A KNOWLEDGE-BASED METHODOLOGY FOR TUNING ANALYTICAL MODELS

R.S. Freedman*
G.R. Stuzin**

ABSTRACT

Many computer-based analytical models for decision-making and forecasting have been developed in recent years, particularly in the areas of economics and finance. Analytic models have an important limitation which has restricted their use: a model cannot anticipate every factor that may be important in making a decision. Some analysts attempt to compensate for this limitation by making heuristic adjustments to the model in order to "tune" the results. Tuning produces a model forecast that is *consistent with intuitive expectations*, and maintains the detail and structure of the analytic model. This is a very difficult task unless the user has expert knowledge of the model and the task domain. This paper describes a new methodology, called knowledge-based tuning, that allows a human analyst and a knowledge-based system to *collaborate* in *adjusting an analytic model*. Such a methodology makes the model more acceptable to a decision-maker, and offers the potential of improving the decisions that either an analyst or a model can make alone. In knowledge-based tuning, subjective judgments about missing factors are specified by the analyst in terms of linguistic variables. These linguistic variables and knowledge of the model error history are used by the tuning system to infer a specific model adjustment. A logic programming system was developed that illustrates the tuning methodology for a macroeconometric forecasting model that empirically demonstrates how the predictability of the model can be improved.

*Department of Computer Science, Polytechnic University
**Department of Computer Science, St. John's University.

To appear in *IEEE Transactions on Systems, Man, and Cybernetics*, Volume 21, Number 2, March, 1991.

Biosketch of Authors

Roy S. Freedman received his BS and MS in Mathematics in 1975, his MS in Electrical Engineering in 1978, and his PhD in Mathematics in 1979, all from Polytechnic University. He joined the Hazeltine Corporation Research Laboratories in 1979, and was responsible for the software engineering of several well-known AI systems. In 1985 he returned to Polytechnic University as an Associate Professor. As a consultant, Dr. Freedman has done extensive work for a broad range of clients in the financial services and telecommunication industries, including the New York Stock Exchange, Equitable Life, MCI, Chemical Bank, Intelligent Technology Inc. (Tokyo), Grumman Aerospace, and Hazeltine Corporation. In 1989, he founded Inductive Solutions, Inc., a firm that produces tools for building embedded knowledge-based systems.

Dr. Freedman has published over thirty papers in artificial intelligence and software engineering as well as two books, *Programming Concepts with the Ada Language* and *Programming with APSE Software Tools,* both now published by McGraw-Hill. He has been an Associate Editor of the journal *IEEE Expert* for the past four years, and is a member of the IEEE Computer Society, AAAI, ACM, and SIAM.

Dr. Freedman can be reached at the Department of Electrical Engineering and Computer Science at Polytechnic University, 333 Jay Street, Brooklyn, NY 11201 or at Inductive Solutions, Inc., 380 Rector Place, Suite 4A, New York, NewYork 10280.

Gerald J. Stuzin is an Associate Professor of Computer Science at St. John's University. He obtained his Ph.D. from Polytechnic University in 1989, and an MBA in Economics from New York University in 1972.

Since 1969, Dr. Stuzin has worked as an economic consultant at American Can Company, CBS, Inc., Union Carbide, and the International Forecast Group. He is presently a consultant at Business Planning Systems, an economics consulting firm. His current interests are in developing intelligent analytic tools.

Dr. Stuzin can be reached at the Department of Computer Science at St. John’s University, Jamaica, New York 11471.

I. Introduction

Many analytic models for decision-making have been developed in recent years. For our purposes, a model is a dynamical system

$$\begin{aligned}\mathbf{x}(n) &= \mathbf{A}(n-1)\mathbf{x}(n-1) + \mathbf{B}(n-1)\mathbf{u}(n-1) \\ \mathbf{y}(n) &= \mathbf{C}(n)\mathbf{x}(n) + \mathbf{D}(n)\mathbf{u}(n)\end{aligned} \tag{1}$$

where, at period n, $\mathbf{x}(n)$ is a vector in $\mathbf{R}^{\mathbf{m}}$ of state variables, $\mathbf{u}(n)$ is a vector in $\mathbf{R}^{\mathbf{k}}$ of control variables, and $\mathbf{y}(n)$ is a vector in $\mathbf{R}^{\mathbf{p}}$ of observed model variables. Here, $\mathbf{A}$ is an $\mathbf{m}$x$\mathbf{m}$ matrix, $\mathbf{B}$ is an $\mathbf{m}$x$\mathbf{k}$ matrix, $\mathbf{C}$ is a $\mathbf{p}$x$\mathbf{m}$ matrix, and $\mathbf{D}$ is a $\mathbf{p}$x$\mathbf{m}$ matrix. These matrices are assumed to be known. Usually, they are estimated by statistical methods.

The solution to (1) can be computed recursively given an initial state $\mathbf{x}(i)$, i = n-1 from

$$\begin{aligned}\mathbf{y}(n) &= \mathbf{C}(n)\mathbf{A}(n-1)\mathbf{x}(n-1) \\ &\quad + [\mathbf{C}(n)\,\mathbf{B}(n-1)\mathbf{u}(n-1) + \mathbf{D}(n)\mathbf{u}(n)] \\ &= \mathbf{C}(n)\mathbf{A}(n-1)\mathbf{x}(n-1) + \mathbf{c}(n)\end{aligned} \tag{2}$$

The expression $\mathbf{c}(n)$ in (2) that is independent of the model variables is called the *constant term*. The constant term is also usually estimated by statistical methods. Intuitively, this term relates the averaged effect of "excluded" model variables (called *factors*) to model variables. Formally, factors correspond to the effect of the control variables $\mathbf{u}(n)$ on the model variables. As the model evolves, factors may be formally introduced as model variables $\mathbf{y}(n)$ in a new model.

Model (1) approximates the behavior of a more complex "real-world" system

$$\begin{aligned}\mathbf{x}^*(n) &= \mathbf{F}^*(\mathbf{x}^*(n-1), \mathbf{u}^*(n-1), n-1) \\ \mathbf{y}^*(n) &= \mathbf{G}^*(\mathbf{x}^*(n),n)\end{aligned} \tag{3}$$

where functions $\mathbf{F}^*$ and $\mathbf{G}^*$ are not known. Given a metric $\mathbf{d}$, the model error at time n is defined to be

$$e(n) = \mathbf{d}(\mathbf{y}^*(n), \mathbf{y}(n)) \tag{4}$$

Model quality and sensitivity analysis is performed by evaluating the convergence and ergodic properties of the error term. Sensitivity analysis is evaluated by determining the effects of changing the initial conditions of the state variables.

Tuning is a process that is concerned with creating a *new model* $\mathbf{y}^t(n+1)$ of $\mathbf{y}^*(n+1)$ in terms of some function **f** of the historical errors e(n) and the predicted value of the old model:

$$\mathbf{y}^t(n+1) = \mathbf{y}(n+1) + \mathbf{f}(e(n)) \tag{5}$$

Tuning can be considered to create a *model adjustment* of one or more components of the constant term **c**(n) in Equation (2). The function **f** depends on the subjective judgements of the model users and model experts, and on the metric **d** that is used to define the historical error. Because of variable interdependencies, this model adjustment process results in a new set of equations to be solved. For example, after an assessment that the value of y_6 is "too high," an analyst may decide in a heuristic way to change the constant term for y_1 from 0.445 to 0.449 and the constant term for y_5 from -0.988 to -0.776. The justification for these changes can be based on the analyst's judgement regarding the effects of particular factors that are excluded as model variables. These adjusted values would then be propagated through the model and the new values would be assessed again. The tuning process can be iterated and stops when the model values are consistent with the analyst's judgement and intuition.

The rationale for tuning an analytic model is that *judgment* may serve to compensate for the following unavoidable deficiencies in the model:

- inadequate theory due to missing model variables and relationships
- short-term disturbances
- data revisions

Another motive for tuning a model is to produce a forecast of $\mathbf{y}^*(n)$ that is *consistent with intuitive expectations*, while maintaining the detail and at least some of the structure that an analytic model has to offer. As Pindyck and Rubinfeld [1] observe:

> There is a ...method that is often used to make minor adjustments in ... models, particularly those that are used for forecasting purposes. This method is called 'tuning' and consists of making small changes in some of the model's coefficients,...so as to improve the ability of the model to forecast.
> [Tuning has] come to be used in large... forecasting models, particularly those constructed for commercial or business applications (often they are adjusted to keep the forecast "in line" with intuitive forecasts - thus to some extent negating the predictions of the model). Needless to say, [these adjustments] can easily be misused (and often are).

Evans [2] comments on how common the practice is in econometrics:

> ...I am sure it is no secret that virtually everyone who uses an econometric model for forecasting does so only after he has adjusted the constant terms in some or even all of the stochastic equations.... Adjust the constant terms, incorporating as a guideline the average residuals of the previous period, but using judgment to adjust the residuals further.

This is typical in many other applications of algorithmic models: even though some models often outperform human decision-makers, the model outputs are not generally well-accepted. Many analysts will usually reject the model's conclusions, especially if they are presented with binary choice between the total acceptance of the model's decision and total rejection. Some reasons for this are:

- Even the best models may on occasion produce decisions much worse than a human analyst would, because some important factors have not been included.
- The models utilize uncertain theory as well as uncertain data.
- The analyst's risk preference in dealing with uncertain outcomes may differ from that of the model.
- The analyst's role is trivialized if decisions are solely provided by the model.
- Models provide precision at the expense of intuition and common sense.

What is needed is a methodology that allows a human decision maker and a knowledge-based system to *collaborate* in *adjusting the model* by explicitly producing model adjustments. Instead of explicitly changing constants and coefficients directly in the model, an analyst can have the following kind of interaction [3]:

Analyst: I think the value for consumption is too low.

Computer: I was told that consumption depends on the factors consumer debt and consumer confidence. The specific rules are:

1. Consumer debt is extremely important to consumption.
2. Consumer confidence is very important to consumption.

Would you like to enter your assessment about these factors?

Analyst: In this forecast for consumption, I think consumer confidence has a strong, positive impact. I think consumer debt has a very strong, positive impact.

Computer: Are there any other factors that you consider significant in this forecast?

Analyst: Yes. I think another factor you should consider is the stock market. The recent stock market crash will have a significant negative impact on consumption.

Computer: The percent change in the forecast for consumption based on a new forecast is -17%. Shall we continue?

The advantages to such an approach is to provide documented decisions consistent with user intuition. Ad hoc tuning can be replaced by an integrated analytic model/knowledge-based system that can explain and justify its model adjustments. Such a methodology makes the model more acceptable to an analyst, and offers the potential of improving the decisions that either an analyst or a model can make alone. Moreover, such a system can be used to elicit new domain knowledge for model evaluation and improvement [4]. A data-flow diagram illustrating this process is shown in Figure 1.

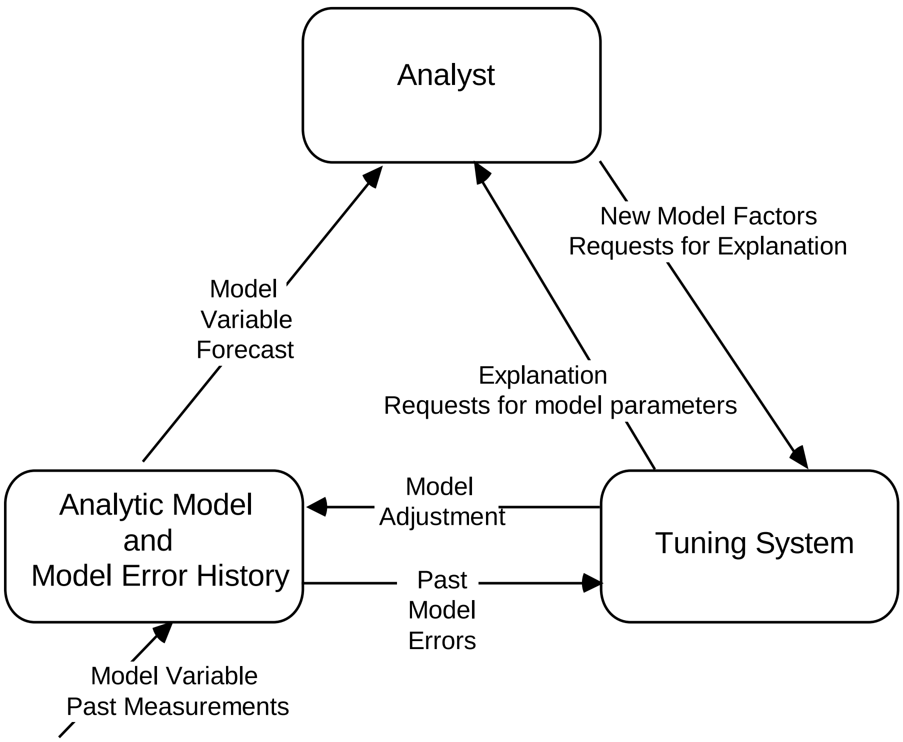

Figure 1. Data Flow for the Tuning Process

We have developed a knowledge-based tuning methodology that integrates a mathematical model, a knowledge-based system and a human evaluator. The mathematical model provides computational power and the underlying theory. The knowledge-base represents expert domain knowledge on managing and making subjective model adjustments to the model computations. The evaluator provides the domain knowledge to insure that the final result has practical usefulness.

In our research, we have used the models associated with economic forecasting as a case study. Economic forecasting is one of the four example "analytic languages" that was discussed in the context of judgement and analytic knowledge elicitation in [4], and is also an analytic language that we are most familiar (one of the authors was an economic consultant for twenty years). The specific model that we chose to illustrate our methodology is the macroeconometric model described in [1].

Requirements for the knowledge representations for tuning are described in Section II. The knowledge-based tuning methodology is described and demonstrated in Section III, where we also discuss the relationship between tuning and sensitivity analysis. In Section IV, we show an example that demonstrates how knowledge-based tuning can be used to help analysts improve the predictability of an econometric model.

II. Knowledge Representations for Tuning

A. Previous Approaches to Tuning

The desirability of developing techniques by which humans and computers collaborate in making decisions, rather than the decision being made by one or the other, has been recognized for some time. In 1961, Yntema and Torgerson [5] questioned how to combine the analytical speed of the computer with the "good sense" of the human user, without sacrificing too much of either. They proposed to let the machine make the decisions according to simple rules, but require the analyst to monitor the result and change the machine's answer if the analyst finds the results too foolish.

Computer models have become much more complex since 1961. However, it is still the case that all abstract models are only approximations and that optimization achieved with respect to the model is not the same as optimization with respect to the real world. This is particularly true in the case of forecasting. Ultimately, only a human can judge if the discrepancy between the real world and the model is large or small.

In 1964, Shepard[6] proposed an approach similar to that of Yntema and Torgerson. He noted the possibility of achieving *subjective* optimality by decomposing a decision process into a human effort and a computer effort. The human would be responsible for a set of elementary comparisons with respect to the underlying subjective variables. The computer would deal with the algorithmic process of combining the judgments. Similar division of labor ideas were also proposed in 1968 in the context of a Probabilistic Information Processing System [7] where humans would be used to estimate likliehood ratios and the computer would be used to compute payoff matrices.

More recently, Zimmer [8] discussed the possibilities of man/computer collaboration in forecasting. He suggested to elicit expert rules for qualitative predictions and combine inferences based on these rules with the results of quantitative forecasts. Rules that resulted in subjective predictions would have to be formally described, and algorithms would have to be developed to translate qualitative judgments into analytic parameters. His conclusion was that a system that integrates qualitative and quantitative techniques in this way would also increase the acceptability of forecasts.

The possibility of combining analytic techniques with ideas from artificial intelligence leads to a new kind of intelligent analytic tool. They are not so much intelligent assistants as they are *collaborators.*

From the perspective of knowledge-based systems, an intelligent analytic collaborator should relieve an analyst of routine computation and data handling. A collaborator should should also explain its reasoning. In this regard, collaborators are similar to apprentices [9] and a tutors [10] in that they compare user behavior to expert behavior and attempt to minimize the difference by negotiation. In the our methodology, tuning heuristics determine what constitutes "minimum difference." The objective of the negotiating process is to influence the user's behavior, making it as rational as possible from the perspective of the domain expert knowledge.

A general program for an "artificial laboratory" of such tools was also proposed in [11], where the assistance provided analysts was classified into three components: *model developers and representers*; *model testers*; and *model refiners*. Such tools may be used as collaborators in scientific discovery (model creation) as well as collaborators in model utilization.
In this context, our approach to knowledge-based tuning provides a new example of an intelligent computational tool that assists in the *refinement* of the model.

Examples of intelligent computational tools that assist in the development and representation of models are discussed in [12]. These systems collaborate with analysts in understanding and displaying some qualitative characteristic (like stability or periodicity of the solutions of differential equations).

Examples of intelligent computational tools that assist in the testing and maintenance of the model have also been demonstrate. Expert system methods and spreadsheet-based algebraic models were integrated to provide (model verification) advice on sensitivity analysis for financial problems [13]. In another example, [14] shows how a rule-based system can maintain the correspondence between model knowledge and semantic constraints that are important to the problem, but are not represented in the model.

B. Tuning Analytic Models Used in Decision Support Systems

Decision support systems might be described either as man machine problem-solving systems, or as interactive computer systems that assist a person in making decisions. These systems are most valuable in problems that are complex and quantitative enough to make computers useful, but still require a considerable amount of human judgment. Typically, in these problems what constitutes an "optimum" solution is ultimately a subjective determination. The concept of "satisficing" is applicable here [15]. The decision-making exercise ends when the analyst is satisfied with the decision.

Several mathematical models are currently being used to advantage in decision support systems. These include linear and non-linear programming, game theory, decision analysis, utility theory, queuing theory and time series analysis. Typical applications of the models include inventory policy, production scheduling, facility location, capital allocation and forecasting. The discussion below focuses on issues relevant to the application of decision support systems that are based on tuned analytic models.

A typical mathematical model for decision-making accepts parameter values as inputs and computes outputs which constitute the "decision." Experience to date shows that they can perform better in some domains than can knowledge-based systems. Most model-based decision support system lack the desirable features of a knowledge-based system, namely:

- the ability to accept linguistic input.
- the ability to add or delete chunks of knowledge.
- the ability to provide explanations and guidance on proceeding to the user's goal.

Tuning such a system can only be justified when users of the model have knowledge that bears on the decision which is not an input to the model, and is not part of the model computations. It is this extra-model information which drives the development of the tuning system. The representation of tuning knowledge has three major steps:

1. The determination of what type of user knowledge must be represented.
2. The representation of methods for incorporating the user knowledge into the decision process in order to adjust the model computations.
3. The development of an interactive architecture for man-model collaboration.

The first step is an exercise in knowledge engineering. Expert decision-makers in the specific problem domain, particularly those with experience using the model, will be the best source of the type of knowledge the tuning system should accept as input. However, since novice users will have the greatest need for guidance in using the model, an understanding of how they work must be reflected in the tuning system's knowledge base.

In the second step, an expert in the particular mathematical model must specify a method for modifying the model computations to reflect the information specified during the first step. This amounts to specifying the error metric **d** and the error evaluation function **f** (Equations 4 and 5 in Section I). We believe that the tuning heuristics specified in Section III can be applicable to any model, as long as certain domain specific parameters are also incorporated in the representation.

The design of an appropriate interactive system goes to the heart of the tuning process. Tuning a model is essentially a trial and evaluate activity. The model user does not know his "goal" in advance. This is to be expected. When a person must decide among a number of alternatives, each involving many factors, he cannot fully anticipate the consequences of his choices. However, when he sees the results of two such choices, he may very well be able to say which he prefers. The process is one of approximation with feedback correction. The decision-maker continues the process until he reaches a "satisfactory" outcome.

Probably the most difficult task in constructing a tuning system is the development of the knowledge for interacting with the user in order to insure an adequate feedback loop. The system must be controlled by the margin of error (as perceived by the user) with reference to an external goal, but the goal may be changing. Figure 2 illustrates the structure of this feedback representation. The analyst acts as evaluator, where error is defined in Equation (4) of Section I. The model errors are obtained by comparing the model variable predictions with the actual variable measurements. The historical error performance of the model is used by the analyst to create model adjustments, which, according to the analyst, reflects a "better" forecast as is viewed as a *model refinement*: if the model were perfect there would be no need for tuning.

Usually, the probability distributions of the model forecast errors are assumed to normal with zero mean. Since the errors are not biased, the errors themselves cannot be used to improve the forecast. Only "extra-model" information (such as that acquired through tuning) can be used to adjust the errors. Consequently, tuning (as an example of model refinement) can be viewed as the first step in the development of *new models* and *new model representations*, based on the results of the old model, the historical model errors, and the new parameters based on the tuning process.

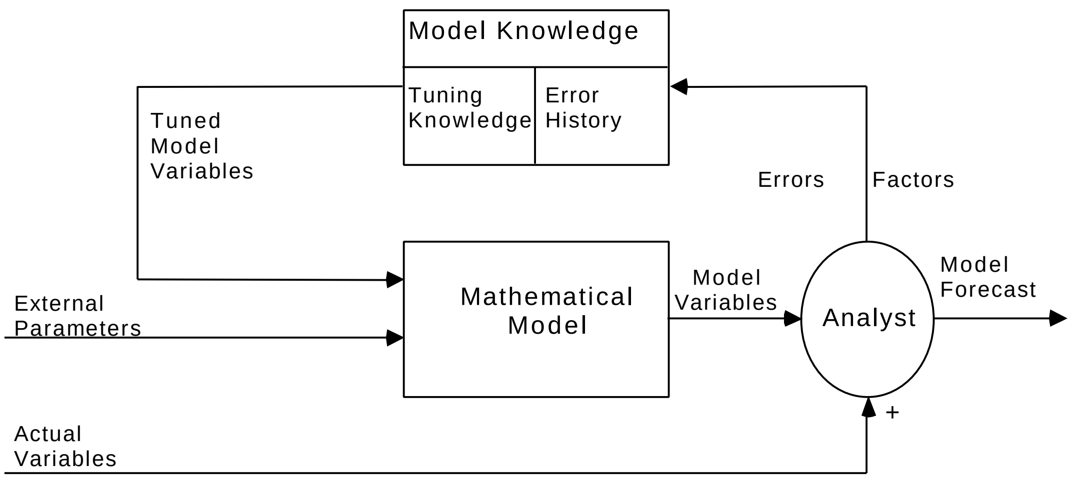

Figure 2. Tuning as a Feedback System

In designing the feedback loop, it must be remembered that as the tuning proceeds, the analyst may change goals and either retract old information or add information in a non-monotonic manner. Consequently, the tuning system must be stable in the feedback sense [16], and converge to a satisfactory outcome. Since the model itself is usually not designed to be stable *when tuned*, the burden for maintaining stability and validity falls to the rules of the tuning system and the user. The factors that will impact on stability correspond to the different types of knowledge in the system. The tuning system will very likely be so complex that insuring stability by purely analytic methods will not be possible. If desired, stability behavior can be demonstrated by simulation exptertiments. In fact, it will probably be necessary to tune the tuning knowledge-base, particularly in regard to the domain-specific parameters for **f**, in light of simulation experiments.

C. Difficulties in Representing Subjective Tuning Knowledge

The design of a knowledge-based tuning system is predicated on the assumption that although the user's subjective input is indispensable, it is also quite fallible and it should be used selectively. The assumption of the fallibility of human judgment in decision-making is based on numerous studies. Among the psychological tendencies which have been reported are:

Anchoring. This is the tendency not to stray from an initial judgment even when confronted with conflicting evidence. Experiments have shown [17] that the amount of probability revision made by the subjects, as indicated by the difference between posterior and prior probabilities, is consistently smaller than would be prescribed by Bayes' theorem. In other words, the maximum information possibly derived from experience is greater than what is actually learned. Subjects are reluctant to revise their opinion in light of experience. This may be related to what psychologists call "cognitive dissonance" [18], a theory explaining the tendency to come down excessively heavily on one side or the other when confronted with conflicting evidence.

Inconsistency. Given quantities A, B, and C, consistent behavior would require a subject to treat them as though they satisfied the following two properties:

1. Exclusivity of comparison.
   Either A > B *or* A < B *or* A = B.

2. Transitivity of comparison.
   If A > B *and* B > C *then* A > C.

However, violations of both properties have been seen. If a pair of alternatives is presented to a subject many times, successive presentations being well separated by other choices, a given subject does not necessarily choose the same alternative each time [19]. Sometimes the subject claimed that A > B and at other times that B > A. Shanteau [20] described a classic experiment in which "experts" were asked to judge samples of produce. When judged a second time, the experts frequently made different assessments. Edwards [21] reviews experiments in which subjects violated the transitive property in making choices and suggests that they arise from the inability of people to focus on the dimension in question, i.e., they are distracted by some other dimension.

Selectivity. This refers to using only a portion of the information available. Commonly, people use only those pieces of information that come readily to mind. People make poor decisions when they must take into account a number of attributes simultaneously: decision-makers may be aware of many different factors, but it is seldom more than one or two that they consider at any one time [6]. Similarly, it is observed that expert judgments are based on little information [20]. One reason for this is that experts are often influenced by irrelevant information.

Fallacy. This refers to the improper use of probabilistic reasoning. Common errors include
conservatism (the failure to revise prior probabilities sufficiently based on new information [17]) and calibration (the discrepancy between subjective probability and objective probability).

Representativeness. This refers to the focusing on how closely a hypothesis matches the most recent information to the exclusion of generally available information [22].

Other issues concerned with measuring the accuracy of knowledge is discussed in [23].

D. Representing Expert Knowledge

Knowledge-based tuning utilizes domain-specific knowledge in a way that is somewhat different from the utilization of domain-specific knowledge in expert systems.The important differences between the tuning and expert system utilizations of domain-specific knowledge are:

Non-monotonicity. In an expert system it is assumed that the rules are correct, at least to some specified degree of probability or confidence, and that there is enough knowledge to produce a satisfactory solution to the problem. In a tuning system the rules are tentative and assumed to be incomplete. It is expected that the user will supplement and revise the knowledge.

Integration. Expert systems are typically "stand alone" systems, solving problems that are important in their own right. A tuning system is embedded in a larger system which incorporates a mathematical model. It is the problem solved by the larger system which is of primary interest.

Autonomy vs. Collaboration. Expert systems are designed to perform a task ordinarily performed by a human expert. The objective is to emulate expert knowledge and inference procedures autonomously. Applied to decision-making situations, a classical expert system would obtain data from the system user, determine an optimum decision and explain its reasoning. A physician using a medical expert system, for example, enters information about a patient in response to questions and obtains a therapy recommendation and explanation.

Decision-makers do not wish to turn over control of a decision entirely to a computer. Just as a decision-maker is disinclined to surrender control of a decision to a mathematical model, he would not wish to surrender control to an expert system that tunes the model.

A tuning system must avoid imposing different normative reasoning procedures on analysts.

Availability of Expertise. In some decision-making areas it is not possible to develop a set of rules for an expert system that will produce satisfactory results: there may be no experts with sufficient knowledge. The knowledge base required would be extremely large in order to anticipate all the possible causes for tuning and all the appropriate model adjustments [24]. A tuning system must integrate the different sources of knowledge from experts, mathematical models and analysts.

E. Representation of Uncertain Knowledge

The underlying assumption in tuning a model is that an analyst has useful, albeit uncertain, knowledge that the model does not have. We have already discussed the limitations that humans have in dealing with uncertain knowledge. The problem now addressed is concerned with eliciting the maximum useful domain knowledge. The key to the problem lies in the methods of measurement used in eliciting the domain knowledge.

The way in which humans measure and describe their sensations has been a concern of psychologists for many years, and is the subject of the branch of psychology known as psychophysics [25]. A person's description of his response to a stimulus (his subjective measurement) is most commonly made relative to a standard scale. Research in psychophysics has established that the usefulness (in terms of accuracy and consistency) of a person's measurement depends greatly on the scale selected.

The principal decision to be made in measuring a human's tuning knowledge is whether to use numbers or words. Although numbers have the desirable properties of being precise and easy to manipulate, there are other considerations. The potential advantages of using words for measurement underlie the introduction of the linguistic variables used in fuzzy logic [26].

The usefulness of linguistic variables form the standpoint of human psychology is given justification in [26]. Linguistic input is superior to numerical input in "fuzzy" situations. For example, it has been found that a higher degree of response consistency is obtained if people are allowed to give imprecise verbal response about a fuzzy concept than if they are forced to give a numerical grading. Of course, there must be a methodology for defining and computing with the linguistic variables.

It is inevitable that any methodology for computing with linguistic variables will be somewhat subjective. For example, for tuning an econometric model, we can use linguistic variables to describe the relationship (as defined by an expert) between a causative factor and an affected variable. For example, "X is very important to Y" means that X has a certain potential for causing a change in Y. The strength of X at a given point in time is input by the analyst and is also expressed linguistically: "X is extremely strong". The analyst also uses linguistic variables to comment on a forecast value: " Y is significantly too low".

A problem related to the representation of uncertainty is the problem concerned with the combination of uncertainty. Knowledge-based tuning must employ heuristics to combine different sources of uncertain information, in order to produce a model adjustment reflecting a combination of sources.

A number of methods have been utilized by expert systems to reason with knowledge that is correct only to some specified degree of probability or confidence[27-29]. The feasibility of using a particular method will depend on the availability of the information that the method requires. A knowledge-based tuning system should not increase the burden on the analyst by requiring information that the analyst does not ordinarily have. Not only will this increased burden discourage the use of the system, but this kind of knowledge is likely to be unreliable. A knowledge-based tuning system should exploit the knowledge that is available in the form that it is used, and should minimize the requirements for a normative uncertainty calculus.

F. Logic as a Conceptualization for Knowledge-Based Tuning

A key representation decision in developing a knowledge-based system is the selection of a uniform paradigm for the conceptualization of knowledge. The paradigm must be adequate for representing subjective as well as objective knowledge. It is important to have a notation that is clear and allows for easy addition and modification of the knowledge. Furthermore, it should be easy to obtain solutions to problems and explanations of the solutions.

We advocate the logic for the basic representational paradigm for tuning. The specific benefits of a logic conceptualization are summarized as follows:

Factual Clarity. When domain knowledge consists of a number of facts expressing relationships between variables, the facts can be easily represented by unit clauses in logic. For example, consider the economic fact:

"Consumer confidence is an important factor influencing the level of consumption expenditures."

This can be expressed as a unit clause in logic using the functor *factor* as:

factor(consumer_confidence, important, consumption).

In logic programming, querying and deducing the logical consequences of such clauses is straightforward.

Flexibility. It must be easy to mix linguistic and numeric values. Knowledge-based tuning is concerned with both linguistic and numeric knowledge. It is convenient in logic to express functions to transform from one form of knowledge to the other. This can be done by utilizing a combination of the declarative and procedural interpretations of logic.

Non-monotonicity. The meta-language capability of logic is adequate for expressing the addition or deletion of tuning knowledge interactively while maintaining consistency.

Integrability. The integration of domain knowledge and meta-knowledge is natural and transparent. Knowledge of the problem domain by itself is not adequate for a knowledge-based tuning system. Meta-knowledge, to enable the system to convert user input into useful domain knowledge and to interact appropriately with the user is also necessary. Expressing meta-knowledge is difficult in expert system shells and in conventional programming languages, but is straightforward in logic programming.

Uncertainty. There must be sufficient flexibility to handle special heuristics for expressing and computing with uncertain knowledge. Most expert system shells have built-in methods for dealing with uncertainty. However, knowledge-based tuning must also employ its own heuristics for expressing and computing with uncertainty. The meta-language capability of logic is well-suited for expressing the tuning heuristics efficiently.

III. A Methodology for Tuning

A. Representing Model Adjustment Knowledge

There are three categories of model adjustment knowledge. The first category relates excluded model variables (called *factors*) to model variables. With respect to the notation in Section I, factors can correspond to the introduction of a new variable $u_{k+1}(n)$ to the model (and an increase in dimensionality in the model).

Factors can be of two types. The first type, called *standard factors*, are those factors due to expert knowledge about model variables. These factors are recorded in the knowledge-base *a priori*. The second type are called *user factors*. These factors are elicited from the user interactively.

Here are some examples in econometric modelling:

> The level of *consumer debt* is very important to the variable *consumption.*
> The performance of the *stock market* is important to the variable *consumption.*

These assertions are represented as the unit clauses

```
factor(consumer_debt, very_important, consumption).
factor(stock_market, important, consumption).
```

The second category relates model variables to each other. This knowledge can be inferred from the actual model. For example, by inspecting the model we can determine that:

> Consumption expenditures depend on non-residential investment, residential investment, and inventory investment.

This can be represented as the unit clause

```
variables_affecting(consumption,
                    [non-residential, residential, inventory]).
```

where the square brackets represent a list of objects.

The third category represents the historical forecast errors made by the model. An example of this is the following assertion:

> The maximum historical error in forecasting the expenditures for consumption is 9. 75.

This is represented as the unit clause

```
max_error_hist(consumption, 9. 75) .
```

Clauses are also used to express the heuristics that map the linguistic terms into numerical values. For user factors, the term *Significance* describes the current impact of a factor on a variable as estimated by the *user*. For example:

> A factor that is *extremely significant negatively for consumption* will cause an error that equals -50% of the maximum error for consumption.

For standard factors, significance depends on the *Importance* that an expert assigns that factor, as well as the *Strength* that the user assigns to the factor. For example

> A *very important* factor for consumption can cause an error as great as 50% of the historical maximum for consumption.
>
> A factor for consumption that is *very strongly positive* is at 80% of its maximum possible effect.

The heuristic that we use is

$$Significance = Importance \cdot Strength$$

The difference between the concepts of "importance" and "significance" is that the former describes the expert's opinion of the potential for effect, while the latter is the user's opinion of the actual effect.

The subjective appraisal of factors and the evaluation of the current output (forecast) are also represented with linquistic variables.

The forecast for consumption is much too low.

The use of linguistic expression of knowledge is a key element in our method for tuning. Our method uses seven intervals to map linguistic values into numerical values. The placement of the intervals was made by trial and error and then modified on the basis of tuning experiments with experts.

The use of seven ranks conforms to the results of experiments by psychologists. For example, Miller [30] stated if a subject is asked to order stimuli according to the magnitude of a given stimulus, then the subject will be able to use only five to nine ranks efficiently.

The objective of a tuning session is to utilize available model knowledge to anticipate the error in the model forecast and make constant adjustments accordingly. Since we assume that the historical forecast errors of the untuned model are known, we utilize these errors to establish a scale in mapping the importance of qualitative verbal information in anticipating forecast error. For example, if the sample (historical) forecast error for an economic variable is s (which can be evaluated in terms of the standard deviation of the error) then our heuristic is that the importance of a piece of information should be measured in units of s. The heuristic of scaling based on s can also be justified in terms of standard probabilistic inequalities. If we assume that the forecast error in normally distributed about the mean, then a value of ±2s provides a 95% confidence interval for the maximum error about the mean, and gives a good initial estimate of the maximum error to be expected. We denote the absolute value of the maximum expected error by M. Our tuning rule is that verbal qualitative descriptions of standard and user factor effects are ultimately translated into a percentage of M. We note that the errors themselves cannot be used to improve the forecast, Since the errors are not biased. Only "extra-model" information (such as that acquired through tuning) can be used to adjust the errors.

For many applications (in particular, in econometrics), the use of a normal distribution is justified by the fact that the variables are in most cases assumed to be normally distributed. In fact, for models constructed by using multiple linear regression techniques, the forecast error is normal with mean 0.

The linguistic variables used are allowed to vary over seven values. Consequently, for tuning model variable y_k at period n, the constant adjustment ? (n) for a single factor is

$$? (n) = \textit{Significance}\ (n) \cdot M(n)$$

for tuning with a single *user* factor;

? (n) = *Importance*(n) ·*Strength (n)* ·M(n)
for tuning with a single *standard* factor

Significance is one of the seven values $\pm r_1$, $\pm r_2$, $\pm r_3$, 0, with $0 < r_1 < r_2 < r_3 = 1$. The terms used for the values and the transformations between the terms and numerical values for the subjective assessment of *Significance* are {Extremely-Significant-Positive, Very-Significant-Positive, Significant-Positive, Neutral, Significant-Negative, Very-Significant-Negative, Extremely-Significant-Negative}, corresponding to { r_1, r_2, r_3, 0, $-r_1$, $-r_2$, $-r_3$}.

The selection of the r_k is a matter of expert judgment. Only experience can establish the usefulness of the selection, and, no doubt, different users will feel differently about the results obtained. It would be a simple matter to customize our technique for different users by modifying the values of r_k.

For the econometric model, experiments were made to examine the impact of changing the values of the r_k. It was found that in this model, the system performance does not seem overly sensitive to the choice of values. For this model, the values used are {r_1, r_2, r_3} = {0.5, 0.25, 0.125}.
The effect of this is that an "extremely significant" factor causes a constant adjustment of 50% of the maximum allowed. This was acceptable according to the economic forecasting experts that we interviewed. For the econometric model, "Very significant" was initially chosen to have an effect one-half as great as "extremely significant", and "significant" an effect one-half as great as "very significant". The doubling of effects in this scheme should help to avoid conflicts in the interpretation of the terms.

The terms used for the values and the transformations between the terms and numerical values for the subjective assessment of *Importance* are {Extremely-Important-Positive, Very-Important-Positive, Important-Positive, Neutral, Important-Negative, Very-Important-Negative, Extremely-Important-Negative}. For this model, the values used are {0.5, 0.25, 0.125, 0, 0, 0, 0}. Only three of the seven intervals are used in this case: in *our* judgement, information on factors that are not important in a positive way should not be utilized in making constant adjustments.

The constant for "extremely important" for a standard factor corresponds to "extremely significant" for a user factor and was chosen as 50% for the same reason.

When the potential effect of a factor is multiplied by the *Strength* of the factor, the actual effect is obtained. The scale used for the values and the transformations between the terms and numerical values for the subjective assessment of *Strength* is the same as that for *Significance*.

The terms used for the values and the transformations between the terms and numerical values for the subjective assessment of *user evaluation of the forecast model* may also be used to iteratively converge on an appropriate model adjustment. Convergence is based on a "coarse tuning" and a "fine tuning" paradigm. In coarse tuning, the scale is used to modify the forecast variable using the r_k values for *Significance*. In fine tuning, the parameter M is itself tuned by a variable *Evaluation*, where

$$M^{t+1} = M^t + \mathit{Evaluation} \cdot M^t$$

Here M^t is the value of the tuned maximum after the t-th tuning iteration. The terms used for the values and the transformations between the terms and numerical values for the subjective assessment of *Importance* are {Greatly-Too-High, Substantially-Too-High, Slightly-Too-High, Neutral, Slightly-Too-Low, Substantially-Too-Low, Greatly-Too-Low}. For the econometric model, experiments were made to examine the impact of changing the values of the r_k. For this model, the values used are $\{r_1, r_2, r_3\} = \{0.2, 0.1, 0.05\}$

Consequently, tuning a model variable y_k with a single user factor at iteration t is reduced to computing

$$?^t = \mathit{Significance} \cdot M^t$$
$$M^{t+1} = M^t + \mathit{Evaluation} \cdot M^t$$

where -1=*Significance* =1and -1= *Evaluation* =1 are computed from the scaled values of the r_k as described above which are input by the user. The process of coarse and fine tuning are illustrated in Figure 3.

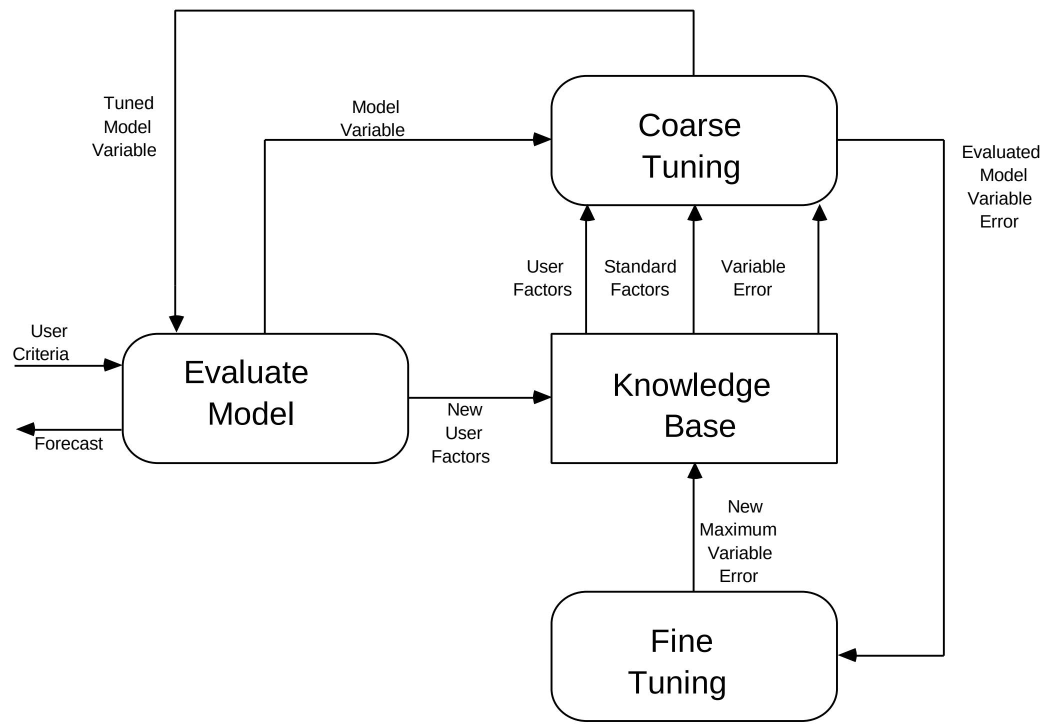

Figure 3. Coarse and Fine Tuning.

B. Rules for Combining Single Model Adjustments into a Net Model Adjustment

In this case, for a particular model variable, a model user has identified more than one factor that are not explicit in the model, but which is believed will cause the forecast of the model variable to be in error.

Let F_1, F_2,...F_n be a set of factors for a model variable whose maximum forecast error is M. We can compute a set of individual model adjustments for each *Significance* $_1$, *Significance* $_2$,... *Significance* $_n$. The problem here is to find a function H, such that the model adjustment can be computed as

$$y^t_k \qquad = y_k + H(\mathit{Significance}_1, \ldots, \mathit{Significance}_n) \cdot M^t$$

where $-1 = H(s_1, \ldots, s_n) = 1$ for $-1 = s_j = 1$.

Our requirements for H are based on the following observations:

1. A simple sum of the separate errors may violate the bound on H.

2. Users are frequently unable to estimate the combined effect of all the factors. The reasons for this were discussed in Section II.

3. $H(s_1, ..., s_n) = \max s_j$, for $s_j > 0$ and $H(s_1, ..., s_n) = \min s_j$, for $s_j < 0$.
   This heuristic specifies that the combination of errors should be greater than the largest error, if all errors are of the same sign.

4. For two errors of different signs, $H(s_1, s_2) = s_1 + s_2$.
   This provides for the "balancing" of positive and negative errors.

Our method specifies the function H in terms of a variation of Bernoulli combination that is used in probabilistic logic [29] .

We define the function H in terms of a binary function that is associative and commutative:

$$
\begin{aligned}
H(a,b) &= a + b - a \cdot b, && \text{if } a > 0 \text{ and } b > 0. \\
&= a + b + a \cdot b, && \text{if } a < 0 \text{ and } b < 0. \\
&= a + b, && \text{if } a * b < 0.
\end{aligned}
$$

For more than two factors, H is applied recursively, combining positive and negative terms separately and then algebraically summing the two results. Thus, given a partition of $\{s_1, ..., s_n\}$ into positive and negative subsets: $\{s_{p1}, ..., s_{pm}\}$ and $\{s_{n1}, ..., s_{nm}\}$, we form

$$
\begin{aligned}
H(s_{p1}, ..., s_{pm}) &= H(s_{pm}, H(s_1, ..., s_{pm-1})) \\
H(s_{n1}, ..., s_{nm}) &= H(s_{nm}, H(s_1, ..., s_{nm-1}))
\end{aligned}
$$

so that

$$
\begin{aligned}
H(s_1, ..., s_n) &= H(H(s_{p1}, ..., s_{pm}), H(s_{n1}, ..., s_{nm})) \\
&= H(s_{p1}, ..., s_{pm}) + H(s_{n1}, ..., s_{nm})
\end{aligned}
$$

The correct sign is then applied to the result (+ for positive errors, - for negative errors).

H can be compared to other systems for combining uncertainty. In probabilistic logic, the "measure of confidence" of the truth of a statement is expressed as the expected value of the probability of the statement over all consistent probability assignments [28]. Let a denote the measure of confidence of statement A and b denote the measure of confidence of statement B. It is well known [28] that

(i) $\max(a,b) = c(A \text{ or } B)$

(ii) $\max(a,b) = a + b - a \cdot b = \min(a+ b, 1)$

The inequalities in (ii) can be considered to correspond to the assumptions of statistical dependence, independence, and mutual exclusivity of A and B. If we consider H(a,b) = c(a,b) to be a measure of confidence that a certain amount of error will occur that is due to A or B, our heuristic for combining the positive and negative estimates corresponds to the independence assumption of the second inequality. If we assume dependence a and b, that is, if H(a,b) = max(c(a),c(b)), then the heuristic for combining the positive and negative estimates corresponds to the dependence assumption of the second inequality. This is the uncertainty calculus used in fuzzy logic. Our definition of h is also similar to the combining function used in MYCIN [27].

The function h can be generalized. In fact, Frank[31] showed that for any associative function
$F^*(x,y) = x + y - F(x,y)$ with 0=x=1 and 0=y=1, F must be of the form

$$F(x,y) = \log_s[1+ (s^x-1)(s^y-1)/(s-1)]$$

Our rule corresponds to the limiting case s=1.

C. Tuning versus Sensitivity Analysis

Tuning should not be confused with sensitivity analysis. Sensitivity analysis (sometimes called "what if" analysis) is the study of the behavior of a mathematical model with respect to defining parameters, with the objective of learning about the problem domain and, thereby, improving decision-making. The method of sensitivity analysis may be analytical, as in the case of linear programming, or experimental (by simulation) when the model is very complex, as in spreadsheet analysis. An advisory system for sensitivity analysis utilizing knowledge-based methods has been developed by Apte and Dionne [13]. In the case of macroeconometric models, sensitivity analysis is used for "policy analysis", i.e., examining the effect of government policies such as money supply and spending level. This would mean solving the model for various values of the model variables for money supply and spending.

In tuning, the objective is to *modify the model* so that it reflects the knowledge of the user, i.e., the model user does not believe that the model adequately represents the problem domain and, therefore, makes adjustments in the model computations. The purpose of a knowledge-based system for tuning is to assist the model user in tuning by making expert knowledge available to him. A hypothetical example may serve to further illustrate the distinction.

Consider an investment model designed to give an investor an optimum mix of investments at a specified level of risk. Such a model may utilize techniques such as mathematical programming and portfolio theory, although to a user it may be considered a "black box". Inputs to such a model would typically include current yields on the different available investments and a measure of their price volatility. The investor would also indicate the amount to be invested and his risk preference.

Given the model described above, the investor can perform a sensitivity analysis by varying the inputs, but he cannot interact with the model in a way that would allow him to indicate that he thinks the portfolio selected is "too high" or "too low" in some investment. A tuning interface for the model, however, could convert such comments into quantitative relationships in the model, e.g., by changing the risk or yield calculations, or by adding constraints on the amounts of particular investments.

IV. Example: Tuning a Macroeconometric Model

A. TUNES

An econometric model is a mathematical representation of economic behavior, arrived at by using statistical methods. Typically, such models are developed by using multiple linear regression to determine the coefficients of linear equations hypothesized by economic theory. The models range in size from six equations to more than one thousand equations.

Few, if any economic forecasters accept the "pure" forecast produced by macroeconometric models [2]. Even the most complex model must exclude a considerable amount of information that is important in the "real world". Much of the "judgment" that motivates an expert to tune a model stems from that information and its relationship to the model equations. In the practical application of econometric models for forecasting, many forecasters believe that if they make "adjustments" to the constant terms in the stochastic equations, then they can obtain better forecasts. As Armstrong [32] states,..."the more important the forecast, the greater is the likelihood that subjective methods will be used. It is in this area that [surrender of control] is threatening to the stakeholder. "

Tuning a large macroeconometric model is a difficult exercise. A user of such a model must not only be in possession of current economic information, he must know the structure of the model and understand the relationship of the current economic information to the model. Finally, the user must be familiar with the computer system requirements for running the model. The complexities are such that a novice user will frequently be dissatisfied with the results he obtains. It is natural, therefore, to expect that he would benefit from a computer program that would provide him with tuning expertise.

We have developed a prototype logic programming system called TUNES [3] that illustrates our knowledge-based methodology for tuning analytic models. The tuning system is written in Prolog. The model is based on a textbook example [1] of a macroeconometric model. The model has also been represented in Prolog to facilitate communication between the tuning system and the analytical system.

The economic rules stored in the knowledge base for TUNES were created by one of the authors (who was an economic consultant for twenty years). The rules are representative of knowledge used by economists when they tune macroeconometric models. The knowledge is highly subjective, and experts will differ greatly in the rules they employ. The tuning rules for standard factors are based on statistical indicators reported by the US Government. For example, one rule states that consumer debt (a statistical indicator which measures how much money consumers owe) has the potential for causing an "extremely important" adjustment in the model's forecast for consumption. The information about the strength of this factor is entered by the analyst. The mapping of the linquistic values into quantitative values is described in Section III.

TUNES does not utilize a natural language interface. However, the menu system employs the use of linguistic variables. The ability to add or delete chunks of user or standard factor knowledge is expressed by means of the assert and retract commands in Prolog. The ability to provide explanations and guidance on proceeding to the user's goal is also provided by record the Prolog backtracking interpreter. Improvements in both of these capabilities are being considered for the next version of TUNES.

B. Example of a Tuned Forecast

In this session, a user indicated that there were two standard factors and four user factors that affects consumption. The user also indicated that the forecast was too low: consequently, this caused the fine tuning component to make further adjustments. The following provides the trace of an explanation for the tuned forecast for the variable consumption:

*With Respect to consumption
The base forecast is 560.6
You included the following standard factors:

consumer_confidence
TUNES considers consumer_confidence to be very important.
YOU consider consumer_confidence to be significantly positive.

consumer_debt
TUNES considers consumer_debt to be extremely important to consumption.
YOU consider consumer_debt to be significantly positive.

You indicated the following new factors:

drought
YOU consider drought to be significantly negative.
stock_market
YOU consider stock_market to be extremely significantly positive.

The constant adjustment resulting from the standard and user factors was 8.4.

*You requested to fine tune the variable consumption reflecting your forecast evaluation of substantially too low.

To do this, TUNES increased the constant adjustment from 8.4 to 9.3.

The new forecast based on current values for all constant adjustments is:

Variable		Actual	%Change from Base Forecast
GNP	gross_national_product	996.2	3.3
C	consumption	570.5	2.6
IIN	change_in_inventories	18.9	57.8
INR	non_residential_investment	101.7	6.5
IR	residential_investment	32.2	-1.2
YD	disposable_income	871.4	6.5

C. Predictability of Tuned Systems

An experiment was performed with TUNES to test its ability to perform its fundamental objective: to convert an analyst's "extra-model" information into an appropriate quantitative impact on a model forecast. A secondary objective is to evaluate the forecasting ability of a tuned econometric model. It must be emphasized that this secondary objective will be met if the extra-model information is correct and relevant, and if the analyst uses TUNES in a rational manner. Consequently, the results obtained in any experiment with TUNES is dependent on the particular analyst and on the information available.

In our experiment, three analysts were provided with "extra-model" information to be used to tune a given econometric model. They tuned the model with and without TUNES. The analysts were a businessman with experience in economic forecasting (User 1); and two university students who were familiar with econometric modeling. All analysts should be considered novices rather than experts in tuning. The information provided is shown in Figures 4 and 5.

	Year:Quarter					
	72:4			73:1		
Variable	*Forecast*	*Actual*	*%²*	*Forecast*	*Actual*	*%²*
C	536.4	542.2	1.07	543.8	552.90	1.65
INR	83.1	87.2	4.70	86	97.20	11.52
IR	32.5	34.8	6.61	33.3	35.00	4.86
IIN	2.9	8.8	67.05	5.8	7.30	20.55
GNP	919	937.2	1.94	937.5	961.00	2.45

	73:2			73:3		
Variable	*Forecast*	*Actual*	*%²*	*Forecast*	*Actual*	*%²*
C	561.70	553.7	1.44	555.90	555.40	0.09
INR	96.30	94.3	2.12	90.20	95.10	5.15
IR	33.30	34.1	2.35	30.40	32.60	6.75
IIN	4.80	7.8	38.46	4.00	8.00	50.00
GNP	968.70	962.4	0.65	952.00	962.50	1.09

Figure 4. Quantitative Information Provided to Forecasters.

1. Consumer debt increased substantially in recent months.
2. New car sales were down moderately in the prior quarter.
3. There was a large increase in the inventory to sales ratio in recent months.
4. There were indications that the Federal Reserve policy was becoming less restrictive.
5. Housing starts were up moderately in recent months.

Figure 5. Qualitative Information Provided to Forecasters.

Since the qualitative information provided in Figure 6 was deliberately selected to be useful, in order to permit successful tuning, we expected that the tuned forecasts would be better than the untuned forecast. This was generally the case (the results of the experiment are shown in Figure 6).

Variable	*Actual Values*	*Untuned Model*	*%²*
C	546.3	560.6	2.62
IIN	20	21.5	7.50
INR	96	100.9	5.10
IR	29.8	32	7.38
GNP	964.9	987.8	2.37
		Mean Error	5.00

	User#1				*User#2*			
Variable	*Alone*	*%²*	*TUNES*	*%²*	*Alone*	*%²*	*TUNES*	*%²*
C	554.6	1.52	554	1.41	560.10	2.53	553.2	1.26
IIN	19.9	0.50	20.2	1.00	20.60	3.00	17.8	11.00
INR	99.7	3.85	100.3	4.48	100.10	4.27	99.6	3.75
IR	33.5	12.42	29.8	0.00	34.20	14.77	29.6	0.67
GNP	980.5	1.62	977.1	1.26	987.90	2.38	973	0.84
	Mean Error	3.98		1.63		5.39		3.50

	User#3			
Variable	*Alone*	*%²*	*TUNES*	*%²*
C	553.4	1.30	552.9	1.21
IIN	9.3	53.50	13.8	31.00
INR	98.7	2.81	99.1	3.23
IR	33.6	12.75	29.5	1.01
GNP	967.8	0.30	968.1	0.33
	Mean Error	14.13		7.36

Figure 6. Comparisons Between Forecast for 1973: 4th Quarter.

It should be noted that the variables being forecast are *not independent:* GNP, for example, is essentially the sum of the other four variables , plus a constant. Moreover, some of the variables are more difficult to tune than others. The variable IIN (change in inventories) is very difficult to forecast (and tune); this explains the large forecast errors obtained by User 2 and User 3.

For practical forecasting applications, some variables are much more important than others. Measuring the *quality* of a set of forecasts for a number of economic variables is therefore somewhat subjective. Standard statistical tests of forecast “correctness” have not been found to be a good guide for selecting a forecast methodology [32,33]. Nevertheless, it may be of interest to consider the results of a t-test for paired differences of the means of the forecast error for the above experiment. Using the above values, we compute the t statistic to have a value of 2.359. A one-sided t-test with 95% significance

(3 degrees of freedom) has a critical value of 2.353. Hence, we can accept the hypothesis that analysts using TUNES can tune models better than analysts that do not use TUNES.

In practice, success in forecasting GNP is the single most important criterion in judging a forecast's quality. Using that criterion, of the six tuned forecasts obtained (two per analyst), five were better than the untuned forecast, and one was essentially the same. This merely confirms the general assumption that tuning can improve model forecasts if analysts are given good information. What is significant is the comparison of the tuning results: all users improved their performance, and all reported that they had spent significantly less time preparing a forecast with automated assisstance then without.

Given the limited scope of the experiment, one cannot definitely conclude that TUNES will provide *more accurate forecasts*. One can argue that the *tuning pro*cess is more effective: the TUNES methodology helps transform extra-model information into model adjustments in a consistent way as opposed to individual ad hoc methods.

V. Conclusion

We have developed a new approach to computer-assisted decision-making, utilizing knowledge-based tuning methods. In this approach, a computer model, a knowledge based program and a model analyst user collaborate in making a decision. The degree of impact of each of the three partners in the decision, as well as the convergence of the decision, is completely controlled by the analyst. The motivation for this approach is the often expressed dissatisfaction with the two approaches most commonly employed now in decision-making: (i) informal methods stress human intuition too much, and (ii) computer modeling eliminates human intuition.

On the other hand, too much reliance on intuition may decrease the predictability of the model. Our tuning methodology explicitly acknowledges this reliance on intuition (and the existence of potential user fallacies in model interpretation), and controls them implicitly:

- The magnitudes of the constant adjustments are constrained by historical knowledge, rather than determined solely by personal judgement: this can help eliminate anchoring.

- The tuning process is based on relationships based on expert knowledge, rather than the more fallible user knowledge. Consequently, reasoning will be more *consistent* (with respect to an expert).

- The tuning heuristics for combining constant adjustments due to separate factors improves user's ability to combine information.

It is believed that the tuning methodology, as demonstrated for economic forcasting, shows the feasibility of the new approach for developing another type of intelligent analytic tools that help users modify models.

There are a number of possible extensions to the tuning methodology:

1. It would be desirable for a tuning system to have a learning capability with regard to linguistic interpretation. Different analysts use words like "significant", "important", etc. differently and in different contexts. By noting the analyst's responses to the system's actions, a user profile for different contexts could be developed.

2. The tuning knowledge-based approach reflects the expertise of a single expert. Conceivably, an analyst could find that expertise unsatisfactory. It would be useful to have alternate knowledge-bases available that reflect the expertise of named experts. The analyst could switch between experts as desired and observe the effects of tuning according to a particular expert..

3. The tuning system attempts to justify the forecasting of novice forecasters by utilizing expert knowledge and by structuring the sequence of inputs. Nonetheless, an unsophisticated user might enter input or take actions that violate the consistency of the model. A knowledge base could be included that would be used to detect these errors and issue warning statements. Moreover, a scenario generator can be developed to allow the user to ask what facts would justify a particular forecast. The use of logic programming techniques make this feasible, although developing the appropriate knowledge base would be difficult.

VII. References

[1] R. S. Pindyck and D. L. Rubinfeld, *Econometric Models and Economic Forecasts*, McGraw-Hill, 1976.

[2] M. K. Evans, "Econometric Models, " in *Methods and Techniques of Business Forecasting*, Butler, Kavesh, Platt (eds.), Prentice-Hall, Inc., 1974.

[3] G. Stuzin, "A Knowledge-Based System for Tuning A Macroeconometric Model," *Polytechnic Notes on Articial Intelligence*, vol. 7, 1989.

[4] B. Fischoff, "Eliciting Knowledge for Analytical Representation," *IEEE Trans. Syst. Man Cybern.*, vol. SMC-19, no. 3, pp. 448-461, 1989.

[5] D. B. Yntema and W. S. Torgerson, "Man-Computer Cooperation in Decisions Requiring Common Sense", *IRE Transactions on Human Factors in Electronics*, HFE-2 (1961), pp 20 - 26. Reprinted in W. Edwards and A. Tversky, *Decision Making*, Penguin Books, 1967.

[6] R. N. Shepard, "On Subjectively Optimum Selections among Multi-attribute Alternatives", in M. W. Shelley and G. L Bryan, *Human Judgments and Optimality*, John Wiley, 1964, pp 257 - 281. Reprinted in W. Edwards and A. Tversky, *Decision Making*, Penguin Books, 1967.

[7] W. Edwards, L. Phillips, W. Hays, B. Goodman, "Probabilistic Information Processing Systems: Design and Evaluation," *IEEE Trans. Syst. Sci. Cybern.*, vol. SSC-4, no. 3, pp. 248-265, 1968.

[8] A. C. Zimmer, "A Model for the Interpretation of Verbal Predictions", in *Developments in Expert Systems*, M. J. Coombs (ed.), Academic Press, 1984.

[9] A. Barr and E. A. Feigenbaum, *The Handbook of Artificial Intelligence,* vol. 2, Heuristech Press, 1982.

[10] W. J. Clancey, *Knowledge-Based Tutoring*, MIT Press, 1987.

[11] M. Lacy, "Artificial Laboratories," *AI Magazine*, vol. 10, no. 2, 1989, pp. 43-48.

[12] H. Abelson, M. Eisenberg, M. Halfant, J. Katznelson, E. Sacks, G. Sussman, J. Wisdom, K. Yip, "Intelligence in Scientific Computing," *Communications of the ACM*, vol. 32, no. 5, May 1989, pp. 546-561.

[13] C. Apte and R. Dionne, "Building numerical sensitivity analysis using a knowledge based approach," *IEEE Expert*, vol. 3, no. 2, Summer 1988, pp. 371 - 378.

[14] V. Dhar and A. Croker, "Knowledge-Based Decision Support in Business: Issues and a Solution," *IEEE Expert*, vol. 3, no. 1, Spring 1988, pp. 53-62.

[15] H. A. Simon, *The Science of the Artificial*, MIT Press, 1981.

[16] R. Bellman, *Adaptive Control Processes*, Princeton University Press, 1961.

[17] L. Phillips and W. Edwards, "Conservatism in a Simple Probability Inference Task", Journal of Experimental Psychology, vol. 72 (1966), part 3, pp 346 - 354. Reprinted in W. Edwards and A. Tversky, *Decision Making*, Penguin Books, 1967.

[18] L. Festinger, *A Theory of Cognitive Dissonance*, Stanford University Press, 1957.

[19] R. D. Luce, "Psychological Studies of Risky Decision Making" in G. B. Strother (ed.), *Social Science Approaches to Business Behavior*, R. D. Irwin Inc., 1962, pp 141 - 161. Reprinted in W. Edwards and A. Tversky, *Decision Making*, Penguin Books, 1967.

[20] J. Shanteau, "Psychological Characteristics of Expert Decision Makers" in *Expert Judgment and Expert Systems,* Mumpower, Phillips, Renn, Uppuluri (eds.),Springer-Verlag, 1987, pp. 289 - 304.

[21] W. Edwards, "The Theory of Decision-Making", *Psychol. Bull.*, vol. 51 (1954) no. 4, pp. 380-417. Reprinted in W. Edwards and A. Tversky, *Decision Making*, Penguin Books, 1967

[22] R. F. Hink and D. L. Woods, "How Humans Process Uncertain Knowledge: An Introduction for Knowledge Engineers", *AI Magazine*, Fall, 1987, pp 41 - 53.
[23] L. Adelman, "Measurement Issues in Knowledge Engineering," *IEEE Trans. Syst. Man Cybern.*, vol. SMC-19, no. 3, pp. 483-488, 1989.

[24] W. Rouse, J. Hammer, C. Lewis, "On Capturing Human Skills and Knowledge: Algorithmic Approaches to Model Identification," *IEEE Trans. Syst. Man Cybern.*, vol. SMC-19, no. 3, pp. 558-573, 1989.

[25] S. Stevens, *Psychophysics: Introduction to its Perceptional, Neural and Social Prospects*, John Wiley & Sons, 1975.

[26] K. J. Schmucker, *Fuzzy Sets, Natural Language Computations, and Risk Analysis*, Computer Science Press,1984.

[27] R. K. Bhatnagar and L. N. Kanal, "Handling Uncertain Information: A Review of Numeric and Non-numeric Methods" in *Uncertainty in Artificial Intelligence*, North-Holland Publishing Co., 1986.

[28] H. Guggenheimer and R. S. Freedman, "Foundations of Probabilistic Logic", *Proceedings of the Tenth International Conference on Artificial Intelligence*, Milano, Italy, August, 1987, pp.939-941.

[29] H. Guggenheimer, "Probabilistic Propositional Logic", *Polytechnic University Notes on Artificial Intelligence* , vol. 4,1987.

[30] G. A. Miller, "The Magical Number Seven, Plus or Minus Two: Some Limits on Our Capacity for Processing Information", *Psychological Review* , vol. 63, pp. 81 - 97, 1956.

[31] M. J. Frank, "On the simultaneous associativity of F(x,y) and x + y - F(x,y)." *Aequat. Math.*, vol. 19, 1979, pp. 194 - 226.

[32] J. Armstrong, *Long-Range Forecasting*, John Wiley & Sons, 1978.

[33] W. Butler, R. Kavesh, R. Platt (eds.), *Methods and Techniques of Business Forecasting*, Prentice-Hall, Englewood Cliffs, 1974.