THE USE OF REFERENCE MODES IN SYSTEM DYNAMICS MODELING
(SUMMARY)

Pieter VanderWerf
Sloan School of Management
Massachusetts Institute of Technology
Cambridge, Massachusetts

Reference modes neatly summarize the real-world problem behavior that motivated a model. If the modeler chooses to include in his reference mode some extra key variables, it also concisely depicts information potentially useful in formulating model structure.

Like all empirical information, reference modes help design theories and refine them. During initial data gathering the construction of a reference mode lends organization to the modeler's efforts. When the modeler is working on model formulation, the reference mode is a standard against which to check mentally the many structural alternatives he sifts through. When a modeler is analyzing his completed model for validity the reference mode provides the first standard against which to judge the simulation output.

There are two types of reference modes, each with its own advantages. A numerical reference mode is an exact plotting of discrete data points of variables from the real system. It is the ultimate judge of accuracy. It shows exactly how the system has behaved; the modeler has the burden of either justifying any differences between model reference mode or changing the model. An impressionistic reference mode is a smoothed version of the numerical that abstracts out the assumed key features of behavior and highlights them. It omits what the modeler believes to be noise. When he is

struggling with initial formulation this streamlined rendering of real system behavior gives him guidance without confusion. It is also a handy tool for explaining the system's problem behavior and the model output to others.

I have developed a few personal rules for constructing a reference mode at the beginning of a modeling project, designed to increase the degree to which the reference mode achieves its potential contribution to a the project. Following them all thoroughly might take more time than it is worth, but they are all points worthy of consideration at the outset of a project. My list recommends that the modeler:

(1) Choose reference mode variables that depict the problem or that intuitively seem likely to illuminate system structure.
(2) Always construct both numerical and impressionistic curves for all variables.
(3) Impose a minimum of restrictions on the initial impressionistic curves.
(4) Collect and retain as much detail as possible.
(5) Update all reference mode curves as new information becomes available.

An example demonstrates the uses of reference modes. In the data gathering for a model of new product diffussion deriving the numerical reference mode forces precise variable definition and reveals some relevant facts about the system. Impressionistic reference modes help evaluate alternative model structures in initial formulation. Once a running version of the model is available, its behavior can be compared to the reference mode for validity testing.

THE USE OF REFERENCE MODES IN SYSTEM DYNAMICS MODELING

Pieter VanderWerf
Sloan School of Management
Massachusetts Institute of Technology
Cambridge, Massachusetts

Introduction

In the first lecture of the first system dynamics course I ever took the professor presented a list of the steps of a modeling project. During the rest of the semester it became apparent to all of us that actual projects never follow the list very closely. But it also became apparent that the list was useful anyway. It helped organize effort, gave direction to a stalled modeler, and provided a checklist of activities to be addressed (if not always accomplished).

Near the beginning of the list -- under Problem Definition -- was the step entitled Construciton of a Reference Mode. Summarizing everything the modeler was trying to do in a simple picture had an appeal that has stayed with me to the present and motivated me to examine its uses in detail. I am not alone in this interest; at least two system dynamicists have written previously of the value of the reference mode in modeling projects.[1,2]

Why do we construct reference modes? To convince ourselves that there really is some phenomenon worth modeling, to be certain of just what that phenomenon is, to get clues to appropriate model structure, and to check the accuracy of the model once it runs.

Why do we fail to construct reference modes? Because we do not recall to do so, because we are unwilling to devote the effort, because we cannot find the necessary information, because we intend the model only as a philosophical construct, not a representation of anything in the real world.

Why should we construct reference modes? And how should we use them? The reference mode is a potentially useful link between a modeling effort and what it proposes to reproduce. Having a strong bias toward the empirical, I consider the influence of the reference mode on a modeler valuable. Indeed, much of what this paper says about the reference mode could be restated in analagous form for all types of empirical information about real systems. When I extol the virtues of the reference mode I am also arguing for the conscientious use of available real-world information in general.

In the following pages I examine in depth the potential role of a reference mode in a modeling project. This entails defining the thing, analyzing its purposes, suggesting how to construct it and employ it to achieve these purposes, and presenting a concrete example. But first I think it would be helpful to outline the philosophy that underlies my opinions. As I have already implied, my view of the reference mode arises from a strong empiricist bias. So as a preface to the paper I state explicitly what this bias implies for dynamic modeling.

A Philosophy of Empiricism for Modelers

For an applied modeler, measures of real system behavior are a guide and a critic. In their role as guide they give clues to what to include in the model. It is easier to sift through a mass of information about a system and pick out the relevant details if we know what the exercise is supposed to reveal to us. In their role as critic they prevent premature confidence in a model. Where simulated and measured real system behavior differ, the burden is on the modeler either to demonstrate that the existing model is adequate despite the discrepancies or to change the model. The differences could be negligible for the purposes of the model. Or the measures of real behavior used could be inaccurate. But if the modeler cannot make a convincing case for either of these possibilities he must admit that the model is flawed. It may be usable and useful, but it is nonetheless demonstrably improvable.

The role of empirical information shifts over the course of a modeling project from predominantly guiding to predominantly critical. The modeler's initial problem is learning about the real system and abstracting a coherent operating representation out of the noise. Simply gathering and organizing available information in the first place acquaints him with the system and prods him to define what aspects of it are of interest to him. Patterns that he can distill from the facts interact with his intuitions of system structure to guide initial formulation. Up through this stage the modeler must ignore the many fine discrepancies between the real system and the model or he will become bogged down. But eventually he will have a fully functioning model and enough intuitive understanding of it to confront details without getting confused. At that point facts about the real system have their highest value as critics. They reveal inaccuracies in model behavior.

Investigating differences between the facts and the model is a major contributor to model improvement. Such differences are a symptom of model inaccuracy, the causes of which we can discover by tracing backwards. Discovering such causes refines the modeler's notion of the circumstances under which the model is valid. The modeler may decide to reformulate his model to eliminate some of the differences and enlarge the realm of its validity. If so, knowing the sources of the discrepancies gives direction to his effort.

The more actively and systematically a modeler seeks out real system behavior contrtadictory of the model output, the better able he is to achieve a robust model and understand its applicability precisely. If a modeler does not look for such model inadequacies, clever colleagues or disastrous attempts to apply the model are the more likely to uncover them for him. Even if a model is intended only to increase the modeler's personal understanding, soliciting and evaluating counterevidence aids the venture.

What is a Reference Mode, Anyway?

The reference mode is a time plot of the key variables from the system being modeled. The variables are "key" in the sense that they illustrate the problem that motivated the model and perhaps give clues to its origins. Usually the plots are of observed historical paths of the variables, either as a connection of discrete points or as a smooth approximation of the general shape of their behavior. In other cases the reference mode or part of the reference mode will consist of the expected or desired future courses of the variables.

The reference mode is one of the real-system descriptors available to the modeler at the outset of his project. Others that might serve the same functions are a verbal description and a set of statistics. The plotted reference mode has certain advantages over these, however. Because it is visual the reference mode quickly gives a precise impression of variable behavior modes. Since it is these behavior modes, rather than specific values, that system dynamicists typically try to reproduce and predict, the reference mode provides a handy guidepost in judging the success of the a system dynamicist's efforts. Verbal description of the same information is often lengthy and confusing; a list of numbers is nearly meaningless until the reader puzzles out their broad pattern of change. This advantage of the reference mode is particularly strong in communication with persons unfamiliar with the model or system: few outsiders will endure and understand a thousand words, but the picture makes a quick impression.

In addition, simple visual impressions have a way of sticking in memory better than verbal or numerical versions of the same ideas. The modeler can conveniently bear a visual behavior mode in mind as he sifts through information on the system; the model user or critic can hold it in memory as he analyzes model structure and output.

Types of Models, Types of Reference Modes

All models are divided into two types: the specific and the generic. A specific model represents a unique system: the economy of the United States, the Sprague Electric Company, the U.S. pork commodity market. A generic model tries to embody the common features of a class of similar systems in an attempt to provide an understanding of them as a group. It may also serve as a starting point for formulation of specific models of individual members of the class. Examples would be an abstract model of a national economy, an industrial production-distribution system, or a commodity market.

All reference modes are also divided into two types: the numerical and the impressionistic. The numerical reference mode is a plotting of historical values, usually connected by straight lines. The impressionistic consists of simplified curves, typically drawn by hand, meant to capture the key features of the behavior pattern of the variables under scrutiny. Drawing an impressionistic curve requires abstraction of interesting features from a set of details. This has the advantage of highlighting those features, but it also loses information about deviations from the hypothesized "pure" behavior mode of the system.

Reference modes of both types can be constructed for both types of models. Specific models have unique numerical reference modes. Definitions of variables and estimates of their actual values may vary, but in theory each variable in a model of the U.S. economy has exactly one historical time path. Yet even for a specific model it is possible to construct an impressionistic reference mode. One need only decide what is either "noise" or explainable but uninteresting variable variation and eliminate these in a new smoothed plot. In the case of a generic model it is generally necessary to construct an impressionistic reference mode. It is unclear what the modeler considers the important behavioral features common to the set of separate systems without one; many of these features will be absent or obscured in some of the individual numerical reference modes. Yet each of these separate systems also has its individual numerical reference mode, and these serve in a sense as numerical reference modes for the parent generic model as well.

When the specific systems intended to be covered by a generic model exhibit widely divergent behavior, more than one impressionistic curve in the reference mode can be useful. These can define an envelope for the set of separate numerical modes. Or, if the separate modes sort readily into a few characteristic behavior patterns, there can be one impressionstic curve for each such subgroup.

Regardless of the type of model, when a reference mode extends into future behavior it is necessarily impressionistic. The future portion may be assigned multiple paths just as can the impressionistic mode of a generic model: one can postulate a range of possible future curves or a few types of curves.

The Purpose of Reference Modes

Like all empirical information the reference mode can serve as a guide and a critic. If a modeler takes the time to construct a reference mode carefully at the outset of a project he will have to organize information in ways that will guide his initial intuition on the problem. Numerical reference modes force him to think about variable definitions and acquaint him with the orders of magnitude he will be dealing with. Deriving an impressionistic mode from this detail requires him to decide precisely what aspects of system behavior define the problem and should be reproduced.

During model formulation, when the modeler is sketching out rough structure, an impressionistic mode is a handy guide. Experienced modelers know what modes of behavior simple structures (at least) are capable of producing. As they weigh alternative formulations in their minds they can reject some and give support to others according to the likelihood that each will reproduce the reference mode.

When the model is developed, the numerical mode is a sharp critic. Where model output deviates from actual experience there is a point for investigation. The modeler can look into these discrepancies to uncover modifications that would make the model more robust. Or he may be satisfied with an increased understanding of the model's limitations. At some point the modeler will be left with deviations so small and numerous that he will want to relegate them to the "noise" category. Yet it may be useful to go a little further at this juncture by estimating the parameters of the noise and thinking about the types of things that are probably behind it.

I once read of a generic model of intra-firm technological diffusion. The model was meant to apply to about a dozen firms that all turned over their capital stock to adopt a particular new production technology. It modeled the process such that predicted diffusion followed a sigmoid growth curve. That is, the percentage of capital converted to the new technology began at zero and initially grew slowly. Growth accelerated in the middle region. Then, as adoption approached 100 percent, it slowed again. The published numerical reference modes for the dozen firms showed that this behavior was accurate in many cases, but by no means all of them. The numerical plot for one firm was in fact exactly the opposite of what the model predicted: adoption rose rapidly initially, then slowed when it was in the 50 percent region, and then suddenly rose rapidly again up to a full 100 percent. The modeler did not comment on this discrepancy anywhere. He merely stated that the model gave a reasonable fit to the numerical reference modes as a group. Had he investigated at least that one anomalous case in more detail he might have learned some things useful to his understanding of real system structure. As it was, he implicitly considered it a noise case.

The reference mode can be useful in a couple of validity tests other than immediate comparison of model output to history. Following the typology of Forrester [3], these are the alternate-structure test and the surprise-behavior test. In an alternate-structure test the reference mode is the standard against which to judge what alternative structures recreate past behavior as well as the original formulation. In a surprise-behavior test, the reference mode is a first check: surprising behavior that appears in the model might be somewhere in the reference mode as well, but have gone unnoticed. In general, the reference mode is useful in any test that requires checking on past behavior.

There is of course much more to validity testing than the mere reproduction of history. But that is the minimum we can ask of any model. A model does not deserve more powerful testing until it can recreate the behavioral phenomena that motivated its construction originally.

Guidelines for Constructing Reference Modes

I have developed a few personal guidelines for constructing reference modes to improve their ability to serve their purposes. Were a modeler to follow every one thoroughly he would have little time left for diagramming and doing model runs: it is possible to refine data indefinitely. So I present this list of suggestions in the spirit that the list of steps of a modeling project was presented to me. It offers the modeler some orgainzing principles for his drawing of reference modes and gives a high standard against which to compare his own work.

Guideline 1: Choose reference mode variables according to the purpose of the model and intuition about basic system structure. There nearly always exists a single variable whose plot neatly summarizes the problem the model is to address. Just what that variable is depends on the precise problem; it might change as the modeler redefines his problem. Plots of extra variables are useful for giving insight into causal structure. To this end a modeler might want to plot the constituent parts of the one variable used to depict the problem, plus any other variables he considers linchpins within the system.

Guideline 2: For any model, construct both numerical and impressionistic curves for all variables. The numerical reference modes are the empirical critics. Constructing them prevents the modeler from rashly adopting the impressionistic reference modes of current folklore. Analyzing them helps identify shortcomings in the final model. The impressionistic reference modes have the advantage of clarity, useful in early formulation. It is nearly always possible and advantageous to construct both.

Guideline 3: Initially, impose a minimum of restrictions on the impressionistic curves. Just as there are advantages to drawing the system boundary as small as possible while still generating the behavior of interest endogenously, there are advantages to leaving the impressionistic reference mode as wide a range as will still convey the behavioral characteristics of interest. A broad band within which the actual variable path can move is sufficiently restrictive in many cases. It is not necessary to include all of the fine oscillations of the numerical reference modes at first, and it may never be. If the problem is, say, a system that overshoots and declines, it may be dangerous to include the irregular twistings of the variable's path in the original impressionistic reference mode. Apparent patterns can be the result of no systematic causal interactions, but only noise. If the modeler includes everything in the impressionistic reference mode he may waste time trying to explain such a spurious pattern instead of identifying the forces behind the broad problem that motivated the project. At the same time, especially in the case of a generic model it is presumptuous to draw a single fine line at the outset. Until the modeler knows more about the causes of variation among the different specific systems he cannot be sure which cases are "normal."

Guideline 4: Collect and retain as much detail as possible. One can always throw out unnecessary data when desirable, as when summarizing project results for a presentation or for one's own reference. But initially it is hard to know what variables and twists in behavior will be of interest. Furthermore, the more detail on real system behavior the modeler has at hand, the more holes he can pick in the model in verification analysis later. Once I have chosen the variables I consider important to include in a reference mode I sometimes spend 2-3 days in the library and on the telephone trying to fill in time plots with data points as densely as possible. For each point I record the source and any reservations or stipulations attached to the estimate. This procedure also helps me get acquainted with the real system and occassionally reveals other variables I should include in the plotting.

Guideline 5: Update reference mode curves as new information becomes available. The updating process itself will yield new insights into variable definition and the nature of the real system. The resulting, more refined reference mode will guide formulation and locate model deficiencies more accurately. In short, regular updating integrates the reference mode into the iterative modeling process: developments in other project work reveal improvements to make in the reference mode, which may later generate new insights for other modeling activities.

Example: A Model of Technological Diffusion

As a concrete example of the construction and use of a reference mode consider a model of the diffusion of a new comsumer product. The model is to address the displacement of an established product with a new technological alternative. The purpose of the model is the purpose of the industry selling the new product: the member firms wish to promote as rapid and complete a displacement of the old product as possible so that their sales rise quickly and high.

The obvious variables to use to summarize the problem are sales of the new and old products. These can be collapsed into one: the fraction of total sales accounted for by the new product. Note, however, that this loses any information about market expansion or contraction. The modeler should bear that in mind in case he later considers changes in market size important.

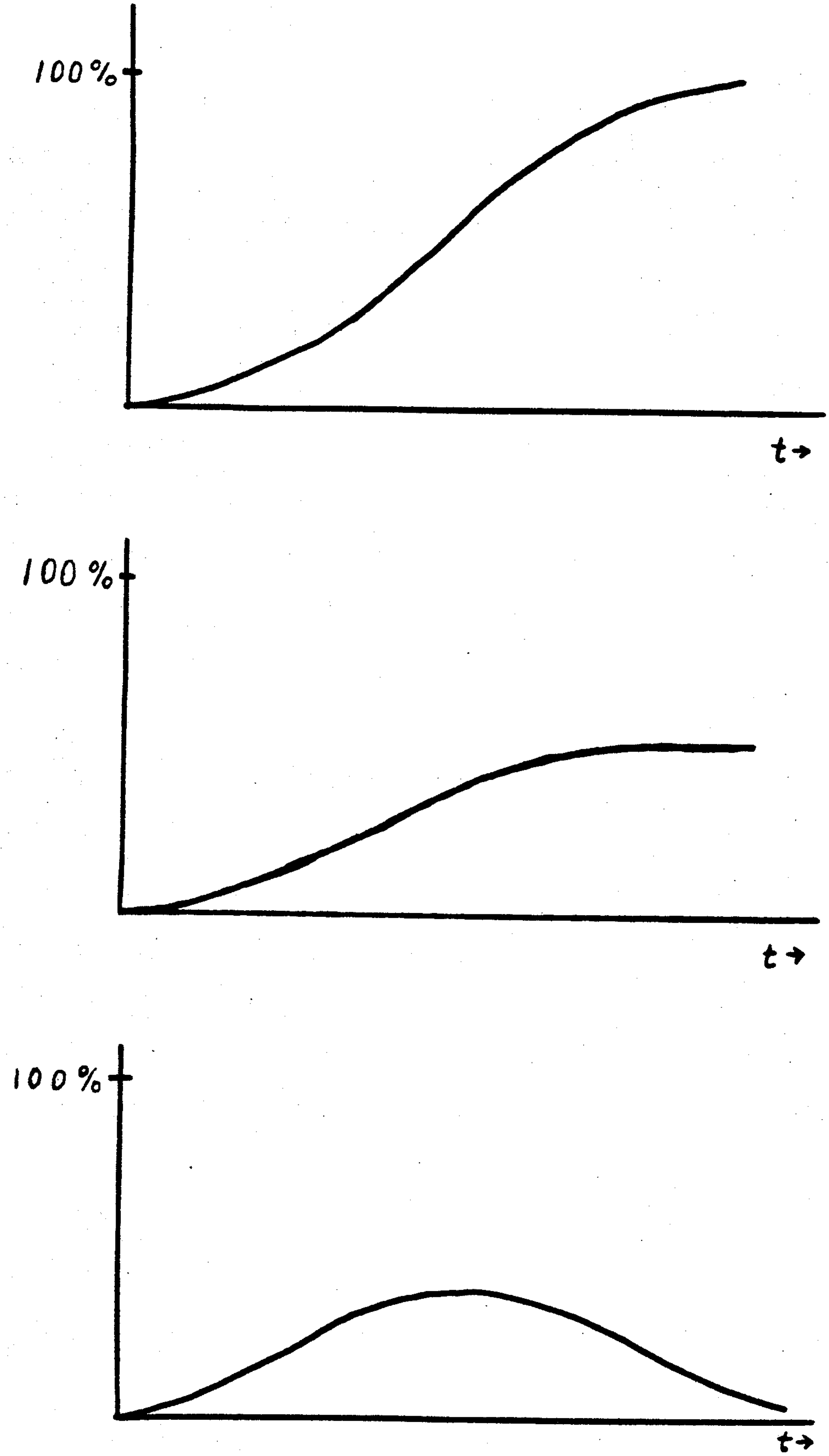
If the model had had different purposes, different variables would have been appropriate to summarize the problem. Imagine that a government agency had commissioned the model because it was interested in the displacement of socially harmful products by more benign ones. Some product in general use might pollute the air when it is used, whereas a new alternative does not. In this case the quantity of interest would be the fraction of the products in use that are of the new variety, not the fraction sold. If the model were the project of an individual firm selling the new product, probably that firm's sales would be the variable used to define the problem, not the industry's sales.

Investigation of past cases of this sort of phenomenon reveals a variety of patterns of behavior. Indeed, this variation is a major motivation for the model. The automobile has nearly completely replaced the horse for personal transportation; radial tires and quartz timepieces threaten to do the same to bias belted tires and mechanical clocks and watches within the next decade. Yet fluorescent light sales rose and plateaued at a market share well below 50 percent. And the rotary-engine automobile, after enjoying a flurry of popularity in the late 60s and early 70s, has all but disappeared from the market. Active solar home heating similarly captured a share of the market in the Southeastern United States during the 1930s and then vanished, only to re-emerge recently under changed market conditions.

The impressionistic reference modes depicting these various behaviors are shown on page 13. These plots are by no means final; they are renderings of reference modes drawn or described by others, used here as a starting point. They as yet have no numerical support.

On the basis of these initial reference modes we can restate model purpose more precisely. The model must illuminate the dynamics of the displacement process sufficiently that the modeler learns what differed to allow complete displacement in some cases and partial or only temporary partial displacement in others. It must also help us find the influences on the speed of displacement. Once these things are done we need to use the model to speculate on policies that would encourage faster and more complete displacement.

This is the point at which to gather data to construct numerical



Impressionistic reference modes for the alternative behaviors of the market share of a new product.

reference modes. If the model is to be generic, we will want data on cases of each variety. If it is to be specific (as it might be if we were interested in a single new product whose life cycle is not yet played out) we will want more extensive data on the one instance of interest, and a smaller amount on some others for ideas on formulation.

Intuition and research suggest two broad factors determining the diffusion of a technologically new product. The first is the inertia of marketing and adoption channels through which the new product flows. It takes people time to learn of an improved product, time to see its advantages, and more time to reach a convenient point for discarding their old product and buying the new. It likewise takes wholesalers and retailers time to decide that the new product is worth the risk of stocking. The second factor is the relative advantage of the new product over the old -- something that can change over time. If the new product is clearly superior it should gradually seep through the system. But more likely, the new product will initially have some advantages and some shortcomings relative to the old. It will gradually undergo improvement, so that it becomes the superior alternative for an ever larger share of the market. Conversely, if the old product undergoes improvement in the meantime it could eliminate or reverse the new product's competitive edge. This might lead to a stagnation in new product growth or its complete disappearance from the market.

On the basis of our preliminary causal analysis, measures of relative technological advantage should be included in the reference mode. "Market inertia " is really just another way of saying that this is a complex system containing delays; it cannot be "plotted" in the usual sense. But the influence of technological improvement is neatly summarized by the time plot of a few variables, so inclusion in the reference mode is appropriate.

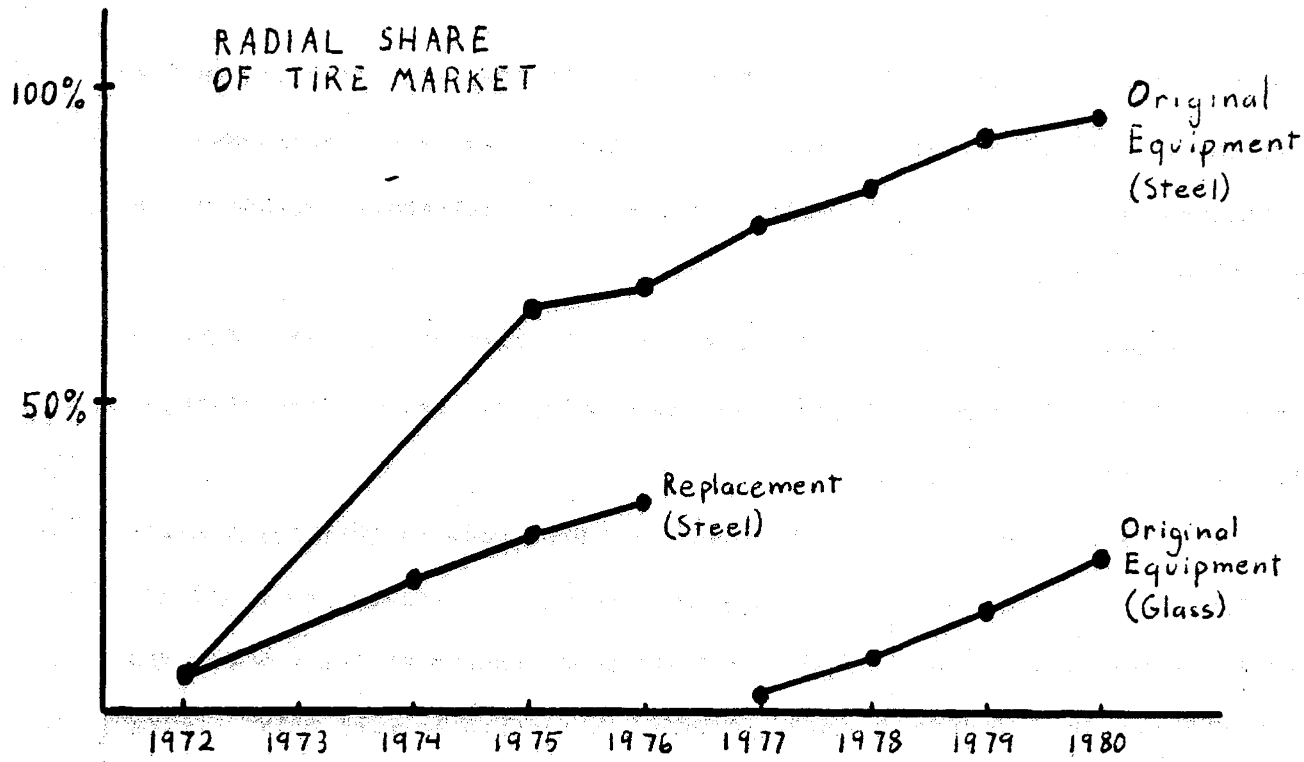
On page 16 is a numerical reference mode for the diffusion of radial tires in the U.S. It includes market share as the variable relating the system to the problem definition. It includes various measures of performance relative to bias belted tires for additional guidance in formulation.

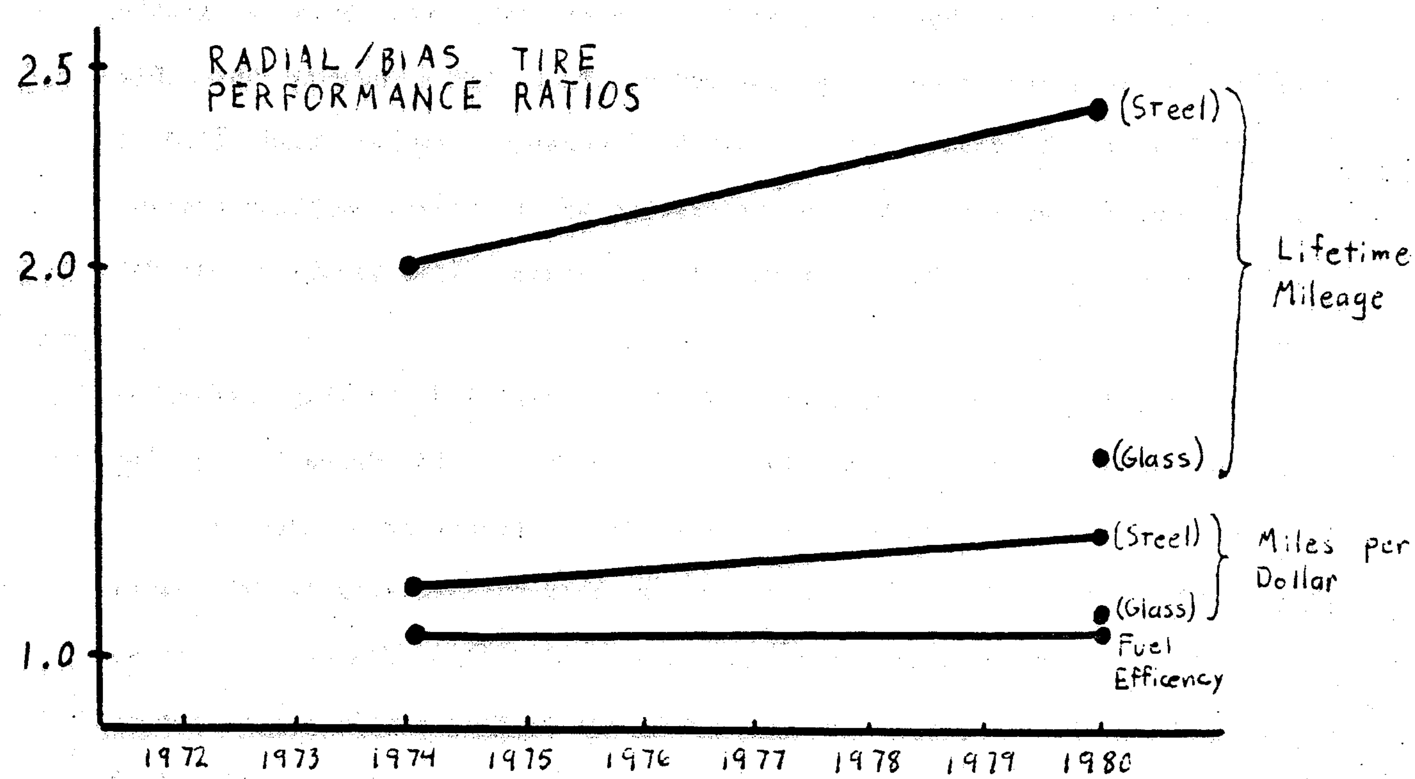
Note here that paucity of data was no barrier to including a variable. Plotting whatever data are available is better than omitting something potentially important.

The first thing we notice from the reference mode is that sales seems to be following the 100 percent displacement pattern. (Sales are divided into two submarkets because of the very different nature of the new car and replacement markets. The importance of this distinction emerged only during research to derive a precise numerical reference mode.)

Next we notice that steel-belted radials improved only a little relative to standard tires over the period of available information. This suggests that in the case of radial tires diffusion results more from a superior product working its way through the marketing and adoption system, not from improvement of the new product that increases its receptive customer base.

There is also an unexpected development depicted in the reference mode that could require incorporation into the model. Searching through the industrial statistics revealed that in the 1970s producers introduced the fiberglass belted radial, which has since captured a growing market share. Further research showed that buyers have been resistant to high-priced steel radials. The fiberglass radial does not have as long a life per dollar as the steel, but its initial cost is lower. This new information forces a decision. We could disaggregate, modeling the fiberglass radial separately. But more likely we would consider it just a variation on the generic radial and resurrect the idea of the influence of technological advance. Some index



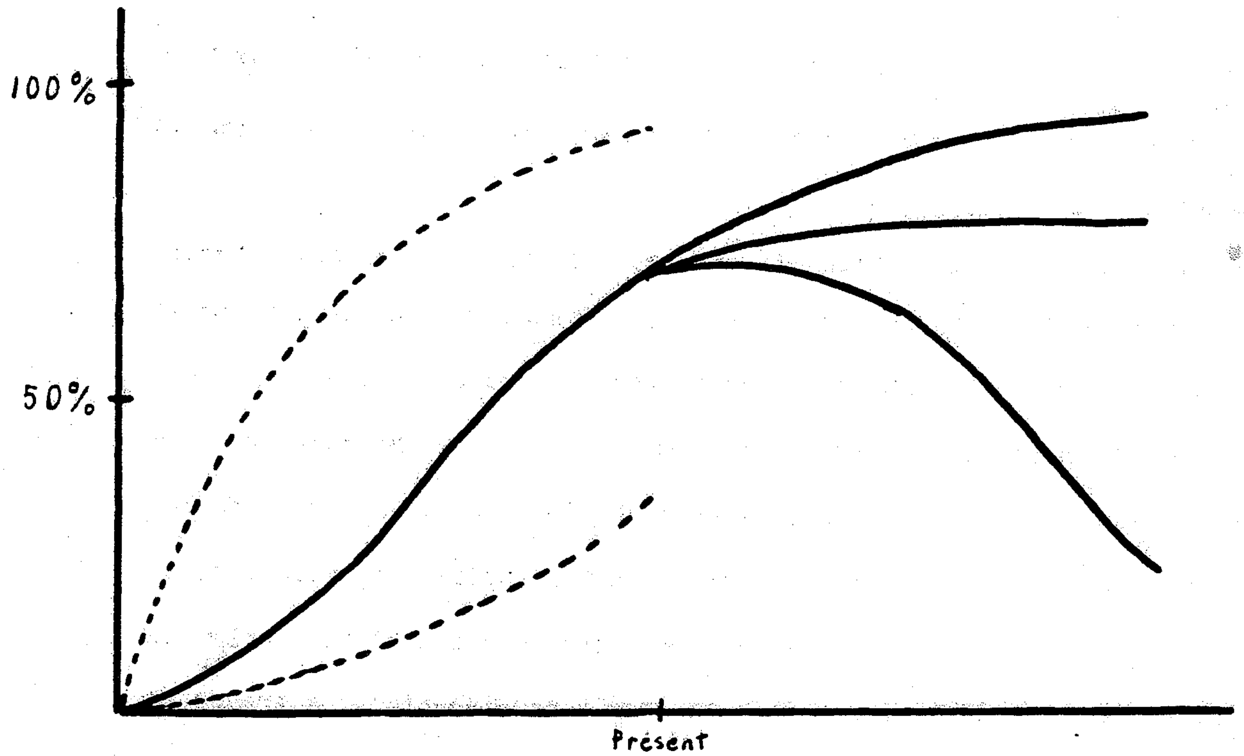


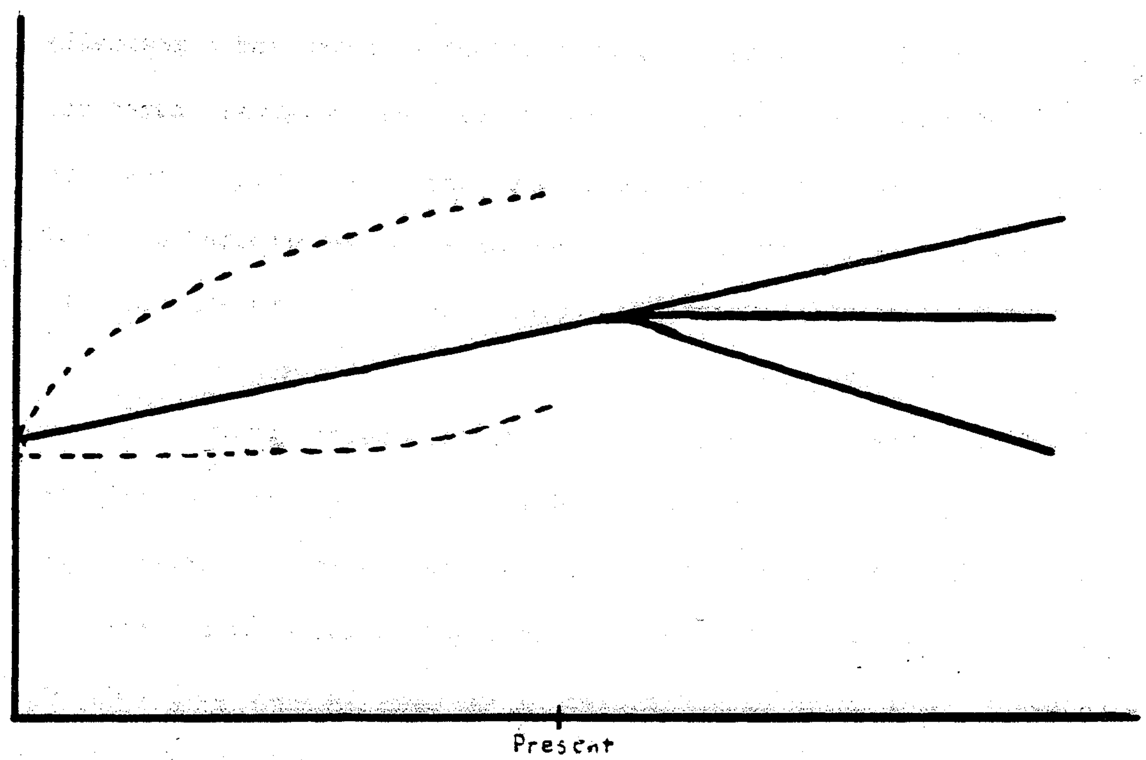
Numerical reference mode for model of the diffusion of radial tires.

of relative utility or attractiveness must be put into the model to show how the development of new types of radials sped their penetration into certain market sectors.

Now consider the derivation of an impressionistic reference mode, shown on page 18, from this jumble of details. Assume that we want to treat the market as one unit. Market share growth appears approximately s-shaped, but we leave a broad band over which we will allow the model to roam initially. Our historical examples show that its future course could follow a variety of patterns, so we draw the three most plausible as extensions of the historical path. We have not even defined the measure(s) of relative attractiveness yet, so it is impossible to draw a precise plot. A gradually rising path depicts our current knowledge of the role of changing technology.

Suppose that we have now constructed a specific model of radial tire diffusion that produces an s-shaped growth in market share and a gradually increasing index of relative utility. We return to the original numerical reference modes to look for discrepancies with the model output. The most obvious is the difference in aggregation. Depending on the expected uses of the model, the modeler might want to reformulate it, breaking out the different tire markets and measures of utility. Another point of disagreement is the temporary slowdown in radial sales growth in the mid-1970s. It is not in the model output. Why is it in the real system? As it turns out, industrty people explain it with a tire workers' strike that cut output temporarily. Most of us would be inclined to shake this event off as noise. But analyzing these discrepancies is intended to make us think, so perhaps we should push the possibility that this one has something to tell us. Perhaps new technologies typically encounter resistance in various forms. Or perhaps expanding output at the rate of the radial tire plants often leads to labor problem. These are possibilities that might deserve





Impressionistic reference modes for model of the diffusion of radial tires. Dotted lines indicate the degree of deviation from past behavior it seems acceptable to allow the initial model. Alternative possible variable paths are plotted for the future.

checking up on. If there is strong evidence that such effects occur regularly and if such deviations from the s-shaped path are important to the model's purpose, reformulation to include this one is warranted.

Next we can check how the technological variables track history. Do the model variables move in the same pattern as in the reference mode?

Are there other model formulations that can also reproduce the growth curve of the reference mode? If the modeler searches actively for such alternatives he increases his chances of uncovering superior structures. If he does find plausible alternatives that reproduce the key behavior, further inspection using other tests (structural verification, symptom generation, etc.) is necessary.

Now assume that the base runs of the model yield unexpected behavior. And is the surprise behavior also hiding in the reference mode somewhere? If so, we have extra support for the model. If the surprise behavior contradicts the reference mode the burden is on the modeler to defend or change the model. If the reference mode does not include the variables or conditions necessary to uncovering that behavior in the real system we must dig for additional empirical information.

Conclusion

Systematic construction and use of a reference mode contributes to the quality of a modeling project. There are other ways to convey the information depicted in a reference mode, but the reference mode's conciseness and the ease with which it can be remembered give it advantages that justify the trouble it takes to prepare one and update it throughout the project.

The knowledge gained simply from construction of a reference mode is itself valuable. It is an activity around which the modeler can organize his initial research of the real-world system. It confronts him with issues of variable definition. It makes him begin to decide what information is relevant to the problem.

During modeling the reference mode provides a ready check against the real system under study. A modeler can use it to prod his imagination when formulating structure and to ferret out model weaknesses when doing validation. It can thus be an integral part of the back-and-forth reasoning process that goes into making and gaining confidence in any model.

What one gets out of a reference mode depends on what one puts into it. Some solid work time devoted to constructing one at the beginning of a project produces a clearer problem focus and a set of variable plots that are valuable later on. While strict adherence to the list of guidelines in this paper may require more time than is warranted, it is a good idea to use it at least as a checklist. A modeler should justify to himself the skipping of any steps.

As an experienced modeler has observed, a reference mode alone will not produce a good model.[4] But including one in a project will probably lead to a better model.

NOTES

[1] Randers, Jorgen. "A framework for Discussion of Model Conceptualization." In The System Dynamics Method. Randers, Jorgen and Leif K. Ervik, eds. Oslo, 1977. Page 448.

[2] Runge, Dale. "The Reference Mode as a Guide to Transparent Causal Structure." Memo D-2460, System Dynamics Group, Massachusetts Institute of Technology. Cambridge, Massachusetts, September 1976. Page 2.

[3] Forrester, Jay W. "Confidence in Models of Social Behavior -- With Emphasis on System Dynamics Models." Memo D-1967, System Dynamics Group, Massachusetts Institute of Technology. Cambridge, Massachusetts, 10 December 1973. Pages 48-53.

[4] Randers, op. cit. Page 448.