Chapter 7
System Models for Policy Analysis

Warren E. Walker and C. Els van Daalen

7.1 Policy Analysis Models

Although quantitative system models are only one of many tools of a policy analyst, they are an important tool. For the policy analyst, the purpose of building and using models is to estimate things that cannot be observed or measured directly.[1] The prime example is impact assessment—estimating the outcomes of a policy that a decisionmaker may consider adopting. Other uses are diagnosis (estimating what factors have the greatest leverage to change a specified outcome or what is the primary source of a given outcome) and forecasting (estimating how a variable is likely to evolve in the future, usually assuming "present trends"). They also may be used as learning tools (to gain an understanding of how the system works, or may work in the future).

Policy analysis models are fundamentally different from most other types of models that scientists and engineers build. Scientists and engineers usually build models to try to obtain a better understanding of one portion of the real world. The better the match between the model and the real world, the better the model is considered to be. Scientific and engineering models can be validated using empirical data. By contrast, policy models are built to provide information to policymakers who are trying to develop policies intended to solve real world problems, usually for a future situation. They are designed to give policymakers information that can help them develop insights into their problem situation and on

[1] Merriam-Webster's Collegiate Dictionary, 10th edition, 1998, includes the following among its many definitions of model: A description or analogy used to help visualize something (as an atom) that cannot be directly observed.

W. E. Walker (✉) · C. E. van Daalen
Faculty of Technology, Policy and Management, Delft University of Technology,
P.O. Box 5015, 2600 GA, Delft, Netherlands
e-mail: w.e.walker@tudelft.nl

W. A. H. Thissen and W. E. Walker (eds.), *Public Policy Analysis*,
International Series in Operations Research & Management Science 179,
DOI: 10.1007/978-1-4614-4602-6_7,

which they can base their policy decisions. The models serve as laboratory environments, to test alternative policies and compare their performance without having to actually implement them to see how they would perform.[2]

Many different modeling methodologies are available to the policy analyst. Greenberger et al. (1976) describe modeling methodologies to support policymaking that developed over the course of the twentieth century. An important source of policy models is the field of Economics (e.g., linear and statistical economics), with e.g., input–output analysis, game theory, cost–benefit analysis, and econometric modeling. Another branch of modeling originates more from the mathematical/physics fields (Operations Research, System Dynamics, Agent Based Modeling). With the arrival of faster and cheaper computing power during the 1960s, model development and use became widespread and the models became larger and more complex.

The quality of a policy analysis model is judged not by how accurately it reflects the real world, but by how well it is able to provide information that enables a decisionmaker to make knowledgeable choices among policy options—i.e., how well the model can help construct and defend an argument about the relative pros and cons of alternative policy options. A relatively crude model that can clearly demonstrate that alternative A performs better than alternative B under both favorable and unfavorable assumptions will probably lead to a better decision than a complex model that can perform only a detailed expected value estimation.

Policy analysis models tradeoff rigor for relevance. In many cases they are intended to be used for screening large numbers of alternative policy options, comparing the outcomes of the alternatives, and/or designing strategies (packages of policy options). This means that they should include a wide range of factors (e.g., technical, financial, social), but not a lot of detail about each of the factors. The outcomes are generally intended for comparative analysis (i.e., relative rankings), so approximate results are sufficient. They must provide sufficient information to map out the decision space—the ranges of values of the various input parameters (policy variables and scenario variables) for which each of the various policy options would be preferred. Implementation planning, engineering, and scientific models are needed for examining fewer alternatives according to a smaller number of factors. However, they are used in situations where absolute values are needed (e.g., numbers of vehicles, kilograms of NOx emitted, etc.), which requires more accurate estimation of the results for each factor (and more fully validated models). Therefore, designing a policy analysis model is a balancing act. There is a tradeoff between breadth and depth. Adding too much depth is a pitfall in developing a policy analysis model. Instead of aggregating, approximating, and simplifying, the modeler includes every factor that s/he thinks might have an influence on the results. But to make the model manageable, the boundaries are pulled in (reducing breadth). When the boundaries are too narrow, the model cannot address all the relevant issues. Figure 7.1 illustrates the scope and level of detail of a policy analysis model in relation to that of a scientific or engineering model.

[2] Of course, engineering design models are built for similar purposes.

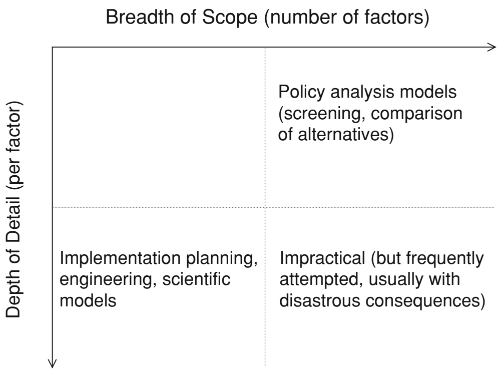

Fig. 7.1 Different types of models have different scopes and levels of detail

A policy analysis model is developed to analyze policies that have not yet been chosen. They have not been (and may never be) implemented, and the impacts cannot be observed directly. Often, though, the theory is suspect, the data have much variation, and even the design of the policy is uncertain. Under such circumstances, it makes no sense to expect to estimate the impacts accurately. Instead, the analyst can use a model to explore the issue (see Chap. 9). A scientific or engineering model will almost always attempt to provide a single estimate of an outcome, perhaps with an error band. A policy model needs to support exploration of possibilities instead of only point predictions.

Once the analyst makes an estimate or draws a conclusion, s/he must persuade the decisionmaker and/or other audiences that it is credible. For this purpose, the model cannot be a “black box”; it must tell a story about how things work in the relevant portion of the world. It must express a set of logical relations, cause-and-effect mechanisms on which to base inferences. The role of a policy analysis model can be shown by looking at the location of the model within the policy analysis framework, which is shown in Fig. 7.2.

A policy analysis system model is a model of the “system domain for policies”. A system model is developed to provide the policymaker(s) and other stakeholders with information about the way the system works presently and to explore the possible consequences of implementing different policies under different future circumstances, which is usually impossible to test in a real situation. The role of such a model is schematized in Fig. 7.3. A policy analyst will first investigate a system as it currently operates (the “base case” or the “validation case”). Following this, the system will be investigated in different future conditions (“reference cases” under different scenarios), and for each future condition, many possible policy changes will be explored (“policy cases”).

Box 7.1 describes an example of a policy analysis model that has been used to investigate different policy options under different possible future conditions.

A classical example of a policy analysis model is a model that can be used by a policymaker to analyze the consequences of a certain change in a physical system, e.g., adding another runway at an airport. This would have an impact on the number of airplanes that can be accommodated and the noise that is produced, but

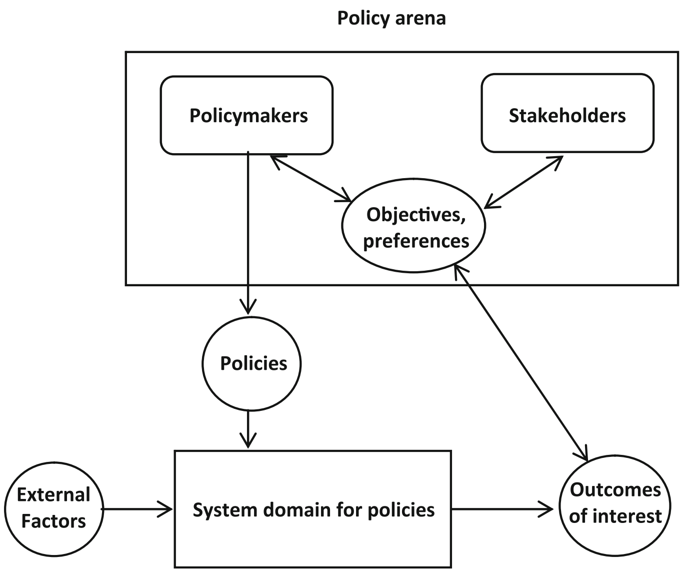

Fig. 7.2 A framework for the rational style of policy analysis

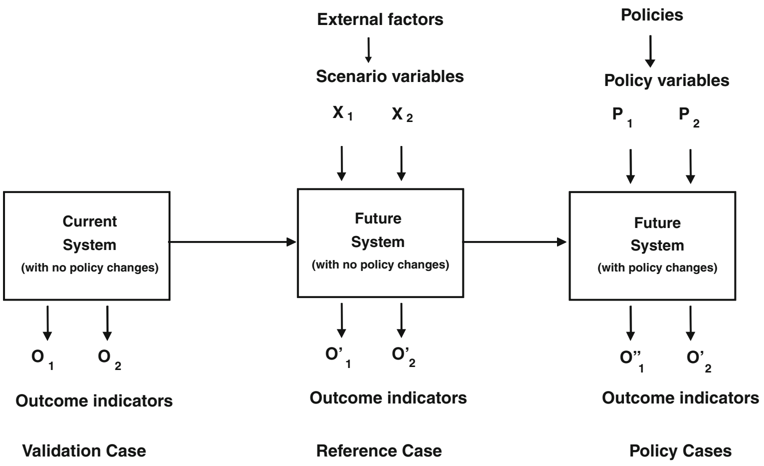

Fig. 7.3 The role of a system model in policy analysis

it will also have socioeconomic consequences a policymaker will be interested in, such as safety, health, and environmental consequences, and costs. The models that can be used for such an analysis will be termed physical system models here, since they consider the policies that impact a physical system. For certain aspects of such a situation, there may specific types of models (e.g., risk and safety) from specific domains; other aspects may require a more general model. A good policy analysis model will allow a decisionmaker to take into account the interests of other stakeholders. The modeler should try to accommodate all relevant outcomes of interest. The outcomes of the models are the criteria that the decisionmaker will use as a basis for decisionmaking, and that the other stakeholders will use to compare the policies being considered. The choice of a preferred policy can be made by weighing the outcomes by their relative importance. In order to select the appropriate method for this type of model, the modeler will have to consider the types of *outcome indictors* that are relevant in the specific situation, and the *characteristics* of the system that is being modeled.

Box 7.1: Example of a policy analysis model for future freight transport in the Netherlands

At the end of the 1980s, the Dutch Government realized that the rapid growth of road freight transport was leading to significant increases in congestion, pollution, and other disbenefits to society. As a result, a broad study was commissioned. This analysis of Freight Options for Road, Water And Rail for the Dutch (FORWARD) was carried out by RAND Europe. It examined 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 (Hillestad et al. 1996). The study included the development of a comprehensive policy analysis model called PACE-FORWARD, which was used to evaluate the performance of a large number of policy options for several economic scenarios extending to the year 2015. In this model, the major modes of inland freight transportation are represented: road, inland shipping, and rail. The model allows the user to choose a policy option and a scenario. It then estimates a wide range of impacts, including the effects of the policy on vehicle emissions, noise, safety, congestion, and the national economy. Equations and data from a number of sources were used to estimate the various impacts. Although the impact modules come from different sources, the architecture provides a structure within which they function together, using consistent assumptions and a common database. The user chooses the impacts to be displayed and how they are displayed. The results are given as percentage changes from a reference case. Results are displayed graphically and in "scorecards". To run the model for a single policy option takes a few seconds on a PC, so the model provides the user with a way to quickly estimate the performance of many policy options as part of the process of formulating a policy. For further information about PACE-FORWARD, see (Carrillo et al. 1996).

7.1.1 Outline of the Remainder of this Chapter

This chapter discusses both building policy analysis models and how to use them. First, the general life cycle of a model is explained. This life cycle applies to both physical system models and actor models. Following this, we concentrate on physical system models (see Fig. 7.2) in order to identify different types of models and their associated modeling methodologies. The chapter closes with some guidelines for the modeler.

7.2 The Life Cycle of a Model

The general life cycle of a model applies to all types of models. There are a variety of ways of specifying the life cycle of a model (see, for example Robinson (2004) and (Balci and Ormsby 2007)). However, in one way or other, they all include the same basic phases:

Planning: Decide on the model's objectives and what is to be estimated. Planning includes defining the system boundaries and selecting the outcome indicators.
Design: Determine the level of aggregation and general form of the model and specify the details to make it relevant to your particular needs.
Implementation: Represent the model in a way that can be executed by the computer.
Calibration and Validation: Build confidence in the model and identify the questions it will be able to address.
Employment: Make use of the model to further the policy analysis.
Documentation: Explain what the model does, how it does it, and why (and to what extent) its results ought to be trusted.

Each of these phases is discussed in more detail below.

7.2.1 Planning

First and foremost, the purpose of models is to help answer the questions you want to ask, and planning can help make this happen. Early in the project, you may not have a clear idea of the specific questions you are going to use the models to help answer. A flexible plan, with due allowance for contingencies, can ensure that you have the time and resources to clarify your ideas. A plan can also keep you from focusing on part of the problem to the exclusion of the remainder. Note that, we speak of models in the plural. It is not necessary, and usually not desirable, to build a single large model that will address all the issues. Instead, think of building a toolkit of many small (but integrated) models (or "sub-models"). Small models have numerous advantages over large models.

A policy analysis model must address the information needs of its users. The first step in designing a policy analysis model should be to assess these needs. Models need to be able to help the policy analyst answer the questions that are asked. This seems obvious, but is harder than it may appear. One problem is that at

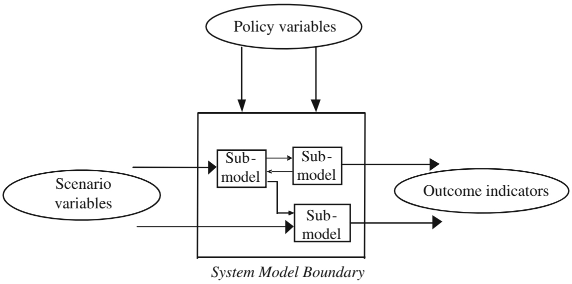

Fig. 7.4 Model diagram with sub-models

the start of model building the policy analyst may not have a clear idea of the specific questions the model will be used to help answer. Take an air quality study, for example. The analyst will clearly need to estimate concentrations of air pollutants from emission rates followed by an estimation of the effects of alternative policies on emissions. So the analyst may set out to build those models. But when the time comes to estimate the outcomes of interest, additional estimations of the financial, social, or health impacts could be needed. This could require building additional models, because the original models were not designed to investigate these aspects.

In order to set the boundaries of the model and to determine exactly what the model should be able to calculate, it is advisable to set up a model diagram. A model diagram is a based on a system diagram (see Chap. 4 and the Appendix). The system diagram is a representation of reality; the model diagram is a representation of the model(s) that will be used to calculate the outcome indicators for the outcomes of interest. In particular, the model (or models) that have to be developed are identified. In addition to system data, the inputs to these models will be policies and exogenous variables.[3] The exogenous variables are variables that the problem owner cannot influence. These may be derived by developing and quantifying scenarios (see Chap. 9). A schematic representation of a model diagram is shown in Fig. 7.4.

Based on the model diagram, the modeler should be able to draw up a set of questions that the model should be able to answer. The questions should include information that is desired to be known about the outcome indicators, which will help in specifying the scope of the model to be built and in selecting the modeling methodology. An example of such a question could be: "What would be the number of takeoffs and landings at a certain airport in the year 2030 for different policies under different external circumstances?".

Table 7.1 indicates questions that should be answered in relation to some of the elements of the outcomes of interest. The final column shows the properties a modeling methodology should possess in order to be of use for estimating the required types of outcomes.

[3] Note that we label the inputs to a model diagram as *variables* instead of *factors* to emphasize the distinction between a model diagram and a system diagram.

Table 7.1 Relationships between Elements of Outcomes and Model Properties

Elements	Question related to element	Resulting model property
Time	What is the time horizon; i.e., what is the length of the time axis (seconds, months, years)?	Operational or strategic
	Are values of the outcomes needed over the whole period or are only the final values needed?	Dynamic or static
Space	Is the spatial component important?	Spatial or non-spatial
Outcome indicators	What outcomes of interest have to be calculated (e.g., technical, economic/financial, social, environmental)?	Identify submodels needed
	What is the level of detail needed in the outcome of interest?	Micro or macro

The model diagram and the modeling questions will also help the modeler to identify the boundary of the model. It is necessary to decide in advance what will and will not be included in the model. The modeler should be very explicit about the boundary and document it well.

Planning should also consider the elapsed time and person-years of effort it will take to build the models, collect the data, etc. A short study with a small budget implies a "quick and dirty" methodology. During the planning phase the analyst should also explicitly consider the way in which the policymaker and other stakeholders will be involved, since it is essential to involve the relevant stakeholders during the modeling process. Recently, a great deal of emphasis has been placed on what is called "collaborative modeling" or "participatory modeling" (see, for example (EWRI 2011)). These approaches help stakeholders with differing perspectives to integrate their interests into the model, and help to gain their acceptance of the model's results. The model enables them to see "the problem" from their own perspective as well as from others' perspectives. The process of working together on a model keeps the focus on "getting the model right", which reduces the focus on personal conflicts. The experience helps to reconcile the facts and to clarify assumptions, while building trust in the policy analysis process and in the model, and helps the participants to build a shared language and to identify and define areas of agreement and disagreement.

7.2.2 Design

In this phase, the model's structure is specified—the equations and other formalisms that establish the relationship between inputs and outputs. There are many ready-made structures, arising from a variety of disciplines (see Sect. 7.3). You can select one of them and modify it for your own purposes. Or you can formulate your model from scratch.

Formulating a model is a balancing act. Detail can be seductive. Instead of aggregating, approximating, and simplifying, you include all the factors you think might have an influence on the results. But then the model becomes enormous, and to make it manageable you may be forced to pull in the boundary (reduce breadth). If you make the boundary too narrow, your model cannot address all the issues of concern.

Most policy situations are so complex that it is easy to become overwhelmed by the "curse of dimensionality". That is, there are so many possible policy options, so many plausible scenarios, and so many outcomes of interest, that it would be difficult to evaluate the complete range of outcomes for each option and several scenarios. One way to deal both efficiently and effectively with this situation is to use a fast policy analysis model to gain insights into the performance of the policy options. A more detailed model might then be used to obtain more information about the performance of the most promising options. Assessments based on the fast policy analysis model, therefore, would be considered as first order approximations in policy discussions. When a promising policy has been identified using the fast model, it will often be necessary to conduct further detailed planning and research in which full account can be taken of the specific circumstances and characteristics of the problem.

There is no requirement that a policy analysis model need be an aggregate version of a more detailed model that might be used later. In fact, because it is fast, it can contain features that would be impossible to include in a high-resolution model. High-resolution models must be limited in scope, lest they become as unwieldy as to be useless. Also, they are intended to be used for different purposes, so their outputs will be different. For example, a transport policy analysis model might have impact assessment submodels for estimating not only the effects of changes in policies and/or changes in scenarios on transport demand (which is often the focus of high-resolution models), but their effects on the national economy, regional economies, land use, and the environment. The tradeoff between depth and breadth, as discussed in the introduction to this chapter, is extremely relevant here (see Fig. 7.1).

The specific uses of policy analysis models imply that they have specific design requirements. This means that, in most cases, the system models should be designed so that:

- it is easy to represent policy changes in terms of policy variables (variables that the models recognize)
- it is easy to change the policy variables (e.g., they are not hardwired into the models)
- it is easy to represent external conditions in terms of scenario variables and structural changes to the system
- it is easy to change scenarios
- submodels are included for estimating outcome indicators for the outcomes of interest.

These design requirements imply the need for a user-friendly graphical user interface—a “policy cockpit” that allows the user to easily specify and examine the results of different policy choices and policy contexts. A key criterion for judging the usefulness of a policy analysis model is its ability to facilitate the exploration of policy and scenario space (i.e., examine a wide range of policy options for a wide variety of scenarios).

An important part of the design phase is to choose a form for the model or models. Examples of modeling methodologies are System Dynamics and discrete event simulation. Specific modeling methodologies have inherent assumptions about the way a system works. One assumption in System Dynamics, for example, is that the time dimension (and, therefore, the variables in the model) is continuous (i.e., not discrete). The issue of selecting a modeling methodology will be dealt with in more detail in Sect. 7.3.

7.2.3 Implementation

This phase involves developing or acquiring the algorithms for computing the outputs from the inputs and implementing them in computer programs. It also includes the mechanics of feeding the inputs to the programs and collecting the outputs as they are generated.

During the implementation phase, the model is represented in such a way that it can be executed by the computer. If a policy analyst chooses a specific modeling methodology, then there are usually specific software tools that facilitate the representation of these models in the computer. An advantage is that these models can usually be built in a relatively short time. The modeler does need to be aware, however, of the (implicit) assumptions related to the tool and the methodology. A policy analyst can also choose to develop a model using generic software (e.g., a spreadsheet). This means that the model has to be developed from first principles, but there are no implicit assumptions.

Some software tools developed over the last 25 years that have made it much easier to build small models include:

- IThink/STELLA, Powersim, and Vensim, for System Dynamics models
- Arena, Promodel, Automod, and Simul8, for discrete event simulations
- General algebraic modeling system (GAMS), for formulating linear and nonlinear programs
- LISREL, for (statistical) structural equation modeling
- Statistical analysis system (SAS) and SPSS, for statistical analysis
- Analytica and Maple, for models consisting largely of algebraic equations
- Excel, for spreadsheet models
- Access (a relational database tool)
- Matlab (a generic tool for mathematical modeling and visualization)
- Python, a high-level programming language that can be used to integrate (or “glue”) existing tools and/or models together.

With the advent of off-the-shelf applications such as these, the architecture of models has changed. Pre-processing (the development of input data) and post-processing (producing data for graphs and tables) used to be programmed as part of the model. Now the model is run and its outputs are dumped into a file. Then the outputs are analyzed using, for example, a standard statistical package. This separation allows more flexibility. Another example is providing scenario (input) data to a System Dynamics model by means of Excel.

Implementation of a model requires quantification of the model's parameters. Determining the parameter values for a policy analysis model is very difficult and time consuming. In a sense, the modeler is not building a representation of the real world, but a representation of observations on the real world. Some of those observations have already been boiled down into a theory (e.g., Newton's laws of motion) that can be incorporated into the representation. Some of them exist as data. But the relationship between theory and the real world, or data and the real world, can be complex.

Data are colored by the circumstances of their collection. Somebody had to decide what to collect, and chose things that seemed important. If the data are collected routinely for administrative purposes, for example, the data elements will be those that the administrator needs, such as employee charges, records showing compliance with regulations, billing records, etc. If a particular data element is costly to collect, chances are it will be collected only occasionally, or a proxy variable (something assumed to be highly correlated with it) will be collected instead. If the people that actually collect and record the data see no benefit for their own jobs, or worse, see a threat (e.g., the manager is checking up on them), the quality of the data is suspect. So, in a sense, there is a model between the real world and the data, which is an additional source of uncertainty in the model's results, even for the base case (see Chap. 9). Experimental data (e.g., clinical trials) will be cleaner. Collecting (and subsequently analyzing) the data is the whole point of the experiment, so attention will be paid to it. But an experiment is controlled. Many factors will be held constant or their range of variation limited. Another way of obtaining data is to use results of more detailed scientific models. The data will then not be based on observations on the real world, but on observations on the output of another model. The resulting model is sometimes called a 'meta-model'.

7.2.4 Calibration and Validation

We want two things from a calibrated and validated model—credibility and the power to predict. But we generally cannot validate our models in this strict sense. *Calibration* establishes the values of otherwise undetermined parameters in the model, using the criterion of a good fit with historical data. There is, however, a great deal of contention about what *validation* is. The most extreme view is that validation should establish that estimates provided by the model are "the same" within specified limits as their real world values would be—a hard thing to do if the model will be used to estimate "things that cannot be directly observed". If we adopt this view, practically no policy analysis model can be validated. A more modest aim is to establish that the model is well grounded on principles or evidence, or able—when considered as an argument or story—to withstand criticism or objection.

We suggest that validation should be the practical exercise of deciding how the model can be used. Just as a rope may be unable to bear much weight and yet be useful, so a model may be fraught with uncertainty and yet provide important support to an argument. The model may provide a bound—a worst case—that can be used in an *a fortiori* argument. It may provide comparisons you can trust (A exceeds B) without providing good absolute estimates. It may rule out some possibilities. In all these cases, the model can help the analyst construct and defend an argument.

The process of modeling can be viewed as one of narrowing possibilities. One is trying to rule out things that cannot happen. If a great deal about the target system is known, it may be possible to narrow things down to a very small range of uncertainty (i.e., things that cannot happen). If a lot less is known, the remaining uncertainty will be greater. "Validation" is an attempt to describe the remaining uncertainty.

7.2.4.1 Calibration

Calibration can be used to derive parameter values from historical data. This establishes the values of otherwise undetermined parameters in the model, using the criterion of a good fit with historical data. Calibration measures how well the model fits the *historical* data. But it will not say anything about the correctness of the model. That is, even a very good fit will neither guarantee that the correct causal relationships have been identified—the points of leverage for the decisionmaker—nor that it can be used to say anything useful about future outcomes. The same data should not be used for both calibrating the model and validating it, since the results of the validation will then be meaningless.

Many real world observations do not come in the form of tables of numbers, but as qualitative data, such as textual material (interviews, field notes, or published descriptions) or images (photographs or drawings). This information may point toward very important factors that are hard to quantify. For example, morale and training are considered to have a major impact on the outcome of a battle. The question is how these factors can be taken into account in a model. Dupuy (1987) developed his "quantified judgement model (QJM)" as a way to do so, although it has not been widely adopted. Although it is difficult to quantify these kinds of factors, it is often essential into take them into account in a policy analysis model. Lack of data should not lead the modeler to ignore factors s/he thinks are important. These are often "soft" factors that are hard to quantify, but should be included in some way.

Often, there will not be a large set of data points for calibration. Instead the model will have been assembled from lower level bits and pieces, and the value of each calibration parameter will have been obtained from a different source. That is, the model represents a *system*, but the data describe different parts of the system. One can not know whether the data derived from experiments on an isolated part of the system are valid in the larger system context.

We explain and illustrate calibration and validation using the following notional model. This model calculates an outcome Z as a function of a policy P, a scenario S, and some calibration parameters C. The policy, scenario, and calibration parameters are drawn from a policy space, a scenario space, and a calibration space respectively. That is:

$$\begin{aligned} &Z = f(P, S, C) \\ &P \in Pspace \\ &S \in Sspace \\ &C \in Cspace \end{aligned} \tag{7.1}$$

In calibrating the model, it is important to consider the model's behavior throughout the ranges over which the policy and scenario variables are expected to be varied. In Eq. (7.1), we denote those ranges as *Pspace* and *Sspace*. This requirement is often overlooked when a large set of data points is available and regression is used for calibration. Regression methods fit the model to the data, and give no weight to the behavior of the model where there are no data points. Yet in a policy study the model is often used to extrapolate beyond the range of the data.[4]

There are two classes of extrapolation to consider. First, you may wish to set a variable in the model to a value outside the range it occupies in the data or has occupied in your experience. For example, you might want to examine what would happen if the price of oil tripled, or if the tax on petrol were doubled.

Second, you may wish to consider changes to variables that do not even appear in the model. For example, consider trying to estimate the effect of a new drug on health. A clinical trial is done in which the new drug cures 63 % of the people in the test group, while the standard treatment cures only 27 %. Does this mean we will see a similar improvement in the population at large? Not necessarily. First, the patients in the trial generally won't represent the full range of people in the population at large. Everybody in the trial will be between 25 and 40 years of age, with no allergies and no comorbidities. Studies have shown that subjects in a clinical trial are more likely than the average patient to comply with the drug regimen specified by the physician (compliance rates are so variable that there is no typical rate, but 50 % is as good a guess as any). When the drug is released for general use, physicians at

[4] This can be a point of contention between the policy analyst and the academic researcher. The purpose of an academic study, after all, is to find the truth of the matter. Extrapolation is mere speculation, and is generally frowned upon. The purpose of a policy study is to decide what to do next, and the analyst does not have the luxury of waiting until the truth is known with reasonable certainty. Extrapolation is necessary.

large may not prescribe it for precisely the same conditions as the physicians running the trial. In other words, the carefully controlled conditions of the trial will not be replicated when the drug is released for general use. When we try to predict the effect of the drug on the general population, we must have a way to extrapolate for changes in the factors that were held constant in the trial. Ensuring that the model extrapolates reasonably well requires that the right kinds of features are built into the model during its formulation.

7.2.4.2 Validation

There is a substantial literature dealing with the classical view of validation, especially of simulation models (e.g., Law and Kelton 1991; Kleijnen 1999). In this view, validation should demonstrate that there is some value for C, the calibration parameters, for which the model agrees reasonably well with reality. That is, for some specified bound B:

$$|Truth(P,S) - f(P,S,Cbase)| < B \quad \forall P \in Pspace, S \in Sspace \tag{7.2}$$

Of course, the bound B must be small enough for the purposes of the study in which the model will be used, or the validation can hardly be counted a success. In addition, the policy and scenario spaces, *Pspace* and *Sspace*, must be rich enough to contain the ranges of policies and scenarios of interest in the analysis.

Validation in this strict sense is hardly ever possible, though some models (or theories) in physics come close. Newton's law of universal gravitation plus his three laws of motion constitute the basis for estimating an enormous range of things that are not measured directly. Even this model, however, has a limited range of validity. Extrapolations of Newton's laws of motion to near light speed are very wrong. Moreover, an engineering model based on Newton's laws may be invalid even though the laws themselves are nearly perfect. Friction, for example, may be dealt with by crude approximations. Notwithstanding these approximations, many engineering models have been validated in the classical sense for specific uses.

However, if the classical view of validation is adopted, no policy analysis model could ever be validated. We can validate models only if the situation is observable and measurable, the underlying structure is constant over time, and the phenomenon permits the collection of sufficient data (Hodges and Dewar 1992). It is the requirement to extrapolate beyond the data that makes validation in the classical sense so problematic for policy analysis models. The reason extrapolation makes validation problematic is that the bound B in condition (7.2) becomes large. For most policies and scenarios you can have no confidence that the model matches reality within a usefully small error.[5] A more modest aim is to establish that

[5] This theme is developed in Bankes (1993), Hodges (1991), and Pilkey and Pilkey-Jarvis (2007).

the model is well grounded on principles or evidence, or able—when considered as an argument or story—to withstand criticism or objection.

This does not render models useless for policy analysis. But it does influence the way models can be used. Models have traditionally been used to predict. But when classical validation is impossible, prediction ceases to make much sense. Instead, a model can be used to explore possibilities and investigate hypotheses—in a word, to develop insight. Skeptics often see this as a cop-out; the model can't do the real job, so the analyst has to invent a justification after the fact for having spent so much time and effort constructing the model. A better way of looking at the issue is to note that, although you cannot build a model that can be classically validated without well-nigh-complete information, there is normally a lot of information, knowledge, and data available that can be used to inform decisionmaking. A research methodology called exploratory modeling and analysis (EMA) aims at utilizing the available knowledge and data by specifying multiple models that are consistent with the available information. Instead of building a single model and treating it as a reliable representation of the information, an ensemble of models is created and the implications of these models are explored. A single model run drawn from this set of models is not a prediction. Rather, it provides a computational experiment that reveals how the world would behave if the assumptions any particular model makes about the various uncertainties were correct. By conducting many such computational experiments, one can explore the implications of the various assumptions. EMA aims at offering support for exploring this set of models across the range of plausible parameter values and drawing valid inferences from this exploration (Bankes 1993; Agusdinata 2008). From analyzing the results of this series of experiments, analysts can draw valid inferences that can be used for decisionmaking, without falling into the pitfall of trying to predict that which is unpredictable. (For further discussion of the use of models for exploratory purposes, see Sect. 9.3.4).

Although strict validity cannot be determined, it is necessary to build confidence in whatever model (or models) are being used. This is done by carrying out a variety of tests. As the first step, it is important for credibility that during the development of the model (or models) all information that is available from a variety of sources is taken into account, including observations, general knowledge, theory, and experience/intuition (Van Horn 1971; Law and Kelton 1991), and that people who are knowledgable about (parts of) the system under study and policymakers are involved throughout the modeling process (Law and Kelton 1991).

A model used for policy analysis cannot be a "black box"; it must tell a story about how things work in the relevant portion of the world. It must express a set of logical relationships—cause-and-effect mechanisms—on which to base inferences. If a model is used to extrapolate the historical data (which is often the case in policy analysis), then the data count for less and the form of the model counts for more. This means that a test in which a comparison is made between model and real system output is not sufficient, and a wider variety of tests is required.

Barlas (1996) distinguishes three types of tests for the assessment of System Dynamics models, but these tests can be extended to policy analysis models in general: direct structure tests, structure-oriented behavior tests, and behavior pattern tests. The last type of test relates to comparing model behavior to the behavior of the system that has been modeled, and the first two investigate the internal structure, or form, of the model. During the validation phase, the model structure is studied first and the model behavior is studied only when the structure is considered to be adequate.

In a direct structure test (Barlas 1996), the model is investigated without running it. Direct structure tests include investigating if equations, parameter values, and/or distributions are consistent with theory and/or available data. Additional tests include checking if equations are robust even for extreme input values, and carrying out a formal inspection or walkthrough.

In a structure-oriented behavior test, the model is run and, by investigating outputs, the structure is studied indirectly (Barlas 1996). One such test is an extreme condition test for which extreme values are entered and the behavior of the model as a result of these values is investigated. In a transport emission model, for example, one could test what would happen in a situation in which there are no vehicles on the road and in which there is an extremely large amount of transport. A sensitivity analysis is also a very important structure-oriented behavior test. If the output is sensitive to a part of the model, then that part requires careful modeling (Law and Kelton 1991), as it is important to the behavior of the model. The more coarse tests are carried out before more detailed tests. For example, an extreme condition test would be carried out before a sensitivity analysis.

The third type of test, a behavior pattern test (Barlas 1996), entails determining if the model output adequately represents the relevant system behavior. The model is tested as a whole and requires an existing system similar to the one modeled (e.g., the current version of the system whose future performance is to be estimated), and seeing whether the model adequately reproduces its outcomes. (This is the "base case" shown in Fig. 7.3). When carrying out a quantitative comparison, classical statistical tests cannot be applied directly, because model and system output are often non-stationary (Van Horn 1971; Law and Kelton 1991). Sterman (2000) and Law and Kelton (1991) describe approaches for quantitative comparison of model output and system output. A further test to investigate if the model adequately represents system behavior is a variant of the Turing test in which experts are presented with model output and system output without knowing the origin, to see if they can distinguish which is which. The above tests relate to comparing model and system output. However, rather than focusing on achieving a good fit with historical data, the validation of a policy analysis model should focus on determining if the model can be used for the purpose for which it has been developed, which is usually to investigate the behavior of a future system.

In many cases, the information that exists for building a model is insufficient to specify a single model that accurately describes system behavior. In this circumstance, models can be constructed that are consistent with the available information, but such models are not unique. Rather than specifying a single model and falsely treating it as a reliable image of the target system, the available information is consistent with a set of models, whose implications for potential decisions may be quite diverse. A single model run drawn from this potentially infinite set of plausible models is not a "prediction"; rather, it provides a computational experiment that reveals how the world would behave if the various guesses any particular model makes about the various unresolvable uncertainties were correct. One use of EMA (see Sect. 9.3.4) is to explore the set of plausible models and examine the implications of the resulting computational experiments. Used in this way, EMA can be understood as searching or sampling over the ensemble of models that are plausible given a priori knowledge. (This use of EMA is described more fully by Kwakkel et al. (2010)).

Sometimes model results can be compared to other models that have been developed in the same or similar fields, and there may also be other models that represent part of the system in more detail, where certain variables can be compared. System experts can also be involved in reviewing the model output for the future situation. However, care should be taken with this, since the reason for building a policy analysis is that it is not known what output to expect from the future system (Law and Kelton 1991). It may also be desirable to carry out complementary research to further increase confidence in the model by investigating model results outside the computer context—for example by conducting experiments or carrying out a field-test (Van Horn 1971).

As explained above, policy models are inherently unvalidatable. Hodges (1991) calls these "bad" models. But, he spells out the "six (or so) things you can do with a bad model". The main point here is that any particular model can be used for specific purposes. These purposes should be made clear by the modeler, and use of the model should then be limited to these purposes.

7.2.5 Employment

The findings during the validation phase will limit the ways in which the model should be employed. A policy analysis model will have to be run numerous times, because the reference case and alternative policies will have to be investigated under different external future conditions (i.e., for different scenarios). A model can estimate the absolute values of the indicators, or their values relative to a baseline or reference case. Estimating relative differences from a reference can be more useful and reliable (for analytical purposes, the policy analyst usually cares mainly about how alternatives compare with each other and with the reference case).

Traditional policy analysis assumes that models will be used for prediction (and, sometimes, even for optimization). One standard approach is cost–benefit analysis, for which the analyst calculates all costs and benefits in monetary terms, and selects the policy with the highest excess of benefits over costs. Another approach is cost-effectiveness analysis, where the various effectiveness measures are estimated in their natural units (e.g., number of fatalities, tons of CO_2, etc.). Uncertainty, more than any other circumstance, constrains the proper use of a model. A model with very little uncertainty can be used to accurately estimate the impacts of a policy (prediction), or compare many policies to find the one with the most favorable impacts (optimization). A model with large uncertainties in the relationships or values of the parameters, however, should not be used for either purpose. In this case, even if a model estimates that Policy A has a better outcome than Policy B, one cannot be confident that things will turn out that way in reality.

Traditional policy analysis does recognize the existence of uncertainty, but it assumes one can deal with it within the prediction/optimization paradigm. Two common ways of dealing with uncertainty using predictive models are (1) incorporate uncertainty into a utility function, and (2) find bounds in the outcomes of interest through the use of sensitivity analysis (see Box 7.2).

The traditional method of employing a model runs the model only a small number of times. Exploratory modeling (Bankes 1993) is a method for employing a model that calls for hundreds, even thousands of runs. You may wish to locate really bad regions of the input space—places you want your policy to avoid. You may want to design a robust policy—one that copes reasonably well with a range of possible futures. Or you may want to design an adaptive policy—one that leaves options open. These alternative (non-traditional) ways of using policy analysis models to design policies are described in detail in Chap. 9.

Box 7.2: Example of the use of sensitivity analysis

Caulkins et al. (1997) used sensitivity analysis to generate a figure like Fig. 7.5. The analysts built a model to estimate the reduction in kilograms of drugs consumed per million dollars invested in one of three policies, "More Treatment", "Longer Sentences", and "More Arrests". The figure maps the regions where each of the policies performs best as a function of the values of two key calibration parameters, the cost of making an arrest and the value to a drug dealer of avoiding a year in prison. (The other calibration parameters were all held at their baseline values.)

In the Caulkins study, the analysts made no attempt to develop probability distributions. Instead, they estimated baselines for the scenario *Sbase* (shown by the star). For each policy they ran their model for many scenarios *S* in a neighborhood of *Sbase*. They made the neighborhood large enough so that in their judgment it included all scenarios that were reasonably likely. They observed that the baseline assumptions lie well inside the region where the "More Treatment" policy is the best. They argue that it is unlikely that the true assumptions will be enough different from the baseline to make the "Longer Sentences" policy preferable. They are less sure that "More Treatment" is truly better than "More Arrests." It overstates the conclusions of Caulkins et al. only modestly to say that they regard the "More Treatment" policy as very likely to be the choice that performs best on this impact.

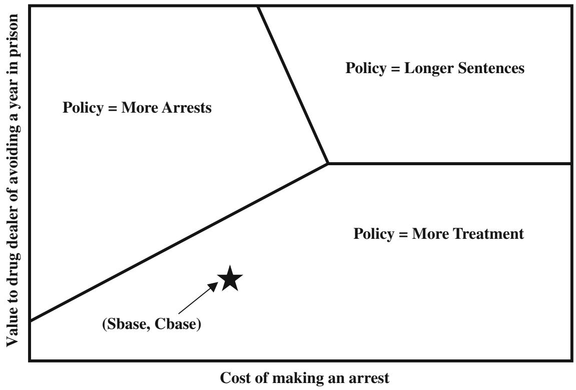

Fig. 7.5 Which program is most cost-effective at reducing drug consumption?

7.2.6 Documentation

There are basically three types of model documentation:

(1). *Executive summary:* This type of documentation motivates the model and describes it in non-technical terms. It is a concise description to help policymakers and other stakeholders understand how the model can be used for policy analysis. It buttresses the model's credibility, and is always necessary.
(2). *Users' manual*: This type of documentation gives complete instructions for collecting data and operating the program. It also presents the mathematical details underlying the model's calculations.
(3). *Program description*: This type of documentation is designed primarily for computer programmers. It includes file specifications, installation instructions, etc. It is useful for maintaining and modifying the model. It should discuss sources of calibration data and the details of whatever formal validation exercises were attempted.

If the model is small and won't be used in any follow-on projects, the last two kinds of documentation are often omitted. Documentation is very costly. But, it is an essential part of any policy study.

7.3 Tools and Templates for Building Physical System Models

Many different mathematical forms of system models have been employed, arising from different disciplines. Interestingly, this list would have looked pretty much the same 25 years ago as it does now. It is very useful to become familiar and comfortable with all of these model forms. On the other hand, one should be wary of using these tools. Each tool tells a story. It can predispose the modeler to see things in a certain way, and obscure alternative ways of looking at things. This reduces the effort when it is the right thing to do, but it can cost additional effort, or lead to wrong conclusions, when it is not. Ackoff (1974) bemoaned the decline of operations research (OR) as follows:

> "By the mid-1960s most OR courses in American universities were given by academics who had never practiced it. They and their students were text-book products engaging in impure research couched in the language, but not the reality, of the real world. As a result, OR came to be identified with the use of mathematical models and algorithms rather than the ability to formulate management problems, solve them, and implement and maintain their solutions in turbulent environments…[P]ractitioners decreasingly took problematic situations as they came, but increasingly sought, selected, and distorted them so favored techniques could be applied to them".

This is old advice—let the problem determine the tools to be used. The trouble with the advice is that these powerful, sophisticated tools exist, and it would be a waste to ignore them. But it is useful to step back from time to time and question the assumptions that are built into the tools and to ask what might be done instead.

7.3.1 Tools from Operations Research(OR)

Operations Research (OR) provides a long list of templates that you can use for your models. Each template is taught with a motivational story or two that suggests the kind of subject matter that it can best represent. A clever practitioner can apply a template to subjects that seem quite remote from those given in the motivational stories, for it is the form and not the substance that counts.

The templates specify which quantities are to be inputs and which are outputs. So they are much more tailored to asking particular kinds of questions than they are to asking questions about particular subject areas. Here is a partial list from a classic OR textbook (Hillier and Lieberman 2005):

- Linear programming—allocates limited resources among competing activities in the best possible (i.e., "optimal") way.
- Nonlinear programming—like linear programming, except the various functions need not be linear. For example, the cost of producing an item may decrease the more items you produce.
- Integer programming—like linear programming, except activities come in discrete packages. You must buy zero units or one unit; you can't buy half a unit.
- Dynamic programming—systematic procedure for determining the optimal combination of decisions when decisions must be made sequentially and early decisions limit later options.
- Decision analysis—like dynamic programming, it addresses sequential decision problems. But it is simpler and less general.
- Game theory—addresses competitive situations in which multiple players make decisions that affect each other's payoffs.

- Queuing theory—addresses situations in which somebody or something waits in line for a service. For example, you may wish to explore the effect of investing in more capacity on the average waiting time.
- Inventory theory—examines situations in which a stock of items is held to cover uncertain future orders. Typically you might determine the reorder policy that minimizes the expected cost of holding inventory when business happens to be slow plus the cost of having to backorder items when business happens to be brisk.
- Discrete event simulation—a very general approach for studying just about any dynamic system. You describe the system in term of individual events involving basic entities, and run the simulation (operate the system for a period of time) to see what happens. Typically you try to make the simulation model a faithful representation of the relevant aspects of the system, and then do experiments on the model as if it were the real system. Some simulation models are entirely automated. Others have a man-in-the-loop, and can be called games (e.g., war games). Modeling decisions is very hard, and the man-in-the-loop can make the decisions instead of the modeler developing an algorithm to do so.

7.3.2 Tools and Models from Other Disciplines

Not everybody comes to policy analysis with an operations research background. Some come from subject matter disciplines, and in many cases those disciplines have their own preferred model forms. They have theories to tell them what factors are important, what categories and definitions are appropriate, what is a cause, and what is an effect.

Physicists and engineers have traditionally built and used models of differential or partial differential equations, such as meteorological models. These are sometimes called simulations, but *continuous* simulations to distinguish them from the discrete event variety (Seinfeld 1986). Such continuous simulations are being increasingly used by policy analysts to model socioeconomic systems, such as urban areas, water basins, etc.

A generic continuous modeling method that may be used for organizational and socioeconomic problems is System Dynamics (Forrester 1961; Sterman 2000). In System Dynamics, a system is considered in terms of its underlying "flows". For example, flows of people, money, material, orders, and information can be recognized (Roberts 1978). These flows can accumulate in stocks. The stock-flow structure of a system is represented, and simulation of the model generates the behavior of the outcomes of interest over time. The concept of feedback is essential to System Dynamics (Forrester 1961). Feedback loops may have an important influence on system behavior (e.g., positive feedback creating a vicious cycle, or negative feedback that can regulate a system). Feedback can be used in explaining system behavior and in designing policies aimed at influencing system behavior.

Cost analysis and its relatives cost–benefit analysis and cost-effectiveness analysis offer ways to look at costs that are derived partly from economics and partly from accounting. They recommend discounting to compare future costs with present ones, and ways of valuing all sorts of things (e.g., human life and health) in terms of money (Perkins 1994; Gold 1996; Hopkins 1992; Jones-Lee 1976).

Statistics provides a number of model forms. Generally they provide ways to transform data so the undetermined parameters in an estimating relation can be estimated by linear regression. The social sciences rely heavily on these models, as does medical research (Armiger 1995; Morton and Rolph 2000).

Economists study the allocation of scarce resources to alternative uses. Microeconomics (the study of the behavior of individuals—households or firms—when resources are scarce) defines impacts such as consumer surplus (a social welfare impact). Elasticity data let you estimate changes in demand in response to changes in price (Landsburg 2011). Macroeconomics studies the consequences of the behavior of countless individuals for an economy as a whole. Subjects of study include the effects of economic and fiscal policy on inflation, employment, and economic growth (Mankiw 1997).

7.4 Modeling Guidelines

This section discusses a number of important issues that should be addressed by the modeler in the design and development of a model.

7.4.1 Keep the Model Simple and Transparent

One of the most important guidelines for a policy analysis model is that the model must be kept simple and easy to explain. The analyst must keep in mind that s/he is going to have to explain the results and methodology to a policymaker who will generally not be familiar with advanced mathematics. The simpler the model the easier it will be to explain and the better the chance that the policymaker will understand the analysis. As Quade points out: "The most convincing analysis is one that a nontechnician can think through". (Quade 1989, pp. 362–363). It may seem attractive to use a large model. If the policy analysis study intersects an area with a strong academic tradition, there will often be a standard model. These models tend to be conceptually clean and simple, but often lead to very large and data-hungry models when implemented, and may have long execution times. So there is an incentive to look for simpler formulations that will not lead to such large models. But analysts who do so may face the skepticism (even hostility) of academic experts. The analyst must be prepared to demonstrate that the simple model gets approximately the answers a more complicated model would have gotten, and that the accuracy is sufficient for the analysis being carried out and the problem being addressed.

It is important to limit the number of variables in a model, for as the number of variables increases the model becomes larger, requires more data, and is harder to use. This explosion of variables is called “the curse of dimensionality”. It is vital to contain this explosion. There are different ways to do this, which are discussed in the a few of the following subsections.

7.4.2 Aggregate

A model will often have to categorize something, e.g., income levels, ages, ethnicities, locations. As more categories are created, the model’s hunger for data is increased. Too few categories, however, will limit the questions that can be addressed. Often there will be conventional categories that can be used in the model. If there are too many categories, these can be aggregated. To use the example of population, you can keep track of age in 1-year categories, or 5-year categories, or even monthly categories. Replacing 1-year categories by 5-year categories reduces the number of instances by a factor of five. If categories on three dimensions (e.g., age, income, and zip code) are aggregated by a factor of five each, the number of instances is reduced by a factor of $5 \times 5 \times 5 = 125$.

7.4.3 Limit the Depth of the Model

Work at RAND has examined the advantages and difficulties of deliberately building a hierarchical structure into models (Davis and Bigelow 1998). The analyst arranges the variables in a network, in which there is a node for each variable, and a directed link for each causal influence. That is, if the analyst thinks that a change in variable A will cause a change in variable B, other things being held constant, a link is drawn from A to B. Outcome variables are usually put at the top and the factor variables (i.e., variables representing factors that can influence the outcomes) below. Usually the variables fall naturally into a hierarchy, where each variable occupies a level in the hierarchy equal to the number of links between it and an outcome. This expansion of variables, defining them in terms of lower level variables, is what is meant here by adding depth to the model. Clearly, adding depth to a model increases its size. It makes the model more difficult to use for analysis. Fewer cases can be run, so fewer policies can be tested against fewer scenarios. The size of the model can be limited by truncating the hierarchy, and building the model using variables from a relatively high level rather than going clear down to the bottom.

7.4.4 Estimate Only the Outcomes that are Necessary

Outcome indicators are as likely as any other model elements to proliferate. But a policymaker will rarely be interested in all the detail; instead, the modeler will want to estimate summary measures or specific outcomes. In air quality, a modeler might look at the concentration of a pollutant on the worst day of the year, and at average pollutant emissions. In energy, the modeler could look at peak demand and average demand. In a school system, the modeler could look at dropout rates, truancy rates, literacy, and proportion of students held back a year. In a suite of models for strategic planning at airports that was built at the Delft University of Technology (Walker et al. 2003), outcomes were calculated for the peak hours of the 1st, 5th, 10th, 15th, 20th, and 25th peak days of the year, since peak periods determine the airport's infrastructure requirements. Of course, a model can be built to estimate all of the real world observations, which can then be used to calculate summary statistics. (For example, the airport models could have been run to estimate outcomes for every day of the year). But the model will be smaller (and will run more quickly) if it is built to estimate the summary statistics (e.g., peak day outcomes) directly.

7.4.5 Use Meta-Models

On occasion, the analyst may want to build a policy analysis model for a domain in which there is a large model that has become institutionalized (e.g., econometric, transportation, or climate models), although the model itself may be too extensive to use in the analysis. It is possible to build "meta-models" (also called "repro models" or "fast simple models")—small models that reproduce (approximately) the aspects of a large model's behavior that are relevant for the policy issue at hand. Instead of using the large model directly, the large model (called the target model) is treated as the object to be modeled. The meta-model can then be used either as a freestanding model or as a subroutine within another model. This approach allows the analyst to cite the large model as authority (assuming it is trusted), but reap the advantages of smallness. The meta-model may also be extended beyond the circumstances dealt with in the target model, e.g., by extrapolating to data values not considered in the available outputs from the target model, or by adding or reinterpreting variables.

The process of building a meta-model is no different from building any other model. The only difference is that it is a model of a model, rather than a model of a portion of the real world. One formulates and implements it, then calibrates and validates it. Of course, one should not expect the meta-model to agree exactly with the target model, any more than one expects a perfect match of a model with the real world. Box 7.3 describes one such meta-model.

Box 7.3: The demand response model (DRM) in the SUMMA project

> The sustainable mobility, policy measures and assessment (SUMMA) project (SUMMA Consortium 2005) was a policy analysis project carried out by RAND Europe for the European Commission under its Fifth Framework Programme. One of its objectives was to assess transport policy measures for promoting sustainable transport and mobility in Europe. Part of the project's analysis was carried out with the help of a meta-model called the demand response model (DRM). The DRM includes no transport networks. But, it was based on the outputs from the Dutch national model (LMS), the Norwegian national model (NTM-4), The Italian national model (SISD), the Danish National Model, and the Swedish National Model, which are highly detailed models that include network specifications for each of the transport modes.

7.4.6 Beware of Oversimplification

A model should be as small and simple as appropriate, *but no simpler*. The reason the guidelines mentioned above provide ways to control the size of a model is because people tend to build models that are larger and more complex than necessary more than they do the reverse. But beware that there are limits.

Aggregating time steps in a time-stepped discrete event simulation model can produce perhaps the largest aggregation errors of all. It is typical (though not necessary) to assume that rates remain constant during a time step. Vehicles or units move at constant velocities. Supplies arrive at a constant rate. If the size of the time step is doubled, the simulation takes one step for every two it used to take. Instead of taking one step, adjusting the rates, and then taking the second step, the model takes only one large step. It arrives at different states for different step sizes. For the next step, then, the model will start out in different states. The trajectories followed by the model will thus depend on the size of the time step.

7.4.7 Be Aware of Implicit Assumptions in Existing Tools and Templates

While using an existing method or model form can save time, it can also impose biases and preconceptions on the model. Some of the most ubiquitous assumptions are that models are linear (or can easily be transformed into a linear form) and that residual errors are distributed normally (or according to one of the other exponential-based distributions). Statistical modeling deserves a cautionary mention here. Correlation is not causation, and causation—the story behind the model—is what identifies the points of leverage for the decisionmaker.

A number of the assumptions implicit in certain methods and techniques have been discussed in Sect. 7.3. The modeler must watch out for hidden assumptions in some model types, e.g., rationality and market equilibrium in economic models, and the appropriateness of discounting in cost analysis. The modeler must be aware of assumptions and of the perspective on the system that is introduced by using existing methods or model forms.

The effect of type on model size (as measured by, e.g., the amount of data it will require or the amount of code that must be written) is also of importance here. Of course, if the model already exists, along with the appropriate data libraries, the analyst may choose to use it. This is often the position a research or consulting group is trying to get itself into when it invests in building a large model.

7.4.8 Build a Toolkit

Build a toolkit of models rather than a large, comprehensive model. This keeps the analyst at the center of the analysis, since the analyst is the one using the tools. Somebody must be the user of a large, comprehensive model as well, of course, but the large model will automate much more of the analysis process. Decisions about how to design alternatives, or how to weigh one impact against another, will often be implemented as algorithms in the large model.

Conduct the analysis in stages, so as to allow a change of course as necessary. Build the methodology in small pieces—as small models that can be used in stand-alone mode to aid different stages of the analysis.

7.4.9 Cooperate Closely with the Policymaker and Develop a Good Story

The modeler must keep in mind that the model is being developed for the purpose of decision support. This requires the modeler to involve the policymaker and other participants in the decisionmaking process during the whole process of model development, from planning until execution (EWRI 2011). In general, modeling not only involves the model as an end product, the entire modeling process is also a learning process.

In addition, a good story is an important element of the representational core. The policymaker will not accept a model's result "because the computer said so". There must be an intuitively satisfying explanation—though it can be largely qualitative—for why the model results come out the way they do, and why the real world should be expected to act similarly.

7.4.10 Let the Problem Determine the Tools that are Used

Last but not least, the problem situation should determine the methods and tools used for modeling. This is well known advice, and a number of considerations for this were discussed in Sect. 7.3, but it is easier said than done. Analysts have a certain background and may have more experience in using certain methods than others. Although it is easiest to use a familiar method, it is essential to step back to investigate what other possibilities exist and what the advantages and disadvantages of alternative methods might be.

References

Ackoff RL (1974) The future of operational research is past. J Oper Res Soc 30(2):93–104

Agusdinata DB (2008) Exploratory modeling and analysis: a promising method to deal with deep uncertainty. Ph.D. thesis, Delft University of Technology, Delft

Armiger G (1995) Handbook of statistical modeling for the social and behavioral sciences. Plenum Press, NY

Balci O, Ormsby WF (2007) Conceptual modelling for designing large-scale simulations. J Simul 1:175–186

Bankes S (1993) Exploratory modeling for policy analysis. Oper Res 43(3):435–449

Barlas Y (1996) Formal aspects of model validity and validation in system dynamics. Syst Dyn Rev 12(3):183–210

Carrillo MJ, Hillestad RJ, Twaalfhoven PGJ, Bolten JG, van de Riet OAWT, Walker WE (1996) PACE-FORWARD: policy analytic and computational environment for dutch freight transport, MR-732-EAC/VW, RAND, Santa Monica

Caulkins JP, Rydell CP, Schwabe WL, Chiesa J (1997) Mandatory minimum drug sentences: throwing away the key or the taxpayers' money?, MR-923-RWJ, RAND, Santa Monica

Davis PK, Bigelow JH (1998) Experiments in multiresolution modeling, MR-1004-DARPA, RAND, Santa Monica

Dupuy TN (1987) Understanding war: history and theory of combat. Paragon House Publishers, NY

Environmental & Water Resources Institute (EWRI) (2011). Collaborative modeling for decision support in water resources: principles and best practices, Report 2011-R-03, U.S. Army Corps of Engineers, Institute for Water Resources, Alexandria, Virginia

Forrester JW (1961) Industrial dynamics. MIT Press, Cambridge

Gold MR (ed) (1996) Cost effectiveness in health and medicine. Oxford University Press, Oxford

Greenberger M, Crenson MA, Crissey BL (1976) Models in the policy process: public decision making in the computer era. Russell Sage Foundation, NY

Hillestad RJ, Walker WE, Carrillo MJ, Bolten JG, Twaalfhoven PGJ, van de Riet OAWT (1996) FORWARD—freight options for road, water, and rail for the dutch: final report, MR-736-EAC/VW, RAND, Santa Monica

Hillier FS, Lieberman GJ (2005) Introduction to operations research, 8th edn. McGraw-Hill, Boston

Hodges JS (1991) Six (or so) things you can do with a bad model. Oper Res 39(3):355–365

Hodges JS, Dewar JA (1992) Is it you or your model talking? a framework for model validation, R-4114-AF/A/OSD, RAND, Santa Monica

Hopkins A (ed) (1992) Measures of the quality of life, and the uses to which such measures may be put, Royal College of Physicians of London

Jones-Lee MW (1976) The value of life: an economic analysis. University of Chicago Press, Chicago

Kleijnen JPC (1999) Validation of models: statistical techniques and data availability". 1999 Winter Simulation Conference Proceedings, Phoenix, AZ, 5-8 Dec 1999, pp 647–654

Kwakkel JH, Walker WE, Marchau VAWJ, From predictive modeling to exploratory modeling: how to use non-predictive models for decisionmaking under deep uncertainty. In: Proceedings of the 25th mini-euro conference, University of Coimbra, Portugal, 15-17 April 2010 (ISBN 978-989-95055-3-7)

Landsburg SE (2011) Price theory and applications, 8th edn. South-Western Cengage Learning, Mason, Ohio

Law AM, Kelton WD (1991) Simulation modeling and analysis, 2nd edn. McGraw-Hill, Boston

Mankiw NG (1997) Macroeconomics, 3rd edn. Worth Publishers, NY

Morton S, Rolph J (2000) Public policy and statistics: case studies from RAND. Springer, NY

Perkins F (1994) Practical cost benefit analysis: basic concepts and applications. Macmillan Education, Australia

Pilkey OH, Pilkey-Jarvis L (2007) Useless arithmetic: why environmental scientists can't predict the future. Columbia University Press, NY

Roberts EB (ed) (1978) Managerial applications of system dynamics. MIT Press, Cambridge

Robinson S (2004) Simulation: the practice of model development and use. Wiley, UK

Quade ES (1989) Analysis for public decisions, 3rd edn. Elsevier, NY

Seinfeld JH (1986) Atmospheric chemistry and physics of air pollution. Wiley, NY

Sterman JD (2000) Business dynamics: systems thinking and modeling for a complex world. McGraw-Hill, Boston

SUMMA Consortium (2005). Final publishable report, report for the European Commission (DG-TREN), Leiden. [http://www.tmleuven.be/project/summa/summa-d8.pdf]

Van Horn RL (1971) Validation of simulation results. Manage Sci 17:247–258

Walker WE, Lang NA, Keur J, Visser HG, Wijnen RAA, Kohse U, Veldhuis J, De Haan ARC (2003) An organizational decision support system for airport strategic exploration. In: Bui T, Sroka H, Stanek S, Goluchowski J (eds) DSS in the uncertainty of the internet age. Publisher of the Karol Adamiecki University of Economics in Katowice, Katowice, pp 435–452