Use of the Analysis of Variance Method for Model Order Selection of N-FIR Models

Ingela Lind

Department of Electrical Engineering
Linköping University, S-581 83 Linköping, Sweden
WWW: http://www.control.isy.liu.se
Email: ingela@isy.liu.se

August 16, 1999

Report no.: LiTH-ISY-R-2177
Submitted to CCSSE

Technical reports from the Automatic Control group in Linköping are available by anonymous ftp at the address ftp.control.isy.liu.se. This report is contained in the pdf file 2177.pdf.
Use of the Analysis of Variance Method for Model Order Selection of N-FIR Models

Ingela Lind
August 16, 1999

Abstract

Identification of non-linear FIR-models is studied. In particular the selection of model structure, i.e., to find the contributing input time lags, has been examined. A common method, exhaustive search among models with all possible combinations of the input time lags, has some undesired drawbacks, as a tendency that the minimization algorithm gets stuck in local minima and heavy computations. To avoid these drawbacks we need to know the model structure prior to identifying a model.

In this report we show that a statistical method, the multivariate analysis of variance, is a good alternative to exhaustive search in the identification of the structure of non-linear FIR-models.

We can reduce the risks of getting an erroneous model structure due to the non-convexity of the minimization problems, reduce the computation time needed and also get a good estimate of how far we can enhance the fit of the desired model.

1 Problem Description

Assume that a non-linear FIR model describes the measurements $y_t$ from a system with input $u_t$, that is,

$$y_t = g(u_t, u_{t-T}, u_{t-2T}, ..., u_{t-kT}) + e_t. \quad (1)$$

The value of $k$ is unknown in addition to which time lags of $u_t$ that contributes to the value of $y_t$ and $g$ is an unknown static non-linear function of up to $k + 1$ variables.

A common method used to estimate the function $g$ is to parameterize it as a non-linear neural network. This method has some drawbacks:

Since the model order is not known in advance, it is necessary to assume a largest $k$ and estimate a network of each combination of the time lags of $u_t$ under that assumption. Of these models, the one with the best performance on validation data is chosen. Each model takes time to estimate and depending on the assumption of $k$, there can be many models.

During the estimation of the neural network, the optimization algorithm can stop in a local minimum, which can lead to a model with wrong assumptions on relevant input time lags coming out best.

Both these drawbacks can be avoided if the structure of the model is known in advance, i.e., the contributing time lags and $k$ are known. In order to achieve
this knowledge, it is possible to use the Multivariate Analysis of Variance method (MANOVA) from statistics [4, 3]. This way, we only need to estimate one model, thereby reducing the computation time considerably, and need not worry about getting an erroneous model structure because of the optimization algorithm has been stuck in a local minimum.

2 Exhaustive Search for Finding Model Structure

The idea behind the exhaustive search method to find the model structure is to enumerate all possible combinations of the input time lags and estimate a model for each such combination. When all the models are estimated, their prediction performance are compared with some good measure, e.g., the root mean square error between measured and predicted output. The one model that has the best performance is chosen as the model for the system. If we are lucky, this model has the same structure as the system we want to identify the structure of. This method does not distinguish between the task of finding the model structure and the task of finding a model, thereby a lot of tuning is done to improve models before we know if they are going to be used or not.

The models used to parameterize the system depend on our assumptions on how the system is working, i.e., if it is linear, non-linear, dynamic or static etc.. If we had been identifying a linear system there would not have been any big problems with using the exhaustive search method to find the correct structure. The identification methods for finding good linear models are efficient and reliable. Of course, we could still do the wrong choices sometimes, due to the non-deterministic nature of the system. In this report we have been trying to identify the structure of non-linear systems though, and that is where the problems start. There are no really efficient and completely reliable methods to estimate a non-linear model of a non-linear system. It is common to use neural networks to do it, but these suffer from the lack of search algorithms that guarantees that the global minima is found. They also take lots of computations to find a minima at all. So, we cannot be sure that the global minima is found for each of the models we compare and can thereby not be sure that we have found the model structure that describes the system best.

3 The Multivariate Analysis of Variance

The statistical analysis method MANOVA is a widely spread tool for finding out which factors contribute to given measurements. Though common in medicine and quality control applications, it does not seem to have been tried in system identification applications.

The method is based on hypothesis tests with F-distributed test variables computed from the residual quadratic sum. Below, one variant of the method is stated in a statistical framework for two factors. The complexity grows rapidly with the number of factors.

Assume that we have a batch of $abn$ observations, corresponding to the $ab$ treatment/level combinations of the factors $A$ and $B$. $A$ has $a$ different levels, $B$ has $b$ different levels and the experiment is repeated $n$ times. The measurements
are made in random order to be sure to avoid effects of time dependency etc.
The observations may be described by a linear statistical model,

\[ y_{ijk} = \mu + \tau_i + \beta_j + (\tau \beta)_{ij} + \epsilon_{ijk}, \]  

(2)

where \( i = 1, \ldots, a \), \( j = 1, \ldots, b \), \( k = 1, \ldots, n \), \( \mu \) is the overall mean effect, \( \tau_i \) is the effect of the \( i \)th level of the factor \( A \), \( \beta_j \) is the effect of the \( j \)th level of the factor \( B \), \( (\tau \beta)_{ij} \) is the effect of the interaction between the \( i \)th level of factor \( A \) and the \( j \)th level of the factor \( B \) and \( \epsilon_{ijk} \) is a random error component. This means that we assume that there are deterministic effects from the level of the factors, and that the stochastic effects are totally due to measurement noise.

The treatment effects are defined to be fixed deviations from the overall mean, so \( \sum_{i=1}^{a} \tau_i = 0 \), \( \sum_{j=1}^{b} \beta_j = 0 \), \( \sum_{i=1}^{a} (\tau \beta)_{ij} = 0 \), \( \forall j \) and \( \sum_{j=1}^{b} (\tau \beta)_{ij} = 0 \), \( \forall i \).

We are interested in testing the following hypotheses regarding the treatment effects:

\[ H_{0AB} : (\tau \beta)_{ij} = 0, \forall i, j, \]  

(3)

that is, the measurements do not depend on the level combination of factors \( A \) and \( B \) (interaction effect), against

\[ H_{1AB} : \text{at least one } (\tau \beta)_{ij} \neq 0, \]  

(4)

that there is at least one significant interaction effect. If the null hypothesis is accepted, we are also interested in testing:

\[ H_{0A} : \tau_1 = \tau_2 = \ldots = \tau_a = 0, \]  

(5)

\[ H_{0B} : \beta_1 = \beta_2 = \ldots = \beta_b = 0, \]  

(6)

against, respectively,

\[ H_{1A} : \text{at least one } \tau_i \neq 0, \]  

(7)

\[ H_{1B} : \text{at least one } \beta_j \neq 0. \]  

(8)

If any of the null hypotheses is rejected, we assume that the factor involved does have some effect on the measurements. To do the hypothesis tests we use a two-factor analysis of variance. We compute the overall mean, the cell means and the means over the factor levels. Let

\[ \bar{y}_{..} = \frac{1}{abn} \sum_{i=1}^{a} \sum_{j=1}^{b} \sum_{k=1}^{n} y_{ijk}, \]  

(9)

\[ \bar{y}_{i..} = \frac{1}{bn} \sum_{j=1}^{b} \sum_{k=1}^{n} y_{ijk}, \]  

(10)

\[ \bar{y}_{.j.} = \frac{1}{an} \sum_{i=1}^{a} \sum_{k=1}^{n} y_{ijk}, \]  

(11)

\[ \bar{y}_{ij.} = \frac{1}{n} \sum_{k=1}^{n} y_{ijk}. \]  

(12)
The total residual quadratic sum can be written

\[
SS_T = \sum_{i=1}^{a} \sum_{j=1}^{b} \sum_{k=1}^{n} (y_{ijk} - \bar{y}...)^2 \\
= \sum_{i=1}^{a} bn(\bar{y}_{i.} - \bar{y}...)^2 + \sum_{j=1}^{b} an(\bar{y}_{.j} - \bar{y}...)^2 \\
+ \sum_{i=1}^{a} \sum_{j=1}^{b} n(\bar{y}_{ij.} - \bar{y}_{i.} - \bar{y}_{.j} + \bar{y}...)^2 + \sum_{i=1}^{a} \sum_{j=1}^{b} \sum_{k=1}^{n} (y_{ijk} - \bar{y}_{ij.})^2 \\
= SS_A + SS_B + SS_{AB} + SE, \quad (13)
\]

where

\[
SS_A = \sum_{i=1}^{a} bn(\bar{y}_{i.} - \bar{y}...)^2, \quad (14)
\]

\[
SS_B = \sum_{j=1}^{b} an(\bar{y}_{.j} - \bar{y}...)^2, \quad (15)
\]

\[
SS_{AB} = \sum_{i=1}^{a} \sum_{j=1}^{b} n(\bar{y}_{ij.} - \bar{y}_{i.} - \bar{y}_{.j} + \bar{y}...)^2 \quad (16)
\]

and

\[
SS_E = \sum_{i=1}^{a} \sum_{j=1}^{b} \sum_{k=1}^{n} (y_{ijk} - \bar{y}_{ij.})^2. \quad (17)
\]

From a theorem by Cochran [4, page 59] it is possible to show that

- the stochastic variables \( SS_A, SS_B, SS_{AB} \) and \( SS_E \) are independent,
- the stochastic variable \( \frac{1}{\sigma^2} \cdot SS_E \sim \chi^2(ab(n-1)) \),
- if \( \tau_1 = \ldots = \tau_a = 0 \), then \( \frac{1}{\sigma^2} \cdot SS_A \sim \chi^2(a-1) \),
- if \( \beta_1 = \ldots = \beta_b = 0 \), then \( \frac{1}{\sigma^2} \cdot SS_B \sim \chi^2(b-1) \),
- if \( (\tau\beta)_{ij} = 0, \forall i, j \), then \( \frac{1}{\sigma^2} \cdot SS_{AB} \sim \chi^2((a-1)(b-1)) \).

These observations are used to design test variables to test the interesting hypotheses. The test variable associated with factor \( A \) is chosen as

\[
v_A = \frac{SS_A/(a-1)}{SS_E/(ab(n-1))}. \quad (18)
\]

If \( H_{0A} \) is true, then \( v_A \sim F(a-1, ab(n-1)) \), i.e., \( v_A \) is \( F \)-distributed with \( a-1 \) and \( ab(n-1) \) degrees of freedom. The null-hypothesis is rejected if we get a large value of \( v_A \), that is, we reject \( H_{0A} \) if \( v_A > c \), where \( c \) is taken from an \( F(a-1, ab(n-1)) \)-table and \( \alpha \) denotes the level of significance (the probability to reject \( H_0 \) though \( H_0 \) is true). \( H_{0B} \) and \( H_{0AB} \) are tested analogously. We can also estimate the standard deviation, \( \sigma \), associated with the random error component, as \( \hat{\sigma}^2 = \frac{SS_E}{ab(n-1)} \). The degrees of freedom are \( ab(n-1) \). Note that \( n \) needs to be larger than 1 to do the analysis with all interaction effects.
The results from the hypotheses testing can be used to determine which factors have effect on the measurements and if there are interaction effects between different factors.

The most important modelling simplifications made are the assumptions that the variance is constant through the batch and that the random error component is Gaussian distributed. The F-tests are quite robust against violations against both assumptions [3, Chapter 7]. If the levels of the input cannot be regarded as fully deterministic or if the effects of the input are partly stochastic it is better to use another variant of the analysis of variance method [4, 5].

Now, to use MANOVA in a system identification application, we can consider the non-linear FIR-model

\[ y_t = g(u_t, u_{t-1}) + e_t. \] (19)

\( y_t \) is seen as the measurements, \( u_t \) as factor A and \( u_{t-1} \) as factor B. This structure gives that \( a = b = m \), the number of levels we allow \( u_t \) to assume.

4 Experimental Design

We want to design an experiment to determine which time lags have effect on the output of a system, assumed to be described by a non-linear FIR model, 
\[ y_t = g(u_t, u_{t-1}, \ldots, u_{t-k}) + e_t, \] where \( y \) is the output, \( g \) is an unknown function of the input and \( e \) is additive Gaussian white noise with variance \( \sigma^2 \). The result will both give us the important time lags and the main structure of the function \( g \), e.g., with \( k \) assumed equal to 3,

\[ g(u_t, u_{t-1}, u_{t-2}, u_{t-3}) = g_1(u_t, u_{t-2}) + g_2(u_{t-2}, u_{t-3}) + g_3(u_{t-1}). \] (20)

To do identification experiments with the described variant of MANOVA we need to construct a signal \( u_t \) that contains all possible combinations of levels of the time lags we want to examine. Assume that we have \( p \) time lags we want to examine, \( m \) levels of \( u_t \) and need \( n \) repetitions of each combination. The most straightforward way is to enumerate all possible sequences of length \( p \), with different level combinations of the time lags, repeat them \( n \) times, and sort them in random order. This gives us a sequence \( u_t \) of length \( p mn^p \). In this longer sequence, the short sequences will occur more than \( n \) times each, and we use only every \( p \)th measurement of the output from the system. It is also possible to create a periodic signal \( u_t \) of length \( n(m^p - 1) \), with the help of an \( p \)-stage, \( m \)-level shift register [1, page 41]. In this signal all except one of the possible combinations occur \( n \) times and we use all the output from the system. Measurements of the missing combination (all time lags assuming the lowest level) might be easy to add afterwards, since only a constant input signal is needed.

We have two structure identification methods to compare:

1. **Exhaustive search/non-linear neural network estimation**
   Assume largest possible \( k \).
   Estimate a non-linear neural network model for each combination of time lags fulfilling that assumption.
   Simulate all models and compare with validation data.
   The model with best performance, mean square error fit, is chosen.
2. MANOVA/non-linear neural network estimation

Assume largest possible \( k \).
Perform MANOVA for \( k \)-factor design on input/output data.
Choose model structure according to MANOVA.
Estimate a non-linear neural network model with the given structure.
Check performance on validation data and if better model is required, use more data or another type of network.

5 Test Specification and Results

To compare the different methods, we test their performance on a list of different functions that fits into

\[
y_t = g(u_t, u_{t-1}, u_{t-2}) + e_t,
\]

see Table 1. That is, we have assumed that \( k = 2 \) and will perform a three-way analysis of variance (\( p = 3 \)). For method 1 we use estimation data where the level of \( u_t \) is uniformly distributed between -2 and 5 (\( m = 4 \)). \( u_t \) is then filtered through \( g \) and Gaussian distributed noise with mean 0 and variance 1 is added to obtain \( y_t \).

For method 2 we use a signal \( u_t \) where the level of \( u_t \) can assume the values -2, 1, 3 and 5. All possible combinations of these levels in the sequence \( u_{t-2}, u_{t-1}, u_t \) are constructed. The signal is repeated \( n = 3 \) times. Then the output \( y_t \) is computed according to Equation 21 and the list of functions in Table 1. The input/output data are used both for the MANOVA test and for the estimation of the corresponding network. For the MANOVA test we have chosen the level of significance \( \alpha = 0.01 \). The implementation of the MANOVA test is made by us.

The number of input/output data, \( N \), is equal for both methods and set to \( N = n \cdot m^p = 3 \cdot 4^3 = 192 \).

The validation data is chosen in the same way as the estimation data for method 1. This means that the model chosen by method 1 is likely to get better performance on the validation data than the model chosen by method 2. The choice to estimate the neural network model in method 2 with the input/output data used for the MANOVA test is motivated by the wish to keep the number of needed measurements on the system as low as possible.

The networks used have two layers with 10 tansig neurons in the first layer and one purelin neuron in the second layer. We vary the number of input time lags between one and three and get seven different combinations of possible time lags. This means there are 31 to 51 parameters to estimate, depending on the number of input time lags included. We give the minimization algorithm one chance to find the global minimum on the estimation data, which might be too few chances to get a good result. The minimization algorithm used is Levenberg-Marquardt (MatLab’s Neural Network Toolbox is used).

The only performance criteria used in this test is the root mean square error (RMS) between measured output and simulated output from the model.

The implementation of the MANOVA test used in this experiment does not employ all useful features of the theory. It is possible to get the test more efficient and to gain more information by using it.
As we can see from Table 1, the MANOVA method is much better at identifying the correct factors (input time lags) that should be included in the model.

The experiment gives rise to a few questions: Will MANOVA perform better than the exhaustive search method in all situations? How much better? What happens if we try to do our best with exhaustive search and MANOVA respectively on estimation data and compare the methods on an untouched set of data? Is the type of network used unsuited for identifying some systems (e.g., the ones involving exponential functions)?

5.1 Further Tests

To do a 'proper' exhaustive search with neural networks we need to redesign the test. Assume that we have two sets of input/output data for the system we want to identify. Also assume that the datasets are suited for analysis with MANOVA, that is, all possible level combinations of the input are included. We put the second data set aside for validation use. From these data sets we want to find what input time lags are relevant, and if possible, a good model for the system. We will try the same identification methods as before with some modifications.

1. Exhaustive search Divide the first data set into estimation data and verification data. Construct neural networks for all possible combinations of time lags under the assumption of largest k. For each such combination, construct networks with different numbers of parameters (e.g. 5, 10 and 15 neurons in the first layer). Start with random network parameters and estimate the parameters on the estimation data. Start over 4 or 5 times with new random network parameters to try to avoid getting stuck in a
local minima. Of all these estimated networks, chose the one with the smallest RMS on the verification data.

2. **MANOVA** Use all of the first data set and perform MANOVA. This gives a model (which input time lags contribute to measurements) and a model structure (the interaction pattern of the inputs). In this report, we do not use the information about the model structure to improve the estimation of the model, that is, only the information about the contributing input time lags is used. Construct neural networks with different numbers of neurons (5, 10, 15) for the chosen model. Estimate as in the exhaustive search and choose the network with the smallest RMS on the verification data.

Note that one benefit of the MANOVA test is that we get a good estimate of the noise variance more or less automatically. This information can be used to determine when the minimization algorithm has failed to find a minima close to the global minima. If the RMS is much larger than the MANOVA estimated standard deviation we know that we have got a bad model. Try to restart the estimation process more times with new random network parameters to find a better local minima. If this does not lead to smaller RMS try to get more estimation data or change to another type of network. In the following test, this feature is not used.

Last, we compare the network chosen by exhaustive search and the network chosen by MANOVA on the validation data.

We are interested in how often the two methods can pinpoint the correct input time lags, and for MANOVA, also the correct model structure, for a specific function. We test this by running 100 Monte Carlo simulations of our test functions from Table 1. We also include an additional test function, a random network function, which is the same type of network as the ones we try to estimate the function $g$ with. We vary the signal to noise ratio and the level of significance for the MANOVA method and let the number of repeated measurements, $n = 4$. The results are collected in Table 2 and Table 3. The theoretical probability of finding the correct model structure is explained in Section 6.

From Table 2 and Table 3 we can draw the conclusion that the MANOVA method is much better at spotting what input time lags contribute to the output than the exhaustive search method. The results for the first function, $u_t - 0.03u_{t-2}$, shows that it is important to have not too small a signal to noise ratio (SNR). If the SNR is increased by a factor 4 the theoretical probability of finding the correct model structure by MANOVA increases from 4% to 89%.

The difference in performance between the two methods becomes more profound when the functions have a more non-linear behaviour, e.g., exponential functions. This indicates that the network we have been working with does not handle this kind of functions very well, which we can confirm by looking at the RMS values on the validation data.

In Table 3 the better performance for MANOVA as compared to Table 2 is mostly due to the increased significance level, except for the first function, where the decrease in noise variance is important to explain the better performance. We can also see that the decrease in noise variance does not affect the performance for the exhaustive search method either, except for the first function.
Table 2: Results from Monte Carlo simulations to test structure identification methods. Stated are the number of correctly chosen models out of 100. $n = 4$, $\sigma = 1$ and $\alpha = 0.01$. The theoretical upper limit of finding the correct model structure by MANOVA is computed as in Section 6.
<table>
<thead>
<tr>
<th>Function</th>
<th>Exhaustive search</th>
<th>MANOVA input time lags</th>
<th>MANOVA model structure</th>
<th>Theor. model structure</th>
</tr>
</thead>
<tbody>
<tr>
<td>$u_t - 0.03u_{t-2}$</td>
<td>94</td>
<td>100</td>
<td>100</td>
<td>99.95</td>
</tr>
<tr>
<td>$\ln</td>
<td>u_t</td>
<td>+ u_{t-1} + e^{u_{t-2}}$</td>
<td>78</td>
<td>100</td>
</tr>
<tr>
<td>$u_{t-1} \cdot [u_t + \frac{1}{u_{t-2}}]$</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>99.98</td>
</tr>
<tr>
<td>$\text{sgn}(u_{t-1})$</td>
<td>80</td>
<td>100</td>
<td>100</td>
<td>99.94</td>
</tr>
<tr>
<td>$\text{sgn}(u_{t-1}) \cdot u_{t-2}$</td>
<td>92</td>
<td>100</td>
<td>100</td>
<td>99.96</td>
</tr>
<tr>
<td>$\text{sgn}(u_{t-1}) \cdot u_t \cdot u_{t-2}$</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
</tr>
<tr>
<td>$\ln</td>
<td>u_{t-1} + u_{t-2}</td>
<td>$</td>
<td>94</td>
<td>100</td>
</tr>
<tr>
<td>$\ln</td>
<td>u_{t-1} + u_{t-2}</td>
<td>$</td>
<td>82</td>
<td>100</td>
</tr>
<tr>
<td>$u_{t-2} \cdot \ln</td>
<td>u_{t-1}</td>
<td>$</td>
<td>95</td>
<td>100</td>
</tr>
<tr>
<td>$u_{t-2} \cdot \ln</td>
<td>u_{t-1}</td>
<td>$</td>
<td>56</td>
<td>100</td>
</tr>
<tr>
<td>$u_{t-2} \cdot (\ln</td>
<td>u_{t-1}</td>
<td>)^3$</td>
<td>91</td>
<td>99</td>
</tr>
<tr>
<td>$</td>
<td>u_{t-2}</td>
<td>\cdot e^{u_{t-1}}$</td>
<td>54</td>
<td>100</td>
</tr>
<tr>
<td>$u_{t-2} \cdot e^{u_{t-1}}$</td>
<td>49</td>
<td>100</td>
<td>100</td>
<td>99.96</td>
</tr>
<tr>
<td>$u_{t-2} \cdot e^{u_{t-1}-0.03u_t}$</td>
<td>58</td>
<td>100</td>
<td>100</td>
<td>100</td>
</tr>
<tr>
<td>$</td>
<td>u_t</td>
<td>$</td>
<td>73</td>
<td>100</td>
</tr>
<tr>
<td>network</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$g(u_{t-1}, u_{t-2})$</td>
<td>94</td>
<td>100</td>
<td>100</td>
<td>-</td>
</tr>
<tr>
<td>TOTAL</td>
<td>80.6</td>
<td>99.9</td>
<td>99.9</td>
<td>99.96</td>
</tr>
</tbody>
</table>

Table 3: Results from Monte Carlo simulations to test structure identification methods. Stated are the number of correctly chosen models out of 100. $n = 4$, $\sigma = 0.0001$ and $\alpha = 0.0001$. The theoretical upper limit of finding the correct model structure by MANOVA is computed as in Section 6.
6 Significance and Power of MANOVA

In order to decide the appropriate amount of measurements necessary to gain an acceptable performance of the hypothesis tests, we need to know how to calculate the power of the tests. There are two measures of performance often used:

\[ 1 - \text{significance} = \alpha = P(H_0 \text{ rejected} | H_0 \text{ true}) \]  
\[ 1 - \text{power} = \beta = 1 - P(H_0 \text{ rejected} | H_0 \text{ false}) \]

We want both \( \alpha \) and \( \beta \) to be small. Do not confuse \( \beta \) with the effects \( \beta_j \). We use the wanted \( \alpha \) to calculate the critical limit for the test variable \( v \), that is, we regard \( \alpha \) as a design parameter. It is harder to get a value of \( \beta \), since we need an assumption of in what way the null hypothesis is false and the distribution of the test variable according to this assumption. For the hypothesis test associated with factor \( A \) in the two-way analysis of variance we have that [5]

\[ v_A = \frac{SS_A/(a-1)}{SS_E/ab(n-1)} \sim F(a-1, ab(a-1)) \text{ if } H_0 \text{ true} \]  
\[ v_A = \frac{SS_A/(a-1)}{SS_E/ab(n-1)} \sim \text{non-central } F(a-1, ab(a-1), \delta) \text{ if } H_0 \text{ false} \]

where the two first parameters in the \( F \)-distribution are the degrees of freedom and the third, \( \delta \), is a non-centrality parameter with

\[ \delta = na \sum_{i=1}^{a} \frac{\tau_i^2}{\sigma^2}, \]

which is closely related to the signal to noise ratio through \( \tau_i \). The formula for \( \delta \) depends on how many factors are included in the test and which interaction effect is tested, see [3, p 201]. The power of the test depends on the number of repetitions of the measurements, \( n \), the number of levels of \( u_t \), \( m \), and the deviation from the null hypothesis we want to test. The power is different for the tests of main effects and for the tests of interaction effects of different orders.

As an example of how to compute the power of the hypothesis tests we will describe how the theoretical probability of finding the correct model structure in Table 2 and Table 3 is computed. We use the first function, \( u_t - 0.03u_{t-2} \), for the demonstration.

**Example:** We are using a three-way analysis of variance and want to find what inputs have a significant effect on the output and if they interact. Input/output data from the function \( u_t - 0.03u_{t-2} \) are examined. We have 4 different levels that \( u_t \) can assume, and each measurement is repeated 4 times, that is, \( m = 4 \) and \( n = 4 \). The level of significance, \( 1 - \alpha \), equals 0.99 in the test and the noise is Gaussian with standard deviation 1. The factor \( A \) is associated with \( u_t \) and \( B \) with \( u_t \).

To find the correct model structure we need to:

- accept the null hypotheses for the interaction effects \( ABC, AB, AC, BC \) and \( B \), and
A level

-2  -1.94  -2.03  -2.09  -2.15  -2.0525  -3.75
C level  1.06  0.97  0.91  0.85  0.9475  -0.75
      3.06  2.97  2.91  2.85  2.9475  1.25
      5.06  4.97  4.91  4.85  4.9475  3.25

| means | 1.81 | 1.72 | 1.66 | 1.6  | 1.6975 |
| 1.06  | 0.97 | 0.91 | 0.85 | 0.9475 |
| 5.06  | 4.97 | 4.91 | 4.85 | 4.9475 |

Table 4: Function values, means and effects, \( u_t - 0.03u_{t-2} \).

- reject the null hypotheses for the effects A and C.

The probability to accept the null hypothesis when it is true is given by \( 1 - \alpha \) and the probability to reject the null hypothesis when it is false is given by \( 1 - \beta \). We neglect the fact that the different tests for the null hypotheses are not truly independent, due to the division by the estimated variance instead of the true variance in the test variables. We get an upper level for the probability to find the correct model,

\[
P(\text{find the correct model structure}) \leq (1 - \alpha)^5(1 - \beta_A)(1 - \beta_C). \tag{27}
\]

\( \beta_A \) is given by \( \beta_A = P(v_A < c_A | H_{0A} \text{ false}) \), where \( c_A \) is the critical limit with confidence level \( \alpha \) for the test variable \( v_A \), which belongs to the distribution

\[
v_A \sim \text{non-central } F(m - 1, m^3(n - 1), \delta_A) \tag{28}
\]

with

\[
\delta_A = nm^2 \sum_{i=1}^{m} \frac{\tau_i^2}{\sigma^2} \tag{29}
\]

when \( H_{0A} \) is false. To find the critical limit \( c_A \) we also need the distribution for \( v_A \) when \( H_{0A} \) is true,

\[
v_A \sim F(m - 1, m^3(n - 1)). \tag{30}
\]

See also Figure 1(a) and 1(b).

The value of \( \beta_C \) is computed analogously. The deterministic values for all factor combinations, mean values and factor effects are given in Table 4. It is easy to compute the effects for factor A, \( \tau_i \), \( i = 1, \ldots, 4 \) and for factor B, \( \gamma_k \), \( k = 1, \ldots, 4 \). We get \( \delta_A = 1.54 \) and \( \delta_C = 1713 \), and use tables to find the corresponding values \( \beta_A = 0.95 \) and \( \beta_C = 0 \). The result is that the probability to find the correct model structure is 4.2%. We can also verify that

\[
P(\text{find only factor } C) \leq (1 - \alpha)^6(1 - \beta_C) = 0.94, \tag{31}
\]

which means that we are very likely to assume that only factor C, \( u_t \), explains the output from the function.

7 Remarks

The variant of the MANOVA method described in this report is known as the fixed effects model, due to the assumption that the effects from the input to the
(a) Distribution for $v_A$ when $H_{0A}$ true.

(b) Distribution for $v_A$ when $H_{0A}$ false.

Figure 1: Distributions for $v_A$. 
output are deterministic and that the levels of the input are fixed. In addition to these presumptions we also have assumed that there is Gaussian distributed additive noise.

If these assumptions are not valid, there are other variants of the MANOVA method that can be useful, e.g., the random effects model or the mixed model. If the noise is not Gaussian distributed it is sometimes possible to find a way to transform data which leads to better results. Other problems that can arise are:

- too great complexity due to a large number of possible inputs,
- too low signal to noise ratio,
- too few repetitions of measurements to perform the full analysis,
- missing data,
- etc..

Most of these problems, and some others, are discussed by Montgomery [4] and/or Scheffé [5]. Hocking [2] has an approach to the method which clearly handles the problems with missing data.

8 Conclusions

We can conclude from experiments that the multivariate analysis of variance method is a good alternative to exhaustive search for identifying what input contribute to the output in non-linear finite impulse response (N-FIR) models.

The MANOVA method manages much better to identify the correct input signals (contributing time lags) and reduces the number of erroneous models from 19.4-24.4% to 0.1-9.7% when compared to the exhaustive search method. The main source of the differences is that the minimization algorithm gets stuck in local minima, due to the non-convexity of the identification problems. MANOVA is also computationally much more effective. The MANOVA method (with estimation of a model) is at least 14 times faster in finding appropriate input time lags and a corresponding model. When the complexity increases, due to more possible input time lags, it is an even greater advantage to use the MANOVA method.

References


