Multi-Step Prediction Theory

W. C. Cave and R. E Wassmer†, January 24, 2015

INTRODUCTION

Applications requiring multi-step prediction can be considered an extension of control theory. But other than formulating the equations, a theory of multi-step prediction is generally not treated, see [1], [10], and particularly [14]. This is because most work in engineering is based on mathematical models using difference or differential equations. Missile guidance systems fall into this category, where controls are typically governed by the smallest time constants of interest and depend upon single step prediction. In these engineering systems, prediction is defined in relation to the following functional categories:

- Smoothing - Functions operating at $T < 0$;
- Filtering - Functions operating at $T = 0$;
- Prediction - Functions operating at $T > 0$.

Current engineering design approaches using single-step prediction focus on filtering to estimate the current state of the system. They do not depend upon predicting multiple steps into the future. Most references to multi-step prediction are outside the field of engineering and not based upon scientific principles or experimental results. Many of these refer to the literature on the scientific theories of Filtering and Smoothing, much of which is based on the work of Kalman, [11], after whom the famous filter is named. Sound theoretical work on multi-step prediction is scarce. The resulting misunderstandings have been described by Athens and Kendrick, [2], and more precisely by Kalman, [12], regarding confusion between filtering (estimation) and prediction.

Potential applications of multi-step prediction are numerous, see for example [13], [5], [6], [7], and [8]. Econometric systems used to make financial decisions are an excellent example; however, many are concerned with short-term estimation, see [3]. Others are approached using time series analysis, [4], using sophisticated forms of curve fitting, where the underlying assumption relies on a complex form of stationarity which may hold for parts of a system as shown below. Weather forecasting is generally treated separately, being interpreted based upon individual measures. To fairly asses the application of multi-step prediction to these categories requires that we differentiate between forecasting and prediction.

Multi-step prediction requires specific measures of accuracy of the system producing the predictions, and a characterization of confidence in the measures themselves. Such measures can be difficult to achieve, particularly if there is not sufficient historic data to characterize both the error and confidence in the measures. This is best understood using specific examples provided below. This paper addresses the theory and corresponding measures required when predicting outcomes multiple time steps into the future.

† The authors are with Prediction Systems, Inc., Spring Lake, NJ.

MULTI-STEP PREDICTION

As used here, *Multi-Step Prediction* implies being able to measure the accuracy of predicted outcomes at multiple time steps in the future. This requires the following:

- Predicted outcomes must be defined in terms of a specified class of possible outcomes.
- A measure of the probability that the predicted outcome will occur - or the probability of error - must be produced based on a recent history of prior predictions.
- The probability of error in the prediction statement must be based on data that has not been seen by those producing the prediction system.
- The probability of error must be accompanied by a confidence level.

Prediction Systems

Figure 1 contains an illustration of a Prediction System designed to predict the future responses of a system whose outputs are observed, see [9]. Design of a prediction system must consider those factors affecting the future outcomes of interest from the system being observed. Some of these factors are observable and can be used as inputs to drive the prediction system. Others that are unobservable may be treated statistically using an estimation subsystem, e.g., a Kalman Filter.

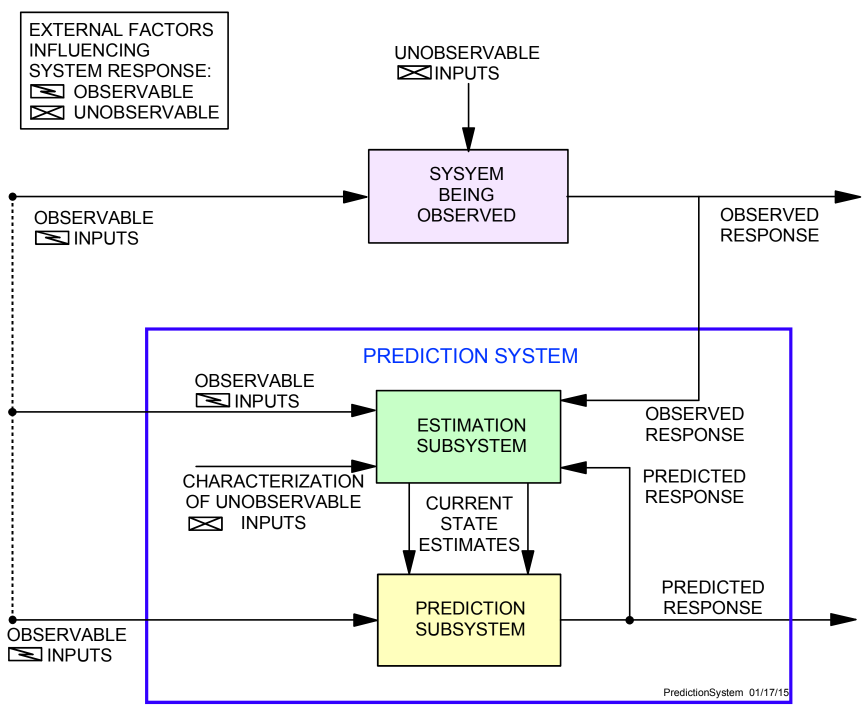

Figure 1. Illustration of a Prediction System.

The purpose of the estimation subsystem is to use all observable data at T, as well as statistical estimates of the unobservable factors to produce optimal estimates of the current state of the system. Given the best estimate of the current state of the system, observable inputs (driving forces) are used by the prediction subsystem (model) to produce predictions of the systems response.

The sequence of residual errors, between predictions and actual system responses are used to measure the accuracy of prediction over a Looking Back Horizon, see [9]. The looking back horizon must contain a sufficient number of time steps to characterize the confidence level in the error probability statement. If sufficient data and time exist, then a *prediction* can be made whose accuracy is characterized as described above. If this is not done, then the outputs produced are considered a *forecast*.

CHARACTERIZING PREDICTION ACCURACY

We start by analyzing the distribution of trajectory variations representing potential multi-step prediction outcomes as they unfold in time. Illustrated in Figure 2, such a set of distributions must be derived from the history of actual outcomes relative to their predicted values. The more narrow the distribution, the more accurate the prior predictions. We note that as time progresses further into the future, the distributions generally widen.

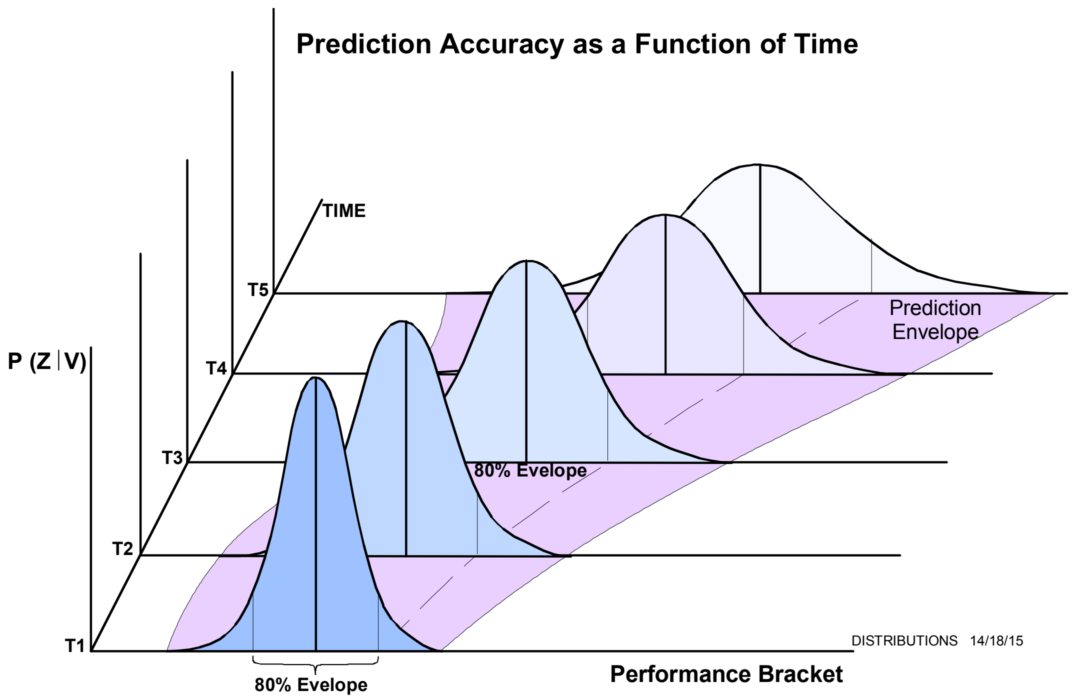

Figure 2. Prediction of future outcomes characterized by a distribution envelope.

The area under the distribution curve is used to determine the probability that the predicted outcome falls within a specified range. As an example, the 80% envelope implies that the actual outcomes will fall within the specified range with an 80% probability.

Specifying A Prediction Envelope

An example of such an envelope is shown in Figure 3. It is composed of a sequence of weekly intervals for each of the prediction horizons $\tau_p = 1, 2, ..., 12$. The prediction statement claims that 80% of the time, future outcomes will fall within the specified envelope. This is illustrated in the confidence level characterization in Figure 4.

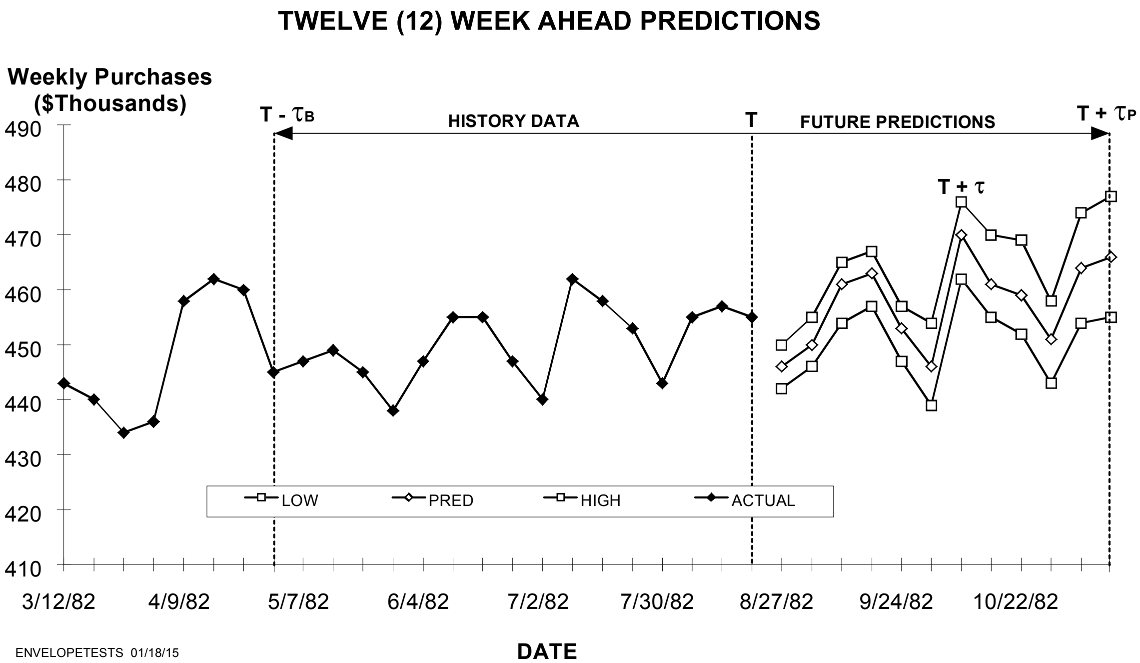

Figure 3. Prediction of future outcomes characterized by a distribution envelope.

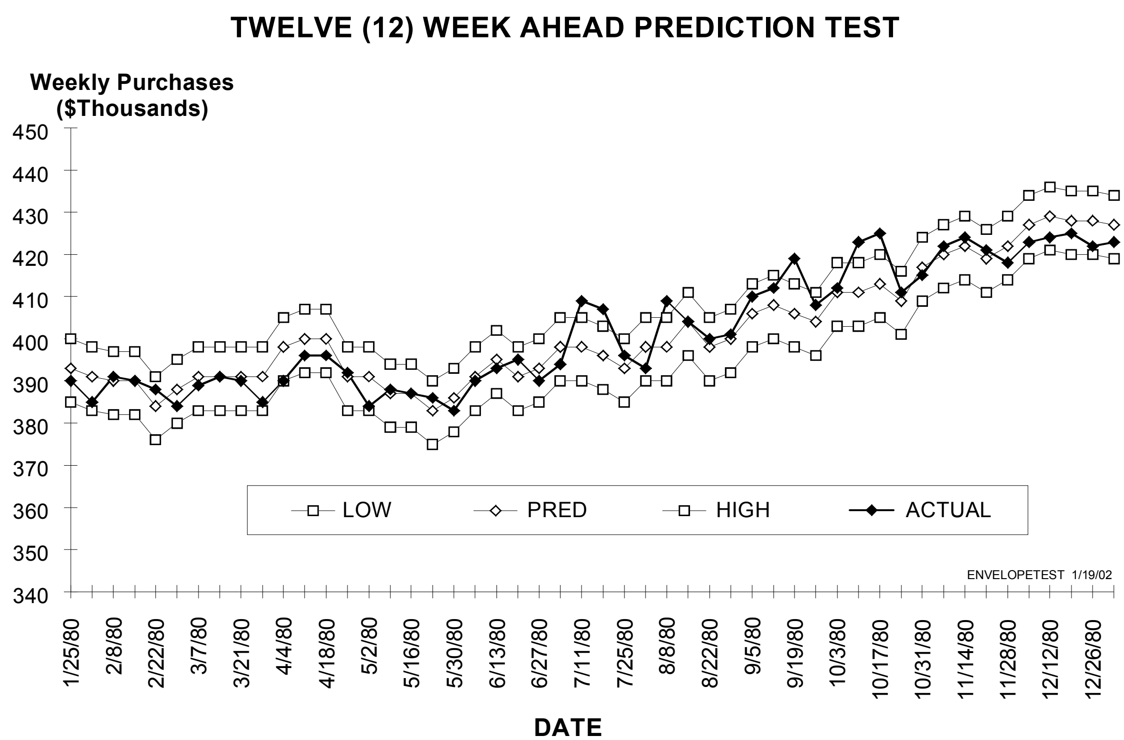

Figure 4. Characterizing the confidence in the prediction statement.

To characterize the statistics for the above problem, the following definitions are used.

τ_p - is the number of future time steps from the current time step to the future time horizon for which the system response is being predicted.

τ_b - is the number of past time steps from, and including, the current time step to the looking back horizon, used to define the probability statements.

n - is the number of mutually exclusive "τ_b" sample sets (ensembles) of history data available for testing the probability statement.

In other words if N is the total number of sample points (weeks) of history data, then

$$n = \frac{N}{\tau_b}$$

As models are improved to ensure the truth of an 80% probability statement (as shown in Figure 4), there may be certain sample sets of τ_b weeks for which it is difficult to support an 80% level. Decisions on looking back horizons must be made by the producer of predictions to ensure acceptance in the market. Certainly multiple measures can be offered for the same predictions.

Measuring Confidence in the Prediction Envelope

From the above we derive the following conclusions. When making statements about the probability that future outcomes will lie within a given envelope, we must pick a specific looking back horizon, τ_b, to test the probability statement. Next, we must consider all possible sample sets from the history data which contain τ_b *contiguous* samples. (There will be $N - \tau_b + 1$.) We can then plot the distribution of the number of times the actual values fell inside the envelope for a given horizon. See Figure 4.

Assuming this distribution is representative of the future, we can compute the probability that the actuals will fall inside the envelope at least 80% of the time. This provides a confidence statement about the 80% probability envelope. For example, we might conclude from Figure 4. that:

$$P\{X \geq 80\%\} = 0.95 .$$

We note that as $\tau_b \rightarrow N$, $\sigma \rightarrow 0$, and μ represents the probability statement that would be perfectly correct for the entire history. Conversely, as $\tau_b \rightarrow 1$, σ expands so that the distribution has finite probabilities at 0 and 100%, and zero probability everywhere else, refer to Figure 5. Ideally, for a τ_b of reasonable size, we would like to see the standard deviation as small as possible. A small standard deviation would indicate that the probability statement varied little from time period to time period. However, to achieve this may require a large value for τ_b, which the market for predictions may question.

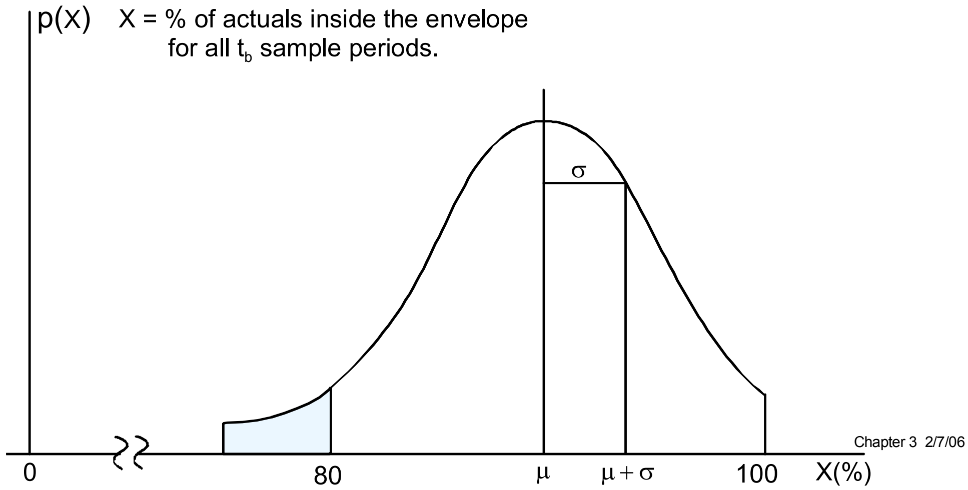

Figure 5. Statistical distribution of the number of times the predictions fall inside the envelope.

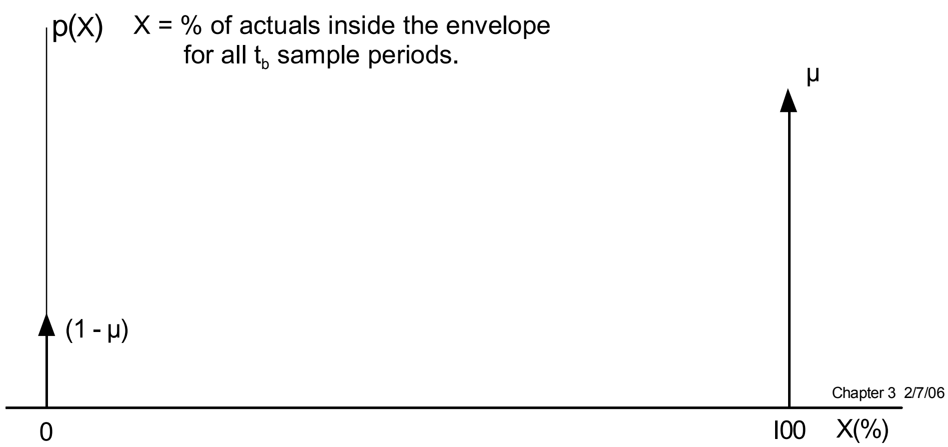

Figure 6. Distribution when the looking back horizon, τ_b, equals one.

A client of the producer of the predictions may point out that the latest predictions are not meeting an accurate probability criteria since, over the last 26 weeks, the actual values have fallen outside of the τ_p = 12 prediction interval (farthest out horizon) 6 times. Therefore, it should have been called a 77% envelope (at best) since actuals were outside slightly more than 23% of the time.

The producer of the predictions may be concerned that 26 weeks is an insufficient time period to characterize the probability statement, and consider longer term records which show that actuals have been inside the 12 step prediction interval better than 80% of the time. In fact, at τ_p = 12, they have been in 81.5% of the time in the prior year in Figure 4.

The client may say that the producer had a great model but, over the past few years, its accuracy has degraded. Since the market is most concerned about current history, a producer must consider how to address it. The first decision to be made is what "looking back" horizon, τ_b, is best used to characterize its probability statement. The shorter the horizon, the more appealing in the market. After much thought, the producer concludes that it must consider horizons on a quarterly basis, and that a single quarter might be watched, but that two quarters (26 weeks) is probably the shortest realistic time period from a "statistical" standpoint.

Our goal is to develop measures of accuracy that also serve to measure consistency of the model for small looking back horizons over long periods of history. This can be accomplished using confidence intervals about the prediction envelope boundaries for a given τ_b. In general, for any given τ_b, we can determine the confidence level (e.g., 95%) for which we will be inside the (80%) envelope. Assuming that the distribution of points in Figure 4 were normal, then maximum consistency can be achieved by minimizing the variance, or the mean absolute deviation, given a desired looking back horizon, τ_b, and probability prediction envelope, e.g., 80% .

A Measure Of Prediction Quality

Using the above definitions, we can pose a measure of *quality of prediction* that accounts for the actual width of the envelope for a given probability (e.g., 80%), see [9]. The following measure applies for a particular forward prediction horizon, τ_p, and looking back horizon, τ_b.

$$Q(\tau_p, \tau_b) = \frac{C*P}{1 + W}$$

where:
- Q is the measure of prediction quality,
- P is the probability that future values will fall within the envelope at a given τ_p (80% in the above examples),
- C is the confidence in the value of the probability statement for a given τ_b (95% in the above examples),
- W is the mean normalized width of the envelope, relative to the actual value, at τ_p.

Using this measure, quality improves (degrades) with increasing (decreasing) probability of being inside the envelope, and with increasing (decreasing) confidence in the probability. It also improves (degrades) as the width of the envelope grows smaller (larger). As the statement of probability of being inside the envelope approaches unity (100%) and the confidence in the statement approaches unity (100%), and the width of the envelope approaches zero, quality approaches unity, and predictions approach certainty.

BUILDING ACCURATE PREDICTION MODELS

The prediction subsystem in Figure 1 is generally composed of a model of the system being observed. It is the modeler's task to build models that maximize the accuracy of predictions. This requires applying additional information to condition the probability statements. Additional information can come from observation data, but typically best comes from knowledge of how a system operates internally. We start by characterizing the basic properties of dynamic systems.

Homogeneous Versus Nonhomogenious Systems

Time-variant systems may be defined by sets of differential or difference equations. These systems may be divided into homogeneous and nonhomogeneous parts. Homogeneous systems are self-contained, i.e., they are not affected by external driving forces that vary independently with time. Normal planetary motion is an example of a homogeneous system. Planetary positions can be predicted based upon the internal mechanics of the system itself. This assumes that no external forces affect the planetary system, e.g., large meteorites hitting a planet.

Homogeneous systems are represented as functions of time. These representations may be complex, but are generally stationary, implying the manner in which they vary with time can be identified with sufficient accuracy for the foreseeable future of interest.

The systems of interest here, e.g., missile guidance, monetary systems, etc., are generally nonstationary. They may have stationary components, but their outcomes are heavily influenced by external forces that are nonstationary. In these cases, one must model how external factors act as leading forces that can be observed in advance of the system's response. This implies modeling how inertial properties of one entity affect those of another. Unless a system has inertial properties whose time constants are sufficiently long, there is little chance of predicting future responses with useful accuracy. Given observation data for the driving forces as they occur, one must be able to represent the behavior of how they affect the inertial properties of the system with sufficient accuracy. This generally requires a substantial understanding of the internal mechanics of the system itself.

Modeling The Effects Of External Driving Forces - An Example

As an example, consider predicting the number of housing completions in a given geographical area months in advance. Such predictions can be used to predict sales of appliances, communication systems, furniture, carpeting, etc., purchased after a house is complete. Actual completions can be measured by certificates of occupancy issued in a given month. The most significant factor affecting housing completions is building permits. These are normally taken out many months prior to completion. We start with an analysis of one month's worth of building permits to determine the resulting distribution of completions for those permits. Let's assume that our investigation yielded an average distribution that took on a shape as shown in Figure 7. Then we could model the resulting distribution as shown where the number of housing units in the distribution equaled the building permits taken out (or some percentage if all did not result in completions). Figure 8 shows the superposition of housing completions due to building permits taken out in months 3 and 10.

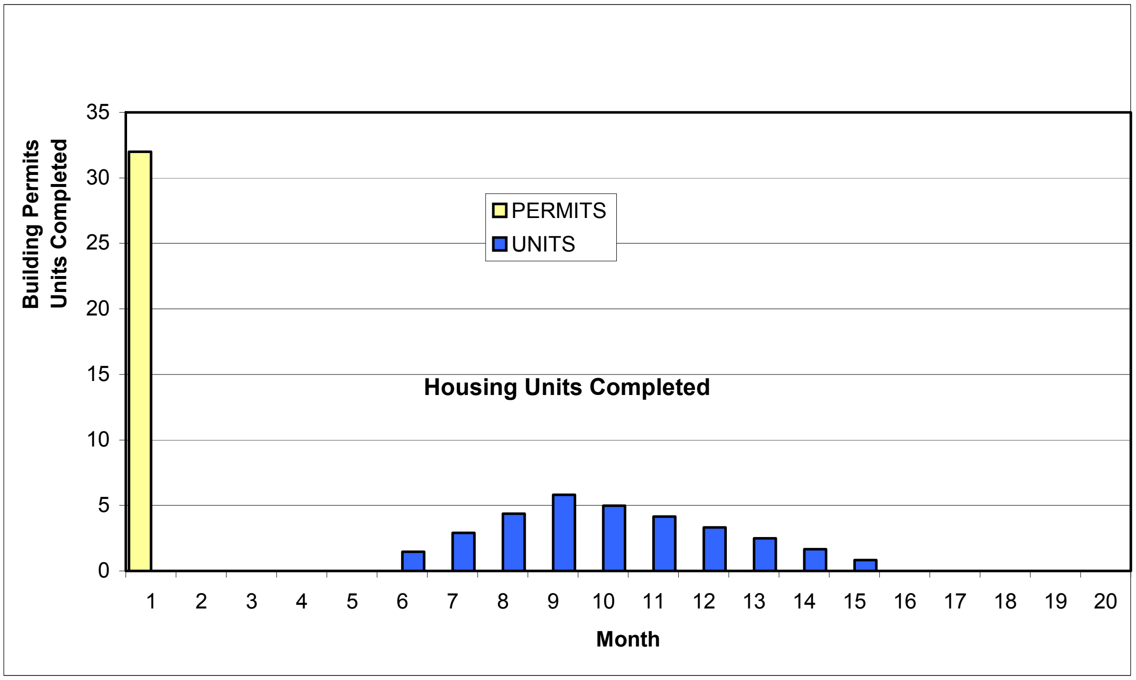

Figure 7. Housing units completetled in months 6 through 15 as a result of building permits in month 1.

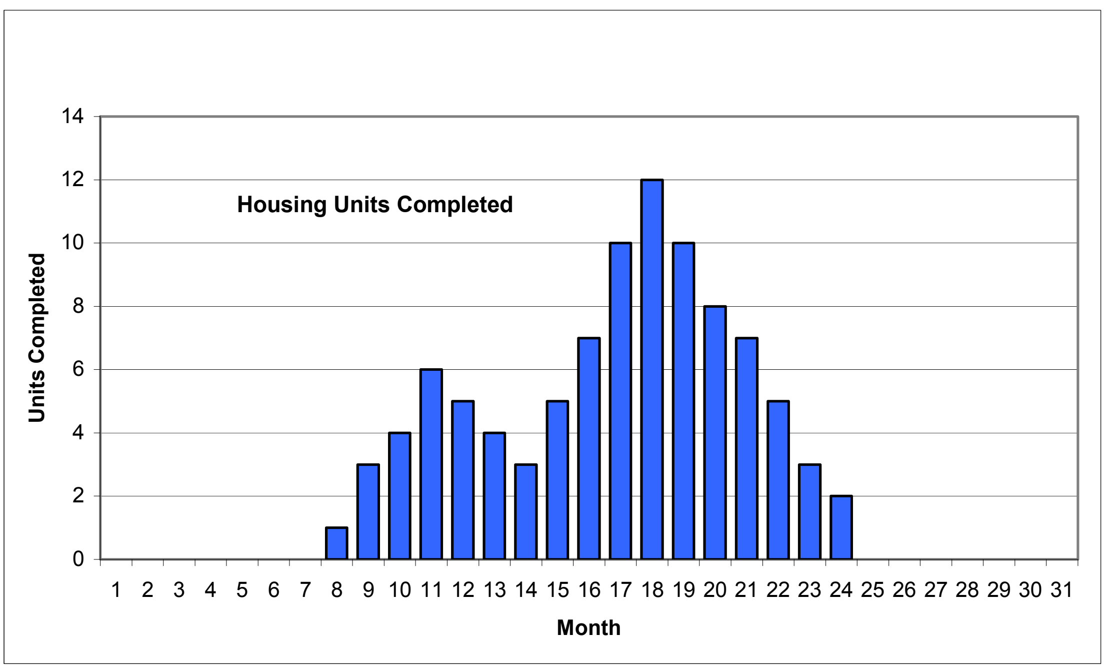

Figure 8. Housing units completed in months 8 through 24 as a result of building Permits in taken out in months 3 and 10.

Modeling Quaisi-Stationary Systems

Characterizing the properties of statistical stationarity in a complex system also requires a substantial understanding of its behavior. One must separate the internal inertial properties of the system from those driven by external forces. Having done so, identifying models that represent the internal stationary components may be achieved only using approaches that go well beyond typical tests for stationarity.

Three years of M1 data, Jan 1981 - Jan 1984, Not Seasonally Adjusted (NSA), are shown in Figure 9 where the data appears to be moving up and down more randomly than that which would result from the input driving forces. Clearly one must look for correlation with other sources. Although the data jumps around in what may at first appear to be a random fashion, it becomes clear that, after special testing, the up and down movement is correlated with the calendar. Thus we will look for correlation with the calendar. Figure 10 shows the data behind the plot in Figure 9.

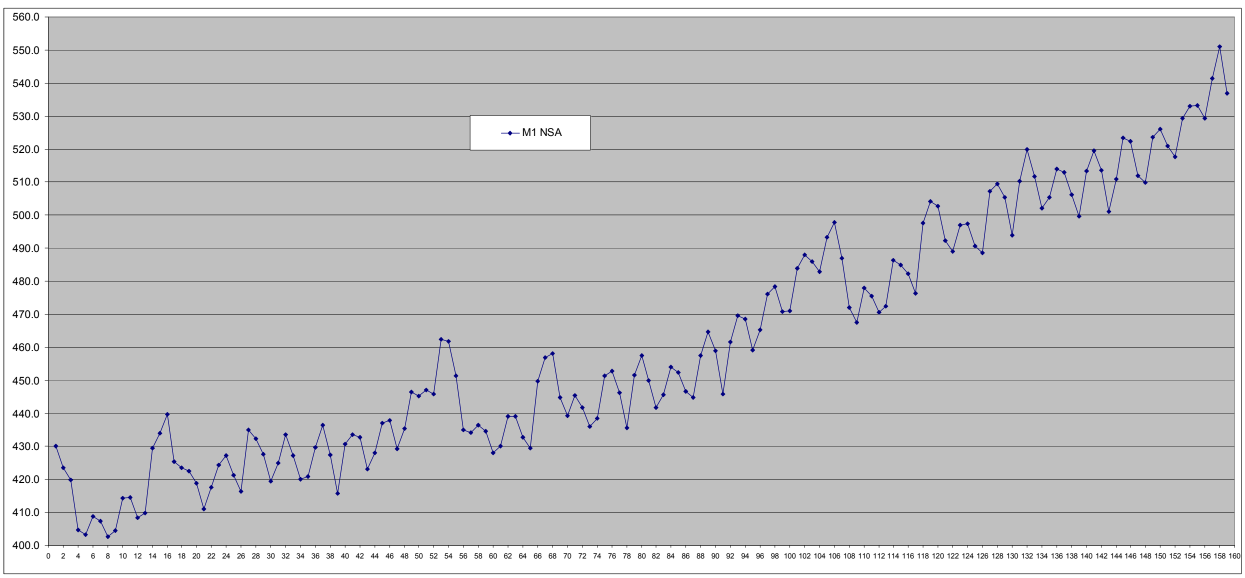

Figure 9. Actual curve appears almost random.

The actual data in Figure 10 has all of the "bottom" points highlighted in yellow. There are 12 of these in each year, each occurring at the transition between months. Four major peaks are highlighted in blue. These peaks occur in the 1st or 2nd week of the beginning of the year. Three major "double" peaks are highlighted in red. These occur the week before and the week after April 15th, tax time.

From the curves, it is clear that a special type of correlation analysis - based upon the calendar - is required to determine coefficients that could be used to improve the accuracy of predictions so that the width of the 80% envelope is as small as possible.

Although the data is produced once a week, it is correlated on a monthly and annual as well as weekly basis, requiring a special correlation analysis. These components can be analyzed independently, where the time scale with the most correlation can be used to pull out that component and redo the correlation analysis on the residual data using the second component.

	DATE					M1 - NSA
1	Jan	5	1981			430.1
2	Jan	12	1981			423.6
3	Jan	19	1981	3	4	419.8
4	Jan	26	1981			404.7
5	Feb	2	1981			403.3
6	Feb	9	1981			408.7
7	Feb	16	1981	3	4	407.4
8	Feb	23	1981			402.6
9	Mar	2	1981			404.5
10	Mar	9	1981			414.3
11	Mar	16	1981	3	5	414.6
12	Mar	23	1981			408.3
13	Mar	30	1981			409.8
14	Apr	6	1981			429.4
15	Apr	13	1981	3	4	433.9
16	Apr	20	1981			439.7
17	Apr	27	1981			425.4
18	May	4	1981			423.5
19	May	11	1981	3	4	422.5
20	May	18	1981			418.9
21	May	25	1981			411.0
22	Jun	1	1981			417.6
23	Jun	8	1981	4	5	424.4
24	Jun	15	1981			427.2
25	Jun	22	1981			421.3
26	Jun	29	1981			416.4
27	Jul	6	1981			435.0
28	Jul	13	1981	3	4	432.3
29	Jul	20	1981			427.7
30	Jul	27	1981			419.5
31	Aug	3	1981			425.0
32	Aug	10	1981	4	5	433.5
33	Aug	17	1981			427.2
34	Aug	24	1981			420.0
35	Aug	31	1981			420.8
36	Sep	7	1981			429.6
37	Sep	14	1981	3	4	436.5
38	Sep	21	1981			427.5
39	Sep	28	1981			415.7
40	Oct	5	1981			430.6
41	Oct	12	1981	3	4	433.5
42	Oct	19	1981			432.8
43	Oct	26	1981			423.2
44	Nov	2	1981			428.0
45	Nov	9	1981	3	5	437.1
46	Nov	16	1981			437.8
47	Nov	23	1981			429.2
48	Nov	30	1981			435.4
49	Dec	7	1981	4	4	446.5
50	Dec	14	1981			445.3
51	Dec	21	1981			447.0
52	Dec	28	1981			445.9
53	Jan	4	1982			462.5
54	Jan	11	1982			461.7
55	Jan	18	1982	3	4	451.4
56	Jan	25	1982			435.0
57	Feb	1	1982			434.2
58	Feb	8	1982			436.5
59	Feb	15	1982	3	4	434.5
60	Feb	22	1982			428.0
61	Mar	1	1982			430.1
62	Mar	8	1982			439.0
63	Mar	15	1982	3	5	439.0
64	Mar	22	1982			432.8
65	Mar	29	1982			429.4
66	Apr	5	1982			449.7
67	Apr	12	1982	4	4	456.9
68	Apr	19	1982			458.2
69	Apr	26	1982			444.9
70	May	3	1982			439.3
71	May	10	1982			445.4
72	May	17	1982	3	5	441.8
73	May	24	1982			436.0
74	May	31	1982			438.4
75	Jun	7	1982			451.4
76	Jun	14	1982	3	4	452.8
77	Jun	21	1982			446.3
78	Jun	28	1982			435.5
79	Jul	5	1982			451.6
80	Jul	12	1982	3	4	457.4
81	Jul	19	1982			450.0
82	Jul	26	1982			441.8
83	Aug	2	1982			445.6
84	Aug	9	1982	4	5	454.0
85	Aug	16	1982			452.4
86	Aug	23	1982			446.7
87	Aug	30	1982			444.8
88	Sep	6	1982			457.5
89	Sep	13	1982	3	4	464.6
90	Sep	20	1982			459.0
91	Sep	27	1982			445.9
92	Oct	4	1982			461.5
93	Oct	11	1982	3	4	469.5
94	Oct	18	1982			468.5
95	Oct	25	1982			459.1
96	Nov	1	1982			465.3
97	Nov	8	1982	4	5	476.1
98	Nov	15	1982			478.3
99	Nov	22	1982			470.8
100	Nov	29	1982			471.0
101	Dec	6	1982			483.9
102	Dec	13	1982	3	4	488.0
103	Dec	20	1982			486.0
104	Dec	27	1982			482.9
105	Jan	3	1983			493.2
106	Jan	10	1983	4	5	497.7
107	Jan	17	1983			486.9
108	Jan	24	1983			472.0
109	Jan	31	1983			467.5
110	Feb	7	1983			477.9
111	Feb	14	1983	3	4	475.5
112	Feb	21	1983			470.5
113	Feb	28	1983			472.5
114	Mar	7	1983			486.3
115	Mar	14	1983	3	4	484.9
116	Mar	21	1983			482.3
117	Mar	28	1983			476.3
118	Apr	4	1983			497.5
119	Apr	11	1983	4	4	504.2
120	Apr	18	1983			502.8
121	Apr	25	1983			492.2
122	May	2	1983			489.1
123	May	9	1983			497.0
124	May	16	1983	3	5	497.4
125	May	23	1983			490.6
126	May	30	1983			488.6
127	Jun	6	1983			507.3
128	Jun	13	1983	3	4	509.4
129	Jun	20	1983			505.3
130	Jun	27	1983			494.0
131	Jul	4	1983			510.2
132	Jul	11	1983	3	4	519.9
133	Jul	18	1983			511.7
134	Jul	25	1983			502.1
135	Aug	1	1983			505.4
136	Aug	8	1983	4	5	513.9
137	Aug	15	1983			513.0
138	Aug	22	1983			506.1
139	Aug	29	1983			499.7
140	Sep	5	1983			513.3
141	Sep	12	1983	3	4	519.5
142	Sep	19	1983			513.5
143	Sep	26	1983			501.1
144	Oct	3	1983			510.9
145	Oct	10	1983	4	5	523.4
146	Oct	17	1983			522.3
147	Oct	24	1983			511.9
148	Oct	31	1983			509.8
149	Nov	7	1983			523.5
150	Nov	14	1983	3	4	526.1
151	Nov	21	1983			520.9
152	Nov	28	1983			517.6
153	Dec	5	1983			529.4
154	Dec	12	1983	3	4	533.0
155	Dec	19	1983			533.2
156	Dec	26	1983			529.4
157	Jan	2	1984			541.3
158	Jan	9	1984	4	5	551.0
159	Jan	16	1984			536.8
160	Jan	23	1984			520.5
161	Jan	30	1984			508.8

Figure 10. Data used to produce time correlation factors.

To identify calendar correlation on observed data, Z(T), one must perform comparisons for a 5 or 7 day week; a 4 or 5 week month; and a 12 month year. One must also determine how to handle transitions when there are holidays, and especially when holidays fall on Friday or Monday, the transition at the end of a week, or transitions at the end of a month or year.

The usual definition of randomness implies no correlation with time, i.e., no autocorrelation. The usual test states that Z(T) is random when the expected value of the inner product of the deviates is sufficiently close to zero for all $\tau > 0$. We will use the notation:

$$E\{Z(T), Z(T+\tau)\} < \delta \approx 0 \quad \text{for all } \tau > 0.$$

where

$$E\{Z(T), Z(T+\tau)\} = \frac{1}{T_T} \cdot \sum_{T=1}^{T_T} [DZ(T) \cdot DZ(T+\tau)] ,$$

$$DZ(T) = [Z(T) - \mu_Z],$$

and μ_Z is the expected value of Z over the period of interest:

$$\mu_Z = E\{ Z(T) \} = \frac{1}{T_T} \cdot \sum_{T=1}^{T_T} Z(T) .$$

Since we are dealing with bounded data sets, we will interpret randomness as follows: Z(T) is *not* random if a transformation C can be found such that for some $\tau > 0$,

$$E \{C[Z(T)], Z(T+\tau)\} \geq \varepsilon_\tau$$

where ε_τ is a sufficiently large value based on judgement. When this is true, Z(T) is predictable to some extent up to τ steps into the future. Otherwise, Z(T) is *apparently random*. The word "apparently" is used to imply that we can never be sure that a data set is random, i.e., how do we know that, if a C cannot be found, one does not exist. This is best explained by way of example. Modeler A uses a standard autocorrelation test and comes up with a value ε_A which is less than ε_τ. Modeler B uses a special "window" to search for autocorrelation and obtains $\varepsilon_B > \varepsilon_A$, but still less than ε_τ. Modeler C uses a special function C which allows for variations in the "period of periodicity" of the data, and comes up with $\varepsilon_C >> \varepsilon_\tau$. (As an example of changing periodicity, the product of two periodic functions with different periods will appear aperiodic over a bounded time frame). We would expect model C to provide substantially more accurate predictions relative to models A and B.

The above examples indicate that what one person perceives to be random in time, another may determine as having a high degree of order with time. In other words, there appears to be no single measure of randomness for a bounded data set.

Probably the best example of this phenomenon is encountered in cryptography. Here one creates ciphers using "pseudo" random codes which, when tested by people from whom information is to be hidden, *appears* to be random. Those having the "key" to decipher the code (i.e., they know the transformation C), can retrieve intelligible data which can contain information relating to future values of the data set, including new keys.

DEFINING THE PREDICTION PROBLEM

With the framework provided above, we can proceed to define the prediction problem.

System Uncertainty

Based upon the above, we define an *uncertain process* as follows. A process, Z(T), is said to *appear random* when no transformation C can be found for which the expected value:

$$E\,[C[Z(T)],\, Z(T+\tau)] \;\geq\; \varepsilon_\tau, \quad \text{for any } \tau > 0.$$

For nonhomogeneous systems, we say that Z(T) is an *uncertain process* relative to driving force vector U(T) when no transformation C can be found such that, for any $\tau > 0$,

$$E\,[C[U(T), Z(T)],\; Z(T+\tau)] \;\geq\; \varepsilon_\tau$$

Neither of these statements implies that a C does not exist, only that it has not been found.

System Predictability

We say Z(T) is a *predictable process* of order τ when a vector of driving forces U and transformation C can be found such that for $\tau > 0$,

$$E\,[C[U(T), Z(T)],\; Z(T+\tau)] \;\geq\; \varepsilon_\tau$$

We note that a process which appears random by standard statistical tests can be predictable since Z(T) can be a delayed function of a purely random process U(T). This represents a generalization of the *Markoff Process,* being conditioned on (nonhomogeneous) driving forces, observed τ states (time steps) back.

Referring back to Figure 7, we see that the process shown is predictable up to 6 steps into the future with potentially little error. If we attempt predictions 7 steps into the future with this model we incur an error, since an impulse at the next (observable) time step will affect the response 6 steps in the future. This is a *prediction* error due to the inherent order of predictability of the system. This must be distinguished from the model or observation error which is generally treated in control theory literature. We are assuming, of course, that the driving force has unpredictable components. When we construct state equations containing error terms, we must incorporate an *additional error term* beyond those reflecting uncertainty in the model and in the data.

Modeling or Estimation Error

The following measure is offered to optimize the choice of U and corresponding transformation C. We want to find C(T) and U(T) such that

$$\Phi(C, U) = D\,[C[U(T), Z(T)],\ Z(T+\tau)]$$

is minimized, where D is some measure of distance (e.g., mean absolute deviation) between the predicted response based on the model,

$$\hat{Z}(T+\tau) = C[U(T), Z(T)] = \hat{Z}(T+\tau|T)$$

and the actual response $Z(T+\tau)$. For example, C and U can be selected to minimize the mean absolute error function

$$(1) \qquad \hat{e}^{-}(C, U, Z) = \hat{e}\,[\hat{Z}(T+\tau|T),\ Z(T+\tau)]$$

$$= \mathbf{E}\left[\ \left|\frac{C[U(T),\ Z(T)] - Z(T+\tau)}{Z(T+\tau)}\right|\ \right]$$

A similar measure would be to minimize the mean square error. We note that the selection of U and C depend, in general, on τ. In practice, one can select the value of τ most critical to the application. Or, some functional combination of $\hat{e}^{-}$ at various values of τ can be used.

However, once we use (1) as a performance measure in an optimization process, then *information at* $T+\tau$ *has been incorporated into the model*. Therefore,

$$(2) \qquad Z(T+\tau) = C[U(T), Z(T), Z(T+\tau)] = Z(T+\tau|T+\tau),$$

is not a true prediction - *it is an estimation* - and any future error measure will be of the form $\hat{e}^{+}(C, U, Z)$.

Correlating Prediction Error to Modeling or Estimation Error

The measure ê used for modeling error can also be used for prediction error. What is important is that the data sets are different. All data up to the current time T can be used to optimize C and U so as to minimize $\hat{e}^{+}(C, U, Z)$, providing an optimal estimate. Future data beyond the current time must be used to measure prediction error. If reductions in modeling error do not correlate to reductions in prediction error, then the modeler has no consistent method for improving model accuracy in a way that reduces prediction error.

To summarize, if the same error function, e.g., $\hat{e}^{-}$ in (1) above, is used to measure both modeling error and prediction error, the difference in the measures is essentially the use of previously available data versus the use of unseen “future” data.

SUMMARY

The problem of building structural (versus statistical) models is currently being faced by practitioners who are trying to produce more accurate forecasts. The problem stems from the most difficult task of translating knowledge of a system's structure into a model, and the subsequent difficulties in verification and validation of executable computer code. Because of these difficulties, most forecasters fall back on statistical approaches, fitting the data with mathematical functions that get extrapolated into the future.

We wish to note the complexity involved in defining and solving the multi-step prediction problem. One is typically trying to predict a vector of observable responses out to some maximum number of time steps (horizon) into the future for which predictions are required. Typically, complex systems are neither linear, homogeneous, nor stationary, so that an understanding of the "mechanics" of the system is necessary to approach such a problem. In practice, one must comprehend these mechanics in order to postulate candidate driving force vectors, and then model the mechanics to produce transformations that relate future values of the response to the driving forces. To accomplish this, it is necessary to define meaningful distance measures to maximize prediction accuracy (minimize prediction error).

To summarize the results described above, the following tabular comparison is offered.

History Data	**Future Data**
$Z(1), ..., Z(T)$	$Z(T+1), ..., Z(T+\tau)$
Modeling (Estimation) Error	**Prediction Error**
$\hat{e}^{+}(C, U, Z)$	$\hat{e}^{-}(C, U, Z)$
$\hat{e}^{+}[\hat{Z}(T\|T), Z(T)]$	$\hat{e}^{-}[\hat{Z}(T+\tau\|T), Z(T+\tau)]$

Referring to the comparisons above, modeling (estimation) error can be measured using a model conditioned on all data up to and including the final measurement time, T. In the case of prediction error, the dynamic model, which is part of the error function, can only be conditioned on information up to the current time, T, which is τ steps back from the final measurement.

When optimizing model parameters to reduce prediction error, a correlation must exist between $\hat{e}^{-}$ and $\hat{e}^{+}$, to ensure that reducing modeling error implies reducing prediction error. Else, the modeler has no criteria for improving a model. It is clear that determination of this correlation can involve substantial amounts of hidden data in order to ensure that the correlation test uses true prediction error, i.e., *it is based on data the modeler has not yet seen*.

If one simply uses a naive function to fit the history data, it is doubtful that the properties of the system will be discovered, no matter how powerful the mathematical techniques used to identify or optimize the curve fitting parameters. However, if a modeler builds a structural model based on an understanding of the mechanics of the system, he need only use the data to validate his model and measure prediction accuracy. Furthermore, the likelihood of correlation, between modeling error and prediction error, will be much higher.

REFERENCES

1. Athans, M. and Falb, P.L., **Optimal Control**, McGraw-Hill, New York, 1966.
2. Athans, M. and Kendrick, D., *Control Theory and Economics: A Survey, Forecast and Speculation*, IEEE Trans, Automatic Control, pp. 518-524, Oct. 1974.
3. Black, Fisher and Myron Sholes, *The Pricing of Options and Corporate Liabilities*, Journal of Political Economy, Vol.81, No. 3, 1973.
4. Box, G. and Jenkins, G., **Time Series Analysis**, Holden Day, San Francisco, CA, 1976
5. Cave, W.C. and Rosenkranz, E., *A Stochastic State Space Model for Prediction of Product Demand*, AFIPS-Conference Proceedings, Volume 48, AFIPS Press, pp. 67-72, 1979.
6. Cave, W.C., *A General Approach To Supply/Demand Market Prediction Using State Space,* Proceedings, Institute of Management Science XXIV International Meeting, Honolulu, Hawaii, July 1979.
7. Cave, W.C. and Guilfoyle, R.H., *A General Approach to Supply/Demand Market Prediction Using the State Space Framework.* First International Forecasting Symposium, Quebec, CD, 1982.
8. Cave, W.C. and T.E. Busch, *The Theoretical Underpinnings Of Predictive Battlespace Awareness (PBA)*, Proceedings of SPIE, Vol 5091, Orlando, FL, April 2003.
9. Cave, W.C., **Prediction Theory for Control Systems**, Prediction Systems, Inc., Spring Lake, NJ, 2014.
10. Gelb, A., Editor, **Applied Optimal Estimation**, MIT Press, Cambridge, MA, 1974.
11. Kalman, R., **A New Approach to Linear Filtering and Prediction Problems**, Trans. ASME, Series D: Journal Basic Engineering 82, pp. 35-45, 1960.
12. Kalman, R., *Keynote Address*, Third International Symposium on Forecasting Program, Philadelphia, PA., pp. 150, 1983.
13. Mehra, R., *Kalman Filters and Their Applications to Forecasting*, TIMS Studies in the Management Sciences 12, North-Holland, 1979.
14. Schweppe, F., **Uncertain Dynamic Systems**, Prentice Hall, Englewood, NJ, 1973.