Proceedings of the 2003 Winter Simulation Conference
S. Chick, P. J. Sánchez, D. Ferrin, and D. J. Morrice, eds.

VISUALIZATION OF PROBABILISTIC BUSINESS MODELS

Lev Virine
Lisa Rapley

Calgary Technology Center
Schlumberger Ltd
600, 322-11 Avenue S.W.
Calgary, AB T2R OC5, CANADA

ABSTRACT

One of the main challenges in the modeling of business problems is to provide the modeler and the user with meaningful visual tools. The business model is usually presented by different types of flow charts and diagrams. If the modeling process is simplified in how it is represented to the user, it improves understanding, as well as, helps to interpret the result of the analysis. This paper discusses a proposed methodology for business modeling and how this process can be applied to real world problems. The formal iterative modeling process includes a Probabilistic Model Description, Domain Model Diagram, and diagrams to define model's calculation logic, sensitivity analysis tools, decision trees, and other tools. The paper also discusses benefits of the unification of specification for the visualization tools. The described methodology is used in decision and risk analysis application Decision Tool Kit.

1 INTRODUCTION

Computer-based business analysis relies on the model, or algorithmic representation of the business process and the associated data. The model can change significantly for multiple business processes and will also vary based on different requirements. This makes flexibility in the process necessary to adapt various approaches to solving problems.

Effective decision making and problem evaluation analyses of a business process requires a clearly defined business model. One of the challenges to accurately defining a business model is the complexity involved in decision analysis, risk and uncertainty management, optimization and portfolio management. It is time consuming to attempt to model a situation and can require multiple attempts before the user can communicate the relationships properly. There can be a disconnect between the business processes when it is translated using technology processes. The key reason for this is the difficulty in being able to visualize the model.

The analysis of the business model usually involves a number of calculations, analysis of alternatives and special conditions, concurrent processes based on multi-dimensional data. The visualization can be accomplished using various diagrams, graphs, and interactive tools.

Currently existing decision and risk analysis methodologies include the visualization activities; however, not all phases of the decision and analysis process are included in any particular visual tool. As a result, the formal decision and risk analysis process can be significantly complicated. The formal process of deriving of one phase based on previous phase is not trivial. The interconnection of the phases is part of the process that becomes lost.

A significant challenge to existing methodologies is the lack of unification. In spite of significant number of different visualization techniques and tools for the probabilistic problems, there isn't a unified approach to the problem of visualization. The visualization techniques used in one model may not applicable for another model. Sometimes authors of a model come up with their own type of diagrams or modify existing style of diagrams. Due to the lack of visualization standards, the modelers and users have difficulties to read and understand the model created by another formal process. This can confuse the meaning of the model and leave wrong assumptions in the process.

With a standard in place, unified visualization tools for probabilistic models can be similar to construction blueprints: all industry specialists will be able to commonly understand diagrams, graphs, and iterative tools without misunderstanding.

This paper discusses the analysis of existing visualization techniques and tools for the probabilistic business models. These techniques have been used to create and visualize models for a number of real problems from different industries. The results, in the form of diagrams, graphs, and interactive tools were presented to number specialists, who were not familiar with the particular model. They were interviewed and an assessment of their understanding of the model was made. Most efficient visualization tools, which can applicable to virtually any industry, were selected based on this research. Finally the specification of a number of unified diagrams was created and process of using them was defined.

This paper will discuss general guidelines and definitions of the various modeling tools. The specifications of the diagrams, as well as, analysis of visualization methods related to optimization are the focus.

2 ANALYSIS OF EXISTING VISUALIZATION TOOLS FOR PROBABILISTIC PROBLEMS

A number of decision and risk analysis processes are defined and actively used (Howard 1988; Skinner 1999). They commonly include the following steps: identifying problem, assessing the business situation, modeling of that situation, evaluation and sensitivity analysis, risk analysis, determination of the value of new information, and making an informed decision (Howard 1988; Skinner 1999). Typically, the initial steps in this process including definition and analysis of business rules, alternatives, and possible risks include only a few limited visualization tools.

In order to simplify the analysis of the visualization tools they will be separated into two main groups. The first one includes diagrams used to represent the model algorithms, data, and associated events. The second group includes decision and risk analysis visualization tools.

The first group of diagrams includes flow charts and state diagrams. The basic flowchart can be used to represent the calculation algorithm (Curtis 1998). There are a number of different permutations of the basic flowchart. Another type of diagram for time-based events is a state diagram. State diagrams are successfully used for the description of time-based technological processes and business processes over a period of time (Cover and Thomas 1991).

Further, a number of different network diagrams have been developed and can be applicable for visualization of probabilistic problems. All of these visual tools are event-based diagrams. In the notation, each node of the diagram represents a particular event (action) or condition. A similar approach is used for diagrams in a number of business modeling software applications. A centerpiece of Extend simulation software by ImagineThat Inc. (Krahl 1999; Clymer 1999) is a modeling diagram, which represents a simulation model using a group of interconnected nodes. Each node may contain both graphical and textual representation of the associated event. Simplification of the model is important because it allows for greater understanding by the users. A similar kind of diagram can be fond in GoldSim software by GolderAssociates GoldSim Technology Group. Developers of GoldSim found a very elegant method to present business model together with a graphical representation of calculations and simulation processes. Another software system, Enterprise Optimizer by River Logic Inc., allows the user to see model events and associated probabilistic data. Diagrams used in all of these software systems are very similar but not completely unified. The underlying interactive processes related to these diagrams is different for each product.

Most business processes have sophisticated logic of data flow. Processes used to describe the flow of data are Bachman diagrams, Chen notation, and Martin notation, which are all included in Microsoft Visio. Another example of a visualization tool is the IDEF family of methods (Liles and Presley 1996). This method includes a complete and well maintained specification and is considered to be a powerful modeling tool. It is a method designed to model the decisions, actions, and activities of an organization or system. IDEFØ was derived from a well-established graphical language, the Structured Analysis and Design Technique (SADT) (SofTech 1976). Effective IDEF models help to organize the analysis of a system and to promote effective communication between the analyst and the user. However, our research found, that an inexperienced IDEF user has difficulties with understanding the model visualized using this family of methods. Ease of use is required for a methodology to be adopted throughout multiple industries to become a standard form of practice.

One of most remarkable examples of unification of visualization tools is Unified Modeling Language (UML) for specifying, visualization, and documenting for object-oriented software development (Fowler and Scott 2000; Jacobson, Booch, and Rumbaugh 1999). This type of standardization is possible for decision and risk analysis notation. UML diagrams consist of class diagrams, activity diagrams, sequence diagrams, collaboration diagrams, use cases, and others. They become an important tool to analyze to the original business problem and collect business requirements. UML diagrams are easy to understand and are well accepted because they are simple to use, are understood by users with different cultural origins (supersede language and industry specific jargon), and display the relationships between different components rather than just the components themselves. However; UML specification does not include the explicit definition of probabilistic methods and tools.

The second group of diagrams include specific tools for decision and risk analysis. This category of diagrams contains decision-framing tools such as strategy tables or influence diagrams, and tools associated with the particular analysis techniques (tornado diagrams, decision trees, etc.). An influence diagram both displays the business problem by presenting relationship between variables and processes (Goodwin and Write 2002; Skinner 1999). The advantage of the influence diagram is that specification is considered a standard and is included in many software applications. Analytica by Lumina Data System will create a probabilistic model based on the influence diagram. One of the shortcomings is that for some business processes, especially for time-related processes, it is difficult to visualize a model using an influence diagram.

The influence diagram demonstrates the relationships of the variables and defines the input types in relationship to the value measure. When this approach was used in fiscal model research and development, most of the users had difficulties understanding processes such as capital depreciation or discounting using influence diagram, because the influence diagram operates mostly with decision and variables, and not time related events.

Other decision analysis tools such as decision trees, tornado diagrams, chart related Monte Carlo simulations (frequency and cumulative probability chart for input distributions and results of calculation) are very well defined. The probabilistic time series diagram is used to represent the uncertain parameters or group of parameters that can change over time. There are also forecasting tools (Yurkiewicz 2003): time-series plots, scatter diagrams, forecast plots, residual plots and others.

Due to their nature the visualization tools they cannot be the initial tool used to visualize the model, but can be part of the interactive process related to model's analysis. These types of diagrams are used in number of software applications: Crystal Ball by Decisioneering Inc, Decision Tools Suite by Palisade Corporation, DPL by Standard & Poor's, DATA by TreeAge Software Inc., DecisionPro by Vanguard Software Corporation and others.

Although decision trees and tornado diagrams are commonly accepted tools with nomenclature associated with them, other tools do not have a unified approach. There are a few of reasons why the unification of some visual tools for the probabilistic model is not yet achieved:

1. Multiple software application vendors use very similar software tools that accomplish the same things with different specifications.
2. Industries work independently and have varying specifications and visualization approaches.
3. The decision analysis and risk management industry was not mature enough to justify significant investment in to the unification of the visualization process.

3 THE MODELING AND SIMULATION WORFLOW

The proposed modeling and simulation process starts with the definition of probabilistic problem. We propose the term Probabilistic Problem Description for this step and associated visual tool. Once established, the overall description of business rules and alternatives (the proposed term is Domain Model) can be created. The Domain Model can be represented in form of a Domain Model Diagram. The next step includes decision and risk analysis methods. This modeling process is iterative; therefore many trials of the steps may be implemented. In most cases the model can be refined based on results of analysis to reflect the business problem more clearly. Results of the particular decision analysis and simulation methods can augment each other. For example, results of sensitivity analysis can be used to generate decision trees.

Recommended improvements between this modeling and simulation process and traditional decision analysis process are:

1. The each step within the proposed process is associated with the particular visual tool or visualization methods. The formal rules can be applied to move from one step of the process to another using visual representation of each step.
2. This methodology is designed to simplify the traditional process. A number of steps within a traditional process can be presented by one visual tool. For example, analysis of business environment, including competitor analysis, analysis of best practices, identification of risk and uncertainties can be presented in Probabilistic Problem Description.
3. Because of limited number of steps within the proposed methodology, there is opportunity to define a specification of each visual tools.

Visual tools for the modeling and simulation workflow are presented on Figure 1.

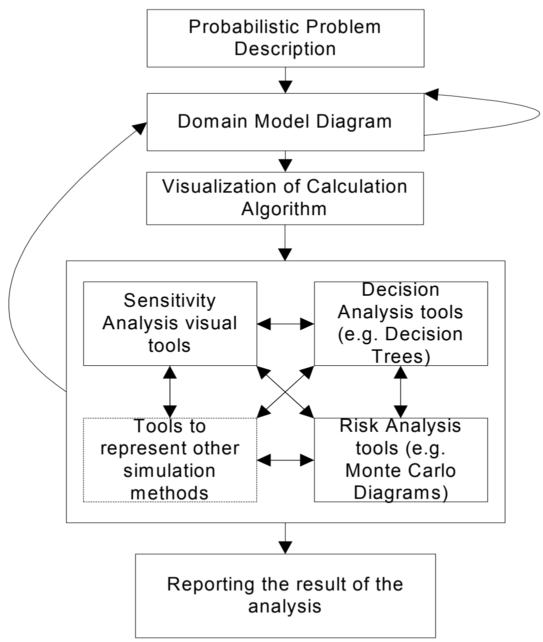

Figure 1: Visual Tools for the Modeling and Simulation Workflow

The proposed methodology has main limitations: it cannot be applicable to comprehensive and large scale decision and risk analysis involved a large number of steps and alternatives. The example of such analysis is an evaluation of new business opportunities for large scale enterprise modeling.

4 PROBABILISTIC PROBLEM DESCRIPTION

In most cases decision and risk management processes involve estimating certain selected parameters of the process in different circumstances. In other words the main goal of decision and risk analysis is to determine how uncertainties and input data effect a selected value measure and indicate a decision direction based on this. Therefore, a model description includes a sequence of calculations with input and output variables, conditions and constraints. Sometimes the model description can be very complicated especially when results of calculation conditionally effect inputs. The task is to describe the model in readable format, decompose the model on separate calculations, and define the workflow of the calculations and simulations.

Here is an example of a probabilistic problem. Oil company X Inc. is exploring an underwater oil reserve. The estimated size of the reserves is based on a geophysical survey. The company is planning to build a small facility (platform) and drill one well. If they do not discover oil in the first well (called a dry hole), the company will abandon future operations in this area. If oil is found, then the facility will extend and two more wells will be drilled. If at least one of them is dry hole, the production will continue on existing facility until the economic limit of the well is reached. If all three wells are producing, the facility will be upgraded and three more wells will be drilled. If production is high, then an additional upgrade of the facility is required. The task is to estimate net present value (NPV) of the complete project, for each outcome and on the each stage based on the timeline.

This process includes many deterministic and probabilistic variables. It also includes calculation of NPV multiple times. The visualization of this situation will make the model more easily understood. Visualization tools will help the user of model input data, run the simulation, and analyze the results. Analysis of existing means of visualization shows that on the initial step the textual description of the process the certain format is most efficient way to represent the model. This textual description is called a Probabilistic Problem Description. It includes step-by-step definition of the process, definition of condition and alternatives, input and output variables. Here is how the Probabilistic Problem Description will look like for this example:

1. Obtain uncertain data about underwater reservoir size and future oil process estimation.
2. Calculate production for the one well based on geological data .
3. Calculate NPV for this well based on production, oil prices, capital expenditure, taxes and royalties including dry hole sunk costs.

4.1 If there is a dry hole determine the cost of abandonment (cost of termination of the well) based on well parameters

4.2.1. If the well is producing oil, determine the cost of facility extension based on production data

4.2.2. Calculate NPV for two other wells based on production, oil prices, capital expenditure, taxes and royalties including dry hole sunk costs for each of them.

4.2.3.1. If the second well is a dry hole determine the cost of abandonment based on well parameters.

4.2.3.2. If the third well is a dry hole determine the cost of abandonment based on well parameters.

4.2.3.3. If both wells are producing, determine the cost of additional facility extension based on production data

4.2.3.4. Calculate NPV for three other wells based on production, oil prices, capital expenditure, taxes and royalties including dry hole costs for each of them.

4.2.3.5. If production is high determine the cost of the additional facility extension

4.2.3.6. Determine the abandonment cost of the facility and all six wells.

The numbering system in the current example represents the alternative IDs. If there is more than one alternative, additional level of numbering will appear. Another feature of this artifact is underlining of the variables: A dashed line underlines input probabilistic variables, a single solid line underlines input deterministic variables and the box will surround the value measure(s).

The Probabilistic Problem Description can be derived for the textual description of the project. But it is different from the project description because it represents steps for analysis rather than steps for project implementation. It basically answers the questions: what steps need to be taken in the process analysis, what variables are involved, and what are the alternatives?

The Probabilistic Problem Description is a very simple tool and has been found useful in defining the model. A diagram is not involved at this point since the Probabilistic Problem Description is required to determine the logical system for the diagram. Therefore, it is considered as a part of the set of visualization tools.

5 DOMAIN MODEL

The Domain Model is an important step in the analysis of business process and generation of the probabilistic model. The main purpose of the Domain Model is to provide model decomposition or separate the model by its calculations and determine input and outputs for each of the steps. The Domain Model is derived from the Probabilistic Problem Description. In addition to the information from Probabilistic Problem Description it also shows explicit relationships between variables and alternative selections in the process.

Domain model is represented by a diagram with a number of boxes. Each box shows a part of the process, and when it is required to have a separate calculation or an alternative. This is a system of influence diagrams but is proposed to be in a more readable format.

Each box contains multiple columns. The left column contains a list of input variables, the middle column or columns contain the list of intermediate variables, and the right column contains list of output variables. The variables can be underlined using the same rules described in the Probabilistic Problem Description. Figure 2 shows the example of Domain Model Diagram. Sometimes it is difficult to generate a complete model drawing. It this case the table of each calculation can be created separately.

The Domain Model also represents the relationships between the calculations. This is noted by the different boxes. There are three basic types of relationships:

1. Conditional relationship between calculations when input data for the next calculation depends on result of previous calculation. It is represented by the diamond symbol (the same as in standard flow chart diagram).
2. Random relationship or probabilistic relationship (based on selected statistical distribution). It means that the selection of alternative will be selected randomly. The meter sign represents this relationship (not shown on Figure 2).
3. Unconditional relationship is represented by the straight line.

The aforementioned relationships represent how the calculations are linked to each other. In addition to this a relationship between separate variables within different calculations. For example well production for two wells described in Probabilistic Problem Description (step 4.2.2) depends on production on Step 3. It this case, selected variables in the boxes will be connected to each other by curved connecting lines and conditional signs (the diamond or the meter) may be inserted at the middle of the line.

The Domain Model is read from left to right with alternative scenarios placed vertically in parallel. This will simplify the conversion of the Domain Model to the decision tree, which is also drawn from the left to the right.

The specification of the Domain Model diagram includes the ability to present a multiple calculation scenarios. It can be accomplished by "three-dimensional" calculation box as show on Figure 3.

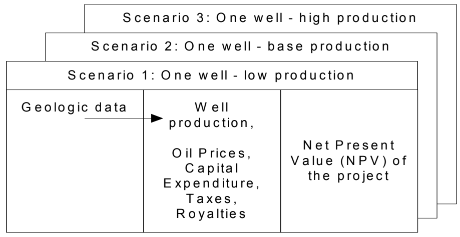

Figure 3: Multiple Scenarios in Domain Model Diagram

Results of the analysis (NPV in the current example) can be combination of NPV values at each stage. There are different ways in which the results can be combined. The results of subsequent calculations can be replaced by result of previous calculations. Also results of analysis can be the sum of results of each calculation, etc. Results of the analysis can be presented in the separate box as shown on Figure 2.

The Domain Model is an important step in the proposed workflow. It helps to generate a model through other diagrams and charts. The Domain Model can be an interactive "white board" or an analytical tool. The group of modeler can easy discuss a model and frame a decision using the diagram that is understandable to the users.

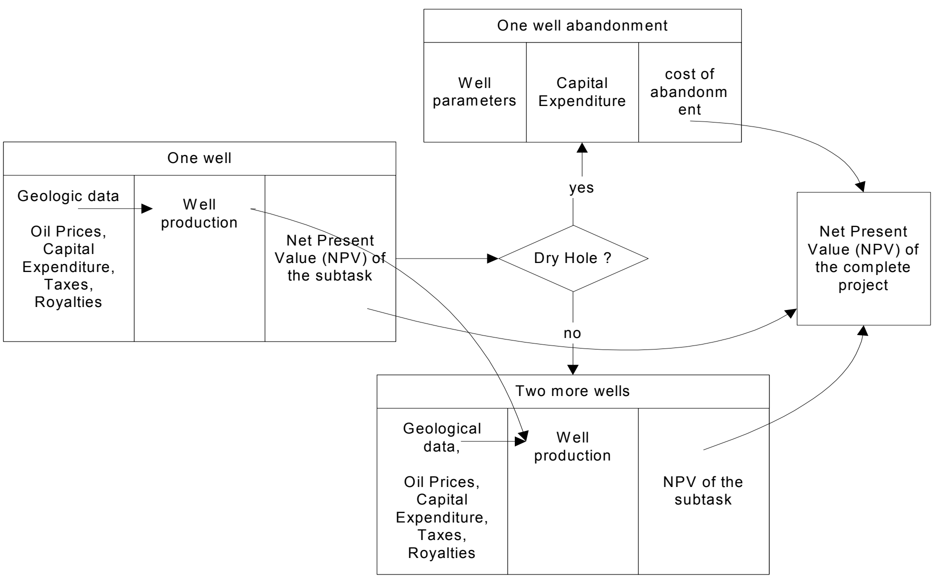

Figure 2: Domain Model Diagram

6 VISUALIZATION OF THE CALCULATION ALGORITHM

Each box within a Domain Model Diagram represents a separate calculation or logically completed algorithmic process. The calculation algorithm can be presented using different tools. Experience shows that the modeler and the user of the probabilistic model are not usually interested in details of the internal calculation when they perform the simulations. For example, in order to calculate Net Present Value of the project, capital depreciation calculation is required. However, the particular mathematical method to compute capital depreciation is not important during the calculation aspect of the simulation.

Four methods can be user to visualize a calculation algorithm for probabilistic data models:

1. Short textual description of the calculation algorithm as a reference or an annotation in to Domain Model Diagram. The reference can be used if a number of calculations use the same calculation algorithm. The annotation can refer to the particular calculation engine. For example, NPV calculation can be viewed using an NPV spreadsheet.
2. Extended textual description of the calculation with description of the step and some mathematical formula's if it is required.
3. Standard flowcharts: Domain Model diagrams will have a reference to the flowchart artifact. The flowchart is useful when a complicated computational logic is involved.
4. State diagrams with reference to the Domain Model Diagram. It can be useful for the time-based processes.

7 DECISION TREES, SENSITIVITY DIAGRAMS, MONTE CARLO DIAGRAMS

The next step in the proposed workflow is the execution and visualization of the particular method of decision or risk analysis. Risk analysis use different methods and different visualization tools, but this paper will describe only the most commonly used techniques.

The sensitivity analysis (or what-if analysis) is used to determine how sensitive is result of analysis to the uncertainty in the input variables. The result is usually presented in the form of tornado diagram. In order to simplify understanding of tornado diagram it is recommended to use it in the form as a presented on Figure 4:

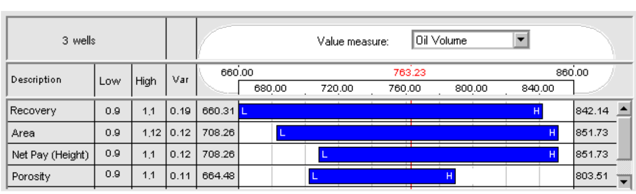

Figure 4: Tornado Diagram with Input Variables

The tornado diagram in this format includes low and high values for the input variables, and the corresponding result of calculation in the same row.

Another visualization tool for the sensitivity analysis is a spider diagram. The spider diagram shows the rate of change of different input variables against the rate of change of an output variable. Spider diagrams can be useful if there is a non-linearity in the relationship between rates of change of inputs and outputs, as in Figure 5.

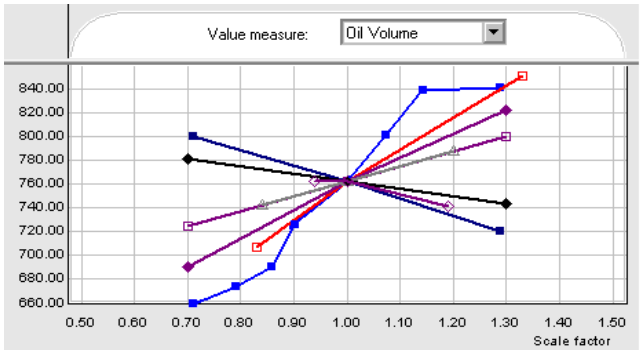

Figure 5: Spider Diagram with Non Linear Data

It allows the user to understand the model behavior more clearly. Another type of sensitivity diagram is the value measure diagram. It is similar to the spider diagram and shows the relationship between rate of change of one selected input and rate of change of multiple compatible output variables. The example is shown on Figure 6.

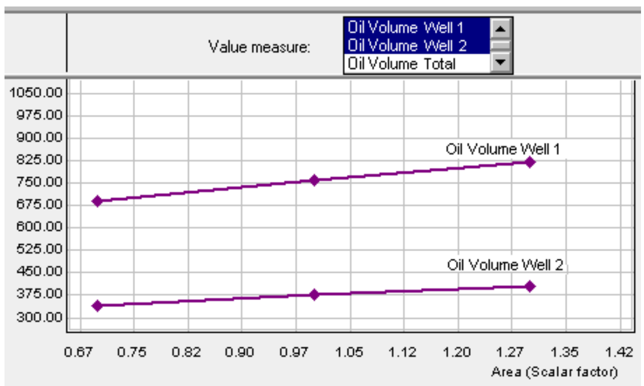

Figure 6: Value Measure Diagram

Input data and results of decision analysis can be presented in form of a decision tree. The Domain Model can be used to simplify the generation of the decision trees. It some cases the order of boxes in the domain model diagram can correspond with order and layout of decision nodes in the decision tree.

Monte Carlo simulation is a powerful risk analysis method, but sometimes visualization of the Monte Carlo simulation can be very difficult to understand. It is recommended to modify the Domain Model Diagram for the visualization of Monte Carlo. In order to generate Monte

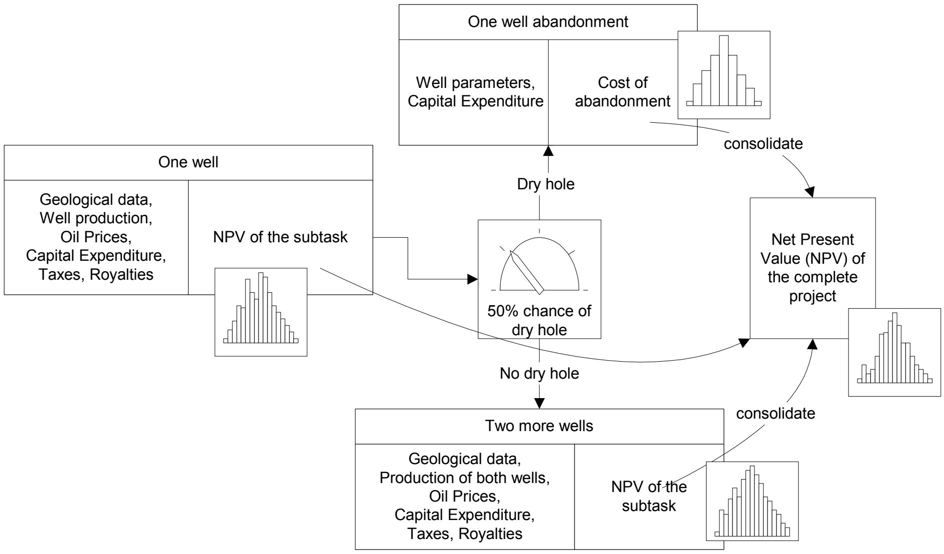

Figure 7: Monte Carlo Diagram

Carlo diagrams based on the Domain Model diagram, the following steps should be preformed:

1. Calculation or logical steps, represented by boxes in Domain Model Diagram by boxes, should be selected for Monte Carlo diagram.
2. Intermediate variables and deterministic variables within boxes can be omitted to simplify the diagram
3. Input distribution for the variables should be defined and can be presented on the Monte Carlo diagram.
4. Results of Monte Carlo simulation can be visualized using frequency charts. Sometimes cumulative probability charts can be used, but can be more confusing than frequency charts. The user of cumulative probability chart can lose understanding of the nature of the results. It is recommended to present final results of the analysis in form of frequency chart within a Monte Carlo diagram and associate it with selected value measure.
5. Monte Carlo diagrams contain a very specific notation – consolidation of results. Results of different calculation (for different boxes) can be consolidated (added to each other), replaced, or another mathematical operation can be performed with them. Connecting lines between value measures for different boxes will present this action.

The Example of Monte Carlo diagram is shown on Figure 7.

A significant number of business models represent time-based or event based problems. It is recommended to define a Probabilistic Problem Description and Domain Model in proper chronological order to simplify understanding of the problem. Results on the analysis can be also represented using time-based chart. Probabilistic time-based chart represent the uncertainty in results in each period of time. The example of probabilistic cash flow chart is presented on Figure 8. Probabilities P10/P50/P90 are presented by separate lines. It is possible to use different colors, shades, and lines for different percentiles to visualize uncertainties.

Another important tool with the probabilistic time-series diagram is the definition of the event, associated with changes in the diagram. The events can be defined by short textual description and shown as a callout on the diagram.

The callout notations will make the model more understandable to the user. This is especially important when the result of calculation is changed due to special condition or the selection of uncertainties.

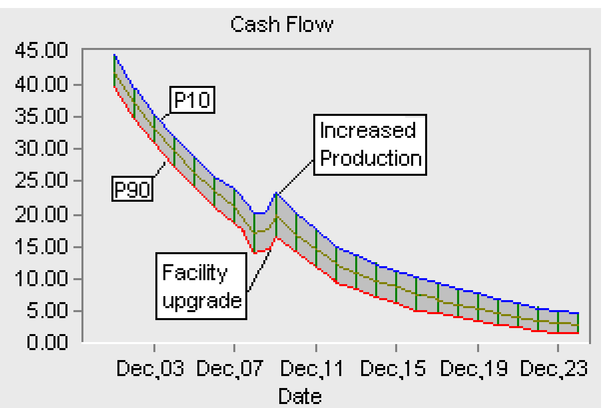

Figure 8: Probabilistic Time Series Plots with Associated Events

8 SOFTWARE APPLICATIONS

The proposed methodology has been used to build and analyze models in oil and gas industry. The process of drawing diagrams including Domain Model and Probabilistic

Problem Description can be done using standard commercial office software. Other diagrams, such as Decision Tree and sensitivity diagrams can be implemented using other kinds of the commercial software.

The workflow, presented in the papers is encapsulated in Schlumberger's Decision Tool Kit software product. The application is stand alone Decision analysis and Risk Management engine. It can work with different calculation engines including Microsoft Excel and includes three integrated models: Sensitivity Engine, Decision Tree, and Visual Monte Carlo tool. The open architecture of the Decision Tool Kit permits the addition of client's calculation engines. It includes the most of visualization tool described in the paper.

CONCLUSIONS

This paper described a proposed workflow that generates a probabilistic business model using a number of visualization methods and tools. The idea is to impose a formal and unified process of the model generation where each step and activity is associated with the specific visualization tool. It significantly simplifies the modeler's work and helps make the model more accurate.

The described concept is encapsulated in the software application Schlumberger's Decision Tool Kit. Future research and analysis is required to refine the proposed methodology to ensure that it is applicable for wider range of industries.

REFERENCES

Clymer, J. R. 1999. Optimization of Simulated System Effectiveness Using Evolutionary Algorithms. *Simulation, Journal of the Society of Computer Simulation*. 73(6): 334-340.

Cover, T., Thomas, J. 1991. *Elements of Information Theory*. Wiley-Interscience.

Curtis, G. 1998. *Business Information Systems - Analysis, Design & Practice*. 3rd Ed. Addison Wesley.

Fowler, M. and Scott, K. 2000. *UML Distilled*. Reading, MA: Addison-Wesley.

Goodwin P., Wright G. 2002. *Decision Analysis for Management Judgment,* 2th Ed. Wiley.

Howard, R. 1988. Decision Analysis: Practice and Promise. *Management Science*, 34: 679-695.

Jacobson, I., Booch, G., and Rumbaugh, J. 1999. *The Unified Software Development Process*. Reading, MA: Addson-Wesley.

Krahl, D. 1999. Modeling with Extend. In *Proceedings of 1999 Winter Simulation Conference*, Phoenix, AZ.

Liles D., Presley A. 1996. Enterprise Modeling within an Enterprise Engineering Framework. In *Proceedings of 1996 Winter Simulation Conference*, San Diego, CA.

Skinner D., 1999. *Introduction to Decision Analysis*. 2nd Ed. Gainesville, FL: Probabilistic Publishing.

SofTech. 1976. *An Introduction to SADT Structured Analysis and Design Technique*. Document No. 9022-78R, Waltham, MA.

Yurkiewicz J. 2003. Forecasting Software Survey, *ORMS Today.* 30(1) 34-46.

AUTHOR BIOGRAPHIES

LEV VIRINE is senior software engineer of Schlumberger Information Solutions, a division of Schlumberger Ltd. He received his post-graduate degree from Moscow State University in 1986. His current research interests include decision analysis, risk management, and their applications for oil and gas industry. His email address is <lvirine@calgary.geoquest.slb.com>

LISA RAPLEY is an experienced project manager currently working in research and development for Schlumberger Information Solutions. She received her MBA at the University of Calgary with a major in Finance. Currently, her interests include investigating the decision making process, risk analysis, problem identification and framing barriers, and the effectiveness of leadership within group projects.