From Data-Poor to Data-Rich
System Dynamics in the Era of Big Data

Erik Pruyt et al.
August, 2014

Abstract

Although SD modeling is sometimes called theory-rich data-poor modeling, it does not mean SD modeling should per definition be data-poor. SD software packages allow one to get data from, and write simulation runs to, databases. Moreover, data is also sometimes used in SD to calibrate parameters or bootstrap parameter ranges. But more could and should be done, especially in the coming era of 'Big Data'. Big data simply refers here to more data than was until recently manageable. Big data often requires data science techniques to make it manageable and useful. There are at least three ways in which big data and data science may play a role in SD: (1) to obtain useful inputs and information from (big) data, (2) to infer plausible theories and model structures from (big) data, and (3) to analyse and interpret model-generated "brute force data". Interestingly, data science techniques that are useful for (1) may also be made useful for (3) and vice versa. There are many application domains in which the combination of SD and big data would be beneficial. Examples, some of which are elaborated here, include policy making with regard to crime fighting, infectious diseases, cyber security, national safety and security, financial stress testing, market assessment and asset management.

1 Introduction

Will 'big data' fundamentally change the world? Or is it just the latest hype? Is it of any interest to the SD modeling and simulation field? Or should it be ignored? Could it affect the field of SD modeling and simulation? And if so, how? These are some of the questions addressed in this paper. But before starting, we need to shed some light on what we mean by big data. Big data simply refers here to a situation in which more data is available than was until recently manageable. Big data often requires data science techniques to make it manageable and useful.

So far, the worlds of data science and SD modeling and simulation have hardly met. Although SD modeling is sometimes called theory-rich data-poor modeling, it does not mean SD modeling per definition is or ought to be data-poor. SD software packages allow one to get data from, and write simulation runs to, databases. Moreover, data is also sometimes used in SD to calibrate parameters or bootstrap parameter ranges. But none of these cases could be labeled big data.

However, in an era of 'big data', there may be opportunities for SD to embrace data science techniques and make use of big data. There are at least three ways in which big data and data science may play a role in SD: (1) to obtain useful inputs and information from (big) data, (2) to infer plausible theories and model structures from (big) data, and (3) to analyses and interpret model-generated "brute force data". Interestingly, data science techniques that are useful for (1) may also be made useful for (3) and vice versa.

More than just being an opportunity to be seized, adopting data science techniques may be necessary, simply because some evolutions/innovations in the SD field result in ever bigger data sets generated by simulation models. Examples of such evolutions/innovations include: spatially specific SD modeling (Ruth and Pieper, 1994; Struben, 2005; BenDor and Kaza, 2012) especially if combined with GIS packages; individual agent-based SD modeling (Castillo and Saysal, 2005; Osgood, 2009; Feola *et al.*, 2012) especially with ABM packages; hybrid ABM-SD modeling and simulation; sampling (Fiddaman, 2002; Ford, 1990); and multi-model multi-method SD modeling and simulation under deep uncertainty (Pruyt and Kwakkel, 2012, 2014; Auping *et al.*, 2012; Moorlag et al., 2014).

There are also many application domains for which the combination of SD and big data or data science would be very beneficial. Examples, some of which are elaborated here, include policy making with regard to crime fighting, infectious diseases, cybersecurity, national safety and security, financial stress testing, market assessment, asset management, and future-oriented technology assessment.

The remainder of the paper is structured as follows: First, a picture of the future of SD and big data is discussed in section 2. Approaches, methods, techniques and tools for SD and big data are subsequently discussed in section 3. Then, some of the aforementioned examples are presented in section 4. And we conclude this paper with a discussion and some conclusions.

2 Modeling and Simulation and (Model-Generated) Big Data

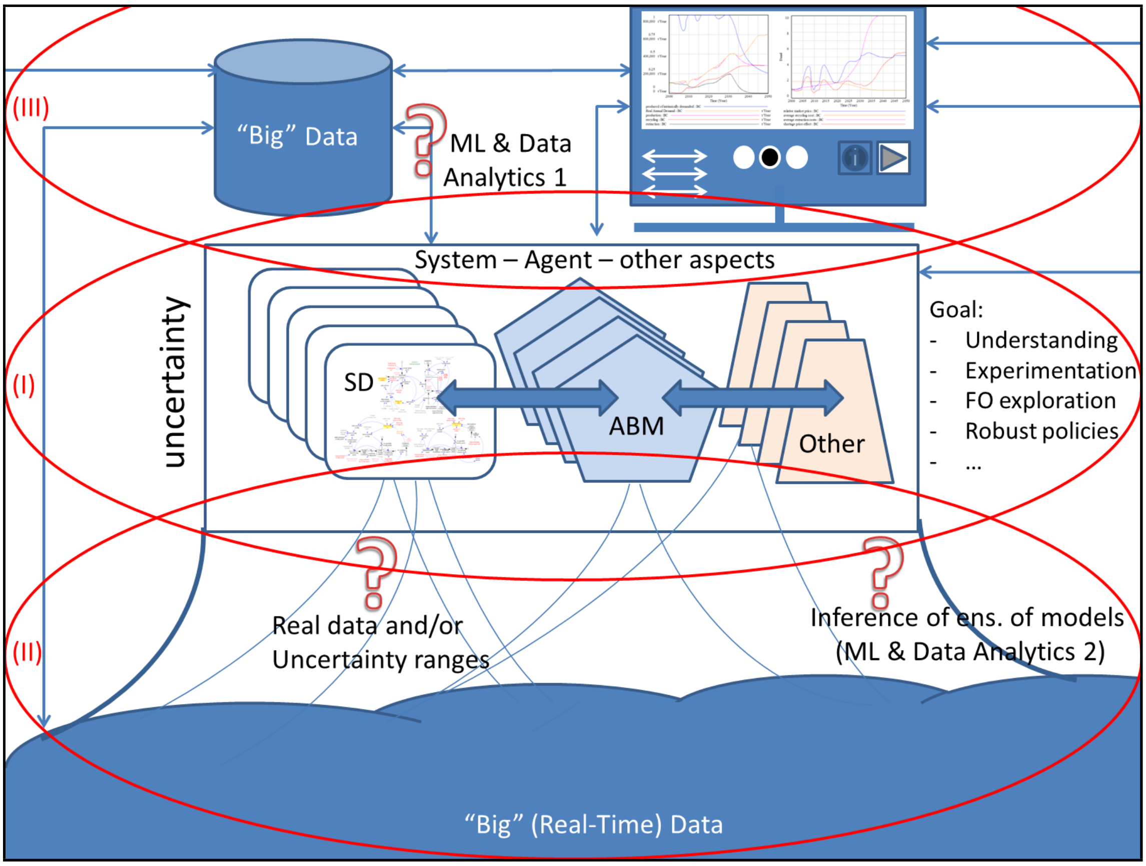


Figure 1: Picture of the near term state of science / long term state of the art

Figure 1 shows how real-world and model-based (big) data may in the future interact with modeling and simulation.

I. In the near future, it will be possible for all system dynamicists to simultaneously use multiple hypotheses, i.e. simulation models from the same or different traditions or hybrids, for different goals including the search for deeper understanding and policy insights, experimentation in a virtual laboratory, future-oriented exploration, and robust policy design and robustness testing under deep uncertainty. Sets of simulation models may be used to represent different perspectives or plausible theories, to deal with methodological uncertainty, or to deal with a plethora of important characteristics (e.g. agent characteristics, feedback and accumulation effects, spatial and network effects). This will most likely lead to more model-generated data, even big data compared to sparse simulation with a single model.

II. Some of these models may be connected to real-time or semi-real time (big) data streams, and some models may even be inferred in part from (big) data sources.

III. Storing the outputs of these simulation models in databases and applying data science techniques may enhance our understanding, may generate policy insights, and may allow to test policy robustness across large multi-dimensional uncertainty spaces.

Approaches, methods, techniques and tools that may enable system dynamics to deal with big data are discussed below, followed by a few examples.

3 Methods, Techniques and Tools for SD & Big Data

There are many data science methods, techniques and tools for dealing with (big) data. Many of them could be used to generate inputs for models or analyze model outputs. However, a comprehensive discussion of these methods, techniques and tools is beyond the scope of this paper. Below, we focus our attention on data science methods/techniques/tools for model-generated data, for generating useful model inputs, and for inferring model structures.

The fundamental differences between data science for model-generated data and data science for real data is that (i) in modeling, cases (i.e. underlying causes) are –or can be– known, (ii) in modeling, there is no missing output data, and (iii) in modeling, it is possible to generate more model-based data if more data is needed.

3.1 Approaches for dealing with big (model-based) data

3.1.1 Smartening to avoid big data

A first approach for dealing with big data –or rather not having to deal with big data– is to develop "smarter" methods, techniques and tools, i.e. methods, techniques and tools that provide more insights and deeper understanding without having to create or analyze too big a data set.

A first example is the development of 'micro-macro modeling' (Fallah-Fini *et al.*, 2013), which allows to include agent properties in SD models, and hence, to avoid ABM which is computationally more expensive and results in relative terms in much larger data sets which are also harder to make sense of.

Similar to the development of formal modeling techniques that smartened the traditional SD approach, are new methods, techniques and tools currently being developed to smarten "brute-force" SD approaches. Instead of sampling from the input space, adaptive sampling approaches are for example being developed to span the output space (Islam and Pruyt, 2014).

Another example of this approach to big data, as illustrated by Miller (1998); Kwakkel and Pruyt (2013a), is to use optimization techniques to perform directed searches as opposed to undirected "brute-force" sampling.

3.1.2 Data set reduction techniques

A second approach for dealing with big data is to use techniques to reduce the data set to manageable proportions, using for example filtering, clustering, searching, and selection methods.

Machine learning techniques like feature selection (Kohavi and John, 1997) or random forests feature selection (Breiman, 2001) could for example be used to limit the number of unimportant generative structures and uncertainties, and hence, the number of simulation runs required to adequately span the uncertainty space.

Time series classification methods like similarity-based time series clustering based on (Yücel, 2012; Yücel and Barlas, 2011) and illustrated in (Kwakkel and Pruyt, 2013a,b), K-means clustering (Islam and Pruyt, 2014), or dynamic time warping (Petitjean, 2011; Rakthanmanon, 2013; Pruyt and Islam, 2014) may for example be used to cluster similar behavior patterns, allowing the selection of a small number of exemplars spanning all behavior patterns.

Alternatively, conceptual clustering (Fisher, 1987; Michalski, 1980) may be used. Conceptual clustering is a technique developed to find natural groupings in the data, and to explain the differences between the groupings. Conceptual clustering differs from standard clustering because it gives clear, actionable rules which distinguish the different groups. While cluster analysis gives the central tendency of a group, conceptual clustering describes the limits or boundaries to the cluster.

Machine learning techniques like PRIM (Friedman and Fisher, 1999; Bryant and Lempert, 2010; Kwakkel and Pruyt, 2013b,a; Kwakkel et al., 2014) and PCA-PRIM (Kwakkel *et al.*, 2013) may be used to identify areas of interest in the multi-dimensional uncertainty space, for example areas with high concentrations of simulation runs with very undesirable dynamics.

Other machine learning techniques like t-SNE (van der Maaten and Hinton, 2008) may be used to cluster across multiple dimensions (see the example in subsection 4.3).

3.1.3 Using methods and techniques for dealing with (model-generated) big data

A third approach for dealing with big data is to use methods and techniques that are appropriate for dealing with (model-generated) big data.

Examples include optimization techniques like stochastic optimization (in Vensim), SOPS (in Powersim), or –even better– (multi-objective) robust optimization (Hamarat *et al.*, 2013, 2014), e.g. to identify policy levers and define policy triggers across large multi-dimensional uncertainty spaces: these methods actually require many simulation runs.

The same is true for methods for testing policy robustness across a wide ranges of uncertainties (Lempert *et al.*, 2003, 2000). These outcomes are usually hard to visualize, unless they are unambiguous. This is similarly the case with other 'big data approaches': needed are new ways to visualize thousands of multi-dimensional outcomes changing over time.

Machine learning techniques may possibly also be used to infer parts of models and sets of plausible models from data. A linear dynamical system (Roweis and Ghahramani, 1999) is a machine learning technique which both models data in the presence of uncertainty, and reproduces a range of dynamical patterns consistent with real-world observations. The technique is related to older approaches in dynamic modeling and control, such as Kalman filtering. The particular appeal of the technique is the potential automated production of an SD model straight from data. The model is locally linear to the data, so additional modeling insight will be necessary to translate the evidence into a full system dynamic model. The modeling approach might also be useful because it structures and parameterizes uncertainty in the data, showing a range of SD parameters which are consistent given the data.

Methods and techniques that could easily be made useful for dealing with model-based big data include Formal Model Analysis techniques (Kampmann and Oliva, 2008, 2009; Saleh *et al.*, 2010), mathematical methods (Kuipers, 2014), statistical screening techniques (Ford and Flynn, 2005; Taylor *et al.*, 2010), statistical pattern testing techniques like the ones in BTS-II.

3.2 Tools for dealing with models and (big) data

Database connectivity and data management procedures of some of the traditional software packages have been improved. For example, Vensim DSS allows to read from and write to databases via ODBC[1].

[1]See the Vensim DSS Supplement (Ventana Systems, 2010, 2011) and this video.

More powerful though is the use of scripting languages like Python (Van Rossum, 1995; McKinney, 2012) to control traditional SD software packages, manage data and perform advanced analysis on stored data. TU Delft's EMA workbench software is an open source tool in python that integrates multi-method simulation with data management, visualization and analysis[2].

Modeling and simulation across platforms is also likely to become reality in a few years time. The XMILE (eXtensible Model Interchange LanguagE) project (Eberlein and Chichakly, 2013; Diker and Allen, 2005) aims at facilitating the storage, sharing, and combination of simulation models across software packages and across (some) modeling schools. It may also allow for easier connections with databases, statistical and analytical software packages, and ICT infrastructure. Or:

> 'XMILE will enable System Dynamics models to be reused with Big Data sets to show how different policies produce different outcomes in complex environments. Models may be stored in cloud-based libraries, shared within and between organizations, and used to communicate different outcomes with common vocabulary. XMILE will also support the integration of system dynamics models and simulations into mainstream analytics software.'[3]

Note however that this is already possible to a large extent by using scripting languages or software packages with scripting capabilities such as the aforementioned EMA Workbench.

4 Examples

4.1 Crime Fighting under Deep Uncertainty

This example relates to crime fighting (see Figure 2). An exploratory SD model and related tools were developed for the police in view of increasing the effectiveness of their fight against high impact crime (HIC). Although HIC comprises robbery, mugging, burglary, intimidation, assault and grievous bodily harm, the pilot reported on in this paper focussed on fighting robbery and burglary. HIC require a systemic perspective and approach: they are characterized by important systemic effects in time and space, such as learning and specialization effects, 'waterbed effects' between different HICs and precincts, accumulations (prison time) and delays (in policing and jurisdiction), preventive effects, and other causal effects (ex post preventive measures). HICs are also characterized by deep uncertainty: a large fraction of these perpetrators is unknown –although it is known that these crimes are mainly committed by opportunistic locals, lone professionals, and mobile criminal teams– and even though their archetypal crime-related habits are known, accurate time and geographically specific predictions cannot be made. At the same time is part of the HIC system well-known and is near-real time information related to these crimes available.

The main goals of this pilot project were to support strategic policymaking under deep uncertainty and to monitor the effectiveness of policies to fight HIC. The SD model was used as an engine behind the interface, to explore effects under deep uncertainty, i.e. to generate and explore model-based big data in view of generating policy insights, and to identify real-world pilots that c/would increase the understanding about the system and effectiveness of interventions, and finally to compare the robustness of interventions across ensembles of thousands of plausible futures. Real world information and insights from the real-world pilots would then be used to improve and update the model. Today, geo-spatial real-world crime related data is available in near real-time and could be updated automatically. An evolved version of this model could thus automatically update the information, and, by doing so, increase the model's value for the strategic decision makers. An even more advanced version may use hybrid models or a multi-method approach, with more attention paid to individuals, individual objects, spatial characteristics, and networks. The latter will further expand the amount of model-generated data.

[2] The EMA workbench can be downloaded for free from, `http://simulation.tbm.tudelft.nl/ema-workbench/contents.html`.

[3] From `https://www.oasis-open.org/committees/tc_home.php?wg_abbrev=xmile` accessed on 18 March 2014.

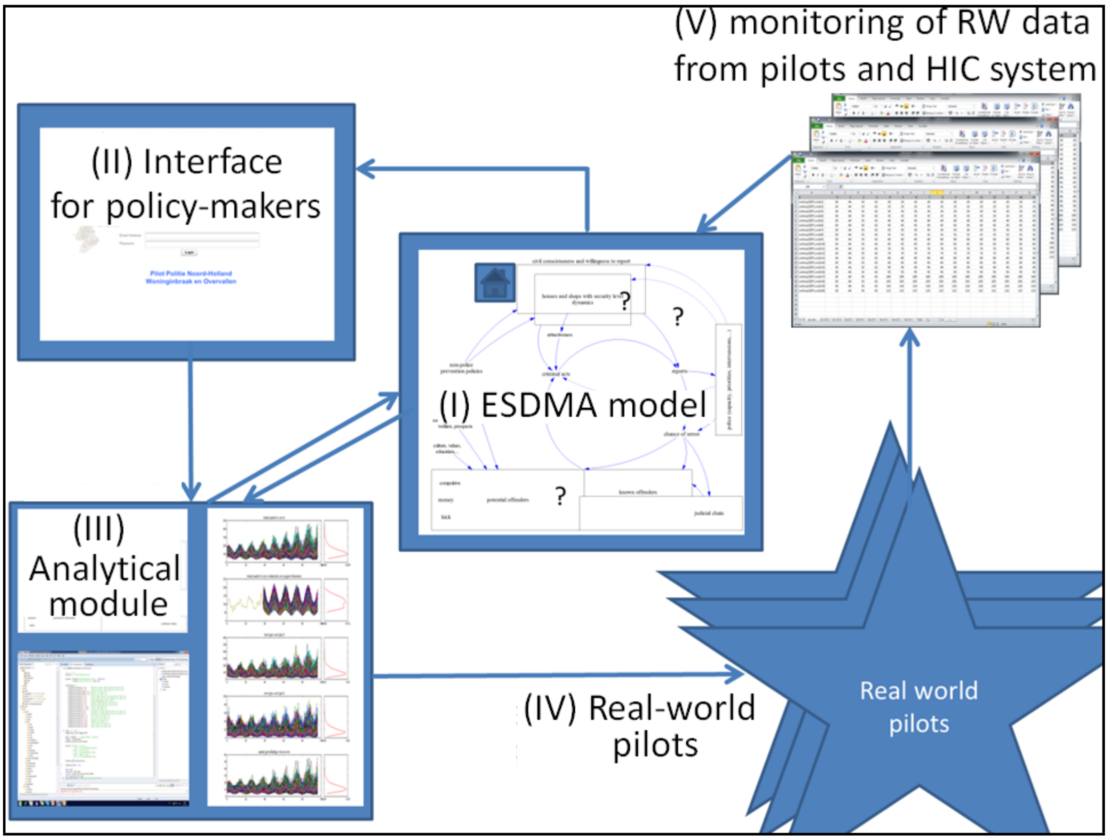


Figure 2: HIC model, interface, analyses under deep uncertainty, real-world pilots, and monitoring of real-world data

4.2 Integrated Risk-Capability Analysis under Deep Uncertainty

The third example deals with integrated risk-capability analysis under deep uncertainty for National Safety and Security as discussed by Pruyt *et al.* (2012) and displayed in Figure 3. This multi-model approach results in three instances of big data. Large ensembles of plausible futures are generated using multiple simulation models for each of a few dozen risk types. Time series clustering and machine learning techniques are used to identify and select a dozen of exemplars – exemplary in terms of outputs and origin. These exemplars are used as inputs for a capability analysis under deep uncertainty, resulting for each exemplar in hundreds to thousands of simulation runs. Again, exemplars are selected from these ensembles. Either they are used to assess the effectiveness of alternative sets of capabilities, or they are used as starting point of an automated search for robust capability sets (using robust optimization techniques). Some risk classes, settings of some of the capabilities, and exogenous uncertainties may in the future be set with real-world data.

But the real big data problem in this case is one of model-generated big data. Smart sampling techniques and time-series classification methods that together allow to identify the largest variety of behavior patterns with the minimal amount of simulations are desirable for this computational approach, for performing an automated multi-hazard capability analysis over many risks is – due to the Multi-Objective Robust Optimization method used– computationally very expensive. Compared to traditional SD, it could be labeled ‘big computing’.

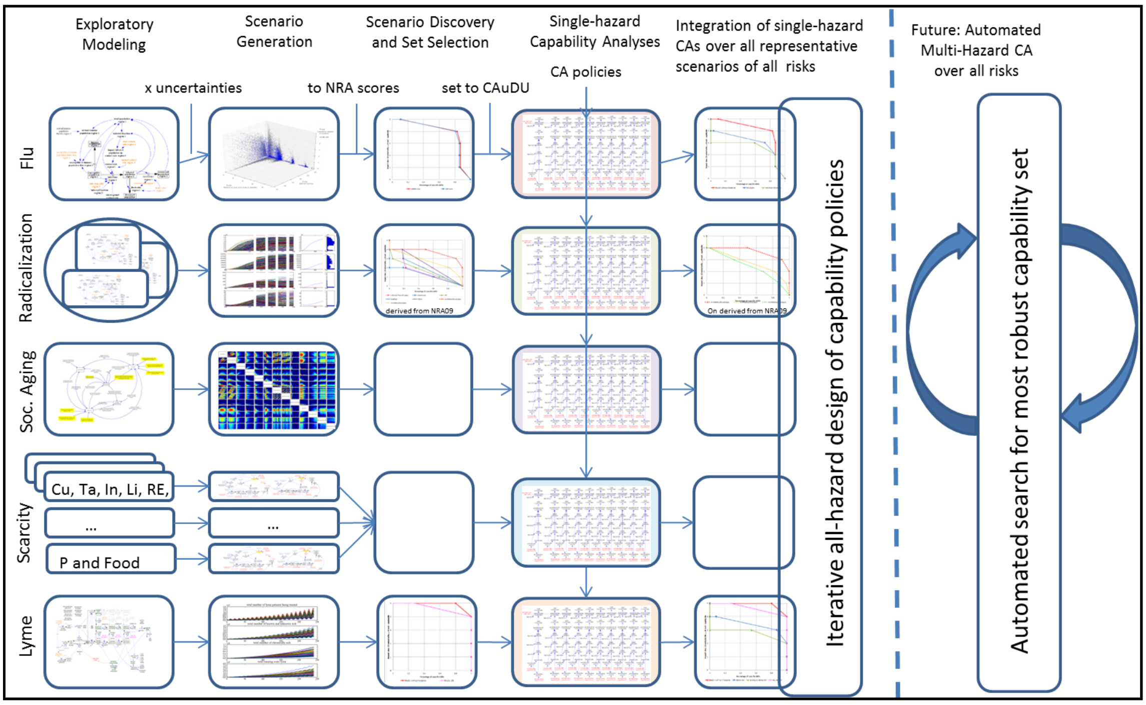


Figure 3: Model-Based Integrated Risk-Capability Analysis (IRCA)

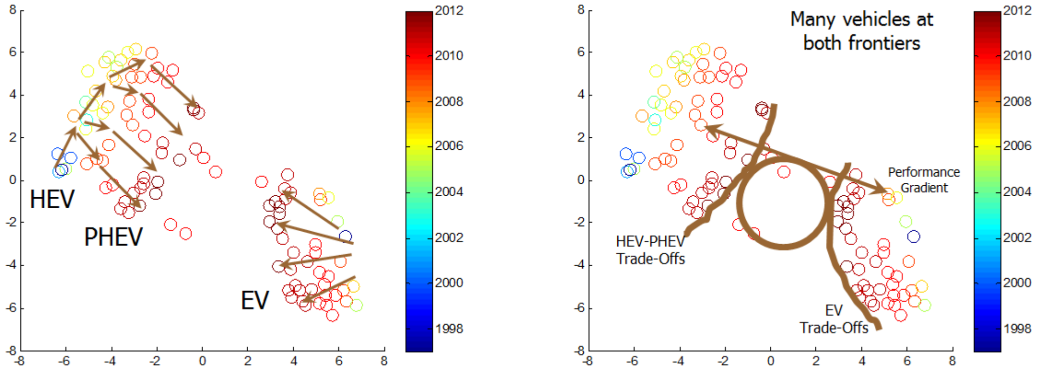
4.3 Future-Oriented Technology Assessment

Machine learning techniques could also be used to inform model/theory building. In this example, based on (Cunningham and Kwakkel, 2014), a machine learning technique called t-NSE (van der Maaten and Hinton, 2008) is used to better understand recent technological developments in Electric Vehicles, Hybrid Electric Vehicles, and Plug-in Hybrid Electric Vehicles. T-NSE is a technique for dimensionality reduction that is particularly well suited for visualizing high-dimensional datasets. It is primarily a visualization technique. Nonetheless it is representative of an emerging trend in the analysis of data as well, known as topological data analysis. Topological data analysis (Carlsson, 2009) enables researchers to capture the broad features of a data set without requiring extensive parametric knowledge of the underlying processes which may have generated the data. This is in sharp contrast to parametric statistics which may require strong assumptions about the data, and may demand an extensive base of prior hypotheses about the data. Topological data analysis approaches may be particularly suitable as a means of structuring SD outputs. These structured outputs may enhance understanding of the dynamic capability of the system, may allow comparison between seemingly very different dynamic trajectories, and may enable the identification of interesting or under-explored regimes of behavior.

Dimensions included in the analysis are: type, weight, output power, battery capacity, acceleration, CO_2 emissions, gas mileage and electric range of the vehicles. These dimensions are first rescaled so that similarities in design performance can be assessed[4]. t-SNE then reduces our 7 dimensions to 2 dimensions + time. Figure 4(a) shows the individual trajectories: HEVs have radically shifted their evolution since 2006. And Figure 4(b) shows the convergence between HEV and PHEV on the one hand and EV on the other. Note that these plots are dimensionless (except for time).

A slightly more detailed analysis of performance gradients shows that PHEV, HEV and EVs are converging: HEVs and PHEVs are on the one hand hybridizing with pure electric vehicles,

[4]The nominal variable ‘type’ is converted to a logical variable (EV or not). Gas mileage and electric range are converted to a ‘functional range’ variable. Data is ranked sorted by column, and the data replaced with its ranks. Repeated values are replaced with an average rank.



(a) Trajectories of HEV, PHEV and EV

(b) Frontiers, trade-offs, and performance gradient

Figure 4: t-SNE applied to data regarding the multi-dimensional development of Electric Vehicles, Hybrid Electric Vehicles, and Plug-in Hybrid Electric Vehicles

slowly shedding reliance on their IC engines. EVs on the other hand display eroding technological goals: they are growing more powerful and lighter in weight, able to cruise for greater ranges on a charge, but their acceleration rates are declining. Finally, CO_2 reduction is the most rapid indicator of change. This information could be used to inform model building or to compare model outcomes with real data. Note also that t-SNE can be applied to multi-dimensional outcomes of simulation models, including their trajectories.

4.4 Monitoring Infectious Diseases

The last case is about (new) infectious diseases, like the 2009 A(H1N1)n flu. This case is described and addressed with Exploratory System Dynamics Modeling and Analysis in (Pruyt and Hamarat, 2010) and used to illustrate by Pruyt *et al.* (2013) that more can be done with SD models.

If deep uncertainty is taken into account seriously, then model-based analyses ought to show many plausible futures. Over time, information will narrow down the ensemble of plausible futures. Near real-time geo-spatial data (twitter, medical records, et cetera) could be used in combination with SD models to reduce the uncertainty space. More real data then results in a reduction of the model-generated data.

5 Discussion and Conclusions

In this paper, we addressed the combination of SD modeling and simulation and data science to deal with so-called 'big data'.

Although their combination is promising, it will also require investments by the SD community. To name a few: Multi-method and hybrid modeling approaches should be further developed in order to make existing modeling and simulation approaches appropriate for dealing with agent-system characteristics, spatial and network aspects, deep uncertainty, and possibly other aspects; Data science and machine learning techniques should be further developed into techniques that can provide useful inputs for simulation models; Data science and mathematical techniques should be further developed into techniques that can be used to infer parts of models from data; easy connectors to databases and other programs need to be provided by software developers; Machine learning algorithms and formal model analysis methods should be further developed into tools that help to generate deeper understanding and generate useful policy insights; Methods and tools should be developed to turn intuitive policymaking into model-based policy design and automated robustness testing; And new analytical approaches and visualization techniques should be developed to make more sense out of models and data.

There are multiple approaches for dealing with big data: three of these approaches were discussed in the paper. Some machine learning techniques have been introduced and their application demonstrated on a few cases.

References

Auping WL, Pruyt E, Kwakkel J. 2012. 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, System Dynamics Society.

Bankes SC. 1993. Exploratory modeling for policy analysis. *Operations Research* **41**(3): 435–449.

Bankes SC. 2002. Tools and Techniques for Developing Policies for Complex and Uncertain Systems. In *Proceedings of the National Academy of Sciences of the United States of America* **99**(3): 7263–7266.

BenDor TK, Kaza N. 2012. A theory of spatial system archetypes. *System Dynamics Review* **28**(): 109–130.

Breiman L. 2001. Random Forests. Machine Learning **45**(1): 5–32

Bryant BP, Lempert RJ. 2010. Thinking inside the box: A participatory, computer-assisted approach to scenario discovery. *Technological Forecasting & Social Change* **77**(1): 34–49.

Carlsson, G. 2009. Topology and Data. *Bulletin (New Series) of the American Mathematical Society*, **46**(2): 255–303.

Castillo D, Saysel AK. 2005. Simulation of common pool resource field experiments: a behavioral model of collective action. *Ecological Economics* **55**(3): 420–436, http://dx.doi.org/10.1016/j.ecolecon.2004.12.014.

Clemson B, Tang Y, Pyne J, Unal R. 1995. Efficient methods for sensitivity analysis. *System Dynamics Review* **11**(1): 31–49.

Cunningham SW, Kwakkel JH. 2014. Technological Frontiers and Embeddings: A Visualization Approach. *Portland International Conference on Management of Engineering and Technology (PICMET)*, Kanazawa, Japan: July 27–31, 2014.

Diker VG, Allen RB. 2005. XMILE: towards an XML interchange language for system dynamics models. *System Dynamics Review* **21**(4): 351–359.

Dogan G. 2007. Bootstrapping for confidence interval estimation and hypothesis testing for parameters of system dynamics models. *System Dynamics Review* **23**(4): 415–436.

Eberlein RL, Chichakly KJ. 2013. XMILE: a new standard for system dynamics. *System Dynamics Review* **29**(): 188–195.

Feola G, Gallati JA, Binder CR. 2012. Exploring behavioural change through an agent-oriented system dynamics model: the use of personal protective equipment among pesticide applicators in Colombia. *System Dynamics Review* **28**(): 69–93.

Fallah-Fini S, Rahmandad R, Chen HJ, Wang Y. 2013. Connecting micro dynamics and population distributions in system dynamics models. *System Dynamics Review* 29(4):197–215

Fiddaman TS. 2002. Exploring policy options with a behavioral climate–economy model. *System Dynamics Review* **18**(2): 243–267.

Fisher DH. 1987. Knowledge acquisition via incremental conceptual clustering. *Machine Learning* 2 (2): 139–172.

Islam, T. and Pruyt, E. 2014. An adaptive sampling method for examining the behavioral spectrum of long-term metal scarcity. In: *Proceedings of the 30th International Conference of the System Dynamics Society*. Delft, NL, System Dynamics Society.

Michalski RS. 1980. Knowledge acquisition through conceptual clustering: A theoretical framework and an algorithm for partitioning data into conjunctive concepts. *International Journal of Policy Analysis and Information Systems* **4**: 219–244.

Roweis S, Ghahramani Z. 1999. A Unifying Review of Linear Gaussian Models. *Neural Computation* **11**, pp.305–345.

Ford A. 1990. Estimating the impact of efficiency standards on the uncertainty of the northwest electric system. *Operations Research* **38**(4):580–597

Ford, A. and Flynn, H. (2005), Statistical screening of system dynamics models. *System Dynamics Review* **21**(1): 273–303

Forrester JW. 2007. System dynamics – the next fifty years. *System Dynamics Review* **23**(2–3): 359–370.

Friedman JH, Fisher NI. 1999. Bump hunting in high-dimensional data. *Statistics and Computing* **9**(2): 123–143.

Groves DG, Lempert RJ. 2007. A new analytic method for finding policy-relevant scenarios. *Global Environmental Change* **17**(1): 73–85.

Hamarat C, Kwakkel JH, Pruyt E. 2013. Adaptive robust design under deep uncertainty. *Technological Forecasting & Social Change* **80**(3): 408–418.

Hamarat C, Kwakkel JH, Pruyt E, Loonen ET. 2014. An exploratory approach for adaptive policymaking by using multi-objective robust optimization. *Simulation Modelling Practice and Theory*. Accepted for publication.

Kampmann CE, Oliva R. 2008. Structural dominance analysis and theory building in system dynamics. Systems Research and Behavioral Science, **25**(4): 505–519.

Kampmann CE, Oliva R. 2009. Analytical methods for structural dominance analysis in system dynamics. In: R. Meyers (Ed.), *Encyclopedia of Complexity and Systems Science*, Springer, New York, p8948–8967

Kohavi R, John GH. 1997. Wrappers for feature subset selection. *Artificial Intelligence* 97(1–2), 273–324

Kuipers J. 2014. Formal Behaviour Classification under Deep Uncertainty: Applying Formal Analysis to System Dynamics. In: *Proceedings of the 30th International Conference of the System Dynamics Society*. Delft, NL, System Dynamics Society.

Kwakkel, J.H., Auping, W.L., Pruyt, E. 2014. Comparing Behavioral Dynamics Across Models: the Case of Copper. In: *Proceedings of the 30th International Conference of the System Dynamics Society*. Delft, NL, System Dynamics Society.

Kwakkel JH, Auping WL, Pruyt E. 2013. Dynamic scenario discovery under deep uncertainty: The future of copper. *Technological Forecasting & Social Change* **80**(4): 789–800.

Kwakkel JH, Pruyt E. 2013a. Using system dynamics for grand challenges: The ESDMA approach. *Systems Research & Behavioral Science*. doi: 10.1002/sres.2225

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

Lempert RJ, Groves DG, Popper SW, Bankes SC. 2006. A general, analytic method for generating robust strategies and narrative scenarios. *Management Science* **52**(4): 514–528.

Lempert RJ, Popper SW, Bankes SC. 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.

Lempert RJ, Schlesinger ME. 2000. Robust strategies for abating climate change. *Climatic Change* **45**(3–4): 387–401.

McKay MD, Beckman RJ, Conover WJ. 2000. A comparison of three methods for selecting values of input variables in the analysis of output from a computer code. *Technometrics* **42**(1): 55–61.

McKinney W. 2012. *Python for Data Analysis: Data Wrangling with Pandas, NumPy, and IPython*. O'Reilly Media, p470

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

Ruben Moorlag, Willem Auping, Erik Pruyt. 2014. Exploring the effects of shale gas development on natural gas markets: a multi-method approach. In: *Proceedings of the 32nd International Conference of the System Dynamics Society*, Delft, NL, System Dynamics Society.

Moxnes E. 2005. Policy sensitivity analysis: simple versus complex fishery models. *System Dynamics Review* **21**(2): 123–145.

Oliva R. 2003. Model calibration as a testing strategy for system dynamics models. *European Journal of Operational Research* **151**(3): p552–568.

Oreskes N, Shrader-Frechette K, Belitz K. 1994. Verification, Validation, and Confirmation of Numerical Models in the Earth Sciences. *Science* **263**(5147): 641–647.

Osgood N. 2009. Lightening the performance burden of individual-based models through dimensional analysis and scale modeling. *System Dynamics Review* **25**(): 101–134.

Petitjean FO, Ketterlin A, Gancarski P. 2011. A global averaging method for dynamic time warping, with applications to clustering. *Pattern Recognition* **44**(3): 678–693 doi:10.1016/j.patcog.2010.09.013

Pruyt E. 2013. Small system dynamics models for big issues: Triple jump towards real-world complexity. TU Delft Library, Delft. Available at `http://simulation.tbm.tudelft.nl/`.

Pruyt E, Hamarat C. 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, System Dynamics Society.

Pruyt E, Hamarat C, Kwakkel JH. 2012. 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, System Dynamics Society.

Pruyt, E. and Islam, T. 2014. The Future of Modeling and Simulation: Beyond Dynamic Complexity and the Current State of Science. In: *Proceedings of the 30th International Conference of the System Dynamics Society.* Delft, NL, System Dynamics Society.

Pruyt E, Kwakkel JH. 2014. Radicalization under Deep Uncertainty: A Multi-Model Exploration of Activism, Extremism, and Terrorism. *System Dynamics Review* **30**(1-2):1–28.

Pruyt E, Kwakkel JH. 2012. 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, System Dynamics Society.

Pruyt E, Kwakkel JH, Hamarat C. 2013. Doing more with models: Illustration of a system dynamics approach for exploring deeply uncertain issues, analyzing models, and designing adaptive robust policies. In *Proceedings of the 31st International Conference of the System Dynamics Society*. Cambridge, MA, System Dynamics Society.

Rakthanmanon T. 2013. Addressing Big Data Time Series: Mining Trillions of Time Series Subsequences Under Dynamic Time Warping. *ACM Transactions on Knowledge Discovery from Data* **7**(3): 10:1–10:31. doi:10.1145/2510000/2500489

Ruth M, Pieper F. 1994. Modeling spatial dynamics of sea-level rise in a coastal area. *System Dynamics Review* **10**(): 375–389.

Saleh M, Oliva R, Kampmann CE, Davidsen PI. 2010. A comprehensive analytical approach for policy analysis of system dynamics models. *European Journal of Operational Research* **203**(3): 673–683

Struben J. 2005. Space matters too! Mutualistic dynamics between hydrogen fuel cell vehicle demand and fueling infrastructure. In *Proceedings of the 2005 International Conference of the System Dynamics Society*. System Dynamics Society.

Taylor TRB, Ford DN, Ford A. 2010. Improving model understanding using statistical screening. *System Dynamics Review* **26**: 73–87.

Van Rossum G. 1995. *Python Reference Manual*. CWI, Amsterdam.

van der Maaten LJP, Hinton GE. 2008. Visualizing High-Dimensional Data Using t-SNE. *Journal of Machine Learning Research* **9**: 2579–2605

Ventana Systems Inc. 2010. Vensim DSS Reference Supplement. Ventana System Inc.

Ventana Systems Inc. 2011. Vensim Reference Manual. Ventana System Inc.

Weaver CP, Lempert RJ, Brown C, Hall JA, Revell D, Sarewitz D. 2013. Improving the contribution of climate model information to decision making: the value and demands of robust decision frameworks. *Wiley Interdisciplinary Reviews: Climate Change* **4**(1): 39–60.

Yücel G. 2012. 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, Switzerland, 22 July–26 July 2012.

Yücel G, Barlas Y. 2011. Automated parameter specification in dynamic feedback models based on behavior pattern features. *System Dynamics Review* **27**(2): 195–215.