Modeling As a Communication Process: Computer Conferencing Offers New Perspectives

JACQUES VALLEE

ABSTRACT

Model building is currently one of the most expensive and difficult applications of computer technology. It is expensive both as a conceptual activity, because building a model requires an abstract understanding of a complex process, and as a computational activity, because it demands massive amounts of data and many operations. Furthermore, under conventional approaches one cannot start using the full power of the computer in modeling until all the components have been identified *and* all the relationships are established. This article introduces an alternative approach in which the computer is used during the earliest stages of model construction. This approach is actually an extension of current technology for computer-based teleconferencing. It incorporates nonhuman "participants"–modules, data bases, or parts of programs–in the modeling process. The advantages of this approach include (1) the opportunity for human participants to test alternative model structures and (2) the capability of the computer to monitor and feed information about the process back to the model director.

Introduction

"Not much is new in modeling technology," remarked Edward Roberts in a recent article [1], and he added that "the key variants of model building approaches were available ten years ago or earlier". Whether or not this means that model building has matured as a technological form or that methodological stagnation exists, it is a fact that there remains a disparity between the expectations of users and their satisfaction with results. This disparity may be traced, in part, to communications problems, and it might be remedied if recent advances in human communication *through* computers were applied to develop what might be called "dynamic modeling." Accordingly, this article examines the potential of computer conferencing systems to deal in novel ways with the three components of a model: the process, the data, and the people who build and use it.

The Components of Modeling: Process, Data, and People

The PROCESS of modeling has been examined in a study completed in 1973 by Mar and his coworkers [2]. They report that, in their observations of 11 modeling projects,

DR. JACQUES VALLEE holds a Ph.D. in computer science and has published over forty articles on information systems in British, French and American scientific journals. He is currently a senior research fellow at the Institute for the Future and the President of INFOMEDIA Corporation, 430 Sherman Avenue, Palo Alto, CA 94306, where he is leading the development of a computer conferencing service.

"results indicated that the *process* is seldom recorded and the technology of interdisciplinary modeling has not been captured." According to the same study, three issues are encountered time and again in the development of models: how to define and bound the model; how to orchestrate the team in constructing and validating the defined model; and how to document and communicate the model or its results. In short, the process of modeling is not well-defined.

The DATA upon which the model operates present several challenges: choice of data sources, validation, and access are the major ones. Also, biases in information sources which are not obvious to designers may undermine the user's confidence in the model. Finally, the needed information may be scattered in many places, as illustrated in a report by the National Commission on Materials Policy:

> The 1960s produced a snowstorm of mineral data, scattered through the literature and in private and public file cases. Although data are available within various governmental and private files to create a national data base, these data are not in a computerized storage and retrieval form which would allow for their utilization [3].

In cases such as these, considerable waste is incurred when an attempt is made to centralize all the files under a single system even before their relevance and value has been examined. In contrast with this situation, an experiment was conducted in June 1973 in which two data bases, residing on different computers, were linked with a teleconference on a national computer network. The participants in the conference could address questions to the data bases in the course of a real-time discussion without investing in any new system development work. The results of that experiment demonstrated the value of human discussion and control of data quality before information channels are frozen [4].

The PEOPLE involved in modeling constitute the third and least understood component. The idea of building models springs from the belief that any process or situation can be conceptualized and given an abstract representation. This belief, founded in the successful application of mathematical models in physics, is seldom challenged. Furthermore, it is assumed that the resulting abstract representation can be shared between the conceptualizer and the "user".

Common industrial experience indicates many flaws in these assumptions. It is possible that not *all* situations or processes can be modeled; when the situation can be modeled, individual differences in cognitive styles and modes of abstraction make the representation a highly personal one; and the usefulness of the end product is limited by the ability to document not only the model but the world view upon which it is based.

These three components of process, data, and people create a high investment cost in modeling. Before any tests can be performed by the end users, considerable time and money must be expended to develop the abstract representation, to identify data sources, to implement model prototypes, and to test them. Naturally, each of these steps can also be viewed as the choice of a branch in a decision tree, and the first actual tests are performed on a system which embodies many a priori decisions. The cost of further changes in the basic structure, if aspects of the concept are found to be wrong, is thus extremely high.

Computer Conferencing: A Structured Communications Environment

If these problems are viewed as communication problems, it becomes clear that computer conferencing, by augmenting the communications environment, can potentially improve the building and use of models. The concept of computer conferencing is currently embodied in several operational systems, such as the PLANET program implemented by the Institute for the Future, the EIES system of the New Jersey Institute of Technology, the TOPICS system developed by Infomedia, and others. Intended to support disseminated groups of experts or researchers, PLANET requires no previous knowledge of computers for its operation and allows users to "attend" meetings and exchange messages from remote terminals located in their homes or offices. The organizer of a PLANET conference lists the names of authorized participants; a facilitator may also participate to keep the discussion focused and to attend to the dynamics of the group. The substance of the conference is found in public messages, which can always be reviewed and searched, and in private messages which disappear from the system once they have been viewed by the recipient.

Such a medium of communication could offer the following features, shown in Table 1, to support modeling efforts.

This potential of computer conferencing is reflected in the current use of PLANET. The main users to date have been research groups at NASA and ERDA, mineral geologists at the U.S. Geological Survey, and groups of educators, climatologists, and other scientists supported by such organizations as the Kettering Foundation and Lilly Endowment. The users have applied this medium of communication to the discussion of policy issues, to the joint development of research reports and position papers, to the planning of distributed experiments and the design and use of data bases. Even in the context of the primitive discussion structure we have described, some of the user groups have also begun to define joint models and to promote their use. One conference, for example, was organized by a geologist whose team has developed components of an energy model of the Northern Great Plains. Some typical entries in the public transcript of their discussion are shown below, with fictitious participant names. At this point, they only address the *structure* of the model. However, later entries will be seen to apply to the *process* itself.

[6] Richard(Org) 9-Jan-76 3:37 PM
This conference is an experiment in the use of teleconferencing as a means of facilitating communication between geographically separated persons working with a given simulation model. To run the model (after leaving PLANET), type "Run NGP". The computer then prompts with "H:" to let you specify which variables you wish to

TABLE 1

Model components	Useful features of a computer conference
Process	Single structure for orchestrating the model design, validating the model, and training users
	Complete record of discussion and iterations in the design phase
	Link among model builders located at distant sites
Data	Access to data sources
	Opportunity to challenge the validity of a given data source and to suggest alternatives
	Ability to monitor actual data use, general statistics on access frequency for various components, quantify the value of information
People	Easy participation of distant colleagues
	Closeness to one's own data sources
	Control of discussion structures
	Quantitative and qualitative information exchange are equally supported

see. The computer will next prompt "changes?", at which point you have the opportunity to change or examine any of the program data (constants, initial values of variables, switches, etc.) before execution of the simulation.

It is apparent that this group is already using the computer conference to train users and to keep others informed of changes being made in the model as the program expands and errors are corrected:

[10] Robert 21-Jan-76 1:24 PM
Annotation was added to the water sector of NGPCO and C(161) was corrected from 50.0 to 5.0.

In a later conference, a group began the discussion of the process itself, and one of the participants used the system to describe a flow diagram:

[99] William 20-Sep-76 8:35 PM
I'm going to try to get a flow chart on the system for you all. The messages will be one geometric figure per message. First will be the form (square, diamond, etc.), then what goes inside the form. Then you can all draw your own flow chart. O.K.?

Such use of computer conferencing, in the absence of a graphics facility, is clearly primitive. However, it does enable us to capture the step-by-step development of the model:

[100] William 20-Sep-76 8:37 PM
Start (this is the "linear array" model, by the way). Form = box–input array.
[101] William 20-Sept-76 8:39 PM
Form 2 = box (arrow from box 1)–input test pair. Form 3 = diamond–is one test pair an anchor? "Yes" line to the left of diamond, "No" line down.

This conference developed very rapidly, and a second user (Dave) began entering an alternative representation for the model. Figure 1 shows what participants would see as they entered this discussion and began to "catch up" with the rest of the group. Dave's model triggered a lively discussion of definitions and relevance. Two entries by a third user who came in the next day show how the model can be questioned, bound, and clarified by a computer conference:

```
You have seen 218 of the 230 entries made so far.  Please indicate
the entry number at which you wish to begin.  If you do not wish to
see any of the past entries, strike only the CR key.

#218

[218]  Dave  27-Sep-76  4:54 PM

Form #5 - diamond:  This form is off to the left of form #4.  If
the difference between weights is greater than X plus Y the subject
has no need to change these "weightings" when feedback is given.
In this case, an arrow from the left of form #5 goes back into form
#1.  If the difference is less than X plus Y, an arrow goes from
the bottom of the diamond to form #6.
```

Fig. 1. Example of the use of computer conferencing to discuss model development.

[223] George(Org) 28-Sep-76 12:17 AM
Just got Dave's flow chart figured out, I think. This looks really good, though I have a couple of questions, probably just misunderstandings; from diamond 4, if difference greater than X, do you go to diamond 8? And only the guess part of diamond 4 goes to diamond 5? Also, would it be okay to have also a "no weightings available" possibility within box 2?

[224] George (Org) 28-Sep-76 12:21 AM
This model of Dave's, which talks only about the acquisition of the weights, is perfectly compatible with the retrieval models thought to be operating during testing. I should have a simple program to you by Wednesday, and we'll try to incorporate Dave's process in it.

There are, of course, several limitations in this type of discussion structure: one which we have already mentioned is the inability to use graphics directly. This is a limitation that future computer networks hope to eliminate soon. The other limitation, which is of more concern to us, has to do with the logical organization: the participants do not currently have the ability to run models directly. Therefore, they cannot observe each model within the context of other formulations. The ideas embodied in the process are separated from the ideas involved in its use. The communication obstacles identified by Mar are given physical reality. To begin to address these obstacles, we turn to Richard Smallwood's conceptual framework for models.

Modeling and Communication

Eight years ago, Richard Smallwood described a conceptual framework [5] in which "model forms" are specified while certain parameters are left unknown. (For example, one might specify that component A will supply data to component B, but the frequency of this exchange of data could vary.) Within this "model space", successive decisions lead to the estimation of the parameters. Thus, several alternative models can be considered simultaneously for the same situation. This extension of the more classical approach (in which a single, fixed model is created) centers on the evaluation of the optimum settings for certain control variables. This situation is typical of modeling efforts we have already observed under real-world conditions through teleconferencing. In particular, such a situation is found in Richard's statement No. 6 quoted above: "you have the opportunity to change or examine . . . constants, initial values of variables, switches, etc".

The idea of dynamically revising the parameters of a family of models has been applied before, notably in a forecasting technique named KSIM. [6] KSIM combines a small group workshop procedure with a mathematical forecasting model and a computer program to generate changes over time. Although KSIM was developed specifically to support a study of water resource development alternatives, it illustrates a general trend toward reintroducing human judgment and expert opinion in the analytical computations usually associated with modeling.

An interesting concept inherent in KSIM (but not in the earlier formulation by Smallwood) is that of combining a mathematical forecasting model with a cross-impact analysis. Computer programs like KSIM present ways of articulating and visualizing what people sense to be the relationships among the interacting components. Dr. Julius Kane, who developed KSIM in the early 1970s, pointed out that "if computers are to be effective instruments of policy, then they must have open channels that can accept subjective data and give it its proper role".

A primary advantage of this modeling framework, then, is a clear division between the task of the computer and that of man. Smallwood writes:

> The crucial tasks of hypothesizing new model forms, developing the cost structure, and supplying the subjective inputs are the man's; the evaluation task is the machine's. The speed and memory capabilities of modern computers allow us to explore and evaluate a much wider spectrum of models than we would be able to otherwise.

Such a dynamic modeling process, even with the help of a computer, is far from a straightforward problem. Smallwood suggested that the following issues be examined: How should one decide whether to invest in additional data? How can value be placed on such data? What mechanism should be used in encoding subjective inputs? What computational tools are needed to aid in analysis? To these problems, one might add the types of concerns that emerge from studies of expert interaction: How can differing basic assumptions be reconciled when a large modeling effort is at stake? How can group judgment be best aggregated? [7] Computer conferencing should enable modelers to find more precise answers to these questions in the same manner already explored in some areas of simulation and gaming [8, 9, 10].

Beyond Computer Conferencing: Interactive Group Modeling

The concept proposed here is to build large models by enabling them to evolve dynamically out of a very flexible meta model which represents a computer analogue to Smallwood's "model space". The meta model can be viewed simply as a computer conference of a special kind in which some participants are nonhuman: they are mathematical routines or data bases which can either be previously developed modules, simulated modules, or even separate computers.

Smallwood has described the modeling process with the four components shown in Fig. 2. We can imagine that each of these would have several subactivities (Fig. 3). In the first stage (the requirements analysis), the initial group, under the leadership of a model director or facilitator, would define the goals and constraints of the model; they would also identify the data sources and the algorithms to be used. Some of these may already be in the form of computer programs. Then the model director would assemble the actual "meta model." It would consist of on-line data bases, mathematical modules, and human

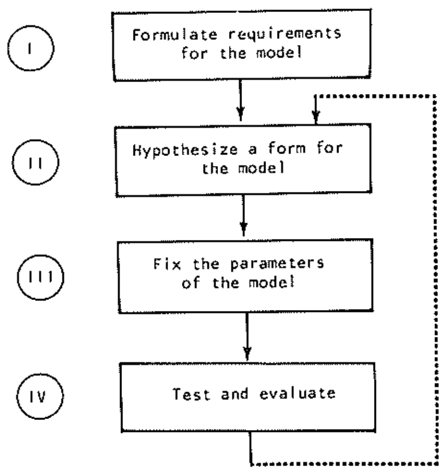

Fig. 2. Modeling process (after Smallwood).

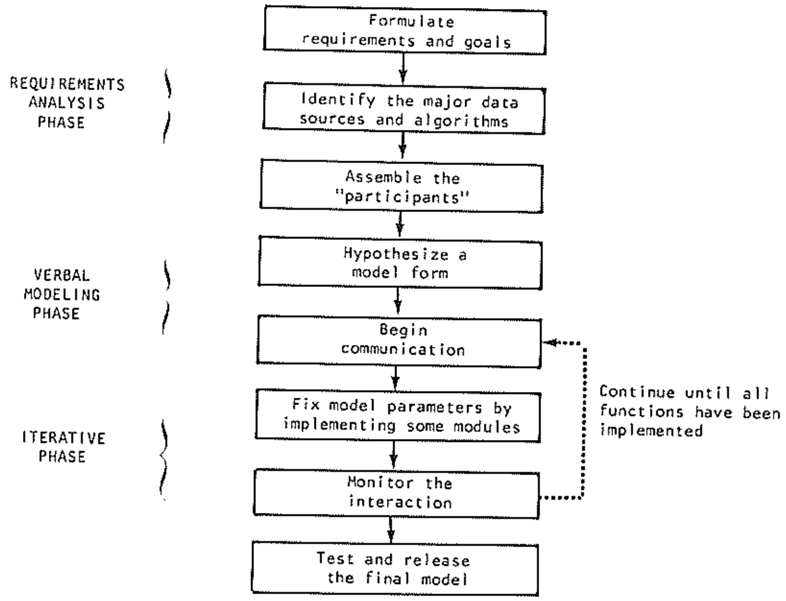

Fig. 3. Generalized process for interactive group modeling.

participants who can either supply missing information, provide quantiative estimates, or simulate parts of the model that have not yet been implemented or even formalized (such as a "social attitudes and values" module in a mining exploration and production model).

The second stage in Smallwood's model could be called the "verbal modeling phase". It would be divided into two activities: setting up an initial interaction pattern among the participants (for example, certain mathematical modules would receive their input from certain data bases); and actual interaction among the participants, that now include data bases.

The final two stages would be executed iteratively. In Stage III, some of the modules that were previously simulated by human participants in Stage II would be implemented as computer programs. Then, in Stage IV, to test and evaluate the model, complete monitor statistics would be examined to guide the optimization process for the next iteration. Stage IV is critical to the "convergence" of the modeling process. It would determine, for instance, which of the existing modules should communicate with others, what the value of information is for each one, how traffic flow should be organized, and, more generally, how optimization should take place.

Finally, the resulting model could be validated by the whole group and actual use could proceed. As an alternative, there might be a separate activity for a separate group to validate the model, as proposed by Scher [11].

Let us try to visualize how such a process might work in an actual modeling conference. We assume that seven sites will be involved in jointly creating a new model. At each site (which we denote as A, B, . . . G), there may be one or more participants equipped with computer terminals. Interaction begins in much the same way as in the actual examples given earlier. In the requirements analysis stage, participants discuss the topology of the situation they are going to represent. They attempt to answer questions such as: What types of data should be selected? What will be their representation? Which

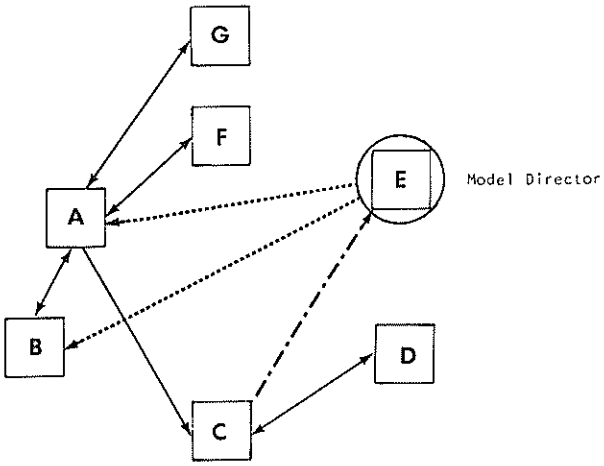

Fig. 4. The "verbal modeling phase": the computer conference before optimization.

operations are relevant? How should they interact? Then, in the verbal modeling stage, the participants themselves begin to *play the parts* of the model components until they can clearly see who needs to interact with whom or what the optimum interface will be (Fig. 4).

In the example of Fig. 5, the iteration phase results in the simplification of the model, component G being eliminated. As parts of the model become well understood, the human roles are replaced by program modules. Here, components B and D are now played by actual data bases that the four remaining human participants can address in the course of their discussion, all interaction becoming part of the public record.

One major advantage of computer conferencing in this situation is its ability to quantitatively monitor group interaction and to display traffic flow to the model leader. This ability could be used here to help in optimizing the topology of the model and to characterize its components. Such statistics as lengths of messages, interarrival times of

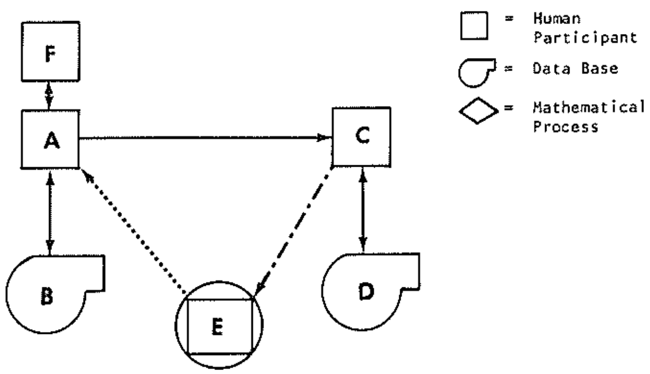

Fig. 5. Iteration phase: the computer conference after optimization. Participants B and D have been replaced by data bases. G has been eliminated. The topology of data flow has been determined, and the role of the director has been clarified.

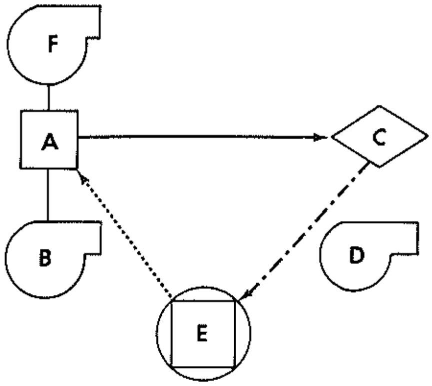

Fig. 6. Computer conference after implementation of several new components. The model has taken the form of a two-person game. Participant C is a mathematical program. Participant F is an on-line file.

requests, and volume of response would be useful in the design of large, interactive data-base systems such as a complex model may involve. (Under conditions of state-of-the-art computer networking, these systems need not reside on the same machine as the computer conference itself.)

In Fig. 6, the model is further simplified; all components except for A and E are now automated. The model can thus be viewed as a two person game. To give a practical reference to this example, components B and F might be data files on employment levels and consumer spending, while D and C might be a data base of certain reserves and a model of a distribution system, respectively. Component A would be a human participant representing the public's response to price changes while E would represent the decision-maker in a position to influence prices.

Figure 7 shows the final state of the model, in which all components except E are now computer processes.

In this example, we have shown a process that converged simply towards a single solution. This outcome may not be the general case. It may, in fact, be desirable to allow the interactive development and testing of alternative models on a competitive basis. Why shouldn't European economists be able to run a model against their own data base in an environment where the interaction is recorded for all to observe, to comment upon, and

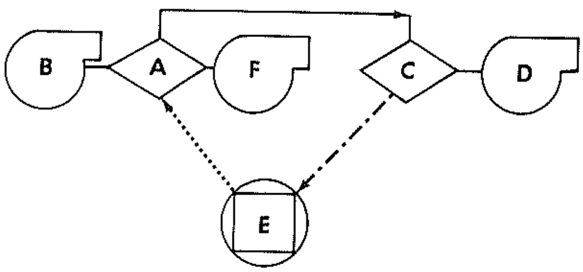

Fig. 7. Final state of the model: A has been replaced by a simulation process. The computer conference has converged towards a mathematical model with a single user.