase multiple scopes have been selected, one or more conceptual models need to be developed for each scope. The modeller will, of course, have to communicate the locations of deep uncertainties in conceptual diagrams.

During the model formulation phase the actual simulation model is formulated and implemented. The modeller has to make choices regarding the way in which uncertainties identified in the conceptualisation phase are expressed in the model or models. For example, the modeller may formulate alternative structures to accommodate the different plausible structures that have been put forward. These structures may resemble different theories, or pragmatic origins if no clear best option for potential formulations exists. In some cases it might be possible to encapsulate these alternative structures within a single model where a parameter determines which structure is active, essentially turning structural uncertainty into parameter uncertainty. If capturing these different formulations in a single model becomes impractical, multiple models will have to be formulated. When parameter values are uncertain, the plausible bandwidths for these parameters need to be defined.

The evaluation phase aims at building confidence in the quality of the model by performing tests and evaluating model results. Standard procedures for exploratory modelling imply performing a large number of runs to explore the consequences of the identified deep uncertainties. This provides the basis for testing whether the model or models are fit for purpose. A set of runs may function as a base ensemble, compared to a base run in consolidative modelling. Runs of interest may be selected to identify which combinations of uncertainties cause them, for example with the Scenario Discovery approach. This further increases the understanding of how and why model inputs map to model outcomes.

The policy testing phase aims at testing and analysing the effects of different policies, alone and in combination, on all plausible model behaviour. In the policy testing phase, acknowledging deep uncertainty changes how policies are tested and what evaluative criteria are used. Policy implementation itself may be uncertain. The effects of a policy may thus be uncertain, as well as the moment of implementation. This may be exacerbated if the power of the problem owner in the system is only limited. This can be approached by making important policy variables uncertain. Policy uncertainty is in that case treated just like other types of structural and parametric uncertainty.

The analysis of exploratory model development makes clear that if deep uncertainty is recognised and acknowledged in early phases, it becomes impossible to disregard that uncertainty in later phases. Further, especially the use of multiple models and structural uncertainties may increase the variety of types of model outcomes found in DMDU analyses.

Effects on EMA approaches

The complexity of exploratory models may render the use of some exploratory modelling approaches, including Scenario Discovery, more difficult. There are three reasons for this. First, classifying the types of time series generated by non-linear models is often problematic. Selection of the most relevant runs to see whether these have common origins is thus often not possible. Since publishing the paper on dynamic scenario discovery (i.e., paper 1), significant advances have been made in the field of time series clustering. Future research should investigate the potential of the resulting new time series clustering approaches for classifying behavioural modes.

Second, Scenario Discovery makes use of tools (e.g., the Patient Rule Induction Method) that do not work appropriately for non-linear models. Two directions of future research are (i) the development of algorithms that allow for the use of non-linear models, and (ii) the use of model variables instead of uncertain parameters as the independent variables in existing algorithms for scenario discovery.

Third, exploratory simulation models generally have relatively high numbers of uncertainties. Reducing these numbers is not always possible, which makes new techniques that allow smarter sampling necessary to avoid having to perform unrealistically high numbers of runs. Future research should investigate the potential of adaptive sampling, or alternatively the potential of sensitivity-analysis based screening methods that do account for interaction effects amongst the uncertainties.

Costs of exploratory modelling

The costs of exploratory modelling are significant. Model development and analysis takes significantly more time if multiple models have to be developed. Performing high numbers of runs increases the computational costs. Finally, analysing all outcomes generated may cause an information overload for the analyst, which obscures sharp observations. Benefits of the approach, however, include the increased richness of insights resulting from this analysis and increased opportunities for new insights. For example, in the case of the geopolitical impact of the shale revolution, most stakeholders did not consider the possibility of falling oil prices as plausible before our analysis was presented to them.

Communication and Reception

The research underpinning this thesis received both negative and positive reactions when communicated to stakeholders in policy discussions, in academic policy domains, and in methodological fields. In policy discussions, EMA based policy research is more difficult to quickly comprehend and is often considered relatively expensive. However, the results were often appreciated, especially if new insights were provided. Further, words like uncertainty, complexity, robustness, and resilience resonated with policy makers given their salience in the policy issues they were coping with. In domain specific fields, my research was sometimes seen as unfit by other researchers using different methods. Positive reactions, however, also came forward when domain specific researchers recognised some useful innovations in exploratory modelling. In methodological fields, especially the SD field, the reaction was mixed as well. In part, negative reactions arose from an overly ambitious and perhaps offensive argumentation line in our papers. Next to this, some consolidative modellers just view the inability to unify model structures into a single, best definition as insufficiently rigorous modelling. On the other hand, framing the work as complementary to existing work has led to some good discussions and well received work.

Conclusions

The reflection on model development and use in my dissertation makes clear that while the DMDU field is rapidly expanding, many challenges remain. The first may be to find interest in exploratory model development. This may increase the depth of understanding that arises from exploring the consequences of all – modelled – uncertainties in complex societal challenges. Next to this, the toolset currently used in EMA approaches has limited capabilities with dynamic non-linear simulation models of complex problems.

Exploratory modelling remains expensive. The many positive reactions, however, from policy makers, policy researchers, and methodologists following on sometimes initial negative reactions, do show that the methods discussed in this dissertation have great promise. Continuous reflection on how to build strong narratives based on exploratory models is thus needed to further increase the acceptance and use of these approaches.

Samenvatting

Simulatiemodellen worden steeds vaker gebruikt om de gevolgen van diepe onzekerheid op complexe maatschappelijke uitdagingen te onderzoeken. Het in eigen gedachten doordenken, ook wel ‘Mentale simulatie’ genoemd, van grote maatschappelijke uitdagingen is vaak onmogelijk door de complexiteit, die meestal gekarakteriseerd wordt door de onderlinge samenhang tussen de verschillende systeemelementen van de problemen. Het is daarom noodzakelijk om simulatiemodellen te gebruiken om de menselijke gedachtenvorming te ondersteunen. Deze maatschappelijke uitdagingen zijn vaak ook onderhevig aan diepe onzekerheid, waardoor het bijvoorbeeld onmogelijk is om tot een gezamenlijk begrip te komen van verschillende onderdelen van het systeem, exogene invloeden op het systeem, of zelfs een gezamenlijke probleemdefinitie. Diepe onzekerheid is door Lempert, Popper en Bankes in 2003 gedefinieerd als omstandigheden “waar analisten niet weten, of beslissers het niet eens kunnen worden over, (1) de geschikte conceptuele modellen die de relaties beschrijven tussen de belangrijkste krachten die de toekomst bepalen, (2) waarschijnlijkheidsverdelingen gebruikt om onzekerheid over belangrijke variabelen en parameters in mathematische formuleringen van deze conceptuele modellen uit te drukken, en/of (3) hoe de waarschijnlijkheid van verschillende uitkomsten te waarderen.

Simulatiemodellen kunnen dan gebruikt worden om systematisch de consequenties van verschillende combinaties van aannames over onzekere factoren of eigenschappen van het probleem en het onderliggende systeem te exploreren. Dit type gebruik van simulatiemodellen werd in 1993 geïntroduceerd als ‘*Exploratory Modelling and Analysis*’ (EMA, ‘verkennende modellering en analyse’). Tegenwoordig is deze benadering de belangrijkste pijler on het wetenschappelijke veld met de naam ‘Decision Making under Deep Uncertainty’ (DMDU, ‘besliskunde onder diepe onzekerheid’).

Er is echter nog verbetering mogelijk in de manier waarop in het meeste DMDU-onderzoek met diepe onzekerheid om wordt gegaan. Momenteel wordt namelijk in het meeste DMDU-onderzoek gebruik gemaakt van bestaande modellen. Deze modellen zijn over het algemeen ontwikkeld voor zogenaamd ‘consolidatieve’ (hier: ‘verenigend’) gebruik: de modelleurs hebben gepoogd bestaande kennis te verenigen om tot een enkele, ‘beste’ modelformulering te komen. De meeste van deze modelleurs zijn het er overigens over eens dat geen enkel model een perfecte weergave van de werkelijkheid is en dat ze dus ook niet gevalideerd kunnen worden in de strikte betekenis van het woord, maar desondanks erkennen zij het bestaan van diepe onzekerheid niet in hun modellen. Als men het echter eens is over de invloed van diepe onzekerheid op het beschouwde probleem, dan is het problematisch om verenigende modellen te gebruiken. In die situaties hebben we

daarom modellen nodig die expliciet voor verkennend gebruik zijn ontwikkeld: modellen die alle diepe onzekerheid die potentieel relevant is voor de onderzoeksvraag of -vragen nadrukkelijk meenemen. Tot op heden bestaat er echter relatief weinig aandacht en sturing voor ontwikkeling en gebruik van specifiek verkennende modellen.

Ik doe in deze dissertatie een eerste poging om beslissende keuzes gedurende de ontwikkeling en het gebruik van verkennende modellen te identificeren, of het maken van deze keuzes te ondersteunen. Dit gebeurd op basis van vier casestudy's die als vijf afzonderlijke wetenschappelijke artikelen zijn gepubliceerd. Het eerste en tweede van deze artikelen gaan over de toekomstige beschikbaarheid van koper. Het eerste ("*Dealing with Multiple Models in System Dynamics*") onderzoekt onder welke omstandigheden drie verschillende modellen met drie verschillende perspectieven op hetzelfde probleem vergelijkbaar of juist verschillend gedrag genereren. Het tweede ("*Dynamic scenario discovery under deep uncertainty: The future of copper*") onderzoekt hoe *Scenario Discovery* toegepast kan worden op data met tijdseries. Scenario Discovery is een methode om scenario's te identificeren uit een grote hoeveelheid runs van een computermodel met behulp van algoritmes die automatisch leren ondersteunen. Het derde artikel ("*The geopolitical impact of the shale revolution*") onderzoekt daarna hoe twee exploratieve modellen met verschillende, maar aan elkaar rakende, toepassingsgebieden tegelijkertijd kunnen worden gebruikt. Ik heb dat gedemonstreerd door te kijken naar de mogelijke gevolgen van de grote toename in productie van schaliegas en -olie in de Verenigde Staten op landen die sterk afhankelijk zijn van olie- en gasexport. Het vierde artikel ("*Societal ageing in the Netherlands*") onderzoekt hoe verschillen in probleemperceptie tussen verschillende belanghebbenden meegenomen kunnen worden in tijdens het verkennend modelleren. We hebben dit toegepast op de impact van vergrijzing in Nederland op enerzijds de houdbaarheid van vergrijzingsgerelateerde collectieve uitgaven en anderzijds de wenselijkheid van vergrijzingsbeleid in de ogen van inwoners van Nederland. Het vijfde artikel, ten slotte, ("Simulating endogenous dynamics of intervention-capacity development") onderzoekt hoe omgegaan kan worden met beleidsonzekerheid. We kijken in dit artikel naar de ontwikkeling van interventiecapaciteiten tijdens de West-Afrikaanse Ebola-uitbraak in 2014. Er bestaat geen sterk overkoepelend thema tussen al deze artikelen, aangezien het meeste achterliggende onderzoek in opdracht van verschillende klanten uitgevoerd is, buiten het feit dat in al het onderzoek gebruik gemaakt is van het zelfde modelleerparadigma en de zelfde onzekerheidsmethodologie.

Methode

Ik heb een diepgaand overzicht samengesteld van de keuzes waar een modelleur tegen aan loopt bij ontwikkeling en gebruik van verkennende modellen door kritisch te reflecteren op de casestudies. Dit overzicht bevat vier hoofdcategorieën. De eerste categorie beschouwt verschillende manieren waarop recht gedaan kan worden aan onzekerheid tijdens de ontwikkeling van verkennende modellen. De tweede categorie beschouwt de problemen die nog bestaan tijdens het gebruik van specifieke methoden en technieken voor verkennende analyse, in het bijzonder de analyse van de resultaten van verkennende simulaties. De derde categorie beschouwt de kosten van verkennend modelmatig onderzoek in vergelijking met verenigend onderzoek. In de vierde, laatste categorie beschouw ik ten slotte de communicatie van EMA-onderzoek aan klanten en belanghebbenden.

Ontwikkeling van verkennende modellen

Ik ben tot de conclusie gekomen dat diepe onzekerheid op iedere fase in de modelontwikkeling invloed heeft. Een model beschouw ik in deze dissertatie als een intern consistente set van vergelijkingen, die niet noodzakelijkerwijs is voorzien van waarden voor de parameters. Modelontwikkeling kan worden opgedeeld in vijf verschillende fases: probleemarticulering, modelconceptualisatie, modelformulering, modelevaluatie en beleidsevaluatie.

De probleemarticuleringsfase richt zich op het verwoorden van het centrale te onderzoeken probleem. Hierbij wordt de gevonden probleemformulering gebruikt om te selecteren welke systeemelementen endogeen (binnen de grenzen van het model) of exogeen (buiten de grenzen van het model) gemodelleerd gaan worden en wat de tijdshorizon van het model is. Onzekerheid over de probleemformulering, bijvoorbeeld in het geval van zogenaamde *wicked problems* of *societal messes*, kan leiden tot meervoudige modelbegrenzing: in die gevallen kan het nodig zijn om verschillende modellen te ontwikkelen om recht te doen aan de verscheidenheid aan ideeën over probleem en systeem. Verder kan het nodig zijn om de modelbegrenzing breder te nemen dan gebruikelijk in traditionele benaderingen, als dit verscheidene verklaringen van hetzelfde probleem toestaat. De begrenzing kan ook nauwer genomen worden, als het noodzakelijk is om de responsie van het model op verschillende, goed gevestigde inputscenario's (bijvoorbeeld klimaatscenario's) te testen, of op andere inputscenario's die gebruikt kunnen worden om de weerbaarheid van het systeem te testen.

In de modelconceptualisatiefase probeert de modelleur om de belangrijkste relaties tussen kernvariabelen te identificeren, wat veelal voortbouwt op mentale modellen van belanghebbenden en experts. Onzekerheid speelt in deze fase een rol door het identificeren van de belangrijkste structurele onzekerheden die in het model opgenomen moeten worden. Als in de probleemarticuleringsfase verscheidene modelbegrenzingen zijn geselecteerd, dan dienen vaak ook verscheidene conceptuele modellen te worden ontwikkeld om recht te doen aan iedere modelbegrenzing. De modelleur heeft hierbij, vanzelfsprekend, de taak om de aanwezigheid en locaties van diepe onzekerheden in conceptuele diagrammen te communiceren.

Gedurende de modelformuleringsfase wordt het daadwerkelijke simulatiemodel geformuleerd en geïmplementeerd. De modelleur moet hierbij keuzes maken over hoe de in de conceptualisatie gevonden onzekerheden uit te drukken in model(len). Een modelleur kan, bijvoorbeeld, alternatieve structuren formuleren recht te doen aan de verschillende plausibele structuren die naar voren zijn gekomen. Deze structuren vertegenwoordigen dan alternatieve theorieën, of hebben een meer pragmatische oorsprong als er geen duidelijke beste formulering bestaat. Soms is het mogelijk om deze alternatieve structuren in een enkel model te vatten, waarbij een parameterwaarde bepaald welke structuur actief is, wat feitelijk van een structurele onzekerheid een parametrische onzekerheid maakt. Als het echter onpraktisch of onmogelijk is om deze verschillende formuleringen in een model op te nemen, dan is het wederom nodig om verscheidene modellen te formuleren. Ten slotte dienen ook de bandbreedtes voor parametrische onzekerheden te worden gedefinieerd in deze fase.

De evaluatiefase richt zich op het opbouwen van vertrouwen in de kwaliteit van het model door een aantal tests uit te voeren en modelresultaten te evalueren. De standaardprocedure van verkennend modelleren schrijft voor dat hierbij een groot aantal modelruns wordt uitgevoerd om de gevolgen van de geïdentificeerde diepe onzekerheden te exploreren. Deze runs vormen de basis van tests om te bepalen of het model (of de modellen) al of niet geschikt zijn voor het beoogde doel. Deze set runs kan ook functioneren als ‘base ensemble’, in tegenstelling tot de ‘base case’, die gevormd wordt door een enkele run en die vaak wordt gebruikt in het verenigend modelleren. Verder kunnen interessante runs worden geselecteerd om te identificeren welke combinaties van onzekerheden tot dat specifieke modelgedrag leiden, bijvoorbeeld met de ‘Scenario Discovery’ benadering. Deze benadering leidt zo weer tot toegenomen begrip van de manier waarop het model input naar output vertaald.

De beleidstestfase richt zich op het testen en analyseren van de effecten van verschillende beleidsopties, op zichzelf en in combinatie, op het geheel van plausibel modelgedrag. In deze fase verandert, indien diepe onzekerheid wordt erkent, de manier waarop beleidsopties worden getest en welke criteria gebruikt worden om ze te evalueren. Beleidsimplementatie zelf kan ook onzeker zijn: zowel de effecten van het beleid als het moment van implementatie. Dit idee kan versterkt worden als de macht van de probleemeigenaar in het betreffende systeem beperkt is. De modelleur en analist kan hiermee omgaan door belangrijke beleidsvariabelen ook onzeker te maken. Beleidsonzekerheid wordt in dat geval precies zoals andere typen van structurele of parametrische onzekerheid behandeld.

De analyse van de ontwikkeling van verkennende modellen maakt duidelijk dat als diepe onzekerheid in de eerste fases wordt herkend en erkend, het onmogelijk wordt om diezelfde onzekerheden niet te beschouwen in latere fases. Daarbij geldt dat in het bijzonder het gebruik van verscheidene modellen en structurele onzekerheden de variëteit aan mogelijke modeluitkomsten gevonden in DMDU-analyses aanzienlijk kan doen toenemen.

Effecten op EMA-benaderingen

De complexiteit van verkennende modellen kan het gebruik van sommige benaderingen voor EMA-methodes, zoals Scenario Discovery, bemoeilijken. Daar zijn drie redenen voor. Ten eerste, het identificeren en classificeren van verschillende types van tijdseries die door niet-lineaire modellen gegenereerd worden, kan erg moeilijk zijn. Selectie van de meest relevante runs om te bekijken of deze een gezamenlijke oorsprong hebben, wordt dan dus vaak onmogelijk. Sinds de publicatie van het tweede paper in deze dissertatie over dynamische Scenario Discovery zijn er significante verbeteringen gemaakt op het gebied van het clusteren van tijdsseries. Toekomstig onderzoek zou moeten beschouwen wat het potentieel is van deze verbeteringen voor het classificeren en beoordelen van verschillende types modelgedrag.

Ten tweede maakt Scenario Discovery gebruik van een aantal tools, zoals de Patient Rule Induction Method, die de slecht werken met niet lineaire modellen. Twee richtingen van toekomstig onderzoek op dit vlak zijn dan ook (i) de ontwikkeling van algoritmes die ook met niet lineaire modellen goed werken en (ii) het gebruik van modelvariabelen in plaats van onzekere parameters als onafhankelijke variabelen in bestaande algoritmes voor Scenario Discovery.

Ten derde hebben verkennende simulatiemodellen vaak een relatief hoog aantal onzekerheden. Het is niet altijd mogelijk om dit aantal terug te brengen, wat nieuwe technieken die slimmer bemonsteren van de onzekerheidsruimte nodig maken om te voorkomen dat onrealistisch hoge aantallen runs gedaan moeten worden. Toekomstig onderzoek zou zich moeten richten op het potentieel van adaptief bemonsteren, of, anders, kijken naar het potentieel van methoden die gebaseerd zijn op basis van gevoeligheidsanalyses die ook interactie-effecten tussen onzekerheden in beschouwing nemen.

Kosten van exploratief modelleren

Verkennend modelleren brengt aanzienlijke kosten met zich mee. Modelontwikkeling en -analyse nemen aanzienlijk meer tijd in beslag als verscheidene modellen moeten worden ontwikkeld. Het grote aan runs laat de kosten op het gebied van rekentijd ook toenemen. Ten slotte kan het grote aantal gegevens wat is gegenereerd tot een overaanbod aan informatie leiden voor de analist, wat scherpe observaties vertroebeld. De voordelen van de benadering bevatten echter de toegenomen rijkdom aan inzichten die voortkomen uit deze analyse en toegenomen mogelijkheden voor nieuwe inzichten. Een voorbeeld hiervan was het onderzoek naar de geopolitieke impact van de schalierevolutie, waarbij de meeste betrokkenen geen rekening hielden met de mogelijkheid van sterk dalende olieprijzen voordat onze analyse aan ze was gepresenteerd.

Communicatie en ontvangst

Het onderzoek dat aan deze dissertatie ten grondslag ligt heeft zowel positieve als negatieve reacties ontvangen tijdens de communicatie aan belanghebbenden in beleidsdiscussies, in academische beleidsdomeinen en in methodologische vakgebieden. In de beleidsdiscussies viel het op dat EMA-gebaseerd onderzoek vaak moeilijker is om snel te bevatten en daarbij te vaak als relatief duur wordt beschouwd. Deze resultaten werden echter vaak gewaardeerd, zeker als nieuwe inzichten werden gepresenteerd. Daarbij resoneren woorden als onzekerheid, complexiteit, robuustheid en weerbaarheid bij beleidsmakers door de urgentie van deze begrippen in de beleidsproblemen waar de beleidsmakers mee te maken hebben. In domeinspecifieke vakgebieden werd mijn onderzoek door onderzoekers die zelf andere methodes gebruikten soms ongeschikt bevonden. Er waren echter ook positieve reacties vanuit onderzoekers die de meerwaarde en bruikbare innovaties in verkennend modelleren herkenden. In methodologische velden, in het bijzonder in het SD-veld, was de reactie ook gemengd. Voor een deel kwamen negatieve reacties voort uit het gebruik van te onduidelijke, en soms zelfs beledigende, betooglijnen in onze papers. Daarbij komt dat een deel van de verenigende modelleurs het niet kunnen verenigen van modelstructuren in een beste modeldefinitie als niet voldoende grondig modelleren ziet. Aan de andere kant bleek dat door juist te wijzen op het feit dat onze benaderingen aanvullend zijn bij bestaand werk, goede discussies tot stand kwamen en het werk goed ontvangen werd.

Conclusies

De reflectie op modelontwikkeling en -gebruik in mijn dissertatie maakt duidelijk dat veel uitdagingen blijven bestaan in het zich snel uitbreidende DMDU-veld. De eerste kan gevonden worden in meer aandacht in de ontwikkeling van verkennende modellen. Betere, echt verkennende modellen kunnen namelijk leiden tot een verdiept begrip voorkomend uit het verkennen van de gevolgen van alle (gemodelleerde) onzekerheden in complexe maatschappelijke uitdagingen. Hier staat wel tegenover dat de methoden en technieken die momenteel gebruikt worden in EMA-benaderingen beperkte mogelijkheden bieden in combinatie met dynamische, niet-lineaire simulatiemodellen van complexe problemen.

Verkennend modelleren blijft duur. De vele positieve reacties van beleidsmakers, beleidsonderzoekers en methodologen die echter volgen op de soms initieel negatievere reactie, tonen echter aan dat de methoden die in deze dissertatie besproken zijn, een grote belofte in zich houden. Continue reflectie op hoe krachtige verhaallijnen opgebouwd kunnen worden die gebaseerd zijn op verkennende modellen is daarom nodig om acceptatie en gebruik van deze methoden verder te laten toenemen.

Preface & acknowledgements

In November 2011, I graduated from the TU Delft as Master of Science (Dutch title: *ingenieur*) on the comparison of three different models about the future of the global copper market. I developed these models due to a semantic misunderstanding between me and my supervisors about using multiple models for exploring potential futures. My supervisors meant multiple parameterisations of the same set of equations, while I interpreted multiple models as different sets of equations (i.e., pragmatically, different model files, regardless of their parameterisation). My misunderstanding met enthusiasm, however, and led to the idea of pursuing a PhD about the development and use of models for exploring the consequences of uncertainty.

As it was not possible to hire me as an “internal” PhD candidate at the Policy Analysis section of TU Delft’s faculty of Technology, Policy and Management, my first supervisor during my master thesis, Erik Pruyt, was able to convince Michel Rademaker and Erik Frinking from the Hague Centre for Strategic Studies (HCSS) to hire me for half a year. During that time, we would try to acquire funds allowing me to pursue my PhD on a policy questions relevant to them and the university. It turned out to be impossible to acquire this money directly, but the half year experiment was interesting enough for HCSS to hire me for a longer time. While working at HCSS, I was able to spend 20% of my time on my scientific work related to my PhD, and most of the other time on HCSS projects. At the same time, I became an external PhD candidate at the TU Delft, where a workplace was made available on Erik Pruyt’s room, and with Erik Pruyt as my daily supervisor for my PhD and Wil Thissen as my promotor. After roughly four and a half years at HCSS, I “returned” to the TU Delft as lecturer and researcher, where I could finish my PhD still as external PhD candidate.

The combination of my work at HCSS and TU Delft made it possible for me to work on actual policy problems, where both clients and colleagues expected results – more or less on time – and I was able to engage in discussions about my work with stakeholders and other policy researchers from HCSS and other organisations. Combining the policy analysis praxis and the scientific reflection on this work was not a trivial task, but an experience I would not have liked to miss.

During my PhD, I received support from many people. First of all, I would like to thank Michel Rademaker and Erik Frinking, together with HCSS founder Rob de Wijk, for giving me the chance to work as the only “quant” at HCSS. You showed an entrepreneurial mind in hiring me, and I am grateful for everything I was able to learn during my time at HCSS. I would also like to thank my colleagues and all interns at HCSS, and especially: the ladies from the back office for all their support; Teun van Dongen and Peter Wijninga, my “roommates” on the Lange Voorhout, for the good

discussions in our room; Marjolein de Ridder, for the opportunity we had to start understanding each other despite our different backgrounds, but with our shared interests of resource scarcity; Josh Polchar, for the many lengthy discussions we had on practically everything; and Sijbren de Jong and Eline Chivot, my fellow internship coordinators (or as someone called it: “internship givers”), for the wonderful time and successful collaboration we had.

In academia, I would like to thank my committee members for the feedback during the last phase of my PhD. I would like to thank Wil Thissen for the continuous support and constructive criticism during the whole period of my PhD. Wil, it is an honour to be promoted by you. Further, Erik Pruyt, for the time we spend in your (later, a bit, our) room, your visions on the advancements of our field, and the fact that without you, I would never have been a proper independent researcher. Jan Kwakkel, you became an informal extra supervisor. I learned a lot from you, especially regarding my writing skills and understanding of academia. Els van Daalen, for being my new “roommate” at the TU and your support. Others, including Warren Walker and Bartel van de Walle and all other Policy Analysis colleagues, for the support, feedback, and enjoyable work atmosphere.

Finally, at home, my family for their support during my PhD. My father, unfortunately deceased, for his inquisitive mind and inspiration during the first fourteen years of my life. I know you would have liked my work. My mother, for supporting me during all these years, especially when I needed it during my years at Applied Physics. Max and Frits, my sons, for the happiness you generate in our home. And last but most definitely not least, Dieneke: thank you. Thank you for helping me finish my work, and starting our family. Without you, I would still be studying at the TU Delft.

1 Introduction

Dealing with societal or grand challenges is enduringly difficult for decision makers. Therefore, supranational institutions, national governments, large funding organisations, and others have identified these challenges as important topics for future-oriented research. Examples include the European Union's Horizon 2020 research funding program of which one section focusses explicitly on societal challenges (European Commission, 2015), the Global Grand Challenges program of the Bill & Melinda Gates foundation on global health and developmental programs (Bill & Melinda Gates Foundation, 2017), and the Grand Challenges programs of both the Canadian and the US governments (Government of Canada, 2017; USAID, 2017). Many of these programs focus on the development of technological solutions for these challenges, but some, especially the Societal Challenges in the Horizon 2020 program, also focus on policy analysis research.

Two linking characteristics of societal challenges are: (1) strong interconnectedness of different parts, which is generally referred to as complexity[1], and (2) high uncertainty about either the structure of the system or future values of key variables due to long time horizons. Next to these characteristics, there may be disagreement between stakeholders on how to evaluate outcomes. Complexity and uncertainty have led to the use of various frameworks and methodologies, including Integrated Water Resource Management (Medema, McIntosh, & Jeffrey, 2008), various qualitative and quantitative scenario approaches (Söderholm, Hildingsson, Johansson, Khan, & Wilhelmsson, 2011), and expert consultations (Hoorens et al., 2013). Next to these approaches, different simulation modelling approaches are being used to address the complexity of societal challenges (e.g., Fiddaman, 2002; Forrester, 2007; Heppenstall, Crooks, Batty, & See, 2012; Kwakkel & Pruyt, 2015). Recently, there has been a renewed interest in computer simulation driven scenario development to deal with the combination of complexity and uncertainty (Lempert, Popper, & Bankes, 2003).

Early attempts to deal with the interconnectedness of elements of societal challenges include the use of physical simulation models for dealing with large scale hydraulics and water management, followed by the development of computer simulation models. In the Netherlands, researchers started constructing scale models of Dutch waterworks in the 1920s based on earlier examples in Germany, Austria, and Sweden (Steenhuis, Voerman, Noyens, & Emmerik, 2015). A bit later, around the time that the first computer was developed, the first – analogue – computer simulation model was developed during the Manhattan Project ("Computing and the Manhattan Project," 2014). The use of computer simulation models continuously

[1] For definitions of important concepts and terms used in this dissertation, please consult the glossary in the appendix.

increased after World War II, in particular for numerically solving large sets of differential equations, for which the early computers were particularly well suited. The development of the General System Theory (Bertalanffy, 1950, 1968) exemplified the application of differential equations – and consequentially also computer simulation – to complex social phenomena.

Concurrently and in connection with the development of computer simulation models, scenario techniques were developed to deal with uncertainty in planning for the long-term future, the other key characteristic of societal challenges. This development started in the fifties, but gained speed in the late sixties of the twentieth century. Multiple scenario 'schools' exist, the most notable of which are Intuitive Logics, La Prospective, and Probabilistic Modified Trends (Bradfield, Wright, Burt, Cairns, & Van der Heijden, 2005). These schools all offer approaches for developing multiple scenarios based on potential values of key variables or indicators of the system at hand. In intuitive logics, these scenarios are communicated qualitatively in the form of narratives, and no probability of occurrence is specified. Other schools develop quantitative scenarios, and do try to indicate their probability of occurrence (Bradfield et al., 2005). Scenario development using Exploratory Modelling and Analysis (EMA) was more recently developed. EMA combines the non-probabilistic nature of intuitive logics scenarios with computational, systematic sampling over the bandwidths of uncertainties influencing the system (Bankes, 1993; Lempert et al., 2003). Selections of individual computer simulation runs, or of similar runs are in this approach used to develop plausible future scenarios.

Broadly speaking, there are three ways to deal with uncertainty in developing strategies for the future, and consequentially in simulation models: ignore, reduce, or embrace. For modelling, ignore implies that the scope of the model is chosen so narrowly that most uncertainties lie outside the boundary, or parameter values and model formulations are implicitly assumed to be known. Needless to say, this way of modelling is not related to any scenario school.

In the second approach, reduce implies 'consolidative modelling' (Bankes, 1993): the modeller will try to find a best solution for potential structural uncertainty by bringing together existing knowledge. A scenario exploration can be done with the model by manually changing input parameters, or sampling over the input parameters and their related probability distributions. Generally, a base case or business as usual scenario is used to depict the most likely future without changes in policy. An early use of the base case concept can be found in La Prospective and Probabilistic Modified Trends schools. Scenario development using consolidative models can thus be related to these schools, although I am not familiar with any formal connection between these scenario schools and consolidative modelling.

In the third approach, embrace, the modeller assumes that it is not possible to reduce at least part of the uncertainty during model development: it is impossible to measure or reason yourself out of this uncertainty. Lempert et al. (2003) refer to this type of uncertainty as ‘deep uncertainty’, and define it as ‘*where analysts do not know, or the parties to a decision cannot agree on, (1) the appropriate conceptual models that describe the relationships among the key driving forces that will shape the long-term future, (2) the probability distributions used to represent uncertainty about key variables and parameters in the mathematical representations of these conceptual models, and/or (3) how to value the desirability of alternative outcomes*’. Deep uncertainty is incorporated in models by specifying bandwidths or different options for uncertain model elements. This last method for dealing with uncertainty makes use of large numbers of computer simulation runs to explore the consequences of combinations of plausible realisations of uncertainties. Selections of these runs can then be used as scenarios, combined with a text which explains what happens in these runs, why it happens (i.e., the combination of values for the various uncertainties causing this type of run behaviour), and why it is plausible that this might happen. It was, therefore, again the advance of computing power that made this development possible. This way of dealing with uncertainty can be seen as a quantitative development of the intuitive logics scenario school.

Exploratory modelling literature has been rapidly expanding since the turn of the century. Most of this literature uses consolidative models to explore the consequences of parametric deep uncertainty. Deep uncertainty, however, can be manifest in other model attributes than just parameters (Kwakkel, Walker, & Marchau, 2010). Therefore, only exploring the consequences of parametric uncertainties reduces the potential bandwidth of scenarios and futures developed with the models. Using these scenarios and futures to test policies for their robustness (i.e., whether a policy functions desirably in all plausible futures) can lead to wrong judgements. These issues make development of ‘true’ exploratory models relevant, as uncertainty should have a more profound impact on exploratory model development and use.

1.1 Research approach

The goal of this dissertation is to illustrate and analyse how deep uncertainty can affect model development and use. I will start with illustrating how deep uncertainty can be handled by presenting a number of cases in which exploratory models have been developed and used. I will then reflect on what lessons can be learned from these cases with respect to model development and use.

As illustration of how deep uncertainty affects model development and use, this dissertation first presents a number of cases in which deep uncertainty is acknowledged in model-based approaches to grand societal challenges. The cases consider the future availability of copper (later referred to as the ‘copper’ case), societal ageing in the Netherlands (‘ageing’), the geopolitical impact of the US’ shale revolution (‘shale’), and the 2014 Ebola outbreak in West-Africa (‘Ebola’). Most of this research was performed during my employment at The Hague Centre for Strategic Studies (HCSS), a Dutch think tank that generally operates on the interplay of international relations and security. As a consequence, the modelling work on the cases has an applied rather than academic character. Further, there is no strong connection between most cases apart from the methods they apply, but as the cases were performed consecutively, each case builds on the experience from earlier cases.

Most research underlying these cases was partly or completely funded by clients, which has as a consequence that no systematic build-up can be found in the cases. Although they are in part related (i.e., three of the four cases presented deal with resource pricing issues), and the way in which the cases have been approached should show at least some increasing experience, the choice for the cases was purely pragmatic and depended on the issues potential clients brought to the table. The cases are presented in five papers, which are presented as chapters in this dissertation. By themselves, these papers – especially the papers about ageing, shale, and Ebola – contribute to the deep uncertainty field by presenting applications of EMA approaches on real world problems for real world clients.

In addition to the cases, I will provide a reflection on the applied methodological improvements, possible further methodological refinements, the costs and added value of the methodology used, and the lessons learned from analysing and communicating the results based on the five cases. The applied and possible further methodological improvements are on two levels. First, I reflect systematically on the consequences of deep uncertainty for model development, as all research presented in this dissertation made use of models specifically developed for exploratory use. I will do this by comparing choices that can be made during exploratory modelling with choices made during consolidative modelling as described in consolidative literature (Sterman, 2000). Second, I look at the consequences of the use of non-linear models in combination with Scenario Discovery. Next, I will address the issue of the impact this research had on policy discussions, also on two different levels. First, I will consider costs (e.g., computational, time resources of analysts, potential of information overload) and benefits (e.g., conclusions that could not have been drawn without the exploratory approach). Second, I will discuss how the research was communicated to and received by stakeholders, and researchers from the domain fields of the applications and methodological fields. I will do this as a reflective practitioner.

1.2 System Dynamics combined with Exploratory Modelling

All research presented in this dissertation makes use of System Dynamics (SD) modelling, which was originally conceived in the 1950s by Jay W. Forrester (Forrester, 1961). The choice for SD was formally motivated by the characteristics of the systems underlying the research: feedback effects, delays, and accumulations. However, besides the formal motivation, the choice for SD was also pragmatic: SD is the modelling discipline I have exclusively used in my own work since my master thesis, which makes me most skilful in this kind of modelling.

I will start with explaining some of the characteristics of SD models, and continue with explaining the suitability of SD for deep uncertainty research.

1.1.1. Technicalities of System Dynamics

SD models are in essence large sets of integral equations which are numerically solved, and can be depicted via SD-specific diagrammatic conventions. Crucial elements in SD models are stocks or levels, which are connected to flows. The behaviour of a stock over time is mathematically defined as an integral equation:

$$s(t) = s(t_0) + \int_{t_0}^{t} f(t) - g(t)\mathrm{d}t, \qquad \text{Eq. 1.1}$$

where $s(t)$ is a stock at time t, $s(t_0)$ the initial value of this stock, $f(t)$ an inflow and $g(t)$ an outflow. Besides stocks and flows, SD also knows auxiliary variables and constants. As the interconnected set of integral equations can become too large to analytically solve, SD languages like Vensim (Ventana Systems, 2010) use numerical integration methods like Euler (Euler, 1768) and Runge-Kutta 4 (Kutta, 1901; Runge, 1895). Fig. 1.1 shows a simple SD model, where constant $s_0 = s(t_0)$, flow $f(t)$ is a function of constant c_1 and $s(t)$, auxiliary $a(t)$ is a function of constant c_2 and $s(t)$, and flow $g(t)$ is a function of constant c_3 and auxiliary $a(t)$.

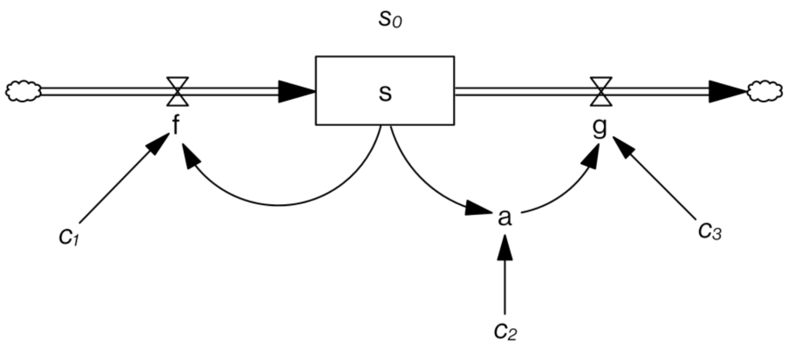


Fig. 1.1. Simple stock-flow structure in SD diagrammatic conventions

The SD modelling elements of stocks, flows, and auxiliary variables made SD a suitable choice for the cases presented in this dissertation. Stocks are used to model accumulation and memory in systems, like stocks of resources in use in the copper case, different population cohorts in the ageing and state stability cases, and the accumulation of stocks of oil in the energy cases. Flows like resource extraction or migration are used to change stock levels. Combined, stocks and flows allow increasing understanding about, for example, how resource prices depend on delays in the development of extraction capacity in reaction to changing demand, or the speed with which the Ebola virus spreads depending on the development of intervention capacities.

Finally, stocks, flows, and auxiliary variables allow feedbacks – crucial and central in SD thinking – to be modelled. Two different types of feedback loops are generally distinguished: balancing and reinforcing. In balancing loops, increase of a variable in a loop will, ceteris paribus, lead in time to a decrease of the same variable. Growth will, therefore, be balanced. For example, if demand for a resource increases, the market price of that resource will increase. As a consequence of the price increase, the demand will decrease in time. In reinforcing loops, increase of a variable in a loop will, ceteris paribus, lead in time to a further increase of the same variable. For example, if a population grows due to a relatively high fertility rate, this will lead to more people in fertile age, who will get more children, which will lead to exponential growth of the population. Together, accumulations, flows, and feedbacks represent the complexity and non-linearity of these systems.

1.1.2.Exploratory SD: ESDMA

Exploratory use of SD models has several major advantages. First, SD models are – on presently available computers – relatively fast to simulate. This makes performing a large number of runs to explore the consequences of uncertainty feasible without significantly reducing the possibilities for much needed iterations. Further, it is relatively easy to incorporate structural uncertainties. Next to this, SD allows to focus on dynamics over time instead of just end-states, which creates a rich picture of plausible system evolutions. Finally, due to the causal structure of SD models, they allow the analyst to look at the structure of the system to explain the different types of behaviour found. As a consequence of these advantages, SD models have been used extensively for exploring the consequences of uncertainty. The specific combination of the EMA approach and SD is also referred to as ‘exploratory system dynamics modelling and analysis’ (ESMDA) (Kwakkel & Pruyt, 2015). This fits the SD philosophy of focussing on behavioural patterns and using the model structure with its underlying assumptions to understand the system’s behaviour, rather than using the model predictively.

Besides my own work, examples of combined EMA and SD use include the use of the Wonderland model (Lempert et al., 2003) on global sustainable development, work on the 2009 Influenza A(H1N1)v pandemic (Pruyt & Hamarat, 2010), terrorism (Pruyt & Kwakkel, 2014), residential energy use (Yücel, 2013), and uncertainties in the Dutch natural gas sector (e.g., Eker & van Daalen, 2015).

1.3 Dissertation setup

In this dissertation, I will first present the set of papers (Table 1) which all demonstrate the use of SD models for researching the consequences of deep uncertainty in societal challenges. The build-up of this dissertation is as follows. The ‘copper’ research (Chapter 2, and 4) was partly performed as part of a master thesis project at the TU Delft, and partly funded by the Platform Material Scarcity. The research on the geopolitical impact of the shale revolution (‘shale’, Chapter 4) and societal ageing (‘ageing’, Chapter 5) was performed for the joint research program ‘Strategy & Change’, which was funded by the Dutch Organisation for Applied Research (TNO) and executed together with HCSS. The ‘Ebola’ research (Chapter 6) was partly funded under the 2015 Strategic Monitor program for the Dutch Ministry of Defence. Each paper is concisely introduced to illustrate its relevance and key contributions to this dissertation.

Table 1. Overview of case chapters with their case names, respective journal publications, original policy reports, and type of uncertainty research.

Ch.	Title	Case	Journal publication	Policy report
2	Dealing with Multiple Models in System Dynamics	Copper	Auping, Pruyt, and Kwakkel (2014)	Auping, Pruyt, Kwakkel, and Rademaker (2012)
3	Dynamic scenario discovery under deep uncertainty	Copper	Kwakkel, Auping, and Pruyt (2013)	Auping, Pruyt, Kwakkel, and Rademaker (2012)
4	The geopolitical impact of the shale revolution	Shale	Auping, Pruyt, De Jong, and Kwakkel (2016)	De Jong, Auping, and Govers (2014)
5	Societal Ageing in the Netherlands	Ageing	Auping, Pruyt, and Kwakkel (2015)	Willem L. Auping, Erik Pruyt, Jan H. Kwakkel, Govert Gijsbers, and Michel Rademaker (2012)
6	Simulating Endogenous Dynamics of Intervention-Capacity Deployment	Ebola	Auping, Pruyt, and Kwakkel (2017)	Auping, Frinking, Coelho, and Ginn (2015)

Finally, in Chapter 7 I will synthesise and reflect on this work. I do this with regard to both the methodology used and the policy contributions produced in my research. I will start with systematically assessing the impact of deep uncertainty on the methodology from the experiences I gained during research of the case studies for both model development and Scenario Discovery. First, I will assess model development in situations of deep uncertainty, and the impact of these models on

Scenario Discovery. Examples from the papers will be used to illustrate how deep uncertainty changes the problem articulation, model conceptualisation, specification, and evaluation (including ensuring model quality), and policy testing. Second, I will assess the impacts the non-linearity of models has on Scenario Discovery. Examples from the papers will be used to explain how I dealt with these issues, and which approaches are promising or not. I will finish by discussing the policy contributions the research underlying these papers produced. I will do this by first looking into the costs and benefits of such approaches for policy analysis. This entails issues like the time needed by the research team for model development and interpretation of the results, and the issue of information overload during the analysis of the generated results. I will then discuss how I communicated my conclusions to, and the reception of my conclusions by, communities of policy makers and stakeholders, and domain or methodology oriented scientists.

References

Auping, W. L., Frinking, E., Coelho, R., & Ginn, J. (2015). *Epidemics and Strategic Timing: Integrating Epidemic Risk and Intervention Capabilities in the Case of the 2014 West-Africa Ebola Outbreak*. Retrieved from The Hague: http://www.hcss.nl/reports/download/187/3089/

Auping, W. L., Pruyt, E., De Jong, S., & Kwakkel, J. H. (2016). The geopolitical impact of the shale revolution: Exploring consequences on energy prices and rentier states. *Energy Policy, 98*(2016), 390-399. doi:10.1016/j.enpol.2016.08.032

Auping, W. L., Pruyt, E., & Kwakkel, J. H. (2014). Dealing with Multiple Models in System Dynamics: Perspectives on the Future of Copper. *International Journal of System Dynamics Applications, 3*(4), 17-35. doi:10.4018/ijsda.2014100102

Auping, W. L., Pruyt, E., & Kwakkel, J. H. (2015). Societal Ageing in the Netherlands: A Robust System Dynamics Approach. *Systems Research and Behavioral Science, 32*(4), 485-501. doi:10.1002/sres.2340

Auping, W. L., Pruyt, E., & Kwakkel, J. H. (2017). Simulating endogenous dynamics of intervention-capacity deployment: Ebola outbreak in Liberia. *International Journal of Systems Science: Operations & Logistics, 4*(1), 53-67. doi:10.1080/23302674.2015.1128576

Auping, W. L., Pruyt, E., Kwakkel, J. H., Gijsbers, G., & Rademaker, M. (2012). *Aging: Uncertainties and Solutions. An exploration of scenarios, problems and solutions with respect to the affordability of societal aging* (1363). Retrieved from The Hague: http://static.hcss.nl/files/uploads/1363.pdf

Auping, W. L., Pruyt, E., Kwakkel, J. H., & Rademaker, M. (2012). *Futures for Copper*. Retrieved from The Hague: http://www.hcss.nl/reports/download/97/1242/

Bankes, S. C. (1993). Exploratory Modeling for Policy Analysis. *Operations Research, 41*(3), 435-449. doi:10.1287/opre.41.3.435

Bertalanffy, L. v. (1950). An Outline of General System Theory. *British Journal for the Philosophy of Science, 1*, 114-129. Retrieved from http://www.isnature.org/Events/2009/Summer/r/Bertalanffy1950-GST_Outline_SELECT.pdf

Bertalanffy, L. v. (1968). *General Systems Theory*. New York: George Braziller.

Bill & Melinda Gates Foundation. (2017). About Grand Challenges. Retrieved from https://gcgh.grandchallenges.org/about

Bradfield, R., Wright, G., Burt, G., Cairns, G., & Van der Heijden, K. (2005). The origins and evolution of scenario techniques in long range business planning. *Futures, 37*(8), 795-812. doi:10.1016/j.futures.2005.01.003

Computing and the Manhattan Project. (2014). Retrieved from http://www.atomicheritage.org/history/computing-and-manhattan-project

De Jong, S., Auping, W. L., & Govers, J. (2014). *The Geopolitics of Shale Gas*. Retrieved from The Hague: http://static.hcss.nl/files/uploads/2180.pdf

Eker, S., & van Daalen, E. (2015). A model-based analysis of biomethane production in the Netherlands and the effectiveness of the subsidization policy under uncertainty. *Energy Policy, 82*, 178-196. doi:10.1016/j.enpol.2015.03.019

Euler, L. (1768). *Institutionum calculi integralis*. Petropoli: Impensis Academiae Imperialis Scientiarum.

European Commission. (2015, 4/12/2015). Societal Challenges. *Horizon 2020.* Retrieved from https://ec.europa.eu/programmes/horizon2020/en/h2020-section/societal-challenges

Fiddaman, T. S. (2002). Exploring policy options with a behavioral climate-economy model. *System Dynamics Review, 18*(2), 243-267. doi:10.1002/sdr.241

Forrester, J. W. (1961). *Industrial Dynamics*. Cambridge, MA: MIT Press.

Forrester, J. W. (2007). System Dynamics—the next fifty years. *System Dynamics Review, 23*(2/3), 359-370. doi:10.1002/sdr.381

Government of Canada. (2017). About Grand Challenges Canada. Retrieved from http://www.grandchallenges.ca/who-we-are/

Heppenstall, A. J., Crooks, A. T., Batty, M., & See, L. M. (2012). Reflections and Conclusions: Geographical Models to Address Grand Challenges. In A. J. Heppenstall, A. T. Crooks, L. M. See, & M. Batty (Eds.), *Agent-Based Models of Geographical Systems* (pp. 739-747). Dordrecht: Springer Netherlands.

Hoorens, S., Guerin, B., Ghez, J., Schweppenstedde, D., Hellgren, T., Horvath, V., . . . Kobzar, S. (2013). *Europe's Societal Challenges: An analysis of global societal trends to 2030 and their impact on the EU*. Retrieved from Brussels: https://www.rand.org/pubs/research_reports/RR479.html

Kutta, M. W. (1901). Beitrag zur näherungsweisen Integration totaler Differentialgleichungen. *Z. Math. Phys., 46*, 435-453.

Kwakkel, J. H., Auping, W. L., & Pruyt, E. (2013). Dynamic scenario discovery under deep uncertainty: The future of copper. *Technological Forecasting & Social Change, 80*(4), 789-800. doi:10.1016/j.techfore.2012.09.012

Kwakkel, J. H., & Pruyt, E. (2015). Using System Dynamics for Grand Challenges: The ESDMA Approach. *Systems Research and Behavioral Science, 32*(3), 358-375. doi:10.1002/sres.2225

Kwakkel, J. H., Walker, W. E., & Marchau, V. A. W. J. (2010). Classifying and communicating uncertainties in model-based policy analysis. *International Journal of Technology, Policy and Management, 10*(4), 299-315. doi:10.1504/IJTPM.2010.036918

Lempert, R. J., Popper, S. W., & Bankes, S. C. (2003). *Shaping the Next One Hundred Years: New Methods for Quantitative, Long-Term Policy Analysis* (MR-1626). Retrieved from Santa Monica, CA, USA: http://www.rand.org/pubs/monograph_reports/2007/MR1626.pdf

Medema, W., McIntosh, B. S., & Jeffrey, P. J. (2008). From Premise to Practice: a Critical Assessment of Integrated Water Resources Management and Adaptive Management Approaches in the Water Sector. *Ecology and Society, 13*(2), 29 [online]. doi:10.5751/ES-02611-130229

Pruyt, E., & Hamarat, C. (2010, 2010). *The Influenza A(H1N1)v Pandemic: An Exploratory System Dynamics Approach.* Paper presented at the 28th International Conference of the System Dynamics Society.

Pruyt, E., & Kwakkel, J. H. (2014). Radicalization under Deep Uncertainty: A Multi-Model Exploration of Activism, Extremism, and Terrorism. *System Dynamics Review, 30*(1-2), 1-28. doi:10.1002/sdr.1510

Runge, C. D. T. (1895). Über die numerische Auflösung von Differentialgleichungen. *Math. Ann., 46*, 167-178.

Söderholm, P., Hildingsson, R., Johansson, B., Khan, J., & Wilhelmsson, F. (2011). Governing the transition to low-carbon futures: A critical survey of energy scenarios for 2050. *Futures, 43*(10), 1105-1116. doi:10.1016/j.futures.2011.07.009

Steenhuis, M., Voerman, L., Noyens, M., & Emmerik, J. (2015). *Waterloopkundig laboratorium. Cultuurhistorische duiding, ruimtelijke analyse en essentiële principes*. Retrieved from Schiedam, The Netherlands: http://publications.deltares.nl/WeL1929.pdf

Sterman, J. D. (2000). *Business Dynamics: Systems Thinking and Modeling for a Complex World*. New York: McGraw.

USAID. (2017). Grand Challenges for Development. Retrieved from https://www.usaid.gov/grandchallenges

Ventana Systems. (2010). *Vensim Reference Manual*. Harvard, MA: Ventana Systems.

Yücel, G. (2013). Extent of inertia caused by the existing building stock against an energy transition in the Netherlands. *Energy and Buildings, 56*, 134-145. doi:10.1016/j.enbuild.2012.09.022

2 Dealing with Multiple Models

When dealing with complex and uncertain problem situations, multiple perspectives often exist on how to conceptualise the system. If these perspectives are distinctive enough, it is impossible to unite them into a single model. While it is possible to choose only one of these potential perspectives, it was recognised decades ago that the choice itself may have consequences for the simulated model behaviour (Cole in Meadows, Richardson, & Bruckmann, 1982, p. 205). Therefore, the existence of different perspectives can be seen as a form of deep uncertainty (Lempert et al., 2003), making it potentially useful to represent the different perspectives on a system in a set of models, in support of developing robust policies (Lempert, Groves, Popper, & Bankes, 2006). As choices regarding the different perspectives to be included in modelling are made in the beginning of the research, they will affect many of the choices made in later phases of model development.

In this paper, I explore to what extent models with different perspectives generate both similar and different behaviour. This paper presents a comparison of the behaviour of models that are structurally different due to the perspectives on the copper system (global top-down, global bottom-up, and regional top-down) they represent. We do this by a model-by-model comparison of the most similar and the most different behaviour, while the input space that is shared across the three models is kept the same.

The existence of multiple perspectives on problem and/or system will be discussed in the first section of the synthesis chapter, where I discuss the problem articulation and its consequences for the selection of the models' scope.

This paper was originally published as:

Auping, W. L., Pruyt, E., & Kwakkel, J. H. (2014). Dealing with Multiple Models in System Dynamics: Perspectives on the Future of Copper. *International Journal of System Dynamics Applications, 3*(4), 17-35. doi:10.4018/ijsda.2014100102

Dealing with Multiple Models in System Dynamics: Perspectives on the Future of Copper

Willem L. Auping, The Hague Centre for Strategic Studies, The Hague, Netherlands

Erik Pruyt, Delft University of Technology, Delft, Netherlands

Jan H. Kwakkel, Delft University of Technology, Delft, Netherlands

Keywords: Copper, Metal scarcity, Multi-perspective, Multi-Model, Deep uncertainty, Aggregation, System Dynamics

Abstract

This paper introduces an approach to compare simulation runs from multiple SD simulation models. Three dynamic hypotheses regarding the uncertain long-term copper availability are introduced and used to illustrate the new approach. They correspond to three different perspectives on the copper system (global top-down, global bottom-up, and regional top-down). Although each of these models allows to generate a wealth of behavioural patterns, we focus in this paper on the differences in trajectories caused by different models for identical values and settings of shared parameters and assumptions, not on differences in behavioural patterns caused by each of the models. Hence, differences in trajectories between the three models are identified, quantified, and classified based on a quantified measure of difference. For these models, small differences between the trajectories are only found in stable runs, while the alternative perspectives are largely responsible for medium to large differences. Hence, it is concluded that multiple dynamic hypotheses may have to be modelled when dealing with uncertain issues.

1. Introduction

More than 30 years after the first explicit calls for multi-model work in System Dynamics (SD), there are only a few examples of multi-model SD work (Kwakkel et al., 2013; Moorlag, Auping, & Pruyt, 2014; Moxnes, 2005; Pruyt & Kwakkel, 2014), and there are hardly any techniques and tools available for performing multi-model SD analyses. To contribute to filling this gap, this paper proposes a first method to compare runs from multiple simulation models regarding the same issue or system. Using this method to compare runs generated with three different SD models of the global copper system, we test here whether differences in trajectories generated with different models were primarily due to model uncertainty or to parametric uncertainty, and hence, whether a multi-model approach is needed in the first place. The three alternative copper models presented here correspond to different perspectives regarding the copper system. That is, they were developed from a global Bottom-up perspective, a Regional top-down perspective, and – what is mostly used in SD – a global Top-down perspective. Although the pair-wise comparison method proposed and illustrated in this paper does not offer a rationale for judging which perspective or model is more valid, it could offer a rationale for using multiple perspectives, and hence, multiple models, for policy analytical purposes. This method might help practitioners decide whether alternative competing hypotheses might have to be considered during policy analysis. Hence, this paper adopts an exploratory research agenda to test the efficacy of multi-model SD analysis in the copper industry, and does not purport to provide specific actionable solutions with regard to copper scarcity.

There is a long tradition of modelling resource depletion and scarcity in SD. The limits to growth study (Meadows, Meadows, Randers, & Behrens, 1972) is probably the most well-known example. Many SD studies combine geological, technological, and economic aspects of mineral depletion (Davidsen, Sterman, & Richardson, 1987; Kwakkel & Pruyt, 2015; Pruyt, 2010; Sterman & Richardson, 1985; Sterman, Richardson, & Davidsen, 1988; Van Vuuren, Strengers, & De Vries, 1999). Other SD studies focus on specific metals, like the platinum group metals (Alonso, Field, & Kirchain, 2008) or magnesium (Urbance, Field, Kirchain, Roth, & Clark, 2002), and are mostly linked to specific metal uses, such as electronics (Alonso et al., 2008) or the automotive industry (Urbance et al., 2002). Copper markets and their interaction with aluminium markets have also been studied by several system dynamicists (Auping, 2011; Ballmer, 1960; Schlager, 1961).

Copper is the most common of the geochemically scarce elements (Gordon, Koopmans, Nordhaus, & Skinner, 1987, p. 2). Hence, the case of potential copper scarcity is often used as an example in studies examining metal scarcity, for example in Ayres, Ayres, and Råde (2002); Gerst and Graedel (2008); Gómez, Guzmán, and Tilton (2007); Gordon, Bertram, and Graedel (2006); Kapur (2006); Nassar et al. (2012); Ruhrberg (2006) Although most recent criticality reports mainly focus on minor metals, like lithium (Angerer, Marscheider-Weidemann, Wendl, & Wietschel, 2009) and the rare earth metals (European Commission, 2011), copper is included in most criticality assessments too.

In spite of all these studies, copper scarcity is still an important and actual research area: Annual copper demand is continuously growing (ICSG, 2010b), while copper prices are historically high (LME, 2011). There seem to be two causes for recent high prices: the growing demand for minerals and metals in rapidly developing economies like China and India (European Commission, 2011) and the growing demand for minerals and metals as a result of energy transitions (Kleijn & van der Voet, 2010). Given the substantial economic impact of high prices, which could be seen as (temporary) economic scarcity, more research related to the potential influences of changes in regional demand and changes in demand related to particular uses seems to be needed.

2. Modelling the uncertain copper system

In spite of the fact that the structure of the copper system is well-documented, it is also deeply uncertain. That is, different perspectives on copper demand –from top-down to bottom-up and from global to regional– are described in the literature (Gordon, Bertram, & Graedel, 2007; Meadows et al., 1982, pp. 205, 274-275; Tilton & Lagos, 2007). The 'top-down perspective' assumes copper demand is determined by the size of the population and the wealth per capita. In the 'bottom-up perspective, copper demand is determined by different uses and their autonomous development. The 'global perspective' assumes there is a free global copper market where demand of those who are prepared to pay more is satisfied first. The 'regional perspective' assumes that copper markets are not free and global, and that copper demand is fulfilled first and foremost in regions with sufficient supply.

Over thirty years ago, Cole already argued in Meadows et al. (1982, p. 205) that "[w]hether a 'top-down' or 'bottom-up' approach is chosen […] may affect the results[, for s]imple recursive calculation of global or regional aggregates broken down by sector often gives surprisingly different results from systematically building up the global or regional aggregates from the sector or subsector levels". If modelling different perspectives indeed leads to different behavioural patterns, which possibly expand the set of plausible long term scenarios of the copper system, then different perspectives may have to be modelled, explored and used (Pruyt, 2014). The hypothesis that different models of the copper system generate different behavioural pattern for the same settings and sets of parameter values –and hence, that a multi-model approach is needed if multiple equally valuable perspectives exist– will be tested in this paper by comparing runs generated with three different models of the copper system over the intersection of their input spaces (i.e., with identical settings and values for shared variables and parameters).

Table 1. Matrix of the copper models given the uses and regions perspectives.

<table>
<tr><td colspan="2" rowspan="2">Dimension</td><td colspan="2">Regions</td></tr>
<tr><td>1</td><td>3</td></tr>
<tr><td rowspan="2">Uses</td><td>1</td><td>Top-down</td><td>Regional</td></tr>
<tr><td>6</td><td>Bottom-up</td><td>Complete</td></tr>
</table>

The three models used to test this hypothesis are a global top-down model, a global bottom-up model, and a regional top-down model. The matrix in Table 1 shows they can be seen as the result of crossing two sets of perspectives, on the one hand perspectives with regard to copper uses (Angerer, Mohring, Marscheider-Weidemann, & Wietschel, 2010) and on the other hand perspectives with regard to how global/regional mining, refining and consumption of copper is (ICSG, 2010b), as well as the number of regions and uses dealt with in these models. However, the fourth possible model, combining regional and bottom-up disaggregations, is not considered in this study, as the combination led to a unmanageable amount of uncertain input parameters (300+), making a thorough uncertainty analysis impossible.

Table 2. Major uncertainties in the copper system

Uncertainty	Type of uncertainty	Description
Capacity development	Model uncertainty	The capacity for (deep sea) mines, smelters and refineries
Demand development	Model uncertainty	The intrinsic demand for copper, i.e. the demand without effects due to price and substitution
Economic growth	(Dynamic) parametric uncertainty	The growth of the GDP globally or regionally
Ore grade development	Model uncertainty	The ore grade declines with mining of copper
Price of energy development	(Dynamic) parametric uncertainty	The price of the energy needed for copper production
Prices of substitutes development	(Dynamic) parametric uncertainty	The development of the price for substitutes for copper use
Resources/resource base	Model uncertainty	What amount of copper is ultimately recoverable from the earth's crust

Other important uncertainties related to the copper system that are included in these models are the development of ore grades, energy prices, prices of substitutes, economic growth, infrastructure and capacities, and the resource base. Table 2 specifies how these uncertainties are dealt with. Some of these uncertainties are in turn composed of other uncertain elements, for example demand development from a top-down perspective is calculated from global population scenarios (UNPD, 2011), economic development, and the relation between copper demand and GDP per capita (Wouters & Bol, 2009, p. 18).

3. The Models

We will now introduce the model structures of the Top-down, Bottom-up and Regional models, starting with the smallest model, which is the Top-down model. The main differences with the Bottom-up and Regional models, which can be seen as expanded and sub-scripted versions of the Top-Down model, are then briefly explained in relation to the Top-down model. As the sub-scripted variables, combined with alternative structures for demand (in case of the Bottom-up model), and inter-regional transport (in the case of the Regional model), lead to an extension of available feedbacks, the models each link to a different perspective, and therefore constitute different dynamic hypotheses. Full models and the associated model documentations will be provided upon request.

3.1. The Top-down Model

In the Top-down copper model, the intrinsic demand is calculated by looking at the development of the world population, the GDP per capita and the effect of GDP on copper demand. Figure 1 shows three of the most important balancing feedback loops of this model (i.e., a supply loop, a demand loop, and a substitution loop).

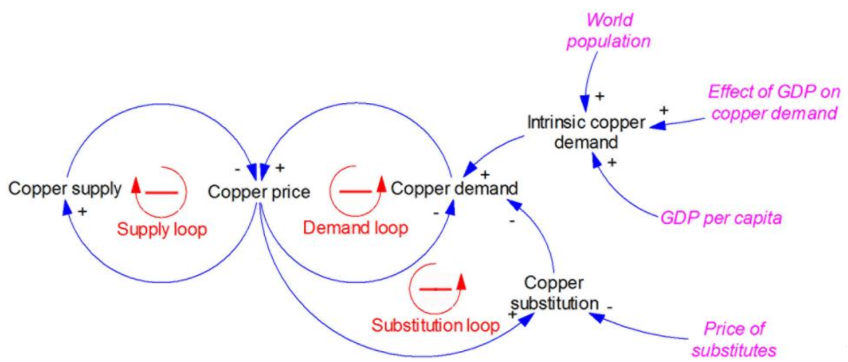


Figure 1. High level CLD of the Top-down copper model. Major uncertainties in italics.

3.1.1. Copper stocks

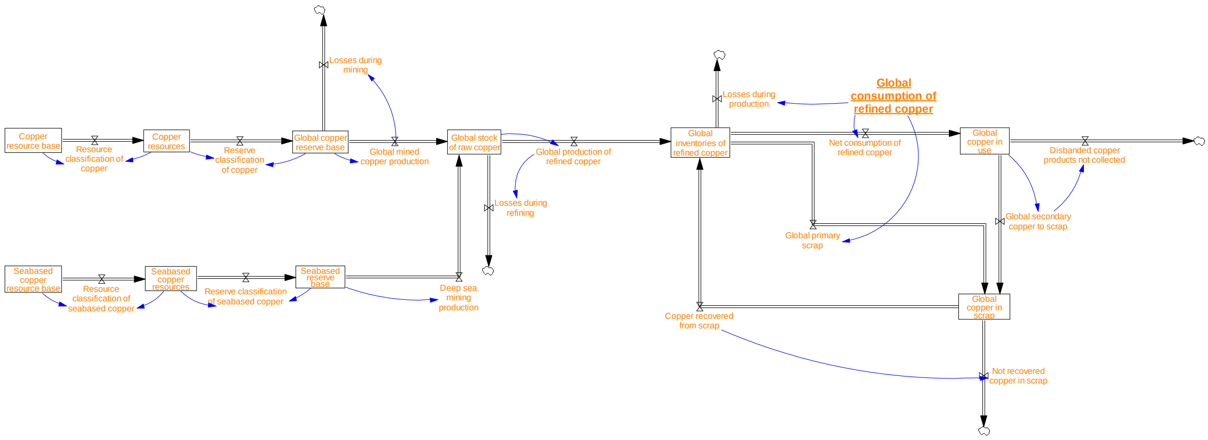


Figure 2. Supply chain of copper in use in the copper stocks sub-model

Real world copper stocks and flows are modelled here by means of a stock flow structure (see Figure 2) linking resource base, resources, reserve base, mining and refining, global consumption and copper use (ICSG, 2010b; Lossin, 2005). When global copper in use reaches the end of its lifetime, it is partially collected via a global secondary copper to scrap flow and recycled as copper recovered from scrap.

<table>
<tr><td rowspan="3"></td><td colspan="3">Identified resources</td><td colspan="2">Undiscovered resources</td></tr>
<tr><td colspan="2">Demonstrated</td><td rowspan="2">Inferred</td><td colspan="2">Probability range</td></tr>
<tr><td>Measured</td><td>Indicated</td><td>Hypothetical</td><td>Speculative</td></tr>
<tr><td>Economic</td><td colspan="2" rowspan="3">Reserve Base</td><td rowspan="3">Inferred Reserve Base</td><td colspan="2" rowspan="3">Resources</td></tr>
<tr><td>Marginally economic</td></tr>
<tr><td>Sub-Economic</td></tr>
<tr><td>Other occurrences</td><td colspan="5">Resource base</td></tr>
</table>

Figure 3. Relation between reserves and resources. Based on the McKelvey Box (McKelvey, 1973)

In literature about copper scarcity, there is a polemic around the total recoverable amount of copper. The relevance of the resource base for the availability of copper in relation to the development of the ore grade of copper (Gordon et al., 2007; Tilton & Lagos, 2007). In this research, we assume –following Tilton and Lagos– that both the price of copper in terms of the amount of energy needed to mine copper as well as the price of energy ultimately define how much copper could be mined. This approach can also be referred to as Tiltons opportunity costs paradigm (Tilton, 1996).

The structure of resources and reserves largely follows the McKelvey classification (McKelvey, 1973), displayed in Figure 3, although some simplifications have been made with respect to JORC classification rules (JORC, 2004). That is, we do not distinguish between reserve base and reserves. The difference between reserve base and reserves relates to the cut-off ore grade, which is the lowest ore grade that can be mined at a particular price. Not differentiating between them allows modelling the system without an explicit cut-off grade for copper ore. The relation between resource base and resources is furthermore reduced to the economic relationship. We also assume that deep sea mining could develop if the marginal costs of copper are higher than the marginal costs of deep sea copper.

The amount of copper mined or refined depends on the capacities from the mine, smelting and refinery capacity sub-model and possibly by a forecast of the copper demand. The global consumption of refined copper is mainly determined by the total demand for copper and the availability of copper, which largely depends on the global inventories of refined copper which in turn corresponds to real inventories of copper traders (Giyose, Manjra, Magoro, & Warren, 2011). The inflow into this stock consists of primary copper (global production of refined copper) and secondary copper (copper recovered from scrap) flows. The Recycling Input Rate (RIR) is calculated by dividing the latter variable by the sum of both. The amount of copper recovered from scrap contains both primary scrap and secondary scrap, where primary scrap originates from the production process of copper products, and secondary scrap from End of Life (EOL) products. We assume that primary scrap is fully recovered, albeit delayed. The secondary scrap on the other hand depends on the collection rate of copper products and the recycling efficiency score. We assume the latter variable depends on the average ore grade of land based copper, the copper grade in EOL goods and its variance.

3.1.2. Mine, smelting, and refinery capacity

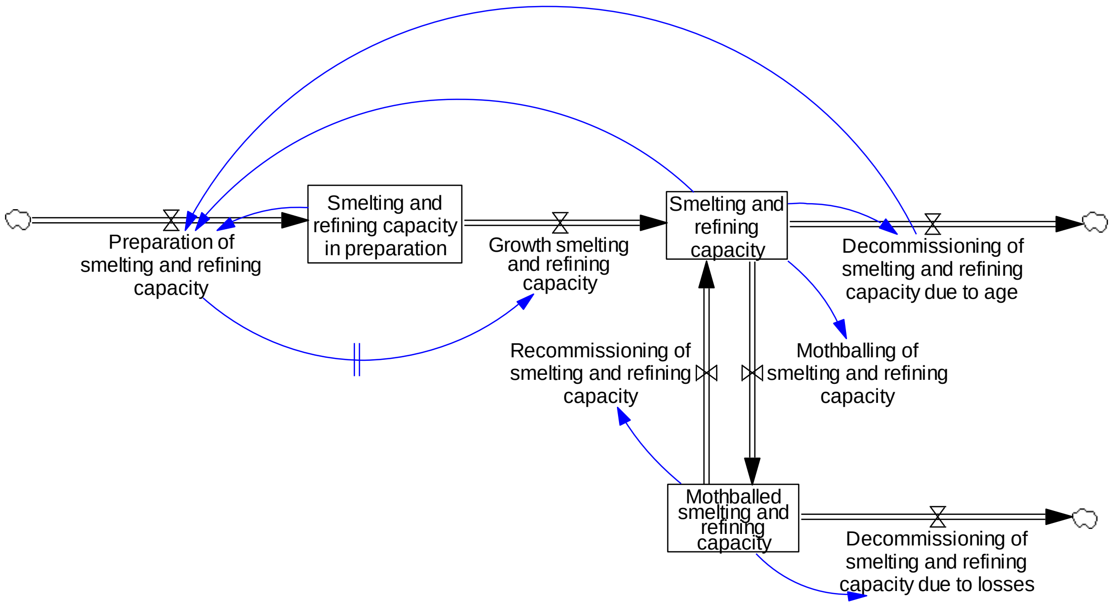


Figure 4. The basic stock-flow structure of mining, smelting, and refinery capacity as used in the copper models

The ‘Mine, smelting, and refinery capacity’ sub-model contains the structures that determine growth and decline of copper mining capacity (Figure 4), of smelting and refining capacity, and of deep sea mining capacity. A similar three-stock structure was used for all three capacities allowing for temporary and permanent closure (Abdel Sabour & Poulin, 2010). Furthermore, learning effects in the mature copper market are assumed to be negligible compared to cost increases due to declining ore grades.

Recycling capacity is not modelled separately, since the data (ICSG, 2010a, 2011) shows that smelter capacity and refinery capacity are actually used for recycling purposes too. The smelting and refinery capacity is therefore larger than the world copper mining capacity. The potential recycling input rate is used in the ‘copper stocks’ sub-model to divide the refinery capacity between flows of primary (mined) and secondary (recycled) copper.

3.1.3. Copper demand

The total demand for copper is influenced by the intrinsic global demand for copper, the total availability of copper, and the relation between copper and aluminium prices which we assume is representative for all copper substitutes. It is assumed that this effect can change the demand in the short term as well as in the long term through accumulation of the effect over a longer period of time. The sum of the short term and long term effects define the maximum decrease or increase in demand. Total demand integrates intrinsic demand, price and substitution effects. The intrinsic demand depends on the average global GDP per capita and the copper use related to GDP. The GDP itself is modelled by five different growth scenarios. Four different United Nations scenarios are used for the world population (UNPD, 2011). Four distinct lookup functions are used for the copper use related to GDP (Wouters & Bol, 2009, p. 18).

Substitution of copper demand takes place when the ‘real copper price’ (i.e., the copper price corrected for inflation, see Svedberg & Tilton, 2006) –which equals the marginal costs of copper when intrinsic global copper demand and total availability of copper balance– is such that it is cheaper to use the substitute than the substituted metal. For this reason, we used a substitution threshold, which takes into account the weight of aluminium required for replacing copper. The substitution effect, which is uncertain, could be amplified or attenuated. Different uses are modelled by using different threshold values (Gordon et al., 1987, pp. 66, 67), although similar price developments for the different substitutes are assumed. A two-stock substitution structure keeps track of the amount of demand that is substituted.

3.1.4. Economics of copper

Finally, in the ‘economics of copper’ sub-model, the marginal costs of copper and marginal costs deep of sea copper are calculated by taking the decline in ore grades for both copper origins into account, which lead to higher energy demand for copper beneficiation, as well as developments in the energy price. Potential profits, both short-term and long-term, are calculated using the marginal costs and either the forecasted copper price or the actual copper price in the model. These potential profit calculations cause changes to mine capacities, and smelter and refinery capacities. Both marginal costs are calculated by looking at the cumulative mined copper and an ore grade which corresponds to that amount. For any particular ore grade, a certain amount of energy is needed, which, together with some other cost factors, determines the marginal costs. The marginal costs are compared to the copper price by either a forecasted value, calculated with the first and second order derivatives, or current costs and prices. Potential profits have both short term and long term effects, similarly to the effects on demand explained above, on the development of new capacities.

3.2. The Bottom-Up Model

In the Bottom-up model, intrinsic demand is calculated by looking at the (quasi) autonomous increase in demand for separate uses of copper (Figure 5).

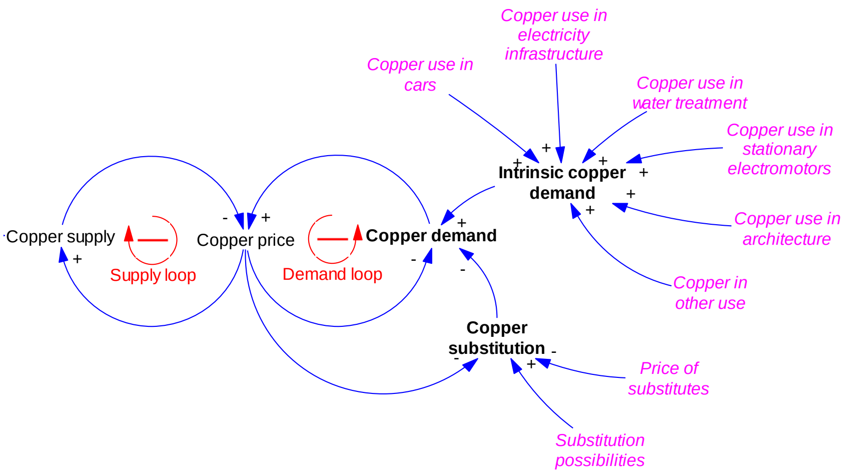


Figure 5. High level CLD of the Bottom-up copper model. Major uncertainties in italics, subscripted sub-models bold.

3.2.1. Differences with the Top-down model

In the bottom-up demand additional structures are used to include six copper uses (Figure 5). All variables related to the use of copper are therefore subscripted, just as the total demand for copper and the flows to and from it, and the substitution threshold values. Here, substitution depends on substitution possibilities for different uses and the price of substitutes, and different uses have different substitution thresholds. The different uses in the bottom-up demand sub-model thus develop autonomously, and, combined, form the intrinsic copper demand.

3.2.2. Bottom-up demand

Following Angerer et al. (2010), the model includes copper use in the automotive sector, in electricity infrastructure, in water treatment, in stationary electro motors, in architecture, and in other uses. Especially the automotive sector and infrastructure are strongly linked to the transition towards sustainable energy. We used the “dominance” and “pluralism” scenarios of electric vehicles for the automotive industry, developed by the Fraunhofer ISI, for the automotive sector (Angerer et al., 2010, pp. 18, 19). The amount of copper per vehicle depends heavily on the degree to which the vehicle has electric propulsion. For the development of the electricity infrastructure a scenario related to the development of decentralised sustainable energy sources presented by Kleijn and van der Voet (2010) was used.

3.3. The Regional Model

The Regional model, with a top-down approach for the intrinsic demand pays particular attention to the geopolitical side of the copper system. The regions used here have primarily resource dependant boundaries. Region 1 is money rich (Europe, N-America, Oceania and Japan), while region 2 is population rich (Asia without the CIS and Asean-10), and region 3 is resource rich (Africa, S-America, and Asean-10).

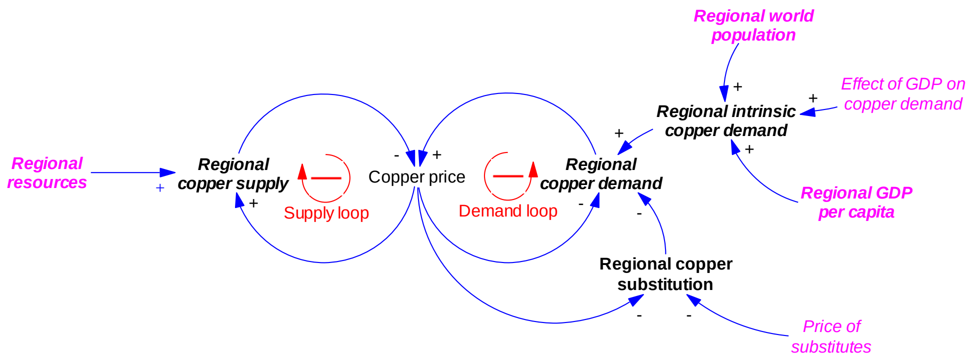


Figure 6. High level CLD of the Regional copper model. Major uncertainties are in italics, regional sub-models are bold.

3.3.1. Differences with the Top-down model

Following variables, including supply variables, are subscripted (see Figure 6): all capacities, demand variables (including the economic situation and the population scenarios), marginal costs, and variables related to the economics of copper. This model also includes import and export flows for different copper fabricates.

Deep sea mining is more difficult to regionalise, especially when deep sea mining takes place in international waters (McKelvey, 1980; United, 1982). In the Regional model this issue is solved by means of regionalised mining concessions. The reserve base is then part of the regionalised mining concession. And the regional ‘preference’ to develop deep sea mining depends on the regional GDP per capita.

3.3.2. Copper transport

Import and export of copper products are modelled in the copper transport sub-model. The regional surplus and deficit for raw copper, the regional surplus and deficit for copper scrap, and the regional surplus and deficit for refined copper are calculated there. Regional surpluses are exported to regions with deficits. Both the exports and imports are allocated proportionate to the regional GDP per capita: export is allocated from regions with a lower GDP per capita and import to regions with a higher GDP per capita.

4. Differences in Scenarios/Runs between Pairs of Models

Differences in behaviour between SD models or model runs for a given outcome may be explained by differences in parameter values and/or structures. The differences in behavioural patterns between two structurally different models could thus be minimized by using the same parameter values for runs that are to be compared. Using TU Delft's open source EMA Workbench software[2] to perform multi-model SD simulation under deep uncertainty, this can be achieved by defining uncertain run settings for all models via the intersections of the model parameter sets.

In the approach presented here, the absolute difference for each time step between two runs from different models is then summed over all time steps and divided by the number of time steps to calculate the average difference. We define the model intersection as the intersection of the input spaces of two models. In set theory, the intersection I of two sets X and Y is defined as the part of the sets that is an element of both sets, hence $I = X \cap Y$. The relative complement C_X of set X in set Y is the part of X that is not part of Y, hence $C_X = X \backslash Y$. The difference D in behaviour between two models with parameter sets A and B can thus be explained both from the complement of the inputs C_A and C_B and the structural differences between the models. The differences for each time step t are defined here, for a given outcome, as:

$$d_{XY,i}(t) = \frac{f_X\left(t, I_{XY,i}, C_{X,i}\right) - f_Y\left(t, I_{XY,i}, C_{Y,i}\right)}{f_X\left(t, I_{XY,i}, C_{X,i}\right)}.$$

Hence, the average absolute difference D can defined, with n as the number of time steps, as:

$$D = \left|\overline{d_{XY,\imath}(t)}\right| = \sum_{t=0}^{t=n} \frac{\left|d_{XY,i}(t)\right|}{n}.$$

Since there are three models in this study, there are also three comparisons to be made: Bottom-up – Regional (Figure 7), Bottom-up – Top-down (Figure 8), and Regional – Top-down (Figure 9). The differences between the runs of these pairs of models are calculated for the outcome *real copper price*. The *real copper price* is an important and volatile performance indicator of periodic imbalance between supply and demand due to delays in the system, and thus of economic scarcity. Although physical scarcity may not be an imminent threat, economic scarcity may well be; copper prices have continued to be high during the last years of economic slow-down, while copper price were high during the decade before which was characterised by globally strong economic growth. The real copper price is also one of the most visible and influential factors in the real world copper system. Hence, we chose the real copper price as key performance indicator in this analysis.

[2] http://simulation.tbm.tudelft.nl/ema-workbench/download.html

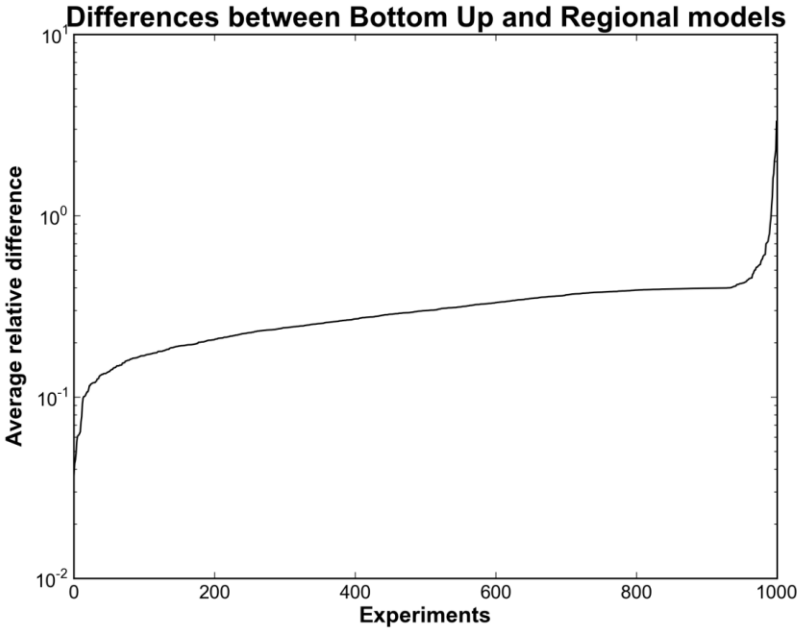


Figure 7. The differences between runs for the Bottom-up - Regional intersection. The differences are ordered from small (left) to large (right).

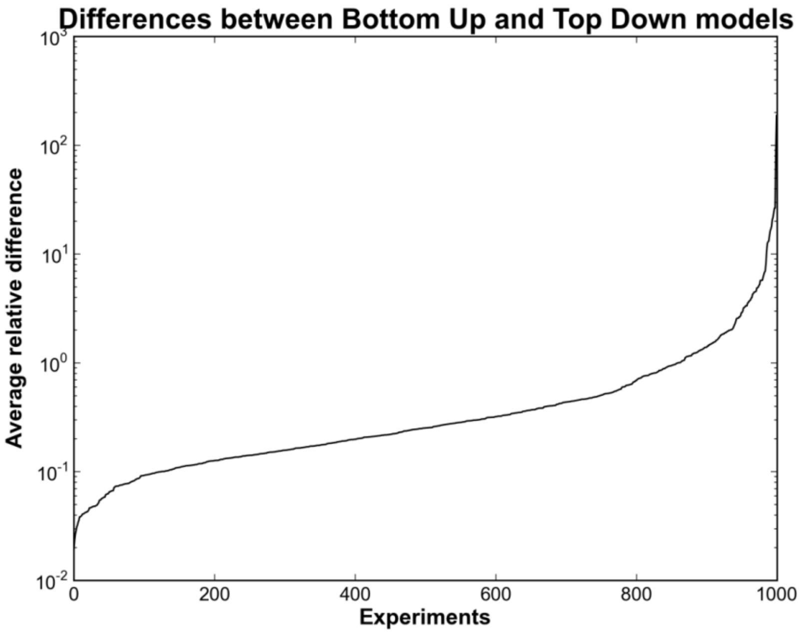


Figure 8. The differences between runs for the Bottom-up – Top-down intersection. The differences are ordered from small (left) to large (right).

The differences for all runs of all intersections in Figure 7, Figure 8 and Figure 9 show that not a single pair of runs shows exactly the same behaviour (i.e., less than 1% difference), while the majority of pairs of runs show average differences of more than 10%. Comparing the Regional and Top-down models, more than 90% of the runs have average differences larger than 100%. Since higher differences can be explained by very low values for the copper price in one of the models, frequently occurring low prices in the Regional model explain part of the high values. This can be explained by the specific structure of the Regional model. In a regionalised copper system, the market is assumed to be hindered by the fact that regional demand is first satisfied by the supply. As a consequence, the response between changes in demand and supply is lagged, which results in lower or higher copper prices than seen in the other models. The difference with the copper price seen in the Top-down model is normalised to the Regional copper price, explaining the higher average differences. This is also visible in Figure 12, and may also explain the difference in form of the average differences on the intersection of the Bottom-up and Regional models, compared to the intersection between the Bottom-up and the Top-down models.

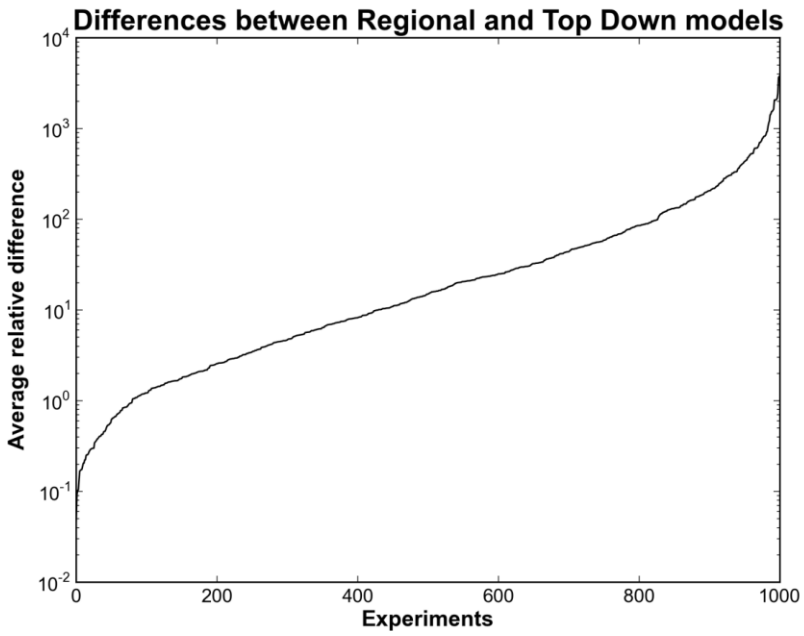


Figure 9. The differences between runs for the Regional – Top-down intersection. The differences are ordered from small (left) to large (right).

Figure 10, Figure 11, and Figure 12 display the trajectories for each pair of models for three pairs of runs, more precisely for the smallest, median and largest differences, in view of visually exploring the differences in trajectories between the pairs of models.

The results for the smallest differences suggest that small differences only occur in situations of stable equilibrium. Only in quasi equilibrium are the trajectories of all pairs of models similar. To confirm this idea, we looked at the Vensim model parameterisations that generated these specific runs. By doing so, we were able to assess to what extent certain behaviour was caused primarily by the specific model parameters, or by the structural difference between the models.

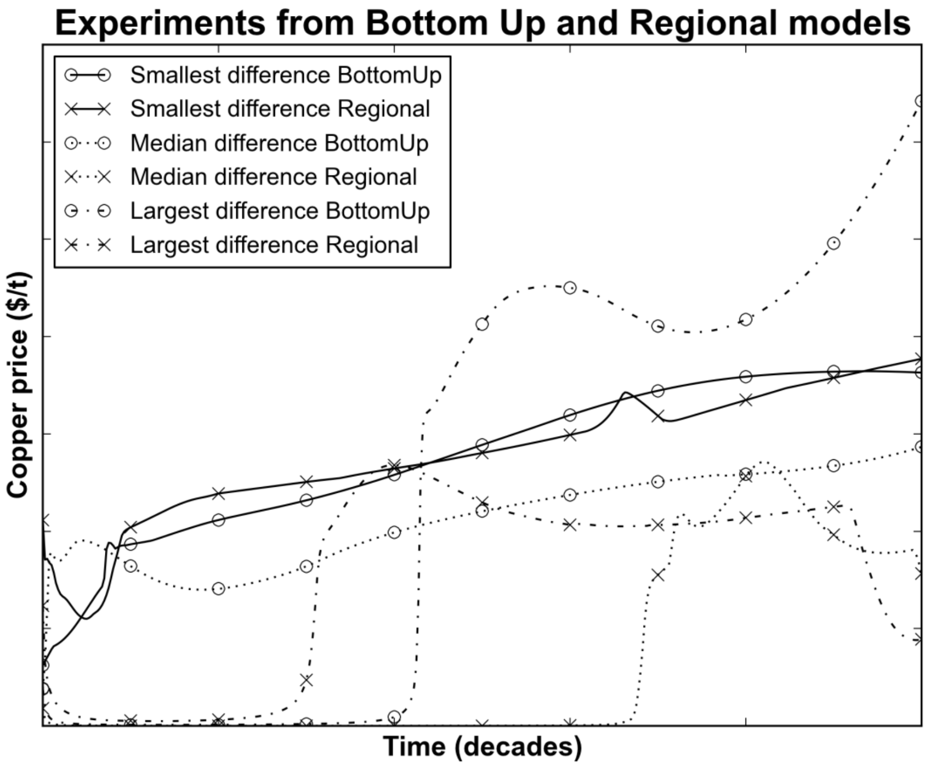


Figure 10. Three runs for the intersection of the Bottom-up and Regional models.

The Bottom-up and Regional runs for the median difference of this intersection show different behavioural patterns (Figure 10). In the Bottom-up model, the price develops slowly after some initial disturbance, while in the Regional model the price stays at a low level for a longer period of time, until at approximately two thirds of the model run time, the price suddenly rises. This difference in behaviour is caused by different developments of the global copper consumption. In this case, the consumption in the Bottom-up model gradually decreases for most of the simulation, whereas in the Regional model, the copper consumption rises exponentially, as may be expected for low copper prices, until sudden scarcity causes a large bust in copper consumption at the first signs of strong price rises. This typical behaviour of the Regional copper model is caused by the slower reaction of the primary production sector due to regionalisation.

The behaviour mode corresponding to the largest difference between the Bottom-up and Regional models are actually quite similar: periods with low copper prices are followed by periods with higher copper prices. The prices in the Bottom-up model are so low that the relative differences are largest. These differences are caused by the differences in copper demand: The Bottom-up model has global copper consumption which is almost three times as small as the global copper consumption in the Regional model for comparable copper supplies. The Bottom-up demand, which is the sum of new copper needed for different uses, is calculated in a very different way than the – regionalised – top-down approach in the Regional model. These differences are thus consequences of the different demand structures in both models and the difference measure used here.

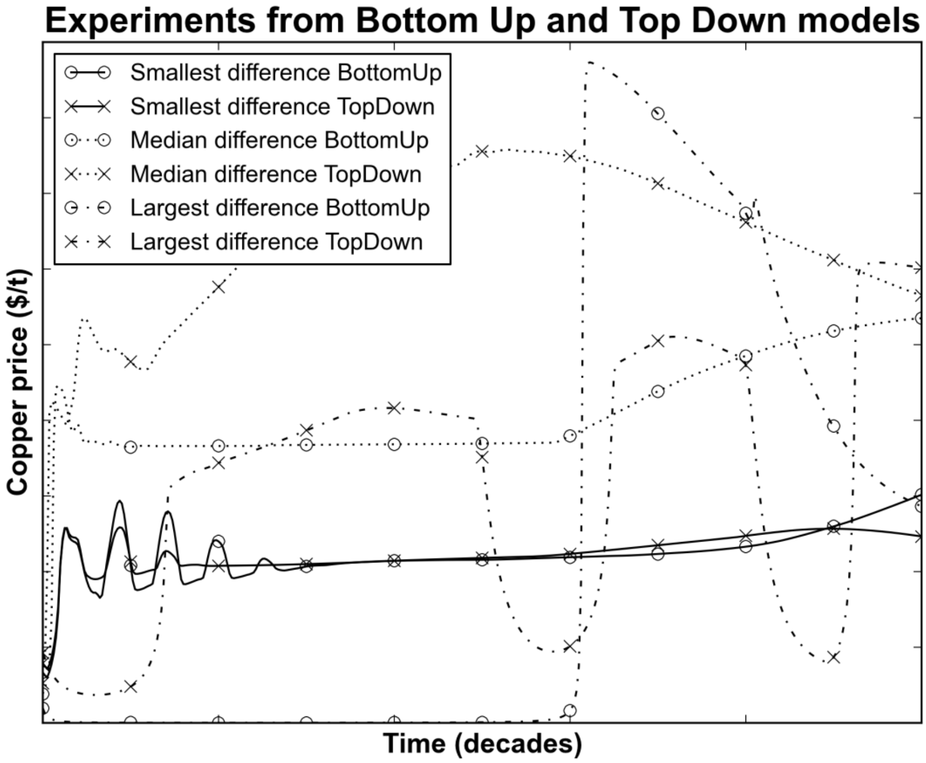


Figure 11. Three runs for the intersection of the Bottom-up and Top-down models.

The behavioural patterns corresponding to the median differences between the Bottom-up and Top-down models are not very different. However, further investigation of these model runs shows that the differences are caused by very different structures. In the Top-down model, the global copper consumption decreases to a very low level over the model run time. The decline in demand is caused by a very strong substitution of copper in that same period. Substitution is also seen in the Bottom-up model, but at a lower rate, since it is spread out over time for different uses, resulting in a slower substitution rate over time.

The largest difference is again explained mainly by the low copper demand compared to the supply in the Bottom-up model. The Top-down model has a higher demand compared to the Bottom-up model, but the imbalance between demand and supply is also present. The smaller difference causes more frequent price oscillations, while the price does not reach a level as low as in the Bottom-up case. Thus, the differences can be explained by the different ways of calculating the copper demand, while for these cases, both models function roughly the same way.

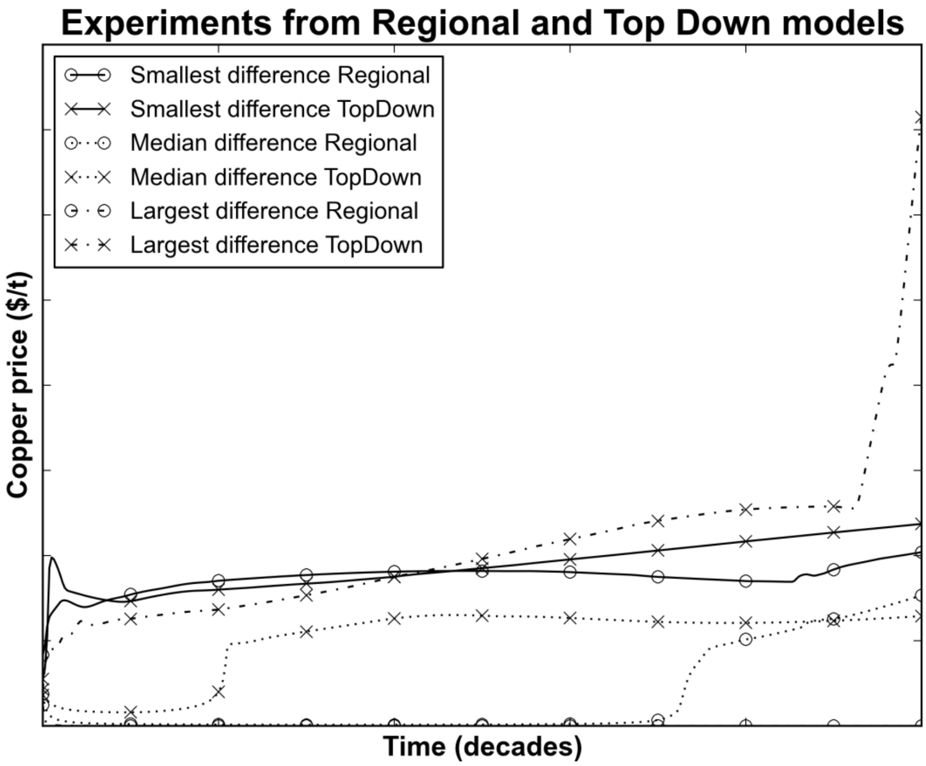


Figure 12. Three runs for the intersection of the Regional and Top-down models.

The median difference of the Regional – Top-down intersection can again be explained by the difference in supply system reaction time. The Regional model has a long period of relatively large supply compared, which is followed by a collapse of the consumption when the supply stocks are outrun by demand. In the Top-down model the same situation exists in the first decade of the model run time. However, due to the different situation this results in a quicker adaptation of demand, followed by an early bust. A period of relative smooth price behaviour follows after the bust.

The same happens with the cases with the largest difference. However, in the Regional case the moment of bust does not come during the run time, while in the Top-down model, the run starts with the bust. In the Top-down the bust is followed by a long period of scarcity, followed by a period with high scarcity at the end of the run time. In short, the supply dominance in the Regional case does not end before the end of the run time, while the Top-down model switches from to another states before the end of the run.

5. Discussion and Conclusions

In this paper, we introduced three different copper models developed from a global top-down perspective, a global bottom-up perspective, and a regional top-down perspective. Performing pair-wise comparisons, we assessed whether differences in trajectories generated with different models were primarily due to model uncertainty or to disjoint parameter uncertainty (i.e., uncertainty pertaining to parameters not shared by both models). The goal was not to replicate the current state of the system or provide forecasts based upon it, but rather to explore the full uncertainty space, that is to generate a plethora of plausible behavioural patterns that could be generated under all sorts of different circumstances. Doing so, we found that the differences in model behaviour are largely explained for these copper models by alternative model structures (i.e., by the difference in model structures). This is not necessarily the case: Moxnes (2005) found for example that, in the case of two fisheries models, different model structures are less important than alternative nonlinear functions. Note, however, that models that differ in terms of one or more important functions are in fact different models (Lane, 1998, 2000).

Although Meadows et al. (1982) argue that examining issues from multiple perspectives is a central principle of the systems view of the world, Sterman (1994, p. 310) notices that “[u]nfortunately, [scientists, professionals and laypeople] do not generate sufficient alternative explanations or consider enough rival hypotheses.” But even if people would be able to generate alternative plausible models, would there be problems related to model choice and model use. Sterman (1994, p. 310) argues with respect to econometric methods that: “In practice the data are too scarce and the plausible alternative specifications too numerous for [models] to discriminate among competing theories [for t]he same data often support wildly divergent models equally well, and conclusions based on such models are not robust.” SD modelling faces similar problems. Choosing between different plausible SD models may be problematic (Moxnes, 2005), or even impossible. And if choosing one perspective over another affects the outcomes of modelling studies and multiple perspectives are plausible, then policy analysis requires alternative perspectives to be included. Hence, testing for differences may be necessary, since it is unknown before simulation and comparison whether different perspectives substantially affect behavioural patterns and policy conclusions. However, the lack of differences between runs drawn from multiple models is not necessarily a valid argument in favour of selecting only one plausible model. Different models may provide alternative explanations (i.e., alternative dynamic hypotheses), for the same patterns of behaviour, which may need to be addressed for designing adaptive robust policies (Hamarat, Kwakkel, & Pruyt, 2013; Hamarat, Kwakkel, Pruyt, & Loonen, 2014). For robust policy design, instead of addressing one causal explanation of undesirable behaviours, all causal explanations of the same undesirable behaviours may need to be addressed.

Alternatively, one may circumvent this ontological-epistemological problem by using existing tools, like TU Delft’s open source EMA workbench, to simultaneously run and analyse alternative models, either over the intersection or over the union of their uncertainty spaces. An ex-ante multi-model approach assumes that alternative plausible models add value by adding additional insights. For example, the Regional model adds the insight that regionalisation of the copper market may lead to slower reaction times of the system, resulting in more extreme behaviour of prices.

It would be good, according to Bremer in Meadows et al. (1982, p. 231), to “know whether the overall behaviour of the model is strongly affected by our choice of one rather than another”, both for choosing a model and for using multiple models. This paper shows that it is possible, despite some remaining technical problems and difficulties, to perform multi-model simulation and investigate whether model behaviour is strongly affected by our choice of levels of aggregation or perspectives.

The scripts[3] developed for this paper can be used to that purpose. Moreover, they may also be useful for informing model choice and for facilitating multi-model simulation and analysis by supporting the identification of differences in model outputs and the search for their causes. Other uses can easily be envisioned. One example would be to assess the added value of alternative models in terms of the new behavioural patterns or old behavioural patterns with new mechanisms that could be added to the already existing set.

A remaining difficulty in multi-model behaviour comparison relates to the fact that different models developed from different perspectives do not necessarily share exactly the same parameters and other inputs, which complicates the identification of the origins of differences in behavioural patterns. Since not all parameters and other inputs for the copper cases were the same, differences in behaviour could not be attributed exclusively to structural differences between the models. However, the example runs in our analysis showed that the differences in behaviour were at least partly caused by the different models. Note that there is room for improvement in terms of the metric used to identify differences between trajectories from pairs of models. Using absolute differences between trajectories, oscillatory behaviours in counter phase may show large differences for similar modes of behaviour. This problem could be addressed with behavioural metrics as explored in Kwakkel, Auping, and Pruyt (2014). Another problem related to the metric used in this paper is that the relative differences are sensitive to the size of the denominator: small denominators cause large relative differences. Comparison based on behavioural patterns, for example using time series classification as in Kwakkel and Pruyt (2015) or Kwakkel et al. (2013) may solve some problems, but not all. New methods, techniques and tools need to be developed to compare models and behaviours and provide deeper analytical insight under deep uncertainty.

References

Abdel Sabour, S. A., & Poulin, R. (2010). Mine expansion decisions under uncertainty. *International Journal of Mining, Reclamation and Environment, 24*(4), 340-349. doi:10.1080/17480931003664991

[3] The Python scripts are available upon request.

Alonso, E., Field, F. R., & Kirchain, R. E. (2008). *A Case Study of the Availability of Platinum Group Metals for Electronics Manufacturers*. Paper presented at the IEEE International Symposium on Electronics and the Environment, San Francisco, CA.

Angerer, G., Marscheider-Weidemann, F., Wendl, M., & Wietschel, M. (2009). *Lithium für Zukunftstechnologien*. Retrieved from Karlsruhe: https://www.isi.fraunhofer.de/content/dam/isi/dokumente/ccn/2009/Lithium_fuer_Zukunftstechnologien.pdf

Angerer, G., Mohring, A., Marscheider-Weidemann, F., & Wietschel, M. (2010). *Kupfer für Zukunftstechnologien*. Retrieved from Karlsruhe: https://www.isi.fraunhofer.de/content/dam/isi/dokumente/cce/2010/Bericht_FSEM_Cu.pdf

Auping, W. L. (2011). *The Uncertain Future of Copper: An Exploratory System Dynamics Model and Analysis of the global copper system in the next 40 years*. Retrieved from Delft: http://simulation.tbm.tudelft.nl/RESEARCH/Auping2011small.pdf

Ayres, R. U., Ayres, L. W., & Råde, I. (2002). *The Life Cycle of Copper, Its Co-Products and Byproducts* (Vol. 13): Springer Netherlands.

Ballmer, R. W. (1960). *Copper Market Fluctuations. An Industrial Dynamics Study.* (MSc), MIT, Cambridge, MA.

Davidsen, P., Sterman, J. D., & Richardson, G. P. (1987). *A Petroleum Life Cycle Model for the United States with Endogenous Technology, Exploration, Recovery, and Demand*. Paper presented at the 1987 International Conference of the System Dynamics Society, China.

European Commission. (2011). Tackling the challenges in commodity markets and on raw materials.

Gerst, M. D., & Graedel, T. E. (2008). In-Use Stocks of Metals: Status and Implications. *Environmental Science & Technology, 42*(19), 7038-7045. doi:10.1021/es800420p

Giyose, H., Manjra, A., Magoro, B., & Warren, M. (2011). The Investment Value of Resource Equities. 27. http://www.firstavenue.co.za/sites/default/files/downloads/The_Investment_Value_(03.08.2011).pdf Retrieved from http://www.firstavenue.co.za/sites/default/files/downloads/The_Investment_Value_(03.08.2011).pdf

Gómez, F., Guzmán, J. I., & Tilton, J. E. (2007). Copper recycling and scrap availability. *Resources Policy, 32*(4), 183-190. doi:10.1016/j.resourpol.2007.08.002

Gordon, R. B., Bertram, M., & Graedel, T. E. (2006). Metal stocks and sustainability. *Proceedings of the National Academy of Sciences, 103*(5), 1209-1214. doi:10.1073/pnas.0509498103

Gordon, R. B., Bertram, M., & Graedel, T. E. (2007). On the sustainability of metal supplies: A response to Tilton and Lagos. *Resources Policy, 32*(1-2), 24-28. doi:10.1016/j.resourpol.2007.04.002

Gordon, R. B., Koopmans, T. C., Nordhaus, W. D., & Skinner, B. J. (1987). *Towards a New Iron Age? Quantitative Modeling of Resource Exhaustion*. Cambridge, Massachusetts: Harvard University Press.

Hamarat, C., Kwakkel, J. H., & Pruyt, E. (2013). Adaptive Robust Design under Deep Uncertainty. *Technological Forecasting & Social Change, 80*(3), 408-418. doi:10.1016/j.techfore.2012.10.004

Hamarat, C., Kwakkel, J. H., Pruyt, E., & Loonen, E. T. (2014). An exploratory approach for adaptive policymaking by using multi-objective robust optimization. *Simulation Modelling Practice and Theory, 46*, 25-39. doi:10.1016/j.simpat.2014.02.008

ICSG. (2010a). *Release of ICSG 2010 Statistical Yearbook*. Retrieved from

ICSG. (2010b). *The World Copper Factbook*. Lisbon.

ICSG. (2011). *Copper: Preliminary Data for January 2011*. Retrieved from

JORC. (2004). *The JORC Code - Australasian Code for Reporting of Exploration Results, Mineral Resources and Ore Reserves*. Retrieved from

Kapur, A. (2006). The future of the red metal: discards, energy, water, residues, and depletion. *Progress in Industrial Ecology, An Int. J., 3*(3), 209-236. doi:10.1504/PIE.2006.011130

Kleijn, R., & van der Voet, E. (2010). Resource constraints in a hydrogen economy based on renewable energy sources: An exploration. *Renewable and Sustainable Energy Reviews, 14*(9), 2784-2795. doi:10.1016/j.rser.2010.07.066

Kwakkel, J. H., Auping, W. L., & Pruyt, E. (2013). Dynamic scenario discovery under deep uncertainty: The future of copper. *Technological Forecasting & Social Change, 80*(4), 789-800. doi:10.1016/j.techfore.2012.09.012

Kwakkel, J. H., Auping, W. L., & Pruyt, E. (2014). *Comparing Behavioral Dynamics Across Models: the Case of Copper*. Paper presented at the 32nd International Conference of the System Dynamics Society, Delft, The Netherlands.

Kwakkel, J. H., & Pruyt, E. (2015). Using System Dynamics for Grand Challenges: The ESDMA Approach. *Systems Research and Behavioral Science, 32*(3), 358-375. doi:10.1002/sres.2225

Lane, D. C. (1998). Can we have confidence in generic structures? *The Journal of the Operational Research Society, 49*(9), 936-947. doi:10.2307/3010167

Lane, D. C. (2000). Diagramming conventions in system dynamics. *Journal of the Operational Research Society, 51*(2), 241-245.

Lempert, R. J., Groves, D. G., Popper, S. W., & Bankes, S. C. (2006). A General, Analytic Method for Generating Robust Strategies and Narrative Scenarios. *Management Science, 52*(4), 514-528. doi:10.1287/mnsc.1050.0472

Lempert, R. J., Popper, S. W., & Bankes, S. C. (2003). *Shaping the Next One Hundred Years: New Methods for Quantitative, Long-Term Policy Analysis* (MR-1626). Retrieved from Santa Monica, CA, USA: http://www.rand.org/pubs/monograph_reports/2007/MR1626.pdf

LME. (2011, 2011). Copper graphs. Retrieved from http://www.lme.com/copper_graphs.asp

Lossin, A. (2005). Copper *Ullman's Encyclopedia of Industrial Chemistry*. Weinheim: Wiley-VCH Verlag GmbH & Co. KGaA.

McKelvey, V. E. (1973). *Mineral Resource Estimates and Public Policy*. Retrieved from

McKelvey, V. E. (1980). Seabed Mlinerals and the Law of the Sea. *Science Magazine, 209*(4455), 464-472. doi:10.1126/science.209.4455.464

Meadows, D. H., Meadows, D. L., Randers, J., & Behrens, W. W., III. (1972). *The Limits to Growth: A Report for the Club of Rome's Project on the Predicament of Mankind*. New York: Universe Books.

Meadows, D. H., Richardson, J., & Bruckmann, G. (Eds.). (1982). *Groping in the dark*. Bristol: J.W. Arrowsmith Limited.

Moorlag, R. A. C. W. M., Auping, W. L., & Pruyt, E. (2014). *Exploring the Effects of Shale Gas Development on Natural Gas Markets: a Multi-method Approach*. Paper presented at the 32nd International Conference of the System Dynamics Society, Delft, Netherlands.

Moxnes, E. (2005). Policy sensitivity analysis: simple versus complex fishery models. *System Dynamics Review, 21*(2), 123-145. doi:10.1002/sdr.311

Nassar, N. T., Barr, R., Browning, M., Diao, Z., Friedlander, E., Harper, E. M., . . . Graedel, T. E. (2012). Criticality of the Geological Copper Family. *Environmental Science & Technology, 46*(2), 1071-1078. doi:10.1021/es203535w

Pruyt, E. (2010). *Scarcity of Minerals and Metals: A Generic Exploratory System Dynamics Model*. Paper presented at the 18th International Conference of the System Dynamics Society, Seoul, Korea.

Pruyt, E. (2014). *Uncertainty and System Dynamics*. Paper presented at the 32nd International Conference of the System Dynamics Society, Delft, The Netherlands.

Pruyt, E., & Kwakkel, J. H. (2014). Radicalization under Deep Uncertainty: A Multi-Model Exploration of Activism, Extremism, and Terrorism. *System Dynamics Review, 30*(1-2), 1-28. doi:10.1002/sdr.1510

Ruhrberg, M. (2006). Assessing the recycling efficiency of copper from end-of-life products in Western Europe. *Resources, Conservation and Recycling, 48*(2), 141-165. doi:10.1016/j.resconrec.2006.01.003

Schlager, K. J. (1961). *A Systems Analysis of the Copper and Aluminum Industries: An Industrial Dynamics Study.* (MSc), MIT, Cambridge, MA.

Sterman, J. D. (1994). Learning in and about complex systems. *System Dynamics Review, 10*(2-3), 291-330. doi:10.1002/sdr.4260100214

Sterman, J. D., & Richardson, G. P. (1985). An Experiment to Evaluate Models for Estimating Fossil Fuel Resources. *Journal of Forecasting, 4*(2), 197-226.

Sterman, J. D., Richardson, G. P., & Davidsen, P. (1988). Modeling the Estimation of Petroleum Resources in the United States. *Technological Forecasting & Social Change, 33*(3), 219-249. doi:10.1016/0040-1625(88)90015-7

Svedberg, P., & Tilton, J. (2006). The real, real price of nonrenewable resources: copper 1870–2000. *World Development, 34*(3), 501-519. doi:10.1016/j.worlddev.2005.07.018

Tilton, J. E. (1996). Exhaustible resources and sustainable development: two different paradigms. *Resources Policy, 22*(1-2), 91-97. doi:10.1016/S0301-4207(96)00024-4

Tilton, J. E., & Lagos, G. (2007). Assessing the long-run availability of copper. *Resources Policy, 32*(1-2), 19-23. doi:10.1016/j.resourpol.2007.04.001

United, N. (1982). United Nations Convention on the Law of the Sea.

UNPD. (2011, 2011). World Population Prospects: The 2010 Revision. Retrieved from http://esa.un.org/unpd/wpp/index.htm

Urbance, R., Field, F. R., Kirchain, R. E., Roth, R., & Clark, J. (2002). Market model simulation: The impact of increased automotive interest in magnesium. *Journal of the Minerals, Metals and Materials Society, 54*(8), 25-33. doi:10.1007/BF02711863

Van Vuuren, D. P., Strengers, B. J., & De Vries, H. J. M. (1999). Long-term perspectives on world metal use—a system-dynamics model. *Resources Policy, 25*(4), 239-255. doi:10.1016/S0301-4207(99)00031-8

Wouters, H., & Bol, D. (2009). *Material Scarcity. An M2i study*. Retrieved from Delft: http://www.m2i.nl/images/stories/m2i%20material_scarcity%20report.pdf

3 Dynamic Scenario Discovery under Deep Uncertainty

In the commonly used definition of scenario within Scenario Discovery and Robust Decision Making, scenarios are defined as future states of the world in which particular policies perform poorly. There are two problems with this understanding of scenarios. First, it is often not the static state of a system that is of interest, but the different plausible behaviour patterns leading to this state. In these situations, it is thus necessary to be able to distinguish different types of behaviour over time. The dynamics of complex real-world problems can be simulated using highly non-linear simulation models. The non-linearity of such models however complicates the use of existing algorithms for Scenario Discovery, as these algorithms work best if there is a more or less orthogonal mapping of inputs to outputs. Second, the design and testing of policies can take advantage of knowing which uncertain factors and policy levers are most influential in determining the type of behaviour.

In this paper, we present an extension of Scenario Discovery. First, we understand scenarios not as future states of the world in which policies perform poorly, but as behaviour over time which is of interest to various stakeholders. We operationalize this by using behavioural pattern feature clustering. Next, we pre-process the model input space using Principal Components Analysis prior to using the PRIM algorithm (i.e., PCA-PRIM) to relate parts of the input space to the interesting part of the output space.

This combined approach allowed us to identify regions in the model input space responsible for generating behaviours of interest for about 50% of all such behaviours. The combined approach was not entirely successful because first, distinguishing between different types of oscillatory behaviour remained difficult, and second, the non-linearities in the models stretch the capability of the PRIM algorithm.

In the synthesis chapter, I will discuss these and similar lessons in the second section on the impact of complexity of models on Scenario Discovery.

This paper was originally published as:

Kwakkel, J. H., Auping, W. L., & Pruyt, E. (2013). Dynamic scenario discovery under deep uncertainty: The future of copper. *Technological Forecasting & Social Change, 80*(4), 789-800. doi:10.1016/j.techfore.2012.09.012

My contributions to this paper were: I developed the SD models, wrote the sections regarding the models (3.1, 3.2), designed the classification of principle components as used in the PCA-PRIM analysis, and interpreted the results of the PCA-PRIM analysis.

Dynamic scenario discovery under deep uncertainty: the future of copper

Jan H. Kwakkel, Faculty of Technology, Policy & Management, Delft University of Technology, Delft, The Netherlands

Willem L. Auping, The Hague Centre for Strategic Studies, The Hague, The Netherlands,

Erik Pruyt, Faculty of Technology, Policy & Management, Delft University of Technology, Delft, The Netherlands

Abstract

Scenarios are commonly used to communicate and characterize uncertainty in many policy fields. One of the main challenges of scenario approaches is that analysts have to try and capture the full breadth of uncertainty about the future in a small set of scenarios. In the presence of deep uncertainty, this is even more challenging. Scenario discovery is a model-based technique inspired by the scenario logic school that addresses this challenge. In scenario discovery, an ensemble of model runs is created that encompass the various uncertainties perceived by the actors involved in particular decision making situations. The ensemble is subsequently screened to identify runs of interest, and their conditions for occurring are identified through machine learning. Here, we extend scenario discovery to cope with dynamics over time. To this end, a time series clustering approach is applied to the ensemble of model runs in order to identify different types of dynamics. The types of dynamics are subsequently analyzed to identify dynamics that are of interest, and their causes for occurrence are revealed. This dynamic scenario discovery approach is illustrated with a case about copper scarcity.

KEYWORDS: scenario discovery, exploratory modeling and analysis, system dynamics, deep uncertainty, metal scarcity

1. Introduction

Scenarios provide a commonly used means to communicate and characterize uncertainty in many decision support applications. There exists a plethora of scenario definitions, typologies, and methodologies (Börjeson, Höjer, Dreborg, Ekvall, & Finnveden, 2006; Bradfield et al., 2005). A distinction can be made between the *La Prospective* school developed in France, the Probabilistic Modified Trends school originated at RAND, and the intuitive logic school typically associated with the work of Shell (Bradfield et al., 2005). In the evaluative literature, one of the reported problems of traditional scenario approaches is that they often struggle in case of problems that involve a variety of actors with quite diverse world views (European

Environmental Agency, 2009) or when there is a lacking consensus (van 't Klooster & van Asselt, 2006). Scenario approaches also struggle with anticipating rare events (Goodwin & Wright, 2010) and grapple with the multiplicity of plausible futures (Popper, Griffin, Berrebi, Light, & Daehner, 2009). The challenge of traditional scenario approaches is that analysts have to try and capture the full breadth of the uncertainty about the future in a small set of scenarios that need to be intelligible and useful to both the actors involved in the scenario development process and analysts supporting this process (Bryant & Lempert, 2010; Schwartz, 1991; van der Heijden, 1996). Developing or identifying a handful of scenarios, that fully represent all plausible futures is difficult. Communicating, and using more than a handful of representative scenarios is equally difficult and may even be counterproductive (van der Heijden, 1996). The intuitive logic school addresses these problems through the identification of the factors that are both highly uncertain and can have a profound impact on the decision problem at hand (Schwartz, 1991). However, this works mainly if the group of involved actors is relatively small, their interests and concerns are known, and overlap to a certain extent (Bradfield et al., 2005). Moreover, how to best represent the diversity contained in all the uncertain factors in a small set of scenarios, is a continuing challenge (Groves & Lempert, 2007).

Recently, an approach called scenario discovery (Bryant & Lempert, 2010; Groves & Lempert, 2007; Lempert et al., 2006) has been put forward as a technique that can be used for developing scenarios for problems that involve a large number of actors with quite diverging world views and values and where there are many uncertain factors. Scenario discovery is a model driven approach that builds on the intuitive logic school (Bryant & Lempert, 2010). Scenario discovery builds on earlier work on using models for decision making under deep uncertainty (Bankes, 1993; Lempert, Bryant, & Bankes, 2008; Lempert et al., 2003). It starts from an ensemble of model runs that is analyzed in order to identify runs that are of particular interest. Next, these runs of interest are analyzed to reveal the combinations of factors responsible for generating them. The documented cases of scenario discovery have used a single model with a small set of uncertain parameters as the basis for generating the ensemble of runs. For example, (Bryant & Lempert, 2010) uses a model with 8 uncertain parameters, and (Groves & Lempert, 2007) uses a model with 20 uncertain parameters, and the identification of interesting runs in both cases is based on the terminal values of each individual run of outcome indicators related to policy performance (Bryant & Lempert, 2010; Groves & Lempert, 2007).

In this paper, we extend the scenario discovery approach conceptually, technically, and practically. Conceptually, we understand scenarios not as states of the world but as developments over time. Technically, this implies that the machine learning techniques usually applied in scenario discovery cannot be applied straightforwardly. To overcome this problem, we use time series clustering for the identification of sets of behaviors over time, thus transforming time series results to scalar values that can be used as input to the various machine learning techniques that can be used for scenario discovery. Practically, we extend scenario discovery by working with two structurally distinct models that share only a subset of the uncertain factors, and jointly cover significantly more uncertain parameters than earlier applications of scenario discovery. These practical extensions pose additional challenges in the design of the computational experiments and the analysis of the results.

To illustrate our extended scenario discovery approach, we apply it to the problem of copper scarcity. There has been a growing attention to mineral and metal scarcity, but this attention has been focused mainly on lithium, rare earth metals and other metals characterized by supply risks due the limited number of countries where it is mined. However, base metals can also suffer from scarcity, as evidenced by the copper price which has been on a high level since 2005 (Index Mundi, 2011), resulting in phenomena like the theft of copper wiring. Crisis behavior in the copper market may have profound impacts on society beyond increased copper theft, and may be particularly worrisome with regard to a transition towards more sustainable energy systems (Kleijn & van der Voet, 2010). The main aim of the case was therefore to identify the various ways in which the copper system – comprised of supply, demand, recycling, and substitution – could evolve, the kinds of dynamics that could occur, the undesirable price dynamics, and the causes for their occurrence.

The physical side of the copper system is well documented (e.g. ICSG, 2010b; Lossin, 2005) and does not contain much uncertainty. However, with respect to the way in which demand should be represented, there are profoundly diverging views: there are those who argue that copper demand should be modeled at a high level of aggregation as a function of world population, while others argue that one should use a bottom up approach from the various types of usages to the overall demand (Gordon et al., 2006, 2007; Tilton & Lagos, 2007). As argued by Cole, “whether a 'top-down' or 'bottom-up' approach is chosen, however, may affect the results. Simple recursive calculation of global or regional aggregates broken down by sector often gives surprisingly different results from systematically building up the global or regional aggregates from the sector or subsector levels” (Meadows et al., 1982, p. 205). Other sources of uncertainty are the development of the ore grade (ICSG, 2010b; Skinner, 1976), the impacts of substitution behavior (Gordon et al., 1987), and various geopolitical developments, such as the growing copper demand in developing economies (ICSG, 2010b; Rademaker & Kooroshy, 2010). The various uncertainties are captured in two distinct simulation models. One represents a bottom up modeling of demand, while the other represents a top down modeling of demand. The supply system is essentially the same in both models. The behavior of these models is explored across a wide range of parametric uncertainties using Latin Hypercube Sampling. The results are clustered using a time series clustering approach, and subsequently analyzed using the Patient Rule Induction Method (Friedman & Fisher, 1999), a particular machine learning technique. Exemplars of undesirable dynamics are identified, and their conditions for occurring derived.

In the next section, we review the current scenario discovery approach and outline where and how we have extended it to cope with dynamics over time. In Section 3, we illustrate this modified scenario discovery approach through the case of copper scarcity. Section 4 discusses these results from a methodological point of view. Section 5 presents the main conclusions.

2. Dynamic Scenario Discovery

Scenario discovery addresses problems encountered when trying to develop model-based scenarios for problems that involve a large number of actors with diverging world views and values, or that are characterized by a very large number of uncertain factors. Typical for such problems is that the analysts do not know, or the parties to a decision cannot agree on (1) the appropriate conceptual models that describe the relationships among the key driving forces that shape the long-term future, (2) the probability distributions used to represent uncertainty about key variables and parameters in the mathematical representations of these conceptual models, and/or (3) how to value the desirability of alternative outcomes (Lempert et al., 2003). This is also called decision making under deep uncertainty, or severe uncertainty (Ben-Haim, 2006; Lempert et al., 2003). In the presence of a lack of knowledge or disagreement related to the model representation of a system and the evaluation of outcomes, enumeration of multiple alternatives for how (aspects of) the system work or are to be parameterized and how to value outcomes may still be possible, without being able to rank order these alternatives in terms of how likely or plausible they are judged to be (Kwakkel et al., 2010)

2.1. Exploratory Modeling and Analysis

Scenario discovery builds on earlier work on using models for decision making under deep uncertainty (Bankes, 1993; Lempert et al., 2008; Lempert et al., 2003). Under deep uncertainty, it is not possible to develop a single model that accurately represents the system of interest. Exploratory Modeling and Analysis (EMA) (Bankes, 1993) provides an alternative way of using the available information, data, and knowledge. An ensemble of models, consistent with the available knowledge, data, and information is developed. A single model run drawn from this potentially infinite ensemble of models provides a computational experiment that reveals how the world would behave if the various guesses that particular model makes about the various uncertainties were correct. The behavioral landscape of this ensemble is explored using a series of computational experiments, and the behavioral landscape is analyzed using a variety of machine learning techniques. Scenario discovery can be understood as a particular application of EMA, where one tries to identify the combinations of uncertainties that produce regions of interest in the behavioral landscape of model outcomes. Typically, in scenario discovery, one looks for areas that represent vulnerabilities to proposed policies by looking at the terminal values of various outcomes of interest for the individual runs (Bryant & Lempert, 2010). There is, however, no theoretical or conceptual reason why other criteria cannot be used for defining the regions of interest, nor is it necessary to restrict the identification of regions of interest to terminal values of outcomes of interest for the individual runs.

Important steps in EMA are to (i) conceptualize the decision problem and the associated uncertainties; (ii) develop an ensemble of fast and simple models of the system of interest; (iii) specify the uncertainties that are to be explored. Next, depending on the purpose for which EMA is applied, various subsequent steps are possible. In the case of scenario discovery, the typical subsequent steps are to (iv) analyze the behavioral landscape resulting from (iii); (v) identify the combinations of uncertainties from which regions of interest in the behavioral landscape originate; (vi) assess these combinations of uncertainties using various model quality metrics and related machine learning techniques for assessing model quality (Bryant & Lempert, 2010); (vii) qualitatively or quantitatively communicate the typical futures in these regions of interest, i.e. exemplary scenarios, and the combinations of uncertainties from which the regions of interest in the behavioral landscape originate to the actors involved in the decision making problem. Based on these scenarios and regions, In interaction with these actors, a process of policy formulation and reformulation can start. Candidate policies can be tested by iterating through the previous steps.

2.2. Algorithms for Dynamic Scenario Discovery

Both the analysis of the behavioral landscape and the identification of uncertainties from which particular regions of interest in the behavioral landscape originate can utilize various machine learning techniques. In this paper, we are interested in dynamics over time. Thus, we analyze the behavioral landscape using time series clustering. From this analysis emerges a set of dynamics, some of which are judged to be of interest. In order to identify the combinations of uncertainties responsible for generating the behavior of each cluster of interest, we utilize a modified version of the Patient Rule Induction Method (PRIM) (Friedman & Fisher, 1999), which is typically used in scenario discovery (Bryant & Lempert, 2010; Lempert et al., 2008).

2.2.1. Coping with dynamics: time series clustering

The goal of clustering is to organize an unlabeled data set into homogenous groups, in which the similarity within the group is minimized and the dissimilarity between groups is maximized (Theodoridis, 2003; Warren Liao, 2005). Typically, clustering approaches are applied to static data (Warren Liao, 2005). Static clustering approaches can be divided into five families: partitioning methods, hierarchical methods, density based methods, grid-based methods, and model-based methods (Han & Kamber, 2001). In general, time series clustering approaches try to modify existing clustering approaches for static data so that they can cope with time series data. Either the algorithm is modified to deal with the raw time series data, or the time series are processed in such a way that static clustering methods can be used directly. A substantial portion of the research on time series clustering focuses on modifying the similarity measure used in a clustering method to handle time series data (Keogh & Kasetty, 2003). To assess the efficacy of hierarchical clustering methods in capturing similarity, the use of dendrograms is recommended (Keogh & Kasetty, 2003). A review of the state of the art in time series clustering can be found in (Warren Liao, 2005).

In this paper, we adopt an agglomerative hierarchical clustering approach. That is, we start by positioning each time series in its own cluster, and then hierarchically merge each cluster into larger and larger clusters (Warren Liao, 2005). The advantage of this approach is that it produces an ordering of similarity. The user can experiment with different similarity thresholds, producing a different number of final clusters, which aids the discovery of the types of dynamics.

2.2.2. Identifying undesirable regions: the Patient Rule Induction Method (PRIM)

After identifying regions of interest in the behavioral landscape of model outcomes, one wants to identify where in the model input space these regions of interest originate. That is, one wants to close the loop from uncertainties to behavior of interest and from behavior of interest back to the uncertainties. For this final step, PRIM is typically used in scenario discovery (Bryant & Lempert, 2010; Groves & Lempert, 2007; Lempert et al., 2008).

PRIM can be used for data analytic questions, where the analyst tries to find combinations of values for input variables that result in similar characteristic values for the outcome variables. Specifically, one seeks a set of subspaces of the model input space within which the values of the output variables are considerably different from their average values over the entire domain. PRIM describes these subspaces in the form of 'boxes' of the model input space. This results in a very concise representation, for typically only a limited set of dimensions of the model input space is restricted. That is, a subspace is characterized by upper and/or lower limits on only a few input dimensions. Still, interpretation of such a PRIM box can be challenging for the analyst because of the multi-dimensional character of the subspace. That is, in interpreting the results one has to account for all the restricted dimensions simultaneously. Another issue with PRIM is that the subspaces found by PRIM can overlap, further hampering interpretability. Although the original version of PRIM requires that the input data are continuous or categorical (Friedman & Fisher, 1999), it can be modified to deal with ordinal data (Chong & Jun, 2008).

The default objective function used by PRIM aims at maximizing the mean of the cases inside a box. The algorithm generates a set of candidate boxes and then selects that candidate box that maximizes this mean. In the case of categorical or ordinal data, there is a problem with this approach, for the various candidate boxes do not contain the same amount of data. Thus, the comparison is biased. Moreover, this defeats one of the principal strengths of PRIM, namely its patient (or lenient) character, expressed in removing or adding only a few cases at a time. To address this problem, we modified the objective function used by PRIM. Instead of selecting the candidate box that maximizes the mean, we modified the algorithm to select the box where the increase in the mean divided by the change of the cardinality of the cases inside the box is largest. Thus, the gain in the mean is offset against the change in the number of cases inside a box.

In recent work on using PRIM for scenario discovery, it was shown that the results of PRIM could be improved by applying a preprocessing step based on Principal Components Analysis (PCA) (Dalal, Han, Lempert, Jaycocks, & Hackbarth, 2013). This preprocessing step involves identifying the experiments of interest. Next, the covariance matrix of these experiments is derived (Weisstein, 2012). Then, the eigenvectors and eigenvalues of this covariance matrix are identified through Singular Value Decomposition. The eigenvectors are sorted based on the eigenvalues. The resulting eigenspace is subsequently used for rotating all the experiments. That is, the experiments are rotated to the eigenspace of the covariance matrix of the experiments of interest. To ease interpretation of the results, instead of rotating the uncertain parameters jointly, one can cluster the uncertain parameters, and derive a rotation for each cluster separately. This extension of PRIM, known as PCA-PRIM, is applied here. We exclude categorical dimensions from this rotation.

Some computational support for scenario discovery exists. The Comprehensive R Archive Network (CRAN) contains a scenario discovery toolkit package. However, this implementation of PRIM does not handle ordinal data and categorical data correctly, and the pasting phase of the algorithm is not consistent with the algorithm as described by (Friedman & Fisher, 1999). It is not uncommon to encounter categorical or ordinal data in scenario discovery. In scenario discovery, the model input space is defined by the uncertainties explored using EMA. Not all these uncertainties will be real valued. For example, if there is uncertainty about (an aspect of) model formulation, this can result in multiple distinct model formulations, which are distinct categories. We therefore re-implemented the PRIM algorithm to handle the three data types correctly. In this revised implementation we modified the pasting part of the algorithm to be consistent with (Friedman & Fisher, 1999). This re-implementation also allowed us to modify the objective function used in PRIM.

3. Illustrating Dynamic Scenario Discovery: the case of copper scarcity

The aim of this case study is to explore how the copper price can develop in the future and, through dynamic scenario discovery, identify plausible *undesirable* future dynamics of the copper price. To this end, we first conceptualize the copper system and the main uncertainties affecting the future evolution of the copper price. Second, we present an ensemble of system dynamics models that can be used to explore over these uncertainties. Third, we discuss the design of experiments to explore the behavior of this ensemble. Fourth, we apply the outlined dynamic scenario discovery approach to the results of the computational experiments.

3.1. Conceptualizing the problem and the associated uncertainties

We have chosen to focus in this case study on discovering plausible undesirable dynamics of the copper price. What is deemed undesirable differs from one actor to another. Broadly speaking, periods of high prices are undesirable for consumers, periods of low prices are undesirable to producers, and very rapid changes in price are undesirable to both. Later, in analyzing the clusters of dynamics, we will return to these forms of undesirability in order to characterize the dynamics of the identified clusters.

The dynamics of the copper price arise out of an interaction between supply and demand. The supply side of the copper industry, and in particular the supply chain, is extensively documented (ICSG, 2010b; Lossin, 2005). The chain starts with either the amount of the resource available or the resource base, although there is some discussion about which of these two to use in estimating the maximum amount of available copper (Gordon et al., 2006, 2007; Tilton & Lagos, 2007). The available resources are translated into reserves and a reserve base. The reserves are the fraction of the available resources that can be mined economically; the remainder is the reserve base (JORC, 2004; McKelvey, 1973). Part of the reserve base can become part of the reserves due to copper price and technological changes. Copper reserves are transformed into copper via mining, smelting, and refining. Different mining techniques have associated cost structures. The process of smelting and refining depends on the type of copper ore, creating two final products: cathode copper and ingots. Detailed descriptions for the mining, smelting, and refining of copper, and their available capacity are readily available in the literature (ICSG, 2010b; Lossin, 2005). With respect to the copper ore that is being mined, the ore grade is declining (ICSG, 2010b), and, as a consequence, the marginal costs of copper mining in terms of energy demand for refining (lower grade) copper are increasing (Sun, Nie, Liu, Wang, & Gong, 2010). How this decline of the ore grade will evolve in the future is uncertain.

With respect to the demand side of the copper system, the uncertainties are much more profound. There is an ongoing debate about why the copper price has been so high. Two main explanations that are offered are economic growth in emerging economies and energy transitions (ICSG, 2010b; Kleijn & van der Voet, 2010). These explanations reflect a different idea about how copper demand arises. It can be seen as a function of world population and GDP, or as a function of different usages of copper. Apart from uncertainties related to demand, there are also deep uncertainties related to the future development of energy prices and the dynamics of substituting copper with other materials. The future energy price affects both the copper price as well as the price of substitutes. In particular, the price of aluminum, a major substitute for copper in electronic applications, is highly dependent on energy price developments. In addition, the dynamics of substitution depend on the available substitutes, the amount of materials needed in case of substitution, and the price of the substitutes in comparison to the copper price (Gordon et al., 1987), each of which can change over time.

3.2. The ensemble of models

The models used for generating future scenarios were constructed using System Dynamics (SD) modeling (Sterman, 2000) and were implemented in Vensim (Ventana Systems Inc., 2011). These models have been designed in such a way that they allow for the exploration of the specified uncertainties, while offering an endogenous explanation for the overall behavior of the copper system in terms of how supply, demand, price, and substitution change over time. To cope with uncertainty about whether demand is a function of population and GDP (i.e. the top down perspective) or whether demand is a function of usages of copper (i.e. the bottom up perspective), we developed two distinct SD models. These two models have a common core with respect to the supply of copper but differ with respect to how demand is modeled. Below we offer a concise description of both models and their common core. For a more detailed description, see (Auping, 2011).

3.2.1. Common core: supply

The copper supply is modeled as a stock-flow structure, from resource base, resources, and reserve base via mining and refining, to the global consumption of refined copper and copper use. When global copper in use has reached the end of its lifetime, it is partially collected via the global secondary copper market as scrap and is recycled as copper recovered from scrap. These stocks and flows constitute the physical and technical backbone of the system (ICSG, 2010b; Lossin, 2005). With respect to the discussion about resource versus resource base (Gordon et al., 2007; Tilton & Lagos, 2007), we assume that the marginal cost of copper is a function of the amount of energy needed to mine it and the price of energy. The price is a function of demand and supply. The difference between the marginal cost and the price ultimately defines how much copper can be mined. The structure of resources and reserves further largely follows the McKelvey classification (McKelvey, 1973). The discovery of new resources is modeled through the semi-autonomous findings of independent exploration or junior companies.

The mining, smelting, and refining capacities are modeled as stocks. Capacity can exist in one of three states: under construction, in use, and mothballed (Abdel Sabour & Poulin, 2010). During poor economic times, capacity is first mothballed and, only after continued losses, decommissioned. Further, due to the maturity of the copper market, learning effects are assumed to be negligible in relation to the increasing costs related to the declining ore grade. No separate capacity for recycling is modeled, since the recycling of copper uses the same smelter and refinery capacity as newly mined copper.

3.2.2. Top down demand

The Causal Loop Diagram (CLD) (Lane, 2008; Sterman, 2000) in Figure 1 shows the main feedback loops of the top down demand sub model at a high aggregation level. In this model, the intrinsic demand is caused by the development of the world population, the development of GDP per capita, and the effect of GDP on the demand for copper. This model allows for exploring the effects of both the growing world population on copper demand and, potentially even more important, the effects of the increase in wealth of the large populations of emerging economies.

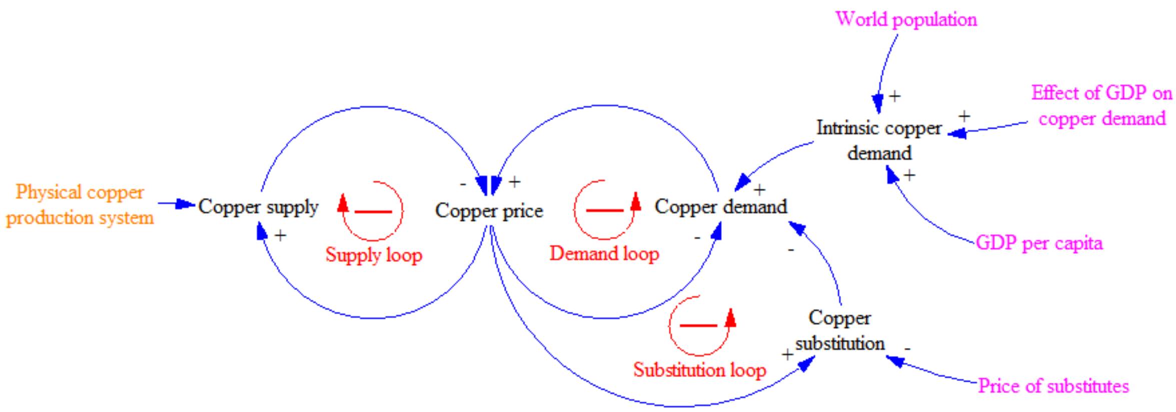


Figure 1. High level CLD of top down model. Arrows indicate a causal influence and the sign indicates the direction of the influence. The circular arrows with a sign in the middle represent feedback loops. The sign indicates whether the loop is positive or negative.

3.2.3. Bottom up demand

A high level CLD of the bottom up demand sub model is shown in Figure 2. In this model, intrinsic demand is a function of the various uses of copper. Following (Angerer et al., 2010), we define six main uses for copper, all of which can have a quasi-independent development over time. Moreover, substitution is based on each of the uses of copper and can behave differently over time for each use. This model allows for exploring, for example, the effects of energy transitions, which among others affect the use of copper in cars (Angerer et al., 2010) and the need for new energy infrastructure to connect renewable energy sources like wind parks and solar plants (Kleijn & van der Voet, 2010). This model has a significantly larger number of uncertain parameters than the top down demand model.

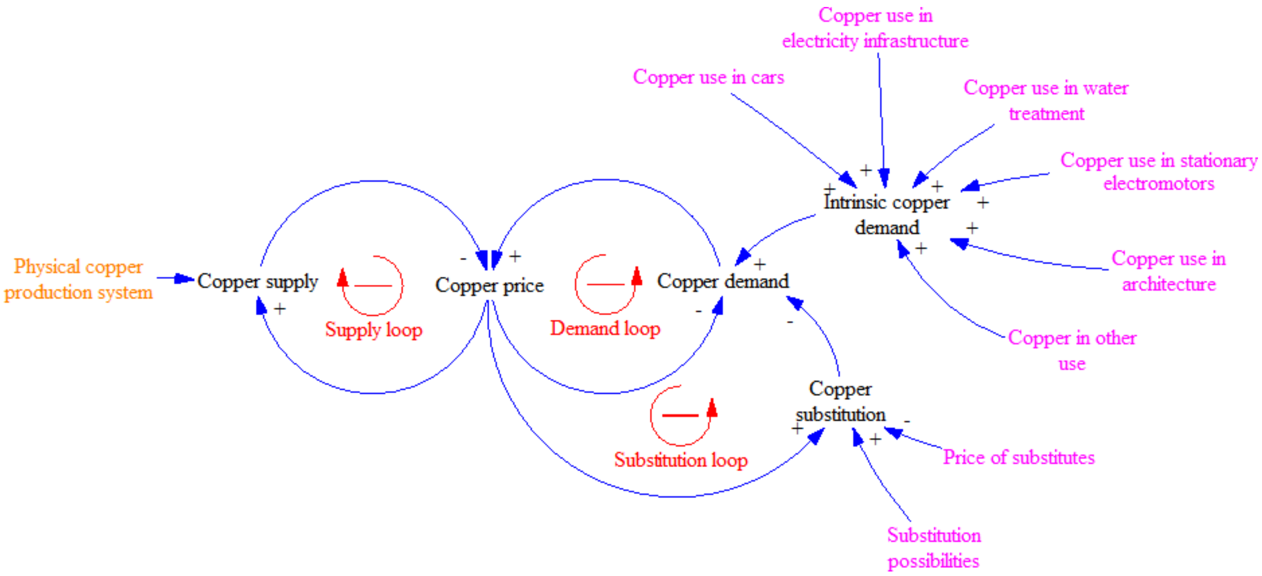


Figure 2. High level CLD of bottom up model. Arrows indicate a causal influence and the sign indicates the direction of the influence. The circular arrows with a sign in the middle represent feedback loops. The sign indicates whether the loop is positive or negative.

3.2.4. Matching supply and demand

Copper supply (available global inventories of refined copper) and demand (intrinsic demand minus substitution) determine the copper price relative to the marginal costs. The copper price influences the total copper demand in two ways: directly, by price elasticity effects on demand, and indirectly by substitution of copper demand (Figure 3). Price and substitution effects both have a short-term and long-term effect on the price. On average, the long-term effect is larger than the short-term effect. When the total demand cannot be fulfilled, the unfulfilled demand is accumulated. This accumulated unfulfilled demand will decline over time, as long as copper scarcity persists. The actual demand for refined copper is thus the demand formed by intrinsic demand, price effects, substitution, and accumulated unfulfilled demand.

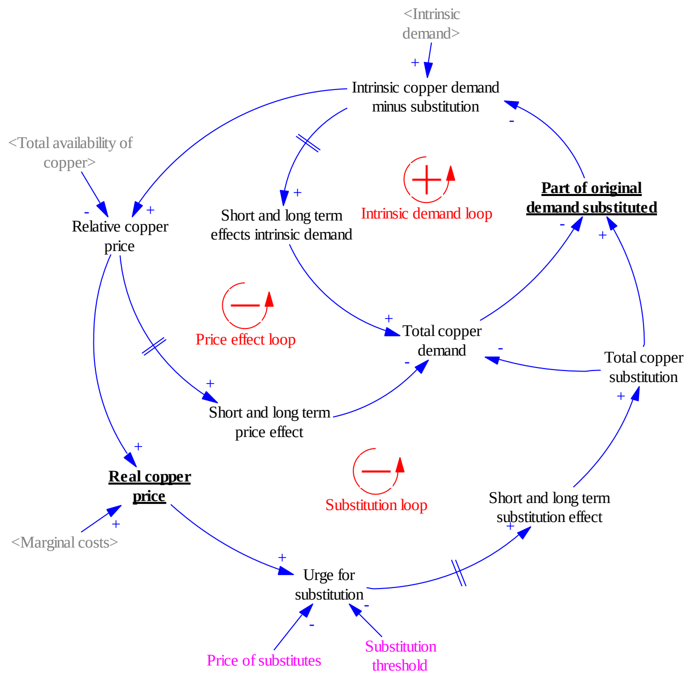


Figure 3. CLD of the matching of demand to the price. Arrows indicate a causal influence and the sign indicates the direction of the influence. Arrows with a strike through indicate delayed causal influences. The circular arrows with a sign in the middle represent feedback loops. The sign indicates whether the loop is positive or negative.

3.3. The design of experiments

In order to explore the behavior of the two models over the uncertainties, a shell written in Python was utilized (Van Rossum, 1995). This 'EMA workbench' can control Vensim through its Dynamic Link Library (Ventana Systems Inc., 2010). The workbench is responsible for generating input values for the various uncertainties, setting these values in the Vensim models, executing the Vensim models, and storing the results. The workbench supports parallel processing, reducing the required computational time.

The top down model contains 31 uncertain factors, and the bottom up model contains 57 uncertain factors. In order to compare the extent to which the way of modeling demand has an impact on the results, we first identified the uncertainties that are shared by the two models. We generated a Latin Hypercube sample containing 2500 experiments for these 12 shared uncertainties. Next, for each model, the uncertainties unique to that model were identified, and a Latin Hypercube was generated for these uncertainties. By combining these two sets of Latin Hypercube samples for each model, a complete set of experiments for each model is derived. This results in a dataset containing 5000 computational experiments.

3.4. Analysis of Results

3.4.1. Time Series Clustering

In order to identify the types of dynamics present in the ensemble of model results, we use an agglomerative hierarchical clustering approach with a specifically tailored metric for identifying similar dynamics. The dynamic behavior over time can be understood as being a concatenation of atomic behavior patterns (D. N. Ford, 1999). The atomic behavior mode is based on the sign (positive, negative, and zero) of the slope and curvature. Thus nine atomic behavior modes exist. Each time series is transformed into a concatenation of atomic behavior modes. Next, similarity between two time series is based on comparing these concatenations. The distance between two dynamics is then the average deviation across the entire concatenation. This approach is essentially an extension of the behavior pattern features discussed in (Yücel & Barlas, 2011). See (Yücel, 2012) for more a detailed elaboration on this similarity metric.

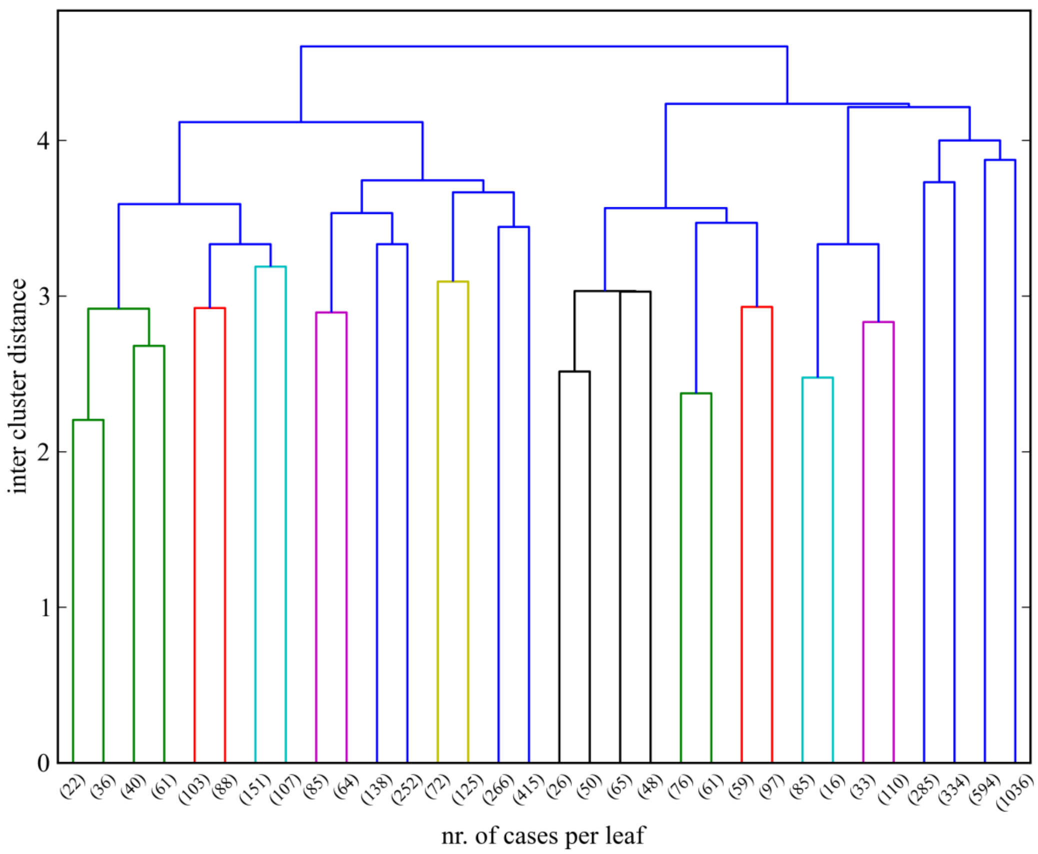


Figure 4: Dendrogram of the clustering of the dynamics of the copper price. The y-axis specifies the inter cluster distance, and the x-axis indicates the number of computational experiments allocated to the lowest level in the dendrograms.

Figure 4 shows the dendrograms (Sibson, 1973) resulting from the agglomerative clustering with the behavioral distance metric. We have shown only the top four levels. For explorative purposes, we chose to extract 30 clusters and review their dynamics. This offers a balance between choosing too few clusters and potentially missing insight into the variety of possible dynamics generated by the 5000 experiments, and choosing too many clusters resulting in information overload. In order to review the dynamics of each cluster, we identified the computational experiment that had the lowest average distance to the other experiments in the same cluster and used it as the exemplar for the dynamics in that cluster. In this way, we are able to move from 5000 computational experiments to 30 exemplar dynamics.

Figure 5 shows these 30 exemplars. To ease the interpretation of the plots, we plotted only 10 exemplars jointly, and scaled the values so that they all have the same starting point. The y-axis is shared across the plots. In the top and bottom plot we see exemplars that show a slow growth of the copper price over time. We also observe several exemplars that show sharp price changes.

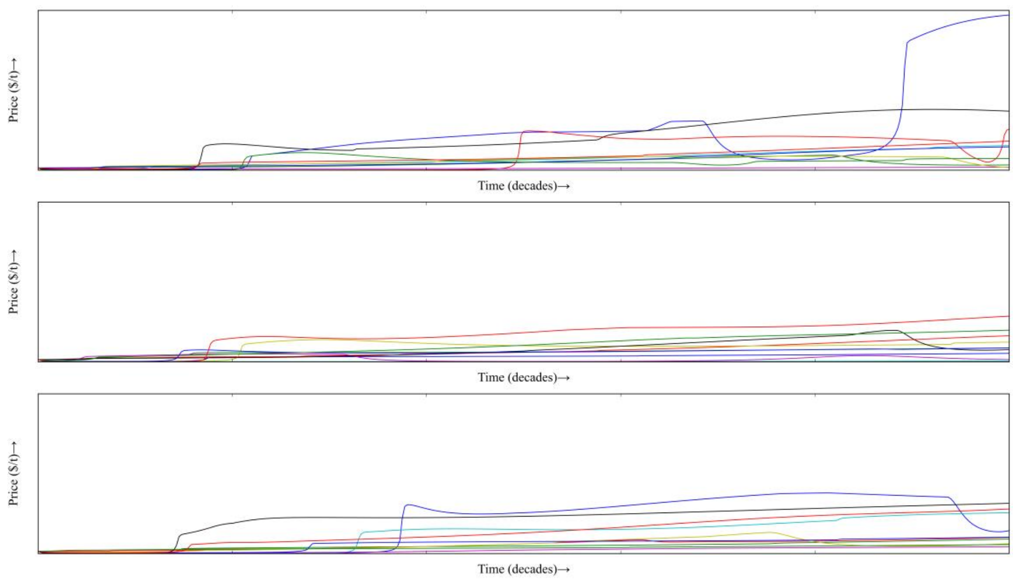


Figure 5: Thirty exemplars of the dynamics of the copper price

3.4.2. Scenario Discovery

Having identified the behavioral landscape through time series clustering, the next step is to specify which of the identified dynamics are undesirable. We assume that high prices are undesirable for consumers, low prices are undesirable to producers, and (rapid) price changes are undesirable to both. The time series clustering reveals several clusters with sharp price fluctuations. For illustrative purposes, we see regions in the behavioral landscape that show sharp changes in price as the regions of interest. Other criteria could also be used (see e.g., McInerney, Lempert, & Keller, 2012; van der Pas, Walker, Marchau, Van Wee, & Agusdinata, 2010). To identify these regions, we use a simple rule. If the absolute value of the first order derivative of the price is higher than 2 anywhere over the run, the run is classified as a run of interest. Next, we use PCA-PRIM (Dalal et al., 2013) to identify where in the uncertainty space these undesirable dynamics originate. This choice for PCA-PRIM was due to the fact that without the PCA preprocessing step, PRIM failed to identify subspaces in the uncertainty space from which the regions of interest in the behavioral landscape originate. Given that the experiments for both models are different, we applied PCA-PRIM to both sets of experiments separately.

3.4.2.1 Top Down Model

For the top down model, PRIM identified two boxes which jointly contain 55% of the experiments of interest. These two boxes are primarily characterized by six restricted principal components. Table shows the loadings of these six principal components on the uncertain parameters. These loadings indicate the strength and direction of the relation between the principal component and the uncertainty. For example, ‘minimum usage smelting and refining capacity’ is strongly positively related to the ‘Supply_3’ principal component, while ‘initial value GDP per capita’ is negatively related to the ‘Demand_6’ principal component. As can be seen, we clustered the uncertain parameters into parameters related to the capacity subsystem, the demand subsystem the supply subsystem, and the price subsystem.

With respect to capacity, the results indicate that the main driver for undesirable price fluctuations coming from the capacity subsystem are the permit time for additional smelting and refining capacity, the initial value for the smelting and refining capacity, and the lifetime of the mines. The underlying dynamic here is that, if there is a misbalance between the mining capacity and the refining and smelting capacity, there can be a buildup of raw copper. But, due to a lack of smelting and refining capacity and the fact that it can take a long time to expand this capacity, it takes a very long time before there is sufficient refining capacity. The long permit time results in an overshoot of new smelting and refining capacity, thus resulting in a sudden increase of refined copper becoming available on the market. This in turn, triggers a rapid drop of the price. Because of the overshoot in capacity, the stock of raw copper is depleted, reducing utilization of the capacity, in turn causing the mothballing of this capacity, and potentially even its dismantling. This effect of the capacity system is aggravated by the supply subsystem, which contributes to price fluctuations if all or a large fraction of the mining and smelting capacity is continuously used.

The demand subsystem contributes to undesirable price fluctuations if the economy is shrinking and the initial value for GDP per capita is low. Intrinsic demand for copper is highly sensitive to small changes in GDP (this is essentially the meaning of ‘normalization value GDP’) (Wouters & Bol, 2009). That is, copper demand rapidly decreases. However, historically the expectation is that demand will continue to increase (represented by the ‘initial value long term effect intrinsic demand’), and it takes a while before this expectation is changed (‘long-term increased demand due to intrinsic demand’). Hence, in case of a falling intrinsic demand and the expectation of continuing growth of demand, the demand subsystem contributes to price fluctuations.

Table 1: Rotation matrix for the top down model. The columns are the principal components. The values indicate the loadings of the uncertainties on these components. Loadings indicate the weight by which the normalized value for an uncertainty needs to be multiplied in order to get the principal component score. Thus, the higher the loading, the stronger the relation, and the sign indicates the direction of the relation. Only the principal components that are most important according to the PRIM results are shown. Irrelevant principal components and their related uncertainties have been removed.

	Capacity_4	Demand_5	Demand_6	Demand_7	Supply_3	Price_1
Average smelting and refining capacity permit term	0.59					
Initial mining capacity	-0.22					
Initial mining capacity in preparation	-0.19					
Initial smelting and refining capacity	0.50					
Average mine lifetime	0.55					
Initial value GDP per capita		-0.16	-0.81	-0.41		
Initial value long term effect intrinsic demand		0.43	-0.20	0.05		
Long term increase demand due to intrinsic demand		-0.44	0.08	0.07		
Base economic growth		-0.72	0.13	0.05		
Amplification factor of intrinsic demand		-0.05	-0.23	0.11		
Amplification factor of relative price effect		-0.10	0.13	0.03		
Normalisation value GDP		-0.02	-0.36	0.90		
Long term copper price elasticity		0.25	0.29	0.03		
Minimum usage smelting and refining capacity					0.58	
Initial inventories of refined copper					-0.06	
Switch between using all capacity, or capacity based on forecast					-0.80	
Forecasting time horizon					-0.11	
Price amplification factor						-0.86
Production time						0.51

The price subsystem contributes to price fluctuations if the relative price of copper is highly sensitive and the production time is low. The production time affects the time it takes for the inventories of copper to progress through the market. The lower this value, the more the amount of copper fluctuates, which in turn is aggravated by the 'price amplification factor'.

3.4.2.2 Bottom Up Model

For the bottom up model, PRIM identified two boxes that jointly contain almost 50% of the experiments of interest. These two boxes are primarily characterized by four principal components, which are related to supply, demand, recycling, and price. The contribution of the price subsystem to the price fluctuations is identical to its contribution in the top down model. In contrast to the top down model, there are no principal components related to the capacity subsystem. The supply subsystem contributes to price volatility if the time horizon used in forecasting future demand levels is quite short in combination with a high value for the minimum usage of smelting and refining capacity. The demand subsystem contributes to price volatility if the price elasticity is low, the long term effects of demand and price are quite strong, and the initial value for GDP per capita is low. This corresponds to an inert demand sector that only slowly adjusts to price changes. The recycling subsystem contributes to price volatility if the initial amount of copper in use is quite high. This means that, over the simulation, a large amount of copper suddenly becomes available via recycling. This lowers the price, but demand responds only slowly, resulting in a prolonged period of low prices. During this period, mining capacity is mothballed and potentially removed. By the time demand starts to respond to the low price, mining capacity and smelting and refining capacity has been reduced, thus resulting in a situation with more demand than supply, driving up the price.

Table 2: Rotation matrix for the bottom up model. The columns are the principal components. The values indicate the loadings of the uncertainties on these components. Loadings indicate the weight by which the normalized value for an uncertainty needs to be multiplied in order to get the principal component score. Thus, the higher the loading, the stronger the relation, and the sign indicates the direction of the relation. Only the principal components that are most important according to the PRIM results are shown. Irrelevant principal components and their related uncertainties have been removed.

	Supply_3	Demand_15	Recycling_10	Price_1
Forecasting time horizon	-0.56			
Minimum usage smelting and refining capacity	0.72			
Part of resource base sea based	0.39			
Threshold for junior companies to start deep sea reserve base development	0.13			
Goal for copper in infrastructure 2050		0.13		
Initial value GDP per capita		-0.39		
Number of cars in 2000		0.09		
Cu in vehicles BEV		0.14		
Long term effect on demand period		0.37		

	Supply_3	Demand_15	Recycling_10	Price_1
Long term increase demand due to intrinsic demand		-0.28		
Base economic growth		0.17		
Initial value long term effect price		0.45		
Amplification factor of intrinsic demand		-0.19		
Switch new cars		0.21		
Cu in vehicles cityBEV		0.02		
Long term effect intrinsic demand period		-0.21		
Switch World population		-0.12		
Amplification factor of relative price effect		-0.04		
Long term copper price elasticity		-0.43		
Growth of copper in architecture		-0.17		
Collection rate copper products Water Treatment			0.02	
Average lifetime of copper in use Architecture			-0.02	
Collection rate copper products Automotive			-0.04	
Percentage copper recovered from scrap			-0.24	
Initial value global copper in scrap Stationary Electro motors			-0.06	
Initial use grade for Water Treatment			-0.02	
Initial total copper in use			0.95	
Average lifetime of copper in use Other Use			-0.14	
Copper score during treatment Water Treatment			-0.03	
Copper score during treatment Energy Infrastructure			-0.05	
Initial value global copper in scrap Other Use			-0.07	
Price amplification factor				-0.92
Production time				0.39

4. Discussion

The aim of this paper was to illustrate how scenario discovery can be used to develop dynamic scenarios. To this end, we utilized a behavioral time series clustering algorithm to identify the behaviors of interest. Next, PCA-PRIM was used to identify the causes for the behaviors of interest. Looking at the final results, it appears that such a technique is able to identify a subset of uncertainties that produce a particular dynamic. In the case of the top down model, we started with 31 uncertain factors, and we were able to identify a substantially smaller subset of uncertainties that is responsible for a particular type of undesirable behavior. Similarly, in case of the bottom up model, we started with 57 uncertainties and were able to reduce it also to a substantially lower number of uncertainties responsible for generating the undesirable dynamics. Still, in both cases, only around 50% of the cases of interest have been accounted for. It thus appears that the dynamics of the model are due to a very particular dependency among the uncertainties. So, the idea of tracing back the behavior of interest to a region of the model input space does not hold for all of the regions of interest.

The identified scenarios suggest that given the current functioning of the copper system, sharp price fluctuations are indeed a plausible future. That is, it is conceivable that the prices will come down quickly and stay there for a period of time, before rising sharply again. Thus, the assumption that the current high price level is necessarily here to stay is not justified in light of these results. With respect to the identified undesirable dynamics and their underlying causes, there are several policy implications. There is a mismatch between the production of raw copper from mining and the production of copper in scrap through recycling and the smelting and refining capacity in the top down model, and to a lesser extent in the bottom up model. This mismatch results in oscillations in the supply of refined copper. To counteract these oscillations, there is a need for supply chain management. Another policy implication is that given the uncertainty related to future demand, and the impact of a declining demand on the price in case of the bottom up model, there is a need to create flexibility in the development of new mining capacity and new smelting and refining capacity. The use of real options or outside the box thinking with respect to the identification of new reserves are ways of addressing this need for flexibility (Abdel Sabour & Poulin, 2010; Tapscott & Williams, 2006). A last important policy implication is that price oscillations can be counteracted through the establishment of strategic reserves that are used in a counter cyclical way. Thus, copper is bought when prices fall below a particular level and sold when the prices rise above a particular level.

To improve the interpretability of the results, we experimented with various techniques. First, we decided to use the time series clustering only as a means of interpreting the model outcomes and derive from it a rule for identifying the cases of interest, rather than using the clusters themselves as cases of interest. This improved the results substantially. Second, we used a preprocessing step based on Principal Components Analysis prior to applying PRIM. Where PRIM without this preprocessing step was unable to identify a region in the model input space, PCA-PRIM was able to identify regions. These regions identified by PCA-PRIM are in the eigenspace rather than in the model input space, thus necessitating an additional interpretation step based on the loadings of the model input parameters on the principal components – that is, the dimensions that span the eigenspace.

One way for further improving the coverage, the fraction of all the cases of interest in a single prim box, would be to modify the objective function used in PRIM. In the current version, the objective function aims at maximizing the mean of the cases inside a box divided by the change of the cardinality of the cases inside the box. Assuming that a binary classification is used to identify cases of interest, this is the same as maximizing the density –the fraction of cases of interest in the box divided by the total number of cases in the box (Groves & Lempert, 2007)– of the box. However, it is possible to modify the algorithm to also take into account the change in coverage–the fraction of cases of interest in the box divided by the total number of cases of interest (Groves & Lempert, 2007) . That is, one could select the box that increases the density the most relative to the decrease in coverage. This way, the algorithm would search for boxes that have both a relatively high coverage and relatively high density. Conceptually, this is closely related to a modification of PRIM known as 'flexible PRIM', where one can make a tradeoff between mass and mean (Chong & Jun, 2008). A related way in which the results could be improved is by adding some form of penalty to the objective function based on the number of restricted dimensions. Ideally, one would like to identify a small manageable subset of the uncertain factors that is responsible for a large number of the undesirable dynamics. In the approach outlined by Bryant and Lempert (2010); Lempert et al. (2008), this is handled by making the PRIM process interactive. A user can then make his or her own tradeoff between increasing density and the number of restricted dimensions. This increases the interpretability of the results, but it makes reproducibility more difficult. Moreover, it introduces some form of expert judgment into the procedure. However, it would be convenient if PRIM were able to be used in an automatic mode, where the tradeoffs between density, coverage, and the number of restricted dimensions are pre-specified and reproducible.

Another direction for improving the results is in the time series clustering. We used a behavioral metric introduced in (Yücel, 2012). This metric helped in identifying plausible dynamics. However, as indicated in (Yücel, 2012), the metric is less suited to oscillatory dynamics. Moreover, a more fine grained analysis of the types of dynamics could further improve the results. So far, we have lumped all cases of sharp changes in price together. A more fine grained approach could be to separate sharp price fluctuations that become worse over time from those cases where the price fluctuations die out.

A final direction for improving the results could be through model simplification. The two models used in this case are quite large and describe the copper system in detail. A simplified model at a higher level of aggregation, which still contains the most elemental feedback loops and the supply chain stock flow structures, could conceivably produce results very similar to the results reported here. However, by producing these results with a simpler model, the number of uncertainties over which the exploration is done, and the interpretation of their results, would be reduced substantially. Moreover, model simplification could help in explaining and communicating better the results to decision-makers and other stakeholders in the copper system.

5. Conclusions

Scenarios are a powerful and frequently employed means for communicating and characterizing uncertainty in decision making. The literature on scenarios and approaches for the development of scenarios is wide and diverse. Various challenges and issues are acknowledged in this literature, including how to cope with actors with quite diverse worldviews, how to develop scenarios when there is a lack of consensus or agreement between stakeholders, and how to best represent the richness and multiplicity of plausible futures in a finite and small set of scenarios. Scenario discovery is a technique that can be employed to cope with some of these challenges. In particular, scenario discovery can be used to identify a small set of scenarios of interest given a very large and diverse set of uncertainties. Scenario discovery also has the potential to be of use if there is a lack of consensus or a disagreement among the actors, thus helping in the development of scenarios in the presence of deep uncertainty. Scenario discovery could also be used to identify a small set of scenarios that is fully representative, both in terms of behavior and in terms of origins in the multi-dimensional uncertainty space, for a large ensemble of scenarios or specific subsets of the large ensemble.

In the literature on scenario discovery, scenarios have been treated as future states of the world in which a candidate policy was performing poorly (Bryant & Lempert, 2010; Groves & Lempert, 2007; Lempert et al., 2006). The future states of the world were generated by a single model with a few uncertain parameters over which a parameter sweep was executed. We extended this approach in multiple ways. First, we broadened the notion of scenarios in the context of scenario discovery to cover any region of the model outcome space that is of interest for some reason. Moreover, we were interested in the dynamics over time instead of a future state of the world. The introduction of dynamics over time necessitated the need for the clustering of time series based on their dynamic behavior pattern, thus allowing the use of the same machine learning techniques that had already been used successfully in the context of scenario discovery. Second, we applied scenario discovery to the case of copper scarcity, where there was deep uncertainty about how to model an aspect of the system, resulting in the development of two distinct models. Moreover, these two models jointly cover a significantly larger number of uncertainties than previously reported applications of scenario discovery (Bryant & Lempert, 2010; Groves & Lempert, 2007).

In the copper case, we used a time series clustering approach to identify behaviorally distinct clusters. In light of the identified behaviors, we specified a rule for classifying results as being of interest or not. For both models, we identified where in the model input space these behaviors of interest originated. We showed that undesirable price fluctuations can result from a mismatch of capacity in the copper production chain. In case of the bottom up model, this creates problems if there is a shock to the system due to a rapid increase of copper scrap due to recycling. Another important contributing factor to undesirable fluctuations of the price is the situation where copper demand starts to drop, while the expectation remains that copper demand will grow. The case showed how scenario discovery can be used to identify a small subset of uncertain factors that is responsible for a particular dynamic over the time period of interest. However, there is room for improvement of the results. In particular, the approach for identifying the combinations of uncertainties responsible for a particular dynamic can be improved, for example by fine tuning it in such a way that it searches for a small set of uncertainties that have both a high coverage and high density.

References

Abdel Sabour, S. A., & Poulin, R. (2010). Mine expansion decisions under uncertainty. *International Journal of Mining, Reclamation and Environment, 24*(4), 340-349. doi:10.1080/17480931003664991

Angerer, G., Mohring, A., Marscheider-Weidemann, F., & Wietschel, M. (2010). *Kupfer für Zukunftstechnologien*. Retrieved from Karlsruhe: https://www.isi.fraunhofer.de/content/dam/isi/dokumente/cce/2010/Bericht_FSEM_Cu.pdf

Auping, W. L. (2011). *The Uncertain Future of Copper: An Exploratory System Dynamics Model and Analysis of the global copper system in the next 40 years*. Retrieved from Delft: http://simulation.tbm.tudelft.nl/RESEARCH/Auping2011small.pdf

Bankes, S. C. (1993). Exploratory Modeling for Policy Analysis. *Operations Research, 41*(3), 435-449. doi:10.1287/opre.41.3.435

Ben-Haim, Y. (2006). *Info-Gap Decision Theory. Decisions Under Severe Uncertainty* (2nd ed.): Academic Press.

Börjeson, L., Höjer, M., Dreborg, K.-H., Ekvall, T., & Finnveden, G. (2006). Scenario types and techniques: Towards a user's guide. *Futures, 38*(7), 723-739. doi:10.1016/j.futures.2005.12.002

Bradfield, R., Wright, G., Burt, G., Cairns, G., & Van der Heijden, K. (2005). The origins and evolution of scenario techniques in long range business planning. *Futures, 37*(8), 795-812. doi:10.1016/j.futures.2005.01.003

Bryant, B. P., & Lempert, R. J. (2010). Thinking inside the box: A participatory, computer-assisted approach to scenario discovery. *Technological Forecasting & Social Change, 77*, 34-49. doi:10.1016/j.techfore.2009.08.002

Chong, I.-G., & Jun, C.-H. (2008). Flexible patient rule induction method for optimizing process variables in discrete type. *Expert Systems with Applications, 34*(4), 3014-3020. doi:10.1016/j.eswa.2007.05.047

Dalal, S., Han, B., Lempert, R. J., Jaycocks, A., & Hackbarth, A. (2013). Improving scenario discovery using orthogonal rotations. *Environmental Modelling & Software, 48*, 49-64. doi:10.1016/j.envsoft.2013.05.013

European Environmental Agency. (2009). *Looking back on looking forward: a review of evaluative scenario literature*. Retrieved from Luxemburg: https://www.eea.europa.eu/publications/looking-back-on-looking-forward-a-review-of-evaluative-scenario-literature

Ford, D. N. (1999). A behavioral approach to feedback loop dominance analysis. *System Dynamics Review, 15*(1), 3-36. doi:10.1002/(SICI)1099-1727(199921)15:1<3::AID-SDR159>3.0.CO;2-P

Friedman, J. H., & Fisher, N. I. (1999). Bump hunting in high-dimensional data. *Statistics and Computing, 9*(2), 123-143. doi:10.1023/A:1008894516817

Goodwin, P., & Wright, G. (2010). The limits of forecasting methods in anticipating rare events. *Technological Forecasting and Social Change, 77*(3), 355-368. doi:10.1016/j.techfore.2009.10.008

Gordon, R. B., Bertram, M., & Graedel, T. E. (2006). Metal stocks and sustainability. *Proceedings of the National Academy of Sciences, 103*(5), 1209-1214. doi:10.1073/pnas.0509498103

Gordon, R. B., Bertram, M., & Graedel, T. E. (2007). On the sustainability of metal supplies: A response to Tilton and Lagos. *Resources Policy, 32*(1-2), 24-28. doi:10.1016/j.resourpol.2007.04.002

Gordon, R. B., Koopmans, T. C., Nordhaus, W. D., & Skinner, B. J. (1987). *Towards a New Iron Age? Quantitative Modeling of Resource Exhaustion*. Cambridge, Massachusetts: Harvard University Press.

Groves, D. G., & Lempert, R. J. (2007). A new analytic method for finding policy-relevant scenarios. *Global Environmental Change, 17*, 73-85. doi:10.1016/j.gloenvcha.2006.11.006

Han, J., & Kamber, M. (2001). *Data Mining: Concepts and Techniques* (1st ed.). San Francisco, CA: Morgan Kaufmann.

ICSG. (2010b). *The World Copper Factbook*. Lisbon.

Index Mundi. (2011, 2012

2012/03/03/). Copper, grade A cathode Monthly Price - US Dollars per Metric Ton. *Commodities, Metals.* Retrieved from http://www.indexmundi.com/commodities/?commodity=copper&months=360

JORC. (2004). *The JORC Code - Australasian Code for Reporting of Exploration Results, Mineral Resources and Ore Reserves*. Retrieved from

Keogh, E., & Kasetty, S. (2003). On the Need for Time Series Data Mining Benchmarks: A Survey and Empirical Demonstration. *Data Mining and Knowledge Discovery, 7*(4), 349-371. doi:10.1023/A:1024988512476

Kleijn, R., & van der Voet, E. (2010). Resource constraints in a hydrogen economy based on renewable energy sources: An exploration. *Renewable and Sustainable Energy Reviews, 14*(9), 2784-2795. doi:10.1016/j.rser.2010.07.066

Kwakkel, J. H., Walker, W. E., & Marchau, V. A. W. J. (2010). Classifying and communicating uncertainties in model-based policy analysis. *International Journal of Technology, Policy and Management, 10*(4), 299-315. doi:10.1504/IJTPM.2010.036918

Lane, D. C. (2008). The emergence and use of diagramming in system dynamics: a critical account. *Systems Research and Behavioral Science, 25*(1), 3-23. doi:10.1002/sres.826

Lempert, R. J., Bryant, B. P., & Bankes, S. C. (2008). *Comparing Algorithms for Scenario Discovery*. Retrieved from Santa Monica, CA.: http://www.rand.org/pubs/working_papers/WR557.html

Lempert, R. J., Groves, D. G., Popper, S. W., & Bankes, S. C. (2006). A General, Analytic Method for Generating Robust Strategies and Narrative Scenarios. *Management Science, 52*(4), 514-528. doi:10.1287/mnsc.1050.0472

Lempert, R. J., Popper, S. W., & Bankes, S. C. (2003). *Shaping the Next One Hundred Years: New Methods for Quantitative, Long-Term Policy Analysis* (MR-1626). Retrieved from Santa Monica, CA, USA: http://www.rand.org/pubs/monograph_reports/2007/MR1626.pdf

Lossin, A. (2005). Copper *Ullman's Encyclopedia of Industrial Chemistry*. Weinheim: Wiley-VCH Verlag GmbH & Co. KGaA.

McInerney, D. J., Lempert, R., & Keller, K. (2012). What are robust strategies in the face of uncertain climate threshold responses? *Climatic Change, 112*(3-4), 547-568. doi:10.1007/s10584-011-0377-1

McKelvey, V. E. (1973). *Mineral Resource Estimates and Public Policy*. Retrieved from

Meadows, D. H., Richardson, J., & Bruckmann, G. (Eds.). (1982). *Groping in the dark*. Bristol: J.W. Arrowsmith Limited.

Popper, S. W., Griffin, J., Berrebi, C., Light, T., & Daehner, E. M. (2009). *Natural Gas and Israel’s Energy Future: A Strategic Analysis Under Conditions of Deep Uncertainty*. Retrieved from Santa Monica, CA: http://www.rand.org/pubs/technical_reports/TR747.html

Rademaker, M., & Kooroshy, J. (2010). *The Global Challenge of Mineral Scarcity.* Paper presented at the Enriching the Planet, Empowering Europe - Optimising the use of natural resources for a more sustainable economy, The Hague, The Netherlands.

Schwartz, P. (1991). *The Art of the Long View: Planning for the Future in an Uncertain World* (1st ed.). Chichester: Doubleday.

Sibson, R. (1973). SLINK: An optimally efficient algorithm for the single-link cluster method. *The Computer Journal, 16*(1), 30-34. doi:10.1093/comjnl/16.1.30

Skinner, B. J. (1976). A Second Iron Age Ahead? *American Scientist, 64*(3), 258-269.

Sterman, J. D. (2000). *Business Dynamics: Systems Thinking and Modeling for a Complex World*. New York: McGraw.

Sun, B. X., Nie, Z. R., Liu, Y., Wang, Z. H., & Gong, X. Z. (2010). Exergy-Based Model of the Depletion of Mineral Resources. *Materials Science Forum, 650*, 1-8. doi:10.4028/www.scientific.net/MSF.650.1

Tapscott, D., & Williams, A. D. (2006). *Wikinomics: how mass collaboration changes everything* (1st ed.). New York City, NY: Portfolio.

Theodoridis, S. (2003). Pattern Recognition. In H. Bidgoli (Ed.), *Encyclopedia of Information Systems* (pp. 459-479). New York: Elsevier.

Tilton, J. E., & Lagos, G. (2007). Assessing the long-run availability of copper. *Resources Policy, 32*(1-2), 19-23. doi:10.1016/j.resourpol.2007.04.001

van 't Klooster, S. A., & van Asselt, M. B. A. (2006). Practising the scenario-axes technique. *Futures, 38*(1), 15-30. doi:10.1016/j.futures.2005.04.019

van der Heijden, K. (1996). *Scenarios. The Art of Strategic Conversation*. Chichester: John Wiley & Sons Ltd.

van der Pas, J. W. G. M., Walker, W. E., Marchau, V. A. W. J., Van Wee, G. P., & Agusdinata, D. B. (2010). Exploratory MCDA for handling deep uncertainties: the case of intelligent speed adaptation implementation. *Journal of Multi-Criteria Decision Analysis, 17*(1-2), 1-23. doi:10.1002/mcda.450

Van Rossum, G. (1995). *Python Reference Manual*. Retrieved from https://ir.cwi.nl/pub/5008/05008D.pdf

Ventana Systems Inc. (2010). *Vensim DSS Reference Supplement*. Retrieved from

Ventana Systems Inc. (2011). *Vensim Reference Manual*. Retrieved from

Warren Liao, T. (2005). Clustering of time series data—a survey. *Pattern Recognition, 38*(11), 1857-1874. doi:10.1016/j.patcog.2005.01.025

Weisstein, E. W. (2012). Covariance Matrix. *MathWorld—A Wolfram Web Resource.* Retrieved from http://mathworld.wolfram.com/CovarianceMatrix.html

Wouters, H., & Bol, D. (2009). *Material Scarcity. An M2i study*. Retrieved from Delft: http://www.m2i.nl/images/stories/m2i%20material_scarcity%20report.pdf

Yücel, G. (2012). *A novel way to measure (dis)similarity between model behaviors based on dynamic pattern features*. Paper presented at the 30th International Conference of the System Dynamics Society, St. Gallen, Switzerland.

Yücel, G., & Barlas, Y. (2011). Automated parameter specification in dynamic feedback models based on behavior pattern features. *System Dynamics Review, 27*(2), 195-215. doi:10.1002/sdr.457

4 The geopolitical impact of the shale revolution

Many societal challenges simultaneously take place at different geographical levels. For example, problems linked to the availability of natural resources take place on global markets, in user countries, and in producer countries. This makes it hard to develop a single model capturing all relevant aspects of the problem at hand. If different models are made to accommodate these different scopes, it becomes necessary to transfer run information – based on the shared problem – between models.

This paper addresses two methodological issues, which originate in the fact that the problem addressed in this paper plays on two different geographic levels: the global energy market and national economies. The first issue is how to change the model scope when dealing with the different geographic levels. We addressed this by making use of two models that were used in series, rather than the parallel multi-model use introduced in the copper cases. Serial multi-model use combined with exploratory modelling is still very little used, and as such this case is a good example of its use. The second issue is the selection of scenarios from the global model to feed into the national models. Scenario Discovery was impractical, as we were looking for representative examples of behaviour patterns over time and not primarily spaces of the input space which explained those behaviours. In this case, we approached the scenario selection by exploring the edges of the output space generated by the global model. In practice, this meant that we looked at the oil prices in situations with the highest shares of particular primary energy types in the energy mix, combined with a number of runs which showed highest oil price volatility.

When communicating the findings to the stakeholder group supervising the research process, they struggled with evaluating the model results. While the model results showed an almost unavoidable dip in oil prices due to the shale revolution, many experts and stakeholders believed during presentations of this research in 2013 and 2014, that declining oil prices were not a plausible future. As reality corroborated our findings after July 2014 with a strong decline in oil price levels, this case is a good illustration of the added value of EMA in finding plausible behaviours not or hardly considered by domain experts.

In the synthesis, the choices made here are discussed with regard to the impact of EMA on model development, the impact of complexity of models on Scenario Discovery, and the costs and benefits of exploratory modelling.

This paper was originally published as:

Auping, W. L., Pruyt, E., De Jong, S., & Kwakkel, J. H. (2016). The geopolitical impact of the shale revolution: Exploring consequences on energy prices and rentier states. *Energy Policy, 98*(2016), 390-399. doi:10.1016/j.enpol.2016.08.032

The geopolitical impact of the shale revolution: Exploring consequences on energy prices and rentier states

Willem L. Auping, The Hague Centre for Strategic Studies, The Netherlands; Policy Analysis Section, Faculty of Technology, Policy and Management, Delft University of Technology, The Netherlands

Erik Pruyt, Policy Analysis Section, Faculty of Technology, Policy and Management, Delft University of Technology, The Netherlands

Sijbren de Jong, The Hague Centre for Strategic Studies, The Hague, The Netherlands

Jan H. Kwakkel, Policy Analysis Section, Faculty of Technology, Policy and Management, Delft University of Technology, The Netherlands

Highlights

- We quantitatively explore geopolitical consequences of the shale gas revolution
- We use a multi-model approach to generate and use energy price scenarios
- Simulations show that current low oil prices could be part of a hog cycle
- The shale gas boom was an early warning for the drop in oil prices
- Low prices due to shale gas can reduce internal stability in rentier states

Abstract

While the shale gas revolution was largely a US' affair, it affects the global energy system. In this paper, we look at the effects of this spectacular increase in natural gas, and oil, extraction capacity can have on the mix of primary energy sources, on energy prices, and through that on internal political stability of rentier states. We use two exploratory simulation models investigating the consequences of the combination of both complexity and uncertainty in relation to the global energy system and state stability. Our simulations show that shale gas developments could be seen as part of a long term hog-cycle, with a short term drop in oil prices if unconventional supply substitutes demand for oil. These lower oil prices may lead to instability in rentier states neighbouring the EU, especially when dependence on oil and gas income is high, youth bulges are present, or buffers like sovereign wealth funds are too limited to bridge the negative economic effects of temporary low oil prices.

Keywords Shale revolution; Energy mix; Energy prices; State stability; System Dynamics; Exploratory Modelling

1. Introduction

In recent years, a spectacular rise in natural gas extraction capacity from unconventional resources has dramatically changed the US' energy landscape, turning the country into a natural gas exporter. This development is often referred to as the 'shale gas revolution' and was made possible by the process of hydraulic fracturing, or 'fracking'. As a consequence of the shale gas revolution, US' gas prices have dropped significantly, giving a competitive advantage to the US' industry. The spectacular rise in extraction of shale oil resources only adds to that advantage.

The shale gas revolution was thus far largely an US' affair, as outside of Northern America hardly any commercial exploitation of shale gas resources took place. This can be primarily explained by institutional differences between the US and other countries (Kuuskraa, Stevens, & Moodhe, 2013; Tian, Wang, Krupnick, & Liu, 2014, p. 11), as significant technically recoverable shale resources can be found outside of North America (Kuuskraa et al., 2013). Notwithstanding the fact that the shale revolution has not spread across the world (Melikoglu, 2014), the large-scale extraction of shale deposits has affected the global energy system through LNG trading, which is still a minor part of global natural gas trade (BP, 2015), and through substitution of other, easier transportable primary energy sources. The impact of the shale revolution on global energy markets is the starting point of this research.

Research regarding the impact of the shale revolution mainly focussed on direct extraction effects like the effects of shale gas drilling on the environment (e.g., Baranzelli, Vandecasteele, Ribeiro Barranco, & Mari i Rivero, 2015; Jenner & Lamadrid, 2013; Kargbo, Wilhelm, & Campbell, 2010; Meng & Ashby, 2014; Olmstead, Muehlenbachs, Shih, Chu, & Krupnick, 2013), the public support for shale gas and fracking (e.g., Boudet et al., 2014; Jaspal, Nerlich, & Lemańczyk, 2014; Perry, 2012), and the impact on the local economy (e.g., Asche, Oglend, & Osmundsen, 2012; Kinnaman, 2011; J. Lee, 2015). In some research, the economic impact was related to energy security (e.g., Jaspal et al., 2014; Richter & Holz, 2015; Victor, Nichols, & Balash, 2014). Others studied the impact of shale gas exploration on energy prices, mainly for oil and gas (e.g., Asche et al., 2012; De Silva, Simons, & Stevens, 2016). Asche et al. (2012) acknowledged that it would be interesting to look at the interplay with coal, another primary energy source, which we do in this paper. However, to the best of our knowledge, more indirect consequences of the shale revolution have not been investigated.

One of these potential indirect effects is the impact of the shale revolution on intra-state stability of major oil and gas exporting countries, also referred to as 'rentier states' (Mahdavy, 1970), through changing oil and gas prices. Price fluctuations may have consequences for the financial-economic stability of rentier states due to the dependence on resource rents (World Bank, 2011) for supporting the economy and government spending. That is, fluctuations in resource prices may influence the development of the local economies in oil and gas exporting countries. In turn, worsening economic conditions are known to have an impact on population discontent, potentially leading to intra-state instability (Collier & Hoeffler, 2004; Ross, 2004). Similar indirect effects may occur due to structural changes in the global energy system induced by climate mitigation policies. However, these effects are complex and uncertain.

Both the global energy system and the relation between resource income and instability are highly complex and deeply uncertain. Feedback effects add to dynamic complexity (Sterman, 2000). An example of a feedback effect in the global energy system is the interaction between supply and demand which results in resource price dynamics. On a country scale, decreasing resource prices may lead to increasing unemployment and a reduction of purchasing power, which may cause frustration among the population and reduce internal stability. This feedback loop is closed if state instability in turn affects resource extraction. Both the global system and national systems are also characterized by deep uncertainty (Lempert et al., 2003). Situations are deeply uncertain if they are characterized by important presently irreducible uncertainties related to how issues could or should be modelled, likelihoods of inputs and outcomes, and the desirability of outcomes. To give examples: on a global scale, the strength of the feedback effect between prices and demand is deeply uncertain, while on a national scale, the influence of population discontent on a country’s polity is also deeply uncertain.

Complexity and uncertainty impede mental simulation of both the global energy system and state stability (Sterman, 1994). Quantitative simulation may enable one to deal with these issues though. Since the 1950s, modelling and simulation approaches have been developed and used to support policy-makers and decision-makers addressing complex issues. Since the early 1990s (Bankes, 1993), model-based methodologies and techniques have been developed to simulate sets of models under deep uncertainty.

In this paper, we use simulation models to explore the indirect consequences of the shale revolution on the global energy system and rentier state stability. For this purpose, we apply a Systems-of-Systems (DeLaurentis & Callaway, 2004) multi-model approach for dealing with complexity and uncertainty. We use a global energy-mix model for supply, demand, and trade of, and substitution between six primary energy sources to generate oil and gas price scenarios. In these scenarios, the focus lies on price scenarios that fall outside the scope of more traditional forecasts of energy prices using a base-case (e.g., IEA, 2012). These scenarios are subsequently used as input for a country-stability model, focussing on economic discontent (i.e., 'greed' in Collier & Hoeffler, 2004). As such, the price scenarios are used for ‘stress testing’ rentier state country stability, more specific those in the vicinity of the European Union (EU). These countries are Algeria, Azerbaijan, Kazakhstan, Qatar, Russia, and Saudi Arabia.

The setup of this paper is as follows. In section 2, we explain the use of Exploratory Modelling and System Dynamics in this study, the model structures of the energy-mix model and the country-stability model, and the metrics for choosing 14 price scenarios. Based on these scenarios, we present the results on country stability by taking Algeria and Russia as examples in section 3. Finally, we discuss the results of this approach in section 4, and draw conclusions regarding the geopolitical consequences of the shale gas revolution in section 5.

2. Methods

In this research, we use an exploratory modelling scenario approach. First, we simulate and investigate the consequences of the shale gas revolution to generate global oil and regional gas price scenarios. Second, a subset of these price scenarios is used to stress-test intra-stability of rentier states in the vicinity of Europe. In this section, we introduce this model-based scenario approach (section 2.1), as well as the modelling and simulation method (section 2.2), and the two models used in this research (section 2.3). At the end of this section we explain the research setup in more detail (section 2.4).

2.1. Exploratory Modelling

Exploratory Modelling is a research methodology that uses computational experiments to analyse deeply uncertain issues (Bankes, 1993; Bankes, Walker, & Kwakkel, 2013; Kwakkel & Pruyt, 2013; Lempert et al., 2003). It consists of a set of the development of plausible quantitative simulation models and associated uncertainties, the process of exploiting the information contained in such a set through a large number of computational experiments, the analysis of the results of these experiments, and the testing of promising policies for policy robustness (Bankes, 1993).

In exploratory modelling, models are used to generate a wide variety of what-if scenarios, which is an important use case of simulation models (Oreskes, Shrader-Frechette, & Belitz, 1994). These what-if scenarios are usually generated such that they comprehensively cover presently irreducible uncertainties. Exploratory modelling, therefore, does not focus on generating a base case, but instead on generating a bandwidth of plausible futures, including the circumstances (i.e., ranges of specific uncertainties) for which these occur.

2.2. System Dynamics

The simulation models used in this research are System Dynamics simulation models. System Dynamics (SD) is a modelling and simulation method to describe, model, simulate, and analyse dynamically complex issues or systems (Forrester, 1961; Pruyt, 2013; Sterman, 2000). The SD approach was first proposed and developed by Jay W. Forrester in the late 1950s. SD aims to provide a holistic and systemic view of an issue under study and its interconnections to its environment, and simulate and analyse the resulting system dynamics over time. More specifically, SD is a method for modelling and simulating dynamically complex systems or issues characterized by feedback and accumulation effects (Sterman, 2000).

Together, feedback and accumulation effects generate dynamically complex behaviour both inside SD models, and, so it is assumed by System Dynamicists, in real systems (Pruyt, 2015). Using SD models can, therefore, be useful for dealing with complex systems characterized by important feedback and accumulation effects. SD modelling is mostly used to model core system structures or core structures underlying issues, to simulate their resulting behaviour, and to study the link between the underlying causal structure of issues and models and the resulting behaviour. SD models, which are mostly relatively small and manageable, can be used for experimental or exploratory purposes too.

There are many SD models regarding energy systems. Well-known examples are the Limits to Growth studies (Meadows et al., 1972), studies regarding national energy transitions (Naill, 1977, 1992; Sterman, 1981), power plant construction and electricity generation (A. Ford, 1999), and externalities of energy economics (Fiddaman, 1997), but there are also more recent examples (e.g., Chyong Chi, Nuttall, & Reiner, 2009; Osorio & van Ackere, 2016). In two cases this included the use of SD for exploratory modelling and the design of robust policies (Eker & van Daalen, 2015; Hamarat et al., 2013; Hamarat et al., 2014). While Hosseini and Shakouri G (2016) used oil price scenarios as input to an SD model, to our knowledge there are no SD study beyond our line of research in which energy models are used to generate scenarios related to energy price developments.

There are also SD models regarding social unrest. For example, Wils, Kamiya, and Choucri (1998) presented a model to simulate and assess the development of internal and external pressure related to resource use. Further, Anderson (2011) used an SD model for looking at the effects of counterinsurgency policy in relation to public support and other factors. Finally, Pruyt and Kwakkel (2014) simultaneously used three SD models to simulate the rise of activism, extremism, and terrorism. In none of these models, however, external price scenarios were used for ‘stress testing’ state stability.

2.3. Model descriptions

In this research, we use two SD models. Scenarios developed with the first model provide input for the second model (Fig. 1). We now discuss these models on a high level of aggregation (model descriptions are provided in the supplementary material). Both models were extensively verified and validated by means of partial model tests, unit checks, sessions with stakeholders, and extreme value tests. In order to assess the effects of long delays in the system, such as developments in extraction capacity in the energy-mix model or demographic effects in the country-stability model, we simulated both models for the time period between 2010 and 2050.

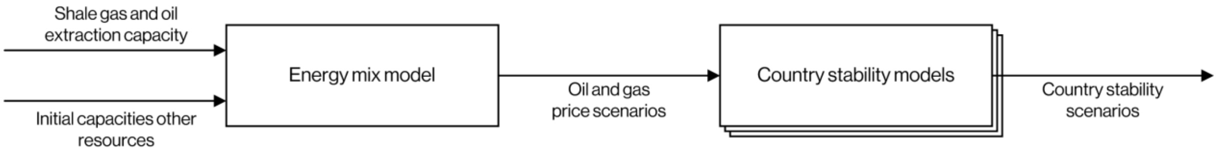


Fig. 1. Research design for this study, containing two SD models. The country-stability model is parametrised for six different countries. Model parametrisations also includes other uncertainties besides the inputs shown in this figure.

2.3.1. The energy-mix model

The energy-mix model is subdivided in 5 sub-models, which are interlinked (Fig. 2). We look at the demand development, supply development, prices of the different primary energy sources, costs development of the supply, and trade between the different regions. We included six primary energy sources (i.e., oil, natural gas, coal, nuclear, biofuels, and other renewables), in line with the definitions provided by the EIA (2015). The development of demand, supply, and the prices of the six energy sources are important given the feedback effects connecting supply and demand through prices. The extraction costs sub-model is important for simulating the effects of depletion on extraction costs and the development of the costs of renewables. Finally, as a greater availability of natural gas may lead to a larger share of LNG entering global markets, it is important to consider trade between the different regions of the tradable resources, in this case gas (LNG), oil, coal, and biofuels. Trade of the two remaining primary energy sources (i.e., nuclear and other renewables) is thus not considered here.

In the model, 4 different regions are defined: Northern America (i.e., US and Canada), Europe and adjacent regions (i.e., Europe, non-European CIS, Middle East, and North Africa), the Far East (i.e., China, India, Japan, and South Korea), and the rest of the world. The first two regions are grouped bearing the availability of overland gas pipelines in mind. The Far East, which is presently a major user of LNG (BP, 2015), is included as a separate region given the fact that pipeline infrastructure to other regions is very limited. The effects of political instability on energy supply is not considered in the energy-mix model. Furthermore, policy measures aimed at changing the composition of the energy mix are considered only as a driver for the development of renewable energy capacity.

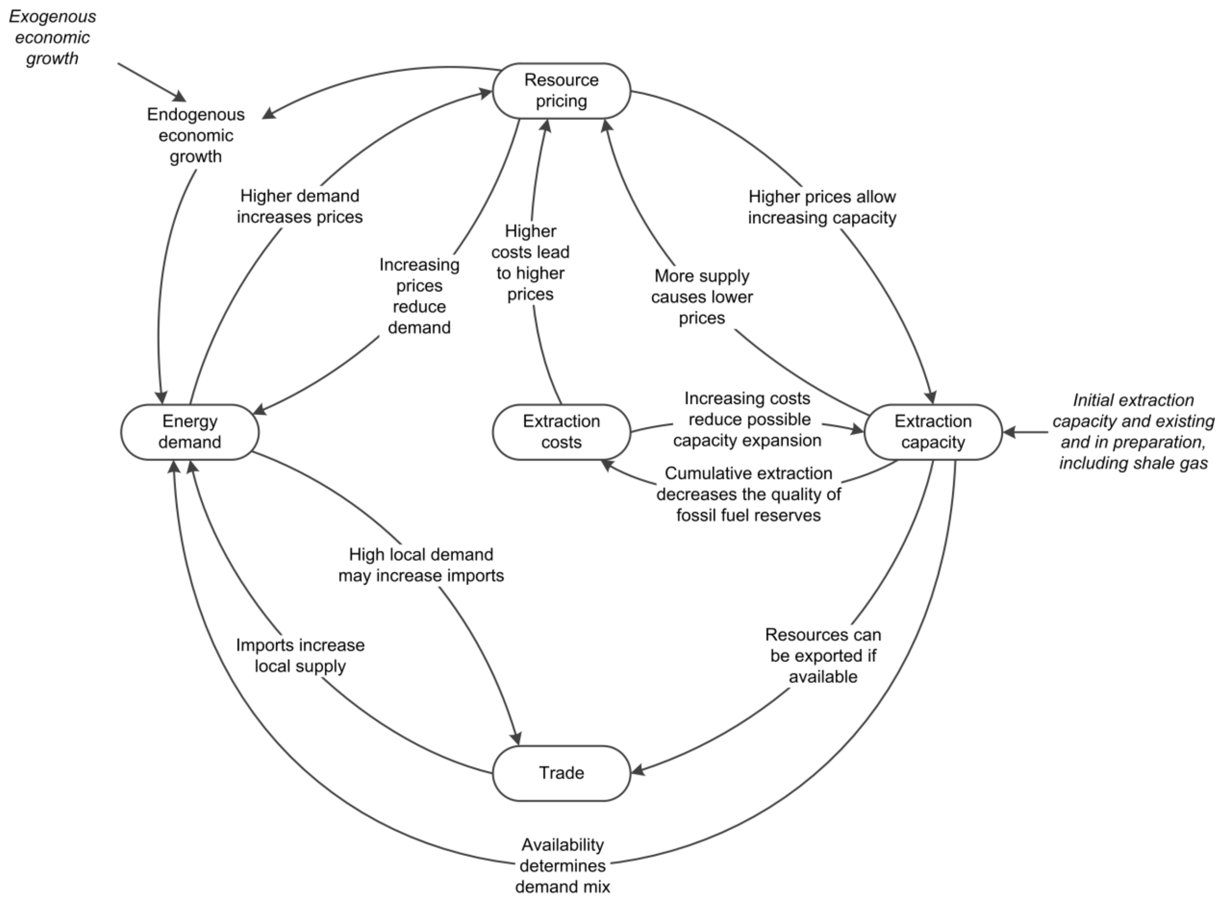


Fig. 2. Sub-system diagram (Morecroft, 1982; Sterman, 2000, pp. 99-102) of the energy-mix model. Sub-models are displayed in a rounded box. Important initial conditions are shown in *italics.*

2.3.2. Country-stability model

The country-stability model also consists of 5 interlinked sub-models (Fig. 3). These sub-models group variables related to the development of resources, the economy, the population, national institutions, and instability. In the resources sub-model of this model, the development of the extraction of oil and gas is a function of endogenous cost developments and exogenous energy price scenarios generated with the energy-mix model. The economy is influenced by resource income (*i.e.*, resource prices times resource extraction capacities), an exogenous economic growth factor constant over the run time, and endogenous negative economic effects of political instability. The economic sub-model also contains modules for the available workforce and work, and for purchasing power. When the amount of work available is lower than the workforce, the rest of the workforce is unemployed. In this way, male youth unemployment can be calculated, which is a well-known factor causing frustration and internal instability (Cincotta, Engelman, & Anastasion, 2003; Urdal, 2006).

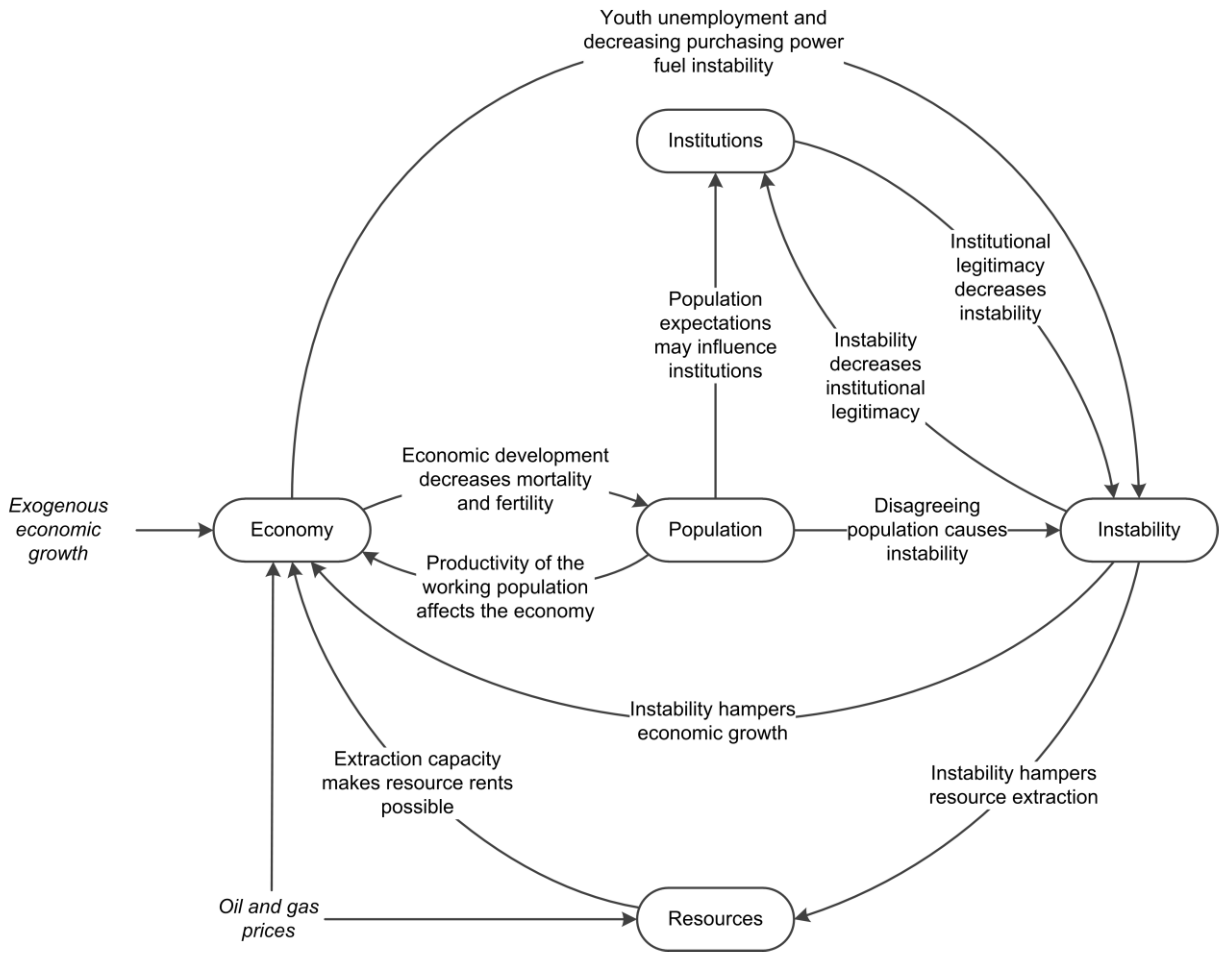


Fig. 3. Sub-system diagram of the country-stability model. Sub-models are displayed in a rounded box. External trends and important initial conditions are shown in *italics*.

The population sub-model contains an endogenous population development model within which fertility and mortality are a function of GDP per capita, and the population aging chain is sub-divided in 5-year cohorts. Next to the population composition, we calculate the education level of the population both for the average population and the young population. We assume that rising education levels increase the democratic expectations of the population, although we treat the exact relation between education and polity expectations as deeply uncertain.

In the institutions sub-model, we take the government type into account following the Polity IV scoring system (Marshall, Gurr, & Jaggers, 2014). In the Polity IV score, countries' polities are measured between -10 (i.e., fully autocratic) and +10 (i.e., fully democratic). Statistical analysis shows that autocracies are five times as stable as countries with a polity score of 0, and that democracies are 10 times as stable as countries with a polity score of 0 (i.e., partial democracies or anocracies) (Marshall et al., 2014). In our model, the government type follows the expectations of the population relative to the average educational level when internal tension is low, and becomes more autocratic when tension is high. The institutions sub-model further captures government legitimacy as function of absence of violence, and government financial reserves and expenditure. If financial expenditure is too high, and government debt becomes untenable, existing food and fuel subsidies are cut. On the other hand, if there is abundant income, sovereign wealth funds (Sovereign Wealth Fund Institute, 2013) may be developed, which increase the resilience of the nation by acting as a buffer for temporarily lower resource income.

Finally, the instability sub-model contains an ageing chain to capture the level of frustration of citizens (i.e., those who support the government , non-activist opposition, activist opposition, and extremist opposition) in line with other SD models on instability and terrorism (Anderson, 2011; Pruyt & Kwakkel, 2014; Sterman, 2000). Origins of frustration are, in this model, economic in nature (e.g., unemployment, especially male youth unemployment, and purchasing power). This corresponds with the 'greed' aspect of state instability (Collier & Hoeffler, 2004). The size of the security forces works as balancing factor for civil frustration and instability. The ratio between the strength of the security forces and the extremist cohort of the population is used as proxy for the level of political unrest.

2.4. Research setup

First, we generated 1000 runs with the energy-mix model. The model was integrated with the Runge-Kutta 4 numerical integration method with automatically adjusted step size. Sampling of the 118 uncertainties happened with Latin Hypercube (LH) sampling, assuring that each run covers a different part of the total uncertainty space. While 1000 runs may seem rather limited given the large number of uncertainties, it proved sufficient for the goal of this research, which was to generate a limited number of sufficiently different energy price scenarios, and not to exhaustively explore the complete behavioural space of the energy-mix model.

Table 1

Metrics for selecting oil and gas price scenarios from the energy-mix model runs.

Scenario number	**Scenario metric**
1	Highest coal share
2	Highest gas share
3	Highest other renewables share
4	Highest oil share
5	Highest biofuels share
6	Volatile scenario
7	Volatile scenario
8	Volatile scenario

We selected eight scenarios representing plausible, yet extreme, situations in terms of the energy mix from the 1000 runs. Each scenario represents an internally consistent, plausible future for how energy mix, absolute and relative regional demand shares, and energy prices may evolve between 2010 and 2050. The selection criteria for these scenarios (Table) were chosen to maximise the bandwidth of potential significant demand fluctuations, leading to a relatively broad selection of price scenarios. To extend this bandwidth further, we also included the 3 scenarios with the most volatile oil price dynamics.

Second, 100 separate cases created with the country-stability model with the same integration method as used for the energy-mix model. These cases were parametrised using LH sampling for 60 general and 13 country specific uncertainties. A selection of important country specific uncertainties can be found in Table 2. Again, the number of runs may seem limited given the number of uncertainties. It proved sufficient, however, for the goal of this part of the research, which was to look at the impact of different energy price scenarios across potential development paths of rentier states. Finally, for each of the 100 cases, we tested the effects of the oil and gas prices represented by the eight scenarios, together with two dynamic price scenarios representing a moderately decreasing and increasing oil and gas price, and a reference scenario representing a constant oil and gas price. Therefore, in total 1100 runs were performed per country.

Table 2

Data for the year 2010 for resource turn-over, government type, youth unemployment, and the size of the sovereign wealth funds for the seven rentier states studied. These data functioned as initial conditions in the country-stability model.

Country	**Oil and gas income (% of GDP)**[a]	**Polity IV score**[b]	**Unemployment, youth male (% of male labor force ages 15-24)**[c]	**Sovereign wealth fund (% of GDP)**[d]
Algeria	45	2	19.1	47
Azerbaijan	63	-7	17.0	67
Kazakhstan	37	-6	4.8	48
Qatar	40	-10	11.0	90
Russia	29	4	16.9	11
Saudi Arabia	67	-10	23.6	160

[a] Author's own calculations: defined here as (resource prices [$/bbtu] x resource extraction capacity [bbtu/year])/GDP [$/year]
[b] Marshall, Gurr, and Jaggers (2016)
[c] World Bank (2016)
[d] Sovereign Wealth Fund Institute (2013)

The impact of oil and gas prices on country stability is largely a one way process. However, as instability impacts the development of GDP and resource extraction capacity, this effect is reinforcing. Some other, minor feedback effects exist between stability and economy. Examples are the effects of population size on the fertility and mortality levels, which may cause a deadlock situation with high population growth and too little economic development. Another example is the effect of immigration on the workforce, and the effect of the available workforce on immigration. A last example occurs when the regime is susceptible to the discrepancy between on the one hand the democratic expectations of the population, and on the other hand the present polity. However, instability may again counteract this development if the government reacts in a more autocratic way to a crisis in the country.

Resource prices may influence countries' economies in many, potentially counteracting, ways. For example, price increases have a positive effect on government finances, and create more employment, but they have an adverse effect on purchasing power of the population. It depends on the specific conditions in a country whether the positive or the negative effects will be dominant.

3. Results

In this section, we discuss the results generated with both the models. In section 3.1, we analyse the oil price dynamics generated with the energy-mix models, followed by the selection of the eight oil and gas price scenarios. In section 3.2, we analyse the effects of these scenarios on rentier state stability as simulated with the country-stability model.

3.1. Effects on global energy markets

The oil price dynamics (Fig. 4) generated with the energy-mix model show an initial dip for all runs generated with energy-mix model. Analysis of this dip reveals a clear connection to the overcapacity due to the US' shale gas revolution. While the depth of the dip differs greatly between runs, it shows that, based on the assumptions underpinning the model, it is impossible to not have a temporary decrease in oil prices. The more direct causes of the dip lie in substitution effects. The abundant supply of natural gas finds its way into the global energy market by substituting other primary energy sources. For the energy sources that allow for accumulation of over-production, such as in the case of oil, there can be a more significant long-term price effect than with natural gas alone.

Besides the initial dip, however, the dynamics show a wide variety of behaviours. Roughly half the runs primarily show oscillatory volatile behaviour. Oscillatory volatile behaviour is measured here by the number of times each time series crosses its own mean value combined with the length of the line. Using this measure, runs that are primarily characterized by oscillatory volatile behaviour and runs with long-term increasing oil prices are classified differently, even if the runs with long-term increasing oil prices are characterized by oscillatory behaviours too. Runs that are primarily characterized by this behaviour represent those situations in which the oil price periodically oscillates between relatively low oil prices, roughly half or even less of the 2010 price level and high levels like the 2010 level.

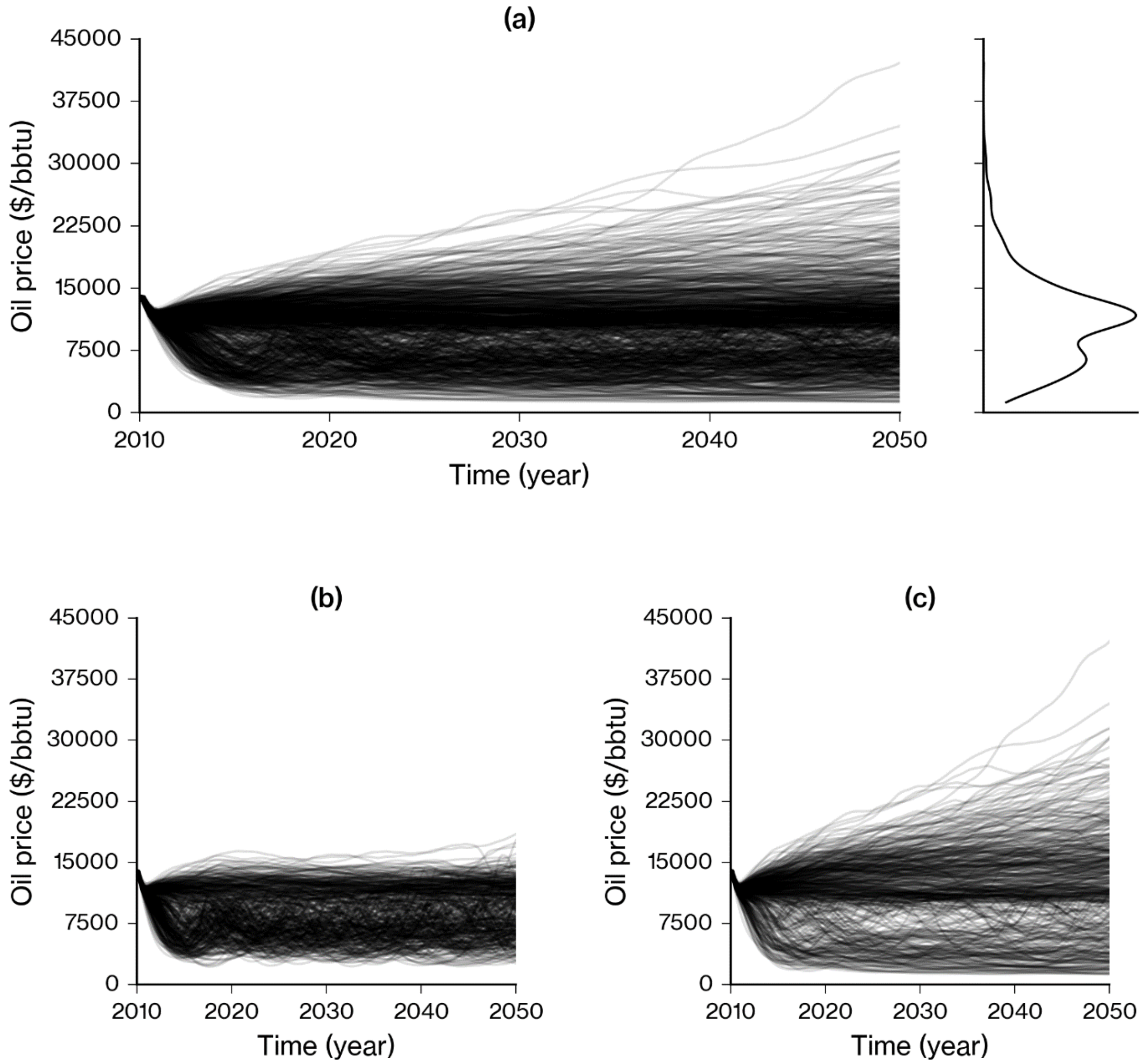


Fig. 4. 1000 runs for the energy-mix model (*a*), the 500 runs with most oscillatory volatile behaviour (*b*), and the 500 runs with least oscillatory volatile behaviour (*c*).

The volatile runs resemble to hog cycles (Hanau, 1928). The periods between the ups and downs are thus related to the delay time for new extraction capacity development, which is found to be generally between eight and fifteen years (MinesQC, 2016). This is consistent with the observed periodicity in the volatile runs. Similar dynamics are not found for the gas price, as massive accumulation of overproduction of gas is assumed to be too expensive. Consequentially, gas prices show far less long-term volatility. Similar dynamics are found for many openly traded resources, but it could be argued that due to OPEC's market power in last decades, hog cycles were less of a problem in the oil market.

The runs with the least oscillatory volatile behaviour mostly show oil prices staying at a price level similar to the 2010 level or rising prices. The rising prices are caused mainly by a combination of continuously rising demand and rising resource extraction costs. Finally, for a last set of runs, the shale gas revolution leads to permanently lower price levels. This behaviour occurs when decoupling – of energy demand and the size of the economy – outgrows economic growth.

We selected eight runs as price scenarios for the second step of the analysis by using the metrics listed in Table . The result of this step for both the oil and the gas price is visible in Fig. 5. By choosing both extreme energy mix scenarios and volatile scenarios, we covered the dynamic behaviour space of all runs relatively well. Finally, the figure shows that the oil price is more volatile than the gas price in all scenarios.

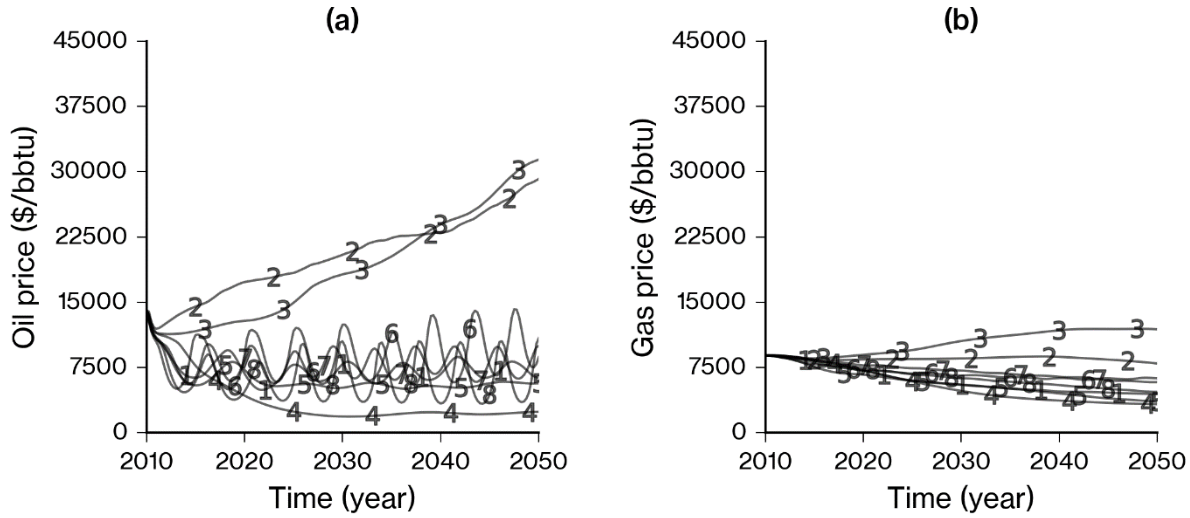


Fig. 5. Selected oil (*a*) and gas (*b*) price scenarios.

3.2. Effects on rentier states

Fig. 6 shows for how many of the 100 cases per country the state stability improved for all time steps (i.e., desirable) or on average (i.e., mostly desirable), or deteriorated for all time steps (i.e., undesirable) or on average (i.e., mostly undesirable) compared to a constant oil and gas price.

All six countries in the analysis become more stable when oil prices increase over time. The importance of the oil price can be explained by the fact that only Qatar receives an equal amount of income from gas extraction, while having the opportunity to expand that share of resource income. All other countries earn considerably more resource income from oil. All countries experience more internal instability when oil prices decrease over time.

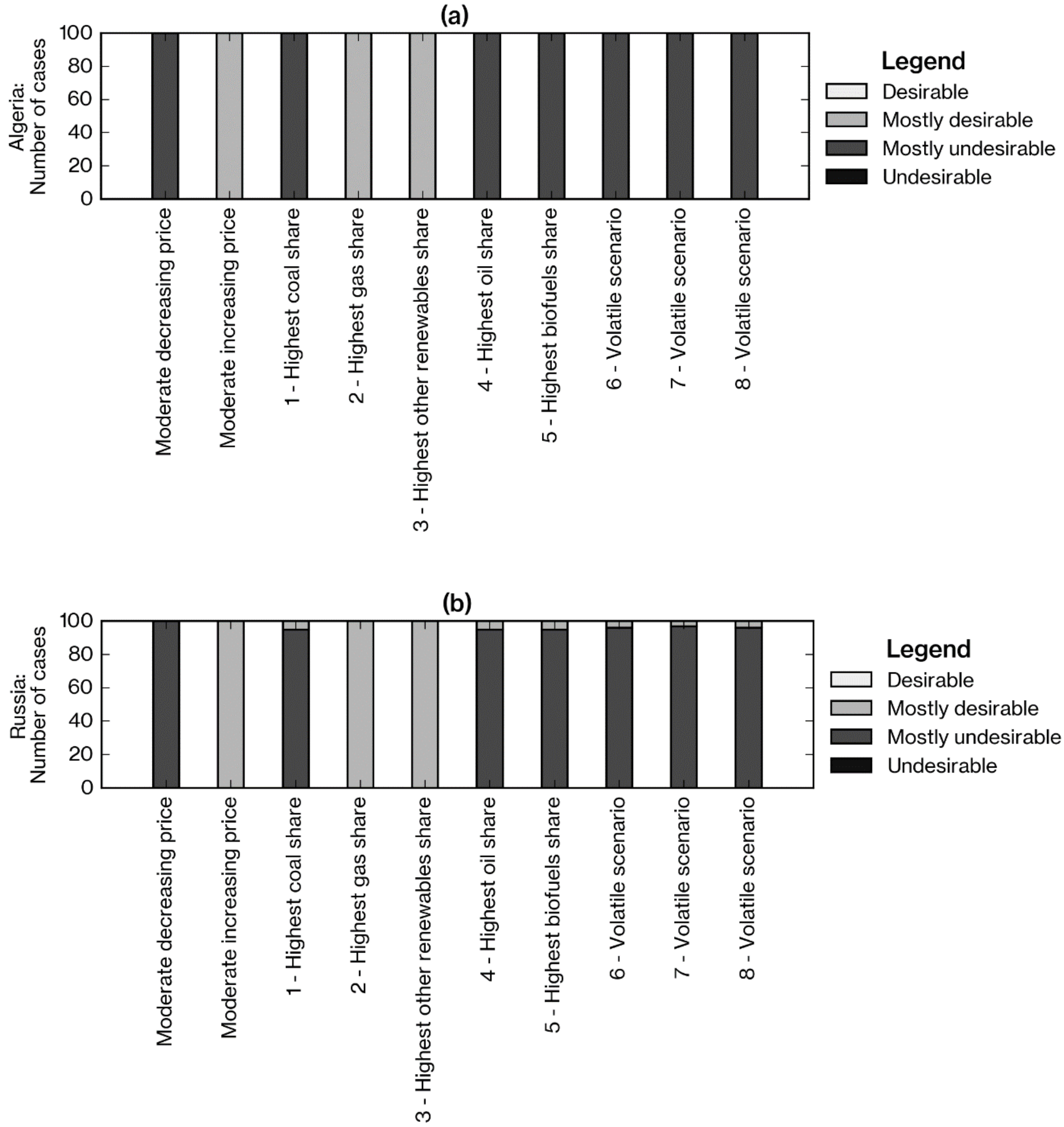


Fig. 6. Effects of the price scenarios on state stability of Algeria (*a*) and Russia (*b*).

Especially Algeria and Russia are vulnerable to the effects of the shale gas revolution, which causes in any case at least a periodic downturn in oil prices. In both countries, the partial democracy is less stable than the more autocratic regimes of Azerbaijan, Kazakhstan, Qatar, and Saudi Arabia. Further, the causes for vulnerability of Russia and Algeria are quite different. In the case of Algeria, it is especially male youth unemployment, combined with a very young population, which makes the country vulnerable. Russia, on the other hand, has a more aged population, but has – compared to the size of the economy – a relatively modest sovereign wealth fund. Russia thus lacks the options to survive one or more prolonged periods of low oil prices.

Although, the other countries analysed in this study are less vulnerable due to either limited youth unemployment (e.g., Kazakhstan and Qatar), or the more significant size of their sovereign wealth fund (e.g., Azerbaijan, Qatar, and Saudi Arabia), they are certainly not immune to the effects of a prolonged downturn in oil prices. The austerity measures in Saudi Arabia (Kerr, 2015), the downgrade of Azerbaijan's credit rating (Agayev, 2016), and the massive currency devaluation in Kazakhstan (Farchy, 2015) are illustrative of the difficulties these countries already experience. However, their overall resilience may prove to be higher than that of Algeria and Russia.

4. Discussion

This research was originally performed in 2013. Earlier versions of the work reported here were presented at the Dutch Ministry of Foreign affairs in the fall of 2013, and at the NATO headquarters in the spring of 2014. The attendees of these presentations found the idea of potentially decreasing oil prices hard to accept, and rather believed that they would continue to rise. Recent price developments, however, corroborate the results we presented in this paper. The belief that oil prices would continue to rise are consistent with the fixed-stock paradigm regarding resource availability, which is in contrast to the opportunity costs paradigm underlying our energy-mix model (Tilton, 1996). As opposing paradigms can be seen as a form of deep uncertainty, it fits our line of research to not oppose ideas like these, but rather to include alternative perspectives and look at our findings as plausible futures, which may inform the development of more robust policies.

The scenarios we generated only provide a limited view on the future of energy prices, as the simulation models used in this research are necessarily simplifications of reality. One example of such a limitation is the fact that we disregarded the existence of strategic reserves in our model. When these are taken into account, the price dynamics, especially the initial shale induced dip, may be delayed over time. This could explain why the real decrease in oil prices only happened from the second half of 2014 on, instead of immediately as in our simulation results.

A further limitation of our research can be found in the fact that we did only very rudimentarily take climate and energy policies into account. Exploring our results, we found that decoupling of economic growth and energy demand has, *ceteris paribus*, a profound negative effect on energy price levels. Again, especially oil prices are vulnerable to these developments. As decoupling due to increased energy efficiency is arguably at the core of climate mitigation policies, it is to be expected that these policies will, similar to the shale-gas revolution, have a profound effect on long-term state stability. As emission targets may also provide a structural change in the global energy system, this may be a far longer-lasting effect than the surge in previously unconventional energy sources. Therefore, countries less vulnerable to periodically lower oil prices due to well-developed sovereign wealth funds, but highly dependent on income from oil and gas, may be especially vulnerable to climate mitigation policy developments. Saudi Arabia is a good example of such a rentier state. These countries should urgently start up a transition process of economic reconversion. Further research with a similar research methodology could increase knowledge about secondary effects similar to those of climate and energy policies.

The country-stability model is, just like the global energy-mix model, a simplification of reality. While much research exists that argues that economic factors as discussed in this paper are relevant and important for understanding the onset of internal instability, there are other factors, not considered here, which may either mitigate or reinforce this effect. An example of a mitigation factor may be nationalism, potentially induced by interstate war. An example of reinforcing factors may be grievance related issues like ethnic diversity, combined with a government that only represents a part of the ethnic groups in a country.

5. Conclusions and policy implications

In this paper we investigated potential indirect effect of the US' shale revolution on state stability of rentier states. To this purpose, we built and used two SD simulation models, one for the global energy mix consisting of six primary energy sources, and one to assess the impact of oil and gas prices on internal stability. The SD models were used to explore the consequences of uncertainties combined with the complexity and non-linear behaviour characterising both energy markets and state stability.

We found that the shale revolution has a periodic negative effect on oil prices in any simulated case. These lower prices can be part of a hog cycle in the energy market, where periods with relatively low oil prices alternate with periods with relatively high oil prices. An example of such a high price period is the pre-2014 period, while current price levels are an example of a period with low prices.

As the surge in shale extraction capacity depended on the period with high prices, and the high price levels were incompatible with the amount of new capacity that became available, the shale gas revolution can be explained as an early warning indicator and partial cause of the 2014-2015 drop in energy prices. This is in itself an indirect effect, as the volatility of oil prices can be explained by a systematic difference between oil trade and gas trade: the surplus of oil is accumulated, while a surplus of natural gas supply is mostly flared. The urge for natural gas suppliers to partly substitute oil demand for their surplus is thus very high.

Low oil prices caused by the shale revolution can have a profound impact on state stability. This effect is caused by the negative economic effects of reduced price levels, which in turn affect male youth unemployment levels and purchasing power. These are underlying causes for population discontent and political instability. Buffers like sovereign wealth funds and immigration workers increase the resilience of countries to these developments.

We found that between the six countries investigated in our research, especially Russia and Algeria are vulnerable to lower oil price levels due to the shale gas boom. The partial democratic ('anocratic') polity of these countries increases their vulnerability. Not all causes are shared by these countries, however, as for Algeria the high male youth unemployment and the youth bulge are detrimental, while for Russia the limited size of the sovereign wealth funds as part of the GDP is problematic to politically survive periodic low oil prices. The other countries were found to be less vulnerable to the lower price effects, mostly as their buffer capacity is larger and their polity more stable (Azerbaijan, Kazakhstan, Qatar, and Saudi Arabia).

The policy implications of these conclusions pertain mostly to the security and international relations domains. We believe, however, that indirect consequences of energy policies should be taken into account in the design of these energy policies. The increased turmoil around Europe's borders is already a difficult policy problem. While this has no direct link to energy policies in itself, the plausibility of energy price related unrest does mean that investment in measures for dealing with these circumstances, both militarily and non-militarily, should be taken sooner rather than later.

Acknowledgements

This research was carried out in the context of the joint Strategy & Change program of HCSS and The Netherlands Institute for Applied Sciences (TNO), in close cooperation with an advisory group composed of TNO energy experts. All funding was provided by TNO. We would like to thank these experts, specifically René Peters, Cyril Widdershoven, and Rob Weterings, for their input during the research. Further, we would like to thank Joris Govers, Ruben Moorlag, Teun van Dongen, Tim Sweijs, and Emily Knowles for their assistance in the research process at HCSS. Finally, we would like to thank Tushith Islam (TU Delft) for his help during the analysis.

References

Agayev, Z. (2016). Azerbaijan credit rating cut to junk by Fitch on oil rout impact. *Bloomberg News*. Retrieved from http://www.bloomberg.com/news/articles/2016-02-27/azerbaijan-credit-rating-cut-to-junk-by-fitch-on-oil-rout-impact

Anderson, E. G. (2011). A dynamic model of counterinsurgency policy including the effects of intelligence, public security, popular support, and insurgent experience. *System Dynamics Review, 27*(2), 111-141. doi:10.1002/sdr.443

Asche, F., Oglend, A., & Osmundsen, P. (2012). Gas versus oil prices the impact of shale gas. *Energy Policy, 47*(3), 117-124. doi:10.1016/j.enpol.2012.04.033

Bankes, S. C. (1993). Exploratory Modeling for Policy Analysis. *Operations Research, 41*(3), 435-449. doi:10.1287/opre.41.3.435

Bankes, S. C., Walker, W. E., & Kwakkel, J. H. (2013). Exploratory Modeling and Analysis. In G. S. & M. C. Fu (Eds.), *Encyclopedia of Operations Research and Management Science* (3rd ed.). Berlin, Germany: Springer.

Baranzelli, C., Vandecasteele, I., Ribeiro Barranco, R., & Mari i Rivero, I. (2015). Scenarios for shale gas development and their related land use impacts in the Baltic Basin, Northern Poland. *Energy Policy, 84*(2015), 80-95. doi:10.1016/j.enpol.2015.04.032

Boudet, H., Clarke, C., Bugden, D., Maibach, E., Roser-Renouf, C., & Leiserowitz, A. (2014). "Fracking" controversy and communication: Using national survey data to understand public perceptions of hydrolic fracturing. *Energy Policy, 65*(2014), 57-67. doi:10.1016/j.enpol.2013.10.017

BP. (2015). *BP Statistical Review of World Energy*. Retrieved from London: UK: http://www.bp.com/content/dam/bp/pdf/energy-economics/statistical-review-2015/bp-statistical-review-of-world-energy-2015-full-report.pdf

Chyong Chi, K., Nuttall, W. J., & Reiner, D. M. (2009). Dynamics of the UK natural gas industry: System dynamics modelling and long-term energy policy analysis. *Technological Forecasting and Social Change, 76*(3), 339-357. doi:10.1016/j.techfore.2008.06.002

Cincotta, R., Engelman, R., & Anastasion, D. (2003). *The Security Demographic: Population and Civil Conflict after the Cold War*. Retrieved from Washington, DC: http://resourcecentre.savethechildren.se/sites/default/files/documents/2599.pdf

Collier, P., & Hoeffler, A. (2004). Greed and Grievance in Civil War. *Oxford Economic Papers, 56*(4), 563-595. doi:10.1093/oep/gpf064

De Silva, P. N. K., Simons, S. J. R., & Stevens, P. (2016). Economic impact analysis of natural gas development and the policy implications. *Energy Policy, 88*, 639-651. doi:10.1016/j.enpol.2015.09.006

DeLaurentis, D., & Callaway, R. K. C. (2004). A System-of-Systems Perspective for Public Policy Decisions. *Review of Policy Research, 21*(6), 829-837. doi:10.1111/j.1541-1338.2004.00111.x

EIA. (2015). *Annual Energy Outlook 2015*. Retrieved from Washington, DC: http://www.eia.gov/forecasts/aeo/pdf/0383(2015).pdf

Eker, S., & van Daalen, E. (2015). A model-based analysis of biomethane production in the Netherlands and the effectiveness of the subsidization policy under uncertainty. *Energy Policy, 82*, 178-196. doi:10.1016/j.enpol.2015.03.019

Farchy, J. (2015). Currenct devaluation places Kazakhstan central bank under pressure. *Financial Times*. Retrieved from http://www.ft.com/intl/cms/s/0/48495ba8-579f-11e5-a28b-50226830d644.html#axzz46umMMbEe

Fiddaman, T. S. (1997). *Feedback Complexity in Integrated Climate-Economy Models.* (PhD dissertation), Massachusetts Institute of Technology, Cambridge, MA. Retrieved from http://www.metasd.com/papers/Fiddaman%20Dissertation%20Climate%20FREE.PDF

Ford, A. (1999). Cycles in competitive electricity markets: a simulation study of the western United States. *Energy Policy, 27*(11), 637-658. doi:10.1016/S0301-4215(99)00050-6

Forrester, J. W. (1961). *Industrial Dynamics*. Cambridge, MA: MIT Press.

Hamarat, C., Kwakkel, J. H., & Pruyt, E. (2013). Adaptive Robust Design under Deep Uncertainty. *Technological Forecasting & Social Change, 80*(3), 408-418. doi:10.1016/j.techfore.2012.10.004

Hamarat, C., Kwakkel, J. H., Pruyt, E., & Loonen, E. T. (2014). An exploratory approach for adaptive policymaking by using multi-objective robust optimization. *Simulation Modelling Practice and Theory, 46*, 25-39. doi:10.1016/j.simpat.2014.02.008

Hanau, A. (1928). *Die Prognose der Schweinepreise*. Berlin: Verlag von Reimar Hobbing.

Hosseini, S. H., & Shakouri G, H. (2016). A study on the future of unconventional oil development under different oil price scenarios: A system dynamics approach. *Energy Policy, 91*, 64-74. doi:10.1016/j.enpol.2015.12.027

IEA. (2012). *World Energy Outlook 2012*. Retrieved from Paris:

Jaspal, R., Nerlich, B., & Lemańczyk, S. (2014). Fracking in the Polish press: Geopolitics and national identity. *Energy Policy, 74*(November 2014), 253-261. doi:10.1016/j.enpol.2014.09.007

Jenner, S., & Lamadrid, A. J. (2013). Shale gas vs. coal: Policy implications from environmental impact comparisons of shale gas, conventional gas, and coal on air, water, and land in the US. *Energy Policy, 53*(February 2013), 442-453. doi:10.1016/j.enpol.2012.11.010

Kargbo, D. M., Wilhelm, R. G., & Campbell, D. J. (2010). Natural Gas Plays in the Marcellus Shale: Challenges and Potential Opportunities. *Environmental Science & Technology, 44*(15), 5679–5684. doi:10.1021/es903811p

Kerr, S. (2015). Saudis unveil radical austerity programme. *Financial Times*. Retrieved from http://www.ft.com/intl/cms/s/0/a5f89f36-ad7e-11e5-b955-1a1d298b6250.html#axzz46umMMbEe

Kinnaman, T. C. (2011). The economic impact of shale gas extraction: A review of existing studies. *Ecological Economics, 70*(7), 1243-1249. doi:10.1016/j.ecolecon.2011.02.005

Kuuskraa, V. A., Stevens, S. H., & Moodhe, K. (2013). *Technically Recoverable Shale Oil and Shale Gas Resources: An Assessment Of 137 Shale Formations in 41 Countries Outside the United States*. Retrieved from Washington: http://www.eia.gov/analysis/studies/worldshalegas/

Kwakkel, J. H., & Pruyt, E. (2013). Exploratory modeling and analysis, an approach for model-based foresight under deep uncertainty. *Technological Forecasting & Social Change, 80*(3), 419-431. doi:10.1016/j.techfore.2012.10.005

Lee, J. (2015). The regional economic impact of oil and gas extraction in Texas. *Energy Policy, 87*, 60-71. doi:10.1016/j.enpol.2015.08.032

Lempert, R. J., Popper, S. W., & Bankes, S. C. (2003). *Shaping the Next One Hundred Years: New Methods for Quantitative, Long-Term Policy Analysis* (MR-1626). Retrieved from Santa Monica, CA, USA: http://www.rand.org/pubs/monograph_reports/2007/MR1626.pdf

Mahdavy, H. (1970). The Pattern and Problems of Economic Development in Rentier States: the Case of Iran. In M. A. Cook (Ed.), *Studies in the Economic History of the Middle East*. Oxford: UK: Oxford University Press.

Marshall, M. G., Gurr, T. R., & Jaggers, K. (2014). *Polity IV Project. Political Regime Characteristics and Transitions, 1800-2013. Dataset Users' Manual*. Retrieved from Vienna, VA: http://www.systemicpeace.org/inscr/p4manualv2013.pdf

Marshall, M. G., Gurr, T. R., & Jaggers, K. (2016, 6 May 2014). The Polity Project. Retrieved from http://www.systemicpeace.org/polityproject.html

Meadows, D. H., Meadows, D. L., Randers, J., & Behrens, W. W., III. (1972). *The Limits to Growth: A Report for the Club of Rome's Project on the Predicament of Mankind*. New York: Universe Books.

Melikoglu, M. (2014). Shale gas: Analysis of its role in the global energy market. *Renewable and Sustainable Energy Reviews, 37*, 460-468. doi:10.1016/j.rser.2014.05.002

Meng, Q., & Ashby, S. (2014). Distance: A critical aspect for environmental impact assessment of hydraulic fracking. *The Extractive Industries and Society, 1*(2), 124-126. doi:10.1016/j.exis.2014.07.004

MinesQC. (2016). What are the stages in the process of mine development? Retrieved from http://minesqc.com/en/blog/article/2016/02/16/what-are-the-stages-in-the-process-of-mine-development/

Morecroft, J. D. W. (1982). A Critical Review of Diagramming Tools for Conceptualizing Feedback System Models. *Dynamica, 8*(1), 20-29.

Naill, R. F. (1977). *Managing the Energy Transition: a System Dynamics Search for Alternatives to Oil and Gas*. Cambridge, MA: Ballinger Publishing Company.

Naill, R. F. (1992). A system dynamics model for national energy policy planning. *System Dynamics Review, 8*(1), 1-19. doi:10.1002/sdr.4260080102

Olmstead, S. M., Muehlenbachs, L. A., Shih, J. S., Chu, Z., & Krupnick, A. J. (2013). Shale gas development impacts on surface water quality in Pennsylvania. *Proc Natl Acad Sci U S A, 110*(13), 4962-4967. doi:10.1073/pnas.1213871110

Oreskes, N., Shrader-Frechette, K., & Belitz, K. (1994). Verification, Validation, and Confirmation of Numerical Models in the Earth Sciences. *Science Magazine, 263*(5147), 641-646. doi:10.1126/science.263.5147.641

Osorio, S., & van Ackere, A. (2016). From nuclear phase-out to renewable energies in the swiss electricity market. *Energy Policy, 93*, 8-22. doi:10.1016/j.enpol.2016.02.043

Perry, S. L. (2012). Development, Land Use, and Collective Trauma: The Marcellus Shale Gas Boom in Rural Pennsylvania. *Culture, Agricultura, Food and Environment, 34*(1), 81-92. doi:10.1111/j.2153-9561.2012.01066.x

Pruyt, E. (2013). *Small System Dynamics Models for Big Issues: Triple Jump towards Real-World Complexity*

Pruyt, E. (2015). From Building a Model to Adaptive Robust Decision Making Using Systems Modeling. In M. Janssen, M. A. Wimmer, & A. Deljoo (Eds.), *Policy Practice and Digital Science. Integrating Complex Systems, Social Simulation and Public Administration in Policy Research* (Vol. 10, pp. 75-93). Switzerland: Springer International Publishing.

Pruyt, E., & Kwakkel, J. H. (2014). Radicalization under Deep Uncertainty: A Multi-Model Exploration of Activism, Extremism, and Terrorism. *System Dynamics Review, 30*(1-2), 1-28. doi:10.1002/sdr.1510

Richter, P. M., & Holz, F. (2015). All quiet on the eastern front? Disruption scenarios of Russian natural gas supply to Europe. *Energy Policy, 80*, 177-189. doi:10.1016/j.enpol.2015.01.024

Ross, M. L. (2004). What Do We Know about Natural Resources and Civil War? *Journal of Peace Research, 41*(3), 337-356. doi:10.1177/0022343304043773

Sovereign Wealth Fund Institute. (2013). Sovereign Wealth Fund Rankings. Retrieved from http://www.swfinstitute.org/fund-rankings/

Sterman, J. D. (1981). *The Energy Transition and the Economy: A System Dynamics Approach.* (Doctoral), Massachusetts Institute of Technology, Cambridge, MA.

Sterman, J. D. (1994). Learning in and about complex systems. *System Dynamics Review, 10*(2-3), 291-330. doi:10.1002/sdr.4260100214

Sterman, J. D. (2000). *Business Dynamics: Systems Thinking and Modeling for a Complex World*. New York: McGraw.

Tian, L., Wang, Z., Krupnick, A., & Liu, X. (2014). Stimulating shale gas development in China: A comparison with the US experience. *Energy Policy, 75*(S.1), 109-116. doi:10.1016/j.enpol.2014.07.025

Tilton, J. E. (1996). Exhaustible resources and sustainable development: two different paradigms. *Resources Policy, 22*(1-2), 91-97. doi:10.1016/S0301-4207(96)00024-4

Urdal, H. (2006). A Clash of Generations? Youth Bulges and Political Violence. *International Studies Quarterly, 50*, 607-629. doi:10.1111/j.1468-2478.2006.00416.x

Victor, N., Nichols, C., & Balash, P. (2014). The impacts of shale gas supply and climate policies on energy security: The U.S. energy system analysis based on MARKAL model. *Energy Strategy Reviews, 5*(2), 26-41.

Wils, A., Kamiya, M., & Choucri, N. (1998). Threats to sustainability: simulating conflict within and between nations. *System Dynamics Review, 14*, 129-162. doi:10.1002/(SICI)1099-1727(199822/23)14:2/3<129::AID-SDR146>3.0.CO;2-6

World Bank. (2011). *The Changing Wealth of Nations: Measuring Sustainable Development in the New Millennium*. Retrieved from Washington, DC:

http://siteresources.worldbank.org/ENVIRONMENT/Resources/ChangingWealthNations.pdf

World Bank. (2016). Unemployment, youth male (% of male labor force ages 15-24) (modeled ILO estimate). Retrieved from http://data.worldbank.org/indicator/SL.UEM.1524.MA.ZS

5 Societal Ageing in the Netherlands

The paper on societal ageing in the Netherlands approaches two different methodological issues. The first issue regards how to deal with divergence in ideas about the desirability of policies. The second issue regards how to choose a model's scope, where uncertainty may lead to a broader scope than usual, or can lead to the use of external driving scenarios to test the system's response (this is considered to be a bad modelling practice by some researchers in the System Dynamics field).

Different actors can have different views on the desirability of policies. For example, in the case of societal ageing in the Netherlands, part of the population finds increasing the formal retirement age highly undesirable, while other parts of the population believe that this exact same policy is an absolute necessity. This uncertainty on the desirability of policies, or even on the nature of the actual problem, is characteristic of wicked problems or societal messes. In these situations, different stakeholders disagree on whether policies themselves, and/or policy effects are desirable or not. It may result in situations where no joint view on the problem and solution can be developed. This is relevant, as in traditional (consolidative) modelling it is generally stated that one should model the problem and not the system, which entails that only those system elements that are relevant for the problem should be modelled. If the problem is uncertain itself, this becomes problematic, and a broader approach could be desired accommodating a variety of problem perceptions.

In this paper, we used a model that has a broader scope than most models on societal ageing, yet is exogenously driven by scenarios, which is uncommon in SD. The model provides a good example of how uncertainty may require a broader boundary selection in the problem articulation phase in order to test the effects of relevant external evolutions like changing labour productivity and female fertility. In most other (SD) research on societal ageing, the scope is limited to the population model. In this case however, we assessed the impact of societal ageing on the economy, curative and long-term care, public spending, and public agreement with policies. Further, it is a good example of the use of exogenous input scenarios for generating well-interpretable Scenario Discovery results. In conventional SD modelling, the focus lies on endogenously explaining problematic behaviour, but in this case it turned out to be very useful to see how different plausible trends in combination result in unsustainable public spending. Finally, this paper provides a nice example of the application of RDM to a grand challenge.

In the synthesis of this dissertation, I will reflect on the methodological issues brought forward by this paper with regard to problem articulation, evaluation of model runs, the classification of behaviours, and the reception of the methodology.

This paper was originally published as:

Auping, W. L., Pruyt, E., & Kwakkel, J. H. (2015). Societal Ageing in the Netherlands: A Robust System Dynamics Approach. *Systems Research and Behavioral Science, 32*(4), 485-501. doi:10.1002/sres.2340

Societal Ageing in the Netherlands: A Robust System Dynamics Approach

Willem L. Auping, The Hague Centre for Strategic Studies, The Netherlands; Policy Analysis Section, Faculty of Technology, Policy and Management, Delft University of Technology, The Netherlands

Erik Pruyt, Policy Analysis Section, Faculty of Technology, Policy and Management, Delft University of Technology, The Netherlands

Jan H. Kwakkel, Policy Analysis Section, Faculty of Technology, Policy and Management, Delft University of Technology, The Netherlands

Abstract

Societal ageing is a messy problem with diverging stakeholder views regarding the desirability of policy measures. In this paper, we use a System Dynamics model representing the Dutch demographic and social security system to investigate if and when Dutch governmental retirement and health care contributions become unaffordable. Following the Robust Decision Making approach, we then design and test policies for reducing the societal costs of ageing, taking into account societal support for these policies. We find that unaffordable societal ageing costs are mainly caused by declining productivity levels and increasing life expectancy: permanent increases in productivity are required to sustain the current social security level. We also find that the recent Dutch retirement age policy is insufficiently robust and lacks public support. Focusing on increasing the actual instead of the formal retirement age may generate more public support.

Keywords: Societal ageing, Fiscal sustainability, Messy problems, System Dynamics, Robust Decision Making

1. Introduction

In most developed societies, for example Japan and Western European countries like the Netherlands, the percentage of older people in society and the life expectancy of older people are increasing simultaneously. This ‘double societal ageing’ process is expected to put significant pressure on healthcare budgets, resulting in rising pension costs, and decreasing fiscal sustainability (European Commission, 2012).

Societal ageing related problems are hard to solve for they are wicked (Rittel & Webber, 1973) or messy (Ackoff, 1974). A common characteristic of messy problems is their complexity. In case of societal ageing, this complexity is caused by diverging stakeholders views, uncertain future developments such as the autonomous development of life expectancy, and systemic complexity arising from the interplay of feedback mechanisms, accumulations, and delays within the system. Stakeholder perspectives on how to address the fiscal sustainability of double societal ageing diverge: many governments around the world seem to prefer countering some of these effects by raising the retirement age, while many employees perceive this solution as highly undesirable. As a result, attempts at addressing societal ageing related problems have stalled in many democratic societies. An example of a feedback mechanism is the positive effect of effective health care on average life expectancy, which results in higher health care costs. The existence of feedback effects and delays make System Dynamics (SD) (Forrester, 1961) a suitable tool for aiding decision making related to societal ageing. However, more than just SD may be needed.

In the case of messy problems, aspects of the understanding of the system may be contested. Such issues are characterised by ‘deep uncertainty’. Deep uncertainty is defined as ‘the condition in which analysts do not know or the parties to a decision cannot agree upon (1) the appropriate models to describe interactions among a system’s variables, (2) the probability distributions to represent uncertainty about key parameters in the models, or (3) how to value the desirability of alternative outcomes’ (Lempert et al., 2003, pp. 3, 4). There is, therefore, a need for approaches that can help in handling deep uncertainty.

When confronted with deep uncertainty, exploring the consequences of different assumptions about the system and the desirability of outcomes could be an option. Robust Decision Making (RDM) (Lempert et al., 2006), is a model-based policy analysis approach that builds on Exploratory Modelling (Bankes, 1993), and adapts it to support the design of policies that perform satisfactorily across the space of possible assumptions. RDM has been applied to contested strategic planning problems in a variety of fields, such as climate change (e.g., Groves, 2006; Lempert, Schlesinger, & Bankes, 1996; Popper et al., 2009). RDM has not yet been applied to fiscal sustainability issues related to demographic ageing outside our line of research.

The fiscal sustainability of societal ageing has been studied extensively. Many of these studies acknowledge the inherent uncertainty of demographic forecasts (Keilman, 2008) and incorporate extensive sensitivity analyses (e.g., Alho, 2014; CBO, 2005; Keilman, 2005; Lassila, Valkonen, & Alho, 2014; R. Lee & Tuljapurkar, 1998; Meyerson & Sabelhaus, 2000). Besides demographic uncertainty, both fiscal conditions and fiscal impacts are acknowledged to be fundamentally uncertain (e.g., see European Commission, 2012, pp. 20, 27-34, 53-60). Although these studies have not overlooked the importance of nonlinear relations, SD modelling offers additional advantages. First, it allows for incorporating feedback effects. Second, in combination with an RDM approach, it provides a straightforward way of testing policy robustness across a multitude of plausible futures. Finally, RDM allows for considering public support for potential policies aimed at increasing fiscal sustainability.

According to Forrester (2007), big issues like the future of social security and health care systems are domains where system dynamicists can have a profound impact on government. However, Forrester expressed his disappointment about the limited number of SD studies regarding such ‘big issues’. Although extensive demographic research (e.g., see Sterman, 2000, pp. 474-485) and aging research has been done in the SD field, these studies either focused on the effects of demographic change on care (Brailsford et al., 2012; J. P. Thompson, Riley, Eberlein, & Matchar, 2012), or on how to best model aging populations (Eberlein & Thompson, 2013; Sutrisno & Handel, 2012). To the best of our knowledge, no SD applications beyond our line of research have looked explicitly at the interplay between demographic developments and fiscal sustainability of health care and pension systems. In this line of research, we take a broad perspective on demographic change and societal ageing to assess how government contributions to retirement funding and health care would develop without policy change.

In this paper, we look for robust policy options that simultaneously address the impact of societal ageing on collective spending, and are broadly acceptable to different strata of the population. In the second section, we provide more information on RDM. In the third section, we introduce the problem of societal ageing in the Netherlands, and discuss the model. In the fourth section, we present simulation results and their analysis. In the fifth section, we present a set of policy options aimed at reducing the costs related to societal ageing, and assess their robustness. The final section presents our conclusions.

2. Robust decision making

RDM has been developed over the last 15 years. The RDM approach uses multiple views on the future to support a thorough investigation of modelling results that helps to identify a plan (1) that is robust, (2) that avoids most situations in which the plan would fail to meet its goals, and (3) that clarifies the conditions under which the plan would fail to meet its goals (Groves & Lempert, 2007; Lempert et al., 2003). RDM consists of the following steps (Hall et al., 2012; Keefe, 2012; Walker, Haasnoot, & Kwakkel, 2013):

1. **Scoping:** determine the scope of the analysis by identifying exogenous uncertainties, policies, key relationships, and performance metrics; construct simulation model(s) that relate(s) actions to consequences.
2. **Simulation and Scenario Discovery:** identify a candidate policy to evaluate, run it across an ensemble of scenarios, and identify vulnerabilities of the candidate policy (*i.e.*, which combinations of uncertainties cause the policy to fail to meet the goals);

3. **Policy design:** identify hedging actions (modifying existing policies or defining new ones) to address these vulnerabilities. Repeat steps 2 and 3 for additional candidate policies.

Sampling techniques are commonly used in the second step, just like in automated Sensitivity Analysis. A difference with standard Sensitivity Analysis is that, in RDM, the focus is on ensembles of simulation runs: Under deep uncertainty, there is no base case to start from – possibly a base ensemble. The number of uncertainties and the size of uncertainty ranges explored in RDM are also very large. To deal with the resulting uncertainty space, it is necessary, in RDM, to combine sampling and simulation with other methods and techniques.

Central to RDM is Scenario Discovery (Bryant & Lempert, 2010), a novel approach aimed at addressing the challenges of characterizing and communicating deep uncertainty associated with simulation models. The basic idea is that the consequences of the various deep uncertainties associated with a simulation model are systematically explored through conducting series of computational experiments (Bankes et al., 2013), and that the resulting model-generated data is analysed to identify regions in the uncertainty space that are of interest (Bryant & Lempert, 2010; Kwakkel et al., 2013). These identified regions can subsequently be communicated by means of scenarios.

In Scenario Discovery, a binary classification of the model results is typically used, based on whether policy objectives are reached or not. Subsequently, a statistical rule induction algorithm is used to find the combination of uncertainties that jointly produce a large number of simulation runs where the objectives are not being met. For that purpose, we used the Patient Rule Induction Method (Friedman & Fisher, 1999).

3. Scope of Dutch societal ageing

3.1. Ageing model

The development of Dutch demography and its consequences on government spending on retirement funds, curative care, and long-term care are captured in a SD model. The sub-system diagram in **Figure** (Morecroft, 1982; Sterman, 2000, pp. 99-102) displays the main sub-systems the model is composed of (*i.e.*, demography, economy, social security, health care, and policy perception). The model was developed based on interviews of, and sessions with, experts from the Netherlands Organisation for Applied Scientific Research (TNO) and the Netherlands National Institute for Public Health and the Environment (RIVM) with different areas of expertise, including health care, health care economics, health care organization, labour market economics, labour productivity, and the built environment. The model is a further elaboration of the model by Pruyt and Logtens (2015). The model contains 1137 variables and 80 stocks, some of which are subscripted, and is implemented in Vensim (Ventana Systems, 2010).

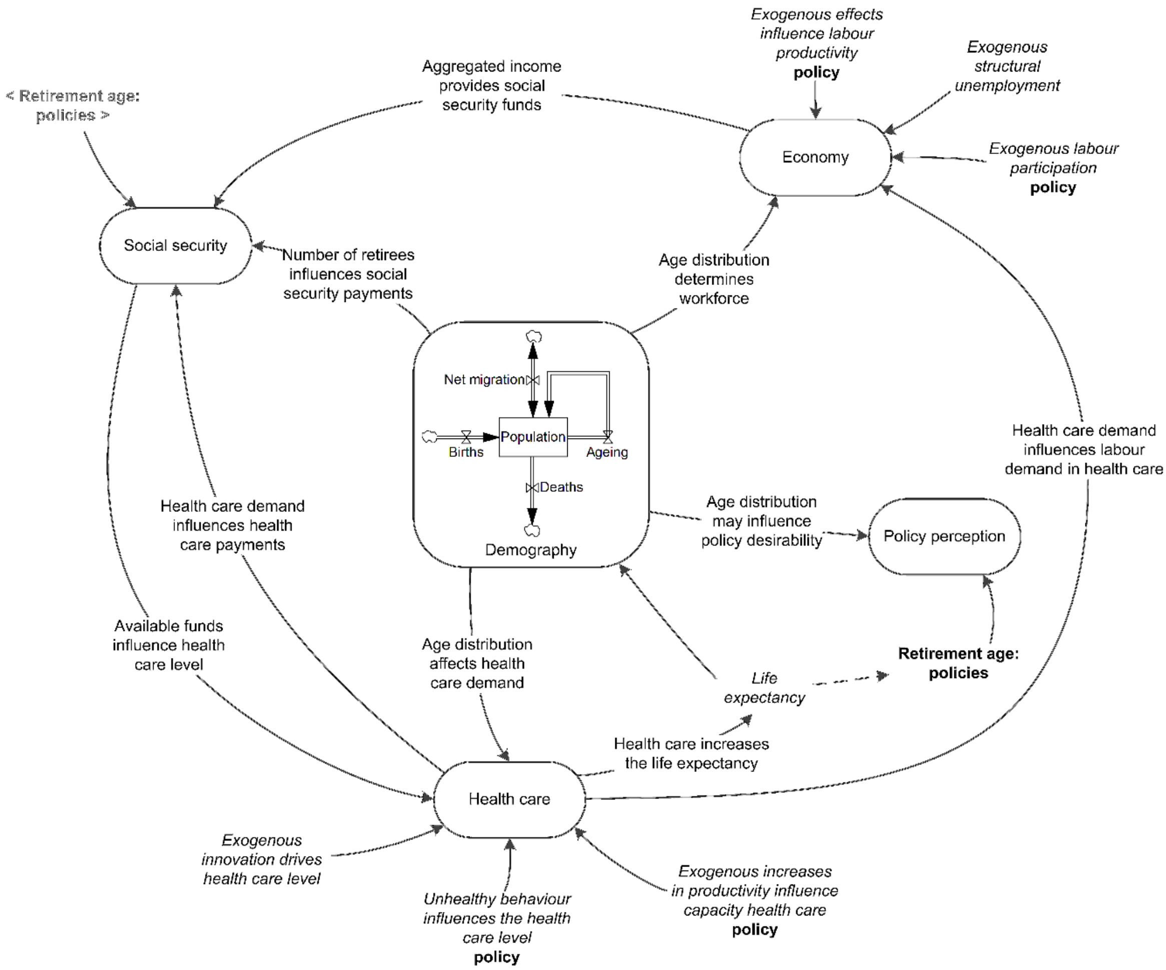


Figure 1. Sub-system diagram of the model used in this study. Exogenous time series in *italics*, policy options in **bold**. The stock-flow structure of the ageing chain is visible in the demographics sub-system. After Pruyt and Logtens (2015)

The demographic sub-system is composed of separate ageing chains for men and women with one year age cohorts (see stock-flow in). The demographic composition influences the economic situation via the composition of the working population, which is defined as all age cohorts between 18 and the formal retirement age. The sum of all age cohorts above the formal retirement age determines the demand for social security benefits, like the basic state pension ‘AOW’ (i.e., translated: general retirement act).

Unhealthy behaviour determines the average unhealthy life expectancy. Demand for long-term care in full time equivalents (fte) is calculated by taking the size of the unhealthy age cohorts, multiplying this with the average costs of long-term care and the fraction spent on wages, and dividing this by the average productivity in long-term care. The demand for curative care in fte is assumed to follow a historic pattern per age and sex (Slobbe, Smit, Groen, Poos, & Kommer, 2007, p. 36). Demand for cure personnel is calculated by dividing the curative care personnel costs by the productivity in curative care. Combined, these two types of care determine total health care costs.

The economic sub-system contains all economic variables. The potential size of the workforce is determined by dividing GDP by the labour productivity. GDP growth is affected if labour supply is smaller than labour demand. The GDP is used to determine the size of the necessary collective expenditure on health care as well as retirement funds. However, automatic feedback to limit this expenditure as part of the GDP is not included endogenously as this would necessitate a policy change.

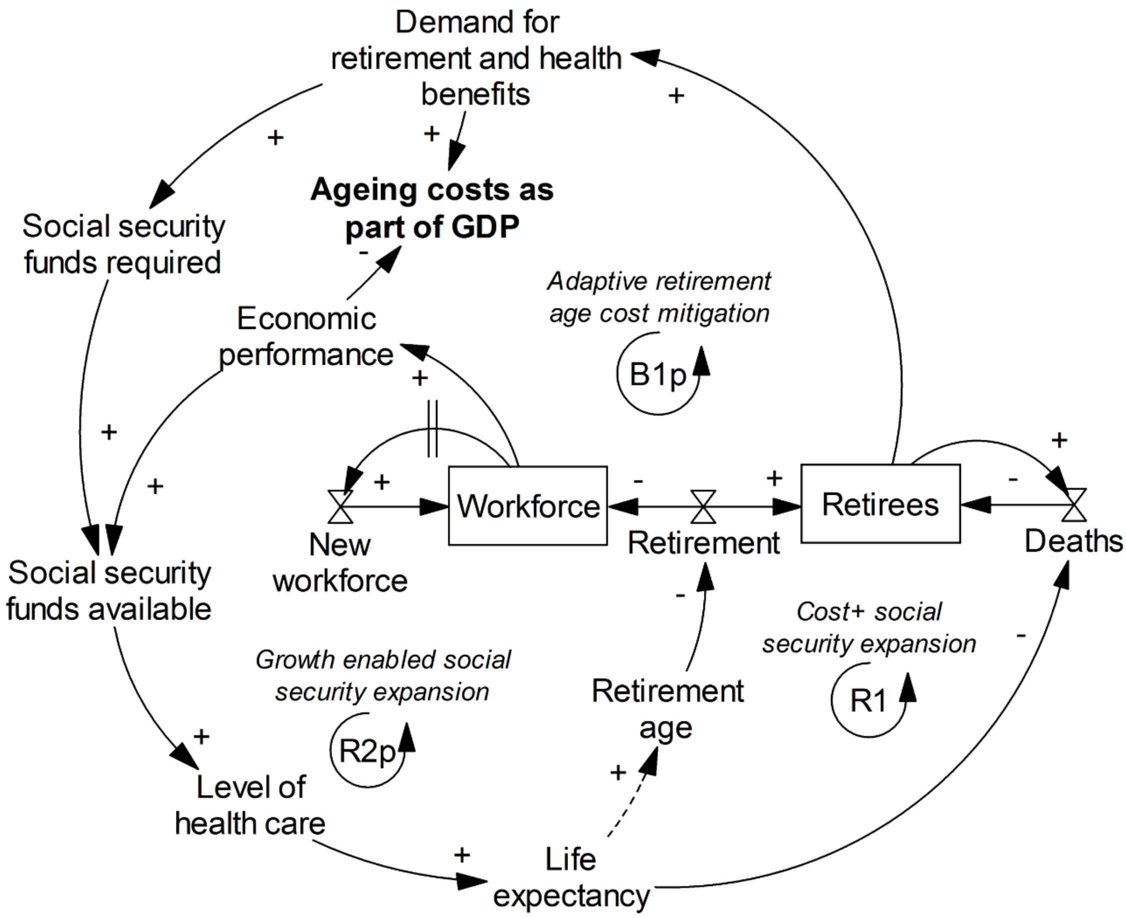


Figure 2. High aggregation level stocks, flows, and causal loops of the ageing model

The most important feedback loops in the model are reinforcing, and act by the influence a high level of health care has on life expectancy (R1 and R2p in Figure 2). The cost-plus social security expansion loop (R1) captures the effect by which more retirees demand more social security funds, which results in more health care expenses, and, ceteris paribus, higher life expectancy, lower death rate, and more retirees. The adaptive retirement age cost mitigation loop (B1p) is a policy loop that mitigates R1 by keeping the number of retirees within sustainable limits. Note, however, that the policy link from life expectancy to retirement age also activates a reinforcing policy loop (R2p): higher life expectancy results in higher retirement age, higher GDP, and higher amounts of social security funds being available, which, in turn, result in more and better health care, and thus, higher life expectancy.

As a result, AOW and health care expenditure both show exponential growth in many scenarios. However, the extent of the growth is unknown due to many uncertain future developments (italics in Figure 1). They are used for evolutions like the productivity of the working population, long-term care, and curative care, and the productivity of both men and women per age. For each of these uncertain future developments, we specified sets of plausible time series, since keeping them constant over a long time horizon would be highly unrealistic. The growth of expenditure on health care and retirement benefits is not endogenously capped by implicit policy options: closing loops by means of new policies is treated as an explicit choice to be made by policy maker. The sets of exogenous time series are used for simulating the model across combinations of different plausible evolutions. After multiple simulations, runs with undesirable behaviour can then be selected, and their input parameters found using the Scenario Discovery approach.

3.2. Public support for policies

To incorporate public support for retirement age policy changes, we modelled the retirement age in multiple ways, namely the formal retirement age, the actual retirement age or leaving age, a delayed formal retirement age, and a forecasted future retirement age for each cohort. They correspond to alternative perceptions with regard to retirement age policies. The first perception is defined as the age at which retirement benefits commence. The second is the average age at which either men or women stop working, and leave the working population. It may, therefore, take time before society accepts the fact that the retirement age has been raised. Third, the different age cohorts may change their expectation regarding their own retirement age due to changes of the retirement age in recent years, providing an expectation for the retirement age for each age cohort (i.e., a cohort forecast).

By combining the formal retirement age with each of the three other perceptions, three metrics for public support related to retirement age changes are modelled (see Table 1), where the maximum public support is defined as 1 (full support) and the minimum as 0 (no support). In the first metric, we calculate for each age cohort above 18 the amount of years they still have to work with both the formal retirement age and the delayed retirement age. The weighted average across relevant age cohorts of the relative differences between these two figures is assumed to correspond to the relative decrease in public support.

Table 1. Metrics used in determining public support

Name	Metric
Formal and delayed retirement age comparison	Formal retirement age compared to the delayed perception of the retirement age in the last period, divided by the number of years left to the formal retirement age
Formal and actual retirement age comparison	Formal retirement age compared to actual retirement age, divided by the number of years left to the formal retirement age
Formal and forecasted retirement age comparison	Formal retirement age compared to the age cohort specific expectation of the formal retirement age, which is based on the average change in retirement age in the recent past

In the second metric, the formal retirement age is compared to the actual retirement age. The rest of this metric is similar to the first metric, although, as the actual retirement ages for men and women are not necessarily the same, the perception is calculated for both the male and the female working population. The total perception is then the weighted average of perceptions for each male and female age cohort.

Finally, in the third metric, each age cohort makes a forecast for the expected retirement age dependent on the average change in the retirement age in recent years. This is calculated as the difference between the formal and the delayed retirement age, relative to the delayed retirement age. The relative increase in years to work is calculated by taking the relative difference between the extrapolated average number of years individuals need to work and the formal retirement age. This relative increase is, for each age cohort equal to the average relative change in the retirement age.

4. Simulating Dutch societal ageing and Scenario Discovery

Using Latin Hypercube Sampling, the model was simulated 1000 times between 2010 and 2060 to generate an ensemble of cases. Each case is a selection of values for uncertainties and assumptions about the future state of the system. In the simulation setup, we considered 40 uncertainties, including parametric uncertainties (30), delay orders (3), trend uncertainties (15), and structural uncertainties (2). An example of a parametric uncertainty is the average wage level rise in a given period.

A delay order uncertainty represents a 1^{st}, 3^{rd}, 10^{th}, or 1000^{th} (pipeline) order delay. A trend uncertainty is a plausible set of exogenous time series, for example regarding the future development of the male life expectancy. Finally, structural uncertainties are, here, different plausible formulas within the SD model, selected using switch variables, for example, the public support metrics.

In order to classify the cases in terms of their desirability, we distinguish between desirable, undesirable, and unaffordable societal ageing. The current limit of public expenditure is approximately 50% of GDP. This corresponds to the present size of Dutch collective spending, which is a historically grown political choice (Bos, 2006). The theoretical upper limit lies at 100% of GDP, as it is impossible for a society to spend more on health care and pensions. The actual limit is probably much lower, given that, for example, food, housing, education, and culture also account for a large proportion of Dutch public expenditure. Since most of the costs of ageing are collectively funded, this is also a clear limit. Expenditure on health care and government funding of the state pension above 50% of GDP is thus practically impossible in the Dutch political context.

The question then remains which limits apply to expenditure on ageing-related costs, as a fraction of GDP, and its desirability. In recent years, relative public spending on health care increased. If this trend continues in the future, it will lead to undesirable public finances, as costs of care will supplant other government expenditures. For this reason, it is important to establish sustainable limits, within which health care costs and other ageing-related costs as a fraction of GDP should remain.

For classifying the individual model runs, we distinguish four scenario classes. The first class corresponds to the situation in which care costs decrease relative to the 2010 level of around 10% of GDP, which is viewed as highly desirable. The second class corresponds to the situation where the costs increase to a level between 10% and 25%, which is viewed as acceptable. The third class corresponds to cost increases to a level between 25% and 50% of GDP, which requires additional measures to be implemented. Finally, in the fourth class ageing costs rise above 50% of GDP, which is the current size of public expenditure. The simulation runs within this class represent cases with unaffordable costs of societal ageing. Each of the scenario classes represents a narrative that can be used by policy makers for assessing the necessity and use of the proposed policies, without failing to acknowledge the existing uncertainty.

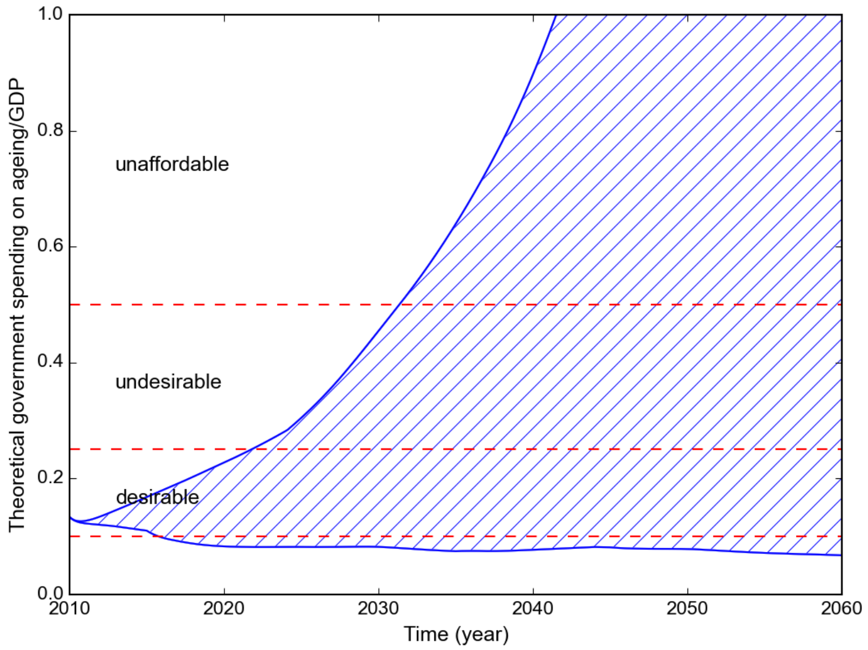


Figure 3. Bandwidth of developments of government contribution to AOW as part of GDP. This is the sum of the necessary government contribution

Figure 3 **s**hows the bandwidth between which the necessary government costs of the retirement and health care may develop relative to GDP. The boundaries between the desired, sustainable, and undesirable categories are indicated by dashed lines. This figure shows that a subset of cases results in unaffordable costs of societal ageing.

We performed Scenario Discovery in order to reveal which combinations of uncertainties result in a situation where the sum of health care costs and retirement funding will become unaffordable. We found that all runs with the sum of health care costs and retirement funding in the unaffordable range resembled a situation in which labour productivity of the Dutch population decreases over time. This has a negative effect on the evolution of GDP, resulting in situations where the costs relative to GDP increase most.

Other important drivers of ageing costs are increased life expectancy, especially of men, increased unhealthy behaviour, and continuously low labour participation of older employees. Increasing male life expectancy is important, as male life expectancy is presently lower than female life expectancy, although it may catch up. The development of life expectancy strongly affects the development of retirement costs, while unhealthy behaviour affects the development of health costs. Finally, scenarios in which labour participation of the older population further decreases have a negative effect on income. This amplifies the effect of a decrease in labour productivity, since more people will be needed to generate the same output, while with low labour participation, the labour availability will be lower.

4.1. Model validation

We performed various tests to assess whether the model was fit for purpose (Forrester & Senge, 1980; Sterman, 2000). During the conceptualisation and specification phase, standard verification checks, like unit checks and extreme value tests, were performed. In addition, we organized workshops with domain and methodological experts from Delft University of Technology, The Hague Centre for Strategic Studies, TNO, PricewaterhouseCoopers, and the Dutch Ministry of Social Affairs in order to validate the model structure (Logtens, 2011). Finally, we compared the bandwidth of population scenarios generated by the model with forecasts provided by Statistics Netherlands (CBS, 2010). The bandwidth of the population size in this study is slightly broader than those produced by Statistics Netherlands (Figure 4).

However, the distribution of the population across different work-related age cohorts has more influence on the affordability of societal ageing, is (Figure 4). An important indicator in that respect is the demographic pressure, which is the sum of the population under 20 years old and the population above 65 years old relative to the population between 20 and 65 years old. In other words, it is a proxy for the proportion of the non-working population compared to the working population (CBS, 2015). The behavioural mode of the demographic pressure generated with our model is similar to the forecasts of Statistics Netherlands (CBS, 2010). However, the spread of our outcomes is, especially in the period between 2040 and 2060, much wider. This can be explained by the broader bandwidth of life expectancies in our research, and by the different approach for dealing with immigration.

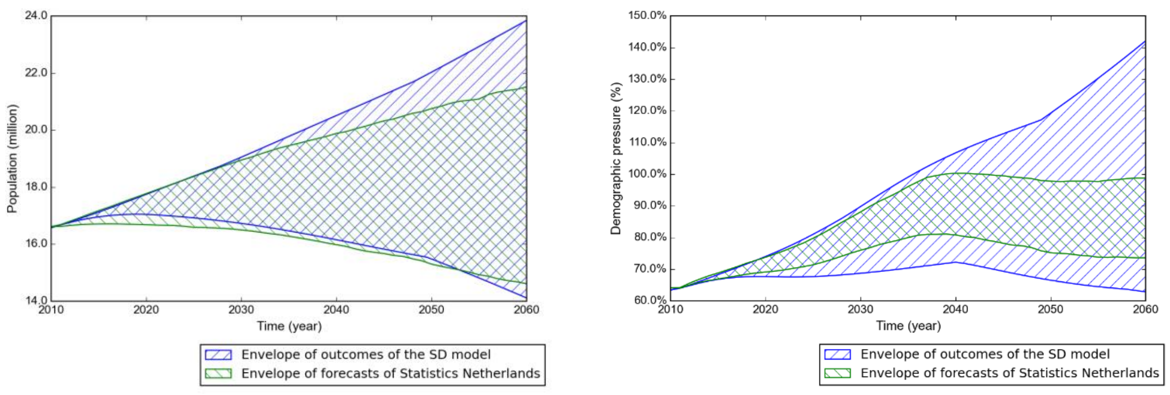


Figure 4. Bandwidth of scenarios for the Dutch population size (*left*) and the demographic pressure (*right*) generated in this study compared to the scenarios by Statistics Netherlands

The model can produce health care cost figures that are higher than the GDP. Although this is both practically and theoretically infeasible, given the intended exploratory purpose of this study, not protecting the model allowed for uncovering hidden assumptions, identifying extremely undesirable evolutions, and determining valid uncertainty spaces. As implicit policies are not taken into account, the bandwidth of public expenditure on retirement and health care costs is not restricted. As such, the model can produce values for theoretical health care costs that are theoretically or politically impossible. Given the purpose of exploring the potential of infeasible ageing related costs if no action would be taken, theoretical or practically impossible evolutions of health care expenditure values do not undermine the usefulness of the model. The stakeholders and clients, therefore, considered these scenarios important indicators for the unsustainability of the current health care and pension system, while the broad bandwidths of outcomes did not undermine their confidence in the model. In similar settings with the Netherlands Court of Audit, the stakeholders also acknowledged the validity of the model outcomes given the then existing set of policies.

5. Policies for societal ageing

5.1. Policy design

The policy designed and tested in this study are based both on policies that were suggested in the public debate, and on the results from our Scenario Discovery approach. Important drivers for unaffordable ageing related costs these policies have to address include life expectancy, labour productivity, unhealthy behaviour, and labour participation of older people. In total, 10 different policy options are simulated (Table 2). All policy options are implemented in the model from 2015 on, except for policies that adapt the retirement age over time. These policy options were all tested, separately and in various combinations, for their effectiveness in reducing the government expenditure on ageing, and the public support for these measures.

The policies can be divided in four categories: policies with regard to the retirement age, the prevention of unhealthy behaviour, policy options for employers, and policies with regard to labour productivity. For the retirement age, we incorporate three different policy options, adaptively following the life expectancy. These policies thus provide a balancing feedback of the effects of increases in longevity on fiscal sustainability. The prevention of unhealthy behaviour leads to the reduction of the part of the population with unhealthy behaviour (*i.e.*, smoking, obesity, inactivity, and heavy drinking). By doing so, they reduce the unhealthy life years, and consequently the costs of long-term care. However, they also increase life expectancy, which leads to increased spending on retirement benefits.

Table 2. Policy options to counteract the undesired effects of demographic ageing.

Category	Name	Description
Life expectancy	Retirement age according to pension agreement	In 2020 to 66 years, in 2025 to 67 years. Thereafter coupling with life expectancy according to the formula V = (L – 18.26) – (P – 65) ("Wet verhoging pensioenleeftijd, extra verhoging AOW en flexibilisering ingangsdatum AOW," 2011).
	Retirement age according to Dutch stability program 2012	Between 2013 and 2019 the retirement age will gradually be raised to 66, and thereafter gradually to 67 by 2023. From 2024 onwards we assume that the increase will proceed as planned in the pension agreement, but with a delay of 10 years instead of 11 years.
	Robust retirement age	Periodically adjusted, adaptive retirement age at 85% of life expectancy
Preve ntion	Prevention of unhealthy behaviour	0.5%/year reduction in the part of the population with unhealthy habits
Options employers	Equalize age preference employers	10%/year improvement in the relative age-preference of employers
	Increase relative number of hours per older employee	2%/year reduction of portion not worked full time for workers over 45 years
	Increase labour participation	2%/year reduction in portion not in full employment for workers over 45 years
Labour productivity	Increase productivity Dutch workforce	2%/year increase in average labour productivity relative to reference
	Increase productivity in curative care	2%/year increase of productivity in curative care compared to reference
	Increase productivity in long-term care	2%/year increase in productivity in long-term care compared to reference

Three different policy options are potential strategies for employers. They can equalize the age preference for older workers, making them as attractive as younger employees, which, in turn, results in later labour market exits. Further, these measures lead to an increase in the hours worked per older employee, and an increase in labour participation (*i.e.*, more people working at higher ages). Finally, labour productivity can be increased for the total working population, long-term care, and curative care. While especially the latter two are difficult to operationalize, it is important to test their effectiveness before expenses are made towards reaching these goals. In testing the robustness of the policies, we combine each of the three retirement age policies with all other policy options. A more elaborate discussion of these policy options can be found in Willem L. Auping, Erik Pruyt, Jan H. Kwakkel, Govert Gijsbers, and Michel Rademaker (2012).

5.2. Policy robustness

Policies are considered to be robust here if they generally improve the desirability of plausible futures without decreasing the desirability of any specific future (Lempert et al., 2006). One way of making policies robust is to make them adapt to how the future is actually unfolding. The retirement age policies presented here adapt to the average life expectancy of the Dutch population. It is interesting to see within which bounds the retirement age can develop, given the different plausible developments of the life expectancy. The bandwidth of life expectancy trends for both women and men (Figure 5) is mainly driven by a set of plausible exogenous trends, but is also affected endogenously by unhealthy behaviour, for prevention of unhealthy behaviour results in a slight increase in life expectancy.

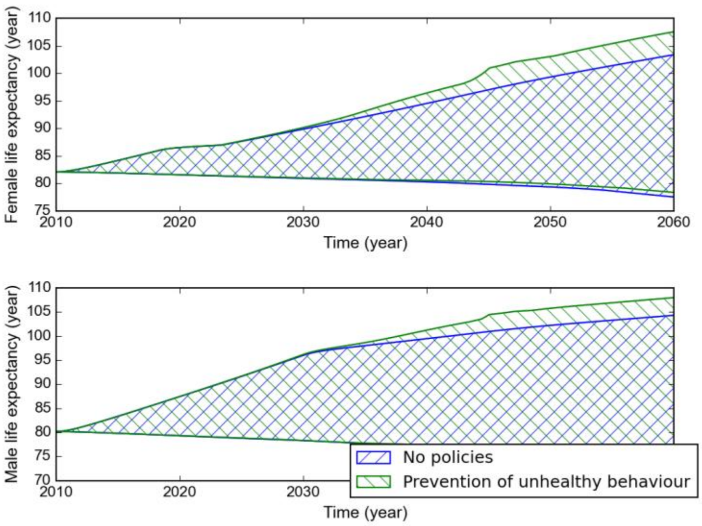


Figure 5. Scenario space for the development of the life expectancy of women (*above*) and men (*under*), for the base ensemble and the ensemble with the reduced unhealthy behaviour policy implemented.

These retirement policies adapt the retirement age as a function of the weighted average life expectancies of the male and female populations. Since the retirement age policies from both the pension agreement ("Wet verhoging pensioenleeftijd, extra verhoging AOW en flexibilisering ingangsdatum AOW," 2011) and the stability agreement (Jager, 2012) are limited to a 1 year increase per 5 years, the extent of the increase is limited to a retirement age of 74 in 2060. This implies that these policies cannot adapt to a life expectancy higher than 92 years (or 18.26 plus the retirement age of 74). The 'robust retirement age' policy tested here does not have this limitation, and is, therefore, able to adapt to higher life expectancies, but will also decrease in case of declining life expectancy. The range of the robust retirement age is consequentially much broader (Figure 6), which increases policy robustness.

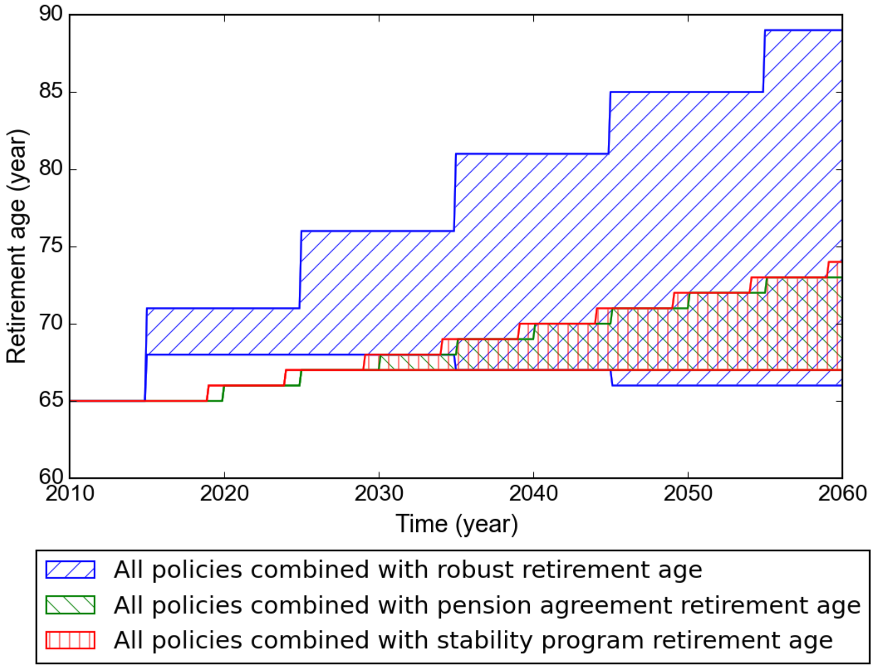


Figure 6. Bandwidth of the retirement age for the three retirement age policies combined with all other policies

5.2.1. Policy effectiveness

The effectiveness of the different retirement age policies largely depends on their ability to cope with changing life expectancies. In unaffordable scenarios, increasing life expectancy plays an important role. The retirement age plans from the pension agreement and the stability program, partially alleviate the cost pressures in these scenarios (Figure 7). However, the robust retirement age policy, which allows complete adaptation to the life expectancy, decreases the costs of the retirement fund significantly better. Due to the fact that decreasing labour productivity and low labour participation also play an important role in unaffordable scenarios, the government contribution might nevertheless still grow considerably compared to the present situation.

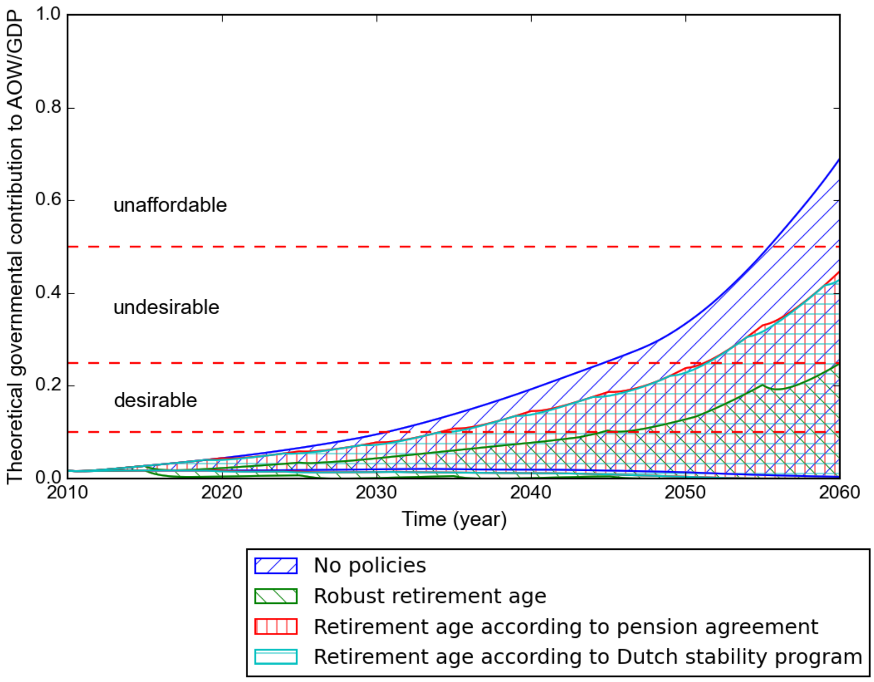


Figure 7. Effects of retirement policies on the necessary government contribution to AOW costs relative to GDP

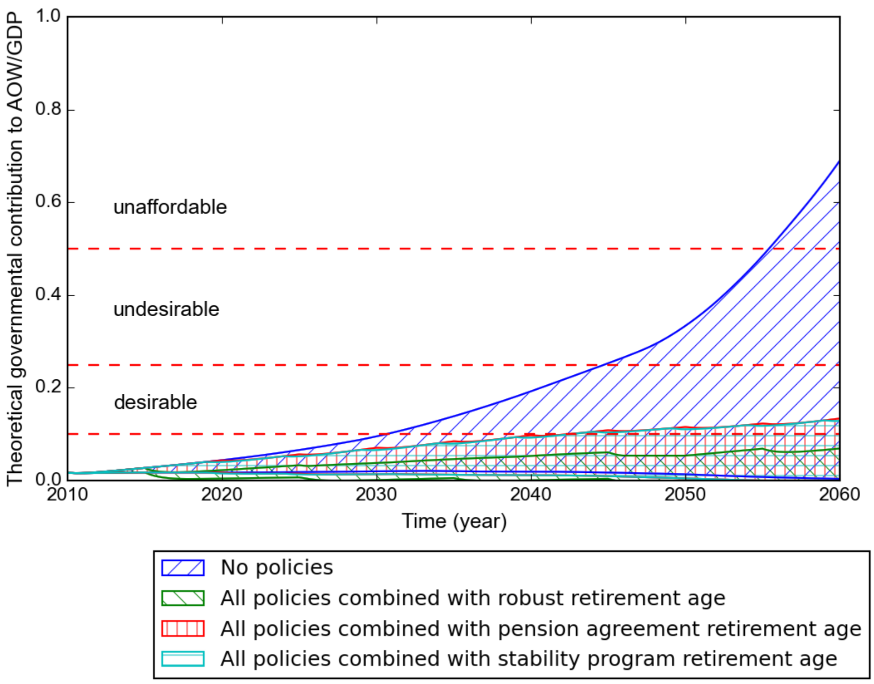


Figure 8. Effects of all policies combined with the retirement policies on the necessary government contribution to AOW costs relative to GDP

As shown in Figure 8, combining the different policies clearly is most effective, bringing the government contribution to AOW back within desirable limits in all cases. This is a logical consequence of strong reinforcing effects between labour participation and retirement age, and the age preference for older workers. That is, when most older people stop working before their formal retirement age, as was widespread in the Netherlands until a few years ago, raising the retirement age will only have a limited effect. There would not be more people working even though pension payments would be reduced.

As indicated by our Scenario Discovery results, a declining trend for the labour productivity of the Dutch workforce will have the strongest mitigating effect on total societal ageing costs. As the largest share of potentially unaffordable cases is caused by high health care costs, it is important to know how effective the different labour productivity policies are in reducing these costs. Figure 9 shows that increasing overall Dutch labour productivity is the most effective policy. This is understandable, as this policy has a positive effect on the development of the GDP, hence reducing the total net costs.

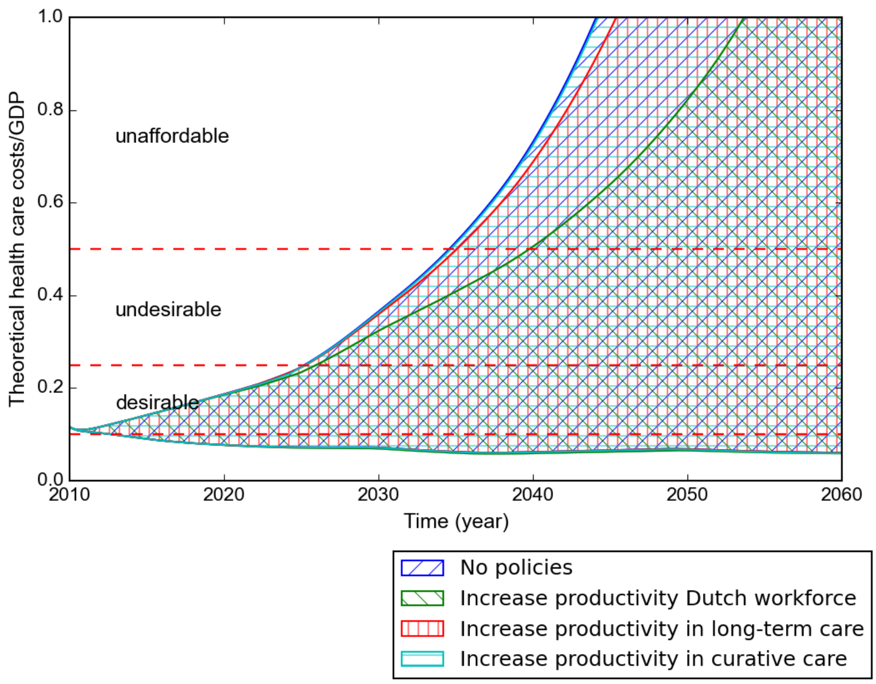


Figure 9. Effects of productivity increases on the costs of health care relative to GDP

The differences between the combinations of all policies with the different retirement age policies (Figure 10**)** is caused by the fact that changing the retirement age policy in case of decreasing labour productivity has a positive effect on the economy, as potential labour shortages will be smaller when more older people are working. That is, due to reduced labour productivity, more employees are needed to uphold economic welfare.

Figure 10. Effects of all policies combined with the retirement policies on health costs relative to GDP

5.2.2. Public support for retirement policies

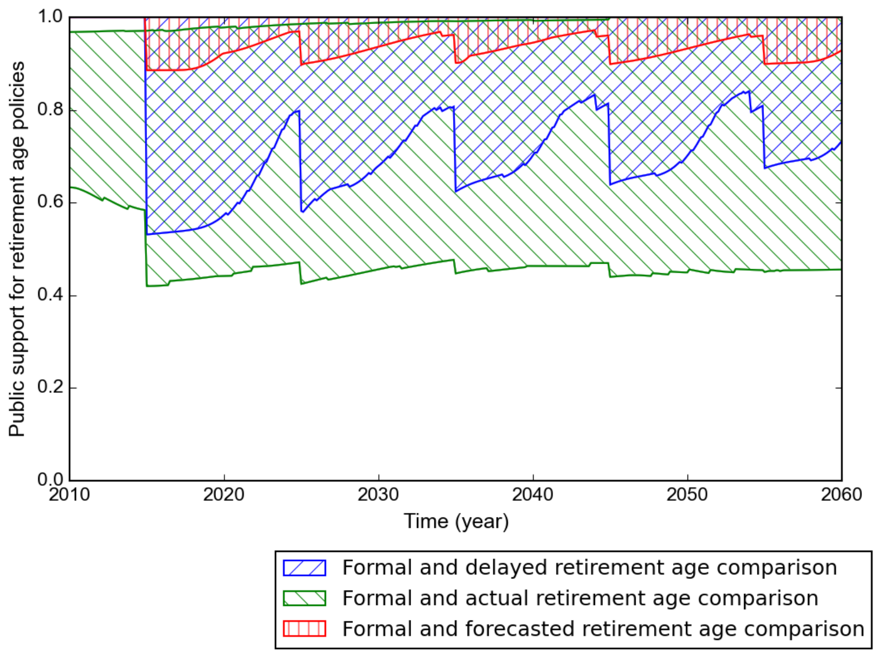


Figure 11. Influence of the three different metrics on the public support with the robust retirement age policy

Public support for the retirement age policies depends largely on the different ways in which parts of the population perceive the retirement age change, and thus, the different metrics used to calculate public support. The effect of these different metrics is shown in Figure 11. It displays public support for the robust retirement age policy option calculated by the three different metrics used. If the population assesses the retirement age by the actual retirement age, public support values are lowest. Hence, increasing labour participation of older people increases the public support for the retirement age in this case (Figure 12).

The lowest public support is explained by the combination of leaving age without a policy to increase the labour participation of older people. The difference between public support for the retirement age policies may increase to a higher level if labour participation of older people increases. If this is combined with an average leaving age above the formal retirement age, there would not be a problem in terms of public support.

This idea is supported by studies advocating 'productive ageing' (e.g., Burr, Caro, & Moorhead, 2002): productive ageing allows for better utilising the productivity potential of older people. Taking the above into consideration, policies aimed at fiscal sustainability with regard to potential costs of demographic ageing should first change behaviours, for example by significantly increasing the leaving age, before changing the formal retirement age. By doing so, there will be far less policy resistance in democratic societies, as vested interests will be less affected.

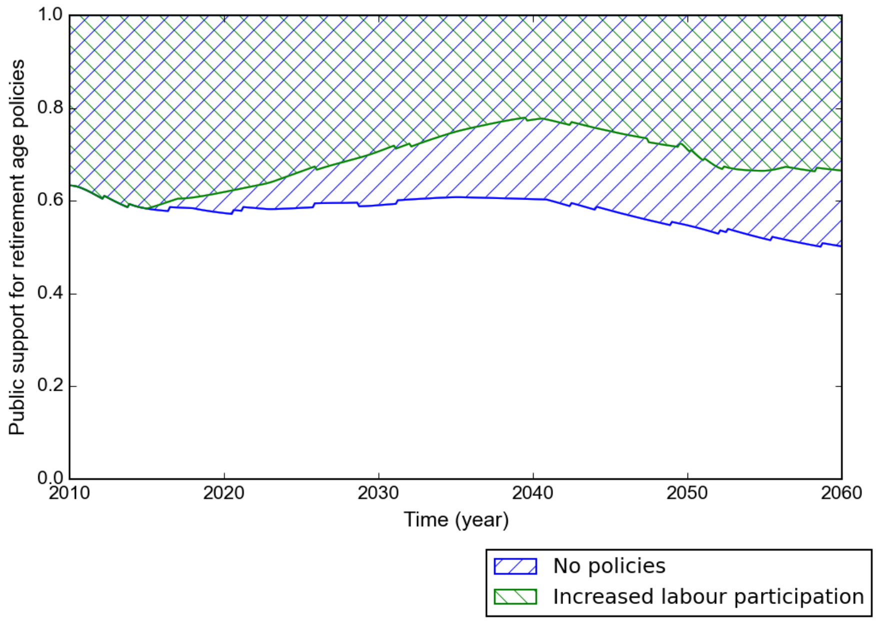


Figure 12. The influence of increased labour participation on the public support for retirement ages

6. Discussion and Conclusions

In this paper we used RDM in combination with SD to address challenges posed by societal ageing. We approached the different perspectives on the desirability of different retirement age policies by classifying cost ranges for government contributions to retirement funding and health care costs, and by modelling different plausible reactions to changes in the retirement age. These reactions may influence the public support for policy reforms believed necessary by many governments, and may result in policy resistance. We believe that this approach has added value, as it provides possibilities to deal with divergent stakeholder perspectives, evaluations, and reactions to policies.

The RDM analysis of the necessary government contributions to ageing costs shows that the set of policies is insufficient for avoiding unaffordable societal costs in all cases. Especially cases with declining labour productivity costs become untenable. Correcting this decline with a policy aimed at stabilising this evolution does not seem to be sufficient: upholding the Dutch social security system in its current form necessitates sustainable economic growth grounded in continued labour productivity increases.

Our model-based approach shows that productivity should be more central in discussions related to societal ageing: raising labour productivity, especially labour productivity in health care, is shown to be even more important than raising the pension age. At the time of this study, this was an important counter-intuitive insight: the then Dutch government was determined to raise the retirement age in order to ‘solve’ societal ageing and cut costs to fight the then economic crisis. However, many of the measures taken to fight the economic crisis actually negatively impacted labour productivity in general and labour productivity in health care, especially in the medium and long term.

Further, lack of public support by the older age cohorts in society prevented taking even more drastic measures regarding the retirement age. Accounting for policy resistance, this issue could be overcome by first changing older people’s behaviour by incentivising significantly higher labour participation and more hours worked. This increases the average age at which people leave the labour market, which, in turn, decreases the perceived undesirability of forcing people to work longer, and thus increases the public support for retirement age policies, compared to increasing the retirement age, which only has limited effects. This is in line with ideas regarding productive ageing, which means that the productivity potential of older people in society should be better utilised.

Regarding the policies aimed at making the retirement age more adaptive to changing life expectancy, which causes ‘double’ societal ageing, it appears that current policies are not robust. However, allowing the retirement age policy to increase the retirement age in a considerably faster rate to make it fiscally robust, may have negative consequences for the cases in which public support is influenced by comparing the formal retirement age with the actual, traditionally lower, retirement age. Given the need for public support in democratic societies, it seems that choosing a more robust retirement age policy may not be feasible given the preferences of the different stakeholders. Focussing on increasing the actual retirement age or leaving age over the formal retirement age instead, may prove a solution in these cases, as it will greatly reduce the potential for public discontent about retirement age policies.

Acknowledgements

The authors would like to thank TNO and HCSS for supporting and funding this research in the 'Strategy & Change' program, HCSS and TU Delft for supporting the case-based development of Exploratory Modelling and Analysis, and Tom Logtens for developing the original societal ageing model.

References

Ackoff, R. L. (1974). *Redesigning the Future: A Systems Approach to Societal Problems*. New York: John Wiley & Sons.

Alho, J. M. (2014). Forecasting demographic forecasts. *International Journal of Forecasting, 30*(4), 1128-1135. doi:10.1016/j.ijforecast.2014.02.005

Auping, W. L., Pruyt, E., Kwakkel, J. H., Gijsbers, G., & Rademaker, M. (2012). *Aging: Uncertainties and Solutions*. Retrieved from The Hague:

Bankes, S. C. (1993). Exploratory Modeling for Policy Analysis. *Operations Research, 41*(3), 435-449. doi:10.1287/opre.41.3.435

Bankes, S. C., Walker, W. E., & Kwakkel, J. H. (2013). Exploratory Modeling and Analysis. In G. S. & M. C. Fu (Eds.), *Encyclopedia of Operations Research and Management Science* (3rd ed.). Berlin, Germany: Springer.

Bos, F. (2006). De Nederlandse collectieve uitgaven in historisch perspectief: CPB.

Brailsford, S. C., Evandrou, M., Luff, R., Shaw, R., Viana, J., Vlachantoni, A., & Willis, R. M. (2012). *Using System Dynamics to Model the Social Care System: Simluation Modeling as the Calalyst in Linking Demography to Care Delivery.* Paper presented at the Winter Simulation Conference.

Bryant, B. P., & Lempert, R. J. (2010). Thinking inside the box: A participatory, computer-assisted approach to scenario discovery. *Technological Forecasting & Social Change, 77*, 34-49. doi:10.1016/j.techfore.2009.08.002

Burr, J. A., Caro, F. G., & Moorhead, J. (2002). Productive aging and civic participation. *Journal of Aging Studies, 16*(2002), 87-105. doi:10.1016/S0890-4065(01)00036-6

CBO. (2005). *Quantifying Uncertainty in the Analysis of Long-Term Social Security Projections*. Retrieved from http://www.cbo.gov/sites/default/files/cbofiles/ftpdocs/68xx/doc6873/11-16-montecarlo.pdf

CBS. (2010, 2010). Kerncijfers van de bevolkingsprognose, 2010-2060. *Statline, Centraal Bureau voor de Statistiek.* Retrieved from http://statline.cbs.nl/StatWeb/publication/?DM=SLNL&PA=80748NED&D1=0,11-12&D2=0-2&D3=a&HDR=G1&STB=T,G2&VW=T

CBS. (2015). Begrippen. Demografische druk. Retrieved from http://www.cbs.nl/nl-NL/menu/methoden/begrippen/default.htm?ConceptID=223

Eberlein, R. L., & Thompson, J. P. (2013). Precise modeling of aging populations *System Dynamics Review, 29*(2), 87-101. doi:10.1002/sdr.1497

European Commission. (2012). *Fiscal Sustainability Report 2012* (8/2012). Retrieved from Brussels: http://ec.europa.eu/economy_finance/publications/european_economy/2012/pdf/ee-2012-8_en.pdf

Forrester, J. W. (1961). *Industrial Dynamics*. Cambridge, MA: MIT Press.

Forrester, J. W. (2007). System Dynamics—the next fifty years. *System Dynamics Review, 23*(2/3), 359-370. doi:10.1002/sdr.381

Forrester, J. W., & Senge, P. M. (1980). Tests for Building Confidence in System Dynamics Models. In A. A. Lagasto, J. W. Forrester, & J. M. Lyneis (Eds.), *TIMS Studies in the Management Sciences* (Vol. 14, pp. 209-228). Amsterdam, Netherlands: North-Holland Publishing Company.

Friedman, J. H., & Fisher, N. I. (1999). Bump hunting in high-dimensional data. *Statistics and Computing, 9*(2), 123-143. doi:10.1023/A:1008894516817

Groves, D. G. (2006). *New Methods for Identifying Robust Long-Term Water Resources Management Strategies for California.* (PhD PhD dissertation), Pardee RAND Graduate School, Santa Monica, CA. Retrieved from http://www.rand.org/pubs/rgs_dissertations/RGSD196.html (RGSD-196)

Groves, D. G., & Lempert, R. J. (2007). A new analytic method for finding policy-relevant scenarios. *Global Environmental Change, 17*, 73-85. doi:10.1016/j.gloenvcha.2006.11.006

Hall, J. W., Lempert, R. J., Keller, K., Hackbarth, A., Mijere, C., & McInerney, D. J. (2012). Robust Climate Policies Under Uncertainty: A Comparison of Robust Decision Making and Info-Gap Methods. *Risk Analysis, 32*(10), 1657-1672. doi:10.1111/j.1539-6924.2012.01802.x

Jager, J. C. d. (2012). *Stabiliteitsprogramma Nederland*. Rijksoverheid Retrieved from http://www.rijksoverheid.nl/bestanden/documenten-en-publicaties/brieven/2012/04/27/geactualiseerde-stabiliteitsprogramma/geactualiseerde-stabiliteitsprogramma.pdf.

Keefe, R. (2012). *Reconsidering California Transport Policies: Reducing Greenhouse Gas Emissions in an Uncertain Future.* (PhD dissertation), Pardee RAND Graduate School. Retrieved from http://www.rand.org/pubs/rgs_dissertations/RGSD292.html

Keilman, N. (2005). The Impact of Demographic Uncertainty on Liabilities for Public Old Age Pensions in Norway. *New Zealand Population Review, 31*(1), 35-50. Retrieved from http://folk.uio.no/keilman/NZPR.pdf

Keilman, N. (2008). European demographic forecasts have not become more accurate during the past 25 years. *Population and Development Review, 34*(1), 137–153.

Kwakkel, J. H., Auping, W. L., & Pruyt, E. (2013). Dynamic scenario discovery under deep uncertainty: The future of copper. *Technological Forecasting & Social Change, 80*(4), 789-800. doi:10.1016/j.techfore.2012.09.012

Lassila, J., Valkonen, T., & Alho, J. M. (2014). Demographic forecasts and fiscal policy rules. *International Journal of Forecasting, In press*. doi:10.1016/j.ijforecast.2014.02.009

Lee, R., & Tuljapurkar, S. (1998). Uncertain Demographic Futures and Social Security Finances. *The American Economic Review, 88*(2), 237-241. Retrieved from http://www.jstor.org/stable/116926

Lempert, R. J., Groves, D. G., Popper, S. W., & Bankes, S. C. (2006). A General, Analytic Method for Generating Robust Strategies and Narrative Scenarios. *Management Science, 52*(4), 514-528. doi:10.1287/mnsc.1050.0472

Lempert, R. J., Popper, S. W., & Bankes, S. C. (2003). *Shaping the Next One Hundred Years: New Methods for Quantitative, Long-Term Policy Analysis* (MR-1626). Retrieved from Santa Monica, CA, USA: http://www.rand.org/pubs/monograph_reports/2007/MR1626.pdf

Lempert, R. J., Schlesinger, M. E., & Bankes, S. C. (1996). When We Don't Know the Cost or the Benefits: Adaptive Strategies for Abating Climate Change. *Climatic Change, 33*, 235-274. doi:10.1007/BF00140248

Logtens, T. (2011). *Demographic aging and its implications for the Dutch welfare state: an Exploratory System Dynamics approach*. Retrieved from Delft:

Meyerson, N., & Sabelhaus, J. (2000). Uncertainty in Social Security Trust Fund Projections. *National Tax Journal, 53*(3), 515-529. Retrieved from https://www.econ.umd.edu/research/papers/522/download/285

Morecroft, J. D. W. (1982). A Critical Review of Diagramming Tools for Conceptualizing Feedback System Models. *Dynamica, 8*(1), 20-29.

Popper, S. W., Griffin, J., Berrebi, C., Light, T., & Daehner, E. M. (2009). *Natural Gas and Israel's Energy Future: A Strategic Analysis Under Conditions of Deep Uncertainty*. Retrieved from Santa Monica, CA: http://www.rand.org/pubs/technical_reports/TR747.html

Pruyt, E., & Logtens, T. (2015). Model-Based Exploration of Societal Aging in the Netherlands. *International Journal of System Dynamics Applications, 4*(1).

Rittel, H. W. J., & Webber, M. M. (1973). Dilemmas in a General Theory of Planning. *Policy Sciences, 4*(1973), 155-169. doi:10.1007/BF01405730

Slobbe, L. C. J., Smit, J. M., Groen, J., Poos, M. J. J. C., & Kommer, G. J. (2007). *Kosten van Ziekten in Nederland. Trends in de Nederlandse Zorguitgaven 1999-2010*. Retrieved from Bilthoven: http://www.rivm.nl/bibliotheek/rapporten/270751023.pdf

Sterman, J. D. (2000). *Business Dynamics: Systems Thinking and Modeling for a Complex World*. New York: McGraw.

Sutrisno, A., & Handel, O. (2012). *Dynamic Aging Population in Germany: A case study about demographic change.* Paper presented at the International Conference of the System Dynamics Society, St. Gallen.

Thompson, J. P., Riley, C. M., Eberlein, R. L., & Matchar, D. B. (2012). Future living arrangements of Singaporeans with age-related dementia. *International Psychogeriatrics, 24*(10), 1592-1599.

Ventana Systems. (2010). *Vensim Reference Manual*. Harvard, MA: Ventana Systems.

Walker, W. E., Haasnoot, M., & Kwakkel, J. H. (2013). Adapt or Perish: A Review of Planning Approaches for Adaptation under Deep Uncertainty. *Sustainability, 5*(3), 955-979. doi:10.3390/su503955

Wet verhoging pensioenleeftijd, extra verhoging AOW en flexibilisering ingangsdatum AOW, 33.046 (2011).

6 Ebola

The implementation and effects of policies can be uncertain, as at least the moment of implementation, the way of implementation, and the exact policies' effects are often unknown. However, in many modelling studies the direct policy effects are assumed to be known. For example, when it comes to bringing epidemics or outbreaks under control, the policies or interventions are generally known. It is unknown though how big their effect will be, how much time it takes to make them operational, and based on what rules the expansion of intervention capacities is based.

Another issue is that models can create futures that are plausible in and of themselves, given the bandwidths of existing uncertainties, but are not corroborated by the way the reality unfolds. For example, a model about virus transmission may have a useful structure, and be able to produce runs how the virus might have propagated, but in reality did not. In consolidative modelling, this may result in adapting the estimate for each parameter. In exploratory modelling, similarly an analyst may reduce the bandwidths of the uncertain parameters in order to force all runs to display behaviour within the actual historic data bounds. Due to the non-linear mapping of inputs to outputs, this may, however, lead to less runs within the historic bounds as well. In that case, the number of plausible explanations of behaviour of interest is also reduced, which is at odds with the goal of exploratory modelling.

In this paper, we introduce an exploratory SEIR (i.e., Susceptible, Exposed, Infected, Recovered) model extended with endogenous intervention capacity development of the 2014 Ebola outbreak in Liberia. This research illustrates – similarly to the ageing research – that exploratory modelling in this case requires a broader scope than conventional transmission models, as alternative explanations besides virus transmission could be important. We did this by modelling the intervention capacities (i.e., the policies) endogenously. Further, it illustrates the importance of policy uncertainty, as the effects of the intervention capacities, and the speed of their implementation, was deeply uncertain. Finally, we eliminated those runs from the set that were out of historic bounds by using available data on how the epidemic was evolving, specifically the reported number of cases combined with the uncertainty around that data. In this way, we did not reduce the number of plausible explanations of runs of interest.

In the synthesis chapter, I will use this paper to discuss policy uncertainty, combined with endogenous policy development, and how to deal with runs that are plausible given the uncertainty bandwidth of input parameters, but are falsified by the way reality unfolds.

This paper was originally published as:

Auping, W. L., Pruyt, E., & Kwakkel, J. H. (2017). Simulating endogenous dynamics of intervention-capacity deployment: Ebola outbreak in Liberia. *International Journal of Systems Science: Operations & Logistics, 4*(1), 53-67. doi:10.1080/23302674.2015.1128576

Simulating Endogenous Dynamics of Intervention-Capacity Deployment: Ebola Outbreak in Liberia

Willem L. Auping, The Hague Centre for Strategic Studies, The Netherlands; Policy Analysis Section, Faculty of Technology, Policy and Management, Delft University of Technology, The Netherlands

Erik Pruyt, Policy Analysis Section, Faculty of Technology, Policy and Management, Delft University of Technology, The Netherlands

Jan H. Kwakkel, Policy Analysis Section, Faculty of Technology, Policy and Management, Delft University of Technology, The Netherlands

Abstract

During the first months, the 2014 outbreak of the Ebola Virus in West Africa was characterized by inadequate intervention capacities. In this paper, we investigate (i) the influence of limited but dynamic intervention capacities and their effect on the effective reproduction number, and (ii) the effects of proactive versus reactive intervention approaches. We use a transmission model extended with dynamical intervention capacities. Taking into account a bandwidth for potential over- and underreporting in reported Ebola Virus Disease cases, the model is used to generate ensembles of plausible scenarios. Next, it is used for testing the effectiveness of more proactive approaches in extending intervention capacities across these scenarios. We show that reactive approaches in extending intervention capacities can lead to continued under-capacity, and, consequently, to an increase of the effective reproduction number and to accelerated EBOV transmission. Proactive approaches, which take deployment delays, doubling times of diseases, and potential underreporting of the number of cases into account, help in limiting the total number of cases and deaths if the effective reproduction number in isolation is lower than the effective reproduction number outside of isolation. If the effective reproduction number in isolation is higher, proactive intervention policies still outperform reactive intervention policies.

Keywords: Ebola Virus Disease, Intervention capacity, Reproduction number, System Dynamics, Scenario Discovery

1. Introduction

The 2014 outbreak of Ebola Virus (EBOV) and, consequently, Ebola Virus Disease (EVD) in Liberia, Sierra Leone, Guinea, Senegal, Mali, Nigeria, Spain, and the United States of America (CDC, 2014; Gire et al., 2014) was by far the largest observed to date (WHO Ebola Response Team, 2014). The number of cases and deaths outnumbered the sum of all previous outbreaks. Where earlier outbreaks took place in rural or otherwise sparsely populated areas (Amblard et al., 1997; Borchert et al., 2011; Bwaka et al., 1999; "Ebola haemorraghic fever in Sudan," 1978; "Ebola haemorrhagic fever in Zaire, 1976," 1978; Okware et al., 2002; Pattyn, 1977; Roddy et al., 2012; Shoemaker et al., 2012), the 2014 outbreak distinguished itself by occurring in densely populated urban areas (WHO Ebola Response Team, 2014).

Dynamic transmission models can be used for intervention capacity planning for epidemics like the 2014 EVD outbreak. During the 2014 outbreak, dynamic transmission models have been used for estimating the basic reproduction number of Ebola, and for projecting the future development of the epidemic (Chowell, Hengartner, Castillo-Chavez, Fenimore, & Hyman, 2004; Lekone & Finkenstädt, 2006; WHO Ebola Response Team, 2014). However, projecting the dynamics of EBOV was, especially during the first few months, complicated by uncertainty about many input factors (Butler, 2014). Examples of uncertain factors include the case fatality ratio (Kucharski & Edmunds, 2014), and the basic reproduction number R_0 (Althaus, 2014; Fisman, Khoo, & Tuite, 2014; WHO Ebola Response Team, 2014). Further, the actual number of cases during the outbreak in West Africa was believed to be considerably higher than the reported number of cases (Meltzer et al., 2014), since the infrastructure to diagnose new cases and identify contamination epicenters was insufficient. The insufficiency of the infrastructure to identify new cases and epicenters of contamination, and the resulting underestimation of the problem, contributed to the continued spreading of the disease (WHO Ebola Response Team, 2014).

With the continued EBOV spreading, capacities like medical staff, hospitals, isolation facilities, and tracing officers were being scaled up dynamically to curb the outbreak. Simultaneously, efforts were initiated to speed up the development and provision of Ebola medication and vaccines. That is, the extent of these capabilities in the region changed significantly over time. These intervention capacities, and their dynamics, therefore need to be incorporated inside transmission models aimed at projecting the future development of the epidemic. This is not new. For example, Bachinsky and Nizolenko (2013) combined a SEIR model (i.e., a model with separate compartments for Susceptible (S), Exposed (E), Infectious (I), and Recovered (R) populations) with constant isolation bed capacities. Studies on influenza also often include the influence of anti-viral medication and vaccination programs (Kenah, Chao, Matrajt, Halloran, & Longini, 2011; Klepac, Bjørnstad, Metcalf, & Grenfell, 2012; Luz, Vanni, Medlock, & Galvani, 2011; McCaw & McVernon, 2007; Moss, McCaw, & McVernon, 2011). However, none of the early Ebola studies made use of detailed dynamic sub-models of endogenous intervention capacity development for a broad range of intervention capacities.

In this paper, we present an extended SEIR model for EBOV propagation that includes intervention capacities endogenously. The model was developed and used early September 2014 to assess capability deployment needs in West Africa. The model presented here is parameterized for Liberia only. The uncertainty by which the EVD outbreak in West Africa was characterized is incorporated by means of large uncertainty ranges. This enables us to evaluate the influence of dynamic limits on EVD interventions on the effective reproduction number. That is, the effective reproduction number is modeled as the result of a SEIR model extended with endogenous intervention capacities. The effective reproduction number, as it is used here, therefore relates to the average number of infections per single infection given the dynamics of the population immunity level and the dynamics of the intervention level.

We explore the dynamics of the model under uncertainty in order to explain how epidemic risk and intervention capacities interact, what the consequences may be of their interaction, and how the use of dynamic transmission models with integrated dynamic intervention capacities can inform planning of intervention capacities during future outbreaks.

The setup of this paper is as follows. First, we present the SEIR model extended with model structures with limiting intervention capacities (i.e., isolation, health workers, tracing officers, and eventually vaccines), and the experimental setup. Second, we discuss the results of our analysis for the cumulative number of cases, the effective reproduction number, and doubling time. Third, we discuss our findings and provide concluding remarks.

2. Methods

We developed a model combining a SEIR core with possible interventions aimed at curbing the Ebola epidemic in West Africa. The model was developed using the System Dynamics (SD) method (Forrester, 1961; Pruyt, 2013; Sterman, 2000) and was used for exploratory purposes (Bryant & Lempert, 2010). SD is a method for modeling and simulating dynamically complex systems or issues characterized by causal relations, feedback loops, accumulations, and delays. SD models are essentially systems of differential equations or integral equations (Lane, 2000). Simulating the dynamic behavior of the modeled system through numeric integration of these equations results in a simulation run displaying the behavior of the modeled system over time. Simulation runs can be used to analyze problems related to the system, and to evaluate the effects of policy interventions in these systems. SD is regularly used to study disease dynamics and health policy (Sterman, 2000; K. M. Thompson & Duintjer Tebbens, 2009). In this particular case, we used it to explore the consequences of the different combinations of uncertainties on the dynamics of the epidemic, and test the effects of different intervention strategies.

2.1. Model description

We started with the traditional SEIR model. The central structure of the model contains state variables, aka stock variables, for the susceptible, exposed, infectious, and recovered sub-populations (Fig.). Mathematically speaking, these stock variables are integral equations. We made several changes to this basic SEIR structure. We divided the infectious population in a critical phase (*infectious population*) and a recovery phase for survivors of the critical phase, where patients may either recover or die. The recovering patients are still infectious. Therefore, they were modeled using a second stock variable, the *infectious survived population*, who are recovering and will survive. We applied this subdivision to both the infectious population in isolation (*isolated infectious population* and *isolated survived population* in Fig.) and the infectious population outside of isolation and treatment centers (*infectious population* and *infectious survived population*).

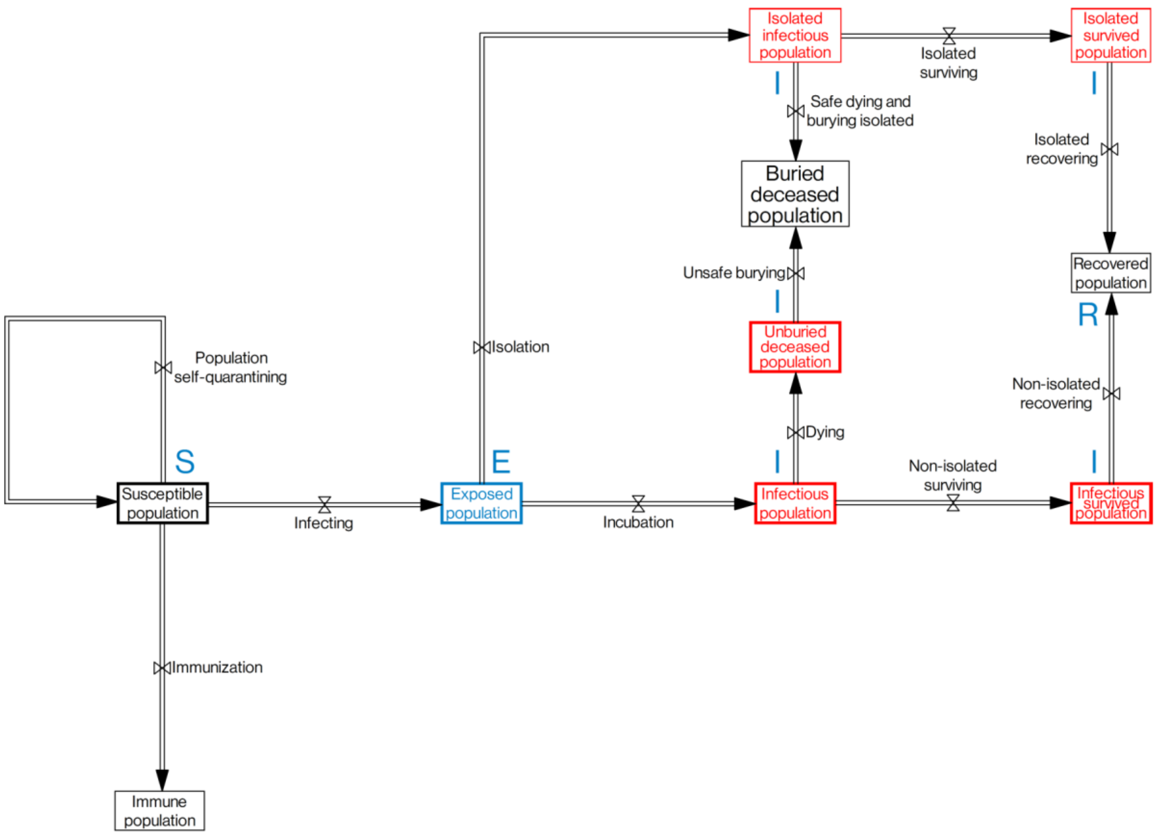


Fig. 1. Stock-flow structure of the extended (other factors and causal relations are not shown) SEIR model containing isolated population stocks, and the immune population due to vaccination. SEIR elements are indicated with their respective letters as well. Subscripted stocks have a bold border, infectious stocks are red, and the exposed population is blue.

Further, we subdivided (i.e., vectorised or subscripted) these population stocks in order to take potential self-quarantining behavior of the population into account. The S, E, and I stocks outside isolation (i.e., *Susceptible population*, *Exposed* *population*, *Infectious population*, and *Infectious survived population*), and the flows between these stocks, contain this subdivision. In Fig. 1, these stock variables have a bold border. Introducing this structure is important, as a succesful societal response to an outbreak leads to a significant decrease in the necessary intervention capacities like treatment and isolation capacity (Pruyt, Auping, & Kwakkel, 2015). Treatment and isolation capacity refers here to Ebola Treatment Centres (ETCs) and to Community Care Centres (CCCs). ETCs provide care to suspected and confirmed cases while attempting to prevent infection of healthcare workers and members of the community. Small CCCs ensure that patients are isolated in areas with insufficient ETC bed capacity or in remote areas without access to ETCs.

The basic SEIR model was further extended with treatment and isolation capacity, not distinguishing between ETCs and CCCs. In our model, this extension consists of two stock variables: one for the critical *Infectious population* and one for the *Infectious survived population*. We further included a stock variable for the *Unburied deceased population*. Finally, we included a stock variable for the *Immune population*, which contains the population that would be vaccinated after vaccines would have become available. This *Immune population*, and the *Recovered population*, are assumed to be no longer susceptible to EBOV.

In our model, intervention capacities are restricted and, unless specified differently, reactive. That is, we included the endogenous dynamic development of the availability of beds in treatment and isolation capacity, health workers, tracing officers, and vaccines. They are adapted to the needs, albeit delayed. This way, the numbers of health workers, tracing officers, and available vaccines increase in response to the dynamics of the epidemic. Scaling up of intervention capacities is delayed. That is why we added stocks for the preparation of intervention capacity and the available intervention capacity, and time delays between these stocks that slow the response to the epidemic (Fraser, Riley, Anderson, & Ferguson, 2004).

For health workers, the possibility of getting infected by EBOV and consequentially dying of EVD (Hewlett & Hewlett, 2005) is taken into consideration, thus reducing their availability. We assume that fully recovered healthcare workers will try to continue their efforts after an extensive recovery time. Further, healthcare workers may be recruited domestically or from outside the region. All additional physicians needed are nevertheless assumed to be foreign for it was assumed that the very small group of domestic physicians were already working full time.[4] Only a small portion of the susceptible population in West Africa is considered suitable for

[4] According to Liberia's ambassador to the United States, Liberia has about 50 doctors — about one for every 90,000 citizens, not counting foreign physicians (see http://www.bbc.com/news/world-africa-29516663 and https://www.washingtonpost.com/world/africa/liberia-already-had-only-a-few-dozen-of-its-own-doctors-then-came-ebola/2014/10/11/dcf87c5c-50ac-11e4-aa5e-7153e466a02d_story.html). The CIA's World Fact Book reports that in 2008 there were 0.01 physicians per 1000 inhabitants (see https://www.cia.gov/library/publications/the-world-factbook/fields/2226.html - last consulted on 10/09/2015).

nursing since they are not trained to protect themselves properly, but a larger part of the recovered population is suitable for nursing, since they are immune. If the medical staff capacity is not sufficient to provide for the necessary isolation capacity, the isolation capacity is limited by the available staff. This corresponds to closing down EVD treatment centers due to lack of staff.

Finally, the effective reproduction number is also included endogenously in the model (i.e., the effective reproduction number is dynamic). The effective reproduction number is calculated here as the daily rate of infections caused by the total infectious population in and outside isolation, multiplied by the average period during which infectious individuals are infectious. The effective reproduction number is thus the weighted average of the basic reproduction numbers in and outside isolation, calculated at each moment in time. It can go down if measures are sufficient, but it can also go up if, over time, intervention measures prove to be insufficient. The effective reproduction number is approximated the model by the ‘reproduction ratio’ which is calculated as the product of the sum of all infections and the sum of the average recovery time of survivors and the average period critical condition, divided by the sum of all infectious patients. The doubling time of the number of cases is approximated by ln2 divided by the fractional growth rate of the number of cumulative cases. The latter variable is calculated as the increase in the number of cases divided by the total cumulative exposed cases. The increase in the number of cases is calculated as the exposed population divided by the incubation period.

2.2. Experimental setup

The model was implemented in the Vensim modeling software (Ventana Systems, 2010) and was parameterized for the Liberian situation. The model contains 161 variables, of which 20 were subdivided for hygienic and normal behaving population, and 35 parameters were considered uncertain. We simulated the model for 400 days, with a time step of 0.25 days using the Runge-Kutta 4 auto numerical integration method. For the 35 uncertain parameters, we used a Latin Hypercube sampling approach, based on uniform distributions with the ranges displayed in Table 1. The parameter ranges are specified in function of the model structure and in relation to other parameter ranges. In this model, the variable ‘vaccinations’ depends, for example, on six variables, one of which is the ‘Vaccination speed’. A vaccination speed of 240 vaccines per person per day then means that, if vaccines are available, 240 people can be vaccinated per medical worker per day.

Some ‘soft’ variables and parameters are included to account for uncertain but plausible effects. For example, the ‘effect of self-quarantining behavior’ represents the effect through which more hygienic behavior causes the infectivity to decrease.

We generated 10,000 samples simulating this model using the open source EMA workbench (EMA Group, 2011) from https://github.com/quaquel/EMAworkbench. The model documentation and model are available as online supplementary materials on https://github.com/ep77/Ebola-Model-with-Endogenous-Response. The model equations are also available as online supplementary material. Visualizations and analyses are provided in an IPython notebook on http://nbviewer.ipython.org/gist/ep77/796491369b0e6fe84b4d.

Table 1. Model inputs considered to be uncertain. Factors for which no references exist, are indicated as assumptions.

Variable name	Unit	Min	Max	References
Average contact rate infectious population	1/Day	0.3	0.9	(Althaus, 2014; WHO, 2014b; WHO Ebola Response Team, 2014)
Average development time isolation facilities	Day	4.2	18.8	Derived from reports like Camacho et al. (2014)
Average extra recovery time survivors	Day	0.5	4.66	(WHO Ebola Response Team, 2014); Derived from analysis
Average time staff active	Day	185	341	Derived from analysis
Average time until burial	Day	0.5	2	(WHO, 2014b)
Average time until return diseased health workers	Day	21	60	Assumption
Average period critical condition	Day	4	9	(WHO Ebola Response Team, 2014)
Case fatality rate in isolation relative to outside isolation	Dimensionless	0.43	0.73	Broad bandwidth around data from 3; Derived from analysis
Case fatality rate outside isolation	Dimensionless	0.45	0.86	(WHO Ebola Response Team, 2014)
Contact rate before funeral	1/Day	0.32	0.97	Derived from (Camacho et al., 2014; WHO Ebola Response Team, 2014)
Contacts to be traced per quarantined patient	Contact/Person	5.47	40	(Bachinsky & Nizolenko, 2013)
Contacts traceable per tracer per day	Contact/ (Person*Day)	10	40	(Bachinsky & Nizolenko, 2013)
Delay time development new vaccines	Day	250	350	Assuming that vaccines will be available in first or second quarter of 2015
Doctors per nurse	Dimensionless	0.12	0.46	Assumption; Derived from analysis
Effect of self-quarantining behavior	Dimensionless	2.28	20	Assumption
Fraction recovered population useful as medical staff	Dimensionless	0.000458	0.043	Assumption; Derived from analysis
Fraction susceptible population useful as medical staff	Dimensionless	1.86E-06	0.000189	Assumption; Derived from analysis
Incubation period	Day	7	15	WHO Ebola Response Team (2014)
Initial exposed population	Person	50	100	WHO (2014b)
Initial isolation capacity	Person	120	600	WHO (2014b)
Initial relative susceptible hygienic population	Dimensionless	0.01	0.2	Assumption
Initial tracing personnel	Person	5	30	Assumption
Initial vaccines in preparation	Vaccine	4	20	Assumption
Lifetime isolation capacity	Day	180	360	Assumption / derived from analysis
Medical staff creating awareness	1/Day	5	100	Assumption
Medical staff per new case	1/Day	0.2	0.5	WHO (2014b)
Preparing time foreign staff	Day	14	60	Assumption; Derived from analysis
Recognition rate diseased	Dimensionless	0.2	0.95	Broad bandwidth around WHO (2014b)
Relative reduction in infectivity due to isolation	Dimensionless	0.7	5	Assumption
Training time new staff	Day	3	10	Assumption
Vaccination speed	Vaccine/ (Person*Day)	50	240	Assumption (estimate)

3. Results

3.1. Scenario selection

The initial ensemble of 10,000 simulations contained a wide range of plausible evolutions of the epidemic. Only a subset of these evolutions was consistent with the outbreak observed in Liberia, as SEIR models can produce simulations of both very lethal and very non-lethal outbreaks. Due to the non-linear nature of these models, outputs of simulations could even fall outside of plausible ranges for combinations of input uncertainties within ranges that are known to be plausible. Therefore, we post-processed the ensemble by selecting only those simulations where the cumulative number of Ebola cases fell within a range of 80% to 250% of the WHO data on the total cumulative number of Ebola cases on 3 September 2014 (WHO, 2014b). Although we have used this model at later moments in time, we present our simulation results calibrated to the WHO data of 3 September 2014, because this ensemble of simulations provides a good illustration of the uncertainty we were facing at the time, as well as the potentially devastating impact of an outbreak without additional policies and changes in behavior. The broad uncertainty range (of 80% to 250%) was used at the time, because it was argued that the WHO data significantly underreported the actual number of EVD cases (Meltzer et al., 2014), although some over-reporting could not be ruled out either. Following this method, we selected 3041 scenarios out of the total of 10,000 simulation runs. This process is depicted in Fig. 2. More visualizations of the initial ensemble, the screening, and subsequent analyses are provided in the online IPython notebook.

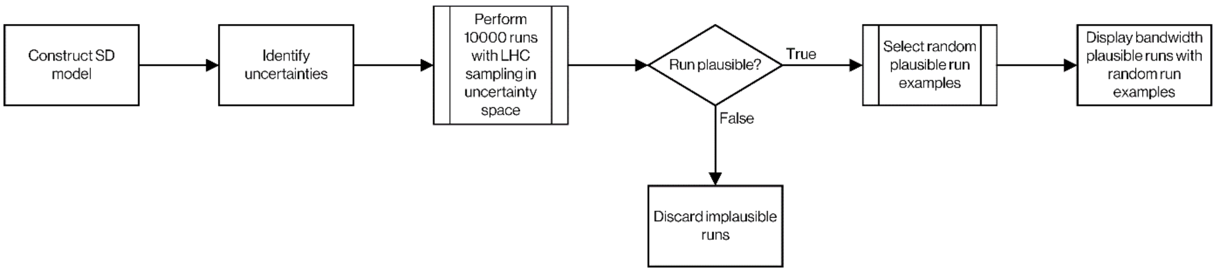


Fig. 2. Flowchart of the experimental setup and post-processing of the ensembles of scenarios

The ensemble (in shaded blue) and a randomly selected set of 30 out of the ensemble of 3041 scenarios are displayed in Fig. 3a and Fig. 3b (these figures differ only in terms of the scales of the y-axes). The ensemble consists of different plausible projections of the simulated number of ‘Actual cases’ (i.e., the total ‘Cumulative exposed cases’ in the model) that were consistent with the WHO data on 3 September 2014. The runs start on 22 June 2014 (t=0), after the WHO reported the first 51 cases in Liberia. In the best-case scenarios, the underreporting of cases is limited due to sufficient tracing capacity. In these scenarios, the effective reproduction number gradually declines as the intervention becomes more effective.

In other scenarios, the tracing capacity is inadequate, which leads to inadequate developments of isolation capacity and medical staff. In these cases, the development of the intervention capacity is overly delayed. The non-isolated population consequently peaks considerably earlier than the isolated population. This results in an order of magnitude difference between the maximum non-isolated infectious and the maximum isolated infectious. The required isolation and treatment capacities are not available in these worst-case scenarios, even if changes in population behavior would be effective (e.g., when part of the diseased actively seek help at treatment centers, even if they were not traced).

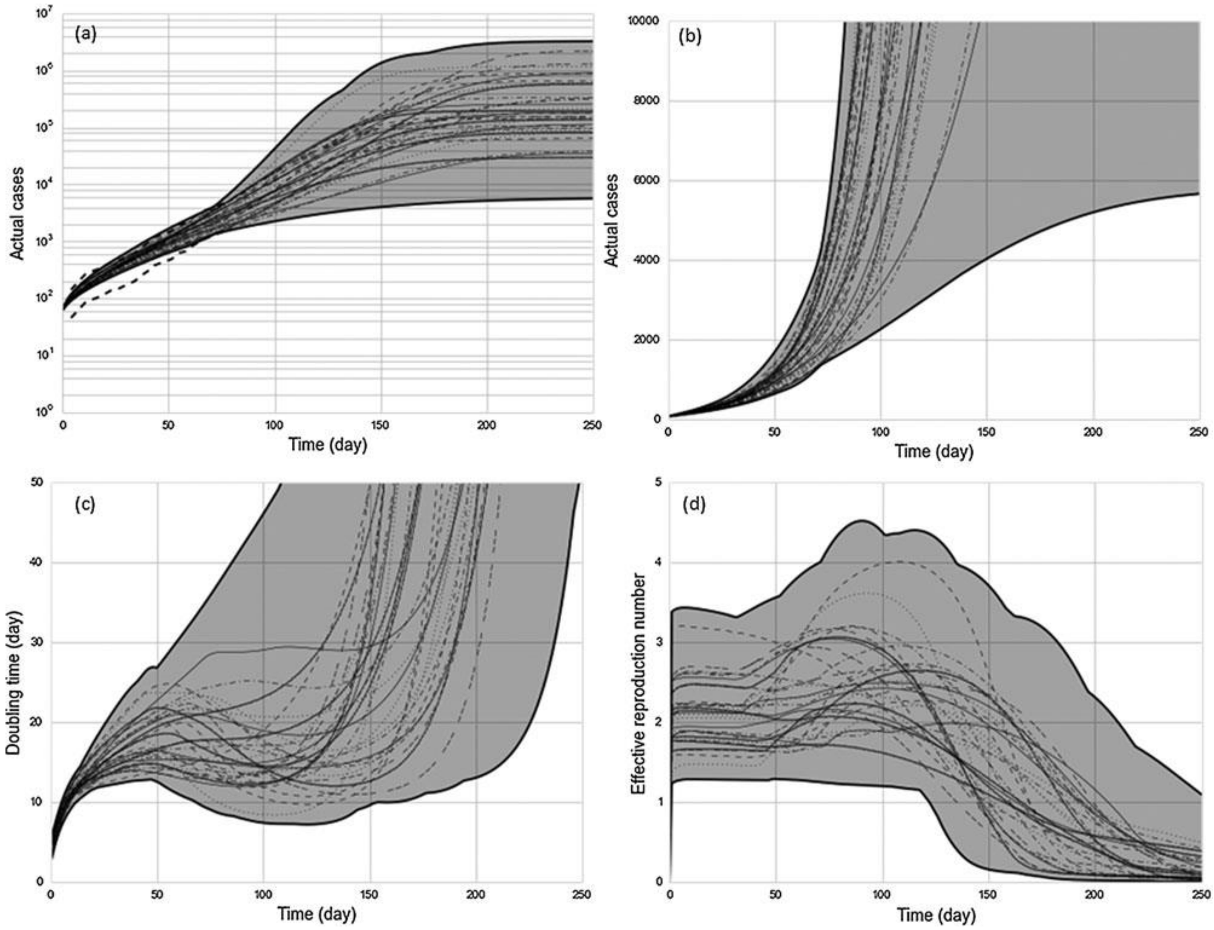


Fig. 3. Dynamics of 30 randomly selected runs and the ensembles for: (a) the cumulative number of cases (i.e., the total ‘Cumulative exposed cases’) on a logarithmic y-axis (with bounds around the historic WHO data displayed with dashed lines), (b) the same cumulative number of cases on a non-logarithmic y-axis , (c) the effective reproduction number, and (d) the doubling time of cases

Limits in the EBOV intervention capability influence the speed with which the virus is transmitted. That is, starting from a situation in which there is a lack of intervention capacity, an increasing lack of intervention capacity may even result in an increase in the speed of virus propagation. However, so do ineffective measures, or a rise in ineffectiveness of measures. That is, the speed of virus transmission may also increase if individuals with EVD who end up in isolation and treatment centers infect more individuals than individuals with EVD who do not end up in isolation and treatment centers. This may for example happen if EVD cases are not recognized as such, if many individuals with similar symptoms – some of whom have EVD and most of whom do not have EVD at first – spend a relatively long time at the same center, if insufficient or ineffective protective measures are taken by non-infected individuals in isolation and treatment centers (e.g., health workers and patients with other diseases with similar symptoms), or if the trip to these centers results in many new infections.

Many problematic scenarios are characterized by at least one of two virus accelerating effects: (i) failure to isolate the large majority of EVD cases which leads to an increase in the reproduction rate of the disease, causing an increase in the effective reproduction number, and (ii) successful isolation but with higher infectivity rates in isolation than outside of isolation which may lead to an increase in the effective reproduction rate of the disease, causing an increase in the effective reproduction number. The results of both accelerating effects are visible in Fig. 3c, which shows how the endogenously modeled effective reproduction number develops in 30 randomly selected scenarios out of the ensemble of 3041 scenarios as well as the ensemble itself. As a consequence, the doubling time of the number of cases declines (Fig. 3d). Finally, when EBOV transmission has peaked, the doubling time of cases rises quickly as the effective reproduction number falls below 1.

It is important to realize that there may be two reasons why scenarios show increased effective reproduction numbers. First, the effective reproduction number is the result of infectious people having contact with their surroundings (e.g., with family members, or with deceased during unsafe burials). If the relative share of the infectious population that cannot be isolated increases, due to limitations in either available beds or available trained and well-equipped staff, then the effective reproduction number could be expected to increase too. Second, many studies estimating the base reproduction number of EBOV or similar diseases assume that intervention capability is not available at the beginning of the epidemic, while its adequacy increases over time (e.g., Chowell et al., 2004; Chowell & Nishiura, 2014). However, that assumption may be wrong. In the case of the 2014 EBOV epidemic in West Africa, for example, it looks as though the adequacy of the intervention capability was first deteriorating over time (compare data in WHO, 2014b)), resulting in dynamics similar to those simulated here.

3.2. Effect of a more proactive approach

Responses to unforeseen outbreaks involving increases of intervention capabilities are mostly delayed. As a result, when new capacities become available they often fall short of the capacity that is actually required, especially when insufficient capacity further increases the speed with which the virus propagates. This is, for example, the case if a lack of tracing officers results in underestimation and underreporting of the speed with which the virus is propagating. Therefore, increasing intervention capacities requires a more proactive approach, for example by trying to anticipate future increases in cases, while taking irreducible delays in the development of new capacities, into account. We therefore introduce the following formula (Eq. 1), which is one way to capture proactive planning:

$$C_{t+1} = c_u * C_{t,des} * \left(1 + \left(\frac{\tau_C}{\tau_2}\right)\right) - C_t \qquad \text{Eq. 1}$$

Where:

C_{t+1} is the capacity to develop;

c_u is the expected underestimation factor of the number of EVD cases;

$C_{t,des}$ is the presently desired capacity;

τ_C is the delay on capacity development;

τ_2 is the doubling time for the number of EVD cases;

C_t is current available capacity.

This formula expresses that while preparing new intervention capacities, one should be prepared for those EVD cases that will arise during the preparation time, as well as the exposed population that will become infectious after the deployment of capacity additions. If the preparation time is relatively short compared to the doubling time, the necessary extra capacity is, therefore, smaller. Existing capacity may be subtracted from the capacity to develop. It should be noted, however, that in the case of underestimation of the number of cases, the desired capacity at that time should also be multiplied with the expected underestimation factor. The potential underestimation factor may be assessed by experts in the field, organizations like MSF or the WHO, or from the literature (WHO Ebola Response Team, 2014).

Fig. 4 shows the effects of a *Reactive response* policy (i.e., the light red envelope), a *Proactive* policy *from on day 110* on (i.e., the light blue envelope), and a *Proactive* policy *from day 72* on (i.e., the light green envelope) on the selected 3041 scenarios, as well as 30 randomly selected scenarios (reactive responses in red, proactive responses from day 110 on in blue, and proactive responses from day 72 on in green). Note that the envelopes are overlapping: overlap of light red and light blue shows as pink-purple, overlap of light green, light blue and light red shows as brown-grey, and overlap of light green and light red shows as yellow-green.

Fig. 4a shows that early post-processing under severe uncertainty (i.e., on 3 September 2014) results in rather similar ensembles in terms of the log-scaled cumulative number of Ebola cases. The underlying reason for this surprising result is that, in our worst case simulations, infectivity in isolation is not necessarily lower than infectivity outside of isolation. More and earlier isolation capacity may be problematic if it is ineffective. Again, the worst cases are either scenarios in which an initial underestimation of the size of the epidemic leads to an early increase in the reproduction number of the virus, or scenarios in which policies that are being implemented are counter-productive. In these worst cases, the EBOV outbreak is hard to curb. Fig. 4b and Fig. 4c nevertheless show that the earlier a more proactive approach is adopted, the earlier the effective reproduction number drops and the doubling time rises. Therefore, adopting a more proactive approach is beneficial across the ensemble, even if measures are not as effective as they could or should be. Adopting an effective proactive approach is what is really needed.

The dominance of proactive approaches becomes clearer when post-processing later in time. Fig. 4d shows the ensembles of the same policies post-processed between 80% and 150% of the number of reported cases on 10 December 2014. That is, all simulation runs that are not in line with the real-world estimates of 10 December 2014, plus/minus a slightly smaller uncertainty interval, are excluded from these ensembles. The upper bound applied in December 2014 is lower than the upper bound applied early September 2014 to account for more reliable data and a reduction in perceived uncertainty. Two observations could be derived from post-processing at this later point in time. First, 65.5% of all runs that are in line with the real-world data are generated with the adaptive policy from day 72 on, compared to 21.5% with the adaptive policy from day 110 on, and 13% with the reactive policy. That is, the adaptive policy from day 72 on corresponds better to what happened in the real-world than the adaptive policy from day 110 on, which, in turn, corresponds better to the real-world data than the reactive policy. This was to be expected given the massive international deployment of intervention capacities that took place in West Africa between September and December 2014. The real-world massive deployment could indeed be argued to have been proactive, because more was planned for than was needed at the moment of planning. Second, the long-term ensemble projections of the proactive approaches are much lower than the long-term ensemble projections of the reactive approach (see the Kernel Density Estimates of the terminal values at the right hand side of Fig. 4d).

The effectiveness of the intervention capacity development approach also largely depends on the phase of the epidemic. Proactive approaches are more effective when applied early in the growth phase of the epidemic. The potential gains are much smaller when the spread of the virus is already decreasing and the doubling time is increasing.

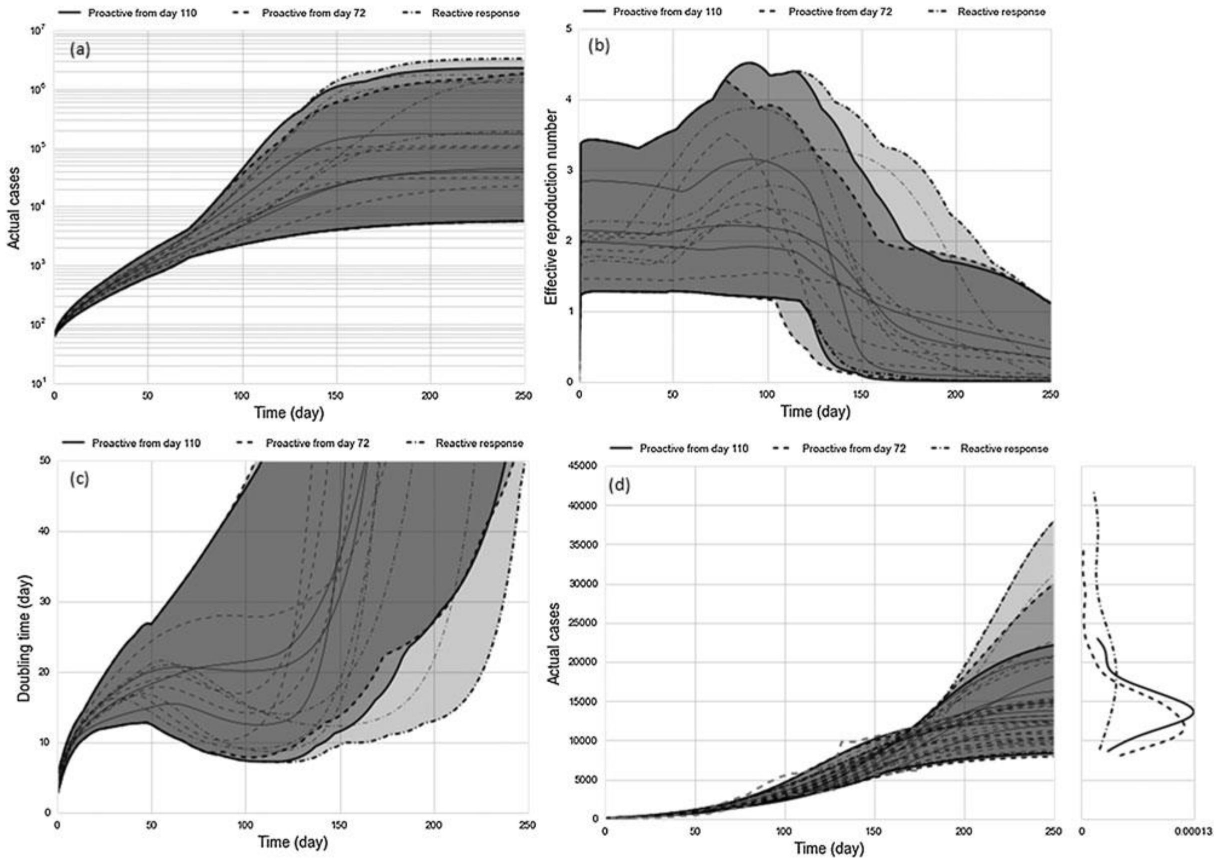


Fig. 4. Dynamics of 30 randomly selected runs and the ensembles of a *Reactive response* (red), a *Proactive* response *from day 110 on* (blue), and a *Proactive* response *from day 72 on* (green) for: (a) the cumulative number of actual cases, (b) the effective reproduction number, (c) the doubling time of the number of cases, and (d) the cumulative number of actual cases post-processed based on WHO data of 10 December 2014. Overlap of light red and light blue shows as pink-purple, overlap of light green, light blue and light red shows as brown-grey, and overlap of light green and light red shows as yellow-green.

4. Discussion

In this paper, we presented a simulation model with a detailed endogenous dynamic response to outbreaks. When developing simulation models to plan the response to an outbreak, it is important to explicitly account for the dynamic development of capabilities and the associated delays in the system. Models that do not include capabilities are likely to overestimate epidemics and may lead to unrealistic planning or calls not to use models for planning epidemic responses (Butler, 2014). However, models with static capabilities or capabilities development without delays are likely to underestimate epidemics and the epidemic responses needed.

We used the simulation model presented in this paper to generate ensembles of scenarios for the spread of the EBOV in Liberia and to project how the epidemic might evolve under deep uncertainty with reactive and proactive policies. Early real-world information was used to inform the model-building, and early real-world data

(August and early September 2014) and late real-world data (10 December 2014) was used to post-process the policy ensembles (i.e., to remove simulation runs that were not compatible with real-world data from the policy ensembles) in order to test how well each of these policy ensembles corresponded to the real data.

Many of the individual scenarios generated by this model were worse than what happened in reality. There are several reasons why the actual disease spread could have been expected to be less dramatic than the worst-case scenarios presented in this theoretical study. First, geographic spread of the population leads to slower virus transmission. Due to geographic spread and spreading of the virus, the real susceptible population at any one time (i.e., the real population-at-risk) was smaller than assumed in our simulation model, and the susceptible and infectious populations outside isolation were assumed to be perfectly mixed. Second, a high upper uncertainty bound was used to select the scenarios for this study. Third, large uniform uncertainty ranges were used for each of the uncertain parameters. Fourth, worst case planning assumptions about the effectiveness of ETCs were included in this study. Finally, this research was not exhaustive in terms of intervention measures considered. For example, essential medical supplies besides the medical staff and bed capacity in isolation were considered here.

A possible limitation of sampling from large uniform uncertainty ranges may be that simulation runs are generated with unrealistic combinations of inputs. Although post-processing introduces some correlation, the ensembles may still contain many implausible scenarios. Since these scenarios are not used as predictions, merely as sets of what-if analyses and as inputs for policy robustness testing, this is, according to us, not a major problem. After all, our focus is on testing the effectiveness of policies across large ensembles of scenarios (i.e., no matter what could happen), especially in case of worse case scenarios. If policies happen to be effective across all cases, even for implausible scenarios, then implausible scenarios do not necessarily need to be identified and eliminated. For example, given the uniform distribution of the ‘relative reduction in infectivity due to isolation’ variable between 0.7 and 5 (see Table 1), many scenarios are simulated in which infectivity rises due to increased isolation of EVD cases. Proactive isolation-oriented policies could be expected to perform poorly for these counter-intuitive scenarios. However, proactive policies seem to perform reasonably well across all scenarios, even across these least surprising or implausible scenarios.

In our model, we have assumed that the intervention capacities developed would not be hindered by lack of resources like skilled medical personnel from foreign countries. Resources are nevertheless limited, both in the model and in reality, due to erroneous planning and due to normal planning and implementation delays.

The same principle nevertheless applies to all capability and resource under-capacities, whatever their cause: Any under-capacity harms the effectiveness of the total intervention capability. That is, the entire intervention capability is as strong as the weakest non-redundant capacity in the chain.

Another limitation of our study relates to the consequences of the real-world geographic spread of virus transmission on real-world capacity planning. In this theoretical paper, we used a homogeneous mixing model, where any expected incidence and any capacity extension affects the whole population equally. In reality, heterogeneity and geographic spread mean that some parts of the population and territory are more heavily affected by the outbreak, which, given inherent uncertainty about the future geographic spreading of the virus, makes it more difficult to foresee where capacity expansions are needed. Although this limitation does not fundamentally alter the general insights of our study, it needs to be taken into account for real-world planning purposes. That is, either these suggested capacity additions are considered to be the absolute minimum capacity additions and estimates are revised upward based on local characteristics and spreading, or geospatial models should be used for real planning purposes. This is especially important in case of heterogeneous spreading in large heterogeneous regions.

5. Conclusions

In this article, we have presented a simulation model with endogenous response related to the 2014 Ebola outbreak in Liberia. Our simulations show that both delayed responses and timely but ineffective measures can cause the effective reproduction number to increase. The consequence of such situations may be that the growth of the actual number of cases accelerates significantly. These findings were derived from an extended SEIR model with endogenously modeled intervention capacities parameterized for the EBOV outbreak in Liberia.

In early September 2014, our research suggested that the effective reproduction number of the 2014 Ebola epidemic could increase compared to the measured effective reproduction number (WHO Ebola Response Team, 2014) if the capacities of the different interventions were not brought to the minimally required level over time. During the first months of the 2014 outbreak in Liberia, which was characterized by a significant shortfall in bed capacity due to a lack of health care staff and a lack of operational bed capacity in Ebola treatment units (WHO, 2014a), intervention capacities were insufficient and ineffective.

This under-capacity may be the result of the reactive response to the initial exponential growth of the number of EVD cases. Early proactive approaches in building up the total spectrum of intervention capacities decrease, on an ensemble level, the final scale of the epidemic, especially if intervention capacities turn out to be effective. More proactive approaches in expanding the intervention capacities may, therefore, help in controlling epidemics like the 2014 West Africa EBOV. Such proactive approaches would at least have to take into account how the development time of these capacities relates to the doubling time of the disease, and the factor by which the measured cases may be underreported (Farrar & Piot, 2014).

References

Althaus, C. L. (2014). Estimating the Reproduction Number of Ebola Virus (EBOV) During the 2014 Outbreak in West Africa. *PLoS Currents Outbreaks, 2014 Sep 2*(1). doi:10.1371/currents.outbreaks.91afb5e0f279e7f29e7056095255b288

Amblard, J., Obiang, P., Edzang, S., Prehaud, C., Bouloy, M., & Guenno, B. L. E. (1997). Identification of the Ebola virus in Gabon in 1994. *The Lancet, 349*(9046), 181-182. doi:10.1016/S0140-6736(05)60984-1

Bachinsky, A. G., & Nizolenko, L. P. (2013). A universal model for predicting dynamics of the epidemics caused by special pathogens. *Biomed Res Int, 2013*, 467078. doi:10.1155/2013/467078

Borchert, M., Mutyaba, I., Van Kerkhove, M. D., Lutwama, J., Luwaga, H., Bisoborwa, G., . . . Van Der Stuyft, P. (2011). Ebola haemorrhagic fever outbreak in Masindi District, Uganda: outbreak description and lessons learned. *BMC Infect Dis, 11*, 357. doi:10.1186/1471-2334-11-357

Bryant, B. P., & Lempert, R. J. (2010). Thinking inside the box: A participatory, computer-assisted approach to scenario discovery. *Technological Forecasting & Social Change, 77*, 34-49. doi:10.1016/j.techfore.2009.08.002

Butler, D. (2014). Models overestimate Ebola cases. *Nature, 515*(6 November 2014), 18. doi:10.1038/515018a

Bwaka, M. A., Bonnet, M.-J., Calain, P., Colebunders, R., De Roo, A., Guimard, Y., . . . Van den Enden, E. (1999). Ebola Hemorrhagic Fever in Kikwit, Democratic Republic of the Congo: Clinical Observations in 103 Patients. *Journal of Infectious Diseases, 179*(Supplement 1), S1-S7. doi:10.1086/514308

Camacho, A., Kucharski, A. J., Funk, S., Breman, J., Piot, P., & Edmunds, W. J. (2014). Potential for large outbreaks of Ebola virus disease. *Epidemics, 9*, 70-78. doi:10.1016/j.epidem.2014.09.003

CDC. (2014). 2014 Ebola Outbreak in West Africa. *Ebola Virus Disease.* Retrieved from http://www.cdc.gov/vhf/ebola/outbreaks/2014-west-africa/index.html

Chowell, G., Hengartner, N. W., Castillo-Chavez, C., Fenimore, P. W., & Hyman, J. M. (2004). The basic reproductive number of Ebola and the effects of public health measures: the cases of Congo and Uganda. *J Theor Biol, 229*(1), 119-126. doi:10.1016/j.jtbi.2004.03.006

Chowell, G., & Nishiura, H. (2014). Transmission dynamics and control of Ebola virus disease (EVD): A Review. *BMC Medicine, 12*, 196. doi:10.1186/s12916-014-0196-0

Ebola haemorraghic fever in Sudan. (1978). *Bulletin of the World Health Organization, 56*(2), 247-270. Retrieved from http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2395561/

Ebola haemorrhagic fever in Zaire, 1976. (1978). *Bulletin of the World Health Organization, 56*(2), 271-293. Retrieved from http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2395567/

EMA Group. (2011, 2011). Exploratory Modelling & Analysis (EMA) Workbench. Retrieved from http://simulation.tbm.tudelft.nl/ema-workbench/contents.html

Farrar, J. J., & Piot, P. (2014). The Ebola Emergency — Immediate Action, Ongoing Strategy. *New England Journal of Medicine, 371*, 1545-1546. doi:10.1056/NEJMe1411471

Fisman, D., Khoo, E., & Tuite, A. (2014). Early Epidemic Dynamics of the West African 2014 Ebola Outbreak: Estimates Derived with a Simple Two-Parameter Model. *PLoS Currents*. doi:10.1371/currents.outbreaks.89c0d3783f36958d96ebbae97348d571

Forrester, J. W. (1961). *Industrial Dynamics*. Cambridge, MA: MIT Press.

Fraser, C., Riley, S., Anderson, R. M., & Ferguson, N. M. (2004). Factors that make an infectious disease outbreak controllable. *Proc Natl Acad Sci U S A, 101*(16), 6146-6151. doi:10.1073/pnas.0307506101

Gire, S. K., Goba, A., Andersen, K. G., Sealfon, R. S. G., Park, D. J., Kanneh, L., . . . Sabeti, P. C. (2014). Genomic surveillance elucidates Ebola virus origin and transmission during the 2014 outbreak. *Science, 345*(6202), 1369-1372. doi:10.1126/science.1259657

Hewlett, B. L., & Hewlett, B. S. (2005). Providing Care and Facing Death: Nursing During Ebola Outbreaks in Central Africa. *Journal of Transcultural Nursing, 16*(4), 289-297. doi:10.1177/1043659605278935

Kenah, E., Chao, D., Matrajt, L., Halloran, M. E., & Longini, I. (2011). The Global Transmission and Control of Influenza. *PLoS One, 6*(5), e19515. doi:10.1371/journal.pone.0019515.g001

Klepac, P., Bjørnstad, O. N., Metcalf, J. E., & Grenfell, B. T. (2012). Optimizing Reactive Responses to Outbreaks of Immunizing Infections: Balancing Case Management and Vaccination. *PLoS One, 7*(8), e41428. doi:10.1371/journal.pone.0041428

Kucharski, A. J., & Edmunds, W. J. (2014). Case fatality rate for Ebola virus disease in west Africa. *The Lancet, 384*(9950), 1260. doi:10.1016/S0140-6736(14)61706-2

Lane, D. C. (2000). Diagramming conventions in system dynamics. *Journal of the Operational Research Society, 51*(2), 241-245.

Lekone, P. E., & Finkenstädt, B. F. (2006). Statistical inference in a stochastic epidemic SEIR model with control intervention: Ebola as a case study. *Biometrics, 62*(4), 1170-1177. doi:10.1111/j.1541-0420.2006.00609.x

Luz, P. M., Vanni, T., Medlock, J., & Galvani, A. (2011). Dengue vector control strategies in an urban setting: an economic modelling assessment. *The Lancet, 377*, 1673-1680. doi:10.1016/S01406736(11)60246-8

McCaw, J. M., & McVernon, J. (2007). Prophylaxis or treatment? Optimal use of an antiviral stockpile during an influenza pandemic. *Math Biosci, 209*(2), 336-360. doi:10.1016/j.mbs.2007.02.003

Meltzer, M. I., Atkins, C. Y., Santibanez, S., Knust, B., Petersen, B. W., Ervin, E. D., . . . Washington, M. L. (2014). Estimating the Future Number of cases in the ebola epidemic - Liberia and Sierra Leone, 2014-2015. *Morbidity and Mortality Weekly Report. Surveillance Summaries (Washington, D.C.: 2002), 63*(3), 1-14. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/25254986

Moss, R., McCaw, J. M., & McVernon, J. (2011). Diagnosis and antiviral intervention strategies for mitigating an influenza epidemic. *PLoS One, 6*(2), e14505. doi:10.1371/journal.pone.0014505

Okware, S. I., Omaswa, F. G., Zaramba, S., Opio, A., Lutwama, J. J., Kamugisha, J., . . . Lamunu, M. (2002). An outbreak of Ebola in Uganda. *Tropical Medicine and International Health, 7*(12), 1068-1075. doi:10.1046/j.1365-3156.2002.00944.x

Pattyn, S. R. (Ed.) (1977). *Ebola Virus Haemorrhagic Fever*. Antwerp, Belgium: Elsevier, North Holland.

Pruyt, E. (2013). *Small System Dynamics Models for Big Issues: Triple Jump towards Real-World Complexity*

Pruyt, E., Auping, W. L., & Kwakkel, J. H. (2015). Ebola in West Africa: Model-Based Exploration of Social Psychological Effects and Interventions. *Systems Research and Behavioral Science, 32*(1), 2-14. doi:10.1002/sres.2329

Roddy, P., Howard, N., Van Kerkhove, M. D., Lutwama, J., Wamala, J., Yoti, Z., . . . Borchert, M. (2012). Clinical Manifestations and Case Management of Ebola Haemorrhagic Fever Caused by a Newly Identified Virus Strain, Bundibugyo, Uganda, 2007–2008. *PLoS One, 7*(12). doi:10.1371/journal.pone.0052986

Shoemaker, T., MacNeil, A., Balinandi, S., Campbell, S., Wamala, J. F., McMullan, L. K., . . . Nichol, S. T. (2012). Reemerging Sudan Ebola Virus Disease in Uganda, 2011. *Emerg Infect Dis, 18*(9). doi:10.3201/eid1809.111536

Sterman, J. D. (2000). *Business Dynamics: Systems Thinking and Modeling for a Complex World*. New York: McGraw.

Thompson, K. M., & Duintjer Tebbens, R. J. (2009). Using system dynamics to develop policies that matter: global management of poliomyelitis and beyond. *System Dynamics Review, 24*(4), 433-449. doi:10.1002/sdr.419

Ventana Systems. (2010). *Vensim Reference Manual*. Harvard, MA: Ventana Systems.

WHO. (2014a). *Ebola Response Roadmap Situation Report - 15 October 2014*. Retrieved from Geneva: http://apps.who.int/iris/bitstream/10665/136508/1/roadmapsitrep15Oct2014.pdf

WHO. (2014b). Ebola virus disease - Disease outbreak news. *Global Alert and Response (GAR).* Retrieved from http://www.who.int/csr/don/archive/disease/ebola/en/

WHO Ebola Response Team. (2014). Ebola Virus Disease in West Africa — The First 9 Months of the Epidemic and Forward Projections. *New England Journal of Medicine, 371*(16), 1481-1495. doi:10.1056/NEJMoa1411100

7 Synthesis

This dissertation presented five papers that build on four case studies. All contribute to relevant policy issues, and aim at the model-based exploration of the consequences of deep uncertainty in complex problems (Kwakkel et al., 2010; Lempert et al., 2003). For that purpose, I made use of newly developed System Dynamics (SD) (Forrester, 1969; Sterman, 2000) simulation models. These models were used in several different approaches rooted in Exploratory Modelling and Analysis (EMA) (Bankes, 1993; Bankes et al., 2013), for example, Scenario Discovery (Bryant & Lempert, 2010) and Robust Decision Making (Lempert et al., 2006). The policy insights conceived in this fashion were communicated to policy-makers and scientists.

Dealing with deep uncertainty has a profound impact on the development and use of simulation models for policy analysis. Therefore, I will look in this chapter at the impact of dealing with deep uncertainty on methodology and policy discussions. Methodologically, taking deep uncertainty into consideration ripples through the entire model development cycle (section 7.1). Moreover, I will reflect on the limitations of some of the existing exploratory modelling techniques when used with non-linear models (section 7.2). Dealing with deep uncertainty also has repercussions for policy discussions with stakeholders. First, the various exploratory modelling techniques come with a variety of additional costs for analysts and stakeholders. Costs include longer lead times for model development and analysis, additional computational costs, and the risk of information overload (section 7.3). Second, I will reflect on my experience in communicating with stakeholders and domain experts regarding the policy relevant insights resulting from applying exploratory modelling (section 7.4).

7.1 Impact of EMA on the model development cycle

An essential part in EMA research is to choose or develop one or more models to be used in the analysis. Most EMA literature until now uses existing models. There may be reasons, however, to develop new models. The most important reason in this context is that many existing models were developed for consolidative model use, which may limit the extent to which the various deep uncertainty relevant to the problem can be explored, in particular if other than just parametric uncertainty exists.

Model development is often perceived to take place in model development cycles. These cycles emphasise the cyclical and iterative nature of model development, and all distinguish multiple phases. As all my work used SD models, I will now specifically focus on SD model development, which one could argue to be partly representative of other simulation-model development cycles. However, I believe that many of my observations and conclusions can be relevant for model development in other modelling paradigms as well.

Model development cycles might be better called model-based problem solving cycles. However, the term ‘model development cycle’ is most often used in literature. Hence, I will consistently use this term in this chapter.

7.1.1 The SD Model Development Cycle Revisited

Several authors (Forrester, 1994; Keating, 1998; Randers, 1980b; Richardson & Pugh, 1981; Sterman, 2000) have defined a model development cycle in SD (Table 7.1). While the different authors have different names for the different phases, all SD model development cycles boil down to more or less the same iterative cycles. In these, limited emphasis is placed on experimentation, simulation under uncertainty, and exploration of simulation results. In practice, there is little use of additional (i.e., non-SD) techniques and tooling to support experimentation and semi-automated exploration. This may be due to the fact that computing was expensive at the time the core SD method was developed and there were no techniques and tools to make sense of many simulation runs. Since this has changed, there may be good reasons to revisit the SD modelling cycle specifically for exploratory modelling.

First, we consider the different versions of the cycle as proposed by different authors in order to compose a version of the cycle suitable as a starting point for discussing the implications of Exploratory modelling.

Table 7.1. SD Model development cycles with their phases by different authors.

Randers	**Richardson & Pugh**	**Forrester**	**Keating**	**Sterman**	**Auping**
Conceptualisation	Problem identification and definition	Describe	Analysis	Problem articulation	Problem articulation
Model formulation	System conceptualisation	Convert	Design	Formulation of a dynamic hypothesis	Conceptualisation
Model testing	Model formulation	Simulate	Formulation	Formulation of a simulation model	Formulation
Implementation/ representation	Analysis of model behaviour	Design	Testing	Testing	Evaluation
	Model evaluation	Educate	Intervention & implementation	Policy design and evaluation	Policy testing
	Policy analysis Model use or implementation	Implement			

While Sterman’s book (Sterman, 2000) represents the common practice in SD, his naming of the phases is very specific for SD. For example, the ‘formulation of the dynamic hypothesis’ in particular is a concept that almost uniquely applies to the SD paradigm, and other SD authors do not use it as a name for a model development phase. Using ‘conceptualisation’ instead for the second phase is applicable for simulation modelling in general and also used by Randers (1980b) and Richardson and Pugh (1981). Further, the way in which Sterman conceptualises ‘testing’ – his fourth phase – may be considered too limited in the context of exploratory modelling. Testing, according to Sterman, refers to the use of different verification and validation tests to check whether a model is able to produce the ‘reference mode’. The reference mode is the relevant problematic behaviour over time selected in Sterman’s problem articulation phase. Since, in Exploratory Modelling, the reference mode itself is uncertain, ‘testing’ whether or not a single reference mode can be reproduced is impossible. Instead, evaluating the set of potential behaviours a model can produce is the relevant concern in EMA. Using ‘evaluation’ instead, similar to Richardson and Pugh (1981), may overcome this issue. To keep the wording in the cycle as consistent as possible, the formulation of a simulation model may also be called ‘formulation’ following on the second, ‘conceptualisation’ phase. Finally, if the fourth phase is already called ‘evaluation’, the last phase may be called ‘policy testing’.

I will, therefore, use the following five phases in the model development cycle in this chapter: problem articulation, conceptualisation, formulation, evaluation, and policy testing. Problem articulation includes the problem, boundary, and time horizon selection, potentially followed by an *ex ante* investigation of possible interesting modes of behaviour as reference modes. Conceptualisation includes the identification and mapping of the main relations in a system, potentially followed by the formulation of a dynamic hypothesis for the link between modes of behaviour and system structure. Formulation includes the specification of model structure and the quantification of parameters. Evaluation includes performing different tests on structure and behaviour of models to ensure the fitness for purpose. Finally, policy testing includes using the model(s) to test which policies may be promising or robust. With this cycle, I do limit myself to model related development and use, while policy implementation as used in Forrester’s and Keating’s cycles is not considered by me.

7.1.2 Problem articulation

The problem articulation phase is generally considered to be the first phase of the model development cycle. The overall goal of this phase is to come to a research design. This design can be obtained by taking the following steps: problem formulation, boundary or scope selection, time horizon selection, and in particular in SD research, the identification of reference modes. Deep uncertainty influences important choices made in especially steps one, two, and four.

Problem formulation

The problem selected at the start of the model development cycle defines the model’s purpose and consequentially has a profound impact on model development. If the problem can be clearly defined, it will leave little uncertainty about the boundary, time horizon, and reference mode. However, problem definition uncertainty occurs in wicked problems (Rittel & Webber, 1973), or in societal messes (Ackoff, 1974, 1979). Wicked problems exist when different perspectives about problem formulation exist which originate from different world views by stakeholders in a system. In other words, in these situations the problem is ambiguous. The conventional, consolidative approach in modelling, which most SD modelling could be characterised by, is to try to unify existing knowledge in these cases (e.g., Forrester, 1987), and, by doing so, reduce the ambiguity of the problem formulation. This process can be facilitated by bringing groups of stakeholders together during the model development cycle, as is done in Group Model Building (Vennix, 1999).

However, if consolidative approaches are either impossible or very hard, exploratory model development provides an alternative approach that acknowledges (deep) uncertainty. This could be the case if the ambiguity about the problem formulation cannot be reduced without favouring one of the world views, perspectives, or interests of the different stakeholders (Kwakkel, Walker, & Haasnoot, 2016). The ambiguity about what exactly the problem is in those situations is a form of deep uncertainty.

Boundary selection

Uncertainty in the problem formulation has direct effects for system boundary selection. In this next step of the problem articulation phase in model development, it is generally recommended that a modeller should not model 'the system', but 'the problem' (e.g., Sterman, 2000, p. 89). This idea is to stay away from trying to model the whole system and all its attributes, but to limit oneself to those parts and aspects that are relevant to the problem at hand. However, if the problem is ambiguous, and the modeller does not want to choose between the different stakeholders' world views, the modeller needs to adjust the scope such that it accommodates these different views.

Ambiguity about problem formulations, multiple problem perceptions, or even completely different world views necessarily affect boundary selection choices. There seem to be two broad ways for dealing with such ambiguity: choosing a broader scope, or developing multiple models. In the first solution, the world views share a mental model about how the system functions while they disagree on the exact problem in the system. The modeller could then develop a single model which has a larger scope than usual to accommodate the different world views. For example, in our work on societal ageing (Chapter 5), which was based on earlier work by Logtens, Pruyt, and Gijsbers (2012), the boundary was set considerably wider than in other SD work on the same topic (cf., Eberlein & Thompson, 2013; Sutrisno & Handel, 2012; J. P. Thompson et al., 2012). Similarly, our exploratory work on the 2014 Ebola epidemic (Chapter 6) had a considerably wider scope than other work on this topic (cf., Chowell et al., 2004; K. M. Thompson & Duintjer Tebbens, 2009; WHO Ebola Response Team, 2014). In both cases, the reason for doing so was to incorporate the effects of uncertainties linked to, but not part of, the core structure of the model. In the case of societal ageing, that core structure comprised an ageing chain, while in the case of Ebola, it comprised a complicated SEIR transmission structure (i.e., an ageing chain of Susceptible, Exposed, Infected, and Recovered population).

However, if it is impossible to accommodate the different perceptions and world views in one model, there may be a need to develop two or more separate models. It is relevant to note here, that I consider a model to be a simplified representation of a system. A computer simulation model specifically refers in this dissertation to an internally connected set of equations, which is not necessarily parametrised. Practically, my use of the word model thus refers to a file made in a specific modelling language that can be simulated.

Creating two or more separate models is the second solution for irreducible problem uncertainty. The best example for such situations in this dissertation would be the copper case; although the two models overlapped for a large deal in scope, the different perspectives resulted in considering different elements of the system for simulating changing copper demand (Chapter 2). The approach chosen earlier by Hoekstra (1998) and others could also be viewed in this light. In his and in similar work, Hoekstra and others used four distinct perspectives inspired by 'cultural theory' to create four different model formulations.

Further, it is conceivable that different world views regarding one problem do not fit in one modelling paradigm. In these situations, it may be useful to develop models about the same issue using different paradigms. This approach was chosen during the shale gas project by trying to model the global primary energy system both in Agent Based Modelling (ABM) (Epstein & Axtell, 1996) and in SD (Moorlag, Pruyt, Auping, & Kwakkel, 2015). While due to the time constraints of this project this approach was not entirely successful for the client, it was very promising.

There are also more practical considerations for choosing an extended or multi-model scope, not necessarily related to wicked problems: for example, if a problem plays on different aggregation levels, like the problems central in the research on the impact of the shale revolution on state stability of traditional oil and gas exporting countries, then multiple models with different scopes may be necessary. In those and similar cases, networks of models can be used, where inputs from one model can be used in other models. This can be seen as a kind of hierarchical or multi-resolution modelling (Davis, 2000; Davis & Bigelow, 1998). Complete multi-resolution modelling is also possible, as is hybrid modelling, where multiple modelling paradigms are used at the same time. Examples of these approaches do exist already for hybrid modelling (Bobashev, Goedecke, Yu, & Epstein, 2007; Rahmandad & Sterman, 2008; Schwarz, 2016; Schwarz & Pruyt, 2016), but it is still far from being common practice.

Exploratory models can have a narrower scope than consolidative models as well, if it is considered useful to test a system's behavioural response to different exogenous input scenarios (e.g., to stress test a water system over a set of climate scenarios). One reason for doing so may be the use of well-established input scenarios, like climate scenarios (e.g., for precipitation) or population scenarios. The difference between models that are purely endogenous or use exogenous input scenarios can be referred to as the difference between 'closed' or 'driven' models (Coyle, 1983). In closed models, behaviour over time originates solely from the structure of the model, while in driven models, input scenarios are used to 'drive' the model to test whether the model will come to different end states or is sufficiently resilient.

Exogenously driven models are at odds with the traditional SD focus on endogenous explanations of problematic behaviour (Forrester & Senge, 1980). Sterman (2000, p. 95) even states that exogenous explanations "simply beg the question". Using exogenous drivers to explore a model's dynamic response under different circumstances can, however, be very relevant. Exploring behaviour over a range of exogenous drivers allows evaluating how a system might react given different plausible future contexts. As SD models represent highly complex systems, a model may react differently on an input if, for example, in two different parameterisations the system before the input was in a similar state. Different states of the system generate even more often different responses to inputs. Therefore, investigating the endogenous response to a range of exogenous inputs is not always "begging the question".

We used exogenous inputs in the research on societal ageing in the Netherlands, where it became clear that the combination of an increased male life expectancy combined with a stable or decreasing labour productivity would lead to unsustainable government finances. In the shale research, it was found that volatile energy prices could lead to situations of increased instability if the period with relatively low prices was long enough to create tipping-point behaviour where a country would have difficulty to recover into a more stable situation due to the strong reinforcing feedback of country stability and economic development (Chapter 4).

Time horizon selection

Part of the difference between consolidative and exploratory modelling in problem and scope selection can be explained by the difference between trying to explain and solve *existing* problems, and trying to identify potential *future* problems. Most SD studies are focussed on explaining existing problems (Lyneis, 2000), although many exceptions to that apparent rule exist (e.g., Meadows et al., 1972). This focus on existing problems seems to be related to a certain reluctance in the field of SD to use SD models for foresight.

To the best of my knowledge, there is very little exploratory modelling research with a pure historic or even archaeological focus. It is conceivable, however, that exploratory modelling can also be used for historic problems, as some historic issues are just as ill-defined as future problems due to a practically irreducible lack of knowledge. Still, all EMA studies until now are forward looking and are aimed at trying to identify potential future problems. As uncertainty can be expected to increase when researching to future, the forward looking nature of this research makes an exploratory approach even more fitting.

Reference mode

The last step, the selection of the reference mode, changes considerably if – instead of a consolidative approach – an exploratory approach is chosen. The reference mode is defined by Sterman (2000, p. 90) as “a set of graphs and other descriptive data showing the development of the problem over time”. The focus of the selected mode fits with the focus on existing problems. In some cases, however, even for consolidative use, hypothesised future behaviour can be used (Randers, 1980b). The goal of the reference mode is to be able to check whether the model is able to simulate the ‘correct’ behaviour.

As exploratory modelling research typically focusses on the future, it may be nonsensical to find a reference mode based on historic behaviour. Further, when exploring what may happen, instead of what happened, focussing on only one mode is too limited. Instead, it is possible to compare expectations and forecasts by other researchers and analysts to the behaviour generated by the to-be-built exploratory simulation model, but also by imagining what might possibly happen. In this situation, other forecasts and imagined futures may be used as reference modes. However, there is one major difference: in exploratory modelling the focus is on generating *at least* all plausible reference modes, instead of the most likely mode or modes. In cases where it is impossible to generate one or more of the possible reference behaviours identified before, modellers should be able to explain, using the model, why this is impossible and why the simulated behaviour is still plausible. This in itself is also considered to be a relevant use of models in exploratory modelling.

Communication about the impact of deep uncertainty on problem articulation

It is always important to communicate the problem articulation used as a starting point for model development with clients and stakeholders, but in situations with innovative, multi-model approaches it is especially important to motivate why and how the use of multiple models contributes to analysing the research problem, and how exactly these models are used together. Therefore, in the copper research, it was important to explain which parts of the models overlapped and which were different due to the different perspectives that were to be modelled (Chapter 2), while the shale research (Chapter 4) included a research design picture.

Consequences of uncertainty on problem articulation

In the problem articulation phase, the possibility of choosing multiple scopes is the most salient difference between exploratory and consolidative modelling. The operationalisation of the model development cycle thus changes already in the first phase (Fig. 7.1). The main reason for choosing multiple scopes is that in wicked problem situations, multiple views on how the system functions may exist. In those situations where it is impossible to unify these different views, using multiple model scopes may be a solution. However, if uncertainty at this phase is limited, there is no real difference with the operationalisation of a consolidative modelling project.

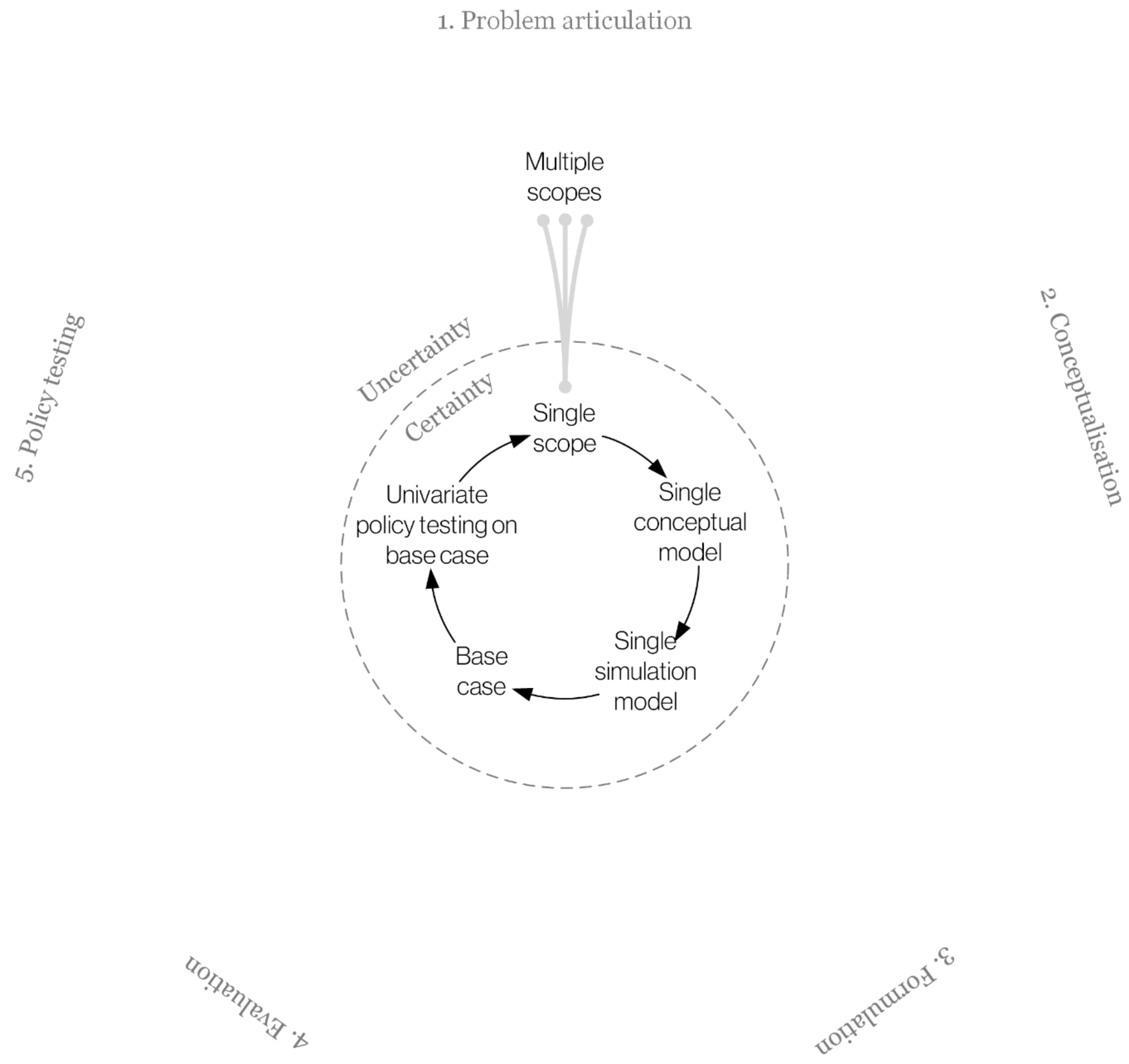


Fig. 7.1. Model development cycle operationalisation expanded for exploratory problem articulation.

7.1.3 Conceptualisation

In the conceptualisation phase, an outline is sketched of the model or models which is (or are) going to be constructed. This makes this phase more modelling paradigm specific than the problem articulation phase. In the SD modelling paradigm, conceptualisation entails that a hypothesis is generated and an attempt is made to make all important relations endogenous to the model. Diagrams are frequently used to facilitate these steps, while updated versions of these same diagrams can be used to communicate basic model structure after the research has been performed. In all these steps, deep uncertainty can play a role and choices made in the problem articulation phase will have knock-on effects.

Consequences of problem articulation choices

Together with the choice of a modelling paradigm, the choice of a multi-model approach has profound effects on the conceptualisation phase. Such choices will often lead to largely separate conceptualisations for each of the models. This is not the case if parallel multi-model use is considered. Parallel multi-models share part of their scope, making parallel conceptualisation of the models, focussing on both similarities and differences of these models, necessary. For example, in the copper case, the interaction between supply and demand is in principle similar for each model, creating a common core in terms of scope. The different perspectives caused the models to be different when it comes to the calculation of the intrinsic demand (bottom-up vs. top-down) and the regional trade (regional vs. top-down).

Identification of main relations

In consolidative modelling, model conceptualisation focusses on identifying the main relations between key variables within the chosen scope. In the case of SD, this comprises the main causal relations and the feedbacks they together cause. In SD thinking, the feedback effects, accumulations, and delays in them cause the system behaviour at hand.

A crucial element in the SD conceptualisation phase is to translate existing mental models (Craik, 1943; Doyle & Ford, 1998, 1999; Johnson-Laird, 1983; Lane, 1999) into causal structures underlying the model under development. As one might expect in the context of this dissertation, I argue that deep uncertainty may occur when different mental models exist about a system. Consolidative SD, in particular group model building (Vennix, 1999), emphasises the role of models in trying to unify different views into a single shared model. However, there may be good reasons not to aim at unifying views. There are numerous examples of paradigmatic differences, which correspond to different mental models, that have not been unified thus far, and may never be unified. For example, in the resource scarcity literature a strict division exists between the 'fixed stock' and 'opportunity costs' paradigms (Tilton, 1996). While some have tried to unify these paradigms (e.g., Chapman & Roberts, 1983), the debate continues (cf., Gordon et al., 2006, 2007; Tilton & Lagos, 2007). Different perspectives on a system may also generate different explanations about the functioning of a system (Cole in Meadows et al., 1982, p. 205). In exploratory modelling, one can accept the disagreement of the mental models and try to unify the different stakeholders at a later phase of the research. This can be done, for example, by finding a set of policies that is robust regardless of the differences in existing mental models.

Mapping of main relations

Most modellers use diagrammatic tools concurrent with the identification of the main relations. These tools provide a visual overview of the model under development, and can be used to communicate the aggregate structure or functioning of the model as well as the chosen scope. In SD, three different types of diagrams are commonly used for this purpose. These are the Causal Loop Diagram (CLD), the Stock Flow Diagram (SFD), and the Sub-System Diagram (SSD) (Morecroft, 1982; Sterman, 2000).

A CLD depicts the aggregate causal structure of a model with the polarity of the causal links between key variables, and the most important feedback loops within the model. An SFD is used to depict the basic stock-flow structure modelled, if this generates better insight in the functioning of the model than a CLD. This is the case for models with a prominent resource supply sub-model (e.g., Chapter 2). An SSD (Morecroft, 1982; Sterman, 2000) shows the main sub-models of which a model is composed and the relations between these sub-models. Its advantage, especially compared to CLDs, is that it gives a better overview of the components of the model. This characteristic makes it better suited to explain the structure of relatively large models. Therefore, if the wickedness of problems has been approached by expanding the scope, SSDs are often better choices than CLDs to explain model structures (e.g., Chapter 4). An additional advantage of SSDs is that they are less technical, making them potentially more acceptable for non-SD audiences.

Uncertainties can be communicated in this phase by supplementing the existing diagrammatic conventions. I suggest to depict uncertain relations in CLDs with different font styles (e.g., italics) and dotted arrows. Depending on the public, different styles of dotted arrows may be used to make more refined distinctions amongst different origins of uncertainties. This will also work in SFDs. Fig. 7.2 shows a CLD example. Please note, that the loop signs are omitted in this diagram for clarity. In sub-system diagrams (SSDs) (Morecroft, 1982), such conventions are less needed as the wording on the arrows connecting the sub-systems already allows for indicating uncertainty.

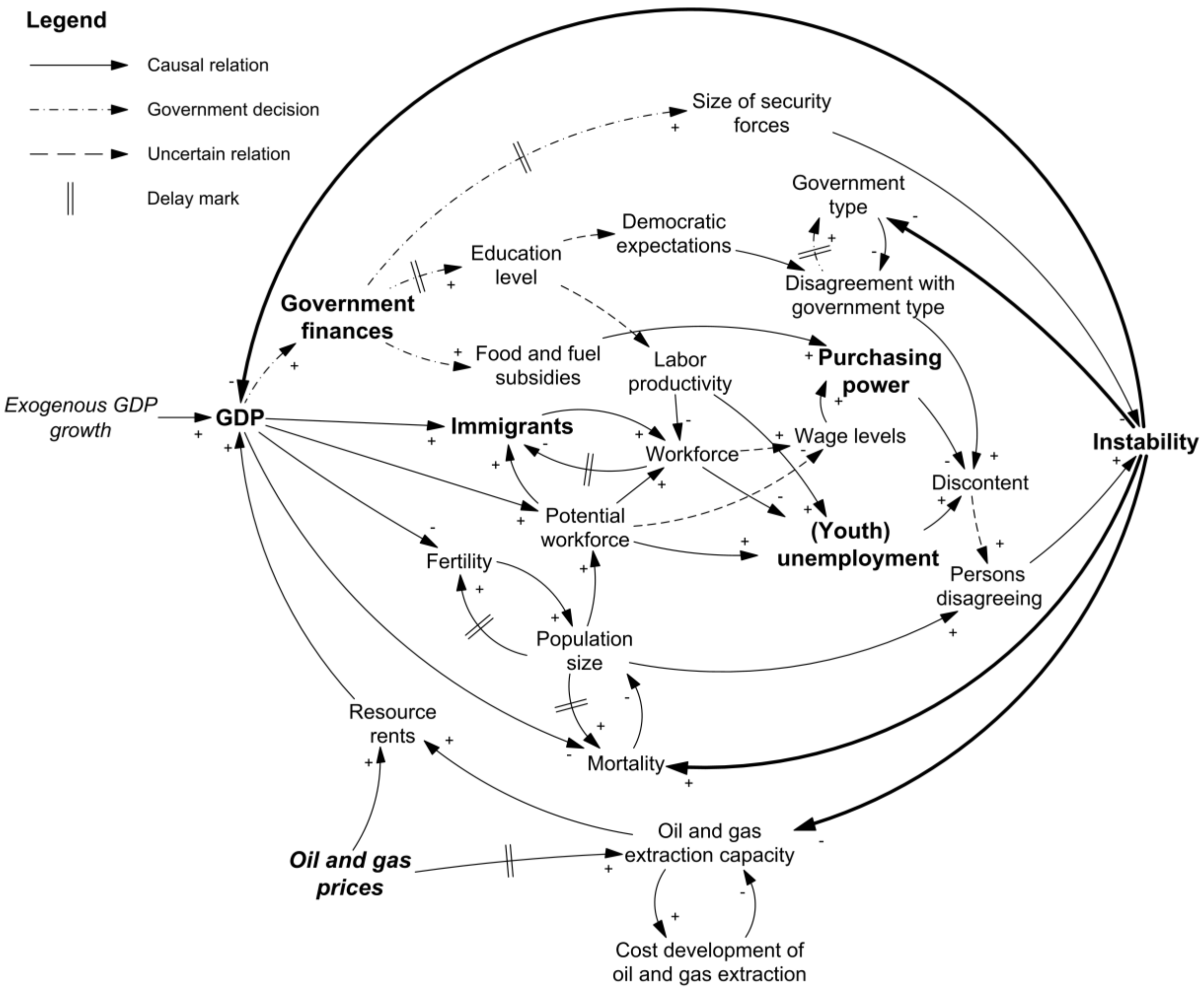


Fig. 7.2. Example of a CLD with explicit uncertainties. Exogenous parametric uncertainties (e.g., exogenous economic growth) and uncertain scenario inputs (e.g., oil and gas prices) are in *italics*, endogenous uncertain relations (e.g., the link between education level and labour productivity) are indicated by *dashed arrows*, and uncertain government decisions are indicated by *dash-dotted arrows*.

Dynamic hypothesis

It is considered good practice in scientific research to present a falsifiable hypothesis *ex-ante* about the outcomes of the research. Modelling research is in that sense no exception, which is why in the SD community a dynamic hypothesis is formulated in the conceptualisation phase. The dynamic hypothesis is a working theory of how the policy issue arises from the structure of the system (Lane, 1993; Randers, 1980a; Sterman, 2000). In consolidative SD practice, this means an explanation of how the conceptualised system structure generates the reference mode from the problem articulation phase.

If deep uncertainty exists about the modelled system, this may have consequences for formulating the dynamic hypothesis. This hypothesis may be influenced by the level of uncertainty (Kwakkel et al., 2010) of both the input data (i.e., all parameters, initial values, and input scenario data to be used for model setup), and the available information about the structure. That is, under deep uncertainty there may be multiple dynamic hypotheses for one reference mode, one hypothesis explaining multiple reference modes, or multiple hypotheses explaining multiple reference modes (Table 7.2).

Table 7.2. Impacts of uncertainty on dynamic hypothesis formulation.

			Uncertainty level of conceptual model: Shallow — *Structure*	Uncertainty level of conceptual model: Deep — *Structures*
Uncertainty level of input data	Shallow	*Behaviour*	One hypothesis explaining how the structure explains the reference mode	Multiple hypotheses about how plausible structures explain one reference mode
	Deep	*Behaviours*	Multiple hypotheses about how the structure explains multiple plausible reference modes	Multiple hypotheses about how plausible structures explain multiple reference modes

Having more than one hypothesis about the relation between structure and behaviour fits the exploratory modelling paradigm quite well, in particular in cases of open exploration of plausible system behaviour. I would recommend, therefore, to list a set of these hypotheses in EMA work. To some extent this is in line with the consistent use of the plural form 'dynamic hypotheses' by Keating (1998). When one or more hypotheses are falsified throughout the wide range of modelled behaviours, this may prove an especially valuable contribution to a policy discussion. In the shale research policy report (De Jong et al., 2014), we did pose a hypothesis about the expected behaviour, which was subsequently falsified. We believed that shale gas would provide gas for gas substitution, lowering natural gas prices in Europe. The simulated behaviour suggested otherwise regardless of the existing uncertainties, however, as it provided a strong argument about a higher volatility of oil prices as a result of the fact that it is easier to accumulate a production surplus of oil than a surplus of natural gas. Consequentially, the oil price, which depends on existing stocks of oil rather than just the current production, becomes more volatile and shows 'hog cycle' (Hanau, 1928) behaviour (Chapter 4). In this case, we could have provided a broader set of hypotheses given the known uncertainties and complexity of the global energy system.

Consequences of uncertainty on conceptualisation

Uncertainty may expand the operationalisation of the model development cycle in two different ways in the conceptualisation phase (Fig. 7.3). If in the problem articulation phase multiple scopes were defined, multiple conceptual models will have to be developed. If uncertainty did not play a role in the problem articulation phase, structural uncertainties in one conceptual model can be the first manifestations of a need for exploratory modelling. Such structural uncertainties can, of course, also be part of multiple conceptual models (not visible in figure). However, it is still possible to have no serious manifestations of uncertainty so far. In that situation, only a single conceptual model without structural uncertainties can be developed.

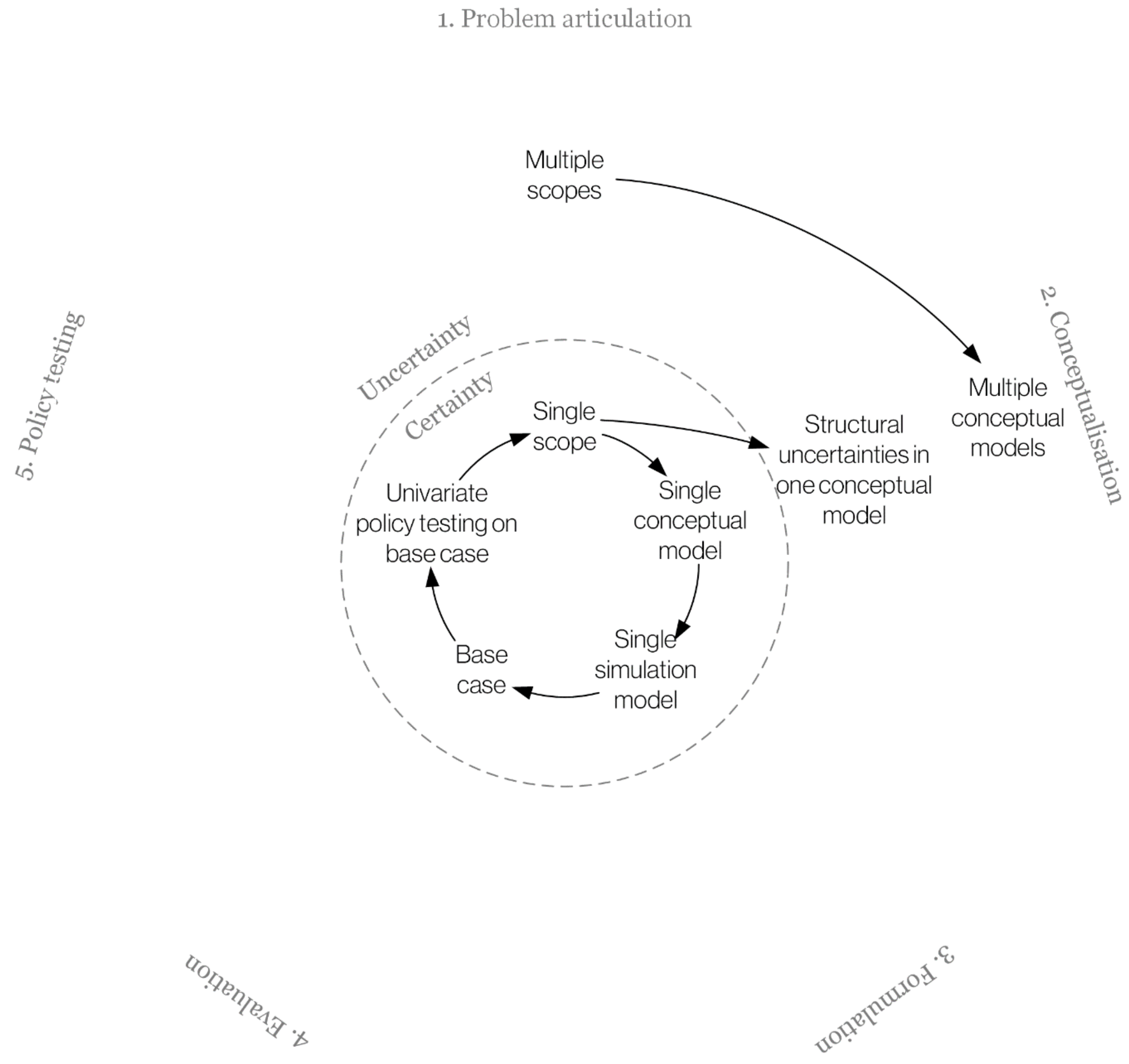


Fig. 7.3. Model development cycle operationalisation expanded for exploratory conceptualisation.

7.1.4 Formulation

In the formulation phase, the model is specified: equations or decision rules are specified, and parameter values are given. The formulation phase is even more modelling paradigm specific. To some extent, it is even modelling language dependent, as a modelling language may make specific types of formulation more suitable compared to alternative modelling languages within one paradigm. For example, Vensim (Ventana Systems, 2010) has a somewhat different implementation of table or lookup functions than Powersim (Powersim Software, 2017). Kwakkel et al. (2010) refer to this issue as model implementation uncertainty.

Consequences of choices in earlier phases

The choices made in the problem articulation and conceptualisation phases have consequences for choices during model formulation. All problem articulation choices that affect conceptualisation do so during formulation as well. This specifically counts for the modelling paradigm choice, and multi-model approaches. In the case of parallel multi-models, the shared common core of the set of models can best be developed first before other unique parts are developed. This way, it is possible to maintain the similarity between these shared model components. This approach was used in my copper research (Chapter 2 and 3). In addition to these effects, structural uncertainties identified in the conceptualisation phase also need to be modelled.

Structure formulation

In consolidative modelling, the modeller defines one equation per variable. Combinations of variables together form unique structures or decision rules in which every element of the system structure is thus modelled exactly once. However, if the conceptualisation reveals major structural uncertainties, this is impossible to maintain. In these situations, the exploratory modeller will have to define different options in parallel (i.e., alternative structures) within the model. In SD this means that for each variable affected by the structural uncertainty, multiple values are simultaneously calculated. A switch variable, constant during the entire run time, determines which of the calculated values is used to influence the rest of the model. In other types of equation-based modelling, similar approaches can be used for dealing with structural uncertainties.

Structural uncertainties may be included in model formulation for more pragmatic reasons (Peterson & Eberlein, 1994), other than theoretical reasons originating in the literature. There are multiple ways of representing some relations or larger structures in models even when, or especially when, there are no explicit discussions in the literature on how these structures should be modelled. For example, what order of delay to use? How many stocks to use to disaggregate an ageing chain? One approach would be to just test the structure at hand with each equation and choose the one that functions best. Another approach would be to try to aggregate the model such that the uncertainty disappears in the aggregation. It is conceivable, however, that there are some remaining situations in which there are only pragmatic reasons for choosing one representation. This can be the case if both produce different, yet plausible results. In these situations, I recommend modelling all plausible representations and use switch variables to explore the consequences of these uncertainties, instead of using a hit-or-miss tactic and choosing among the options.

Parameter formulation

Major differences between consolidative and exploratory modelling occur in parameter formulation. In consolidative modelling, the modeller will try to use a most appropriate value of any parameter specified. Sometimes the parameter values are chosen such that the simulation fits the reference mode best. In exploratory modelling, however, many parameters are considered to be deeply uncertain. In these cases, there is no best value for these parameters, only a parameter range. The modeller may choose to make this explicit during model formulation by referring to the plausible range and it's background in the equation comments.

At the end of the formulation phase, all uncertainties will need to be traceable in later analyses of simulation results. For example, in the evaluation or the policy analysis phases. Therefore, all non-parametric uncertainties, whether they result in multiple models or multiple structures in models, need to be parametrised to make analysis of these results possible, manually or by using computer algorithms. Non-parametric uncertainties in particular should be treated as categorical uncertainties. In this way, different model behaviour due to multiple models or structural uncertainties in models can be traced back to the input assumptions.

Consequences of uncertainty on formulation

Uncertainty may expand the operationalisation of the model development cycle also in terms of differences in model simulation files (Fig. 7.4). Multiple conceptual models will generally lead to an equal number of different simulation model files. If structural uncertainties make the development of a single simulation model file too impractical (e.g., Chapter 2), this may also motivate specifying multiple models. Otherwise, structural uncertainties may be made part of a single model. Sometimes, the problem may make a 'driven' model suitable for the question at hand. In such cases, exogenous scenarios will be part of the model file. Further, it may be possible to have multiple possible formulations for one or more relations in the model or models. However, also after the formulation phase, it may be possible that the uncertainties were not significant enough to make such choices necessary. The exploratory model development operationalisation is then similar to consolidative modelling.

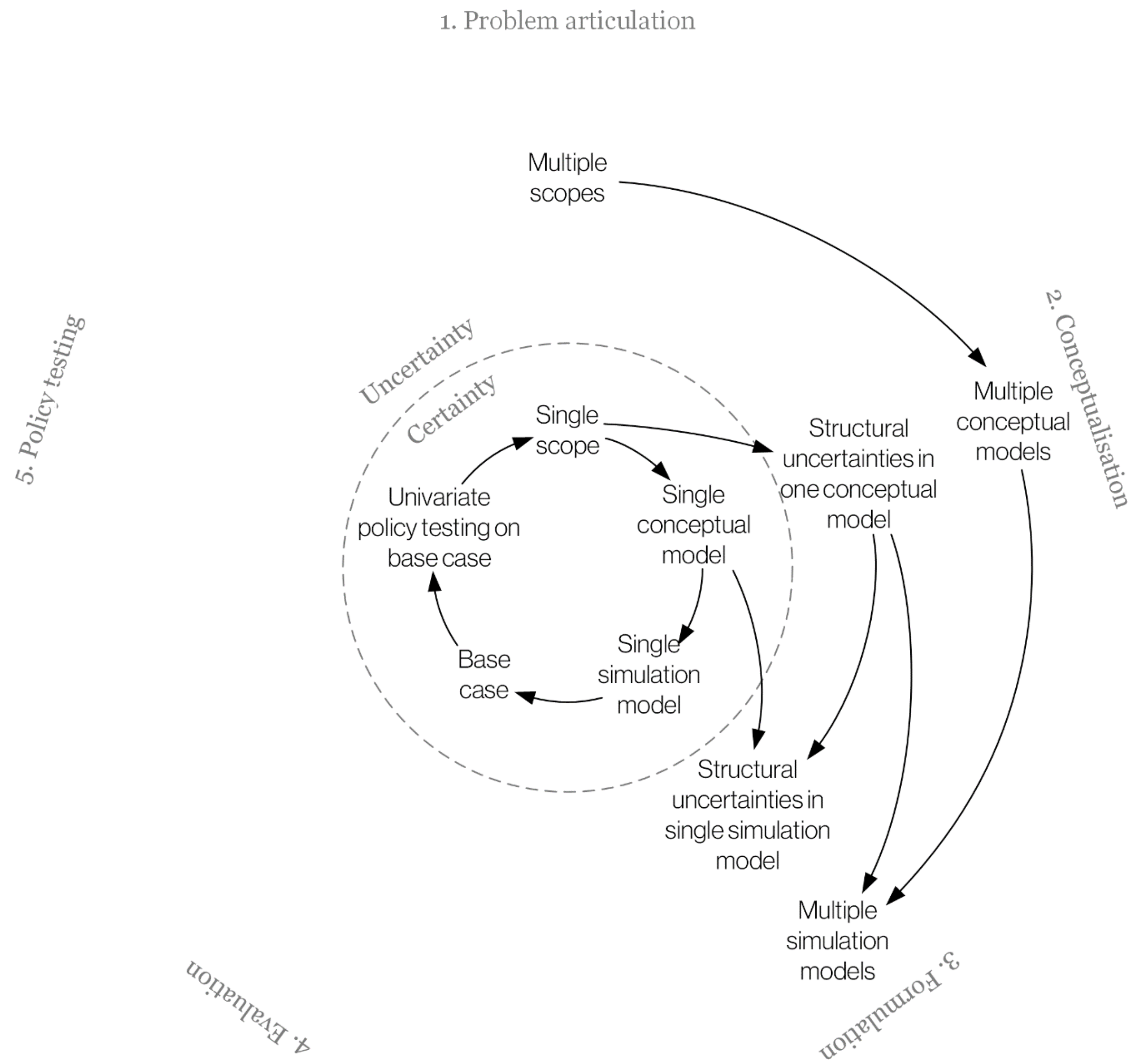


Fig. 7.4. Model development cycle operationalisation expanded for exploratory formulation.

7.1.5 Evaluation

In the evaluation phase, the model is subjected to various tests to assess its quality. These tests respectively focus on whether the model has been correctly constructed (i.e., verification), and whether it is fit for purpose (i.e., validation) (Barlas, 1996; Hodges & Dewar, 1992; Oreskes et al., 1994). The concepts of verification and validation in SD correspond to a large extent to how these terms are used in, for example, operations research and management science (Balci, 1994, 2013; Lane, 1995). The results of the tests can also be communicated to convince stakeholders, policy-makers, or readers of model quality given the purpose.

Once the model or models pass the verification and validation tests adequately, one or more base runs are selected for analysis and communication. This step entails the most striking difference between consolidative and exploratory modelling: generating a base ensemble of runs in contrast to a single base case.

Some of the early literature on exploratory modelling questions whether exploratory models can be validated (Hodges & Dewar, 1992). Part of the discussion lies in the interpretation of the word ‘validation’ itself, which implies that a model can be correct or true. The discussion is thus semantic in nature and does not reflect the operationalisation of the validation steps. There are examples of EMA research in which the validation process seems to reflect consolidative modelling practices (e.g., Fischbach, 2010). Hodges & Dewar’s objection to validation suggests that all such efforts are wasted time, but this can be considered unsatisfactory by clients and stakeholders interested in a quality assessment of the research’s methodology. Moreover, when one understands validation as the process of establishing the fitness of a model for a purpose, the claim that validation is strictly impossible no longer makes sense. Therefore, it is important to assess how traditional verification and validation approaches are altered by deep uncertainty.

As the outcome of the tests performed in the evaluation phase depends on the exploratory modelling purpose, it is important to revisit the question at hand as defined in the problem articulation phase, and how it is different compared to consolidative modelling purposes.

Structural validation

In SD, structural validation tests include structure-verification, parameter-verification, extreme-condition, boundary-adequacy (structure), and dimensional-consistency tests, while behavioural validation tests include behaviour-reproduction, behaviour-prediction, behaviour-anomaly, family-member, surprise-behaviour, extreme-policy, boundary-adequacy (behaviour), and behaviour-sensitivity tests (Forrester & Senge, 1980). Both types of tests are often performed simultaneously and iteratively throughout the modelling cycle. Some of these tests are also considered to be part of model verification. The EMA approach does entail some quite significant changes in the assessment of some of the structural validation test outcomes (e.g., structure-verification, boundary-adequacy, parameter-verification, and extreme-condition tests), while the test descriptions do not need to change. For other tests, for example dimension-consistency tests (i.e., unit checks), there is no difference with consolidative modelling at all. This also holds for the more technical verification tests.

The structure-verification test compares the modelled structure to the observed system. For this test, the outcomes of the conceptualisation phase are most important, especially the underlying mental models. In traditional, consolidative modelling practice, this means that the model tries to unify existing knowledge. In the exploratory approach, this means that all plausible, yet contrasting mental models should be taken into account. Further, literature sometimes agrees on the fact that some structure is unknown, as is the case with the grade-tonnage distribution of metals in the lithosphere (Gerst, 2008). Therefore, in an exploratory modelling approach, one needs to check whether all relevant structural uncertainties have been taken into account.

Similarly, boundary-adequacy tests look both at whether the model contains the structures it needs given the purpose and the aggregation level. These tests are extremely relevant for exploratory modelling, as deep uncertainty has a strong impact on the choices regarding the system's boundary. For exploratory models, these tests entail whether all, or the most important, existing perspectives on the system have been modelled. Further changes come from the intended use of the model. It is always important to be able to map the outputs back to the input space. The existing algorithms for Scenario Discovery, Patient Rule Induction Method (PRIM) (Friedman & Fisher, 1999) and Classification and Regression Trees (CART) (Breiman, Friedman, Stone, & Olshen, 1984), fail in case of the use of highly non-linear models. To work around this issue, the modeller may opt for an exogenously driven model, which may lead to less non-linearity in the model. For example, an exogenously driven model was used in the research on societal ageing (Chapter 5). Alternatively, uncertainty may also lead to expansion of the traditional boundary, as was the case in the research on the Ebola epidemic (Chapter 6; Pruyt et al., 2015). Therefore, an important test should be whether the boundaries chosen actually fit the intended purpose of model use.

The parameter-verification test changes as sometimes uncertainty may exist about the conceptualisation of a parameter. In this test, whether model parameters conceptually and numerically correspond to the system is checked by comparing them to existing knowledge about the system (Forrester & Senge, 1980). In consolidative modelling, each parameter is considered to have a best value. In exploratory modelling, parameters are considered to have at most a bandwidth of values, while determining a best value is impossible. These bandwidths may sometimes be based on different values found in available data or literature, but otherwise it may be possible to derive theoretical bounds of the parameter values. An example may be the value of a country's GDP. Many different ways of calculating a GDP exist, and even given the purpose of a model, it may be hard to determine whether the value given by, for example, the World Bank, the IMF, the UN, or the CIA World Factbook is best. In such a situation, it is possible to include different ways of calculating a variable in the model, determine the lowest and the highest value and consider all values in between as plausible.

In extreme-conditions tests, the modeller tests either whether the model keeps behaving within physical bounds and according to plausible mechanisms when values of one or more variables are outside normal bounds, or under which conditions the model 'breaks' (i.e., a run cannot be completed due to floating point errors) and stops functioning.

The first interpretation of extreme-conditions tests is advocated within SD as an important and valuable test (Forrester & Senge, 1980; Sterman, 2000), and is sometimes referred to as 'reality check' (Barlas, 1996; Peterson & Eberlein, 1994). For example, if a modelled economy is cut off from energy supplies, the economic output should be reduced to practically zero. This test becomes problematic, however, when the model tester misinterprets the meaning of a model variable if the definition of that variable is ambiguous. Ambiguity is a form of deep uncertainty (Kwakkel et al., 2010). Following on economic thought, one might argue that the demand for a good should be zero when prices for that good are extremely high. If the tester artificially sets the price to a very high value, and the demand is not reduced to zero, the tester considers the model to show implausible behaviour. However, in an alternative interpretation demand may also be defined as a need. For example for food, then the demand should not be reduced to zero. Most would agree that children in developing nations that cannot afford food, still have demand for it. Therefore, the use of reality checks without paying attention to the scope and purpose of the model is naive.

The second interpretation of extreme conditions tests tries to find conditions under which the model breaks. This can be done be testing the model for a wide variety of input parameters. The EMA approach provides tools to perform this test in a very systematic way. Especially if the uncertainty space is taken relatively wide (i.e., parameter bandwidths within theoretical, broad bounds and not practical, empirical bounds), the sampling for run generation functions as multivariate extreme-conditions tests. As a consequence, the first sets of runs performed with a new model often result in uncompleted runs.

Behavioural validation

Many of the behavioural tests traditionally used for validation are an integral part of the analysis of the ensemble of model runs generated. The uncertainty analysis in exploratory modelling can be seen as a more extensive form of sensitivity analysis (Herman, Reed, Zeff, & Characklis, 2015; Kleijnen, 1997). With many other behaviour-oriented validation tests, validation is more extensive in exploratory modelling than in consolidative modelling, and aimed at generating insight into all possible behavioural modes, not merely to support the validity of the model for a reference run. As such, the statement of Forrester and Senge (1980) that SD models are relatively inert when it comes to their behavioural modes, is not necessarily true in the context of exploratory modelling using SD.

For ‘behaviour-reproduction’ and ‘behaviour-prediction’ tests (Forrester & Senge, 1980), deep uncertainty means that the runs generated should at least show behaviour that is or can be considered to be characteristic of the system. In consolidative use, the ideas behind these tests do entail the danger of prejudice or stove-piping: the system can be modelled such that the preconceived ideas are made to come out. An example would be to model metal or energy reserves such that they will be finished within the run time, while some researchers believe that this is not plausible. In EMA, however, the models should be able to reproduce all plausible theories, or be explicit about focussing on a subset of the available theories. Trying to model these different world views or perspectives potentially has a large effect on the wealth of behavioural modes generated with the set of models.

Behavioural tests are executed in a more systematic manner in exploratory modelling compared to consolidative modelling. This is especially the case for the behaviour-anomaly and the surprise-behaviour tests (Forrester & Senge, 1980). In both tests one looks for unexpected behaviour, but in the first this regards behaviour considered in conflict with the real system, whereas in the second this regards behaviour that was not expected with current knowledge of the real system, but which proves plausible. As exploratory modelling includes generation of a large set of runs, it provides a systematic way of trying to find as many possible behaviours as possible. These generated model behaviours form input for the behavioural tests. However, model behaviour that can be considered impossible in any future, can still be rejected. The analyst should, in such cases, assess whether such anomalous behaviour warrants a change to the model structure or to the associated uncertainty space.

The EMA methodology also makes use of a single general model in very different parameterisations easier and more systematic. This relates to the family-member test (Forrester & Senge, 1980), which tests whether a model can be applied to a similar situation with partly or completely different parameter values. For example, whether a model about country A can be applied for other countries as well. After all, if a model can already be used within a broad band of uncertainties, it is only a small step to use this same model for a similar class of problem, but in a different context. The parameterisation of one model for multiple countries is an extreme application of the family-member test. Examples of such a practice can be found in in the shale case (Chapter 4). In newer research (Chivot, Auping, De Jong, Rõõs, & Rademaker, 2016; De Jong et al., 2017), we parameterised one general model for 167 different countries.

Sometimes, real-life developments render part of the outcome ensemble obsolete. While it may seem that this renders the model invalid or should lead to a serious reduction of the input space, it is possible that in complex models the same input space can generate runs that produce historic and present day behaviour, as well as plausible behaviour currently not seen in the system. However, different parts of the input space may also generate similar behaviour, which is exactly the opposite and called equifinality (Bertalanffy, 1968). Both equifinality and the opposite were seen in the Ebola case (Chapter 6), where different parts of the input space were able to generate outcomes with behaviour seen in the system. A solution – if real-life developments render part of the outcome ensemble obsolete – is to post-process the runs in order to select only runs satisfying observed conditions for further analysis, like testing policy changes for their robustness. The modeller should thus not try to force the model into the observed behaviour by reducing the uncertainty space, if this reduction leads to fewer runs in the observed outcome space.

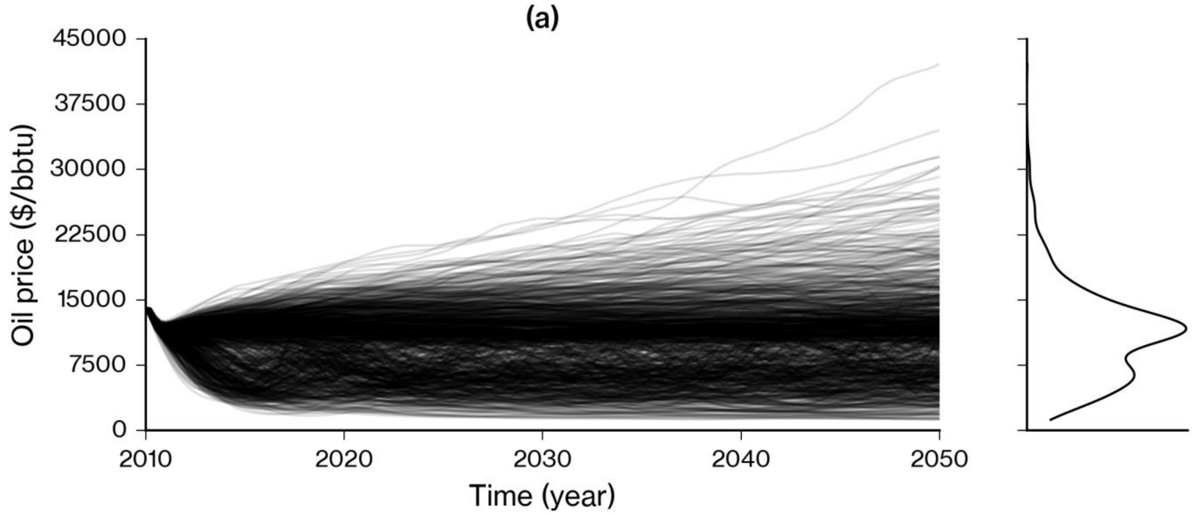


Fig. 7.5. Example of a base ensemble of runs generated by exploratory modelling. On the right, a kernel-density estimate of the distribution of end states is visible. From Chapter 4.

The differences between consolidative and exploratory modelling approaches are most visible during simulation. The differences originate from the fact that in exploratory modelling approaches, no base case is generated, but rather a 'base ensemble' of runs (e.g., Fig. 7.5). The ensemble is generated by sampling over all relevant structural and parametric uncertainties. In practice, this means that structural and model uncertainties are reduced to parametric, categorical uncertainties.

Ensembles are used in exploratory modelling to explore plausible system behaviour and find underlying causes for runs of interest. This process is referred to as Scenario Discovery (Bryant & Lempert, 2010). The underlying causes for the runs of interest generally correspond to specific smaller bounds of a selection of uncertainties within the entire uncertainty space. The identified sub-space of the entire uncertainty space can support the design of new policies. Sometimes, a selection of the whole ensemble is made to function as representatives or 'exemplars' (Islam & Pruyt, 2016; Pruyt & Islam, 2016) to make further analysis more insightful.

Second, ensembles are used to tests policies for their robustness. This is referred to as Robust Decision Making (RDM) (Lempert et al., 2006). Consolidative modelling can also use sensitivity analysis and a selection of runs – often also called 'scenarios' – to test policy robustness. However, it relies on both the manual generation of scenarios for this purpose, and manually testing policy robustness. The exploratory modelling toolkit provides a systematic, automated, and more rigorous approach.

Consequences of uncertainty for evaluation

The evaluation phase is arguably the clearest 'exploratory' modelling phase. In the vast majority of examples of exploratory modelling work thus far, this is the first phase in which uncertainty explicitly enters the analysis. In these cases, the ensemble of runs is generated by a single model, which has been developed in a more traditional, consolidative setting (Fig. 7.6).

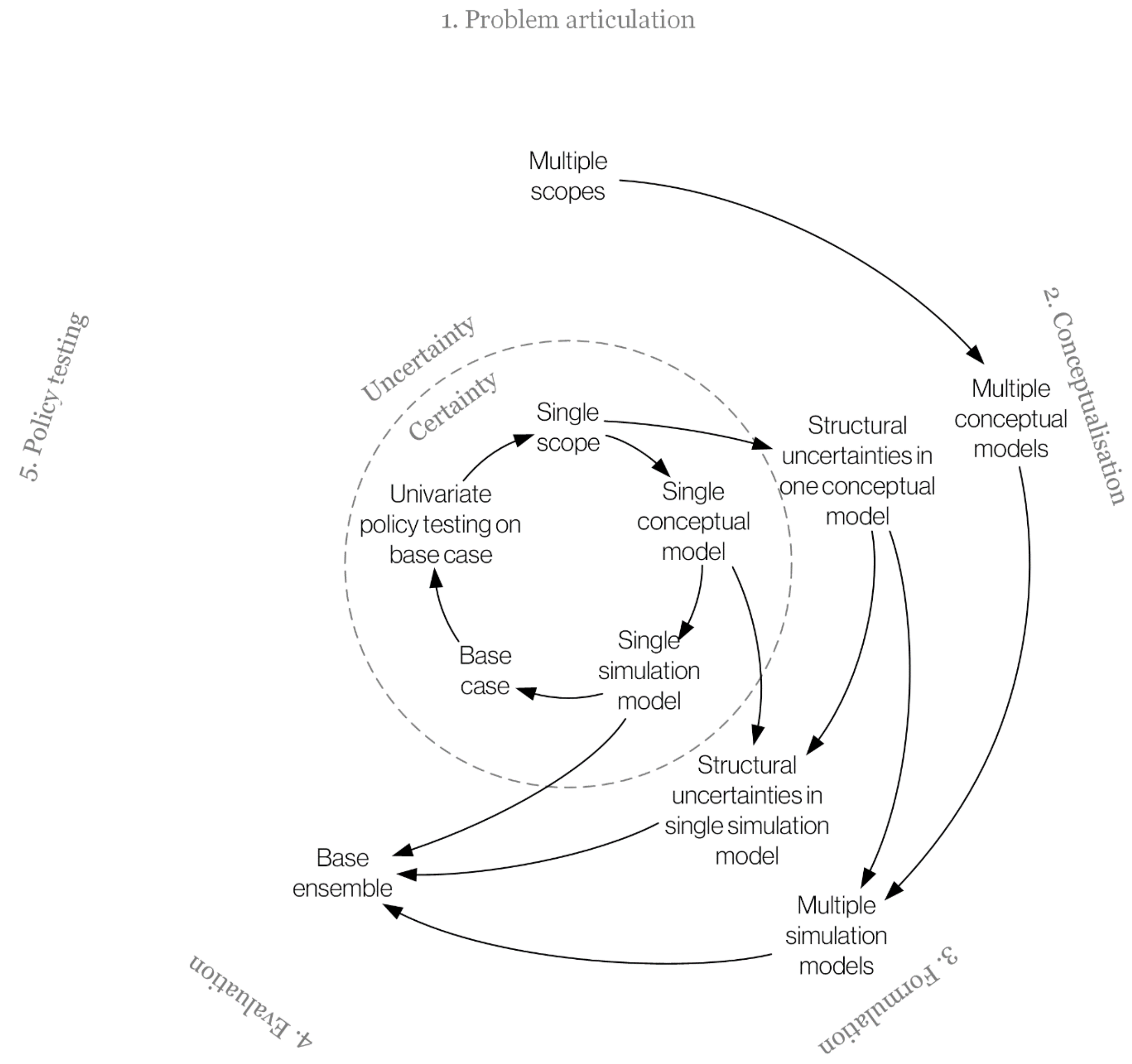


Fig. 7.6. Model development cycle operationalisation expanded for exploratory evaluation.

There may be several reasons for using an existing consolidative model in exploratory modelling. For example, there may be no budget to develop new models, but it is known that the exploration of parametric uncertainties is already meaningful. It is often considered desirable for acceptance of results to use established yet consolidative models. Finally, it may be established that all uncertainties are indeed parametric, making the development of explicitly exploratory models not necessary.

This dissertation demonstrates how structural uncertainties can be included in a single simulation model, as well as how multiple models can be used to generate the base ensemble. Further, I discussed that the exploratory modelling approach does change practices to assure model quality. In fact, many of these tests are an integral part of exploratory modelling practice. As a consequence, I argue that properly designed exploratory models are necessarily thoroughly tested for purpose.

7.1.6 Policy testing

The policy testing phase is the final phase of the modelling cycle and the last phase in which uncertainty is often ignored, while it can be argued that both the effects of policies and their timing are in fact deeply uncertain. In consolidative practices of SD as well as many other simulation modelling paradigms, with or without the EMA approach, it is common to model policy implementation as a discrete shock to the system (i.e., the policy is turned on or off), either at the beginning of the run time, or at one or more discrete moments during the run time. Examples of this practice include policy testing common in SD (e.g., Bleijenbergh, Vennix, Jacobs, & van Engen, 2016; Sterman et al., 2012; Wunderlich, Größler, Zimmermann, & Vennix, 2014), the impacts of interventions in transmission models (e.g., Camacho et al., 2014; Fasina et al., 2014; Fisman et al., 2014), and standard robustness tests in EMA work focussed on time dynamics (Hamarat et al., 2014; Molina-Perez, 2016). Further, policy choices desirable for one actor may be undesirable for another actor. This is also a specific kind of policy uncertainty.

Policy implementation itself is, however, deeply uncertain (Haasnoot, Kwakkel, Walker, & ter Maat, 2013). Policy implementation uncertainty may originate from either the unknown implementation, or an unexpected or unknown system response. This unknown system response may have all kinds of reasons, including political uncertainties, policy resistance, and overall uncertainty regarding the effect of a policy on an indicator. In any case, most policies are not implemented in an on-off fashion, but will start their impact in a gradual and delayed manner. The model implementation of a policy is, however, often modelled in this on-off fashion.

The policy implementation issue is exacerbated if the power of decision makers in the system is relatively limited. In the Ebola case, both policy implementation and system response were uncertain, while the GICMP case illustrates additional implementation issues when the power of decision makers is limited. In the Ebola case (Chapter 6), both the exact implementation (e.g., speed of deployment) of intervention capacities and the impact of the different types of interventions (e.g., the effective reduction of the reproduction rate in isolation facilities) was uncertain. This last effect in particular could have had a tremendous impact on the fact that the 2014 Ebola epidemic went out of control.

Regardless of the source of policy implementation and impact uncertainty, there are multiple ways of dealing with this type of uncertainty in EMA studies. For example, making the policy effects uncertain, or the speed of the policy implementation uncertain.

The first method is to include the uncertain elements of policies in the sampling phase of exploratory modelling. In this way, the policy analysis phase becomes technically indistinguishable from the formulation and evaluation phases.

The second method is to model policy responses in interaction with the system, which is consistent with the view on policies as part of feedback loops in the SD field. Policy implementation uncertainty exists mostly about the timing of policy implementation, where policies delayed may become policies denied. It is closely related to both the 'Integrated Risk-Capabilities Assessment' (Pruyt, 2012) and the 'dynamic adaptive policy pathways' (Haasnoot et al., 2013) approaches. Policy implementation uncertainty was also illustrated in recent work by Zeff, Herman, Reed, and Characklis (2016). All these authors treat the exact effects of an intervention as uncertain. In my work, policy uncertainty was included in the Ebola case (Chapter 6).

Next to policy implementation uncertainty, uncertainty or different opinions about the policy desirability exist. This issue was demonstrated in Chapter 5 with the societal ageing case. In this problem, the solution for untenable expenditure on retirement benefits may be considered to couple the retirement age to the life expectancy. However, this policy is considered undesirable by part of the population, as it increases the uncertainty about their financial situation between the previously perceived retirement age and the final effective retirement age, and the fact that some people look forward to being retired. In situations like these, trying to find robust policies is more complex, as seemingly contradictory interests need to be united. The solution can sometimes be found in additional policies that decrease the negative sentiment regarding the initial policy. In this case, I suggested that first trying to raise the informal "leaving age" (i.e., the age at which people by average stop working), hence making raising of the retirement age more something which only reflects the already existing reality.

Consequences of uncertainty on policy testing

The consolidative way of policy testing – one by one, and as shock to the system rather than modelling the gradual and uncertain response – often relies on univariate policy testing. The best combinations of policies are thus found by manual trial-and-error. By making use of the possibilities offered by exploratory modelling tools, this process can be made more systematic and automated.

Exploratory policy analysis can be made multivariate, where in fact a policy is treated more or less as if it is an ordinary structural uncertainty. The what-if analysis for which the simulation model is used can thus also be extended, as the circumstances under which policies and their combinations are successful, including policy implementation uncertainties, can be made explicit.

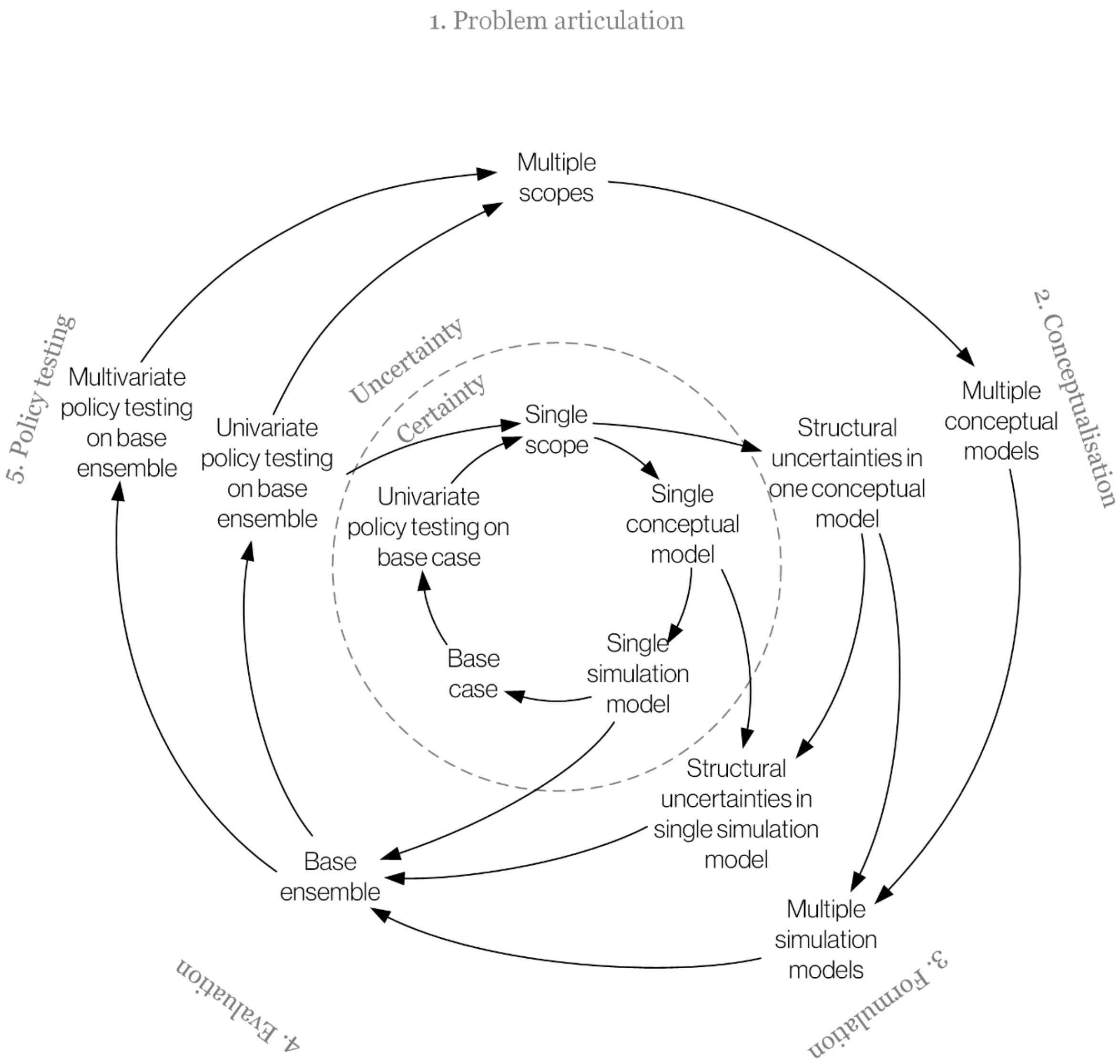


Fig. 7.7. Model development cycle operationalisation expanded for exploratory policy testing.

7.1.7 Conclusions on the complete exploratory model development cycle

Most exploratory modelling research to date acknowledges uncertainty only in evaluation and policy testing. That is, virtually always a single consolidative model is used in an exploratory way. However, the extended, exploratory modelling cycle (Fig. 7.7) makes clear that considering uncertainty is possible – or very important – in earlier phases of model development as well. If choices are made to acknowledge deep uncertainty in one of these earlier phases, this will render reducing the uncertainty by returning to the inner, consolidative modelling cycle in next phase impossible. However, if after performing one whole cycle of model development demonstrates that deep uncertainty had no consequence for the analysis of a specific problem situation, one could choose to ignore it in the communication about the model analysis by focussing on a single run (i.e., a base case).

7.2 Impact of complexity of models on Scenario Discovery

Development of exploratory models is often aimed at model use in EMA approaches like Scenario Discovery and Robust Decision Making. The use of non-linear models in this dissertation for exploratory use was promising, but using them for Scenario Discovery, a common practice in exploratory modelling, was often problematic.

Scenario Discovery aims at trying to link interpretable combinations of input values to those simulation outcomes that represent particularly undesirable futures by making use of computer-learning algorithms (Bryant & Lempert, 2010). Evidently, the same analysis can be used to find regions in the model input space that produce highly desirable futures. In my experience, this approach becomes problematic when used with highly non-linear SD models. There are three origins for these problems: the classification of outcomes, the non-linearity of mapping inputs to outputs in these models, and high dimensionality of the uncertainty space.

7.2.1 Classification of behavioural modes

Simulation models, specifically SD and Agent-Based models (ABM) (Epstein & Axtell, 1996), generate dynamic behaviour patterns over their run time. Classification of such outcomes by their behavioural mode is often very difficult. While some modes of behaviour are easy to classify (e.g., linear or exponential growth or decay, and oscillations with constant amplitude and period), other types of behaviour exhibited by SD models include super-positions of these modes, and even more complicated modes of behaviour.

The most frequently used method of classification in an EMA context is by generating a binary classification of model outcomes using the end (i.e., scalar) values of the runs. This type of classification was used in our ageing research (e.g., Chapter 5). However, the way an end value is reached is often at least as important. Binary end-value classification does not acknowledge this.

In the research on copper price scenarios, we tried to classify the models' behavioural modes (Chapter 2) using behaviour pattern features (Yücel & Barlas, 2011). This approach was unsuitable, however, to distinguish between different oscillatory dynamics, as oscillatory dynamics may have very different causes. First classifying them with behavioural pattern features and then using PRIM on the selection did not generate results that were easily interpretable.

In more recent work (Chapter 4), we classified behaviour by combining different behaviour characteristics (Pruyt & Islam, 2016), like the length of the line or 'roughness' combined with either the maximum value (Islam & Pruyt, 2016), the average value (Pruyt & Islam, 2016), or the roughness combined with the number of times the mean of the line is crossed (Chapter 4). Besides using this approach for selecting runs of interest, it can also be used to increase the sample size of regions in the output space that are apparently under-sampled: a process of 'adaptive sampling' (Bucher, 1988).

The idea of combining adaptive sampling with time-series feature selection is interesting and promising. Yet in the implementation used in this thesis (Chapter 4), it was not sufficient. The samples were not sufficiently large to allow proper mapping of selected outputs back to the input space, while the behaviours were too often quite different, yet classified as similar, or the opposite: very similar, but classified as different.

7.2.2 Linking inputs to non-linear behaviour

Scenario Discovery (Bryant & Lempert, 2010) and RDM (Lempert et al., 2006) make use of computer learning algorithms to link behaviour to parts of the uncertainty input space. PRIM and CART are examples of these algorithms. Both make orthogonal cuts in the input space by limiting the bandwidth of specific input parameters. This results in orthogonal boxes in the input space (Lempert et al., 2008). However, in case of non-linear interactions amongst the uncertain factors acting as input to models, it may become impossible to find orthogonal boxes of input parameters that link to the binary classification of outcomes.

Linking orthogonal boxes of inputs to the behaviour modes exhibited by SD models may be problematic for two reasons: non-linearity and equifinality. Due to the non-linearity, feedbacks, delays, and accumulations in the models, input parameter ranges leading to specific behavioural modes may be dispersed through the entirety of an input parameter's bandwidth. We tried to rotate parts of the input space with a principle component analysis before performing PRIM (Dalal et al., 2013) to overcome these issues (Chapter 3). Rotation does, however, not change the need for orthogonality of the input space if PRIM and CART are to be used.

The potential for equifinality (Bertalanffy, 1968) is also not reduced by the orthogonal cuts of PRIM and CART. The same part of the input space may still generate very different behaviours, while very different parts of the input space may generate the same behaviour in non-linear models. This further complicates performing Scenario Discovery with results generated by non-linear models.

One potential way of solving this issue is to focus more on relations among endogenously modelled variables. This requires more model analysis after runs have been performed, for example by focussing more on the state phase, or phase plots of different combinations of important system performance indicators. This idea has been suggested and demonstrated in the SD field before (e.g., Andersen & Sturis, 1988; Bruckner, Ebeling, & Scharnhorst, 1989; Rasmussen, Mosekilde, & Sterman, 1985; Reiner, Munz, & Weidlich, 1988; Thissen, 1978), but it has not found wide application. We applied this principle in the shale case (Chapter 4), although phase plots were not used as illustrations. We selected energy price dynamics on the basis of the performance of other endogenous performance indicators. This approach makes clear that specific combinations of outcome ranges are impossible. For example, in later work, it appeared to be impossible to have effective climate mitigation policies in combination with a high oil price (De Jong et al., 2017).

7.2.3 Interpretation of high numbers of uncertainties

In cases with a high number of uncertainties (i.e., in practice over 30 uncertainties), the number of dimensions found in PRIM boxes is generally also higher, which makes it harder to interpret the results. These large numbers of uncertainties are often caused by a relatively low model aggregation level that corresponds to aggregation levels found in relevant literature. This issue can be resolved by reducing the number of uncertainties, or by performing enough runs to generate all possible behavioural modes with a sufficiently large sample size for each mode of behaviour.

We have tried to reduce the number of uncertainties by making use of Random Forest-based feature scoring (Breiman, 2001). The result of this approach was that removing uncertain parameters that appeared to have no effect on model behaviour resulted in a reduction in the variety of behavioural modes generated by the model. The assumption based on the random forest scores was thus incorrect. Input factors that did not have an influence by themselves, proved to have a significant influence when combined. Further, besides simply performing more runs in the case of a large number of uncertainties, smarter sampling to generate all possible system behaviour (Pruyt & Islam, 2016) may also be an option. In this method, additional sampling takes place in spaces of interest, trying to find additional behavioural modes not found by standard Latin Hypercube sampling with a reasonable sample size.

7.3 Reflections on costs of exploratory modelling

The costs of exploratory modelling compared to consolidative modelling include longer time needed for model development and analysis, higher computational costs, and an increased risk of information overload.

As evidenced by the extended modelling cycle, explicitly accounting for deep uncertainty during model development means exploring multiple alternative conceptualisations of systems or parts thereof, assessing whether or not these alternative conceptualisations make a difference, and analysing larger amounts of model results. All these different choices lead to extended model development and analysis time.

In addition to that, running one or more models for thousands of different computational experiments, rather than a handful of manually selected scenarios, leads to a proportionate increase in computational costs. It was the relative short run time for a single experiment with an SD model that kept these additional costs within tenable limits. For more computationally intensive model paradigms, however, this is a potentially problematic issue. Another source of increased computational costs arises from the size of the output data sets being generated through exploratory modelling. These data sets are always high-dimensional, often temporal, and sometimes spatially explicit. Analysing these data sets stretches the demands imposed on computing hardware, such as system memory and processing speed. A promising direction for future research is to draw on technological development in the field of stream or real-time data analytics.

Besides the computational requirements for the analysis, the large amounts of data generated may also rapidly cause an information overload for the analyst. Fig. 7.5 is still a conservative picture in exploratory modelling, as the 2000 runs depicted still show a relatively neat set of behavioural modes. However, especially for more highly dimensional pictures, and phase plots, the ease of interpretation quickly deteriorates. Therefore, the analyst needs to rely on subsequent analyses to make sense of outcomes that previously could be directly comprehended. For example, compare a graph with a few lines with the visual spaghetti of Fig. 7.5. Next to this information overload, exploratory modelling easily results in a combinatoric explosion of possibilities, which often create amounts of data that cannot be analysed within one research project. For example, in the shale gas study (Chapter 4), I analysed the impacts of eight different price scenarios and 73 additional uncertainties on the state stability of six oil and gas producing countries. To make this possible, I explicitly limited myself to consider only a single metric of state stability, rather than considering a range of relevant indicators, such as demographics, youth unemployment, purchasing power, the development of the national economy and resource rents, and the situation of government finances and polity. In later research not part of this dissertation (Chivot et al., 2016), this problem was exacerbated as 167 countries were analysed.

7.4 Reflections on communicating policy relevant insights

After model development and use, a policy analyst needs to communicate about the results and underlying research. This communication takes place with stakeholders in policy discussions, and in domain specific and methodological scientific journals.

The goal of the research presented in this dissertation was in all cases to increase awareness of plausible, problematic, and potentially overlooked futures. To find these, we chose to perform open explorations of potential system behaviour, from which problematic behaviours could be selected. Open exploratory modelling has a policy agenda-setting goal (Rogers & Dearing, 1988). This corresponds to the first of four roles of models in the policy life cycle (van Daalen, Dresen, & Janssen, 2002). The agenda-setting goal is emphasized by the fact that even while all research in my dissertation was performed for a client, the client’s intention was never to directly inform new policy development. Agenda setting was further illustrated by several newspaper articles following the research (e.g., De Jong & Auping, 2014a, 2014b).

Regardless of the research objectives of the papers in this dissertation, the underlying work induced both negative and positive reactions. In the following sub sections, I will provide some anecdotal illustrations of these responses and try to assess their causes. It thus seems that for novel methodologies scientists and policy analysts should be careful in picking their battles, and should be trying to focus on a pragmatic argument of complementarity of exploratory approaches compared to existing approaches.

7.4.1 Reception in policy discussions

Policy-makers reacted both negatively and positively when they encountered our EMA based policy research. They reacted negatively for many reasons. First, the EMA methodology often proved to be difficult to understand. To emphasise that, policy-makers often replied that the research was considered to be very innovative, and that due to that fact they needed more time to fully grasp it's added value. In that sense, "very innovative" appeared to be used as a polite, but negative value judgement. Second, it could be noted that quantitative simulation modelling approaches were often considered to be relatively expensive when compared to qualitative research. This cost issue was found to limit accepting research proposals. Third, in some cases, policy-makers and stakeholders found it difficult to accept findings from our research, as they found them counterintuitive. For example, in presentations of the shale gas research in late 2013 and early 2014, attendees commented that they would expect oil prices only to go up, and not to go down. However, the fact that oil prices fell after July 2014 did corroborate our research, and considerably increased acceptance of our findings.

Policy-makers, and other stakeholders, also had some positive reactions to this research. Overall, the policy-makers appreciated the fact that uncertainty was taken seriously. They did not ask for probabilistic statements about the chances of particular scenarios happening. They also recognised that new insights were provided, which they considered to be in part due to the novelty of the approach used. This was the case for the climate mitigation and the Ebola research, where policy-makers were positively surprised by the findings.

In addition to the positive reactions, the exploratory approach fits the lingo and buzz words that policy-makers like to use, making them more inclined to accept the study's results. Examples of these words are both 'uncertainty' and 'complexity', sometimes combined with 'robustness' and 'resilience'. While the policy-makers' definitions of these terms may not be completely consistent with the definitions used by the analysts or researchers, the researcher's definitions do generally resonate. Similarly, it appears as if potential clients in these cases are more apt to grant a project where EMA is the underlying methodology.

In communicating the end result, it proved most successful to focus mostly on the central narrative of the conclusions, instead of the methodological approach underlying the research. In this sense, I often use the analogy of building a house in comparison to exploratory modelling research. The model and model analysis act as the ‘scaffolding’ to build a ‘house’ of one or more narratives which can be communicated. The modelling exercise helps to build consistency in these narratives, as the model underlying the stories needs to be consistent as well. Scenario Discovery, phase plots of different endogenously modelled performance indicators, combined with the system knowledge that arose during model development further strengthens this construction. If the analyst removes the ‘scaffolding’ by not communicating the methodological parts of the research, the narratives should still be plausible by themselves. If the methodological approach is communicated later, it will only strengthen the story, instead of raising the normal questions (e.g., “is this or that in the model?”, “how do you avoid garbage in, garbage out?”, “how did you validate your model?”) which, in my opinion, generally arose when stakeholders did not fully grasp the intricacies of the research’s methodology, and led to less acceptance of the conclusions. The risk you run when choosing this communication strategy is that stakeholders will say that your conclusions are trivial, or that they have come up with them earlier on. In such situations, it is generally sufficient to refer to the fact that no evidence exists in literature or other communications of those judgements.

7.4.2 Reception in domain specific fields

In domain specific fields, reactions were also mixed. Sometimes, policy analysts’ perceptions appeared to depend mostly on the potential of the research to be threatening to their own methods. For example, when I was researching the long-term potential return on innovation investments, a researcher with a background in econometrics was very critical of the approach I used. This was in part due to incompatibility of the terminology used by each of us, and (e.g., the difference between “calibrating” a simulation model with data, and “estimating” an econometric model), perhaps, in part due to the fact that he was working on the same topic with a different methodology. In the end, the project was discontinued due to a lack of progress combined with this criticism. In a preliminary presentation of the shale gas research, other researchers were invited to comment on our findings. In these meetings, some of the attendees used the opportunity to try to discredit the research as much as possible before it was published. This could also be due to the fact that the conclusions of our research (i.e., energy prices could fall as a consequence of the shale revolution) contradicted their findings (i.e., energy prices will rise).

In contrast to the negative reactions, other contributions in our papers were received positively. For example, the reviews of the Ebola paper were very positive and emphasized the innovative side of this research, which modelled disease transmission in feedback with the intervention capacities. Earlier, this innovative side of the research was also recognised by researchers using transmission models with different underlying methodological approaches. These researchers were familiar with SD as a modelling paradigm, acknowledged the omnipresence of uncertainty in modelling of outbreaks, and were aware of potential future collaborations. Similarly, the reviews for the shale gas paper were very positive and regretted that the work was not published earlier. The corroboration of our findings by the oil price decline may have helped here.

7.4.3 Reception in methodological fields

The reception of my papers in methodological fields was mixed as well. In part, SD modellers objected to the use of the uncertainty tools in combination with SD. Part of the objection was based on the perceived idea that exploratory SD models were not sufficiently based on literature, or used too many inputs and had, therefore, too little endogenous feedbacks to justify SD as a modelling paradigm. In addition, the consolidative modelling paradigm dominating the SD field made it difficult for some to grasp the idea of irreducible uncertainty. In a world view found among many consolidative SD modellers, not being able to reduce especially model structure uncertainty is a sign of insufficiently rigorous SD modelling.

Finally, certain arguments used in early exploratory SD papers were considered offensive, and reduced the acceptance of the approach. For example, an early version of the conference paper about copper scarcity included the following paragraph, which was similarly used in other papers from our research group: *"Rather than specifying a single model and falsely treating it as a reliable image of the system of interest, the available information is consistent with a set of models, whose implications for potential decisions may be quite diverse. A single model run drawn from this potentially infinite set of plausible models is not a "prediction"; rather, it provides a computational experiment that reveals how the world would behave if the various guesses any particular model makes about the various irresolvable uncertainties were correct."* (Auping, Pruyt, & Kwakkel, 2012). The first sentence in this paragraph was considered a false straw man argument, as, according to the reviewer, SD does not prescribe the use of a single model as a reliable image of the system. Further, SD is not considered to be a predictive method.

In later papers (e.g., Chapter 5), which were at least in part written for an SD readership, we tried to focus more on the complementarity of the approach followed in our research. The reviews of this work focussed more on traditional SD issues, like the endogenous structural explanation of the model's and system's behaviours. For example, one of the issues was the lack of endogenous policy feedback on potentially untenable or even theoretically impossible spending on societal ageing related health care costs. While we were able to argue why this choice was justified in this case, it was valid criticism. The formal reaction to the paper in a discussant's reaction was very positive about our approach and the added value of using this methodology "for strategic planning, policy design, and performance management in the public sector" (Bianchi, 2015).

Besides these two examples, also in discussions with other colleagues that did not use EMA – or SD – framing our work as complementary to their work always worked better in improving the acceptance of our work in contrast to a more conflictual approach. Finally, focussing on those elements where the approaches do not differ at all increased this mutual understanding.

7.5 Conclusions and discussion

The goal of this dissertation was to illustrate and analyse how deep uncertainty can affect model development and use. I found that including deep uncertainty in model-based policy or system analysis significantly influences model development, use, and communication. Model development becomes in this way part of an uncertainty modelling cycle. When highly non-linear models are used to explore potential future developments, this also complicates Scenario Discovery, a common practice of exploratory modelling. I expect that these findings also apply in other modelling paradigms, especially for other simulation or equation-based modelling paradigms. Finally, the difficulties with the application of Scenario Discovery to results generated with highly non-linear simulation models have not yet been resolved.

Irreducible uncertainty in an early phase of model development leads to irreducible uncertainty in all subsequent phases. This is a key finding of the systematic assessment of the operationalisation of the model development cycle.

Currently, most EMA work only recognises uncertainty in the evaluation and policy analysis phases of the model development cycle. In that practice, a large number of runs is generated in the evaluation phase. These runs are subsequently used in the policy testing phase to assess whether the proposed policies are robust. However, recognising deep uncertainty in earlier phases may lead to very different policy choices. Uncertainty in the problem articulation phase may lead to broader scopes, or even choosing multiple scopes for models, rather than the consolidative approach of a single scope. Uncertainty in the conceptualisation phase will lead to distinguishing large, structural uncertainties, and may lead to developing multiple conceptual models for the research problem at hand. Uncertainty in the formulation phase will lead to the use of switches to alternate over different model structures.

However, even in the two phases of the model development cycle where uncertainty is already being acknowledged in other EMA work, uncertainty results in differences from the consolidative modelling approach. As generating some understanding about model quality and whether models are fit for purpose is crucial in work for clients and stakeholders, verification and validation steps need to be taken, but may considerably change compared to their consolidative application. Generally, it can be concluded that if a model is to be fit for purpose for exploratory modelling uses, it needs to meet more extensive tests than models designed for consolidative use.

Finally, policy uncertainty does exist as well and is caused by uncertainty about implementation and system response. The traditional approach of modelling policies as shocks in the model does not recognise this fact, nor does the univariate and non-systematic testing of policy effects. Solutions for dealing with these last causes of uncertainty include considering policy reactions more as an integral part of system functioning, and making policy formulation technically similar to other parts of the model formulation, and policy analysis similar to model analysis. While this may not be new, exploratory modelling brings to the fore its importance.

These changes increase the costs of the analysis. Relevant costs include longer time needed for model development and analysis, increased computational requirements for performing and analysing large ensembles of computational experiments, and the risk of information overload. An implicit trade-off exists between these additional costs and the additional policy relevant insights that can be achieved solely due to exploratory modelling.

In all the different phases of exploratory model development, it is crucial to communicate to clients and stakeholders about the choices made, but also to scientific peers. Therefore, communication about uncertainties remains important during all phases of model development and use (e.g., Van der Sluijs et al., 2003). First, stakeholders need explanations when different world views regarding their problem or problems may not be reduced to a single model. Second, further structural uncertainties regarding their problems need to be communicated. I have proposed an addition to the diagrammatic conventions of CLDs used in the field of SD to accommodate this, but this proposal is mostly relevant for methodologists. Third, it needs to be clear what ranges exist for parametric uncertainties, and whether all relevant structural uncertainties could and have been modelled.

All cases presented in this dissertation used SD models. This was convenient due to the relatively short simulation time of SD models, the ease of incorporating structural uncertainties, and the well-developed link with exploratory modelling software. The results of this dissertation can be generalised to some extent to other modelling paradigms. Boundary selection is something that needs to happen for any type of modelling, just like conceptualising what is going to be modelled. Formulation of equations is something that needs to happen in any equation-based modelling paradigm. Uncertainty in modelling paradigms like Agent-Based modelling can be acknowledged as well in similar ways. Conclusions about verification and validation are very specific for SD, but models that stand the test of exploratory modelling need to be thoroughly developed, as they will otherwise easily break when uncertainty sampling generates situations outside the envisioned bandwidth of model operation. Finally, for all modelling paradigms that allow explicit feedbacks, policy development may be modelled endogenously and as uncertain as it is in reality.

Finally, using non-linear dynamic models including feedback, accumulation, and delay, does complicate conventional Scenario Discovery practices. These problems occur due to difficulties in making good classifications of model-generated behavioural modes, non-linear transformation of inputs to outputs, and, if the boundary selection and aggregation level demanded, large numbers of input parameters. Solutions for these issues may be found in smarter and more interactive sampling techniques, focussing on the links between endogenous model variables, and especially focussing on the impossibility of specific behaviour given parts of the input space in combination with other endogenous variables.

This research made clear that multiple areas need further attention in future research. With regard to model development, it would be interesting to see whether the suggestions I have made regarding model development can be corroborated and expanded by assessing the impacts of uncertainty in other modelling paradigms than SD. However, the most important advances are to be made in the analysis of results generated with non-linear dynamics in Scenario Discovery. Multiple problems are still unresolved in this aspect. Behavioural classification of dynamics through time can be improved, potentially leading to clearer results of algorithms like PRIM and CART on these clustered results. Related to this, adaptive sampling methods can be improved to allow efficient and effective sampling over the input space to generate a richer view on plausible future behaviour. Further, Scenario Discovery with simulation models can be complemented by more in-depth analyses of the interdependencies amongst different outcomes of interest.

Regardless the still existing difficulties with DMDU, my experience over the few last years demonstrated to me the enormous potential this methodology has for bringing policy modelling forward. The ideas of being able to develop policies that are robust for all stakeholders, and the possibilities to develop strong and transparent narratives by thoroughly understanding the effects of deep uncertainty in non-linear models of complex systems, are incredibly powerful. For that reason, I hope that the findings in this dissertation will help future researchers in strengthening their exploratory research by developing and using thoroughly exploratory models, analysing the output of these models with fitting tools, and effectively communicating their findings to clients and stakeholders alike.

References

Ackoff, R. L. (1974). *Redesigning the Future: A Systems Approach to Societal Problems*. New York: John Wiley & Sons.

Ackoff, R. L. (1979). The Future of Operational Research is Past. *The Journal of the Operational Research Society, 30*(2), 93-104. doi:10.2307/3009290

Andersen, D. F., & Sturis, J. (1988). Chaotic structures in generic management models: Pedagogical principles and examples. *System Dynamics Review, 4*(1-2), 218-245. doi:10.1002/sdr.4260040113

Auping, W. L., Pruyt, E., & Kwakkel, J. H. (2012). *Analysing the Uncertain Future of Copper with Three Exploratory System Dynamics Models* Paper presented at the The 30th International Conference of the System Dynamics Society, St. Gallen. http://www.systemdynamics.org/conferences/2012/proceed/papers/P1385.pdf

Balci, O. (1994). Validation, verification, and testing techniques throughout the life cycle of a simulation study. *Annals of Operations Research, 53*(1), 121-173. doi:10.1007/BF02136828

Balci, O. (2013). Verification, Validation, and Testing of Models. In S. I. Gass & M. C. Fu (Eds.), *Encyclopedia of Operations Research and Management Science* (pp. 1618-1627). Boston, MA: Springer US.

Bankes, S. C. (1993). Exploratory Modeling for Policy Analysis. *Operations Research, 41*(3), 435-449. doi:10.1287/opre.41.3.435

Bankes, S. C., Walker, W. E., & Kwakkel, J. H. (2013). Exploratory Modeling and Analysis. In G. S. & M. C. Fu (Eds.), *Encyclopedia of Operations Research and Management Science* (3rd ed.). Berlin, Germany: Springer.

Barlas, Y. (1996). Formal aspects of model validity and validation in system dynamics. *System Dynamics Review, 12*(3), 183-210.

Bertalanffy, L. v. (1968). *General Systems Theory*. New York: George Braziller.

Bianchi, C. (2015). Enhancing Joined-Up Government and Outcome-Based Performance Management through System Dynamics Modelling to Deal with Wicked Problems: the Case of Societal Ageing. *Systems Research and Behavioral Science, 32*(4), 502-505. doi:10.1002/sres.2341

Bleijenbergh, I., Vennix, J., Jacobs, E., & van Engen, M. (2016). Understanding decision making about balancing two stocks: the faculty gender balancing task. *System Dynamics Review, 32*(1), 6-25. doi:10.1002/sdr.1548

Bobashev, G. V., Goedecke, D. M., Yu, F., & Epstein, J. M. (2007). *A Hybrid Epidemic Model: Combining The Advantages Of Agent-Based And Equation-Based Approaches*. Paper presented at the 2007 Winter Simulation Conference. http://ieeexplore.ieee.org/document/4419767/

Breiman, L. (2001). Random Forests. *Machine Learning, 45*, 5-32.

Breiman, L., Friedman, J. H., Stone, C., & Olshen, R. A. (1984). *Classification and Regression Trees*. Monterey, CA: CRC Press.

Bruckner, E., Ebeling, W., & Scharnhorst, A. (1989). Stochastic dynamics of instabilities in evolutionary systems. *System Dynamics Review, 5*(2), 176-191. doi:10.1002/sdr.4260050206

Bryant, B. P., & Lempert, R. J. (2010). Thinking inside the box: A participatory, computer-assisted approach to scenario discovery. *Technological Forecasting & Social Change, 77*, 34-49. doi:10.1016/j.techfore.2009.08.002

Bucher, C. G. (1988). Adaptive sampling — an iterative fast Monte Carlo procedure. *Structural Safety, 5*(2), 119-126. doi:10.1016/0167-4730(88)90020-3

Camacho, A., Kucharski, A. J., Funk, S., Breman, J., Piot, P., & Edmunds, W. J. (2014). Potential for large outbreaks of Ebola virus disease. *Epidemics, 9*, 70-78. doi:10.1016/j.epidem.2014.09.003

Chapman, P. F., & Roberts, F. (1983). *Metal Resources and Energy* (1st ed.): Butterworth & Co (Publishers) Ltd.

Chivot, E., Auping, W. L., De Jong, S., Rõõs, H., & Rademaker, M. (2016). *The Netherlands in the Global Food System*. Retrieved from The Hague, The

Netherlands: http://hcss.nl/sites/default/files/files/reports/HCSS_Food%20Systems%20Final%20version%2022%20november%202016.pdf

Chowell, G., Hengartner, N. W., Castillo-Chavez, C., Fenimore, P. W., & Hyman, J. M. (2004). The basic reproductive number of Ebola and the effects of public health measures: the cases of Congo and Uganda. *J Theor Biol, 229*(1), 119-126. doi:10.1016/j.jtbi.2004.03.006

Coyle, R. G. (1983). The technical elements of the system dynamics approach. *European Journal of Operational Research, 14*(4), 359-370. doi:10.1016/0377-2217(83)90236-9

Craik, K. J. W. (1943). *The Nature of Explanation*. Cambridge, UK: Cambridge University Press.

Dalal, S., Han, B., Lempert, R. J., Jaycocks, A., & Hackbarth, A. (2013). Improving scenario discovery using orthogonal rotations. *Environmental Modelling & Software, 48*, 49-64. doi:10.1016/j.envsoft.2013.05.013

Davis, P. K. (2000). *Exploratory analysis enabled by multiresolution, multiperspective modeling.* Paper presented at the 2000 Winter Simulation Conference Proceedings, Orlando, FL.

Davis, P. K., & Bigelow, J. H. (1998). *Experiments in Multiresolution Modeling*. Retrieved from Santa Monica, CA: https://www.rand.org/content/dam/rand/pubs/monograph_reports/2007/MR1004.pdf

De Jong, S., & Auping, W. L. (2014a, 22 February 2014). Gaat schaliegas Poetin de kop kosten? *Het Parool*.

De Jong, S., & Auping, W. L. (2014b, 14 maart 2014). Sancties kunnen Poetin wel degelijk hard raken. *De Volkskrant*. Retrieved from http://www.volkskrant.nl/archief/sancties-kunnen-poetin-wel-degelijk-hard-raken~a3612728/

De Jong, S., Auping, W. L., & Govers, J. (2014). *The Geopolitics of Shale Gas*. Retrieved from The Hague: http://static.hcss.nl/files/uploads/2180.pdf

De Jong, S., Auping, W. L., Oosterveld, W. T., Usanov, A., Abdalla, M., Van de Bovenkamp, A., & Frattina della Frattina, C. (2017). *The Geopolitical Impact of Climate Mitigation Policies. How Hydrocarbon Exporting Rentier States and Developing Nations can Prepare for a More Sustainable Future*. Retrieved from The Hague: http://hcss.nl/sites/default/files/files/reports/HCSS_Energytransition.pdf

Doyle, J. K., & Ford, D. N. (1998). Mental models concepts for system dynamics research. *System Dynamics Review, 14*(1), 3-29. doi:10.1002/(SICI)1099-1727(199821)14:1<3::AID-SDR140>3.0.CO;2-K

Doyle, J. K., & Ford, D. N. (1999). Mental models concepts revisited: some clarifications and a reply to Lane. *System Dynamics Review, 15*(4), 411-415. doi:10.1002/(SICI)1099-1727(199924)15:4<411::AID-SDR181>3.0.CO;2-R

Eberlein, R. L., & Thompson, J. P. (2013). Precise modeling of aging populations *System Dynamics Review, 29*(2), 87-101. doi:10.1002/sdr.1497

Epstein, J. M., & Axtell, R. L. (1996). *Growing Artificial Societies: Social Science From the Bottom Up*. Washington, D.C.: Brookings Institution Press.

Fasina, F. O., Shittu, A., Lazarus, D., Tomori, O., Simonsen, L., Viboud, C., & Chowell, G. (2014). Transmission dynamics and control of Ebola virus disease outbreak in Nigeria, July to September 2014. *Eurosurveillance Edition 2014, 19*(40), pii=20920. http://www.eurosurveillance.org/ViewArticle.aspx?ArticleId=20920 Retrieved from http://www.eurosurveillance.org/ViewArticle.aspx?ArticleId=20920

Fischbach, J. R. (2010). *Managing New Orleans Flood Risk in an Uncertain Future Using Non-Structural Risk Mitigation.* (PhD PhD dissertation), Pardee RAND Graduate School. Retrieved from http://www.rand.org/pubs/rgs_dissertations/RGSD262.html (RGSD-262)

Fisman, D., Khoo, E., & Tuite, A. (2014). Early Epidemic Dynamics of the West African 2014 Ebola Outbreak: Estimates Derived with a Simple Two-Parameter Model. *PLoS Currents*. doi:10.1371/currents.outbreaks.89c0d3783f36958d96ebbae97348d571

Forrester, J. W. (1969). *Urban Dynamics*: Pegasus Communications.

Forrester, J. W. (1987). Lessons from system dynamics modeling. *System Dynamics Review, 3*(2), 136-149. doi:10.1002/sdr.4260030205

Forrester, J. W. (1994). System dynamics, systems thinking, and soft OR. *System Dynamics Review, 10*(2-3), 246-256. doi:10.1002/sdr.4260100211

Forrester, J. W., & Senge, P. M. (1980). Tests for Building Confidence in System Dynamics Models. In A. A. Lagasto, J. W. Forrester, & J. M. Lyneis (Eds.), *TIMS Studies in the Management Sciences* (Vol. 14, pp. 209-228). Amsterdam, Netherlands: North-Holland Publishing Company.

Friedman, J. H., & Fisher, N. I. (1999). Bump hunting in high-dimensional data. *Statistics and Computing, 9*(2), 123-143. doi:10.1023/A:1008894516817

Gerst, M. D. (2008). Revisiting the Cumulative Grade-Tonnage Relationship for Major Copper Ore Types. *Ecological Geology, 103*, 615-628. doi:10.2113/gsecongeo.103.3.615

Gordon, R. B., Bertram, M., & Graedel, T. E. (2006). Metal stocks and sustainability. *Proceedings of the National Academy of Sciences, 103*(5), 1209-1214. doi:10.1073/pnas.0509498103

Gordon, R. B., Bertram, M., & Graedel, T. E. (2007). On the sustainability of metal supplies: A response to Tilton and Lagos. *Resources Policy, 32*(1-2), 24-28. doi:10.1016/j.resourpol.2007.04.002

Haasnoot, M., Kwakkel, J. H., Walker, W. E., & ter Maat, J. (2013). Dynamic adaptive policy pathways: A method for crafting robust decisions for a deeply uncertain world. *Global Environmental Change, 23*(2), 485-498. doi:10.1016/j.gloenvcha.2012.12.006

Hamarat, C., Kwakkel, J. H., Pruyt, E., & Loonen, E. T. (2014). An exploratory approach for adaptive policymaking by using multi-objective robust optimization. *Simulation Modelling Practice and Theory, 46*, 25-39. doi:10.1016/j.simpat.2014.02.008

Hanau, A. (1928). *Die Prognose der Schweinepreise*. Berlin: Verlag von Reimar Hobbing.

Herman, J. D., Reed, P. M., Zeff, H. B., & Characklis, G. W. (2015). How Should Robustness Be Defined for Water Systems Planning under Change? *Journal of Water Resources Planning and Management, 141*(10), 04015012. doi:10.1061/(ASCE)WR.1943-5452.0000509

Hodges, J. S., & Dewar, J. A. (1992). *Is It You or Your Model Talking? A Framework for Model Validation*. Retrieved from Santa Monica, CA: http://www.rand.org/content/dam/rand/pubs/reports/2006/R4114.pdf

Hoekstra, A. Y. (1998). *Perspectives on Water. An Integrated Model-Based Exploration of the Future.* (PhD), Delft University of Technology, Delft. Retrieved from http://repository.tudelft.nl/islandora/object/uuid:bb7eaa6b-012e-453c-ab81-b6b40cc2e3ff?collection=research

Islam, T., & Pruyt, E. (2016). Scenario generation using adaptive sampling: The case of resource scarcity. *Environmental Modelling & Software, 79*, 285-299. doi:10.1016/j.envsoft.2015.09.014

Johnson-Laird, P. N. (1983). *Mental Models: Towards a Cognitive Science of Language, Inference, and Consciousness. Cambridge*. Cambridge, MA: Harvard University Press.

Keating, E. H. (1998). *Everything You Ever Wanted to Know about How to Develop A System Dynamics Model, But Were Afraid to Ask*. Paper presented at the 16th International Conference of the System Dynamics Society, Quebec, Canada. http://www.systemdynamics.org/conferences/1998/PROCEED/00024.PDF

Kleijnen, J. P. C. (1997). Sensitivity analysis and related analyses: A review of some statistical techniques. *Journal of Statistical Computation and Simulation, 57*(1-4), 111-142. doi:10.1080/00949659708811805

Kwakkel, J. H., Walker, W. E., & Haasnoot, M. (2016). Coping with the Wickedness of Public Policy Problems: Approaches for Decision Making under Deep Uncertainty. *Journal of Water Resources Planning and Management, 142*(3), 1-5. doi:10.1061/(ASCE)WR.1943-5452.0000626

Kwakkel, J. H., Walker, W. E., & Marchau, V. A. W. J. (2010). Classifying and communicating uncertainties in model-based policy analysis. *International Journal of Technology, Policy and Management, 10*(4), 299-315. doi:10.1504/IJTPM.2010.036918

Lane, D. C. (1993). The road not taken: Observing a process of issue selection and model conceptualization. *System Dynamics Review, 9*(3), 239-264. doi:10.1002/sdr.4260090303

Lane, D. C. (1995). *The Folding Star: A comperative reframing and extension of validity concepts in system dynamics*. Paper presented at the Procedings of the 1995 International System Dynamics Conference, Tokyo, Japan.

Lane, D. C. (1999). Friendly amendment: a commentary on Doyle and Ford's proposed re-definition of ‘mental model’. *System Dynamics Review, 15*(2), 185-194. doi:10.1002/(SICI)1099-1727(199922)15:2<185::AID-SDR166>3.0.CO;2-C

Lempert, R. J., Bryant, B. P., & Bankes, S. C. (2008). *Comparing Algorithms for Scenario Discovery*. Retrieved from Santa Monica, CA.: http://www.rand.org/pubs/working_papers/WR557.html

Lempert, R. J., Groves, D. G., Popper, S. W., & Bankes, S. C. (2006). A General, Analytic Method for Generating Robust Strategies and Narrative Scenarios. *Management Science, 52*(4), 514-528. doi:10.1287/mnsc.1050.0472

Lempert, R. J., Popper, S. W., & Bankes, S. C. (2003). *Shaping the Next One Hundred Years: New Methods for Quantitative, Long-Term Policy Analysis* (MR-1626). Retrieved from Santa Monica, CA, USA: http://www.rand.org/pubs/monograph_reports/2007/MR1626.pdf

Logtens, T., Pruyt, E., & Gijsbers, G. (2012). *Societal Aging in the Netherlands: Exploratory System Dynamics Modeling and Analysis.* Paper presented at the International Conference of the System Dynamics Society, St. Gallen.

Lyneis, J. M. (2000). System dynamics for market forecasting and structural analysis. *System Dynamics Review, 16*(1), 3-25. doi:10.1002/(SICI)1099-1727(200021)16:1<3::AID-SDR183>3.0.CO;2-5

Meadows, D. H., Meadows, D. L., Randers, J., & Behrens, W. W., III. (1972). *The Limits to Growth: A Report for the Club of Rome's Project on the Predicament of Mankind*. New York: Universe Books.

Meadows, D. H., Richardson, J., & Bruckmann, G. (Eds.). (1982). *Groping in the dark*. Bristol: J.W. Arrowsmith Limited.

Molina-Perez, E. (2016). *Directed International Technological Change and Climate Policy: New Methods for Identifying Robust Policies Under Conditions of Deep Uncertainty.* (PhD), RAND Corporation, Santa Monica, CA. Retrieved from http://www.rand.org/pubs/rgs_dissertations/RGSD369.html (RGSD-369)

Moorlag, R. A. C. W. M., Pruyt, E., Auping, W. L., & Kwakkel, J. H. (2015). Shale Gas and Import Dependency: A Multi-Method Model-Based Exploration. *International Journal of System Dynamics Applications, 4*(1), 31-56. doi:10.4018/IJSDA.2015010102

Morecroft, J. D. W. (1982). A Critical Review of Diagramming Tools for Conceptualizing Feedback System Models. *Dynamica, 8*(1), 20-29.

Oreskes, N., Shrader-Frechette, K., & Belitz, K. (1994). Verification, Validation, and Confirmation of Numerical Models in the Earth Sciences. *Science Magazine, 263*(5147), 641-646. doi:10.1126/science.263.5147.641

Peterson, D. W., & Eberlein, R. L. (1994). Reality check: A bridge between systems thinking and system dynamics. *System Dynamics Review, 10*(2-3), 159-174. doi:10.1002/sdr.4260100205

Powersim Software. (2017). Powersim Studio (Version 10). Bergen, Norway.

Pruyt, E. (2012). *State of Art versus State of Science of Integrated Risk-Capability Analysis under Deep Uncertainty*. Paper presented at the Third International Engineering Systems Symposium, Delft University of Technology.

Pruyt, E., Auping, W. L., & Kwakkel, J. H. (2015). Ebola in West Africa: Model-Based Exploration of Social Psychological Effects and Interventions. *Systems Research and Behavioral Science, 32*(1), 2-14. doi:10.1002/sres.2329

Pruyt, E., & Islam, T. (2016). On generating and exploring the behavior space of complex models. *System Dynamics Review, Published online*. doi:10.1002/sdr.1544

Rahmandad, H., & Sterman, J. D. (2008). Heterogeneity and Network Structure in the Dynamics of Diffusion: Comparing Agent-Based and Differential Equation Models. *Management Science, 54*(5), 998-1014. doi:10.1287/mnsc.1070.0787

Randers, J. (1980a). Guidelines for Model Conceptualization. In J. Randers (Ed.), Elements of the System Dynamics Method (pp. 117-139). Cambridge, MA: Productivity Press.

Randers, J. (Ed.) (1980b). *Elements of the System Dynamics Method*. Cambridge, MA: Productivity Press.

Rasmussen, S., Mosekilde, E., & Sterman, J. D. (1985). Bifurcations and chaotic behavior in a simple model of the economic long wave. *System Dynamics Review, 1*(1), 92-110. doi:10.1002/sdr.4260010108

Reiner, R., Munz, M., & Weidlich, W. (1988). Migratory dynamics of interacting subpopulations: Regular and chaotic behavior. *System Dynamics Review, 4*(1-2), 179-199. doi:10.1002/sdr.4260040110

Richardson, G. P., & Pugh, A. L., III. (1981). *Introduction to System Dynamics Modeling with Dynamo*. Cambridge, MA: MIT Press.

Rittel, H. W. J., & Webber, M. M. (1973). Dilemmas in a General Theory of Planning. *Policy Sciences, 4*(1973), 155-169. doi:10.1007/BF01405730

Rogers, E. M., & Dearing, J. W. (1988). Agenda-Setting Research: Where Has It Been, Where Is It Going? *Annals of the International Communication Association, 11*(1), 555-594. doi:10.1080/23808985.1988.11678708

Schwarz, P. (2016). *Modeling and simulating the Zika Outbreak under Deep Uncertainty. A Multi-Model Multi-Resolution Approach.* (MSc), Delft University of Technology, Delft. Retrieved from http://repository.tudelft.nl/islandora/object/uuid:4957df8e-3de1-4b5e-8231-731287a4ede4/datastream/OBJ/download (uuid:4957df8e-3de1-4b5e-8231-731287a4ede4)

Schwarz, P., & Pruyt, E. (2016). *Modeling and simulation the Zika Outbreak under Deep Uncertainty: A Multi-Method Multi-Resolution Approach*. Paper presented at the International System Dynamics Conference, Delft. http://www.systemdynamics.org/conferences/2016/proceed/papers/P1311.pdf

Sterman, J. D. (2000). *Business Dynamics: Systems Thinking and Modeling for a Complex World*. New York: McGraw.

Sterman, J. D., Fiddaman, T., Franck, T., Jones, A., McCauley, S., Rice, P., . . . Siegel, L. (2012). Climate interactive: the C-ROADS climate policy model. *System Dynamics Review, 28*(3), 295-305. doi:10.1002/sdr.1474

Sutrisno, A., & Handel, O. (2012). *Dynamic Aging Population in Germany: A case study about demographic change.* Paper presented at the International Conference of the System Dynamics Society, St. Gallen.

Thissen, W. A. H. (1978). *Investigations into the World3 Model: Lessons for Understanding Complicated Models.* (PhD), Eindhoven University of Technology, Eindhoven.

Thompson, J. P., Riley, C. M., Eberlein, R. L., & Matchar, D. B. (2012). Future living arrangements of Singaporeans with age-related dementia. *International Psychogeriatrics, 24*(10), 1592-1599.

Thompson, K. M., & Duintjer Tebbens, R. J. (2009). Using system dynamics to develop policies that matter: global management of poliomyelitis and beyond. *System Dynamics Review, 24*(4), 433-449. doi:10.1002/sdr.419

Tilton, J. E. (1996). Exhaustible resources and sustainable development: two different paradigms. *Resources Policy, 22*(1-2), 91-97. doi:10.1016/S0301-4207(96)00024-4

Tilton, J. E., & Lagos, G. (2007). Assessing the long-run availability of copper. *Resources Policy, 32*(1-2), 19-23. doi:10.1016/j.resourpol.2007.04.001

van Daalen, C. E., Dresen, L., & Janssen, M. A. (2002). The roles of computer models in the environmental policy life cycle. *Environmental Science & Policy, 5*(3), 221-231. doi:10.1016/S1462-9011(02)00040-0

Van der Sluijs, J. P., Risbey, J. S., Kloprogge, P., Ravetz, J. R., Funtowicz, S. O., Corral Quintana, S., . . . Huijs, S. W. F. (2003). *RIVM/MNP Guidance for uncertainty assessment and communication: detailed guidance*. Retrieved from http://www.nusap.net/downloads/detailedguidance.pdf

Vennix, J. A. M. (1999). Group model-building: tackling messy problems. *System Dynamics Review, 15*(4), 379-401.

Ventana Systems. (2010). *Vensim Reference Manual*. Harvard, MA: Ventana Systems.

WHO Ebola Response Team. (2014). Ebola Virus Disease in West Africa — The First 9 Months of the Epidemic and Forward Projections. *New England Journal of Medicine, 371*(16), 1481-1495. doi:10.1056/NEJMoa1411100

Wunderlich, P., Größler, A., Zimmermann, N., & Vennix, J. A. M. (2014). Managerial influence on the diffusion of innovations within intra-organizational networks. *System Dynamics Review, 30*(3), 161-185. doi:10.1002/sdr.1516

Yücel, G., & Barlas, Y. (2011). Automated parameter specification in dynamic feedback models based on behavior pattern features. *System Dynamics Review, 27*(2), 195-215. doi:10.1002/sdr.457

Zeff, H. B., Herman, J. D., Reed, P. M., & Characklis, G. W. (2016). Cooperative drought adaptation: Integrating infrastructure development, conservation, and water transfers into adaptive policy pathways. *Water Resources Research, 52*. doi:10.1002/2016WR018771

Appendix A Glossary

A

Ageing chain *Methodological.* A one directional set of stocks connected by flows through which a quantity may flow. Population models and transmission models generally use ageing chains, but there are many other examples. *See also* Stock.

Agenda-setting *Methodological.* First phase in a policy life cycle which is aimed at getting an issue on the policy agenda.

Agent-Based Modelling (ABM) *Methodological.* Type of simulation modelling where one or more types of agents and rules how these agents interact with other agents and their environment are defined. Epstein and Axtell (1996).

Aggregation level *Methodological.* The aggregation level of a model determines the amount of detail in model variables and structures. A higher aggregation level means that variables are used which are compounds of lower aggregate variables that are not included in the model. The aggregation level is in part determined by existing world views, perspectives, or mental models regarding a problem and system. *See also* Mental model, Perspective.

Auxiliary variable *Methodological.* Model variable without direct mathematical necessity in SD. Auxiliaries help in making the modelled structure more insightful and linked to the real-world system than when the modeller would only use the mathematically significant stocks and flows. *See also* Flow, Stock.

B

Balancing feedback *Methodological.* Feedback relation involving two or more variables where a higher (lower) value for any given variable will, everything else remaining the same, result in a lower (higher) value of the same variable in the future. *See also* Reinforcing feedback.

Base case *Methodological.* The base case is defined and used in probabilistic scenario approaches like ‘La Prospective’ and ‘Probabilistic Modified Trends’ (including ‘Trend-Impact Analysis’ and ‘Cross-Impact Analysis’) as the most probable scenario, or as a business as usual scenario (Bradfield et al., 2005). In the case of simulation models, it is often used as the single model run used as example of most plausible system behaviour without explicit policy changes. *See also* Behavioural mode, Policy testing, Run, System behaviour.

Behavioural mode *Methodological.* Typical pattern of run time dynamics demonstrated by a model's KPIs. *See also* Exemplar.

Breaking of a model *Methodological*. In a EMA context, a model ‘breaks’ if after sampling over the input space one or more runs cannot be completed (e.g., due to a floating point error), or show behavioural modes which are impossible in the real system. *See also* Floating point error.

Boundary *Methodological. See* System boundary.

Boundary selection *Methodological.* Step in the problem articulation phase where the system boundary is selected. *See also* System boundary, Problem articulation, Scope.

C

Categorical uncertainty *Methodological.* Parametric uncertainty for which a finite set of values can be set. Examples include switches and sets of input uncertainties. *See also* Input scenario, Parametric uncertainty, Switch.

Causal Loop Diagram (CLD) *Methodological.* Diagram showing the causal relations between key variables in a modelled system, the direction of change of these relations, and the main feedbacks between variables. CLDs are often used as conceptual model and to communicate feedback loops responsible for particular types of behaviour. Morecroft (1982); Sterman (2000). *See also* Conceptualisation.

Classification and regression tree (CART) *Methodological.* CART is a data analysis method and classification algorithm that creates a tree like structure by making orthogonal cuts in the input space. Next to PRIM, CART is often used in Scenario Discovery to link bandwidths of uncertainties to particular behaviour or system states (e.g., undesirable futures) in the run ensemble. Breiman et al. (1984); Lempert et al. (2008). *See also* PRIM, Scenario Discovery.

Complexity *Methodological*. Complexity is a system characteristic where due to many interrelations between system elements, no higher rule can be defined that describes the behaviour of the whole system. According to the SD discipline, complexity caused by feedback and delay between system elements, and accumulation in system elements. Complexity often results in non-linearity of system behaviour. *See also* Delay, Feedback.

Conceptual model *Methodological*. Qualitative model or diagram of main relations and elements in a problem and system modelled. It is used to communicate about the link between real world and a quantitative model. In the earlier conceptualisation phase, conceptual models are used to determine which system elements need to be modelled and what their main relations are. *See also* Conceptualisation.

Conceptualisation *Methodological*. Model development phase where the qualitative structure of the problem and system to be researched. Conceptual models or diagrams like CLDs and SSDs can be used in this phase. Further, conceptualisation may also include the formulation of one or dynamic hypotheses. *See also* Conceptual model, Model development cycle.

Consolidative modelling *Methodological*. Development of a simulation model aimed at reducing uncertainties by combining knowledge from literature or stakeholders groups into a single representation of the system of interest. This is a common practice in most conventional modelling as well as group model-building. *See also* Conventional modelling, Group model-building.

D

Deep uncertainty *Methodological*. Lempert et al. (2003) define deep uncertainty as conditions "where analysts do not know, or the parties to a decision cannot agree on, (1) the appropriate conceptual models that describe the relationships among the key driving forces that will shape the long-term future, (2) the probability distributions used to represent uncertainty about key variables and parameters in the mathematical representations of these conceptual models, and/or (3) how to value the desirability of alternative outcomes". Kwakkel et al. (2010) see deep uncertainty as one of four levels of uncertainty (i.e., shallow, medium, and deep uncertainty, and recognised ignorance). Lempert et al. (2003), Kwakkel et al. (2010). *See also* EMA.

Delay *Methodological*. Situation where a variable value depends on a value of another variable at a specific amount of time earlier.

Diagrammatic conventions *Methodological*. Set of rules and conventions to construct a specific type of diagram in order to increase readability and comprehensibility.

Dynamic hypothesis *Methodological*. Theory or hypothesis of how the system structure may explain specific system behaviour. Some authors in the field of SD use dynamic hypothesis as synonym for the model conceptualisation, others use the dynamic hypothesis as an element of model conceptualisation. *See also* Conceptualisation.

E

Econometrics *Methodological*. Scientific discipline in economics which focusses on quantifying the relations between economic quantities by means of statistical techniques.

EMA *Methodological*. *See* Exploratory Modelling and Analysis.

EMA methods *Methodological*. Different methods which make use of EMA. Examples include Scenario Discovery, and Robust Decision Making, Scenario Discovery, RDM. *See also* EMA, RDM, Scenario Discovery.

EMA workbench *Methodological*. Software package specifically designed for EMA. The EMA workbench is agnostic about the modelling paradigm and contains libraries which, for example, enable sampling the input space, performing simulation runs for multiple modelling languages including Excel, Vensim, and

Netlogo, and analysing the results by Scenario Discovery tools . "EMA Workbench documentation" 2016); Kwakkel and Jaxa-Rozen (2016), *See also* CART, EMA, PRIM, RDM, Scenario Discovery.

Ensemble *Methodological*. An ensemble of model runs contains data for each predefined KPI for multiple runs. These runs may function as reference to test the robustness of policies in an RDM approach. The ensemble could thus be seen as the exploratory equivalent of the conventional base case. Bankes (1993) refers in this case to an ‘ensemble of models’. Bankes (1993), *See also* Base case, EMA, RDM.

Equation-based modelling *Methodological*. Any form of model development where the model is specified using mathematical equations.

Equifinality *Methodological*. The principle that the system structure allows multiple means of obtaining the same system state. Bertalanffy (1968).

ESDMA *Methodological*. *See* Exploratory System Dynamics Modelling and Analysis.

Evaluation *Methodological*. Model evaluation is a model development phase which includes performing one or more runs with a model, and analysing these runs. This phase generally includes verification and validation tests. *See also* Model development cycle, Validation, Verification.

Exemplar *Methodological*. Run which is selected to function as example of a larger set of runs with a similar behavioural mode. Islam and Pruyt (2016); Pruyt and Islam (2016). *See also* Behavioural mode.

Exploratory model *Methodological*. Model developed specifically for use in one or more EMA methods. *See also* EMA, EMA methods.

Exploratory Modelling and Analysis (EMA) *Methodological*. “Exploratory Modeling and Analysis (EMA) is a research methodology that uses computational experiments to analyse complex and uncertain systems” (Bankes et al., 2013, p. 532). Practically, this means that in each experiment a high number of samples is made over the input space to parameterise one model or multiple models in order to generate an ensemble of runs. Bankes (1993); Bankes et al. (2013). *See also* Ensemble, Input space, Latin Hypercube sampling, Monte Carlo sampling.

Exploratory System Dynamics Modelling and Analysis (ESDMA) *Methodological*. EMA making use of one or more SD models. Kwakkel and Pruyt (2015).

F

Feedback *Methodological*. Situation where a variable is at least partly dependent on its own value, either presently or in an earlier time step. *See also* Balancing feedback, Reinforcing feedback.

Fixed-stock paradigm Domain specific. Paradigm in the resource scarcity research which assumes that the use of exhaustible, non-renewable resources is ultimately limited mostly by the quantity available. It can be contrasted with the opportunity-costs paradigm. Tilton (1996). *See also* Opportunity-costs paradigm.

Floating point error *Methodological*. Error in SD models where the run is not completed due to too high or too low a value for one or more variables. This may be caused by a division by zero or unlimited exponential growth.

Flow *Methodological*. Variable type in an SD model. A flow is the only type of variable that according to SD principles may influence a stock during the run time. *See also* Auxiliary, Stock.

Forecast *Methodological*. Expectation about a future system state or development. As Lyneis (2000) puts it: "A forecast is a prediction, assumption, or viewpoint on some future event or condition, usually as basis for taking action."

Formulation *Methodological*. Model development phase where equations and values are assigned to model variables. *See also* Model development phase.

Future *Methodological*. Future system state or development. *See also* System state, Undesirable future.

G

Geopolitics *Domain specific*. Domain specific research field, focussed on the effects of geography on international relations and international politics, or how geographical space influences political power.

Grade-tonnage distribution *Domain specific*. See Ore grade distribution,

Grand challenge *Methodological*. High impact, societal problem for which no easy solution exists. Generally, these problems are complex, uncertain, and wicked or messy. *See also* Complexity, Messy problems.

Group model-building *Methodological*. Approach in SD in which models are developed by groups of around 10 people under the lead of a facilitator. The aim of group model-building is often to unite existing knowledge in a group and build consensus about the system structure. Vennix (1999). *See also* Consolidative modelling, System Dynamics.

I

Input scenario *Methodological*. Dynamic input (i.e., time series) to a model.

K

Key Performance Indicator (KPI) *Methodological*. Pre-specified model variable, representing an observable quantity in the real system, of which time series data are saved for each run. All time series of all KPIs in one run, or in set of similar runs, can be considered or classified to form an internally consistent scenario. *See also* Auxiliary, Scenario.

L

Latin Hypercube sampling (LHS) *Methodological*. Latin Hypercube sampling is a sampling method which creates a predefined number of orthogonal, semi-random samples from a multidimensional input space. McKay, Beckman, and Conover (1979).

Level *Methodological*. See Stock.

M

Mental model *Methodological*. "A mental model of a dynamic system is a relatively enduring and accessible, but limited, internal conceptual representation of an external system (historical, existing or projected) whose structure is analogous to the perceived structure of that system." (Doyle & Ford, 1999, p. 414). Craik (1943); Doyle and Ford (1998, 1999); Johnson-Laird (1983); Lane (1999). *See also* Conceptualisation, Dynamic hypothesis.

Model *Methodological*. Simplified representation of a system, specifically those parts of a system which are part of the scope. 'A model' specifically, or variations thereof, is used in this dissertation to refer to a internally connected set of equations, which is not necessarily parametrised. Practically, my use of the word model thus refers to a file made in a specific modelling language that can be simulated. *See also* Scope, System.

Model development cycle *Methodological*. Cycle which consists of all steps or phases necessary to develop and apply a model, or in an exploratory modelling context, a set of models. Generally, a model development cycle is considered to be iterative in nature, where iterations between different phases may occur, or iterations of the whole cycle.

Model development phase *Methodological*. Distinct stage in a model development cycle. Examples of model development phases include boundary selection, conceptualisation, formulation, verification and validation, evaluation, and policy testing. *See also* Model development cycle.

Model development step *Methodological*. Part of one model development phase. *See also* Model development phase.

Modelling cycle *Methodological*. See Model development cycle.

Modelling language *Methodological*. A software implementation of a modelling paradigm designed to allow model development and executing model runs.

Modelling paradigm *Methodological*. Combination of conventions and rules used to represent a system or problem in computer code. Examples mentioned in this dissertation include System Dynamics and Agent-Based Modelling. *See also* System Dynamics, Agent-Based Modelling.

Monte Carlo sampling *Methodological*. Repeated random sampling of the input space to parameterise one model, or the union of multiple parallel models. *See also* Latin Hypercube sampling.

Multivariate analysis *Methodological*. Analysis in which the effects of changing multiple variables at the same time are assessed.

N

Non-linear *Methodological*. Mathematical property that does not satisfy additivity and homogeneity requirements, or taken together as superposition principle which states that $f(ax + by) = af(x) + bf(y)$, where a and b are real, constant, scalar values.

Non-linear behaviour *Methodological*. Behaviour generated by a non-linear function or model, which implies that the behaviour does not satisfy the additivity and homogeneity properties. In contrast, linear behaviour can be represented with a function of the form $f(x) = a + bx$, where a and b are real, constant, scalar values. *See also* Non-linear.

Non-linear model *Methodological*. Model containing equations that are non-linear and representing non-linear systems. Often, these models consist of a set of connected differential or integral equations, making the whole function as a higher order differential or integral equation. As such equations cannot be solved analytically, they require numerical integration. *See also* Non-linear, Non-linear system, Numerical integration.

Numerical integration *Methodological*. Mathematical integration using numerical integration methods, like Euler or Runge-Kutta. Numerical integration is necessary for higher order differential or integral equations.

O

Open exploratory modelling *Methodological*. Exploratory modelling with the goal of finding an as broad as possible set of behavioural modes or plausible futures in the ensemble of runs. *See also* Behavioural mode, EMA, Ensemble, Exploratory modelling, Plausible future.

Opportunity-costs paradigm *Domain specific*. Paradigm in resource scarcity research which assumes that the use of exhaustible, non-renewable resources is ultimately limited by assessment of the available resource quality and related costs compared alternatives. It can be contrasted with the fixed-stock paradigm. Tilton (1996). *See also* Fixed-stock paradigm.

Ore grade distribution *Domain specific*. Relation between quality and quantity of resources in the lithosphere.

P

Paradigm *Methodological*. “A world view underlying the theories and methodology of a particular scientific subject” ("Oxford Dictionaries," 2017). *See also* Mental model.

Parallel multi-model use *Methodological*. Models used in a research design where each model has a partly overlapping scope with all other models. All models are used to generate scenarios for the same KPIs.

Parametric uncertainty *Methodological*. Uncertainty regarding an input parameter value.

Parametrisation *Methodological*. Specific set of model parameter values, or the process of creating and setting these values.

Patient Rule Induction Method (PRIM) *Methodological*. Computer algorithm (which is used in Scenario Discovery) to link a binary selection of values in the output space to a selection of the input space. Friedman and Fisher 1999, *See also* CART, Scenario Discovery.

Perspective *Methodological*. View on a system, generally dependent on the mental model. Different perspectives may affect the aggregation level.

Policy analysis *Methodological*. Research field focussing on identifying potential policy issues and proposing policies which could increase system performance or reduce the undesirability of specific system states.

Policy implementation *Methodological*. The process of operationalising a policy in either reality or a model.

Policy life cycle *Methodological*. Representation of the policy process in a number of stages. van Daalen et al. (2002).

Policy-makers *Other*. “A person responsible for or involved in formulating policies, especially in politics.” ("Oxford Dictionaries," 2017). In reality, these persons are often civil servants responsible for a particular policy domain.

Policy testing *Methodological*. Model development phase where the effects of policies are assessed on either the base case or the base ensemble. *See also* Model development phase.

Population cohort *Domain specific*. Part of a population between certain age limits.

Problem articulation *Methodological*. Model development phase where the research goal or goals are formulated, the boundary or boundaries of the research are selected, potentially a reference mode is selected, and the time horizon is defined. *See also* Boundary selection, Model development phase, Reference mode.

Q

Qualitative model *Methodological*. Purely diagrammatic representation of a system. *See also* Quantitative modelling.

Quantitative model *Methodological*. Mathematically fully specified and parameterised representation of a system. *See also* Equation-based modelling, Simulation modelling.

R

Random forests *Methodological*. Algorithm used to determine a variable's, or variables' importance in a system. Breiman (2001).

RDM *Methodological*. *See* Robust Decision Making.

Reference mode *Methodological*. Potential part of the problem articulation phase. Defined by Sterman (2000, p. 90) as “a set of graphs and other descriptive data showing the development of the problem over time”. Randers (1980b); Sterman (2000). *See also* Problem articulation.

Reinforcing feedback *Methodological.* Feedback relation involving two or more variables where a higher (lower) value for any given variable will, everything else remaining the same, result in a higher (lower) value of the same variable in the future. *See also* Balancing feedback.

Robust Decision Making (RDM) *Methodological*. EMA method where the central point consists of testing the robustness of policies on the ensemble of runs. Lempert et al. (2006). *See also* EMA, Ensemble of runs.

Robustness *Methodological*. Policy characteristic which means that a policy functions desirably in all plausible futures. *See also* RDM.

Run *Methodological*. One simulation of a model. This means that for all specified, non-constant variables new values are calculated for each time step between start and end time of the model simulation. *See also* Model.

Run time *Methodological*. Time between the simulated start time and end time of a simulation run. *See also* Run.

S

Scenario *Methodological*. A scenario can be defined in two different ways in the context of this dissertation. In the narrow definition, it is an internally consistent set of dynamics or time series data for system variables. In the broad definition, it consists of future system states or dynamics towards future system states including the context (e.g., the bandwidth of other uncertainties) of these states or dynamics. Scenarios may be used as input to models, or may be derived from model evaluation. *See also* Input scenario, Output scenario, Serial multi-model use.

Scenario Discovery *Methodological*. Research method and part of the EMA methodology which uses computer learning algorithms like PRIM and CART to identify which part of the input space is responsible for one part of a binary classification of runs. Bryant and Lempert (2010). *See also* CART, EMA, PRIM.

Scope *Methodological*. All system elements which are inside the system boundary.

SD *Methodological*. *See* System Dynamics.

SEIR model Domain specific. Equation-based model containing a sequence of stocks (i.e., ageing chain) representing Susceptible, Exposed, Infected, and Recovered parts of a population. *See also* Ageing chain.

Serial multi-model use *Methodological*. Models used in a research design where outputs or scenarios generated with one model are inputs for another model.

Societal ageing Domain specific. Situation where the average age of people in a society is increasing. This may be caused by an increasing life expectancy, or by a higher share of older people.

Societal messes *Methodological*. "[…] [D]ynamic situations that consist of complex systems of changing problems that interact with each other" (Ackoff, 1979, p. 99). Ackoff (1974). *See also* Wicked problems.

Stock *Methodological*. System variable which accounts for accumulation. *See also* Auxiliary, Flow.

Structural uncertainty *Methodological*. Uncertainty regarding the relation between multiple system elements. In exploratory modelling, this can result in having to define one or more relations between multiple variables with a switch to select one of these structures.

Sub-System Diagram (SSD) *Methodological*. Diagram focussing on the sub-systems an SD model is composed of and the relations between these sub-systems. Morecroft (1982).

Switch *Methodological*. Model parameter or categorical uncertainty designed with the single purpose of switching between plausible system structures or input scenarios. Generally, switches are intended to only have integer values. *See also* Parameter.

System *Methodological*. Internally connected set of elements forming a whole.

System behaviour *Methodological*. Behavioural mode or modes displayed by a system. *See also* Behavioural mode.

System boundary *Methodological*. Border between those elements which are considered inside (i.e., endogenous) and outside the system (i.e., exogenous). The system boundary is generally selected in the problem articulation phase and determines the scope. *See also* Problem articulation, Scope.

System Dynamics (SD) *Methodological*. Modelling paradigm and research field aimed at understanding how particular system behaviour can be explained by feedback effects, and stocks and flows in the systems. Forrester (1961); Sterman (2000).

T

Transmission model *Domain specific*. Quantitative model used in epidemiological research to either simulate infection transmission through a population, or to calculate important epidemiological characteristics like reproduction rate and mortality rate given existing data about an disease outbreak or epidemic.

U

Uncertainty location *Methodological*. Location in the policy analysis framework (e.g., system boundary, conceptual model, computer model, input data, or model implementation) where an uncertainty occurs. Kwakkel et al. (2010); Petersen (2006).

Undesirable future *Methodological*. Future state of a system which is considered undesirable from a policy perspective. *See also* Future, System state.

Univariate analysis *Methodological*. Analysis in which the effects of changing a single variable at the same time are assessed.

V

Validation *Methodological*. Narrow definition outside policy analysis: testing whether a model is a correct or true representation of reality. Use in policy analysis: testing whether a model is fit for purpose, or building confidence that the model has sufficient quality. *See also* Evaluation, Verification.

Vectorised model variables *Methodological*. Instead of having only one possible value for a model variable, vectorised model variables contain a pre-specified number of values which can be seen as a vector. In the SD modelling paradigm, these vectors are defined as subscripts.

Verification *Methodological*. Process of testing whether a system or model is modelled correctly. This can be contrasted with, or considered complementary with validation. *See also* Evaluation, Validation.

W

Wicked problems *Methodological*. According to Rittel, as quoted by Churchman (1967, p. 141), “the term ‘wicked problem’ refer[s] to that class of social system problems which are ill-formulated, where the information is confusing, where there are many clients and decision makers with conflicting values, and where the ramifications in the whole system are thoroughly confusing.” According to Rittel and Webber (1973), wicked problems have at least ten distinguishing properties (Rittel & Webber, 1973). *See also* Messy problems.

References

Ackoff, R. L. (1974). *Redesigning the Future: A Systems Approach to Societal Problems*. New York: John Wiley & Sons.

Ackoff, R. L. (1979). The Future of Operational Research is Past. *The Journal of the Operational Research Society, 30*(2), 93-104. doi:10.2307/3009290

Bankes, S. C. (1993). Exploratory Modeling for Policy Analysis. *Operations Research, 41*(3), 435-449. doi:10.1287/opre.41.3.435

Bankes, S. C., Walker, W. E., & Kwakkel, J. H. (2013). Exploratory Modeling and Analysis. In G. S. & M. C. Fu (Eds.), *Encyclopedia of Operations Research and Management Science* (3rd ed.). Berlin, Germany: Springer.

Bertalanffy, L. v. (1968). *General Systems Theory*. New York: George Braziller.

Bradfield, R., Wright, G., Burt, G., Cairns, G., & Van der Heijden, K. (2005). The origins and evolution of scenario techniques in long range business planning. *Futures, 37*(8), 795-812. doi:10.1016/j.futures.2005.01.003

Breiman, L. (2001). Random Forests. *Machine Learning, 45*, 5-32.

Breiman, L., Friedman, J. H., Stone, C., & Olshen, R. A. (1984). *Classification and Regression Trees*. Monterey, CA: CRC Press.

Bryant, B. P., & Lempert, R. J. (2010). Thinking inside the box: A participatory, computer-assisted approach to scenario discovery. *Technological Forecasting & Social Change, 77*, 34-49. doi:10.1016/j.techfore.2009.08.002

Churchman, C. W. (1967). Wicked problems. *Management Science, 14*(4), 141-142. doi:10.1287/mnsc.14.4.B141

Craik, K. J. W. (1943). *The Nature of Explanation*. Cambridge, UK: Cambridge University Press.

Doyle, J. K., & Ford, D. N. (1998). Mental models concepts for system dynamics research. *System Dynamics Review, 14*(1), 3-29. doi:10.1002/(SICI)1099-1727(199821)14:1<3::AID-SDR140>3.0.CO;2-K

Doyle, J. K., & Ford, D. N. (1999). Mental models concepts revisited: some clarifications and a reply to Lane. *System Dynamics Review, 15*(4), 411-415. doi:10.1002/(SICI)1099-1727(199924)15:4<411::AID-SDR181>3.0.CO;2-R

EMA Workbench documentation. (2016, September 17, 2016). 0.7. Retrieved from http://emaworkbench.readthedocs.io/en/latest/

Epstein, J. M., & Axtell, R. L. (1996). *Growing Artificial Societies: Social Science From the Bottom Up*. Washington, D.C.: Brookings Institution Press.

Forrester, J. W. (1961). *Industrial Dynamics*. Cambridge, MA: MIT Press.

Islam, T., & Pruyt, E. (2016). Scenario generation using adaptive sampling: The case of resource scarcity. *Environmental Modelling & Software, 79*, 285-299. doi:10.1016/j.envsoft.2015.09.014

Johnson-Laird, P. N. (1983). *Mental Models: Towards a Cognitive Science of Language, Inference, and Consciousness. Cambridge*. Cambridge, MA: Harvard University Press.

Kwakkel, J. H., & Jaxa-Rozen, M. (2016). Improving scenario discovery for handling heterogeneous uncertainties and multinomial classified outcomes. *Environmental Modelling & Software, 79*, 311-321. doi:10.1016/j.envsoft.2015.11.020

Kwakkel, J. H., & Pruyt, E. (2015). Using System Dynamics for Grand Challenges: The ESDMA Approach. *Systems Research and Behavioral Science, 32*(3), 358-375. doi:10.1002/sres.2225

Kwakkel, J. H., Walker, W. E., & Marchau, V. A. W. J. (2010). Classifying and communicating uncertainties in model-based policy analysis. *International Journal of Technology, Policy and Management, 10*(4), 299-315. doi:10.1504/IJTPM.2010.036918

Lane, D. C. (1999). Friendly amendment: a commentary on Doyle and Ford's proposed re-definition of ‘mental model’. *System Dynamics Review, 15*(2), 185-194. doi:10.1002/(SICI)1099-1727(199922)15:2<185::AID-SDR166>3.0.CO;2-C

Lempert, R. J., Bryant, B. P., & Bankes, S. C. (2008). *Comparing Algorithms for Scenario Discovery*. Retrieved from Santa Monica, CA.: http://www.rand.org/pubs/working_papers/WR557.html

Lempert, R. J., Groves, D. G., Popper, S. W., & Bankes, S. C. (2006). A General, Analytic Method for Generating Robust Strategies and Narrative Scenarios. *Management Science, 52*(4), 514-528. doi:10.1287/mnsc.1050.0472

Lempert, R. J., Popper, S. W., & Bankes, S. C. (2003). *Shaping the Next One Hundred Years: New Methods for Quantitative, Long-Term Policy Analysis* (MR-1626). Retrieved from Santa Monica, CA, USA: http://www.rand.org/pubs/monograph_reports/2007/MR1626.pdf

McKay, M. D., Beckman, R. J., & Conover, W. J. (1979). A Comparison of Three Methods for Selecting Values of Input Variables in the Analysis of Output from a Computer Code. *Technometrics, 21*(2), 239-245. Retrieved from http://www.jstor.org/stable/1268522

Morecroft, J. D. W. (1982). A Critical Review of Diagramming Tools for Conceptualizing Feedback System Models. *Dynamica, 8*(1), 20-29.

Oxford Dictionaries. (2017). Oxford: Oxford University Press.

Petersen, A. C. (2006). *Simulating Nature. A Philosophical Study of Computer-Simulation Uncertainties and Their Role in Climate Science and Policy Advice.* (PhD), Vrije Universiteit, Amsterdam. Retrieved from http://www.agci.vu.nl/en/Images/Dissertation_VU_Petersen_Simulating_Nature_tcm143-215398.pdf

Pruyt, E., & Islam, T. (2016). On generating and exploring the behavior space of complex models. *System Dynamics Review, Published online*. doi:10.1002/sdr.1544

Randers, J. (Ed.) (1980b). *Elements of the System Dynamics Method*. Cambridge, MA: Productivity Press.

Rittel, H. W. J., & Webber, M. M. (1973). Dilemmas in a General Theory of Planning. *Policy Sciences, 4*(1973), 155-169. doi:10.1007/BF01405730

Sterman, J. D. (2000). *Business Dynamics: Systems Thinking and Modeling for a Complex World*. New York: McGraw.

Tilton, J. E. (1996). Exhaustible resources and sustainable development: two different paradigms. *Resources Policy, 22*(1-2), 91-97. doi:10.1016/S0301-4207(96)00024-4

van Daalen, C. E., Dresen, L., & Janssen, M. A. (2002). The roles of computer models in the environmental policy life cycle. *Environmental Science & Policy, 5*(3), 221-231. doi:10.1016/S1462-9011(02)00040-0

Vennix, J. A. M. (1999). Group model-building: tackling messy problems. *System Dynamics Review, 15*(4), 379-401.

Simulation models are increasingly used for exploring the consequences of deep uncertainty in complex societal issues. The complexity of societal grand challenges, often characterised by the interrelatedness of different elements in the systems underlying these challenges, often renders mental simulation impossible, necessitating the use of simulation models to assist human reasoning. In addition, these grand challenges are typically also subject to deep uncertainty, making it, for example, impossible to come to a shared understanding of parts of the system and exogenous inputs to it, or even a shared problem definition.

Under deep uncertainty, simulation models can be used to explore the consequences of different combinations of assumptions about uncertain factors or attributes of the problem situation and the underlying system. This type of simulation model use was introduced in 1993 as Exploratory Modelling and Analysis (EMA). In more recent years, this approach has become a major underpinning of the Decision Making under Deep Uncertainty (DMDU) field.

The treatment of deep uncertainty in much DMDU research can be improved, however. In most DMDU research to date, pre-existing models are used. These models were generally developed for ‘consolidative’ use: the modellers tried to unify existing knowledge to come a single, ‘best’ model. While most modellers will agree that these models are not perfect representations of reality, and often agree that they as such cannot be validated in the strict sense of the word, these modellers and their models do not acknowledge deep uncertainty. The use of consolidative models is arguably problematic if one agrees that the issue at hand is characterized by deep uncertainty. Therefore, models are needed that are explicitly developed for ‘exploratory’ use: models that explicitly incorporate deep uncertainty potentially relevant for the research question or questions at hand. However, little experience and guidance exists regarding development and use of specifically exploratory models.

In this dissertation, a first attempt is made to identify, and provide guidance for, the critical choices made during the development and use of exploratory models.