



Gönenç Yücel and Els van Daalen (2009)

An Objective-Based Perspective on Assessment of Model-Supported Policy Processes

Journal of Artificial Societies and Social Simulation 12 (4) 3

<<http://jasss.soc.surrey.ac.uk/12/4/3.html>>

Received: 14-Aug-2009 Accepted: 22-Aug-2009 Published: 31-Oct-2009



Abstract

Simulation models, being in use for a long time in natural sciences and engineering domains, are diffusing to a wider context including policy analysis studies. The differences between the nature of the domain of application, as well as the increased variety of usage partially induced by this difference naturally imply new challenges to be overcome. One of these challenges is related to the assessment of the simulation-based outcomes in terms of their reliability and relevance in the policy context being studied. The importance of this assessment is twofold. First of all, it is all about conducting a high quality policy study with effective results. However, the quality of the study does not necessarily imply acceptance of the results by the clients and/or colleagues. This problem of policy analysts increases the importance of such an assessment; an effective assessment may induce the acceptance of the conclusions drawn from the study by the clients and/or colleagues. The main objective of this paper is to introduce an objective-based assessment perspective for simulation model-supported policy studies. As a first step towards such a goal, an objective-based classification of models is introduced. Based on that, we will discuss the importance of different aspects of the assessment for each type. In doing so, we aim to provide a structured discussion that may serve as a sort of methodological guideline to be used by policy analysts, and also by clients.

Keywords: Simulation, Validation, Model Assessment, Policy Analysis, Model Typology



Introduction

1.1 Simulation models have been in use for a long time in the natural sciences and engineering domains as effective research tools. More recently we are observing the diffusion of the tool to a wider context including policy analysis studies, which often focus on problems with an important social dimension. Although the approach is the same on technical grounds (i.e. using computer-based models to imitate the temporal behavior of the system of concern regarding certain aspects), the different nature of the domain of application (i.e. issues with social aspects) brings about new ways of integrating simulation models into the bigger picture of the policy analysis process (e.g. participatory modeling, interactive simulation gaming, etc.). The differences between the nature of the domain of application, as well as the

increased variety of usage partially induced by this difference naturally imply new challenges to be overcome. One of these challenges is related to the assessment of the simulation-based outcomes in terms of their reliability and relevance in the policy context being studied. On the one hand, the reliability and the relevance issues are directly related to the model^[1]. Hence, any discussion on such an assessment has a significant overlap with the issue of validation. On the other hand, it can be said that the reliability of this assessment also corresponds to the way the model is developed, and the relevance aspect more to the way it is used; hence assessment of the whole life cycle of a model is relevant.

- 1.2 The importance of this assessment is twofold. First of all, it is all about conducting a high quality policy study with effective results. However, the quality of the study does not necessarily imply acceptance of the results by the clients and/or colleagues. This problem of policy analysts increases the importance of such an assessment; an effective assessment may induce the acceptance of the conclusions drawn from the study by the clients and/or colleagues. The relevance of an assessment process is not a novel proposition at all, and we don't think that further justification is needed for its importance. Hence, the main discussion of this paper is not about the importance of such an assessment, but it is more related to the way that assessment should be conducted; which aspects are important to be assessed? Where to look in order to build up confidence regarding the reliability of the model and the results obtained by it? In line with this, we will discuss some important aspects of such an assessment process in general, and lay down a set of guidelines for model assessment that can be used by policy analysts as well as clients in this paper.
- 1.3 However, the task is not straightforward. The problem is that it does not seem effective, and also viable, to come up with general answer to the kind of questions phrased above facing the variety of simulation model usage in the policy context. The simplest initial response to such questions would be something like; "depends on the type of simulation-model being considered." Therefore, the primary requirement for making an effective discussion on what is important in terms of assessing the reliability and relevance of model-supported outcomes is a classification of models that may be used for policy analysis. This classification will enable us to focus on classes of models that are more homogenous in terms of their assessment requirements. For that reason, a typology of policy models is developed using the objective for which the model is used as the criterion of classification. Since the objective determines what is relevant and important to a large extent, it is evaluated to be the most appropriate criterion for our purposes.
- 1.4 To sum up, the main objective of this paper is to introduce an objective-based assessment perspective for simulation model-supported policy studies. As a first step towards such a goal, a classification of models to support policy-making based on their objectives is discussed in the following section. Following that, we will give a structured overview of several aspects of assessment that are relevant in general. The section following that will focus on matching the model types in the classification with different aspects of the assessment issue. In doing so, we aim to provide a structured discussion that may serve as a sort of methodological guideline to be used by policy analysts, and also by clients.

An Objective-based Classification of Models in Support of Policy Studies

- 2.1 The preliminary step for a classification of models according to the objectives they are used for is naturally identification of this set of objectives. However, the policy analysis domain covers a wide range of activities conducted with differing primary objectives and perspectives, which makes this preliminary step a challenge by itself. Taking on the challenge, Mayer et al.

(2004) introduce a conceptual framework, i.e. the hexagon framework, that classifies the policy analysis activities in a structured manner. Although the framework focuses on policy analysis activities in general, and does not say anything regarding the tools used (e.g. quantitative models), it constitutes a very appropriate frame to investigate the models used in these activities. Using this frame as the starting point, we will introduce a set of model types used in policy analysis activities, and discuss their characteristics, which mainly differ due to the different objectives these models types serve. In other words, we will map the policy analysis activities classification of Mayer et al. (2004) onto the domain of quantitative models to introduce an objective-based classification of models used in policy analysis processes.

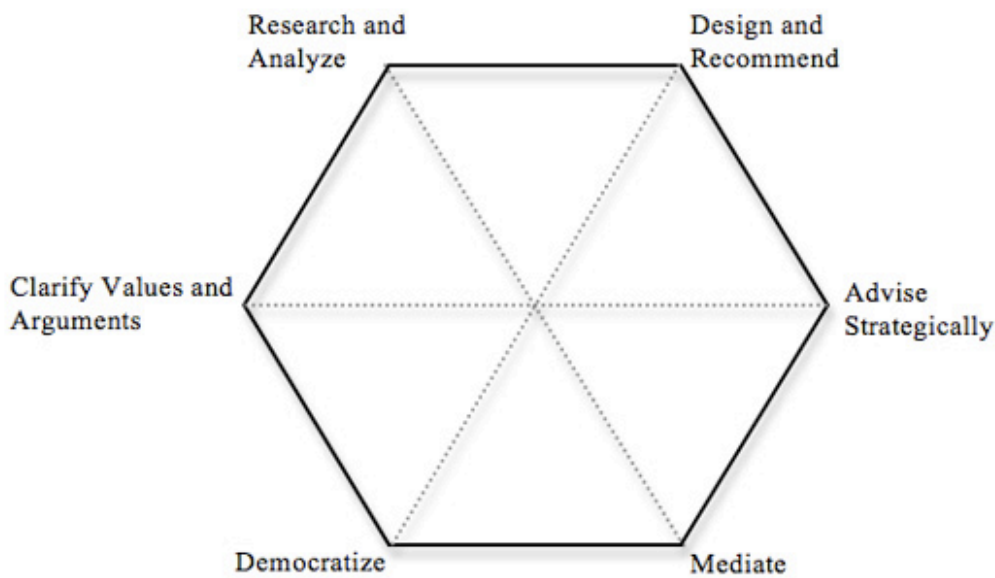


Figure 1 Overview of activities of policy analysis (after Mayer et al. 2004)

2.2 According to the hexagon framework, an analyst providing policy support may carry out six major clusters of activities. These six types of activities are represented as the corners of the hexagon given in Figure 1, hence the name of the conceptual framework.^[2] The six objectives identified in the hexagon model are briefly introduced below.

- *Research and analyze*: This type of activity aims for the generation of knowledge that will be used later for policy purposes. The major objective is to understand certain policy-relevant phenomena, and develop insights about it.
- *Design and recommend*: In certain situations the analyst can assist the decision making process by designing alternative solutions to a problem and analyzing and possibly weighing the consequences of these alternative solutions. The main question here is more about evaluating a set of interventions, or changing the system that is related to the already known phenomena. In other words, there is a certain action orientation that a recommendation is provided.
- *Provide strategic advice*: In certain situations, an analysis can be a strategic, client-oriented activity. The analyst can advise the client on the most effective strategy for achieving certain goals given a certain political constellation, i.e. the nature of the environment in which the client operates, the likely counter-steps of opponents, and so on.
- *Mediate*: A given policy problem may involve multiple-parties that have different views and perspectives regarding the issue. Addressing the problem and coming up with an

effective (i.e. accepted by all parties) policy may require the understanding of the other parties' perspectives. Hence, the task for the policy analyst may be mediating these multiple parties and promoting communication among them in a policymaking or decision-making process.

- *Democratize*: This type of policy analytic activity aims mainly at maintaining the involvement of all related parties in the policy process in order to make it as democratic as possible. This includes assuring the flow of proper information to all stakeholders, and the provision of opportunities for them to have their say regarding the policy issue.
- *Clarify arguments and values*: The main objective in this type of policy analysis activity is the elicitation of mindset, norms and values of the stakeholders involved in the problem at hand. In these situations, the analyst can support or help move forward the decision making process by analyzing the values and argumentation systems that underpin the social and political debate.

2.3 The six corners of the hexagon framework represent different types of activities with differing overall objectives, and also a differing set of approaches used. In that sense, the framework constitutes a promising grounding for our classification of models for supporting policy-making. Using the corners of the hexagon framework, we identified six types of models that correspond to the specific policy objective represented by the corner. These model types are discussed below, and also given in Figure 2 on a modified version of the hexagon representation.

2.4 Before proceeding, it is important to note that the classification is presented as a strictly discrete one. However, we also acknowledge the fact that many models do not purely fit into any of these categories alone, but carry some characteristics of a set of these corners. In that sense the classification given below should be taken as a conceptual framework that is helpful in discussing the differences among modeling-based studies, rather than as a strict and exhaustive classification with mutually exclusive model types.

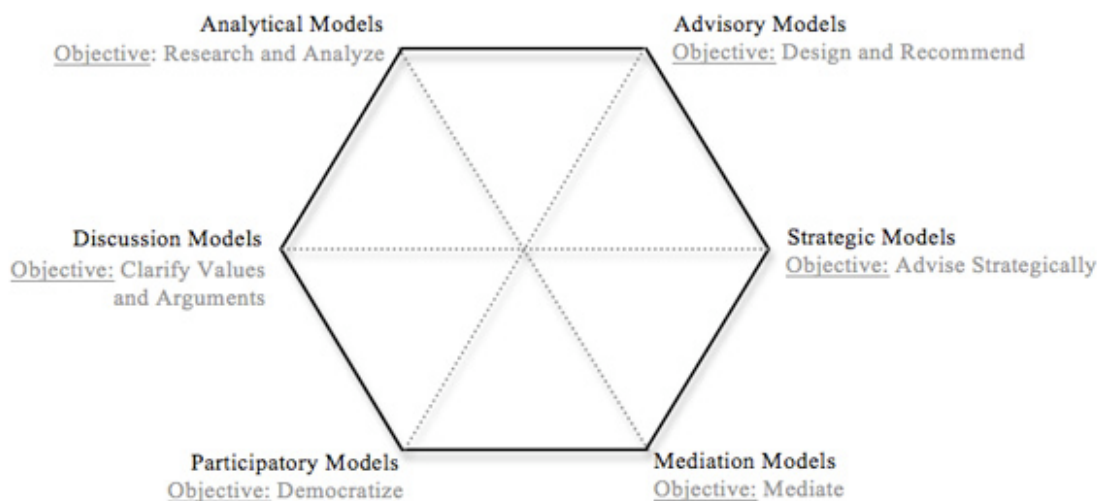


Figure 2. An overview of the types of models for supporting policy-making

a. *Analytical models.*

2.5 This class is composed of models that are constructed in order to support *research and analyze type* of policy analysis activities. The grounding of the model consists of mostly known scientific or empirical facts. Constructed models act as a

tool to develop insight about dynamic phenomena being studied. It can be claimed that these models aim to represent the system as it is. In that sense, a good representation of the 'real system' is the main concern in building this kind of model.

- 2.6 For example, early versions of IMAGE (Integrated Model to Assess the Greenhouse Effect) represent a good example of such models (Rotmans 1990). The model includes, for example, a simplified representation of the carbon cycle and of atmospheric processes. Although the model does not provide any new information about these individual processes, the main purpose is to develop insight about their dynamic interactions. This model is used to calculate future atmospheric greenhouse gas concentrations and the accompanying changes in temperature and sea level for a number of different scenarios of future emissions of greenhouse gases.
- 2.7 An older example of an analytical model is the Urban Dynamics model of Forrester (1969). The model was developed in order to study urban processes as aging industry and residential zones, immigration, emigration, etc. in order to comprehend the urban decay problem experienced in major American cities. The model didn't correspond to a specific city or didn't have the aim for quantitative evaluation of specific policies. The purpose is to understand the dynamic phenomena, i.e. urban decay, better, which in turn may help development of better policies subsequently.
- 2.8 Finally, the EpiSimS model can also be seen as an example of this category (Stroud et al. 2007). The objective of the model is to develop a better understanding about the diffusion of contagious diseases. As in the former examples, the objective is analyzing the underlying causes, mechanisms, conditions and/or processes that influence a dynamic problem. In this case, the model focuses more on the relationship of the disease spread, and the social network structure and demographic conditions, such as household size.

b. *Advisory models.*

- 2.9 Advisory models are used to serve *design and recommend type* of policy activities, which naturally implies a certain action orientation that a recommendation is provided by means of these models. Although this type of model has common aspects with the previous group, when we contrast them it can be seen that these models are more oriented towards studying an action and its impacts within a certain system boundary rather than understanding the system or an observed phenomenon. These actions may involve new policies to be used, and also modifications in the structure of the system related to the problem being studied. Although it is not a discriminatory property of these models, most of the times they focus on a particular problem context, rather than representing a dynamic phenomenon in general.
- 2.10 A typical example of such models is Invert (Stadler et al. 2007). The policy study in which Invert is used aims at identifying cost-effective policies for CO₂ emission reduction in the energy system. Representing the way the energy system works in specific contexts (i.e. seven European regions, especially Austria), the model enables experimentation for alternative policies, and concluding with a cost-effective technology portfolio mix. The model used for the Freight Options for Road, Water And Rail for the Dutch (FORWARD) project is another example of an

advisory model (EAC 1996). The project examines the benefits and costs of a broad range of policy options for mitigating the negative effects of the expected growth in road transport while retaining the economic benefits. The policy model of the transportation system used in this case was developed and used in the identification of some 200 tactics that might be combined into various strategies for improving freight transport. The model enables the design and assessment of policy options for several economic scenarios extending to the year 2015.

- 2.11** A policy model on influenza epidemics discussed in Ferguson et al. (2006) also constitutes a good example and reveals the difference between *analytical* and *advisory models* better. Considering the subject area, the model seems to have significant similarities with the EpiSimS model (Stroud et al. 2007), which is used as an example for *analytical models*. Although both models aim at covering epidemic dynamics, Ferguson et al.'s model (2006) is used more for identification of effective policies for epidemic prevention, rather than for understanding the social, spatial and biological dynamics of the problem.
- 2.12** This category of models is very likely to constitute the most frequently utilized one in the policy analysis context. Some other examples of this category include policy studies on conservation efforts in Amazon (Soares-Filho et al. 2006), agricultural development projects (Saysel et al. 2002), and diffusion of alternative fuel vehicles (Struben 2006).

c. *Strategic models.*

- 2.13** *Strategic models* are models that serve the development of strategic advice to the client for achieving certain goals given a certain political constellation, i.e. the nature of the environment in which the client operates, the likely counter-steps of opponents, and so on. In that respect, this type of model has some common characteristics with the *advisory models* since they somehow serve similar overall purposes, i.e. provide advice for the client. However, instead of focusing on the system wide consequences of a policy action, these models focus more on the reactions of the other actors operating in the same system as the client who takes a certain policy action. These models also have a 'representation' aspect, but what are represented are basically the other actors and their possible actions in the problem context.
- 2.14** The defining characteristics of policy activities in which these models are used are being specific to a client, and being related to the client's strategies. These two aspects limit the publicity of such studies as well as the models used. One of the domains rich in this kind of models is the electricity system. Especially, the liberalization and market design concerns of the states induced a variety of such models used for providing some insights based on possible actions of the generation and transmission companies. For example, Menniti et al. (2008) use game-theoretic models to simulate the possible actions of market players in the presence of multiple producers. A similar study focusing on the US market looks at the issue from the perspective of public authorities in order to reveal effective pricing strategies considering the transmission capacity related decisions of the electricity system actors (Hobbs and Kelly 1992). Exelby and Lucas (1993) use similar models to evaluate potential reactions of electricity generation companies in terms of bidding strategies to different capacity payment policies to be used by the central electricity system operator.

2.15 A rare corporate example of such models in the literature is by Zhao (2001). The model was developed for a corporation in order to assess the consequences of different part return policies considering the possible strategies to be used by the dealers of this corporation. The main focus of the model is to provide an experimental test-bed for conducting cost-benefit analysis for each possible return policy considered by the corporation.

2.16 As can be seen from the highlights of the modeling examples given above, the main focus of this type of models is oriented more towards the behavioral aspects of the other actors in a closed-system. Basically, this constitutes the discriminating characteristic of *strategic models*.

d. *Mediation models*.

2.17 A given policy problem may involve multiple-parties that have different views and perspectives regarding the issue. Solving the problem and coming up with an effective (i.e. accepted by all parties) policy may require the understanding of the other parties' perspectives. Hence, task for the policy analyst can be mediating between these multiple parties and promoting communication among them. The models used for achieving such an objective are labeled as *mediation models*. In the 'classical' use of models, the model is designed to be a representation of a system of concern. However, in the sense they are used in mediation types of activities, the representational aspect loses its significance. The model is designed to be more like a 'boundary object' between the different stakeholders of a conflict, and it operates as a ground for mediation (Zagonel 2002). The representative power of the model with regard to reality is not something aimed at here. The ultimate goal is much more pragmatic. These models can serve mental model alignment, creating agreement about a policy or design, generating commitment about a decision, or clarification of the problem (Andersen et al. 1997).

2.18 In a study about a water resource conflict in Bhutan, a multi-agent system was used as a mediation tool. A conceptual model was set up, based on which a role-playing game was developed and played; the results of which were then used for developing a computerized simulation. This process served to facilitate negotiation between the conflicting stakeholders. The game sessions corresponded to three different types of communication between the villages: intra-village, inter-village (collective), and swapping roles (Gurung et al. 2006). The model used in the policy analysis of Le Bars and Le Grusse (2008) on water resources management also carries similar characteristics. As clearly indicated by the authors, the aim in using the model "is not to find an optimal solution as do models based on linear programming or game theory, but to create models to reach compromise solutions that are acceptable to the different actors.". The model used in this case was developed based on a previously developed model, i.e. Olympe, which is a typical example of an analytical model. Modifying this former model into a mediation environment and using it as such, the authors reveal the conflicting interests of the stakeholders, and also existence of various coalitions in the problem context.

e. *Participatory models*.

2.19 *Participatory models* have significant similarities with *mediation models*. As in the case of *mediation models*, the model itself is used as a medium of interaction.

The most important aspect that differentiates the *mediation* and *participatory models* is that interaction between the stakeholders is not that important for *participatory models*. In the *mediation models* case, the model is an experimental ground to induce some mediation among stakeholders. On the other hand, a *participatory model* is more oriented towards inducing involvement and input to the policy analysis process from different stakeholders.

2.20 Briefly, the goal is to make the internal models/mindsets of the stakeholders about the problem and/or solution alternatives more explicit. This type of models may be built by the participants in a group model building process, or they can be models designed by the policy analysts in order to enable the participation of different stakeholders and express their perspective via interacting with the model.

2.21 Two examples of such models can be seen in the context of traffic problems. In the first case, a *participatory model* was developed for the Regional Transportation Committee (RTC) of Southern Nevada, which is responsible for managing the regional transportation system (Stave 2002). Problems, which had become worse in Las Vegas were traffic congestion and the related issue of air quality. The RTC formed an advisory group of 30 stakeholders who were asked to provide them with policy advice concerning these issues. A group model-building project was conducted. In this way, the model served the purpose of extracting the perspectives of different stakeholders regarding the possible scenarios, policy option and dimensions of the problem. Duijn et al. (2003) also benefit from a model for similar objectives in their work. During their analysis related to congestion and planning issues, the simulations provide valuable input from the stakeholders regarding "... the main policy issues: what are the perceived problems, can we reach an agreement on what our problem is, what are possible and acceptable policy options, and so on."

2.22 As mentioned previously, the proposed classification is discrete one and in reality policy models may serve more than one objective, which means that they may correspond to more than one class in this typology. For example, some of the examples from the companion modeling (ComMod) approach (Barreteau 2003) demonstrate the characteristics of both *mediation* and *participatory models*. The models are used for enhancing mediation between stakeholders, but also serves the objective of involving various stakeholders in the policy making process. In that sense, they also carry the characteristics of *participatory models* to some degree. The works of Barreteau et al. (2001) and D'Aquino et al. (2003) may be seen as such examples.

f. *Discussion models*.

2.23 These are 'models' used to facilitate elicitation of norms and values of the stakeholders. So far, we have not been able to point out a policy analysis activity utilizing simulation models with such an aim. The 'models' used here, which come closest, are conceptual models or mind-maps. Hence, this category of 'models' constitutes a less relevant category in our discussion in the following sections.

Assessment of Model-supported Policy Studies

3.1 Naturally, assessment of a model-supported study will show overlap with verification and validation processes. For certain ways of model usage the quality of the model-supported

policy analysis relies heavily on the quality and appropriateness of the model being used. In these circumstances, the assessment of the overall process, and classical verification and validation processes have a significant overlap. However, in some cases the role of the models in policy analysis activities differs from the classical role of models in natural sciences or engineering. In such cases, the quality of the policy analysis outcomes is significantly influenced not just by the validity of the model, but also by the way it is used. Drawing conclusions using a valid model, but totally inconsistent scenarios and policies during experimentation is an example to show that assessment of overall results demand more than validation of the model. In short, having a valid model may not guarantee the quality of the outcome. Depending on the policy-model type being used, development and usage stages of the model should also be evaluated additional to the validation of the model, as will be discussed below.

- 3.2** Validation, in general, implies a test procedure applied on the model itself, and/or on its output in order to evaluate the adequacy of a model in representing a real system/situation according to the objectives of the study. In short, it is an assessment of the model. In more conventional usages of models (i.e. simulation models in engineering and natural sciences domains), this also assures the quality of the outcomes to a large extent. For example, the way *analytical* or *advisory models* are used shows great similarity with this conventional usage, and it makes sense to equate validation of the model with the assessment of the model-supported policy analysis process. However, models as tools find themselves in different niches in the policy analysis domain where they are expected to serve different purposes. The classification given in the previous section revealed some of these alternative purposes in which being a representation of the real system is not an important aspect sought for. For example in some case, the model is used just for pragmatic purposes and there is no claim that it represents particular real system. In such cases, verification and validation is still important, but not enough for building up confidence for outcomes obtained using the model.
- 3.3** What we are proposing here is that a quality assessment procedure that goes beyond the standard verification and validation is required for assuring a high quality policy study and also for assessing the reliability of the outcomes. Such a process should be based on the stages of the whole process (i.e. development and usage), not only on the model as an object. An overview of stages relevant to a model-supported study is provided in Figure 3. It is clear that not all modeling studies go through all of the processes depicted in the figure. It is basically an overlay of processes from differing modeling studies combined. Hence, the sequence of stages and/or the nature of the task in each stage may be different depending on the type of model. Also, these stages are more often followed in an iterative manner rather than a single-pass sequential manner. However, for the sake of simplicity we have not emphasized this in the representation.

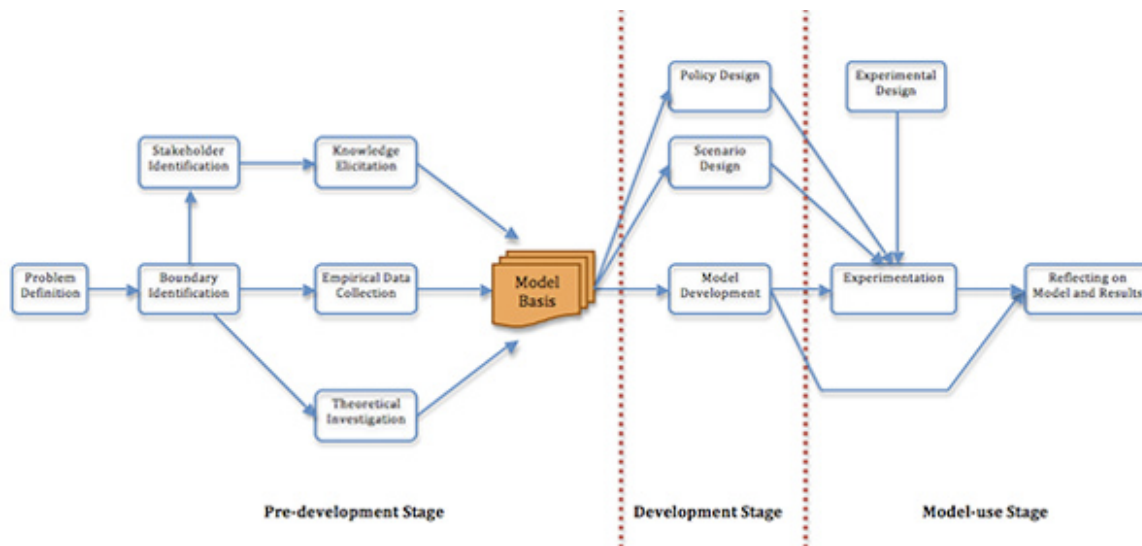


Figure 3. Modeling stages in general

3.4 This overall process overlay provides some clues about the points that should be considered in assessing the quality of the overall process and the outcome. Hence, we will be using the overlay as a guide, and in doing so, we will be paying attention to the fact that some stages shown in the figure might be irrelevant for some modeling studies, and the sequence might be slightly different for the others. These issues will be discussed in more detail in the following section.

a. Boundary assessment.

3.5 This aspect is mainly related to the processes of problem definition and boundary identification. By definition every model is a simplification; hence one of the crucial aspects of model development is determining the factors and relationships to be left out of the model. The boundary assessment of a model is related to the appropriateness of the selected boundary in leaving out factors and relationships. In other terms, it is more about evaluating what is left out of the model, rather than what is included in the model. This aspect is also highlighted as the primary and a very important part of model validation in some validation literature (Forrester and Senge 1980; Sterman 2000), but in general it seems to be underemphasized or overlooked.

3.6 A number of important issues related to this aspect includes exogenous parameter justification and ignored mechanisms/interactions justification. Defining a parameter as being exogenous implies that that variable is assumed to be independent/autonomous of the dynamics of the system being represented by the model. This claim has to be justified by the researchers since it is about ignoring a link between the system being modeled and the variable declared to be exogenous, and this might have a significant impact on the results to be obtained. To give a very bold example, assuming petrol prices are exogenous in a regional energy model, whose share in global energy market is 1% may be justifiable. However, the same assumption will be easily challenged in a model that represents the global energy consumption. A variable to be defined as an exogenous one should be either autonomous from the system being modeled, or the impact of the system's behavior on the exogenous variable should be very slow compared to the time horizon used in the model.

3.7 The second issue is very hard to tackle. It is about justifying the left out mechanisms, relationships, and other aspects of the system. Since the number of

such elements is huge, it is not feasible to consider going over each omission and justify the decision. However, in most of the cases an expert opinion coupled with some intuition helps to point out aspects to be focused on. In a model of social conflict, it is almost certain that the individuals' eye colors are irrelevant and can safely be ignored in the model. Such ignorance sounds common sense and obvious, hence no need for further justification. However, in the existence of multiple theories about factors that may influence conflict, including one factor/relationship in the model and ignoring the other may require some further justification. At some abstract level Dunn (2002) discusses the issue, emphasizing the risk of overlooking causally relevant factors that may be critical in the way problem being studied is influenced. He labels the corresponding process as "context validity" and focuses on the elimination of challenging hypotheses regarding the phenomena being studied. On the more practical side, Forrester and Senge (1980), and Sterman (2000) propose similar tests that should be considered regarding the validity of the selected system boundary. Putting it very simply, they propose to include the aspects that are originally excluded and speculated to be influential regarding the studied phenomena at a testing phase. Then the impact can be evaluated by studying the changes in the model behavior and the conclusions to-be drawn under this modified outcome. If there is no significant alteration, then the decision of excluding that aspect can be justified.

b. *Basis Assessment.*

- 3.8** Referring to the modeling process depiction given in Figure 3, this aspect is related to the information (e.g. information from empirical observation, theory, evidence, or tacit knowledge of actors) used in model development, i.e. model basis. The issue is both related to collecting and filtering the data, as well as deciding on what kind of a source mixture is appropriate for the intended purposes. Brenner and Werker also provide a discussion related to this matter in their 2007 and 2009 papers.
- 3.9** The concept of 'source mixture' may require further clarification. By that we imply the decision regarding to which extent a model has to be based on empirical case-specific data, or about not using case-specific evidence and relying purely on theory, for example. Consider an innovation diffusion model, in which the innovation's costs are assumed to decline due to learning curves (Argote and Epple 1990; Argote 1999). In the diffusion theory literature the percentage of improvement is claimed to be in the range of 15–30% for every doubling of the cumulative experience with the production process. So, it may be acceptable for a modeling study to rely on this theoretical data, if the objective is to conduct some enquiry about the diffusion process in general. However, if the study is all about the probable diffusion dynamics of a particular innovation, such a reliance on purely theoretical data may damage the credibility of the study. In this case, some empirical evidence specific to the innovation being studied may be much more appropriate. This issue is discussed in detail by Boero and Squazzoni (2005). They provide a review of different strategies of gathering data of differing natures (i.e. statistical data, data collected from stakeholders, etc.), and also a classification of models, which they use while discussing what kind of source mixture, may fit which type of model. This dimension is more related to appropriateness of the basis being specific or generic. Another dimension is related to appropriateness of the basis being objective or subjective. This point is much more complicated compared to the previous one, since it is highly related to the epistemological

stance of the policy analyst or assessor. The models used may be based just on subjective information from stakeholders about the system and the problem, or the modelers may aim for an as objective as possible basis relying heavily on theory and empirical observations. Considering that the epistemological question in general (i.e. realism vs. constructivism) is a sort of 'open issue' (Becker et al. 2005), it is hard to provide a guideline about this dimension of the basis development. However, an explicit statement about this type of high-level perspectives that shape the model development is crucial, since it reveals fundamental assumptions that are implicit. This will enable assessment of the study in a specific frame, and also enhance the acceptance of the study in a community sharing similar perspectives (Ahrweiler and Gilbert 2005).

- 3.10** Leaving the 'source mixture' issue aside, the second issue is the collection and filtering of the data from different sources for the model basis. The issue about the empirical and theoretical enquiry seems to be beyond the scope of this piece. However, we believe that information from the stakeholders/participants being a less explicitly discussed issue deserves some elaboration in this respect. Two sub-issues can be identified regarding this point; identification of the participants to be used for knowledge elicitation (i.e. source identification), and the design of the knowledge elicitation process. We will briefly discuss both of these.
- 3.11** The stakeholder/participant identification problem is related to what we can label as 'participant coverage'. This is closely related to the formerly discussed boundary assessment since it is about evaluating the coverage of the model-supported policy analysis process. However, in this case the issue being evaluated is the stakeholders involved in, or consulted during, the modeling process. Very simply put, it is about answering the question of whether all parties related to the problem at hand are considered in the process. Omission of key stakeholders may have differing implications on what kind of policy activity is being carried out. In cases where stakeholders are primarily used as the information/knowledge source regarding the problem, it may lead to an improper representation of the system being modeled. In some others where the involvement of the stakeholders is used for designing policies, it may lead to policies that fail due to unaccounted for stakes and reactions. In that sense, the identification of the stakeholders to be involved in the model-supported policy analysis process needs to be evaluated as a precondition for building confidence in the outcomes of the analysis results.
- 3.12** Unfortunately, there are no well-established standards or guidelines available regarding the selection of stakeholders to be involved. This point is also highlighted by Andersen and Richardson (1997) with respect to the group model building studies, in which group identification is a crucial issue. Despite this void in clear guidelines, there are also some attempts to pinpoint important aspects of the issue of stakeholder selection in policy studies (Rowe and Frewer 2000; Prell et al. 2008).
- 3.13** The design of the knowledge elicitation process in studies where stakeholder participation takes place is also a very important aspect regarding the overall assessment of the study. Even the perfect identification of stakeholders to be involved in the process does not guarantee an effective knowledge elicitation process. The decision regarding the mode of participation and methods to be used to gather the information from the stakeholders is another challenge. The methods used in designing the interaction environment, mode of participation, as

well as the way information gathered is interpreted influences the effectiveness of the process (Luna-Reyes and Andersen 2003; Rouwette 2003; Ramanath and Gilbert 2004; Bots and van Daalen 2008).

3.14 As already mentioned, there are no well-established standards and guidelines, and we also highlighted problems more than providing solution proposals. In that respect, the only way to go seems to follow best practices. However, the point we would like to make regarding the issue is recognition of the importance of this aspect regarding the quality of the outcomes. A recent statement by Moss (2008) also reminds us of the significance of the issue. He proposes that models heavily dependent on the participators' depiction of the system are naturally validated and already approved by the participants/problem owners. This statement is agreeable only when the participant identification and elicitation of knowledge from the participants is done in an acceptable/valid way.

c. *Representational Assessment.*

3.15 This aspect of assessment basically corresponds to 'standard validation', and it is mainly focused on the model and its behavior. It is concerned with the sufficiency of the final model used in representing a system, which includes putting the conceptual model as well as the formalized model under scrutiny. In that sense, most of the validation approaches, and what is understood from the term validation (Barlas 1996; Gilbert and Troitzsch 2005; Moss and Edmonds 2005; Schmid 2005; Moss 2008) can be seen as being related to the representational assessment.

3.16 Validation is one of the fundamental issues regarding simulation modeling. Two strands of works can be identified regarding validation of models developed for systems with social components. The first strand can be seen as more methodological and focuses on the procedures and individual tests to be conducted in order to validate a model (Forrester and Senge 1980; Miser and Quade 1988; Carson and Flood 1990; Balci 1994; Barlas 1996; Sargent 2004; Yilmaz 2006; Wildrum et al. 2007). The second strand is mainly rooted in the agent-based social simulation domain, and it elaborates more on the epistemological aspect of the validation issue (Ahrweiler and Gilbert 2005; Becker et al. 2005; Boero and Squazzoni 2005; Frank and Troitzsch 2005; Küppers and Lenhard 2005). The latter strand is an extension of a more general epistemological discussion in social sciences, and it is an open scientific discussion. Depending on the epistemological perspective (i.e. realism or constructivism) of the modeler and/or audience, what is to be validated and the way it should be validated differs. Although our aim is not to explore and contribute to the epistemological issues regarding simulation modeling, the existence of multiple views affects the discussion on validation. Since it is possible to recognize both perspectives in the model-supported policy studies, we aim to cover both perspectives in our discussion of representational assessment. In order to do so, we will go over various issues regarding validation, but depending on the high-level epistemological assumptions some of those will not be important or relevant for some particular studies. We will try to highlight this point as much as possible.

3.17 At a general level, we can speak about two major types of representation; structural representation assessment and behavioral representation. In cases

where it is implicitly or explicitly assumed that a 'real system' exists out there, and it can be observed and replicated objectively, we can speak about structural representation, or structural validation^[3]. The structural representation is something especially aimed for in modeling studies where the objective implies imitating the processes that lead to the dynamic phenomena being studied. To put it in a simpler manner, structural representation is important when replicating the dynamic phenomena is not enough by itself, but understanding the underlying processes/mechanism are also important (i.e. when goal is not just replication, but also explanation). Models that may be classified as causal–descriptive, theory–like, or transparent/white box constitute the set for which it is important. In cases where this is not the case, the structural representation does not constitute a challenge for the modeler. Validation of the structural representation aspect seems like the other side of the coin compared to the boundary validation. In this case, it is all about validating the interactions, processes and elements in the model. This process is labeled as structural validation (Zeigler 1976; Barlas 1996; Troitzsch 2004), theoretical verification (Takadama et al. 2008), structure assessment (Sterman 2000), conceptual validation (Sargent 2004), etc. by different authors. Depending on the modeling approach to be used, specific procedures to be followed may differ significantly. However, the common notion is the comparison of interactions and behavior rules against the existing knowledge about the system being modeled. This knowledge is a combination of theory, empirical observations or the tacit knowledge of the experts.

3.18 Behavioral representation, or behavioral validation, is the second aspect of representational validation. Again what is being scrutinized is how well the model represents the system of concern, but this time the criterion of evaluation is based on the behavior of the model and the system. The proximity of the model in replicating the system constitutes the criterion of evaluation. However, the 'proximity' assessment may in most cases be subjective and also qualitative. This kind of validity overlaps with the concepts as operational validity (Sargent 2004), external validity (Takadama et al. 2008), behavioral validity (Barlas 1996; Sterman 2000), replicative validity (Zeigler 1976; Troitzsch 2004), etc. discussed in the literature. Going one step further, we will differentiate between three types of proximity: numeric proximity, dynamic pattern proximity, and the terminal pattern proximity.

3.19 The numeric proximity is all about the level of fit between the data collected from the real system and the corresponding model output. For example, for a population model, comparing the real data for years 1995, 2000 and 2005 with the model output and the same time points serves the assessment of this kind of proximity. This is a point–wise comparison since data from two systems at a certain point in time are compared. Among the three types of proximity discussed, this is the easiest one to quantify since it is possible to use various numeric/statistical measures to use as a metric of proximity. The problem with this type of proximity is the fact that it does not assess the dynamics of the system between those points of time being used in comparison.

3.20 The second type of proximity, i.e. dynamic pattern proximity, addresses that issue. In this case, point–wise comparison does not say something very relevant. The objective is to assess whether the model can replicate the characteristics of a variable's dynamic behavior over time. Going back to the population model example, the proximity is judged based on whether the model output resembles

an exponential growth phase followed by a stabilization phase, which is the population dynamics of the country being studied. The numeric proximity is also something desirable, but its priority is lower than the pattern-wise fit in this case. Some statistical measures can be used to quantify this kind of proximity, but their match to the task is questionable. Instead, pattern identification algorithms are proposed for this task, pointing out the potential deficiencies of traditional statistical measures (Yücel and Barlas 2007). However, in most of the cases the judgment of proximity is left to the qualitative judgment of the modeler or the client.

3.21 The third type of proximity is more about the patterns, most likely spatial patterns, being observed at the end of a simulation run. In cases where the phenomena being studied is about emergence of a certain spatial pattern, this type of proximity has to be assessed in order to discuss the validity of the model in terms of its representative qualities. The well-known segregation model of Schelling (Schelling 1971) constitutes a very good case where this type of proximity is vital. The phenomenon being studied is the emergence of segregation in an urban area. The model can be evaluated as successful if it is possible to observe a spatial pattern that resembles the segregation. The different nature of this type of proximity should be clear. It has almost nothing to do with numerical proximity (i.e. it is not important whether 10 or 8 homogenous neighborhoods emerge), and the time trajectory (i.e. when exactly the segregation happens); it is all about assessing a pattern at a point in time. This might be the hardest of the three to quantify. Hence, in most cases the assessment has to be done visually by the modeler, or the clients.

d. *Experimental Setting Assessment.*

3.22 As can be seen in the modeling process diagram in Figure 3, in some cases the model development process itself may lead to conclusions drawn mainly via reflections on the model and the development process, without any experimentation with the model. This is the dominant case in soft systems methodology (Checkland and Scholes 1999), qualitative modeling (Coyle 1996), and similar areas. In other cases, the model is used as a dynamic experimentation ground. As the model itself, the credibility of the outcomes of the study is also dependent on the design of this experimentation process.

3.23 The aspects to be considered in designing the experimental process are naturally different for representational, and for pragmatic models. For representational ones, the points to pay attention to include the determination of contextual settings, scenarios, and policies to be used during this stage. Similar to the boundary identification for developing the model, scenario and policy design processes also demand a boundary study (Forrester and Senge 1980). Especially, in the scenario case the portfolio of scenarios to be tested should cover the plausible and relevant conditions for the system being studied with respect to the problem at hand. The issue is similar on the policy design side. The plausibility of the policy is a primary concern. It is not enough to design a policy that improves the problematic situation, but it should also be acceptable considering the organizational structure and the other constraints of the system being studied. On top of that, the policy design and testing also has additional challenges that put the results to be obtained under scrutiny. Some policies are not just small adjustments to the system while keeping the original structure (e.g.

organizational, technical, social structure). They may include major structural alterations in the system. In these cases, the feasibility of the implementation in the way it is done in the model constitutes a major concern. Any result to be obtained in such experimentation needs to follow a justification (i.e. validation) of the way policy is implemented in the model. Additionally, another potentially important problem is what we will label as 'sleeping structures'. The depiction of the system may be satisfactory given normal conditions. However, a new policy may invoke some reaction in the system that was not visible under the normal conditions, hence excluded from the model. These cases can be seen as some passive interactions/mechanisms becoming active in the system, as a response to the proposed policy (i.e. waking up the sleeping structures). Again, any policy recommendation lacking a proper boundary adequacy validation, considering the new policy implemented, to check for waking up structures will damage the validity of the recommendations, hence the whole policy analysis process. Finally, in most of the cases the models may include probabilistic components, which may yield various levels of changes in the model output among replications. In these cases, conclusions obviously may not be made based on a single run of the model. However, the issue is the number of replications needed, and also the summary statistics to be used in drawing conclusions based on these multiple replications. A lack of careful consideration of the issue may lead to discussions about interesting findings, which may be quite improbable in the real system.

3.24 In the case of pragmatic models, the picture is totally different. The experimentation with these models is used to achieve a goal, such as learning about participants' reactions under certain settings, or helping the participants to recognize the problematic nature of their actions, etc. This includes either direct, or indirect interaction of the participants with the model. In the direct interaction case, the interface, the information available to the participant, the way information is provided to the participant, the way participant can interact with the model may seem like straightforward issues to be decided intuitively. However, even a simple change like whether to provide the instantaneous value of an indicator, or its graph over time may influence the perception on the participant side significantly, and this eventually may lead to different responses. Although it was not documented in detail, a set of such instances was detected in an experimental study conducted by Dalkiran (2006). Apart from this, a point that is relevant to both direct and indirect interaction cases is the briefing part of the process. The amount of information about the experimental model the participants will be given, and the way this information is provided to them may be expected to condition their reactions during the experimentation phase. One way to handle the issue may be including the design choices as independent variables influencing the decision process. This may consist of the same participant or participants from the same homogenous group being exposed to multiple experiments, among which the main change is about the interaction design. With a good design it should be possible to rule out the impact of a particular interaction design on the observed reactions of the participants.

Relevance of Different Assessment Aspects for Different Model Types

4.1 In this section we will try to combine what we have introduced in the last two sections, and discuss the importance of different aspects of assessment with policy analysis cases utilizing different types of models classified in section 2. In the case of the first three model types (i.e.

analytical, advisory and strategic) similarity in the nature of models results in similar important points that influence the outcomes. The other model types (i.e. *mediation, participatory and discussion*) have more significant differences in nature, and the set of important aspects to be discussed will be significantly different from the first three.

4.2 Before we go into the discussion, we like to note that the discussion will be an archetypical one. In other words, while discussing a model type, we will rely on a generalized typification of that certain type of model.

a. *Analytical model assessment.*

- 4.3** As discussed earlier, the main objective in using *analytical models* is developing some explanation and insight regarding a system's dynamic nature. Since these models are seen as simplified copies of the real system, both structural and behavioral representations have the highest importance. First of all, a primary condition for building up confidence in the outcomes of a policy study using an analytical model is the success of the model in replicating the behavior of the system. Although it is possible to have exceptions, in most cases numeric proximity achieved with this kind of models fall short in giving a clear indication of the quality of the overall outcomes of the analysis. The focus should be more on the dynamic behavior. Hence, achieving dynamic pattern proximity or terminal pattern proximity will often be important. However, this is not enough.
- 4.4** Since provision of an explanation is a quality expected from policy studies using this kind of models, behavioral validity of these models is not satisfactory alone. The relationships and action rules that cause this behavior should be validated (i.e. structural validity), since they constitute the explanation sought for. The quality of the provided explanation relies heavily on the internal consistency of the model structure as well as its consistency with the theoretical and/or empirical basis.
- 4.5** This brings us to the issue of basis assessment. These models are based on a theoretical grounding, and/or some empirical data collected from the system. For systems of a purely physical and/or technical nature, assessment of the used theory and data is more or less straightforward; generally commonly accepted theories exist in these domains, and 'objective' data collection is easier and methods for this are well established. It is hard to say the same for systems having social components. This constitutes an important challenge in the assessment, and one that can be easily overlooked. The reliability of the empirical data is very important. Hence the important question to be asked is about the existence of a very probable subjectivity and bias in the empirical data from the social system used to support the validity of the explanation. This should be evaluated by the modeling team, but also the measures taken for obtaining reliable empirical grounding should be documented in order to allow healthy assessment of the process by third parties. Additionally, there is rarely a commonly accepted dominant single theory to build upon regarding social systems. In that respect an objective assessment of such models, as well as study outcomes, is not possible. The assessment can only be conducted with respect to the theoretical stance of the policy team. Also, one related potential risk in such a study is overlooking the possibility of alternative explanations. model this may easily lead to alternative explanations. Depending on the theory used in constructing some parts of the model it may be possible to come up with

alternative explanations. The outcomes of the study should be evaluated in the light of this possibility.

- 4.6** In most cases, these models are multi-disciplinary, which requires the integration of different theories from different fields. Since each theory holds its validity in a certain range of conditions, it is also important to evaluate the consistency of the theories brought together in the model. One way of performing this evaluation is to check the *ceteris paribus* conditions of used theories against each other. Similarly, the assumptions to fill the 'voids' in the model, or used to operationalize the theories used should be assessed carefully for their consistency. More importantly, they should be explicitly reported to allow assessability.
- 4.7** Development of the explanation can be achieved via repeated experiments with the model. There are two important points here. Firstly, the conditions under which the model 'explains' the phenomena are crucial. The explanation can depend too much on these conditions (e.g. state of the environment represented by exogenous parameters) and the results' generalizability is not assured. This point should be considered carefully in the assessment. Secondly, in models having a stochastic nature (e.g. models with some random processes) the explanation may be due to this, rather than to the model structure. Especially, models aiming to explain rare events in the system (e.g. price peaks) are more vulnerable to such a risk. So the true nature of the processes, as well as the extent of the random processes' influence should be evaluated carefully. Otherwise, the conclusions drawn via such a model may be misleading and flawed.

b. *Advisory model assessment.*

- 4.8** Since this type of models is quite similar in nature and in the way they are used to the *analytical models* discussed above, most of issues highlighted for the *analytical models* hold also for *advisory models*. Two aspects create some distinction in terms of assessment. First of all, policy studies in which *advisory models* are used generally focus on a particular system under particular conditions. Secondly, the emphasis is on evaluating the consequences of a certain change on system's behavior in the advisory model usage. This may be a change in system's environmental conditions, or it may be a change regarding the way system operates (i.e. a change in the system structure). These two bring about some differences.
- 4.9** The assessment of the behavioral validity of the model is crucial for the credibility of the results. However, due to the fact that these models represent a particular system, numeric proximity of the model output is likely to be important, since it provides extra evidence regarding the fit between the model and the specific case being studied. Furthermore, the main assumption in relying on the outcomes of such a model is the fact that model behaves the same way as real system does under given conditions. In that sense, the model should be able to replicate and in the same time forecast system behavior, even under structural changes.
- 4.10** If the model is going to be used in an input-output analysis kind of manner, the structural validity of the model loses significance in evaluating the outcomes of the study, since the model is used like a black-box behavior generator. However, if the analysis requires evaluation of alternative system designs or policies that require alteration of the current system structure, structural validity of the model becomes a necessity. The model should resemble the real system also regarding

the structure, so that policy alternatives to be evaluated can be mapped into alterations in the model structure in a reliable way. In short, structural validity of such models is also important in most cases.

- 4.11** As mentioned above, the studies in which these models are used generally focus on specific systems and problems. This makes these studies, as well as models, mainly reliant on the case-specific empirical data. Evaluation of the quality of this data is crucial in building up confidence regarding the overall conclusions drawn using the model outcomes. This includes careful investigation for biases in the data sets, as well as the subjective nature of observations. Additionally, any parameterization that does not rely on empirical observations needs to be explicitly discussed and justified in the particular context on the particular system being studied. A parameter value that is inconsistent with the specific conditions being studied will damage the credibility and quality of the conclusions significantly. Hence, they should be avoided, or justified with sound reasoning in order to attain credibility.
- 4.12** In most cases, an initial set of scenarios and policies constitute the initial problem that is the subject of policy study (e.g. how will system behave under such conditions with such policies?). However, it may be required to expand the set of policies and scenarios considered. In these cases, the relevance of the outcomes mainly depends upon the quality of these new policy and scenarios. One way of assuring quality is to involve problem owners or domain experts in development of these, which almost automatically guarantees the appropriateness of the new set. Otherwise, it may be required to use the problem owner or domain experts for posterior assessment of the developed new scenario or policies. This point should be considered in assessing the relevance and effectiveness of the analysis outcomes.
- 4.13** A major challenge in assessing the quality of the results obtained with an *advisory model* is what we label as 'sleeping structures.' These are relationships or behavior patterns not active in the real system under normal conditions. The problem is that these sleeping structures may become active as a consequence of changing conditions or as a reaction to new policies applied. A model that is used to replicate system behavior under specific conditions need not consider these. However, if the aim is evaluation of changes in the conditions or policies, the analysis should recognize the potential existence of sleeping structures. At the minimum acceptable level, the study should acknowledge the uncertainty in the conclusions drawn from the model due to the risk of such structures. Additionally, extending the system boundary in order to foresee most likely sleeping structures and inclusion of them into the model is another quality that may be sought for in assessing the reliability of the model and outcomes. Finally, a solid discussion ruling out the possibility of having such structures should also be expected as supplementary or partial substitute for the previous ones.

c. *Strategic model assessment.*

- 4.14** As mentioned before, what are mainly represented in this type of models are the reactions of the other actors in the problem context to certain actions of the client. Hence, in most cases, the boundary of the study is already well specified in the problem definition (e.g. which actors to study, which actions to evaluate, etc.). It can be said that these models are generally confined in a well-defined context

in which strategies of actors are evaluated. In that sense, boundary assessment is intuitive. The only check to be conducted is the obvious one; ensuring that the model boundary is consistent with the one depicted by the problem owner. However, in some cases the analyst is also required to question if the given depiction is the most adequate boundary to address the problem.

- 4.15** These models are designed specifically for a client, and the environment that this actor operates in is almost unique. Hence, relying on general theoretical information may hinder the credibility of the study. More intense linkage to the particular case is needed, which makes the information from participants and empirical case-specific observations more important. Assessment of the empirical data that constitutes the model basis is of primary importance. The past behavior patterns should be evaluated in terms of observation quality and the context-dependency. It should be assured that every action of the actors is evaluated with respect to the context it took place. Otherwise may lead to problematic strategic profiles that yield to flawed analysis.
- 4.16** Representational validity of these models is important in the sense that they are expected to replicate strategic behavior of other actors in the context. However, most of the times just replication of the actions/decisions under given conditions is satisfactory, and whether the modeled actors use the same decision procedure as the real ones is not relevant. An example may make the point clear. Consider a game-theoretic model in which actors are depicted as perfectly rational and omniscient ones. This depiction may easily be challenged based on existing theory and empirical studies. However, as long as the actions of the model actors resemble the actions of the real actors in similar conditions, this structural invalidity does not damage the credibility of the study. Hence, it can be said that the behavioral validity of these models has the primary importance, whereas contribution of structural validity is marginal. However, a room for new strategies should be recognized in the model itself or the way it is used. Either the model structure should be flexible enough to allow actors to take actions that were not observed in the past, but are plausible; or the model should be used given an extensive set of 'possible actions set' scenarios. Absence of any of these or a similar approach to consider room for strategy would put the reliability of the results under suspicion.
- 4.17** As one crucial aspect of actor representation is their strategy sets, the other one is their priority and preference structures that determine which strategy they utilize under which conditions. Direct observation of such preference structures is not very likely in strategic studies; hence some indirect elicitation process is conducted in general. The credibility of such a process as well as the obtained actor profiles conditions the outcomes to a large extent. For that reason serious attention is required in evaluation of these, e.g. via expert judgments or else.
- 4.18** Finally, the points discussed formerly for *analytical* and *advisory models* regarding the scenario and policy design processes are equally applicable to the strategic models.

d. *Mediation model assessment.*

- 4.19** The nature of studies in which *mediation models* are used is significantly different from the cases with the former model types. Since the model is used as a boundary object, a medium of interaction, there is almost no claim and

expectation that the model represents a real system. The model is mainly a pragmatic object to be used to support mediation. Naturally, this reduces the significance of classical model validation in assessing the outcomes obtained. The assessment of the outcomes obtained by mediation models is mainly dependent on the way model is used, rather than what is represented by the model.

4.20 However, based on the premise that the model used has the claim of representing a real system, ignoring validation or model evaluation may be misleading. There is still some relevance of the classical validation in evaluating the quality of the study. Although the model does not correspond to a real system, analysts aim to create a copy of a hypothetical decision situation/setting in order to induce effective mediation. Hence, the model should be evaluated with respect to its appropriateness for this objective, or simply with respect to its success in imitating the intended situation. In order to present the problem situation in a simplified manner, it is possible that the policy analyst may utilize some analogies. It is vital to assure that the analogies between the concepts in the model and the real conflict situation are understood correctly by all stakeholders. Otherwise, the mediation process may turn into a model-supported miscommunication session. A related concern regards the setting in which participants interact with each other and also with the model, i.e. experimental setting. The assessment of the model can be done via conventional validation procedures, or alternatively/in addition test groups and evaluating the situation perceived by the participants via debriefing sessions. The latter option of test groups and de-briefing may also serve the purpose of assessing both the model and the experimental setting at once.

4.21 The boundary definition for the problem is crucial since it also automatically points out the relevant stakeholder set to be considered. Since the ultimate aim in using such models to induce some sort of mediation that may enhance designing effective solutions, involvement of all major stakeholders is vital. Hence, proper stakeholder identification and assuring the coverage validity of the study are key issues to be checked in assessing the model and outcomes generated by its means. A misrepresented stakeholder set will yield an outcome, but the effectiveness of the outcome will be very likely to fail in the real situation. The coverage of the model-supported mediation process should be questioned and justified by the analyst team, and this justification is a quality to be sought for in a high quality policy analysis activity utilizing mediation models.

e. *Participatory model assessment.*

4.22 As was clear from their characterizations, *participatory* and *mediation models* have significant overlaps in the way they are used and what is expected from these models as pragmatic objects. In that sense, what is discussed regarding the representational assessment of the model also holds for assessing outcomes of policy activities utilizing *participatory models*; its link to a particular real system or situation may not that important, but its consistency with the situation that is aimed to be created by the policy team should be evaluated carefully.

4.23 Again, the identification of the stakeholders set to be included in the process, i.e. participants, is vital. The quality of the outcomes can depend heavily on the participant set involved, as well as the knowledge elicitation process utilized. Despite a perfect coverage of all knowledgeable and related parties, it is still

possible to end up with a problem definition or system description that is not exactly the same as the stakeholders had in their minds. The stakeholders maybe biased due to the design of the knowledge elicitation process, or the information provided by the stakeholders maybe misinterpreted by the policy analyst. Hence, the credibility of the study depends on the effectiveness of these processes and their reliability. The authors are not aware of a commonly accepted set of guidelines for assuring the reliability of elicitation processes. The best strategy in this situation would be relying on best practices and former experience in the literature in order to pinpoint important aspects that influence the quality of the outcomes, and assess the study based on these. One way to go is to reduce the impact of the subjective interpretation of the policy analysts, and increase the additional information about the reasoning of the participants in their actions as much as possible. A complete reliance on the subjective interpretation of the analyst may indicate serious flaws in the conclusions of the study in general. Alternatively, additional information collected via debriefing sessions about why participants acted or decided in particular ways may reduce the risk of flawed outcomes due to analyst bias.

- 4.24** A related issue may be perceived more like an aesthetic one, but it may have significant impact on the information collected via the model: interface/interaction design. In cases where the participants interact with the model through an analyst (i.e. a participant states his/her decisions, and an operator who is familiar with the model communicates the decision to the model) this is not an issue at all. However, if the participants are directly interacting with the model, it is crucial. Depending on the amount of information provided, the way it is provided, and the way decisions are input, the participant reactions to the same model may differ. Hence, assuring the quality of the results demands evidence that the interface has no impact on the way the participants acted. This can be achieved via using alternative interfaces on identical or very similar participant profiles, and experimentally evaluating that the impact is insignificant. Information collected via debriefing may also provide valuable information regarding the impact of the interface, but its reliability won't be as strong as for the former approach.

Conclusions:

- 5.1** The paper focuses on the assessment of outcomes of a model-supported policy analysis study. However, such a discussion at the general level makes little sense considering the differing natures of models, and the differences in the ways they contribute to the policy study. The first aim of the paper is classifying this variety in classes that will enable discussing differences in model types. This model classification is mainly inspired by the hexagon framework of Mayer et al. (2004), which was developed for classifying policy analysis studies according to their overall objectives. Although the framework of Mayer et al. (2004) does not focus on models, but on policy studies in general, it serves as a good starting point for our purposes. We mapped the policy studies classification of Mayer et al. (2004) onto the domain of quantitative models used to support such activities, and developed a classification of policy models. The classification presented in the paper introduces six classes of models, namely *analytical*, *advisory*, *strategic*, *mediation*, *participatory* and *discussion* models. We also discuss the general aspects that differentiate the different types.
- 5.2** The second contribution of the paper is discussing the various aspects of quality assessment that spans the life-cycle of the model (i.e. development, evaluation, usage), and differences

in relevance of these aspects for different model types identified in our classification. This yielded a type-specific discussion of important aspects that require attention in evaluating the whole model-supported process in terms of the quality and the relevance of the outcomes with respect to the problem at hand. The aspects discussed are related to *boundary*, *basis*, *representational*, and *experimental setting* assessment. These aspects can be utilized in two ways; first of all they are intended to be points to be checked posterior to usage to evaluate the reliability of the outcomes. However, since they are aspects that have significant impact on the quality of the outcomes, they can also be considered as guidelines to be considered prior to the model development process.

- 5.3 The set of aspects discussed in the paper may be seen as intuitive and obvious for experienced policy analysts and modelers. In that respect, our goal is not to 'invent' a totally new perspective or set of procedures to be considered in assessment, but to provide these aspects, which are often overlooked during the development and analysis, in a structured way as a brief set of guidelines.
- 5.4 In the paper, we claim the importance of evaluating different aspects of a model itself, and the way it is used. However, we do not provide a precise prescription of how to conduct these evaluations. In that respect, this work may pose more questions and concerns regarding assessment, rather than providing solutions. That is also one of the intentions of this piece. Such a discussion aims to point out some voids for further research about whether generally accepted procedures can be brought to life to be used for certain aspects of assessment.



Notes

¹ Due to the context-specific nature of the paper, when the term 'model' is used in the following sections of the paper, it should be understood as "simulation models used to supplement/support a policy analysis process."

² The reader may refer to Mayer et al. (2004) for an extensive introduction of the hexagon framework.

³ It is possible to encounter several different classifications (and also set of labels used for types) of validation types, as will be discussed later. In this work, we follow the classification discussed by Barlas (1996).



References

AHRWEILER P and Gilbert N. Caffe Nero: the evaluation of social simulation. *Journal of Artificial Societies and Social Simulation* 2005; 8(4); 14. <http://jasss.soc.surrey.ac.uk/8/4/14.html>

ANDERSEN D F, Richardson G P and Vennix J A M. Group model building: adding more science to the craft. *System Dynamics Review* 1997;13(2); 187–201.

ARGOTE L. *Organizational learning: creating, retaining, and transferring knowledge*. Kluwer Academic: Boston;1999.

ARGOTE L and Epple D. Learning curves in manufacturing. *Science* 1990;247; 920–924.

BALCI O. Validation, verification, and testing techniques throughout the life cycle of a simulation study *Annals of Operations Research* 1994;53; 121–173.

BARLAS Y. Formal aspects of model validity and validation in system dynamics. *System Dynamics Review* 1996;12(3); 183–210.

BARRETEAU O. Our companion modelling approach. *Journal of Artificial Societies and Social Simulation* 2003;6(1); 2 <http://jasss.soc.surrey.ac.uk/6/2/1.html>.

BARRETEAU O, Bousquet F and Attonaty J M. Role–playing games for opening the black box of multi–agent systems: method and lessons of its applications to Senegal River Valley irrigation system. *Journal of Artificial Societies and Social Simulation* 2001;4(2); 5 <http://jasss.soc.surrey.ac.uk/4/2/5.html>.

BECKER J, Niehovens B and Klose K. A framework for epistemological perspectives on simulation. *Journal of Artificial Societies and Social Simulation* 2005;8(4); 1 <http://jasss.soc.surrey.ac.uk/8/4/1.html>.

BOERO R and Squazzoni F. Does empirical embeddedness matter? Methodological issues on agent–based models for analytical social science. *Journal of Artificial Societies and Social Simulation* 2005;8(4); 6 <http://jasss.soc.surrey.ac.uk/8/4/6.html>.

BOTS P W G and van Daalen C E. Participatory model construction and model use in natural resource management: a framework for reflection. *Systemic Practice and Action Research* 2008; 21; 389–407.

BRENNER, T. and C. Werker A Taxonomy of Inference in Simulation Models. *Computational Economics*, 2007; 30, 227–244.

BRENNER, T. and C. Werker Policy Advice Derived From Simulation Models, *Journal of Artificial Societies and Social Simulation* 2009;12(4); 2 <http://jasss.soc.surrey.ac.uk/12/4/2.html>.

CARSON E R and Flood R L. Model validation: philosophy, methodology and examples. *Transactions of the Institute of Measurement & Control* 1990;12(4) 178–185.

CHECKLAND P and Scholes J. *Soft systems methodology in action: a 30–year retrospective*. Wiley: Chichester, Eng. ; New York;1999.

COYLE R G. *System dynamics modelling: a practical approach*. Chapman & Hall: London;1996.

D'AQUINO P, Le Page C, Bousquet F and Bah A. Using self–designed role–playing games and a multi–agent system to empower a local decision–making process for land use management: the SelfCormas experiment in Senegal. *Journal of Artificial Societies and Social Simulation* 2003;6(3); 5 <http://jasss.soc.surrey.ac.uk/6/3/5.html>.

DALKIRAN E. Scuba diving simulator: Testing real time decision making in a feedback environment. 2006. MSc Thesis; Bogazici University, Istanbul.

DUIJN M, Immers L H, Waaldijk F A and Stoelhorst H J. Gaming approach route 26: a combination of computer simulation, design tools and social interaction. *Journal of Artificial Societies and Social Simulation* 2003;6(3); 7 <http://jasss.soc.surrey.ac.uk/6/3/7.html>.

DUNN W N. A pragmatic strategy for discovering and testing threats to the validity of socio–technical experiments. *Simulation Modelling Practice and Theory* 2002;10; 169–194.

EAC (1996). *FORWARD – Freight Options for Road, Water and Rail for the Dutch*, Final Report. RAND: Santa Monica; 1996

EXELBY M J and Lucas N J D. Competition in the UK market for electricity generating capacity.

A game theory analysis. *Energy Policy* 1993;21(4); 348–354.

FERGUSON N M, Cummings D A T, Fraser C, Cajka J C, Cooley P C and Burke D S. Strategies for mitigating an influenza pandemic. *Nature* 2006;442(7101); 448–452.

FORRESTER J W. *Urban dynamics*. M.I.T. Press: Cambridge, Mass.;1969.

FORRESTER J W and Senge P M 1980. Tests for building confidence in system dynamics models. In: A. A. Legasto, J. W. Forrester and J. M. Lyneis (Eds). *System Dynamics*. North-Holland: Amsterdam; 1980.

FRANK U and Troitzsch K. Epistemological perspectives on simulation. *Journal of Artificial Societies and Social Simulation* 2005;8(4); 7 <http://jasss.soc.surrey.ac.uk/8/4/7.html>.

GILBERT N and Troitzsch K. *Simulation for the social scientist*. Open University Press: Berkshire;2005.

GURUNG T R, Bousquet F and Trebil G. Companion modeling, conflict, and institution building: sharing irrigation water in the Lingmuteychu watershed, Bhutan. *Ecology and Society* 2006;11(2); 36.

HOBBS B F and Kelly K A. Using game theory to analyze electric transmission pricing policies in the United States. *European Journal of Operational Research* 1992;56(2); 154–171.

KÜPPERS G and Lenhard J. Validation of simulation: Patterns in the social and natural sciences. *Journal of Artificial Societies and Social Simulation* 2005;8(4); 3 <http://jasss.soc.surrey.ac.uk/8/4/3.html>.

LE Bars M and Le Grusse P. Use of a decision support system and a simulation game to help collective decision-making in water management. *Computers and Electronics in Agriculture* 2008;62(2); 182–189.

LUNA-REYES L F and Andersen D L. Collecting and analyzing qualitative data for system dynamics: methods and models. *System Dynamics Review* 2003;19(4); 271–296.

MAYER I S, van Daalen C E and Bots P W G. Perspectives on policy analyses: a framework for understanding and design. *International Journal of Technology, Policy and Management* 2004;4(2); 169–190.

MENNITI D, Pinnarelli A and Sorrentino N. Simulation of producers behaviour in the electricity market by evolutionary games. *Electric Power Systems Research* 2008;78(3); 475–483.

MISER H J and Quade E S 1988. Validation. In: H. J. Miser and E. S. Quade (Eds). *Handbook of Systems Analysis: Craft Issues and Procedural Choices*. Elsevier Science Publishing: 1988. p. 527–563.

MOSS S. Alternative approaches to the empirical validation of agent-based models. *Journal of Artificial Societies and Social Simulation* 2008;11(1); 5 <http://jasss.soc.surrey.ac.uk/11/1/5.html>.

MOSS S and Edmonds B. Sociology and simulation: statistical and qualitative cross-validation. *American Journal of Sociology* 2005;110(4); 1095–1131.

PRELL C, Hubacek K and Reed M. 'Who's in the network?' when stakeholders influence data analysis. *Systemic Practice and Action Research* 2008; 21; 443–458.

RAMANATH A M and Gilbert N. The design of participatory agent-based social simulations. *Journal of Artificial Societies and Social Simulation* 2004;7(4); 1
<http://jasss.soc.surrey.ac.uk/7/4/1.html>.

ROTMANS J. *IMAGE: An Integrated Model to Assess the Greenhouse Effect*. Kluwer Academic Publishing: Dordrecht;1990.

ROUWETTE E. *Group Model Building as Mutual Persuasion*. 2003. PhD Thesis; Radboud University, Nijmegen.

ROWE G and Frewer L J. Public participation methods: a framework for evaluation. *Science Technology and Human Values* 2000;25(3); 3–29.

SARGENT R G. Validation and verification of simulation models. Winter Simulation Conference 2004.

SAYSEL A K, Barlas Y and Yenigün O. Environmental sustainability in an agricultural development project: a system dynamics approach. *Journal of Environmental Management* 2002;64(3); 247–260.

SCHELLING T C. Dynamic models of segregation. *Journal of Mathematical Sociology* 1971;1; 143–186.

SCHMID A. What is the truth of simulation? *Journal of Artificial Societies and Social Simulation* 2005;8(4); 5 <http://jasss.soc.surrey.ac.uk/8/4/5.html>.

SOARES-FILHO B S, Nepstad D C, Curran L M, Cerqueira G C, Garcia R A, Ramos C A, Voll E, McDonald A, Lefebvre P and Schlesinger P. Modelling conservation in the Amazon basin. *Nature* 2006;440(7083); 520–523.

STADLER M, Kranzl L, Huber C, Haas R and Tsioliariidou E. Policy strategies and paths to promote sustainable energy systems – the dynamic Invert simulation tool. *Energy Policy* 2007;35(1); 597–608.

STAVE K A. Using system dynamics models to improve public participation in environmental decisions. *System Dynamics Review* 2002;18(2); 139–167.

STERMAN J. *Business dynamics: systems thinking and modeling for a complex world*. Irwin/McGraw-Hill: Boston;2000.

STROUD P, Del Valle S, Sydoriak S, Riese J and Mniszewski S. Spatial dynamics of pandemic influenza in a massive artificial society. *Journal of Artificial Societies and Social Simulation* 2007;10(4); 9 <http://jasss.soc.surrey.ac.uk/10/4/9.html>.

STRUBEN J. Essays on transition challenges for alternative propulsion vehicles and transportation systems. 2006. PhD Thesis; M.I.T., Boston.

TAKADAMA K, Kawai T and Koyama Y. Micro- and macro-level validation in agent-based simulation: reproduction of human-like behavior and thinking in a sequential bargaining game. *Journal of Artificial Societies and Social Simulation* 2008;11(2); 9
<http://jasss.soc.surrey.ac.uk/11/2/9.html>.

TROITZSCH K. Validating simulation models. *18th European Simulation Multiconference* 2004.

WILDRUM P, Fagiolo G and Moneta A. Empirical validation of agent-based models: Alternatives and prospects. *Journal of Artificial Societies and Social Simulation* 2007;10(2); 8

<http://jasss.soc.surrey.ac.uk/10/2/8.html>.

YILMAZ L. Validation and verification of social processes within agent-based computational organization models. *Computational and Mathematical Organizational Theory* 2006;12; 283–312.

YÜCEL G and Barlas Y. Pattern-based system design/optimization. *25th International System Dynamics Conference* 2007; Boston, USA.

ZAGONEL A. Model conceptualization in group model building: a review of the literature exploring the tension between representing reality and negotiating a social order *International System Dynamics Conference 2002*; Palermo, Italy.

ZEIGLER B P. *Theory of modelling and simulation*. Wiley: New York;1976.

ZHAO H. Simulation and analysis of dealers' returns distribution strategy. *Winter Simulation Conference* 2001.