# Strathprints Institutional Repository

**Alkaya, Alkan and Grimble, Michael John (2015) Non-linear minimum variance estimation for fault detection systems. Transactions of the Institute of Measurement and Control, 37 (6). pp. 805-812. ISSN 0142-3312 , http://dx.doi.org/10.1177/0142331214548304**

# Nonlinear Minimum Variance Estimation
# For Fault Detection Systems


## Abstract

A novel model-based algorithm for fault detection (FD) in stochastic non-linear systems is proposed. The Nonlinear Minimum Variance (NMV) estimation technique is used to generate a residual signal which is then used to detect actuator and sensor faults in the system. The main advantage of the approach is the simplicity of the nonlinear estimator theory and the straightforward structure of the resulting solution. Simulation examples are presented to illustrate the design procedure and the type of results obtained. The results demonstrate that both actuator and sensor faults can be detected successfully.


## 1. Introduction

The need for high performance, efficiency, safety and reliability in modern engineering systems has focussed interest in the *Fault Detection and Isolation* (*FDI*) problem. A fault is defined as an unexpected change in a system with component malfunction or variation in operating condition. Some faults, if not promptly and properly detected, could turn into unrecoverable failures, causing serious damage and even loss of human lives [1].

In the literature faults can be assume to take place in different parts of a system, and are classified as actuator faults or sensor faults [2]. Actuator faults can represent partial or complete loss of control action. A total actuator fault can occur as a result of a breakage, cut or burned wiring, short-circuit or the presence of foreign body in the actuator [2]. Sensor faults in incorrect outputs from the sensors. They can also be subdivided into partial and total faults.

*Fault Detection* (*FD*) methods can be classified into two major categories; model-based and data-driven approaches [3]. The model-based *Fault Detection Isolation* (*FDI*) approaches include parity space, parameter estimation and observer based approaches. The observer-based FDI method is one of the most effective and has received significant interest from industry [4]. Model based approaches typically rely on two steps: residual generation; the procedure of extracting fault symptoms from the process, and residual evaluation; the procedure of decision making [5]. The residuals are often generated using either an observer; for deterministic models, or an optimal filter for stochastic models.

Observer based FD methods use measurements of the actual signals and estimates of the signals to generate the residual. The residual should be defined to become large when a fault occurs, to avoid false alarms [6], but remain as small as possible due to other uncertainties such as unknown disturbances and modelling errors.

Residual generation approaches have been developed successfully for linear systems. However, much less work has been done for nonlinear systems. This is primarily due to the complexity of nonlinear systems. The area of *FDI* for nonlinear systems is not covered completely yet, so it is worthy of study [7].

There is some existing literature on the use of a nonlinear estimator for fault detection and isolation. The most popular estimator for nonlinear processes is known to be *extended Kalman filter* (*EKF*) [8]. Although widely used, *EKFs* have some deficiencies, including the requirement of differentiability of the state dynamics as well as susceptibility to bias and divergence in the state estimates. The *unscented Kalman filter* (*UKF*), on the contrary, uses the nonlinear model directly instead of linearizing it [9] and hence does not need to calculate the Jacobian and can achieve higher order accuracy. *Particle filters* (*PF*) or Sequential Monte Carlo Methods are considered a general numerical tool to approximate the a posteriori density in nonlinear and non-Gaussian filtering problems. The main drawback with the particle filter is that it is very demanding computationally[10].

In this study, the *Nonlinear Minimum Variance* (*NMV*) estimator is used for the first time to generate a residual signal for fault detection applications. The strong point of this technique is that a general nonlinear operator is used to represent the nonlinearity of the channel or of the measurement sensor. This might involve a set of nonlinear equations or even include a look-up table or be a model obtained from a neural or fuzzy-neural network. The main advantages of proposed estimator is that no on-line linearization is required, as in the extended Kalman filter, and implementation is easy. The cost-function to be minimized is the variance of the estimation error and a relatively simple optimization procedure and solution results [11].

The roadmap for this study is as follows. The derivation of *NMV* estimation method is given in section 2. *NMV* based residual generation for fault detection is described in section 3. The performance of the proposed fault detection method is illustrated by a case study in section 4. Finally the conclusions are summarised in section 5.

## 2. Nonlinear Minimum Variance Estimation

The theory of *Nonlinear Minimum Variance Estimation* (*NMVE*) was intrıduced by Grimble [11] using polynomial system models [12, 13] and later state-equation based models [14, 15]. The *NMVE* technique involves the estimation of a signal that passes through a communications channel having nonlinearities and communication/transport delays [13]. The measurements are assumed to be corrupted by a noise signal, which is correlated with the signal to be estimated. Signal and noise models are assumed to be linear and time-invariant. The *Nonlinear Minimum Variance* (*NMV*) estimator derivation is based on the minimization of the error variance criterion. Consider the system shown in Fig.1, which includes the nonlinear signal channel model and linear measurement noise and signal models.

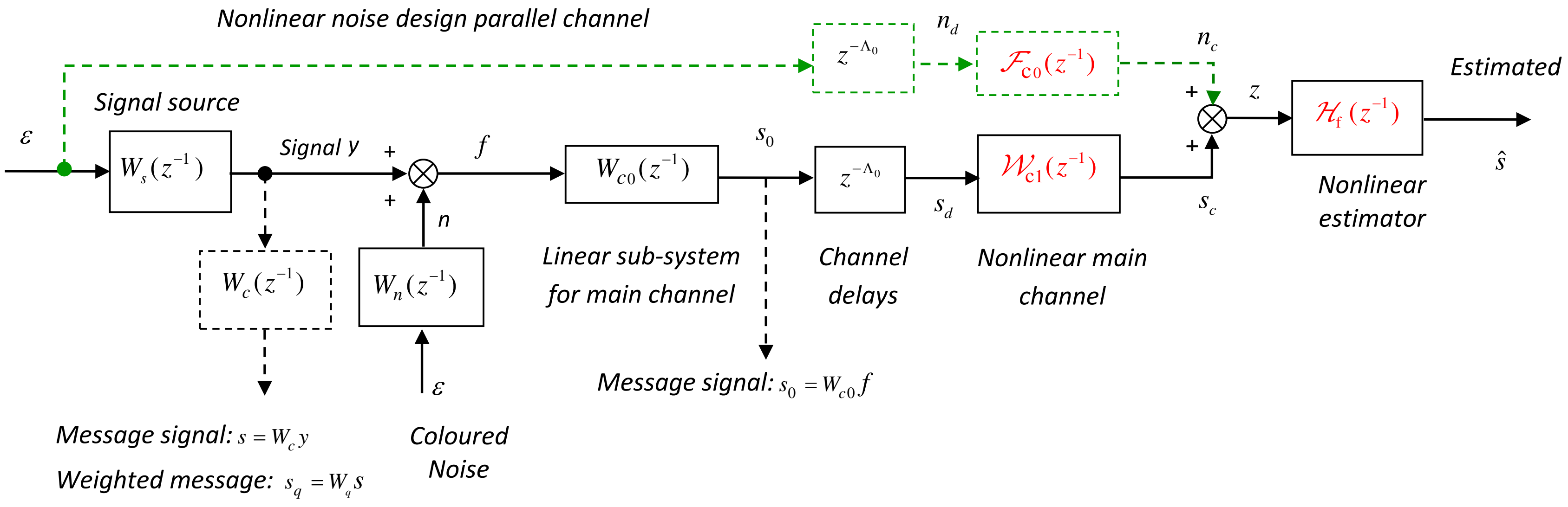


Figure 1: **Signal and Noise Model and Communication Channel Dynamics**

The signal channel model includes the nonlinearities that may involve both linear and nonlinear dynamics. The signal channel dynamics with a delay can be expressed as:

$$\left(\mathcal{W}_{channel} f\right)(t) = \left(\mathcal{W}_{c1} z^{-\Lambda_0} W_{c0} f\right)(t) \tag{1}$$

where $z^{-\Lambda_0}$ denotes a diagonal matrix of the $k$ step delay elements in the signal paths and $\Lambda_0 = kI$. The parallel path dynamics shown in Fig. 1, by a dotted line, can be expressed as:

$$\mathcal{F}_{\mathrm{c}}(z^{-1}) = \mathcal{F}_{\mathrm{c}0}(z^{-1}) z^{-\Lambda_0} \tag{2}$$

This is a fictitious channel, added to provide design tuning options, that can be used to represent uncertainties in channel knowledge, which provides additional design freedom. The combined signal source and noise signal $f(t) \in R^r$ is given as:

$$f(t) = y(t) + n(t) \tag{3}$$

Consider the nonlinear system for the optimal estimation problem illustrated in Fig.1. The input and noise generating processes have an innovations signal model with white noise signal input: $\varepsilon(t) \in R^r$ and it may be assumed to be zero-mean with covariance matrix: $\mathrm{cov}[\varepsilon(t),\ \varepsilon(\tau)] = I\delta_{t\tau}$ where $\delta_{t\tau}$ denotes the *Kronecker delta-function*. The signals shown in the closed-loop system model of Fig.1 may be listed as:

Noise: $$n(t) = W_n \varepsilon(t) \tag{4}$$

Input signal: $$y(t) = W_s \varepsilon(t) \tag{5}$$

Channel input: $$f(t) = y(t) + n(t) \tag{6}$$

Linear channel subsystem: $$s_0(t) = \left(W_{\mathrm{c}0} f\right)(t) \tag{7}$$

Weighted channel interference: $$n_c(t) = (\mathcal{F}_c \varepsilon)(t) \tag{8}$$

Nonlinear channel subsystem: $$s_c(t) = \left(\mathcal{W}_{\mathrm{c}1} s_d\right)(t) \tag{9}$$

Nonlinear channel input: $$s_d(t) = z^{-\Lambda_0} s_0(t) = s_0(t-k) \tag{10}$$

Observations signal: $$z(t) = n_c(t) + s_c(t) \tag{11}$$

Message signal to be estimated: $$s(t) = W_{\mathrm{c}}\, y(t) = W_{\mathrm{c}} W_s \varepsilon(t) \tag{12}$$

Weighted message signal: $$s_q(t) = W_q W_{\mathrm{c}}\, y(t) \tag{13}$$

Estimation error signal: $$\tilde{s}(t \mid t-\ell) = s(t) - \hat{s}(t \mid t-\ell) \tag{14}$$

where $\hat{s}(t \mid t-\ell)$ denotes the estimate of the signal s($t$) at time $t$, given observations z($t$) up to time $t$-$\ell$. Value of $\ell$ may be positive or negative according to the following conditions: $\ell$=0, for estimation; $\ell > 0$, for prediction and $\ell < 0$, for fixed-lag smoothing. The criterion for the nonlinear minimum variance estimator is given below:

$$J = trace\{E\{W_q \tilde{s}(t \mid t-\ell)(W_q \tilde{s}(t \mid t-\ell))^T\}\} \tag{15}$$

where $E\{.\}$ denotes the expectation operator and $W_q$ [16] denotes a linear strictly minimum-phase dynamic cost-function weighting function matrix which is assumed to be strictly minimum phase, square and invertible.

The estimate $\hat{s}(t \mid t-\ell)$ is assumed to be generated from a nonlinear estimator of the form:

$$\hat{s}(t \mid t-\ell) = H_f(t,\ z^{-1}) z(t-\ell) \tag{16}$$

where

$$\mathcal{H}_f(t, z^{-1}) = W_q^{-1} H_0 ( \quad _{c0} \quad _{c1} W_{c0} Y_f )^{-1} \tag{17}$$

where $\mathcal{H}_f(t, z^{-1})$ denotes a minimal realisation of the optimal nonlinear estimator. Since an infinite-time $(t = -\infty)$ problem is of interest therefore no initial condition term is required. The block diagram representation of $\mathcal{H}_f(t, z^{-1})$ will be as shown in Fig. 2.

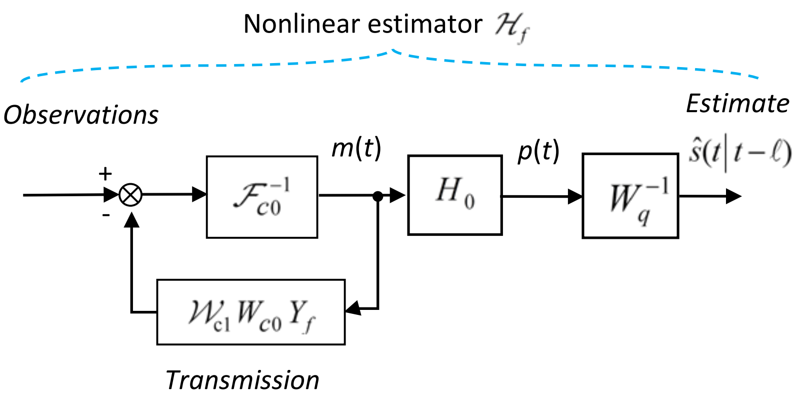


Figure 2: **Implementation of the Nonlinear Estimator**

The terms $H_0, A$ *and* $Y_f$ used in equation (18) can be calculated using the concept of power spectrum for the combined linear models using: $\phi_{ff} = (W_s \quad W_n)(W_s^* \quad W_n^*)$, and where the notation for the adjoint of $W_s$ implies: $W_s^*(z^{-1}) = W_s^T(z)$, and in this case the z denotes zdomain complex number. The generalized spectral-factor: $Y_f$ may be computed using: $Y_f Y_f^* = \phi_{ff}$ ,where $Y_f = A_0^{-1} D_{f0} = D_f A^{-1}$. The system models are assumed such that $D_{f0}$ is strictly Schur polinomial matrix [17, 18] satisfying:

$$D_{f0} D_{f0}^* = (C_s \quad C_n)(C_s^* \quad C_n^*) \tag{18}$$

The right-comprime polynomial matrix model can be defined as:

$$\begin{bmatrix} C_f & D_f \end{bmatrix} A^{-1} = \begin{bmatrix} W_q W_c W_s & Y_f \end{bmatrix} \tag{19}$$

The polynomial operators $H_0$ now may be optained from the minimal degree solution $(H_0, F_0)$, with repect to $F_0$, of the following Diophantine equation:

$$F_0 A + G_0 z^{-k-\ell} = C_f \tag{20}$$

The estimation error can be penalised in a particular frequency range by using a dynamic asymptotically stable weighting function $W_\Omega = A_\Omega^{-1} B_\Omega$, where $A_\Omega \; and \; B_\Omega$ are polynomial matrices. The weighted error involves a linear path at the optimum. In the linear case the modified cost function will have the form (Parceval's theorem does not apply in the nonlinear case):

$$J = trace\left\{ E(W_\Omega e(t|t-\ell))(W_\Omega e(t|t-\ell))^T \right\} = trace\left\{ 1/(2\pi j) \oint_{z=1} (W_\Omega \Phi_{ee} W_\Omega^*) dz/z \right\} \tag{21}$$

### 3. NMVE Based Fault Detection

In nonlinear minimum-variance estimation, the nonlinearities are assumed to be in the signal channel or possibly in a noise channel representing the uncertainty. The simple solution that follows arises because of the assumptions of linearity for the signal generating model and the results obtained here involve only a least-squares type of analysis [19].

The Fault detection techniques are often based on the generation of appropriate residual signals which have to be sensitive to faults themselves but independent of disturbances.

*Model-based FD* methods are based on comparing the behavior of the actual signal and an estimated signal of the system. Typically, it is shown that in the absence of a fault, the observer residual approaches zero. When a fault exists, this residual will be non-zero, and it may therefore serve as a fault indicator.

The block diagram of the proposed nonlinear minimum variance estimator, taking $\ell = 0$, based on residual generation for fault detection, is shown in Fig. 3.

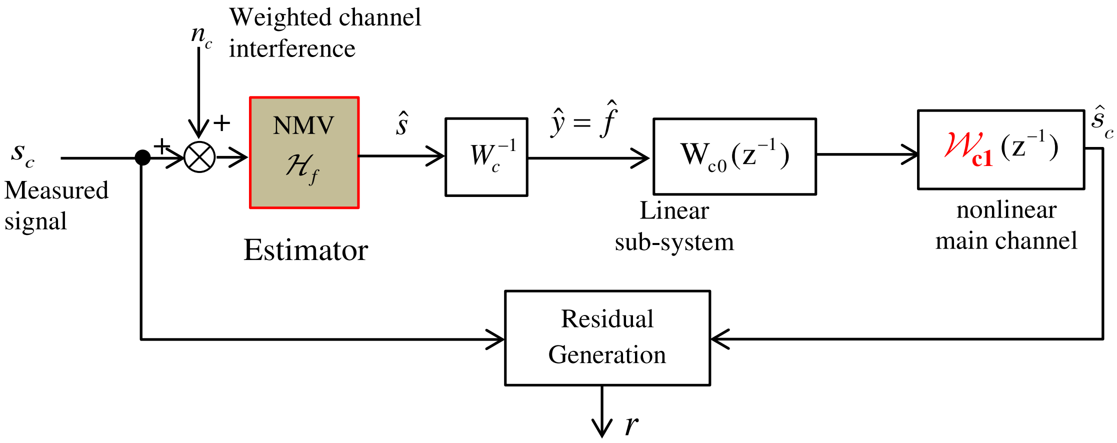


Fig. 3: **NMVE Based Residual Generation Scheme**

The residual signal can be generated by using measured signal $s_c(t)$ and its estimate $\hat{s}_c(t)$, as:

$$r(t) = s_c(t) - \hat{s}_c(t) \tag{22}$$

The *NMV* algorithm estimates the signal $\hat{s}$ so $\hat{s}_c$ might be defined in term of $\hat{s}$ signal by using eqn(6), eqn(7), eqn(9) and eqn(10) as follows

$$\hat{y}(t) = W_c^{-1}\hat{s}(t) \tag{23}$$

$$\hat{f}(t) = \hat{y}(t) = W_c^{-1}\hat{s}(t) \tag{24}$$

$$\hat{s}_c(t) = \mathcal{W}_{c1} W_{c0} \hat{f}(t) \tag{25}$$

Then finally residual signal can be calculated substituting eqn(28) into eqn (23):

$$r(t) = s_c(t) - \mathcal{W}_{c1} W_{c0} W_c^{-1}\hat{s}(t) \tag{26}$$

This residual signal $r$ is going to be checked with a reasonable threshold to detect that a fault has occurred in the system.

When there is a fault at the signal estimation point, the residual becomes

$$r(t) = s_c(t) \quad W_{c1}W_{c0}W_c^{-1}\hat{s}(t) \quad \phi_f \tag{27}$$

$$= W_{c1}W_{c0}W_c^{-1}s(t) - W_{c1}W_{c0}W_c^{-1}\hat{s}(t) + \phi_f \tag{28}$$

If the plant is linear this simplifies as:

$$r(t) = W_{c1}W_{c0}W_c^{-1}\left(s(t) - \hat{s}(t)\right) + \phi_f \tag{29}$$

$$= W_{c1}W_{c0}W_c^{-1}\left(\tilde{s}(t \mid t)\right) \quad \phi_f \tag{30}$$

Where $\phi_f$ is a fault and where $\phi_f \neq 0$ is the output arising from the signal fault. However, it can be only detected if term is large compared with estimation errors and the signal noise $\varepsilon(t)$.

### 3.1. Threshold computation

To achieve a successful fault detection based on the available residual signal, further effort is needed. Residual evaluation and threshold setting are used to distinguish the faults from the disturbances and uncertainties. A decision on the possible occurrence of a fault will then be made by means of a simple comparison between the residual feature and the threshold, as shown in Fig. 4.

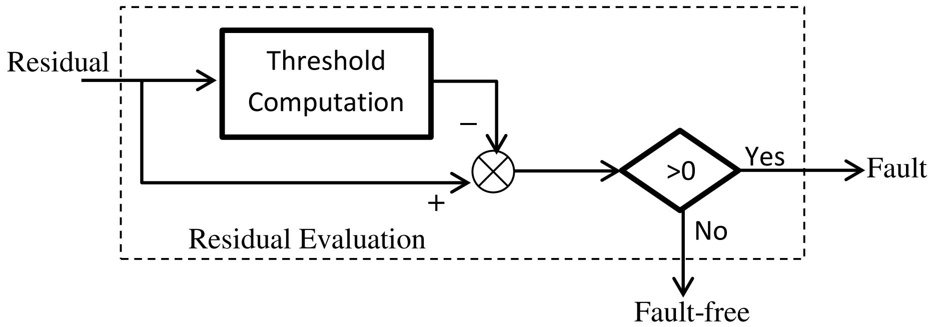


Fig. 4: **Residual evaluation**

In practice, the so-called limit monitoring and trend analysis are, due to their simplicity, widely used for the purpose of fault detection. For a given signal r, the primary form of limit monitoring is

$$\textit{if} \quad r < T_{min} \ \textit{or} \ r > T_{max} \quad \textit{then, Alarm, fault is detected}$$
$$\textit{if} \quad T_{min} \leq r \leq T_{max} \quad \textit{then, No Alarm, fault-free}$$

where $T_{min}$, $T_{max}$ denote the minimum and maximum values of T in the fault-free case. They are the threshold values.

## 4. Design and Simulation Results

The computation of the estimator is relatively straightforward. The polynomial matrix equations can be solved using the Matlab polynomial toolbox PolyX. Given these matrices the estimator may be implemented very neatly, as shown in Fig. 2.

The selection of the uncertainty tuning function $\mathcal{F}_{c0}$ is a dual problem to the selection of optimal control cost function weightings [16]. The requirement for the nonlinear operator is that it should have a stable inverse. A simple starting point is therefore to assume the uncertainty model $\mathcal{F}_{c0}$ is a constant and of a small magnitude. This corresponds to the situation where the uncertainty is simply white noise added at the output of the communications channel before it enters the estimator. Uncertainly is of course often associated with high frequency behavior and hence a simple linear lead term might be used to represent the frequency response of as in the example which follows.

To validate the effectiveness of the NMV filter based fault detection systems, nonlinear SISO system is used as an example. The NMV filter is computed below for the example and a simulation is used to verify the results.We consider a system having the following signal and noise models;

$$W_s = \frac{0.1}{1-0.99z^{-1}}, \qquad W_n = \frac{0.6}{1-0.1z^{-1}}$$

and let weighting $W_q = 1$, $W_c = 0.5/1-0.5z^{-1}$ and Channel delay = $z^{-1}$,so that $\Lambda_0 = k = 1$. The linear channel characteristics are defined as $W_{c0} = 1/(1-0.5z^{-1})$. The static nonlinear characteristic of the system is given in Fig. 5.

The dc-gain and changes in the cut-off frequency of the weighting filter $\mathcal{F}_{c0}$ influences the accuracy of estimation. $\mathcal{F}_{c0}^{-1}$ The tuning function , which is optimized for this example, has the following representation:

$$\mathcal{F}_{c0}^{-1} = \frac{1-0.4z^{-1}}{1-0.1z^{-1}}$$

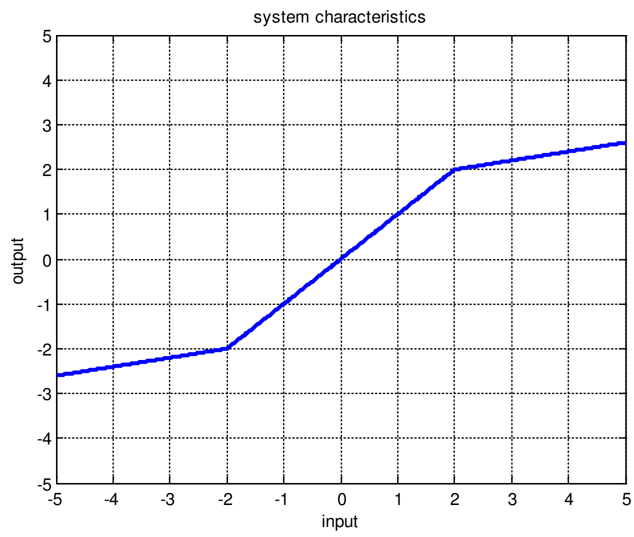


Fig. 5. **Nonlinear Behavior of the Output Sub-System**

The overall system and simulink model of NMV filter for fault detection is as shown in Fig. 6.

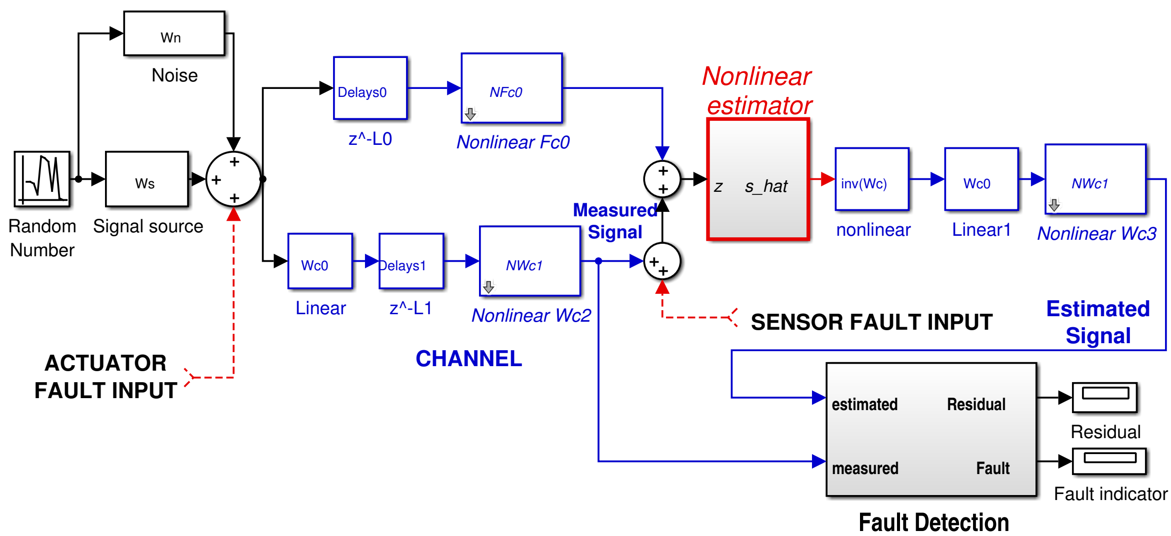


Fig.6: **Simulink Model of NMV Based Fault Detection Systems**

Under normal operation condition (fault-free) measured signal and estimated signal are illustrated in Fig.7. The minimum variance for the NMV estimator is 1.14e-02. Tuning filter response is shown in Fig.8. Calculated residiual signal and confidence level threshold are dedicated in Fig.9. As shown in Fig. 9, residual signal is under the threshold. It means system is under normal operation.

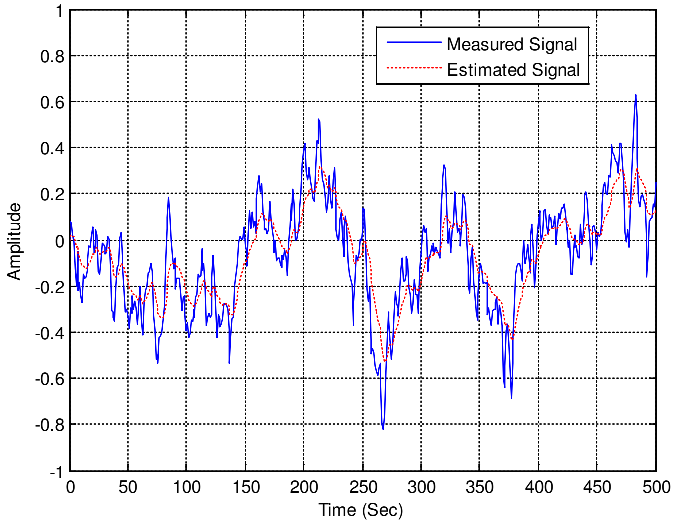


Fig.7. **Measured and Estimated Signal (no fault)**

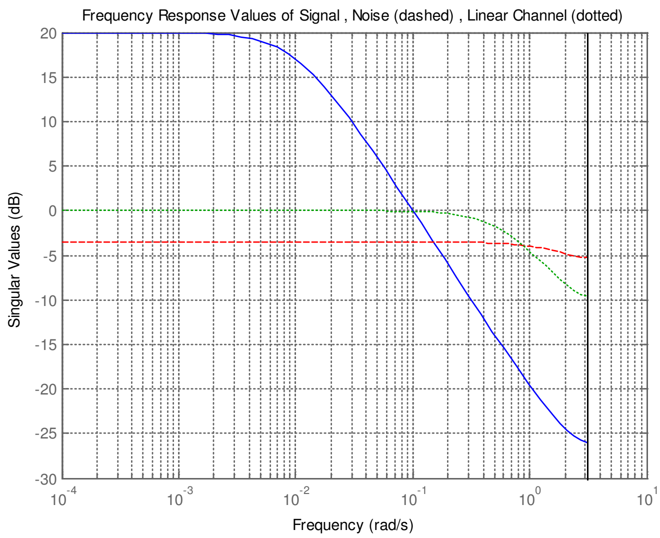


Fig.8. **Tuning Filter Frequency Responses**

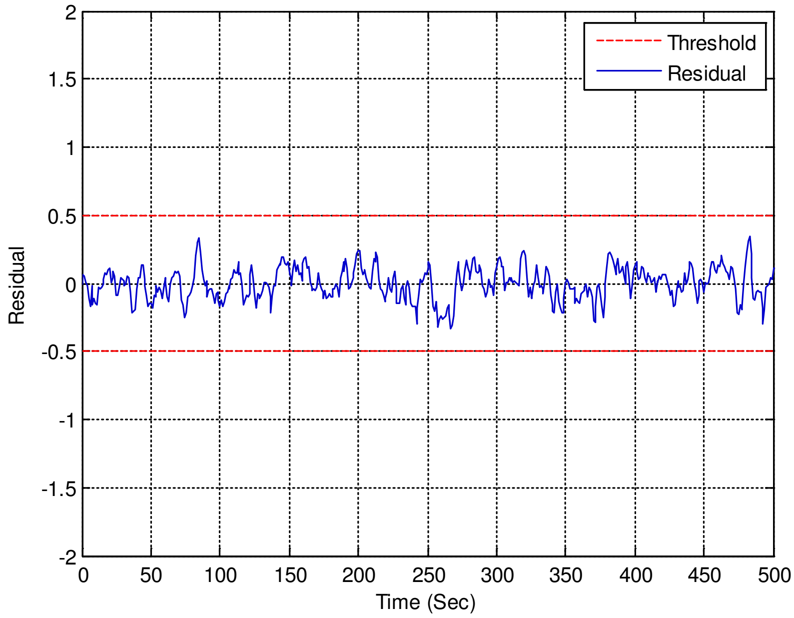


Fig.9. **Residual Signal with Thresholds (no fault)**

Two type of faults are applied to validate the effective of the proposed *NMV* estimator in fault detection implementation.

### 4.1. Sensor Fault

For the sensor fault; the signal shown in Fig. 10 is applied to the 'sensor fault input' of the system as illustrated in Fig. 6 simulink model. The fault is considered as a drift on the measurement sensor. After applied sensor fault, actual signal and estimated signal are illustrated in Fig. 11. Calculated residual signal and confidence level threshold are dedicated in Fig. 12. Fault has been detected successfully with accurate time as shown in Fig. 12.

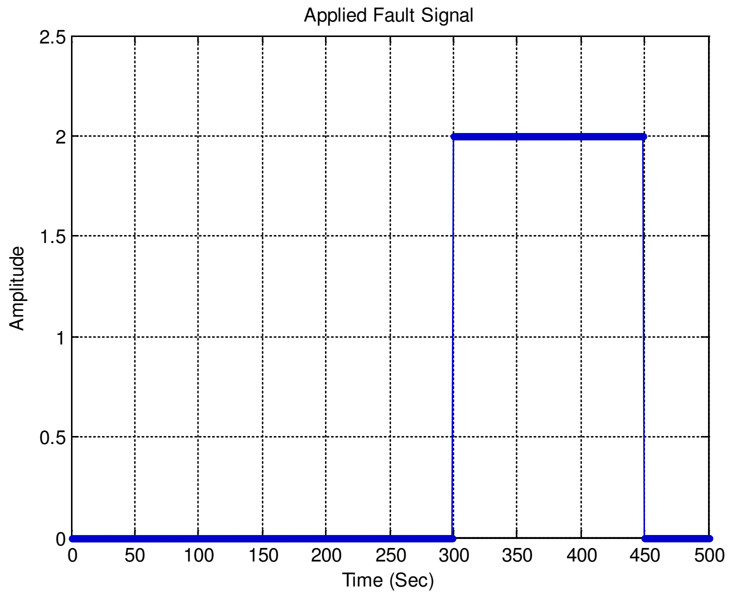


Fig.10. **Applied Fault Signal**

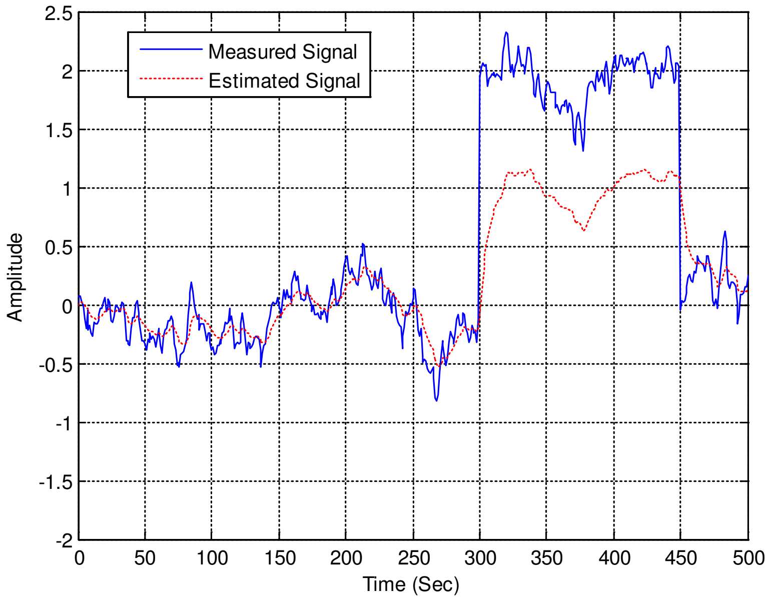


Fig.11. **Actual and Estimated Signal (faulty)**

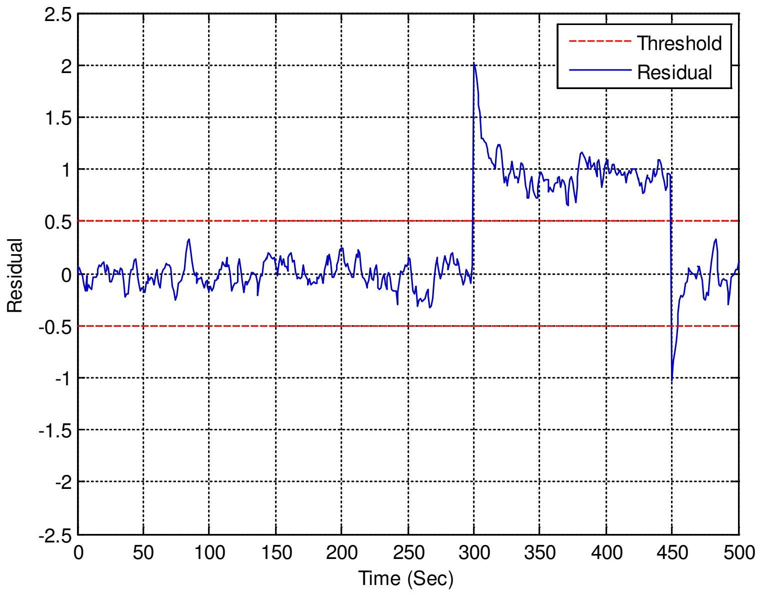


Fig.12. **Residual Signal with Thresholds (faulty)**

### 4.2. Actuator Fault

For the actuator fault; the signal shown in Fig. 13 is applied to the 'actuator fault input' of the system as illustrated in Fig. 6 simulink model. The fault is considered as a lost contact of the actuator input for a while. After the actuator fault is applied, actual signal and estimated signal are as illustrated in Fig. 14. Calculated residual signal and confidence level threshold are dedicated in Fig. 15. Fault has been detected successfully with accurate time as shown in Fig. 15.

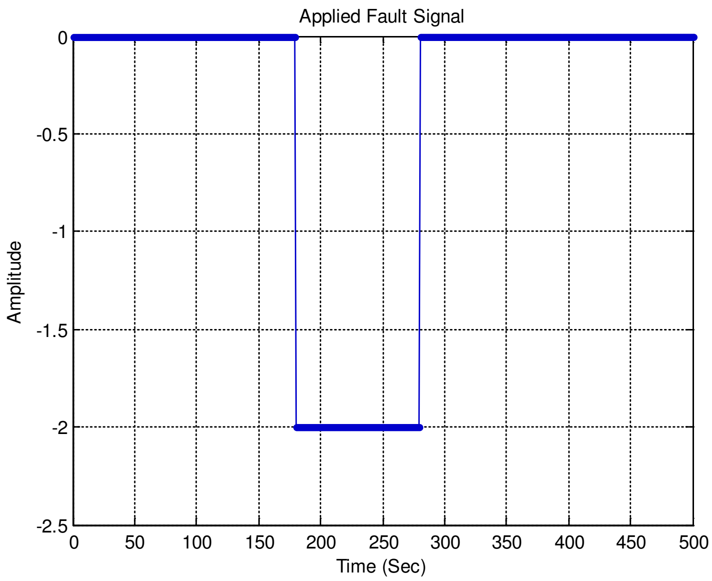


Fig.13. **Applied Fault Signal**

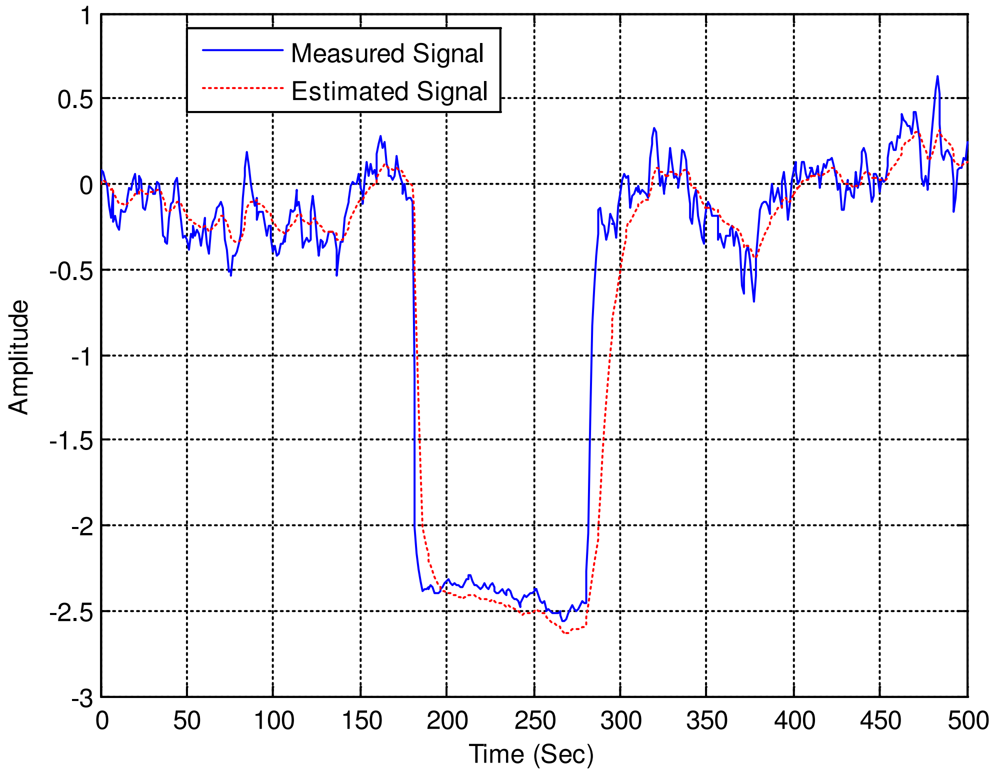


Fig.14. **Actual and Estimated Signal (faulty)**

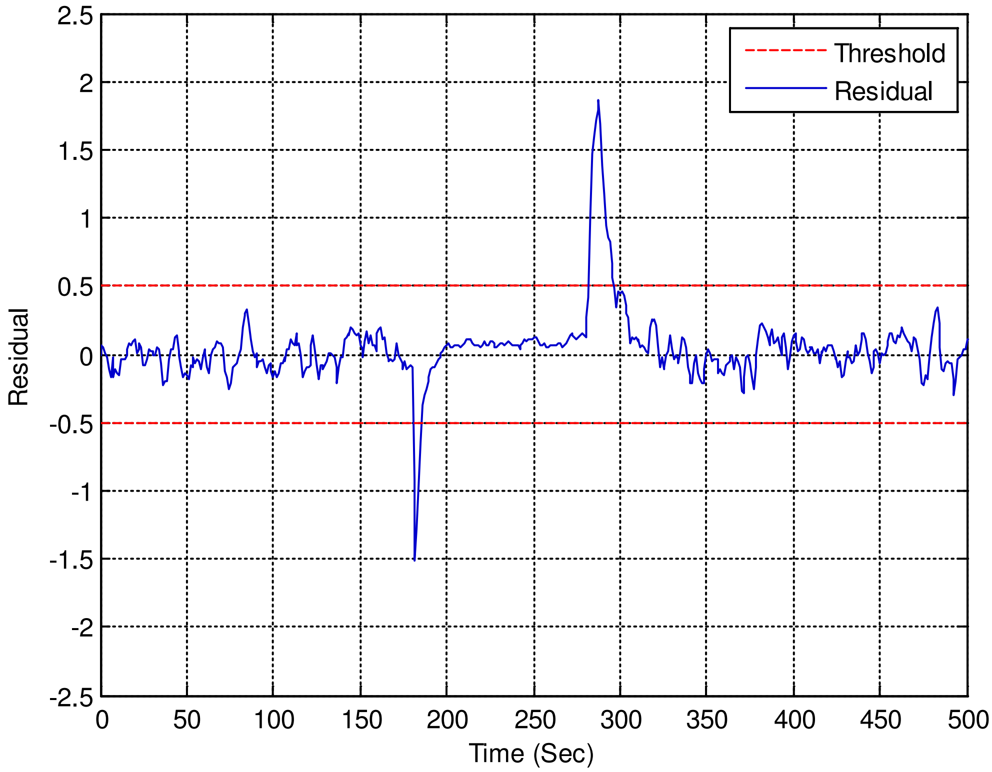


Fig.15. **Residual Signal with Thresholds (faulty)**

## 5. CONCLUSIONS

A *NMV* estimator based fault detection system for nonlinear systems has been developed. The *NMV* estimator is used to generate the residual signal which indicates possible fault conditions in the system. The *NMV* estimator has some benefits relative to some other nonlinear estimators in three respects i.e. it requires less computational cost, easy to implement and to tune. The algorithm is illustrated using the simulation of a nonlinear process control example. The simulation results show that the method has a good performance in detecting faults at either inputs or outputs.

## 6. REFERENCES

[1] Alkaya A, Eker İ. Variance sensitive adaptive threshold-based PCA method for fault detection with experimental application. ISA Transaction ISA Transactions 50 (2011) 287–302.

[2] Isermann R. Fault-diagnosis systems: an introduction from fault detection to fault tolerance. Berlin: Springer; 2006.

[3] Venkatasubramanian VR, Rengaswamy KY, Kavuri SN. A review of process fault detection and diagnosis, part I: quantitative model—based methods. Computers & Chemical Engineering 2003;27:293–311.

[4] Venkatasubramanian VR, Rengaswamy KY, Kavuri SN. A review of process fault detection and diagnosis, part II: qualitative models and search strategies. Computers & Chemical Engineering 2003;2:313–326.

[5] Chow, E. Y., and Willsky, A. S., 1984. Analytical redundancy and the design of robust failure detection systems. IEEE Transactions on Automatic Control, vol. 29(7):603–614.

[6] Hur, S.H., Katebi, R., Taylor, A., Model-based fault monitoring of a plastic film extrusion process. IET Control Theory Appl., 2011, Vol. 5, Iss. 18, pp. 2075–2088.

[7] Alrowaie, F., Gopaluni, R.B., Kwok, K. E., Fault detection and isolation in stochastic non-linear state-space models using particle filters. Control Engineering Practice, 20, (2012), 1016–1032

[8] Gerasimos G. R., A Derivative-Free Kalman Filtering Approach to State Estimation-Based Control of Nonlinear Systems. IEEE Transactıons On Industrial Electronics, Vol. 59, No. 10, 2012, 3987-3997.

[9] Mirzaee, A., Salahshoor, K., Fault diagnosis and accommodation of nonlinear systems based on multiple-model adaptive unscented Kalman filter and switched MPC and H-infinity loop-shaping controller. Journal of Process Control 22 (2012) 626–634.

[10] Shamsher, A. N., Linear and nonlinear polynomial based estimators, PhD thesis, Industrial Control center, Strathclyde University,2009, p(163).

[11] Grimble, M.J.,NMV optimal estimation for nonlinear discrete-time multi-channel systems, 46th IEEE Conf. on Decision and Control,New Orleans, 12–14 December 2007, pp. 4281–4286.

[12] Grimble, M.J.,Time-varying polynomial systems approach to multichannel optimal linear filtering’. Int. Conf. of Acoustics, Speech and Signal Processing, Detroit, MI, 9–12 May 1995, vol. 2, pp. 1500–1503.

[13] Grimble, M.J., Robust industrial control systems: optimal design approach for polynomial systems, John Wiley, Chichester, 2006.

[14] Grimble, M.J, 2011, 'Nonlinear Minimum Variance State Based Estimation for Discrete-Time Multi-Channel Systems', IET Journal on Signal Processing, Volume 5, Issue 4, July, Volume: 5 , Issue 4, Page(s): 365 – 378. Digital Object Identifier: 10.1049/iet-spr.2009.0064.

[15] Grimble M J, 2012, 'Nonlinear Minimum Variance Estimation for State-Dependent Discrete-Time Systems', IET Journal on Signal Processing, Volume 6, Issue 4, June, p. 379 – 391, DOI: 10.1049/iet-spr.2010.0220, Print ISSN 1751-9675, Online ISSN 1751-9683.

[16] Grimble M. J., Nonlinear generalised minimum variance feedback, feedforward and tracking control. Automatica, 2005, 41, 957-969.

[17] Kucera V., Discrete Linear Control, Wiley, Chichester,1979.

[18] Kucera V., Stochastic multivariable control:a polynomial equation approach, IEEE Trans.,1980,AC-25,(5),pp.913-919.

[19] Grimble M.J., Shamsher A. N., Optimal minimum variance estimation for non-linear discrete-time multichannel systems. IET Journal on Signal Processing, 2010, Vol. 4, Iss. 6, pp. 618–629. ISSN: 1751-9675, INSPEC Accession Number: 11674705.