2005 American Control Conference
June 8-10, 2005. Portland, OR, USA

A Novel Approach to Model Determination using the Minimum Model Error Estimation

Jason R. Kolodziej
General Motors Fuel Cell Activities
Honeoye Falls, NY, 14472
jason.kolodziej@gm.com

D. Joseph Mook
Department of Mechanical and Aerospace Engineering
SUNY at Buffalo, Buffalo, NY, 14260
mook@eng.buffalo.edu

Abstract— **The purpose of this paper is to present an algorithm for the combination of a proven nonlinear system identification technique, the Minimum Model Error estimation algorithm (MME) with an Analysis of Variance (ANOVA) correlation routine where a forward stepwise procedure is implemented. The Analysis of Variance approach to model identification is well documented primarily in social science literature but has been sparsely written about for engineering applications. This paper will show a significant improvement in nonlinear model identification when used in conjunction with the MME algorithm.**

I. Introduction

In 1985, Mook and Junkins [1] published, for the first time, a technique called the Minimum Model Error (MME) algorithm, that used a post-process batch algorithm to estimate the states of poorly modeled dynamical systems. The basis of this algorithm was to take discrete time measurement data (which may be noisy), assume a model for the system, and by minimizing a given cost function, determine the unmodeled system dynamics used to correct the assumed model. The motivation for their work was mainly in the aerospace industry, but it has since been applied in many other areas.

MME is not only a state estimation algorithm. In addition to the estimated states, a model correction term is determined that is used to improve the assumed model. Much of the research since the initial work has been devoted to model correction. In 1991, Mook and Stry [5], [6] furthered the previous work by developing a correlation routine that attempted to relate the unknown time-based model correction term with nonlinear (or linear) functions of state. Now that a state-based functional update had been found the assumed model can consequently be updated to result in an improved model of the system.

The proposed algorithm is a modified version of a stepwise regression routine (MSR) that is based on Analysis of Variance (ANOVA). Stepwise regression is well documented in social science and some biological modelling, of which the text Draper and Smith [12] is an excellent reference. However it is noticeably absent in engineering literature. A survey paper by Haber and Unbehauen [2] does make reference to stepwise regression, but only in an input-output system identification scheme, where the measurements are perfect and the model only takes the form of sampled time delays of the input and output. Likewise papers by Leontaritis and Billings [3], and Kukreja, et al. [4] again only seek to identify NARMAX type models which only approximate the nonlinear dynamics, not identify the true dynamics.

The proposed algorithm seeks to extend the the functional correlation routine of the MME algorithm by applying a modified version of the stepwise regression algorithm. The acceptance criteria for admittance to the model becomes more problem dependant rather fixed based on the number of data points.

Section II is a brief outline of the Minimum Model Error estimation algorithm. Section III shows the theoretical development of the modified stepwise regression algorithm. Section IV's intent is to apply the proposed algorithm to the Van der Pol benchmark example. Section V concludes this work and presents some future work.

II. Minimum Model Error Estimation

The Minimum Model Error (MME) estimation algorithm has been the subject of much research since its original development by Mook and Junkins in the mid-1980's. The problem of formulating an accurate mathematical model from state observable noisy measurement data has applications in many engineering disciplines, such as the aerospace industry, structural identification, and robotics.

The remainder of this section is devoted to the equations of the MME algorithm. It serves as only a brief review, for a full theoretical development see either of the papers by Mook [1], [7]. The present correlation routine, along with its problems, is also outlined.

A. The Estimation Problem

Define a system whose state vector dynamics are modelled by the following system of equations,

$$\dot{\underline{x}} = \underline{f}[\underline{x}(t), \underline{u}(t), t] + \underline{d}(t) \tag{1}$$

where $\underline{x}(t) \equiv$ n x 1 state vector, $\underline{f}(t) \equiv$ n x 1 vector of model equations, $\underline{u}(t) \equiv$ n x 1 vector of forcing terms, and $\underline{d}(t) \equiv$ n x 1 vector of unmodelled system dynamics. Given a set of discrete, state-observable measurements, modelled by the system of equations,

$$\underline{\tilde{y}}_k = \underline{g}_k[\underline{x}(t_k), t_k] + \underline{v}_k \qquad k = 1, 2, ..., M \tag{2}$$

0-7803-9098-9/05/$25.00 ©2005 AACC

where $\underline{\tilde{y}}_k \equiv$ m x 1 measurement vector at t_k, $\underline{g}_k \equiv$ m x 1 measurement model at t_k, and $\underline{v}_k \equiv$ gaussian error with zero mean and known covariance, R_k.

Notice the $\underline{d}(t)$ term in (1), this is the vector of unmodeled disturbances. In other words, this is the nonlinear (or linear) terms not included in the original model estimate. The estimation problem is to determine the optimal estimate for $\underline{x}(t_k)$, during the interval $[t_0, t_M]$.

B. The Covariance Constraint

In the MME algorithm the estimate is determined based on the assumption that the measurement-minus-estimate error covariance matrix must match the measurement-minus-truth error covariance matrix. The goal is to estimate the data to within the same variance as the device taking the measurement. Mathematically, this condition is defined by Mook [1] as the Covariance Constraint according to the following formula:

$$E\{[\tilde{y}(t_k) - g(\hat{x}(t_k), t_k)][\tilde{y}(t_k) - g(\hat{x}(t_k), t_k)]^T\} = R_k \quad (3)$$

Therefore, the estimate, $g(\hat{x}(t_k), t_k)$, must match the actual measurements, $\tilde{y}(t_k)$, with the known covariance, R_k, to be correct.

As with any optimization problem one must determine a cost function in which the resulting operation optimizes the parameters desired. The MME method uses the following cost function for determining the state estimates, and later $\underline{d}(t)$. It can be seen from this cost function that there is a continuous minimization of $\underline{d}(t)$ across the interval $[t_1, t_M]$ and the discrete, internal penalty functions based on the covariance constraint, penalizing the measurement estimate for varying from the actual measurement.

$$\begin{aligned} J &= \sum_{k=1}^{M} \left[\underline{\tilde{y}}_k - \underline{g}_k[\underline{\hat{x}}(t_k), t_k]\right]^T R_k^{-1} \left[\underline{\tilde{y}}_k - \underline{g}_k[\underline{\hat{x}}(t_k), t_k]\right] \\ &\quad + \int_{t_1}^{t_M} \underline{d}(t)^T W \underline{d}(t) dt \end{aligned} \quad (4)$$

where, $\underline{\hat{x}}(t) \equiv$ state vector estimate, $W \equiv$ positive definite weighing matrix, and $M =$ number of measurements. It is apparent that the estimate $\underline{\hat{x}}(t)$ and the optimal $\underline{d}(t)$ are required to satisfy the differential equation in (1), thereby incorporating the assumed model with the unknown model terms. The first term in (4) is the weighted sum square of the residuals between the actual measurements and the estimated measurements, and the second term is a weighted integral sum square of the unmodelled disturbance.

By adjusting the weighting matrix, W, it is possible to penalize the model correction term ($\underline{d}^T W \underline{d}$) in this cost function so that the resulting state estimate will vary between the assumed model and the measurement data. The optimum selection of W is chosen based on satisfying the *Covariance Constraint*, or when the estimate covariance equals the measurement covariance.

C. Development of the TPBVP

Consider the cost function J in (4). The problem is to find a smooth, differentiable, unbounded $\underline{d}(t)$ which minimizes this cost function.

An algorithm to deal with an integral in the cost function with a differential equation constraint and discrete penalty terms was developed by Geering [8]. It accounted for internal penalty functions in the Hamilton-Jacobi problem through the modification of the Pontryagin's necessary conditions. Mook and Junkins later extended this work to allow for jump discontinuities, of which a detailed explanation can be found in [1], [7].

Utilizing the Lagrange multiplier technique, the minimization of the cost function results in the following two-point boundary value problem (TPBVP), with the $\underline{\lambda}(t)$ being a vector of Lagrange multipliers. Referring again to control system theory these are usually called the "co-states". To summarize, the TPBVP is then set up as follows:

$$\dot{\underline{x}} = \underline{f}[\underline{x}(t), \underline{u}(t), \underline{d}(t), t] \quad (5)$$

$$\dot{\underline{\lambda}} = -\left[\frac{\partial f}{\partial \underline{x}}\right]^T \underline{\lambda} \quad (6)$$

$$\underline{d} = -\frac{1}{2W}\left[\frac{\partial f}{\partial d}\right]^T \underline{\lambda} \quad (7)$$

with boundary conditions, $\underline{\lambda}(t_o) = \underline{0}$ and $\underline{\lambda}(t_f) = \underline{0}$ or, $\underline{x}(t_o)$ = Measured and $\underline{x}(t_f)$ = Measured.

In order to accommodate the internal penalty terms in the cost function, the co-state $\underline{\lambda}(t)$, and therefore, $\underline{d}(t)$ may have jump discontinuities at each discrete time measurement sample. Therefore, there are two values of $\underline{\lambda}(t_k)$, where k is the measurement sampling time. The following equation is used to update between $\underline{\lambda}(t_k^-)$ and $\underline{\lambda}(t_k^+)$:

$$\underline{\lambda}(t_k^+) = \underline{\lambda}(t_k^-) + 2H_k^T R_k^{-1}\left[\underline{\tilde{y}}_k - \underline{g}(\underline{x}(t_k), t_k)\right] \quad (8)$$

Where, H is defined as,

$$H \equiv \left.\frac{\partial g}{\partial \underline{\hat{x}}}\right|_{\underline{\hat{x}}(t_k)} \quad (9)$$

Many methods have been developed to handle two-point boundary value problems such as this in connection with the MME algorithm, including simple shooting [1], multiple shooting [9], and the Riccati transformation [10]. For the present work multiple shooting is utilized because of its improved performance over simple shooting and its ease of implementation. For a detailed derivation of the multiple shooting algorithm as it applies to the MME TPBVP see Mook and Lew [9], Stry [5].

D. Correlation & Least Squares

Once the TPBVP is solved, thereby concluding the MME algorithm, the result is a smooth estimate of the states, as well as the correction to the assumed model, $\underline{d}(t)$. The final step in the identification process is to interpret the $\underline{d}(t)$ vector. Ideally, the goal is to correlate this vector with some

function of the state estimates to get an improvement on the original model.

The present method of correlating a given set of data with various functions in order to find a functional relationship is by developing a library of correlation functions. The goal is to search the entire library of correlation functions and choose the one with the highest linear correlation coefficient.

If the correlated solution is not sufficient, the estimate is subtracted from the original $\underline{d}(t)$, leaving a modified $\underline{d}(t)$. The correlation algorithm, including the function library, is then applied to this new error vector to find another accurate correlation. The newly correlated function is now added to the old one. The process is repeated until a group of functional relationships are found that sufficiently represents the original error vector. The work by Stry [5] outlines this iterative procedure.

Obviously, if this process is allowed to continue long enough with a large function library an acceptable functional relationship will probably be found. However, a large string of functions detracts from the algorithm's efficiency, and yields model corrections which are typically too complicated for engineering problems, especially control applications.

III. The Proposed Correlation Routine

The primary purpose of this paper is to extend the research on the Minimum Model Error algorithm, namely the correlation routine. The major drawbacks of the present routine is that it only has the ability to add functions to determine the model form. As a result, after one or two functions are added to the model typically only incremental improvements are made with addition of more functions.

Thus this paper presents an algorithm that will add *and* subtract functions according to a proven statistical distribution. At the conclusion of the algorithm a combination of functions that best fits the unknown nonlinearity is found.

A. The Modified Stepwise Regression Algorithm

The initial theoretical basis for the algorithm presented is relatively well documented in scientific literature over the past 50 years. In his 1990 survey paper, Haber [2] outlined many nonlinear identification techniques which utilized input-output data, one of which was the forward-stepwise regression algorithm.

A detailed outline for the forward-stepwise regression technique as it applies to statistical data problems has been written by Draper and Smith [12]. The equations that follow are based on a collection of text on the subject of regression. For brevity, the theoretical basis for multiple regression has been omitted.

1) Phase I: The first step in the algorithm is choosing an F^* threshold for statistical significance. This threshold is the minimum partial F-ratio allowed for a function to be admitted to the model. In general, this selection can be tricky procedure. The selection of F^* is based on a chosen confidence level, the number of degrees of freedom, and the F-distribution.

In order to properly apply the modified stepwise regression technique the following equations need to be applied. The F-ratio for all functions in the current model is:

$$F_M = \frac{MSR}{MSE} \tag{10}$$

Where MSR is the mean square error due to regression and MSE is the mean square error about the regression, and are calculated as follows:

$$MSR = \frac{SSR}{dof_R} \qquad MSE = \frac{SSE}{dof_E} \tag{11}$$

The remaining equations (12)-(15) determine the sum of squares and degrees of freedom for the F-ratio.

$$SSR = \sum_{k=1}^{n} (\hat{Y}_k - \overline{Y})^2 \tag{12}$$

$$SSE = \sum_{k=1}^{n} (\tilde{Y}_k - \hat{Y}_k)^2 \tag{13}$$

$$dof_R = \text{\# of functions in the model} \tag{14}$$

$$dof_E = \text{\# of data points - \# of functions} \tag{15}$$

where, $\tilde{Y}_k \equiv$ data point at k, $\hat{Y}_k \equiv$ model estimate at k, and $\overline{Y} \equiv$ average estimate value $= \frac{1}{n}\sum_{k=1}^{n} \hat{Y}_k$.

In order to determine the model estimate, $\hat{Y}$, simply perform a least-squares calculation for a model consisting of all accepted functions plus a constant.

The next step is to add each function not already in the model individually, least square fitting the expanded model, and determining if there is a statistical significance of increasing the size of the model. This is accomplished by compiling a list of the partial F-ratios of all the functions not in the model, and selecting the largest one for addition into the model, assuming that it is above the threshold value. (16) is the result, where F_A denotes the partial F-ratio of functions to be added that are not currently in the model.

$$F_A = \frac{MSR(X_k|X_{k-1}, X_{k-2}, ..., X_{k-n-1})}{MSE(X_k, X_{k-1}, X_{k-2}, ..., X_{k-n-1})} \tag{16}$$

The mean square error, MSE, is easy to calculate according to (11). MSR is only slightly more difficult to calculate in that it done by determining the improvement on the model error by adding the function, X_k.

$$MSR(X_k|X_{k-1}, X_{k-2}, etc.) = \\ = \frac{SSR(X_k|X_{k-1}, ..., X_{k-n-1})}{1} \tag{17}$$

$$= \frac{SSR(X_k, ..., X_{k-n-1}) - SSR(X_{k-1}, ..., X_{k-n-1})}{1} \tag{18}$$

For the more general case, (18) results in,

$$F_A = (dof_E)\left[\frac{[SSE(X_{k-1}, ..., X_{k-n-1})]}{[SSE(X_k, X_{k-1}, ..., X_{k-n-1})]} - 1\right] \tag{19}$$

Now that a partial F-ratio has been calculated for each function not in the model, a determination as to which function to add must be made. For this procedure the significance level (α) is chosen. 90%, 95%, and 99% are commonly tabulated in almost all introductory statistics books, see Draper [12]. If none of the functions meet this criteria the algorithm is complete and Phase II begins. However, if at least one function is able to be included in the model the algorithm proceeds to the next step, which is to determine if any functions current in the model should be removed.

This calculation is performed in exactly the same manner as (16) except only the functions X_{k-1} through X_{k-n-1} are considered. This results in the following equation, F_R. Note that $k-1$ is only the first of $n-1$ partial F-ratio values to be calculated.

$$F_R = \frac{MSR(X_{k-1}|X_k, X_{k-2}, ..., X_{k-n-1})}{MSE(X_{k-1}|X_k, X_{k-2}, ..., X_{k-n-1})} \tag{20}$$

If it is determined that a function needs to be removed the function with the smallest partial F-ratio is chosen. At this point the removal process is repeated to determine whether another function can be removed.

2) Phase II: Once it is determined that no further functions should be added to the model and none should be removed Phase I of the modified stepwise regression algorithm is complete. Typically, this would signal the end of the modelling. However, it is found that when attempting to combine the algorithm with the MME algorithm the termination of the algorithm became problematic, mainly because of the increased size of the number of the functions possibly added to the model. With the combination of the stepwise regression algorithm and MME, the user-defined function library could easily be extended to thousands (or even millions) of functions. It became evident that functions added to the model (typically at the end of the model building process) are only added because they passed a single criteria of having a partial F-ratio greater then F^*. There is no mechanism for incorporating the relative distribution of partial F-ratios of functions currently in the model. Meaning some functions in the model have partial F-ratios on the order of $> 10^5$ while others were 10^0.

By maintaining a fixed threshold for statistical significance, a unique phenomenon is seen. The number of functions in the model is not based solely on F^*, but rather on the number of functions in the function library. The more functions in the library, the more likely unnecessary functions are added to the model. A larger function library with more diversity typically produces a better fit, but at the expense of a larger model with many superfluous terms having partial F-ratios only slightly greater than F^*.

The idea of the modified stepwise-regression algorithm is simple. It is developed as a means to filter away the extra terms that are added to the model because they are "statistically significant", but not necessary. Essentially, the algorithm employs a floating threshold. The process of adding and removing functions from the model is exactly the same as the original algorithm with hypothesis tests based on partial F-ratio calculations. However, the threshold is moved to one order of magnitude below the function with the largest partial F-ratio. This typically causes at least one function (usually more) in the model to be below the threshold. The lowest function that violates the "new" F^* is removed, as before, with new partial F-ratios calculated for each function remaining in the model, as well as those out of the model. In addition, this re-calculation of the partial F-ratios of the functions in the model results in another "new" F^*.

The process continues until no functions can be removed from the model. At this time the function library is re-checked for possible additions to the model, except now the possibility for admittance to the model is significantly more strict due to the elevated threshold. It is witnessed that by removing these functions a more concise model form is found. It is shown, by considering the ignorance assumed model Van der Pol osciallator from the previous section, that the modified stepwise regression algorithm outperforms the original model determination algorithm.

IV. The Complete System Identification Algorithm

Now that a improved model determination algorithm is available it is advantageous to present the complete system identification algorithm. Figure 1 illustrates the Minimum Model Error estimation with Modified Stepwise Regression algorithm. (MME/MSR). By studying the ignorance assumed model Van der Pol oscillator the overt advantages are seen.

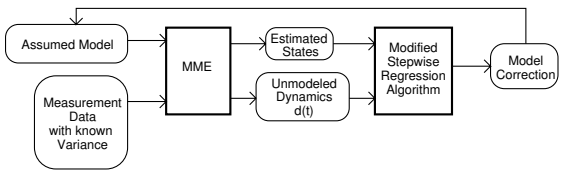

Fig. 1. The Proposed System Identification Algorithm.

A. Van der Pol Example

For this illustrative example, the Van der Pol oscillator is used:

$$\ddot{x} + \alpha(x^2 - 1)\dot{x} + x = 0 \tag{21}$$

where the dots represent the 1^{st} and 2^{nd} time-based derivatives, subject to an initial condition input with only noisy State #1 measurements of known variance $\sigma_M^2 = 0.01$. This means that the unmodeled dynamics term contains all three functions, $\underline{d}(t) = \alpha x^2 \dot{x} - \alpha \dot{x} + x$.

The following conditions are applied: $\alpha = 1.0$, noisy State #1 measurements, simulated at 40Hz, with an initial condition input, $x_0 = [1\ 0]^T$. The measurement data and assumed model are shown in Figure 2, and with W set to five the state estimate is also included.

Func. #	Func.	Func. #	Func.
1-6	x_1^i	13-16	$\lvert x_1^i \rvert$
7-12	x_2^i	17-20	$\lvert x_2^i \rvert$
$-3 \le i \le 3$ $i \neq 0$		$-3 \le i \le 3$ $i \neq 0$ $i \neq$ even	

TABLE I

FUNCTION LIBRARY - VAN DER POL EXAMPLE.

The estimate is excellent in satisfying the covariance constraint $\sigma^2_{meas-est} = 0.0122 \approx \sigma^2_M$. In addition an estimate minus truth variance of $\sigma^2_{truth-est} = 0.001814$ for State #1, and $\sigma^2_{truth-est} = 0.0141$ for the unmeasured State #2 is noted.

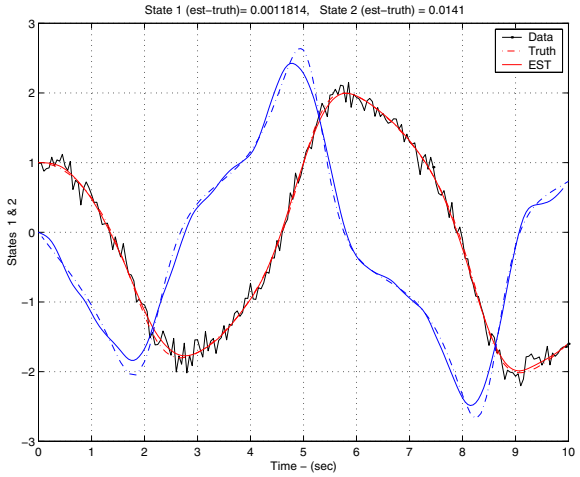

Fig. 2. Estimated States 1 & 2. $\sigma^2_M = 0.01$. W=5.2, Ignorance Assumed Model.

With the results from MME, the modified regression algorithm is now applied. The first task is to compile a function library the results of which are shown in Table I. Even in this modest function library the number of possible model functions is 145. The resulting unmodeled dynamics, $\underline{d}(t)$, for the ignorance assumed model are shown in (22) and represent the final to be determined model of the modified stepwise regression algorithm:

$$\begin{aligned} \underline{d}_{truth}(t) &= -x_1 + x_2 - x_1^2 x_2 \qquad (22) \\ &\equiv (4) \quad (10) \quad (71) \end{aligned}$$

The resulting model fit after the 29^{th}, and final, iteration is also plotted with the data and estimate in Figure 3. The model fit is very good with a truth minus model variance of $\sigma^2_{truth-model} = 0.08114$. While this model fit is a significant improvement over the original correlation routine model, the most impressive result is that the truth model is found.

Observe the iteration history of the modified regression algorithm as shown in Figure 4, and consider the threshold of admittance and removal from the model during Phase I. F^* is chosen to be ten at the beginning of the algorithm and remains constant until no additional functions can be added

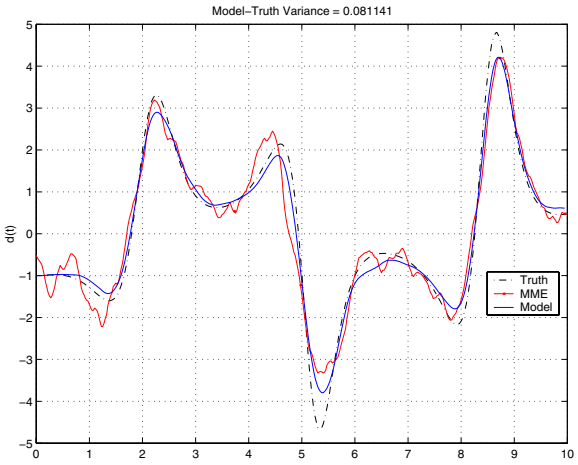

Fig. 3. Unmodeled Dynamics, $d(t)$ - Truth, Estimate - $\sigma^2_{truth-est} = 0.1878$, And MSR Model - $\sigma^2_{truth-model} = 0.08114$ - Van der Pol Example With Ignorance Assumed Model.

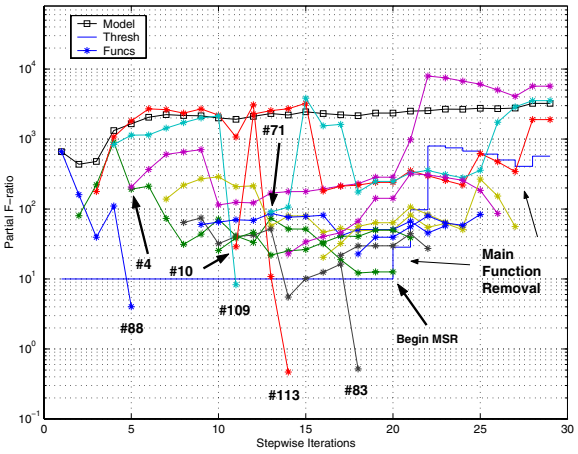

Fig. 4. Modified Stepwise Regression Iteration History - Van der Pol Example With Ignorance Assumed Model.

or removed (through Iteration #20). During this time a great deal of additions and removals take place. Specifically, follow the first function added to the model, Function #88. It is admitted to the model because it contains the highest partial F-ratio of any function not in the model at the 1^{st} iteration. However, by the 5^{th} iteration four additional functions are added to the model and it is shown that Function #88 no longer contributes enough to the model, and is removed. This is an outstanding result! The original correlation routine concludes with this iteration with only the newly added Function #4 being correct.

Throughout the fixed threshold portion of the algorithm four functions are removed from the model because they are deemed to not contribute to the model. An interesting aside that occurs in this example is in Iteration #14. Two functions drop below the threshold during this iteration due to the inclusion of Function #71, a truth function. With only one function allowed to be removed per iteration, the partial

F-ratio of the other function rises enough to no longer be in violation of the threshold.

Coefficient of Determination	0.9607
Multiple Corr. Coefficient	0.9802
Variance of Estimate	0.1226
STD Error of Estimate	0.3502
Intercept	-0.0155

Source of Variation	dof	SS	MS	F-ratio
Due to Regression	3	1188.1	396.0	3230.0
Deviation About Regr	396	48.6	0.123	
Total	399	1236.7		

Functions In The Model			Functions Not In The Model	
Function	Coeff.	Partial F-ratio	Function	Partial F-ratio
4	-0.9874	5688.1	124	128.0
10	0.8125	1895.2	114	109.3
71	-0.8292	3537.6	12	87.8
0	-0.0155		52	46.3

TABLE II

29^{th} ITERATION - 399 DP - FINAL MODEL ANOVA TABLE - VAN DER POL EXAMPLE WITH IGNORANCE ASSUMED MODEL.

At Iteration #20, the standard stepwise regression ends, with 11 functions in the model! (Had $F^* = 2.6$ which corresponds with theory for a three term model with 401 data points. A standard stepwise regression with a modest 145 term function library would have added 17 functions to the model!) From a practical standpoint this is far too many for a simple second-order system. The modified regression algorithm is developed for exactly this purpose. Phase II occurs from Iteration #20 through Iteration #28 where the eleven function model is reduced to a three function model, all of which are part of the truth model from (22). Notice that from Iteration #22 through Iteration #26 only one function is not in violation of the modified F^* threshold, Function #4. As the lower partial F-ratio functions are removed, however, the partial F-ratios of the truth functions rise. At the 29^{th}, and final, iteration only the truth remains. Table II is the final model ANOVA table for this example. Notice that functions not currently in the model have partial F-ratios greater than the original threshold for significance, but yet do not appear in the model. It is clear, however, that the functions *in* the model are much more significant, comparatively, than those *out* of the model. Also notice that the model F-ratio begins to increase slightly during Phase II indicating that the three function model is a better representation of $\underline{d}(t)$ than the eleven term model.

V. CONCLUSIONS & RECOMMENDATIONS

The main purpose of this paper is to extend the current state of the art in model determination. Future work is planned to study the composition of the function library, as well as to study the moving F^* threshold. A relationship between sample rate, system dynamics, and the partial F-ratio calculation would have a lasting effect on the application of this system identification algorithm.

REFERENCES

[1] Mook, D. Joseph, and Junkins, John L., "Minimum Model Error Estimation for Poorly Modeled Dynamioc Systems," *AIAA Journal of Guidance, Vol.11, No.3*, May-June 1988, pgs. 256-261.

[2] Haber R. Unbehauen, H.,"Structure Identification of Nonlinear Dynamic Systems - A Survey on Input/Output Approaches" *Automatica*, Vol. 26, No. 4, pp. 651-667, 1990.

[3] Leontaritis I.J., Billings, S.A.,"Input-output parametric models for non-linear systems part I: deterministic non-linear systems" *International Journal of Control*, 41(2) pp. 303-328, 1985.

[4] Kukreja, S., Galiana, H., Kearney, R.,"Structure Detection of NARMAX Models Using Bootstrap Methods" *Proceedings of the* 38^{th} *Conference on Decision & Control*, December 1999, pp. 1071-1076, 1999.

[5] Stry, Greselda I, and Mook, D. Joseph, "Identification of Nonlinear Dynamic Systems Using Correlation Techniques," *Proc. Struc. and Struc. Dynamics Materials Conference,*, Baltimore, MD, 1991.

[6] Mook, D. Joseph, and Stry, Greselda I., "An Analog Experimental Study of Nonlinear Identification," Nonlinear Dynamics, Vol. 3, No. 1, pp 1-11, 1992.

[7] Mook, D. Joseph, "Estimation and Identification of Nonlinear Dynamic Systems," *AIAA Journal, Vol. 27, No. 7*, July 1989, pgs. 968-974.

[8] Geering, Hans P., "Continuous-Time Optimal Control Theory for Cost Functionals Including Discrete State Penalty Terms," *IEEE Transactions on Automatic Control*, December 1976.

[9] Mook, D. Joseph, and Lew, Jiann-Shuin, "Multiple Shooting Algorithms for Jump-Discontinuous Problems in Optimal Control and Estimation," *IEEE Transactions on Automatic Control, Vol.36, No.8*, August 1991, pgs. 979-983.

[10] Crassidias, John L., Mason, Paul A.C., Mook, D. Joseph, "Riccati Solution for the Minimum Model Error Algorthm" *AIAA Journal of Guidance, Control, and Dynamics*, Vol. 16, No. 6, July 1993, pp. 1181-1183.

[11] Mendenhall, William, Wackerly, Dennis D., Scheaffer, Richard L., *Mathematical Statistics with Applications: Fourth Edition* Duxbury Press. 1990.

[12] Draper, Norman R., Smith, Harry, *Applied Regression Analysis Third Edition*, Prentice Hall, 1998.