Doing more with Models: Illustration of a SD Approach

for exploring deeply uncertain issues, analyzing models, and designing adaptive robust policies

Erik Pruyt*, Jan H. Kwakkel & Caner Hamarat
Delf University of Technology

July 2013

Abstract

Many grand challenges are both dynamically complex and deeply uncertain. Combining System Dynamics with Exploratory Modeling and Analysis allows one to generate, explore, identify and analyze all sorts of plausible scenarios related to such issues, and design and test adaptive policies over many scenarios. This paper explains and illustrates different uses of the resulting computational System Dynamics approach by means of an applied case, the outbreak of a new flu strand like the 2009 A(H1N1)n flu. First, we illustrate the use of this approach for generating and exploring different types of plausible pandemic shocks. Second, we illustrate the use of machine learning techniques to analyze contributions and effects of uncertainties, and discover and select scenarios. Finally, we illustrate the use of this approach for supporting the design of robust adaptive policies in order to be prepared for any new flu outbreak, especially those that really require action.

Introduction

In terms of applications, our research team addresses grand challenges and important issues which are characterized by high degrees of dynamic complexity and deep uncertainty. In this paper we illustrate how developments in sciences involved in model-based decision support can be combined with System Dynamics (SD) modeling and simulation (Forrester 1961; Sterman 2000). Combining them is useful for addressing the combined challenge of dynamic complexity and deep uncertainty through generation, exploration and analysis of many plausible scenarios[1] and through robust optimization of adaptive policies.

The remainder of this paper is structured as follows. First we define *deep* uncertainty and introduce Exploratory Modeling and Analysis for dealing with deep uncertainty, as well as its combination with SD modeling, for dealing with deeply uncertain dynamically complex issues. Then we use a single case to illustrate multiple uses of this approach, more specifically (i) open exploration, (ii) advanced analysis using machine learning algorithms, (iii) open scenario discovery and selection, (iv) directed scenario discovery and selection, (v) adaptive policy design, robust optimization and regret analysis, and (vi) model testing (verification and validation). The case used to illustrate this computational SD approach is the 'A(H1N1)n' case –the 2009-2010 flu pandemic. This case was chosen mainly for explanatory reasons: the model is relatively simple (a small SD101 simulation model[2]), the case is easily understandable (everyone is familiar with the 2009-2010 pandemic), and is used in the tutorial on our web site[3] which explains how to use our

*Corresponding author: Erik Pruyt, Delft University of Technology, Faculty of Technology, Policy and Management, Policy Analysis Section; P.O. Box 5015, 2600 GA Delft, The Netherlands – E-mail: e.pruyt@tudelft.nl

[1]In this paper we use the word 'scenario' for the time evolutionary behavior of a simulation run or computational experiment which is a combination of specific instantiations of uncertainties.

[2]This case is available in the 'small System Dynamics Models for BIG Issues' case book (Pruyt 2013) available for free at `http://simulation.tbm.tudelft.nl`.

[3]The tutorial is available at `http://simulation.tbm.tudelft.nl/ema-workbench/tutorial.html` and an older version of the EMA workbench is available for free at `http://simulation.tbm.tudelft.nl/ema-workbench/download.html`.

EMA Workbench software. Finally we draw some lessons and conclusions for this computational SD approach and, more generally, the SD field.

Dealing with Deeply Uncertain Dynamics

The audience of this paper does not require an introduction to dynamic complexity nor to SD. However, an introduction to uncertainty, deep uncertainty in particular, and Exploratory Modeling and Analysis may be useful.

Deep Uncertainty

In general, uncertainty could be defined as limited knowledge about future, past, or current events. A variety of conceptual schemes, definitions, and typologies of uncertainty have been put forward in different scientific fields (Morgan and Henrion 1990; Hoffman and Hammonds 1994; van Asselt 2000; Walker et al. 2003; Kwakkel et al. 2010b). Three such taxonomies were used by Pruyt (2007) to assess how SD deals with different types of uncertainties. Interestingly, System Dynamicists have assumed for decades that uncertainty is omnipresent and matters to such an extent that models are referred to as 'plausible' models, and SD model results are mostly interpreted in terms of general modes of behavior, not specific point or trajectory predictions or probabilistic outcomes. This stance fits well with Level 4 or deep uncertainty as defined in Table 1 adapted from (Kwakkel et al. 2010b; Kwakkel and Pruyt 2013).

Level of Uncertainty	Description	Approaches for dealing with the level
Level 1: marginal uncertainty	Recognizing that one is not absolutely certain, but that uncertainty is a marginal issue.	Performing sensitivity analyses on model parameters by changing default values with some small fraction.
Level 2: shallow uncertainty	Recognizing that uncertain is more than marginal, and being able to enumerate multiple alternatives and provide probabilities (subjective or objective)	Being able to enumerate multiple possible futures or generate alternative model outcomes, and to specify their probability of occurrence
Level 3: medium uncertainty	Being able to enumerate multiple possibilities and rank possibilities in terms of perceived likelihood. However, how much more likely or unlikely one alternative is compared to another cannot be specified	Being able to enumerate multiple possible futures or alternative model structures, and being able to judge them in terms of perceived likelihood, not in terms of probabilities
Level 4: **deep uncertainty**	**Being able to enumerate/generate multiple possibilities without being able to rank order the possibilities in terms of how likely or plausible they are**	**Being able to enumerate multiple possible futures or specify multiple alternative model structures and generate alternative outcomes, without specifying likelihoods**
Level 5: recognized ignorance	Being unable to enumerate multiple possibilities, because one does not or cannot know the generative mechanisms at play nor the possibilities that may be generated	Fully accepting the possibility of being wrong or being surprised because existing mental and formal models are known to be inadequate

Table 1: Five levels of uncertainty adapted from (Kwakkel and Pruyt 2012b)

Contrasting deep uncertainty to other levels of uncertainty, it could thus be defined as pertaining to those situations in which one could generate or enumerate multiple –several to even millions of– possibilities without being able or willing to rank order the possibilities in terms of how likely or plausible they are judged to be (Kwakkel et al. 2010b). Deep uncertainty could also be defined as pertaining to those situations in which it is not unambiguously clear which of many plausible underlying mechanisms will generate the real-world dynamics, for which it is uncertain which probabilities may be attached to plausible real-world outcomes, and for which different experts and/or policymakers may disagree about the acceptability of the outcomes (Lempert et al. 2003). In other words, models could be used to generate many plausible scenarios if a plausible model –preferably different[4] plausible models– could be specified. Deep uncertainty is of particular inter-

[4]Models are considered different if there are differences in system boundaries, conceptual models, model structures and functions, internal and external parameter values and other input data, and model implementations (e.g. different simulation methods). The word 'model' thus refers to a particular set of assumptions: two different sets of assumptions are therefore two different models.

est to analysts and policymakers dealing with grand challenges and other complex uncertain issues since most of them are indeed characterized by deep uncertainty and/or recognized ignorance.

Exploratory Modeling and Analysis

The definition and explanation of deep uncertainty already suggest that SD models could also be used differently for dealing with deeply uncertain issues – still respecting the SD stance that uncertainty is omnipresent and truly matters, and exact predictions cannot be made: plausible SD models could be used to generate many plausible scenarios which corresponds to a development in model-based decision support, namely the emergence of a different way of developing and using models (Bankes 2009), which we refer to as Exploratory Modeling and Analysis (EMA).

EMA can be useful when relevant information exists that can be exploited by building models, but where this information is insufficient for specifying a single model, i.e. a single set of assumptions. In many such circumstances, multiple models could be constructed that would be consistent with the available information. A single model run drawn from a model or a set of plausible models is then merely a computational experiment that reveals how the system would behave if the various guesses this particular model makes about the various unresolvable uncertainties were correct. Conducting a variety of such computational experiments allows one to explore the implications of various combinations of assumptions.

EMA thus refers to the explicit representation of a set of plausible models, the process of exploiting the information contained in such a set through a large number of computational experiments or very specific direct searches, the analysis of the results of these experiments, and the use of the set of robust policy design (Bankes 1993; Agusdinata 2008).

Important steps in EMA are to (i) conceptualize the decision problem and the associated uncertainties; (ii) develop an ensemble of fast and easily manageable 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. Depending on the particular application or use of EMA, different subsequent steps are possible. In case of an open exploration, aimed at identifying the diversity of dynamics implied by the models and the associated uncertainties, the next steps are (iv) to generate a series of computational experiments, (v) execute these experiments, and through various visualization techniques (vi) develop insight into the types of possible dynamics. In case of a more advanced analysis, the steps of open exploration would be followed up by (vii) defining types of dynamics or classes of outcomes that are for some reason of interest, and (vii) reveal the causes for the occurrence of these types of dynamics or classes of behavior through the application of machine learning algorithms. In case of dynamic scenario discovery (Kwakkel et al. 2013) the typical subsequent steps are to (iv) analyze the behavioral landscape resulting from (iii) through time series clustering; (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 and Lempert 2009); (vii) qualitatively or quantitatively communicate the typical scenarios 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.

Quite a different series of subsequent steps is used in case of directed search, aimed at answering targeted questions such as what is the worst that could happen? In this case, the next steps are (iv) define an objective function that encapsulates the targeted question; (v) perform non-linear optimization; (vi) translate the results from the optimization into an answer to the targeted question.

Both open exploration and directed search uses can be combined, for example for the development of robust adaptive plans or policies. Here, an iterative process is often used based on first identifying the causes for undesirable behavior, translating the resulting insight into possible solutions, and testing the solutions for their efficacy. Directed search techniques can be used to fine-tune actions, assess which combinations of actions are the most efficacious, or help in specifying weak signals that can be monitored for triggering actions only when and if they are needed as in (Hamarat et al. 2013).

The practice of EMA is still being developed. However, it has been applied to a variety of decision making problems. It has been applied to climate change problems in an effort to identify policy options that are on the one hand acceptable to a wide variety of countries, depending on their state of development and their belief about climate change, and on the other hand robust across a wide variety of different plausible future climate change developments (Lempert et al. 2003). Other EMA applications are found in the field of energy generation: Agusdinata (2008) studied how CO_2 emissions could be reduced in the Dutch household sector and (Kwakkel and Yucel 2012) in the electricity sector. A third area in which EMA has been applied is transport planning. Van Der Pas et al. (2010) report a case study related to intelligent speed limiters. EMA has also been applied to the field of airport strategic planning (Kwakkel et al. 2012). Groves and Lempert (2007) report on the use of EMA for addressing water resources management issues in California. Other case studies of the application of EMA in other fields are reported on in (Bankes and Margoliash 1993; Bankes 1994; Lempert et al. 1996; Park and Lempert 1998; Brooks et al. 1999; Bryant and Lempert 2009; Kwakkel et al. 2013). An overview of the field of EMA can be found in (Lempert 2002; Bankes et al. 2002; Bankes 2009) and most recently in (Bankes et al. 2013).

Exploratory System Dynamics Modeling and Analysis

Since EMA is appropriate for systematically exploring and analyzing deep uncertainty and testing the robustness of policies but requires models, SD is appropriate for generating plausible dynamics but requires techniques to handle deep uncertainty, and EMA and SD are philosophically similar, it follows that their combination –which we call Exploratory System Dynamics Modeling and Analysis (E*SD*MA)– is particularly useful for systematically generating, exploring, and analyzing many different plausible dynamics, and for testing the robustness of policies over all sorts of plausible dynamics. Note that ESDMA, in spite of the different label, is just another SD strand.

SD models used for E*SD*MA are easily-manageable models and consequently rather small (Ghaffarzadegan et al. 2011; Pruyt 2010), are slightly more exogenous than traditional SD models (but still largely endogenous), and contain additional SD structures for injecting different types of uncertainties like the ones discussed by Pruyt et al. (2011). Uncertainties we are currently able to deal with include: uncertainties related to initial values and parameters, functions, lookups, generative structures, model formulations, model boundaries, different models, different modeling methods and paradigms, different preferences and perspectives related to different world-views, and different policies with uncertain impacts.

Many real world E*SD*MA studies were recently performed, such as (Pruyt and Coumou 2012) and (Logtens et al. 2012; Auping et al. 2012) in health and societal aging, (Auping et al. 2012) in resource scarcity, (Kwakkel and Slinger 2012; Kwakkel and Timmermans 2012) in water security, et cetera.

We currently build these exploratory SD models in Vensim DSS (Ventana Systems Inc. 2010) and use a shell written in Python (Van Rossum 1995) to generate computational experiments that cover the space spanned by the specified uncertainties. Through our Python shell we force Vensim DSS to execute experiments (i.e. combinations of uncertainties and models) to generate transient simulation runs (scenarios). This shell is also responsible for storing the data when generated so that the ensemble and individual runs can be explored, searched, compared, used for debugging and other purposes. We then use a library of machine learning algorithms coded in Python, C, and C++ integrated in our so-called EMA Workbench to analyze the ensemble of scenarios, and visualize the most interesting findings[5]. We also use various techniques and algorithms, some

[5]Reasons for this particular choice of tooling, are (i) the ease with which different types of uncertainties can be handled and ensembles can be generated, (ii) the ease with which existing algorithms (in Python libraries) can be adapted and used, (iii) the ease with which new algorithms can be developed, tested, used, and compared to existing algorithms, (iv) the ease with which large data sets can be handled and explored, and (v) the possibility of sampling and using algorithms across multiple models, even multiple modeling methods, and policies. Note however that this computational SD approach could also be performed without our EMA Workbench by using commercial

of which have been used already in SD. Many other are new to the field. Some of these tools and techniques will be introduced and illustrated below while illustrating E*SD*MA using one case, but for different purposes. More or less the same is done in (Pruyt and Kwakkel 2012a; Kwakkel and Pruyt 2012), but then with different cases and not just SD models.

Illustration of multiple uses of E*SD*MA

The 2009-2010 A(H1N1)v Pandemic

In the first days, weeks, and months after the first reports about the outbreak of a new flu variant in Mexico and the USA, much remained unknown about the possible dynamics and consequences of this possible epidemic/pandemic of the new flu variant, first known as Swine flu or Mexican flu and known today as new influenza A(H1N1)v. Table 2 shows that, more information became available over time, but still many uncertainties remained. However, even with these uncertainties it was possible to model this flu variant, since it was flu, and flu outbreaks can be modeled.

Date	24 April	30 April	08 May	20 May	12 June	20 July	21 August
Infectivity	unknown	unknown	unknown	unknown	unknown	unknown	unknown
Ro	unknown	unknown	1–2; prob. 1.4–1.9	1–2; prob. 1.4–1.6	–	–	[R up to 2]
Immunity	unknown	unknown	indications (elderly)	idem	idem	idem	idem
Virulence	unknown	unknown	unknown	unknown	unknown	mild and self-limiting	idem
Incubation period	unknown	unknown	long tail? (up to 8d)	–	median 3-4d range 1-7d	idem	idem
CFR (Mex.)	17%?	–	4%?	2%?	0.4–1.8%?	–	–
CFR (USA)	unknown	unknown	0.1%?	0.1%?	0.2%?	0.4%?	–
CFR (UK)	unknown	unknown	unknown	unknown	unknown	0.3%(–1%)?	0.1–0.2%? (0.35% ¬ ex.)
Age distribution	unknown	unknown	older people less affected?	–	skewed tow. younger	idem	idem
Antiviral suscep.	unknown	possible	indications	–	–	–	–
% asympt.	unknown	unknown	unknown	unknown	unknown	indications	33–50% (ass.)
Future?	unknown	unknown	unknown	unknown	unknown	unknown	unknown
Source:	(ECDC 2009g)	(ECDC 2009f)	(ECDC 2009b)	(ECDC 2009c)	(ECDC 2009a)	(ECDC 2009d)	(ECDC 2009e)

Table 2: Information and unknowns provided by the ECDC from 24 April until August 21. CFR stands for Case Fatality Ratio

Many nations were at first particularly concerned about the potential loss of human life, and later –after it became clear that the case fatality ratio was moderately low– about the potentially disruptive effects both on health care systems and societies/economies at large in case large fractions of the (active) population would be immobilized simultaneously by the flu. Hence, we will, in what follows, mainly focus on the deceased population and the highest peak of the fraction of the population that is infected at a given point in time.

Use 1: Open Generation and Exploration

Open generation and exploration can be used to systematically explore plausible models under deep uncertainty. It relies on the careful design of experiments and can use techniques such as Monte Carlo sampling, Latin Hypercube sampling, or factorial methods. An open exploration can be used to answer questions such as 'What kinds of dynamics can this system exhibit? Under what circumstances would this policy possibly do well? Under what circumstances would it possibly fail?' An open exploration provides insight into the full richness of behaviors of a model or an ensemble of models. Hence, E*SD*MA is used in open exploration to generate the full richness of behaviors possible and to create insight into plausible dynamics that could occur. As such it helps to imagine many possible futures and think the unthinkable.

SD software together with advanced mathematical programs.

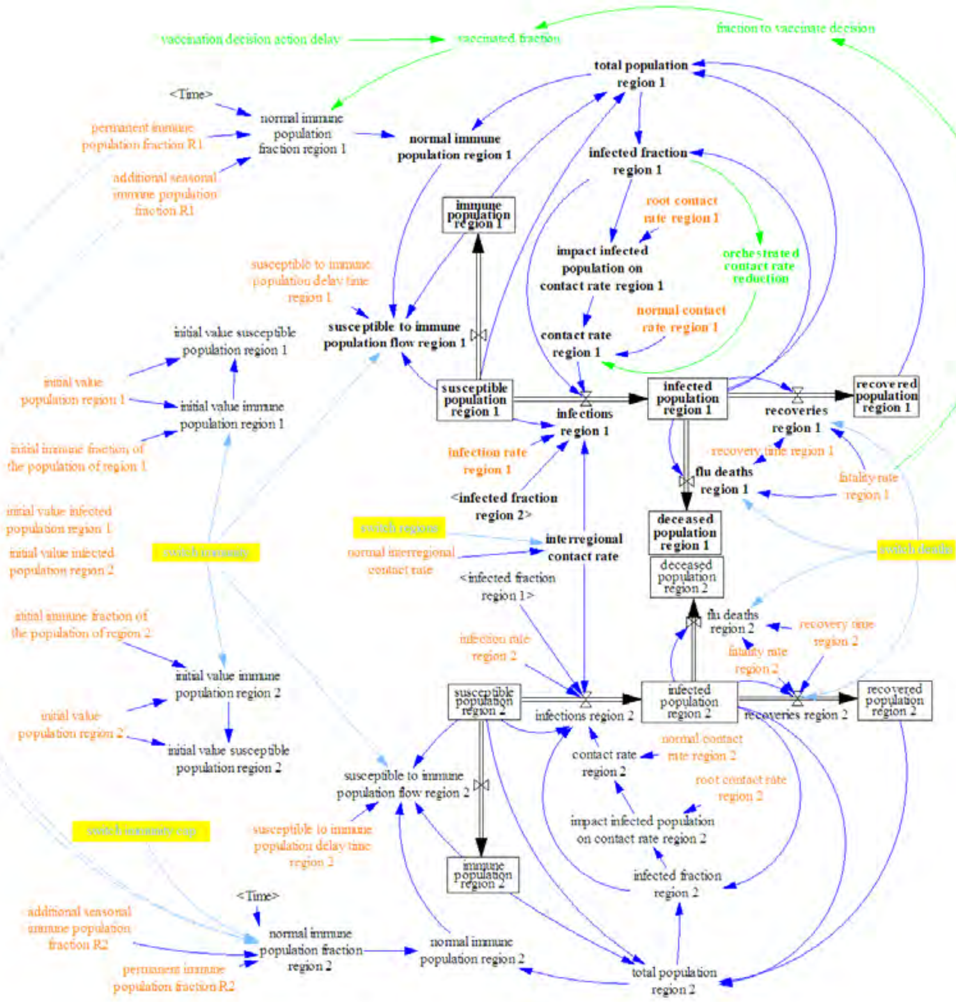


Figure 1: Stock-flow diagram of the ESD two-region flu model with the core structure for region 1 in bold (basic variables are displayed in black connected by dark blue arrows, initial values and parameters in orange, switches to turn structures of/off in light blue and yellow, and policy structures in green)

In the case of A(H1N1)v, or H1N1/09 as it is referred to today, a SD simulation model was developed shortly after the first signs of a potential outbreak were reported in order to foster understanding about the plausible dynamics of the flu outbreak (Pruyt and Hamarat 2010). The model developed at the time, displayed in Figure 1, was small, simple, high-level, data-poor (no complex/special structures nor detailed data beyond crude guestimates), and history-poor given the information in Table 2. The model was used in an ex-ante exploratory way: developments were not waited for and uncertainties were amplified and explored instead of reduced or ignored. In the model, the world is divided into two regions: the Western World, and the densely populated Developing World. For a more elaborate description of the model, see (Pruyt and Hamarat 2010). Table 3 lists the uncertainties used in combination with this flu model. These uncertainties were loosely based on the various unknowns and guestimates as reported by the European Center for Disease Control over the period of early April 2009 up to late August 2009 (Pruyt and Hamarat 2010). The ranges were set somewhat wider, given our explorative purpose and special interest in catastrophic cases. At first, the model was purposefully kept as simple as possible: the core of the model is a SIR model –not a SEIRS model as would be appropriate for a seasonal flu variant–

because of the focus on the first (pandemic) wave, and the very short incubation time (1-2 days). Later, a more refined model was developed to test whether the conclusions obtained with the 'quick and dirty' model would hold when analyzed with a more refined model[6].

Parameter	Lower Limit	Upper Limit
additional seasonal immune population fraction region 1	0.0	0.5
additional seasonal immune population fraction region 2	0.0	0.5
fatality ratio region 1	0.0001	0.1
fatality ratio region 2	0.0001	0.1
initial immune fraction of the population of region 1	0.0	0.5
initial immune fraction of the population of region 2	0.0	0.5
normal interregional contact rate	0.0	0.9
permanent immune population fraction region 1	0.0	0.5
permanent immune population fraction region 2	0.0	0.5
recovery time region 1	0.2	0.8
recovery time region 2	0.2	0.8
root contact rate region 1	1.0	10.0
root contact rate region 2	1.0	10.0
infection ratio region 1	0.0	0.15
infection ratio region 2	0.0	0.15
normal contact rate region 1	10	100
normal contact rate region 2	10	200

Table 3: Parameter ranges for the LHS with parameterised fatality ratios (0.01% – 10%) and reduced –more 'credible'– ranges for the infection ratios (0% – 10%)

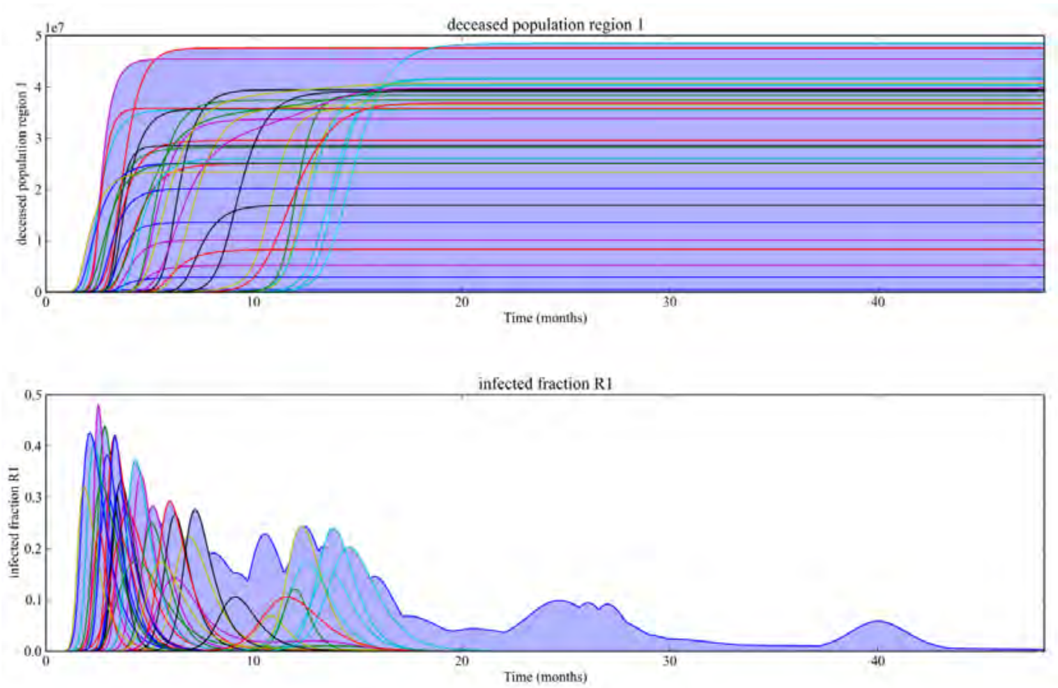
The combination of this model, these uncertainties and the Latin Hypercube sampling plan used, generates an ensemble of thousands of flu scenarios of which the envelope and the 37 worst scenarios are displayed in Figure 2a. These 37 scenarios result from selecting the 20 worst scenarios in terms of deceased population in region 1 (the Western world) and the 20 worst scenarios both in terms of infected fraction in region 1 out of the ensemble of 20000 flu scenarios with 3 scenarios being among the 20 worst in terms of deceased population and peak infected fraction in region 1. These graphs should resonate with System Dynamicists since they show the behavior over time. However, many policymakers are not familiar with behavior over time graphs and may prefer different visualizations. The 20000 runs could for example also be represented in a 3D scatter plot as in Figure 2(b). Since different types of visualizations are useful for different purposes and convey different insights, we use many different types of visualization, such as lines, envelopes, multiplots, heat maps, 3D graphs, interactive graphs, etc.

At the time, we learned from this open exploration that this flu variant could turn into anything from a small flu episode to a catastrophic pandemic, but also that the most catastrophic flu outbreaks would either take place overnight or within approximately one year. Hence, adequate adaptive social distancing measures were needed for dealing with pandemics that would happen before vaccines could be rolled out and vaccination development had to be started up without delay for high priority groups for dealing with pandemics that would happen within the year, but not necessarily for the whole population since the information available at the first signs of the flu variant were too uncertain to be used to justify 100% coverage: Delaying the vaccine stock order decision for the rest of the population in order to gain information for better decision making would have been a good idea at the time if collectively agreed upon by a large coalition of nations.

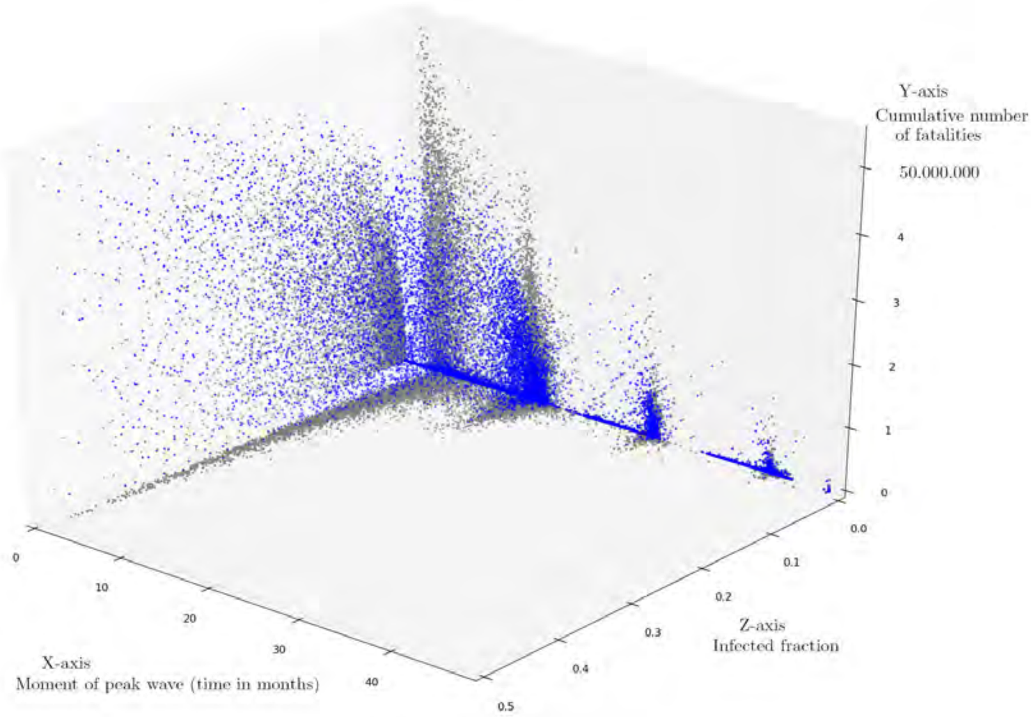
Use 2: Advanced Analyses Using Machine Learning Algorithms

To get an idea of the relative contribution of separate uncertainties with regard to diversity of outcomes, we use Random Forest (Breiman 2001) and Feature Selection (Kohavi and John 1997) algorithms. Since uncertainties do not need to be continuous parameters, it is also possible to use these techniques to explore the relative contribution of structures, loops, policies and models. Table 4 shows the ranking of uncertainties from highest contribution to the number of casualties for 20000 flu scenarios according to the random forest attribute selection and the feature selection algorithms. These relative contributions are for these uncertainties individually, not for combinations thereof.

[6]Both the simple core model and a more extended model can be accessed via (Pruyt 2013)



(a) Envelopes for the deceased population and the infected fraction in region 1 of 20000 scenarios with lines for the 37 worst scenarios in terms of deceased population and/or infected fraction



(b) 3D scatter plot with projections of the LHS 20000. X-axis: 0–48 months; Y-axis: 0–50% infected fraction; Z-axis: 0–50.000.000 fatal cases

Figure 2: Envelopes and lines versus scatter plot of the flu scenarios

This information may for example be used to remove uncertainty ranges (all those with zero and negative values for the random forest attribute selection or extremely low feature selection scores) from generation with computationally expensive techniques (factorial methods) or further analysis.

Random forest attribute selection	RF scores	Feature selection	FS scores
infection rate region 1	18.63	normal contact rate region 1	0.0463
normal contact rate region 1	14.94	fatality ratio region 1	0.0157
recovery time region 1	3.68	recovery time region 2	0.0156
fatality ratio region 1	2.19	root contact ratio region 2	0.0138
permanent immune pop. fraction R1	2.05	recovery time region 1	0.0130
root contact rate region 1	0.65	normal contact rate region 2	0.0069
add. seasonal immune pop. fraction R1	0.20	fatality rate region 2	0.0031
recovery time region 2	0.013	ini. immune fraction of the pop. of R2	0.0031
ini. immune fraction of the pop. of R2	0.01	root contact rate region 1	0.0029
infection rate region 2	0.007	add. seasonal immune pop. fraction R1	0.0024
add. seasonal immune pop. fraction R2	0.006	infection rate region 1	0.0017
normal contact rate region 2	0.004	permanent immune population fraction R1	0.0015
policy	0	normal interregional contact rate	0.0010
model	0	add. seasonal immune pop. fraction R2	0.0005
permanent immune population fraction R2	-0.001	infection rate region 2	0.0004
normal interregional contact rate	-0.002	ini. immune fraction of the pop. of R1	0.0001
fatality rate region 2	-0.002	permanent immune population fraction R2	.00002
ini. immune fraction of the pop. of R1	-0.009	policy	0
root contact ratio region 2	-0.020	model	0

Table 4: Random forest attribute selection and feature selection on 20000 flu scenarios; Policy and model have zero scores because alternatives were not included in the analyses

If the goal is to create insight into the *combinations* of assumptions that produce particular kinds of dynamics or outcomes, then methods and algorithms could be used like the Patient Rule Induction Method (PRIM) (Friedman and Fisher 1999; Lempert et al. 2006; Groves and Lempert 2007). PRIM is useful if one seeks a set of subspaces of the input variable space within which the values of output variables are considerably different from the average value or a classifier threshold over the entire input domain. In the context of this paper, the input space is the uncertainty space. PRIM then generates box-like subspaces (with the fraction of positive matches and the mass of the box relative to the total scenario space) that perform below/above a particular threshold or are characterized by particular features (e.g. acute crisis behavior). PRIM could thus be used to find subspaces in the global uncertainty space that result in highly desirable or undesirable outcomes or dynamics which makes PRIM particularly useful for discovering uncertainty subspaces with catastrophic consequences or behaviors, and identifying the corresponding root causes which allows one to develop adaptive policies consisting of specific adaptive actions for dealing with different sets of plausible futures (see below and (Hamarat et al. 2012)).

	box 1:		rest box	
PRIM box bouding uncertainties	min:	max:	min:	max:
normal contact rate region 1	**59.490**	99.986	10.003	99.986
infection ratio region 1	**0.060**	0.150	0.000	0.150
recovery time region 1	**0.282**	0.750	0.100	0.750
additional seasonal immune pop fraction R1	0.023	0.470	0.000	0.500
fatality ratio region 1	0.010	0.100	0.000	0.100
infection rate region 2	0.014	0.150	0.000	0.150
root contact rate region 1	0.012	4.693	0.014	4.998
permanent immune population fraction R1	0.000	0.478	0.000	0.500
susceptible to immune pop delay time region 1	0.561	1.999	0.501	1.999

Table 5: Prim box ranges for more than 1.5 million flu fatalities

Figure 3 shows the uncertainty space box obtained with PRIM for cases that result in more than 1.5 million deaths. This box consists of the combination of ranges of these particular uncertainties displayed in the second column of Table 5 relative to the full ranges in the last column. This PRIM box, that is box 1, covers more than 37% of all cases with more than 1.5 million deaths, and more than 97% of the runs within this box lead to more than 1.5 million deaths. The most determinant uncertainties in this box are related to uncertainties that determine the infectivity and its speed, not –surprisingly– the fatality ratio which may intuitively look like a more important determinant for the number of deaths on which basis the runs were classified. A more

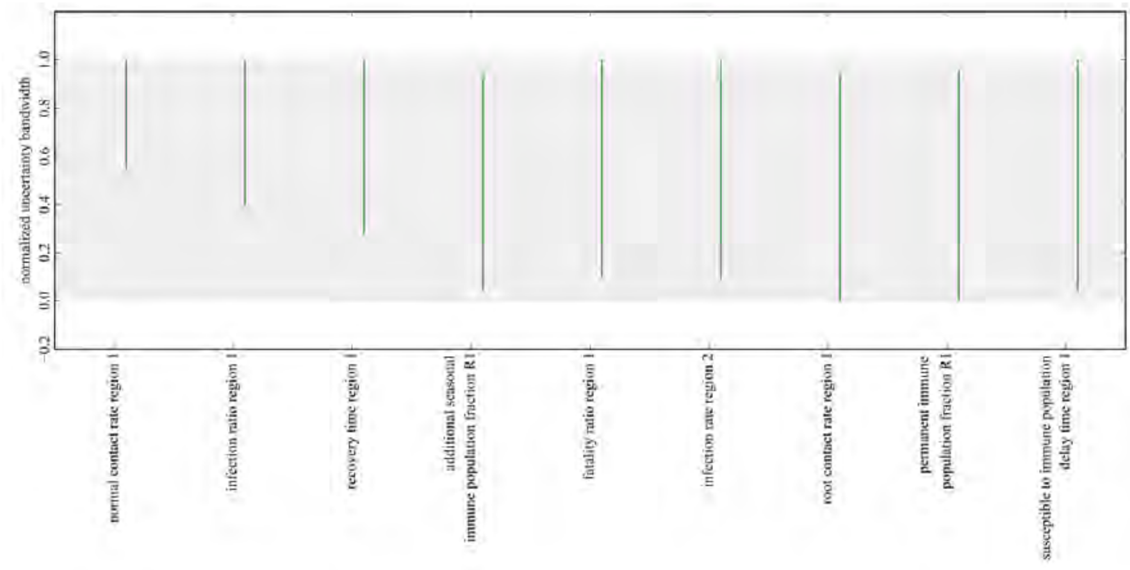


Figure 3: Prim box plot for more than 1.5 million flu deaths in region 1

sophisticated analysis, based on preprocessing the data using Principal Components Analysis as in (Kwakkel et al. 2013), can improve the coverage to 90%.

If the aim is to provide insight into the types of dynamics that could possibly occur, then results from the series of computational experiments were clustered based on the type of dynamics. This requires a form of time series clustering. The goal of clustering in general is to organize an unlabeled data set into homogenous groups where the similarity within the group is minimized and the dissimilarity between groups is maximized (Theodoridis 2003; Liao 2005). 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 (Keogh and Kasetty 2003). A relatively recent review of the state of the art in time-series clustering can be found in (Liao 2005). We currently use 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 (Liao 2005). Similarity of dynamics is determined based on an extension of the behavior pattern features discussed by Yucel and Barlas (2011) and extended further by Yucel (2012). An example of the use of this clusterer is provided in the context of open scenario discovery and selection in the next subsection.

Use 3: Open Scenario Discovery and Selection

Open Scenario Discovery and Selection refers to exploring and analyzing an ensemble generated by means of open generation to identify and select scenarios of particular interest or exemplars. Scenarios could for example be of particular interest because of their own dynamics, outcomes, and/or origin, or because of their representativeness for dynamics, outcomes, and/or origin of a subset of the ensemble. Various techniques could be used for open scenario discovery and selection. Multi-dimensional classification could be used to discover and select (representative) scenarios on multiple outcome indicators. PRIM could be used to discover and select scenarios that are representative of scenarios that share distinctive features (e.g. very undesirable outcomes) and which are highly concentrated in terms of origin in the multidimensional uncertainty space (Bryant and Lempert 2009; Kwakkel et al. 2013). And time series clustering techniques could be used to identify and select scenarios for their (representative) behavior.

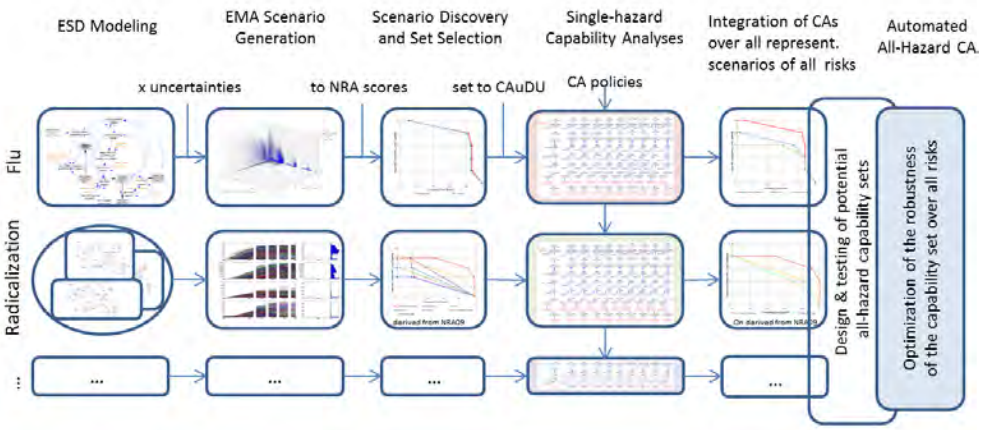


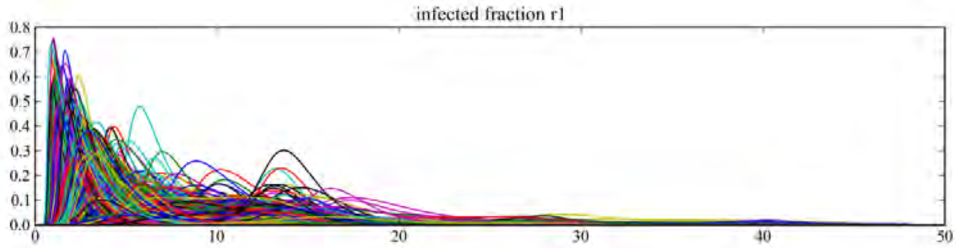
Figure 4: Integrated All-hazard Risk-Capability under deep uncertainty

Open scenario discovery and selection based on the dynamics of multidimensional effects has for example been used on an adapted version of the flu model (with rather catastrophic settings) in the context of an Integrated Risk Capability Analysis (IRCA) for the Netherlands (Pruyt et al. 2012). SD simulation models are used in the model-based IRCA described there and displayed in figure 4 to generate ensembles of thousands of plausible scenarios for each of many different risks. Next, a subset of 100 scenarios that is representative in terms of dynamics, multi-dimensional effect and origin in the multi-dimensional uncertainty space is identified and selected for subsequent use in a capability analysis (CA) model to test the effect of different capability policies under deep uncertainty for all sorts of risks. The aforementioned clusterer was applied to the total National Risk Assessment (NRA) scores (labeled 'total score' in Figure 5a) of 10000 flu scenarios in view of open scenario discovery and selection. Figure 5a shows a lines plot of the evolutions of the infected fraction and Figure 5b the total National Risk Analysis impact for 10000 plausible outbreaks of a new flu virus in the Netherlands. Total NRA impact scores above 0.33 are considered catastrophic, and impact scores between 0.11 and 0.33 are considered very serious. Note that almost all new flu scenarios that are catastrophic happen very fast, that most flu scenarios in this ensemble are very serious and happen slower or build up over time, and that a smaller subset of flu scenarios is classified as less than very serious. Using the clusterer on the total NRA score, 16 different types of behaviors were found (see Figure 5(d)). Two exemplars from each of these 16 different time-series clusters were selected and supplemented with 68 hand-picked exemplars, especially from the largest clusters (proportional to the size of the clusters), resulting in the subset of 100 scenarios displayed in Figure 5(c) selected from the larger ensemble of 1000 runs. Figure 5(e) displays a 'risk envelope diagram'[7], which could be used to plot deeply uncertain risks. This risk envelope diagram shows that the small ensemble (blue line) is indeed representative in terms of the total NRA impact scores of the entire ensemble (red line) and could therefore be used to represent the larger ensemble in the ensuing capability analysis under deep uncertainty.

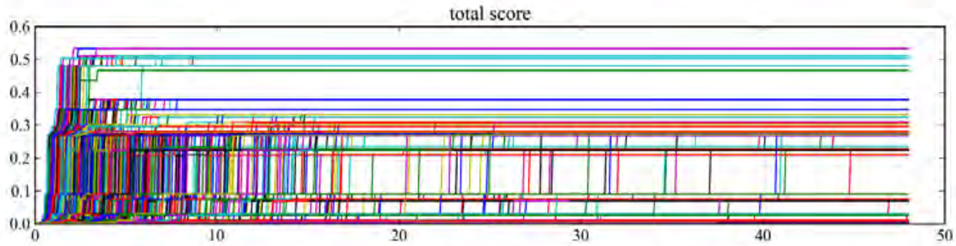
Use 4: Directed Scenario Discovery and Selection

Particular questions can be answered through directed searches. Directed search, in contrast to less refined open exploration, is a search strategy for finding particular cases that are of interest. Directed search can be used to answer questions such as: *What is the worst that could happen? What is the best that could happen? How big is the difference in performance between rival policies?*

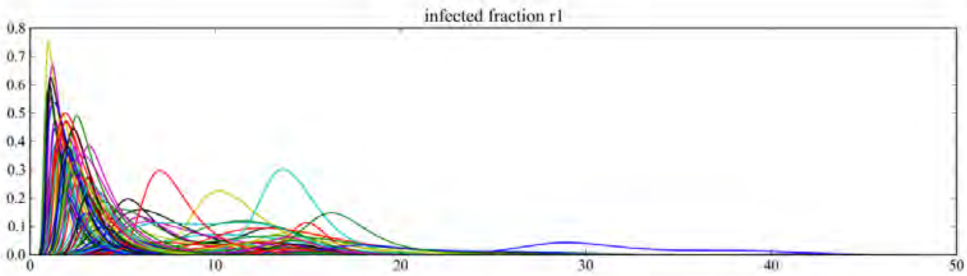
[7]A risk envelope diagram is a risk diagram in which the cumulative relative number of runs in each of the total impact classes starting with the highest impact class are plotted. In other words, 20% of the 1000 risk scenarios have a catastrophic NRA impact, about 83% of these 1000 runs have at least a very serious impact, and 98% have at least a limited impact.



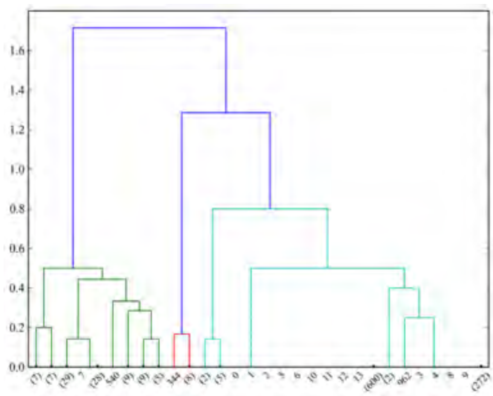
(a) Infected fraction of the Dutch population for 1000 plausible flu scenarios



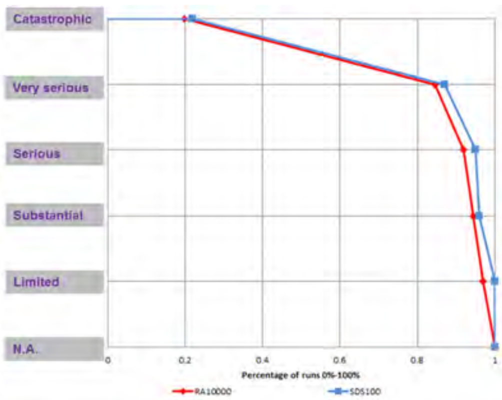
(b) Total NRA scores of the 1000 plausible flu scenarios in the framework of the Dutch National Risk Assessment (NRA)



(c) Small representative ensemble of 100 flu scenarios discovered and selected with time series clustering



(d) Time series clusterer dendrogram (flu)



(e) Risk envelopes diagram (flu)

Figure 5: Open flu scenario discovery and selection for the Dutch IRCA

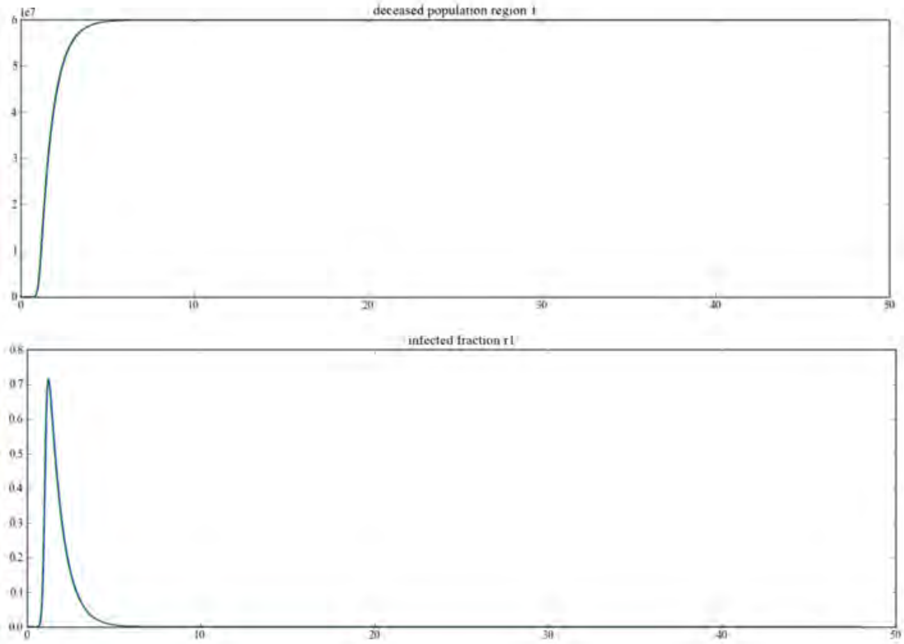


Figure 6: Two almost identical extreme flu cases obtained with two directed searches with different objective functions

A directed search provides detailed insights into the dynamics of specific regions of the full uncertainty space, not the entire uncertainty space. Directed search relies on the use of optimization techniques, such as genetic algorithms and conjugant gradient methods. Active non-linear testing is an example of a directed search strategy (Miller 1998). A suitable optimization algorithm for directed search in the context of SD should be able to cope with the non-linearity of the model, a non-linear objective function, discontinuities in the search space, a search space that is rife with local optima, and noise (Miller 1998). In the context of E*SD*MA, two additional complications are added, namely a potentially very large search space, and a discontinuous search space arising out of the inclusion of variations in e.g. structural equations. On top of this, a suitable optimization algorithm should be economical. That is, it should be able to find the optimum relatively rapidly, without requiring a very large number of computational runs. As argued by Miller (1998), Genetic Algorithms (GA) meet the outlined requirements. Open exploration and directed search can complement each other. For example, if the open exploration reveals that there are distinct types of dynamics, then directed search can be employed to identify more precisely where the boundary is located between these distinct regions.

Relevant questions in the case of A(H1N1)v are: what are the worst cases in terms of total loss of life and worst social disruption and what should be done to address those worst cases? Kwakkel and Pruyt (2013) addressed these questions with directed searches and found two worst case scenarios, displayed in Figure 6, that are almost identical. The first worst case is the maximum number of casualties ('deceased population'). The second worst case is the highest social disruption (i.e. the peak infected fraction). The socially most disruptive case is thus almost identical to the case with the highest number of casualties. In both cases, the flu spreads very quickly. Thus leaving very little time for policymakers to react, let alone leaving time for the development of vaccines. The only type of actions available to decision makers are social distancing related measures that reduce the speed with which the pandemic spreads.

Note that the two runs obtained here are worse than any of the runs in the ensemble generated with open exploration: the directed search scenarios were arrived at through optimization, whereas the open exploration ensemble was generated randomly and 'missed' the exact combination leading to these worst case scenarios. This stresses an important point: one often does not need to generate thousands of runs with 'brute computational force' to answer specific questions. Directed searches or intelligently performed open explorations are often better for answering specific questions. But answers obtained by directed searches may also be too narrow: in the case of an imminent flu pandemic, focussing only on the worst possible pandemic may not be a good idea since that requires a completely different response than in almost any other flu scenario. Note also that directed searches are often but not necessarily faster than open explorations: performing a directed search with multi-objective optimization to find the Pareto front of non-dominated worst case scenarios on multiple dimensions, instead of optimizing a single criterion to find a single worst case scenario on a single criterion as illustrated in this subsection, is computationally expensive, i.e. takes a lot of computing time.

Use 5: From Simple Adaptive to Robust Adaptive Policy Design

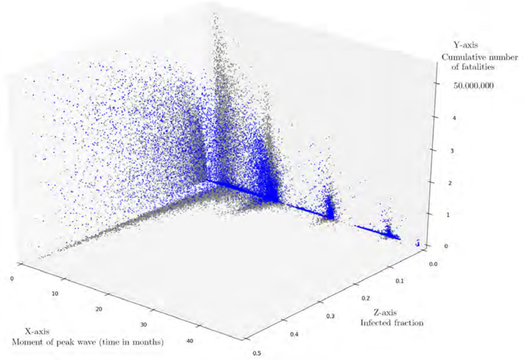
One of the most important uses is the design of effective robust policies in the presence of deep uncertainty, i.e. policies that –given the resources available at hand– lead to acceptable outcomes no matter what happens. Pruyt and Hamarat (2010) intuitively designed various adaptive policies for the flu case and tested them on the full ensemble. Figure 7 shows the effects of some of these policies starting from the no policy ensemble in Figure 7(a). Figure 7(b) shows that an adaptive vaccination policy in function of the *population fatality ratio*[8] is only effective towards the end of, and following, the vaccination campaign. This policy was therefore combined with two adaptive social distance policies that could help out before the end of the vaccination campaign. The first adaptive social distancing policy with which it is combined is based on monitoring of the infected fraction. Figure 7(c) shows that this combined adaptive policy especially helps to reduce impacts in terms of this infected fraction, but not so much the cumulative number of fatalities. The second adaptive social distancing policy with which it is combined is based on monitoring of both the fatality ratio and the infected fraction. Figure 7(d) shows that the latter combined adaptive policy –although still simplistic and solely based on intuition and trial and error testing– significantly reduces the infected fraction as well as the cumulative number of fatalities. However, many unacceptable scenarios remain present.

Hence we reflected on smarter and more refined ways to design adaptive robust policies that are conditional upon the context and robust, i.e. effective in the presence of a wide variety of uncertainties, especially when really needed. For doing so, one needs to know which subensembles of scenarios require (additional) policies, what policies are most effective given the root causes of these subensembles of scenarios, and when these policies would need to be activated. To that purpose, we developed an approach, rooted in the emerging literature on planning under deep uncertainty (Albrechts 2004; Kwakkel et al. 2010a; Walker et al. 2001), called 'Adaptive Robust Design' (Hamarat et al. 2012). A common characteristic of these approaches is the combination of time urgent actions to be taken immediately with pre-specified actions taken in response to how the future unfolds. In order to achieve a robust and adaptive policy design, it is important to correctly specify when to respond with these pre-specified actions. To this end, signposts to track specific information can be defined for monitoring the system. Specific values of these signposts are called triggers and they are triggered when pre-specified conditions occur in the system (Kwakkel et al. 2010a). We perform robust optimization[9] (Ben-Tal and Nemirovski 1998; Ben-Tal and Nemirovski 2000; Bertsimas and Sim 2004) using Genetic Algorithms (Fraser and Burnell 1970; Holland 1975) to determine these trigger values.

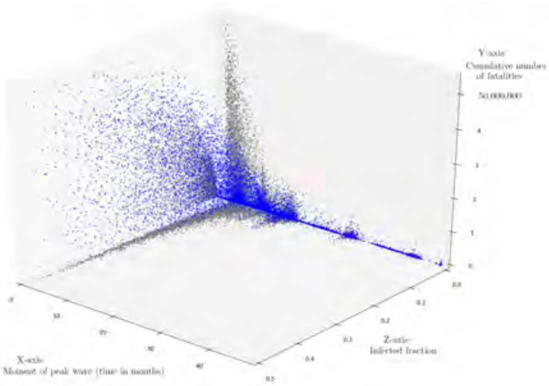
Operationally speaking, our iterative approach combines (i) open exploration to generate all sorts of plausible scenarios, (ii) identification of undesirable scenarios and their root causes (e.g. using PRIM), (iii) design of policy actions that address these root causes and triggers to activate

[8]The vaccination coverage is at least 20% –assumed to be the minimum to start up vaccine development and production. It is assumed that the coverage finally planned for will be based on information regarding the observed population fatality ratio according to pre-specified values. The low-coverage variant shown here assumes a coverage of 60% in case of a total population fatality ratio >2.5%, else the coverage is 20%.

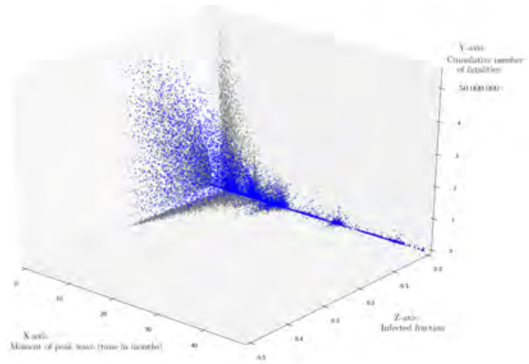
[9]Robust optimization methods aim at finding optimal outcomes in the presence of input parameter uncertainty.



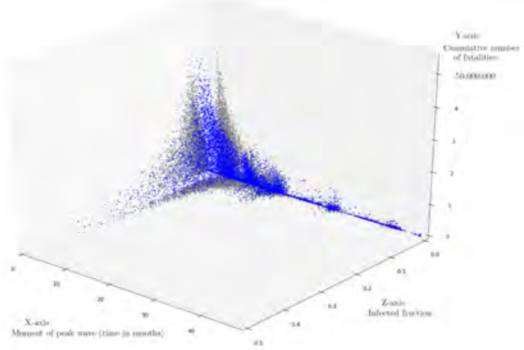
(a) No Policy



(b) Adaptive vaccination based on the fatality ratio



(c) Adaptive vaccination + social distancing based on monitoring of infected fraction



(d) Adaptive vaccination x social distancing based on monitoring of fatality ratio

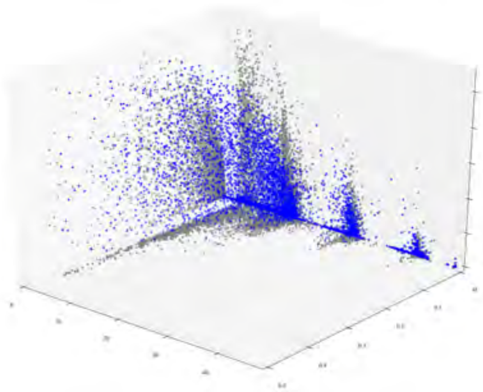
Figure 7: Effect of simple adaptive policies starting from and compared with the No Policy ensemble. Source: (Pruyt and Hamarat 2010)

these policy actions, (iv) robust or stochastic optimization of the strength of the policy actions and the trigger values, (v) testing of the robustly optimized policy on the level of the ensemble, (vi) identification of remaining undesirable futures and their causes, etc. Regret analysis as in (Lempert et al. 2003) could be used to choose one adaptive robust policy if multiple adaptive robust policies would have been developed.

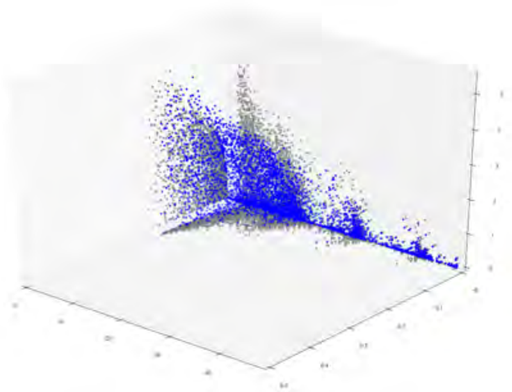
Figure 8 contains outputs generated by Hamarat et al. (2012) applying this approach to the flu case. The basic policy referred to in that figure consists of those actions that are non-regret and time-urgent and are taken from the start. In the adaptive policy two adaptive actions[10] are added to the basic policy. In the optimized adaptive policy, the triggers of these adaptive actions are determined using robust optimization. Comparing these policies and the outcomes with Figure 7c –in which case social distancing depends on the infection rate too– shows that smarter adaptive policies are better, cheaper, and less disruptive than the simple adaptive policies in Figure 8.

A major lesson learned from combining SD with stochastic optimization is that this type of optimization under uncertainty leads to much better results if the policy is adaptive (both in terms of strength and in terms of being triggered) and flexible (different actions are triggered for different circumstances).

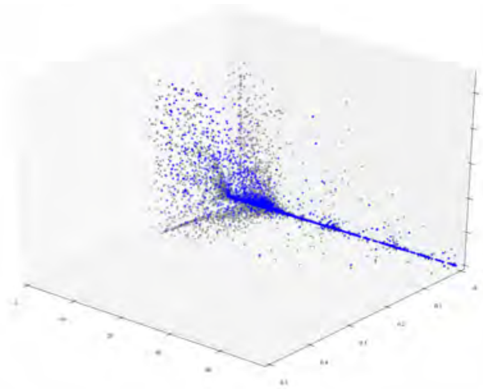
[10](i) For an observed case fatality ratio (cfr) of 0.1%, the vaccination level is increased to 60%. If the observed CFR is 1%, then vaccination level is 80% and for CFR of 10% then the vaccination level is 100%. (ii) If the rate of increase for the infected fraction is positive for three consecutive weeks, then the alert is activated and an additional 50% emergency contact rate reduction is applied.



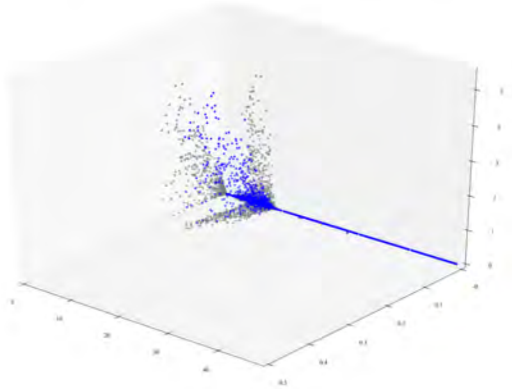
(a) No Policy



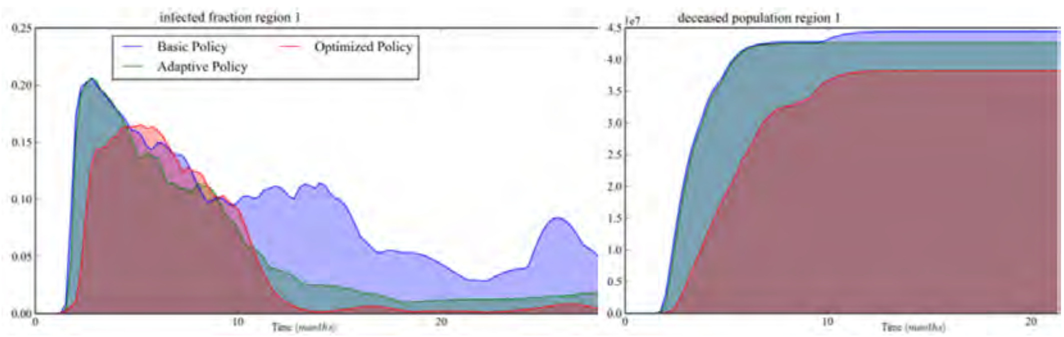
(b) Basic Policy



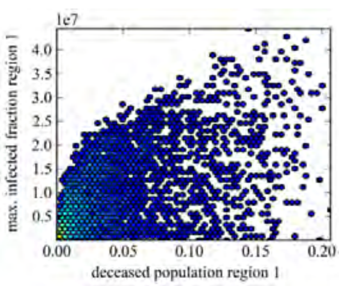
(c) Adaptive Policy



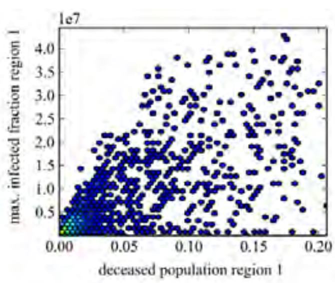
(d) Optimized Adaptive Policy



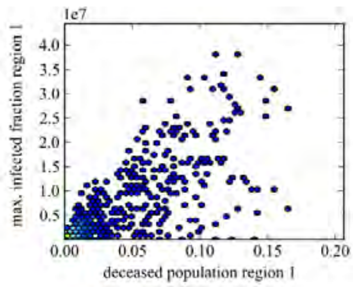
(e) Envelopes of the basic, adaptive, and optimized adaptive policy ensembles



(f) Basic Policy



(g) Adaptive Policy



(h) Optimized Adaptive Policy

Figure 8: Effect of smarter adaptive policy making – comparing the effects of a basic policy, adaptive policy, and robustly optimized adaptive policy: 3D-scatter plots, envelopes and heat maps

Use 6: Model Testing

The EMA workbench is also referred to by some as the 'EMA torture rack' since it is very useful for verification and validation of SD models too. It can be used to perform stress test and try to break models. Each run that requires debugging can be identified in 'interactive mode', singled out of the ensemble, saved[11], and instantiated directly in a Vensim model that could be used to debug that particular run, and hence, the model. Identifying subsets of results with impossible outcomes, followed by the identification of the joint causes of the impossible behavior can be used to remove subensembles or to set constraints on ranges.

The workbench allows performing comprehensive[12] automated sensitivity analysis over the entire parameter space, automated sensitivity analysis over parameterized functions, sensitivity testing with regard to delay times and orders of delays, sensitivity testing with regard to lookup functions with Hearne's method[13] (Hearne 2010; Eker et al. 2011) and other methods, as well as automated extreme condition testing. Other validation tests (like family member tests) can be performed with open exploration and directed searches. Using multiple models allows for multi-model triangulation and multi-model behavioral comparison too. And last but not least could open explorations and directed searches be used to test policy robustness. See for example (Kovari and Pruyt 2012) for policy robustness testing of what started out as a more traditional quick and dirty SD study. Semi-automated loop knock-out analysis was also implemented (Keijser et al. 2012) and other formal model analysis approaches and algorithms are next on the agenda.

Conclusions

In other scientific fields interested in model-based decision support, developments have taken place that can be summarized in terms of using models for systematic exploratory use. Exploratory Modeling and Analysis (EMA) aims at performing such systematic model-based explorations and directed searches. We have tried to explain and illustrate in this paper how SD and EMA can be combined in order to address grand societal challenges that are characterized by both dynamic complexity and deep uncertainty. Addressing such problems requires the systematic exploration of different hypotheses related to model structure, model parametrization, and input uncertainties on the kinds of behavioral dynamics that can occur, as well as directed searches. The resulting E*SD*MA approach is in fact a computational SD approach (Pruyt and Kwakkel 2012a). However, ESDMA makes SD more generally applicable, i.e. it extends the usefulness of SD from dynamically complex issues to *deeply uncertain* dynamically complex issues.

While developing this computational SD approach, we became gradually convinced that some essential SD characteristics of old, may, given the current and near future state of science and computing, become less essential for the practice of SD, whereas other characteristics and ideas may require much more emphasis and development into what they were intended for in the first place.

We believe in that respect that SD models should still be largely endogenous, but in ESDMA, it actually makes sense to 'pollute' SD models with 'open' elements (time series, shocks, et cetera) to bring in exogenous uncertainty, as well as with elements to include uncertainty in the internal functioning of models.

The original idea to resort to qualitative modes of behavior as a way of dealing with the unavoidability of uncertainty may require rethinking. Interpreting model outcomes in terms of modes of behavior may serve many goals but may not be satisfactory policymakers for dealing with real-world issues. And with today's computing power and techniques, we finally have the means to go beyond modes of behavior for dealing with ubiquitous uncertainties.

[11] not just the run, but also all parameter values, functions, and model that generated that particular run

[12] Univariate + 2 Multivariate + 3 Multivariate + . . .

[13] The basic idea of Hearne's method is multiplication of model functions by 'distortion functions' and varying the parameters of these distortion functions in order to obtain various shapes and values of the model functions. Such parameter-based generation of alternative function forms enables automated and extensive uncertainty analysis.

The same is true for the use of a reference case: base or reference cases do not exist under deep uncertainty, there are at most base ensembles, and with today's computing power and advanced visualization and analysis techniques, there is no need for a single reference case nor for a very limited number of scenarios.

Deep uncertainty also puts another question on the table: whether developing and using one plausible SD model is enough. If a model is just plausible, then maybe other plausible models need to be made as well, before robustness could be tested properly. That brings us to the ontological-epistemological stance of traditional SD and the mainstream attempts to try to integrate very different perspectives into one and the same model (Pruyt 2006; Pruyt and Kwakkel 2012a). From our point of view, it makes more sense to model different perspectives separately, design policies that are acceptable for all perspectives, and to test policies over all perspectives/models.

The illustration provided here was purposefully kept as simple as possible, and therefore single-model. Single-model E*SD*MA is exceptional though: in E*SD*MA, it is much more natural to simultaneously use multiple simulation models. Examples of multi-model ESDMA include (Pruyt and Kwakkel 2011; Pruyt and Kwakkel 2012b; Auping et al. 2012).

ESDMA is as much about analysis as it is about modeling, which is why much of our current effort is into developing analytic tools to analyze outcomes and models (via the model outcomes). The current analytic tools already generate a wealth of analytical and policy relevant insights, partially filling another gap in traditional SD, namely the lack of advanced analysis of model outcomes (and models).

We hope in that sense that this paper is also an answer to a criticism we recently received from a few respectable System Dynamicists, namely that this approach is too much a brute-force method that first needs to be refined before being of interest to the System Dynamics community. This approach is not a brute-force method in spite of the fact that more computational power needs to be relied upon than in traditional SD. Although sufficient computational power is needed, this approach should be performed intelligently and with advanced analytical tools, sophisticated techniques, refined analyses and directed searches.

ESDMA also operationalizes the concept of policy robustness by allowing testing robustness of policies and comparing policies over the entire multi-dimensional uncertainty space. Small, manageable and partly open SD models are most appropriate for ESDMA. Since ESDMA allows exploring policy robustness over the entire uncertainty/scenario space and design of adaptive policies addressing the entire scenario space, it actually makes the job of policymakers much lighter. On the one hand, that is also true for the analyst: more analytic tools are available to perform policy relevant analyses. On the other hand does it make the job of a modeler and analyst more difficult, since a larger tool set needs to be mastered, multiple plausible models may need to be developed, analyses need to be performed systematically, complex outputs need to be interpreted, and policies need to be designed iteratively and compared with all other policies.

Another complicating factor is the fact that easy software for doing all of the above is not commercially available yet. Our software, which requires python coding skills, makes it accessible and useable to a select few only. But even if user-friendly software would become available to all, then System Dynamicists also need advanced education in machine learning, data mining, time series clustering, formal modeling techniques, and the like.

However, it is our experience that performing excellent ESDMA may not be enough. Policymakers need to become part of the ESDMA and need to experience uncertainty first hand before being aligning heart and mind (Pruyt 2011). Gaming sessions as in (Pruyt 2011), new types of modeling workshops as in (Logtens et al. 2012), and embedding ESDMA in recurrent policy process with real world pilots to further the understanding of the real world system and reduce uncertainty may be necessary complements to excellent analyses and robust policy advice.

Hence, ESDMA modelers need to have more than just modeling and basic analytic skills: process skills, facilitator skills, modeling skills, programming skills, sampling skills, advanced analytic skills, communication skills are all needed – in other words, super(wo)men or well-functioning teams covering the whole skill set may be required. Hence, more team work and cooperation is needed. Everyone wanting to join our team effort to further the computational science of SD is therefore more than welcome.

References

Agusdinata, D. (2008). *Exploratory Modeling and Analysis. A Promising Method to Deal with Deep Uncertainty*. Phd dissertation, Delft University of Technology, Delft. 3, 4

Albrechts, L. (2004). Strategic (spatial) planning reexamined. *Environment and Planning B: Planning and Design 31*(5), 743–758. 14

Auping, W., E. Pruyt, and J. Kwakkel (2012, July). Analysing the uncertain future of copper with three exploratory system dynamics models. In *Proceedings of the 30th International Conference of the System Dynamics Society*, St.-Gallen, CH. International System Dynamics Society. 18

Auping, W., E. Pruyt, J. Kwakkel, G. Gijsbers, and M. Rademaker (2012). *Aging: uncertainties and solutions. an exploration of scenarios, problems and solutions with respect to the affordability of societal aging*. Number No 2012—10. The Hague: The Hague Center for Strategic Studies. link. 4

Auping, W., E. Pruyt, J. Kwakkel, and M. Rademaker (2012, June). *Futures for Copper: Exploring plausible copper price scenarios and how to act on them*. Number No 15—06—12. The Hague: The Hague Center for Strategic Studies. link to report. 4

Bankes, S. (1993, May-June). Exploratory modeling for policy analysis. *Operations Research 41*(3), 435–449. 3

Bankes, S. (1994). *Artifical Life IV*, Chapter Exploring the Foundations of Artificial Societies: Experiments in Evolving Solutions to N-player Prisoner's Dilemma. Cambridge, MA: MIT Press. 4

Bankes, S. (2009). Models as lab equipment: science from computational experiments. *Comput Math Organ Theory 15*(1), 1–8. 3, 4

Bankes, S., R. Lempert, and S. Popper (2002). Making computational social science effective. *Social Science Computer Review 20*(4), 377–388. 4

Bankes, S. and D. Margoliash (1993). Journal of neurophysiology. *Parametric Modeling of the Temporal Dynamics of Neuronal Responses Using Connectionist Architectures 69*(3), 980–991. 4

Bankes, S., W. Walker, and J. Kwakkel (2013). *Encyclopedia of Operations Research and Management Science* (3rd edn. ed.)., Chapter Exploratory Modeling and Analysis. 4

Ben-Tal, A. and A. Nemirovski (1998). Robust convex optimization. *Mathematics of Operations Research*, 769–805. 14

Ben-Tal, A. and A. Nemirovski (2000). Robust solutions of linear programming problems contaminated with uncertain data. *Mathematical Programming 88*(3), 411–424. 14

Bertsimas, D. and M. Sim (2004). The price of robustness. *Operations Research 52*(1), 35–53. 14

Breiman, L. (2001). Random forests. *Machine Learning 45*, 5–32. 10.1023/A:1010933404324. 7

Brooks, A., B. Bennet, and S. Bankes (1999). An application of exploratory analysis: The weapon mix problem. *Military Operations Research 4*(1), 67–80. 4

Bryant, B. and R. Lempert (2009). Thinking inside the box: A participatory, computer-assisted approach to scenario discovery. *Technological Forecasting & Social Change 77*(2010), 34–49. 3, 4, 10

ECDC (2009a, 12 June). ECDC interim risk assessment. `www.ecdc.europe.eu`. 5

ECDC (2009b, 8 May). ECDC risk assessment. `www.ecdc.europe.eu`. 5

ECDC (2009c, 20 May). ECDC risk assessment. `www.ecdc.europe.eu`. 5

ECDC (2009d, 20 July). ECDC risk assessment. `www.ecdc.europe.eu`. 5

ECDC (2009e, 21 August). ECDC risk assessment. `www.ecdc.europe.eu`. 5

ECDC (2009f, 30 April). ECDC threat assessment. `www.ecdc.europe.eu`. 5

ECDC (2009g, 24 April). ECDC threat assessment – update. `www.ecdc.europe.eu`. 5

Eker, S., J. Slinger, and G. Yucel (2011, July). Investigating an automated method for the sensitivity analysis of functions. In *Proceedings of the 29th International Conference of the System Dynamics Society*, Washington, USA. International System Dynamics Society. 17

Forrester, J. (1961). *Industrial Dynamics*. Cambridge, MA: MIT Press. 1

Fraser, A. and D. Burnell (1970). *Computer models in genetics*. New York: McGraw-Hill. 14

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

Ghaffarzadegan, N. et al. (2011, January–March). How small system dynamics models can help the public policy process. *System Dynamics Review 27*(1), 22–44. 4

Groves, D. G. and R. J. Lempert (2007). A new analytic method for finding policy-relevant scenarios. *Global Environmental Change 17*(1), 73 – 85. Uncertainty and Climate Change Adaptation and Mitigation. 4, 9

Hamarat, C., J. Kwakkel, and E. Pruyt (2012). Adaptive robust design under deep uncertainty. *Accepted for publication in: Technological Forecasting and Social Change*. 9, 14

Hamarat, C., J. H. Kwakkel, and E. Pruyt (2013). Adaptive robust design under deep uncertainty. *Technological Forecasting and Social Change 80*(3), 408–418. 4

Hamarat, C., E. Pruyt, and J. Kwakkel (2012, July). Adaptive policymaking under deep uncertainty: Optimal preparedness for the next pandemic. In *Proceedings of the 30th International Conference of the System Dynamics Society*, St.-Gallen, CH. International System Dynamics Society. 15

Hearne, J. (2010). An automated method for extending sensitivity analysis to model functions. *Natural Resource Modeling*. 17

Hoffman, F. and J. Hammonds (1994). Propagation of uncertainty in risk assessments: the need to distinguish between uncertainty due to lack of knowledge and uncertainty due to variability. *Risk Analysis 14*, 707–712. 2

Holland, J. (1975). *Adaptation in natural and artificial systems: an introductory analysis with applications to biology, control, and artificial intelligence*. Michigan: University of Michigan Press. 14

Keijser, B., J. Kwakkel, and E. Pruyt (2012, July). How to explore and manage the future? Formal model analysis for complex issues under deep uncertainty. In *Proceedings of the 30th International Conference of the System Dynamics Society*, St.-Gallen, CH. International System Dynamics Society. 17

Keogh, E. and S. Kasetty (2003). On the need for time series data mining benchmarks: A survey and empirical demonstration. *Data Mining and Knowledge Discovery 7*(4), 349–371. 10

Kohavi, R. and G. John (1997). Wrappers for feature subset selection. *Artificial Intelligence* ((1-2)), 273–324. 7

Kovari, A. and E. Pruyt (2012, July). Prostitution and human trafficking: A model-based exploration and policy analysis. In *Proceedings of the 30th International Conference of the System Dynamics Society*, St.-Gallen, CH. International System Dynamics Society. 17

Kwakkel, J. and E. Pruyt (2012). Using system dynamics for grand challenges: The ESDMA approach. *Systems Research and Behavioral Science (under review)*. 5

Kwakkel, J. and J. Slinger (2012, July). A system dynamics model-based exploratory analysis of salt water intrusion in coastal aquifers. In *Proceedings of the 30th International Conference of the System Dynamics Society*, St.-Gallen, CH. International System Dynamics Society. 4

Kwakkel, J. and J. Timmermans (2012, July). Exploratory system dynamics: a directed search for limits to global water use. In *Proceedings of the 30th International Conference of the System Dynamics Society*, St.-Gallen, CH. International System Dynamics Society. 4

Kwakkel, J., W. Walker, and V. Marchau (2010a). Assessing the efficacy of adaptive airport strategic planning: Results from computational experiments. *Transport Policy*. (accepted with revisions). 14

Kwakkel, J., W. Walker, and V. Marchau (2010b). Classifying and communicating uncertainties in model-based policy analysis. *International Journal of Technology, Policy and Management 10*, 299–315. 2

Kwakkel, J., W. Walker, and V. Marchau (2012). Assessing the efficacy of adaptive airport strategic planning: Results from computational experiments. *Environment and Planning B: Planning and Design 39*(3), 533–550. 4

Kwakkel, J. and G. Yucel (2012). An exploratory analysis of the dutch electricity system in transition. *Journal of the Knowledge Economy*. (in press). 4

Kwakkel, J. H., W. L. Auping, and E. Pruyt (2013). Dynamic scenario discovery under deep uncertainty: The future of copper. *Technological Forecasting and Social Change 80*(4), 789 – 800. 3, 4, 10

Kwakkel, J. H. and E. Pruyt (2013). Exploratory modeling and analysis, an approach for model-based foresight under deep uncertainty. *Technological Forecasting and Social Change 80*(3), 419 – 431. 2, 13

Lempert, R. (2002). A new decision science for complex systems. *Proceedings of the National Academy of Sciences of the United States of America 99*(3), 7309–7313. 4

Lempert, R., D. Groves, S. Popper, and S. Bankes (2006). A general, analytic method for generating robust strategies and narrative scenarios. *Management Science 52*(4), 514–528. 9

Lempert, R., S. Popper, and S. Bankes (2003). Shaping the next one hundred years: New methods for quantitative, long-term policy analysis. RAND report MR-1626, The RAND Pardee Center, Santa Monica, CA. 2, 4, 15

Lempert, R., M. Schlesinger, and S. Bankes (1996). When we don't know the cost or the benefits: Adaptive strategies for abating climate change. *Climatic Change 33*(2), 235–274. 4

Liao, T. W. (2005). Clustering of time series data - a survey. *Pattern Recognition 38*, 1857–1874. 10

Logtens, T., E. Pruyt, and G. Gijsbers (2012). Societal aging in the netherlands: Exploratory system dynamics modeling and analysis. In *Proceedings of the 30th International Conference of the System Dynamics Society*, St.-Gallen, CH. International System Dynamics Society. 4, 18

Miller, J. (1998). Active nonlinear tests (ants) of complex simulation models. *Management Science 44*, 820–830. 13

Morgan, M. and M. Henrion (1990). *Uncertainty: A Guide to Dealing with Uncertainty in Quantitative Risk and Policy Analysis*. Cambridge: Cambridge University Press. 2

Park, G. and R. Lempert (1998). The class of 2014: Preserving access to california higher education. Technical report, RAND, Santa Monica, CA. 4

Pruyt, E. (2006). What is System Dynamics? A paradigmatic inquiry. In *Proceedings of the 2006 Conference of the System Dynamics Society*, Nijmegen. System Dynamics Society. link. 18

Pruyt, E. (2007). Dealing with Uncertainties? Combining System Dynamics with Multiple Criteria Decision Analysis or with Exploratory Modelling. In *Proceedings of the 25th International Conference of the System Dynamics Society*, Boston, MA, USA. International System Dynamics Society. 2

Pruyt, E. (2010, July). Using Small Models for Big Issues: Exploratory System Dynamics for Insightful Crisis Management. In *Proceedings of the 28th International Conference of the System Dynamics Society*, Seoul, Korea. International System Dynamics Society. 4

Pruyt, E. (2011). Heart meets mind: Smart transition management to smarten energy systems in a deeply uncertain world. In *Proceedings of the 29th International Conference of the System Dynamics Society*, Washington, USA. International System Dynamics Society. 18

Pruyt, E. (2013). *Small System Dynanics Models for Big Issues: Triple Jump towards Real-World Dynamic Complexity* (first ed.). Delft: TU Delft Library. http://simulation.tbm.tudelft.nl. 1, 7

Pruyt, E. and J. Coumou (2012, July). Developing scenarios for deeply uncertain dynamically complex risks: Exploring plausible futures of lyme disease. In *Proceedings of the 30th International Conference of the System Dynamics Society*, St.-Gallen, CH. International System Dynamics Society. 4

Pruyt, E. and C. Hamarat (2010). The Influenza A(H1N1)v Pandemic: An Exploratory System Dynamics Approach. In *Proceedings of the 28th International Conference of the System Dynamics Society*, Seoul, Korea. International System Dynamics Society. 6, 14

Pruyt, E., C. Hamarat, and J. Kwakkel (2012, July). Integrated risk-capability analysis under deep uncertainty: an integrated ESDMA approach. In *Proceedings of the 30th International Conference of the System Dynamics Society*, St.-Gallen, CH. International System Dynamics Society. 11

Pruyt, E. and J. Kwakkel (2011). To radicalize or to deradicalize? Is that the question? In *Proceedings of the 29th International Conference of the System Dynamics Society*, Washington D.C., USA. International System Dynamics Society. 18

Pruyt, E. and J. Kwakkel (2012a, July). A bright future for System Dynamics: From art to computational science and more. In *Proceedings of the 30th International Conference of the System Dynamics Society*, St.-Gallen, CH. International System Dynamics Society. 5, 17, 18

Pruyt, E. and J. Kwakkel (2012b). Radicalization under deep uncertainty: a multi-model exploration of activism, extremism, and terrorism. *System Dynamics Review*. (under review). 18

Pruyt, E., J. Kwakkel, G. Yucel, and C. Hamarat (2011). Energy transitions towards sustainability: A staged exploration of complexity and deep uncertainty. In *Proceedings of the 29th International Conference of the System Dynamics Society*, Washington, USA. International System Dynamics Society. 4

Sterman, J. (2000). *Business dynamics: systems thinking and modeling for a complex world.* Irwin/McGraw-Hill: Boston. 1

Theodoridis, S. (2003). Pattern recongition. *Encyclopedia of Information Systems*, 459–479. 10

van Asselt, M. (2000). *Perspectives on uncertainty and risk: the PRIMA approach to decision support.* Dordrecht: Kluwer Academic. 2

Van Der Pas, J. et al. (2010). Exploratory mcda for handling deep uncertainties: the case of intelligent speed adaptation implementation. *Journal of Multicriteria Decision Analysis 17*(1–2), 1–23. 4

Van Rossum, G. (1995). *Python Reference Manual.* Amsterdam: CWI. 4

Ventana Systems Inc. (2010). *Vensim DSS Reference Supplement.* Ventana System Inc. 4

Walker, W. et al. (2003). Defining uncertainty: a conceptual basis for uncertainty management in model-based decision support. *Integrated Assessment 4*(1), 5–17. 2

Walker, W., S. Rahman, and J. Cave (2001). Adaptive policies, policy analysis, and policymaking. *European Journal of Operational Research 128*, 282–289. 14

Yucel, G. (2012, July). A novel way to measure (dis)similarity between model behaviors based on dynamic pattern features. In *Proceedings of the 30th International Conference of the System Dynamics Society*, St.-Gallen, CH. International System Dynamics Society. 10

Yucel, G. and Y. Barlas (2011, April/June). Automated parameter specification in dynamic feedback models based on behavior pattern features. *System Dynamics Review 27*(2), 195–215. DOI: 10.1002/sdr.457. 10