Preprints of the
9th International Symposium on Advanced Control of Chemical Processes
The International Federation of Automatic Control
June 7-10, 2015, Whistler, British Columbia, Canada

Stochastic Fault Diagnosis using a Generalized Polynomial Chaos Model and Maximum Likelihood

Yuncheng Du*, Thomas A. Duever*[+], Hector Budman*

**Department of Chemical Engineering, University of Waterloo, Waterloo, ON, Canada, N2L 3G1*
(E-mail: hbudman@uwaterloo.ca)
[+]Department of Chemical Engineering, Ryerson University, Toronto, ON, Canada, M5B 2K3
(E-mail: tom.duever@ryerson.ca)

Abstract: A novel approach has been developed to diagnose intermittent stochastic faults by combining a generalized polynomial chaos (gPC) method with maximum likelihood estimation. The gPC is used to propagate stochastic changes in an input variable to a measured output variable from which the fault is to be inferred. The fault detection and diagnosis (FDD) problem is formulated as an inverse problem of identifying the unknown input from a maximum likelihood based fitting of the predicted and measured output variables. Simulation studies compare the proposed method with a Particle Filter (PF) to estimate the value of an unknown feed mass fraction of a chemical process. The proposed method is shown to be significantly more computational efficient and less sensitive to user defined tuning parameters than PF.

***Keywords**:* Fault identification, Uncertainty analysis, Nonlinear model, Particle filter, Estimation theory

1. INTRODUCTION

This paper presents a new method for diagnosing unmeasured stochastic intermittent faults for a nonlinear chemical plant. The proposed approach is based on modelling by generalized polynomial chaos (gPC) expansions and maximum likelihood based estimation. In a previous contribution by the authors it was shown how the gPC models can be used to detect steady state changes in an input variable (fault) (Du, et al., 2014). The novel contribution in this work is a methodology for diagnosing time varying stochastic faults while maintaining the computational complexity at a reasonable level. Unlike many traditional model based methods such as Kalman Filter, the proposed approach explicitly considers the nonlinear behaviour of the process, the stochastic nature of the input faults and their effects on the measured output variables.

The benefit of using the gPC models in parameter estimation problems has been reported (Chen-Charpentier & Stanescu, 2014; Pence, et al., 2011), but there are no reports of gPC in dynamic fault detection problems as proposed in the current study. The available model based alternative to solve this problem is to apply a particle filter (PF). It will be shown that due to the analytical properties of the gPC expansions, the proposed approach is significantly more computationally efficient than the PF, thus making it more suitable for real time implementation in problems of large dimensions (Nagy & Braatz, 2007; Elsheikh, et al., 2014). Also, the proposed algorithm is less sensitive than PF to the user selected tuning parameters.

Assuming a nonlinear system described as follows:

$$\dot{\boldsymbol{x}} = f(t, \boldsymbol{x}, \boldsymbol{u}; \boldsymbol{g}) \tag{1}$$

, where $0 \le t \le t_f$, $\boldsymbol{x} \in R^n$ contains the system states (measured variables) with initial conditions $\boldsymbol{x}(0) = \boldsymbol{x}_0$ over time domain $[0, t_f]$. $\boldsymbol{u}$ denotes the known inputs of the system, and $\boldsymbol{g} \in R^{n_g}$ defines an unknown time varying input that is considered as a fault in this work.

To identify an input fault $\boldsymbol{g}$, a *two-level* fault detection and diagnosis (FDD) algorithm is proposed, for which the occurrence of a major deviation from a specific operating mode is detected in *level-1* of the algorithm by using the probability density functions (PDFs) of measured variables calculated off-line for constant mean values of $\boldsymbol{g}$, and then, in *level-2* of the algorithm, dynamic changes in $\boldsymbol{g}$ are inferred by fitting the measured and gPC model predicted variables based on a maximum likelihood cost. The *level-2* algorithm is only executed when large deviations are detected in the *level-1 algorithm* to reduce the computational burden.

This paper is organized as follows. In section 2, the theoretical background of gPC expansions is presented. The maximum likelihood based fault detection approach is explained in section 3. A nonlinear chemical plant with two continuously stirred tank reactors and a flash tank separator is introduced as a case study in section 4. Analysis and discussion of the results are given in section 5 followed by conclusions in section 6.

2. GENERALIZED POLYNOMIAL CHAOS THEORY

The generalized polynomial chaos (gPC) expansion has been proposed as a modelling method for stochastic systems (Xiu, 2010), which is based on the representation of an arbitrary random variable of interest as a function of a random variable ξ of a known prior distribution. To preserve orthogonality, the basis functions, *e.g.,* Laguerre or Hermite, have to be selected according to the choice of the distribution of ξ (Xiu, 2010).

To quantify the effect of the stochastic input $\boldsymbol{g}$ in (1) on the measured variables $\boldsymbol{x}$, the gPC expansions are employed. According to the gPC theory and the prior information, each

unknown parameter g_i (i = 1, 2,..., n_g) in $\boldsymbol{g}$ is assumed to be a function of a set of random variables $\boldsymbol{\xi} = \{\xi_i\}$:

$$g_i = g_i(\xi_i) \tag{2}$$

,where ξ_i is the i^{th} random variable. The random variables $\{\xi_i\}$ are assumed to be independent and identically distributed. It should be noted that although ξ_i is assumed to follow a standard distribution, the elements in $\boldsymbol{g}(\boldsymbol{\xi})$ may follow a non-standard distribution. The unknown $\boldsymbol{g}(\boldsymbol{\xi})$ and states $\boldsymbol{x}(t,\boldsymbol{\xi})$ in (1) can be approximated in terms of polynomial orthogonal basis functions $\Phi_k(\boldsymbol{\xi})$ as:

$$g(\xi) = \sum_{k=0}^{\infty} g_k \Phi_k(\xi) \tag{3}$$

$$x(t,\xi) = \sum_{k=0}^{\infty} x_k(t) \Phi_k(\xi) \tag{4}$$

, where x_k is the gPC coefficients of measured variables at each time instant t, and $\Phi_k(\boldsymbol{\xi})$ is a multi-dimensional orthogonal polynomial. Since it is assumed in this work that the statistical distribution of the input $\boldsymbol{g}$ is known priori, *i.e.,* the gPC coefficients in (3) are available, the only unknowns are the gPC coefficients x_k of the measured state variables in (4). These coefficients can be estimated with Galerkin projection by projecting equation (1) onto each polynomial chaos basis function $\{\Phi_k(\boldsymbol{\xi})\}$ as (Xiu, 2010):

$$\left\langle \dot{x}(t,\xi), \Phi_k(\xi) \right\rangle = \left\langle f(t, x(t,\xi), u(t), g(\xi)), \Phi_k(\xi) \right\rangle \tag{5}$$

For practical application, (3) and (4) are truncated to a finite number of terms, *i.e., T_N*. Hence, the total number of terms (T_N) in (5) is a function of the number of terms p in (3) that is necessary to represent the a priori known distribution of $\boldsymbol{g}$ and the number of different inputs $\boldsymbol{g}$ (n_g) as follows:

$$T_N = ((n_g + p)!/(n_g!\,p!)) - 1 \tag{6}$$

The inner product in (5) between two vectors $\varphi(\xi)$ and $\varphi'(\xi)$ is defined by:

$$\left\langle \varphi(\xi), \varphi'(\xi) \right\rangle = \int \varphi(\xi)\varphi'(\xi) w(\xi) d\xi \tag{7}$$

, where the integration is conducted over the entire domain of the random variables ξ, and $w(\xi)$is a weighting function, which is chosen for normalization purposes with respect to the type of polynomial basis functions used in the expansion (Xiu, 2010). Statistical moments for the measured variables $\boldsymbol{x}$ represented by the gPCs can be efficiently calculated at any given time instant t as a function of the coefficients x_k of the expansion in (4) as follows:

$$\mathrm{E}(x) = \mathrm{E}[\sum_{i=0}^{p} x_i \Phi_i] = x_0 \mathrm{E}[\Phi_0] + \sum_{i=0}^{p} \mathrm{E}(\Phi_i) = x_0 \tag{8}$$

$$\begin{aligned} \mathrm{var}(x) &= \mathrm{E}[(x - \mathrm{E}(x))^2] = \mathrm{E}[(\sum_{i=1}^{P} x_i \Phi_i - x_0)^2] \\ &= \mathrm{E}[(\sum_{i=1}^{P} x_i \Phi_i)^2] = \mathrm{E}[x_i^2 \mathrm{E}(\Phi_i^2)] \end{aligned} \tag{9}$$

Also, the PDF profiles for the measured variables, $\boldsymbol{x(t)}$, can be easily estimated by sampling from the distribution of $\boldsymbol{\xi}$ and substituting the samples into the expansion in (4). The ability of analytically calculating the mean and variance as per equation (8) and (9) and to rapidly calculate the PDFs of the measured variables are the main rationale for using the gPC since these quantities have to be repeatedly estimated in the algorithms proposed in this work. Thus, the use of gPC to propagate the stochastic variability of inputs onto measured outputs (states) results in significant computational savings.

3. FAULT DIAGNOSIS METHODOLOGY

3.1 Unknown input FDD problem formulation

The unknown input faults $\boldsymbol{g}$ considered in the current work affecting the system described by (1) consist of stochastic perturbations around a specific set of means as described in Fig.1 (a), and given as follows:

$$g = \bar{g}_i + \Delta g_i \;\; (i = 1, \ldots, k) \tag{10}$$

, where $\bar{g}_i$ is a set of constant mean values (operating modes), Δg_i are stochastic variations around each mean value. The statistical distribution of Δg_i is assumed to be known a priori. The changes in the mean values of $\bar{g}_i$ follow a ML-PRS (Multilevel Pseudo Random Signal). The FDD problem is defined as detecting a change in the unknown input mean $\bar{g}_i$ as well as diagnosing around which particular $\bar{g}_i$ the system is being operated. Each particular mean $\bar{g}_i$ will be referred heretofore as to an operating mode, and thus the goal is to diagnose the mean value $\bar{g}_i$ at any given time instant.

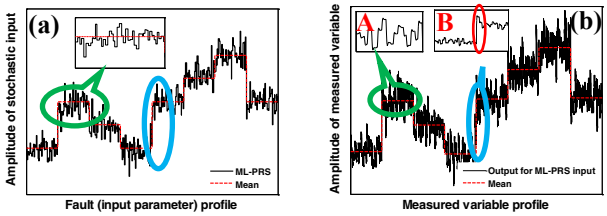

Fig. 1 Fault profiles representing intermittent stochastic fault and resulting measured variable

The FDD algorithm in this work is formulated as a *two-level* procedure composed of:

Level-1 algorithm – For each mean value of $\bar{g}_i$ given in (10) the corresponding PDFs of the measured output variables ($\boldsymbol{x}$) are approximated assuming that the mean remains constant. The PDF profiles of the measured variables are calibrated with simulated noisy measurements. When operating in the neighbourhood of a specific mean $\bar{g}_i$, the input and measured variables are not constant due to the stochastic variations (Δg_i) and the noise on measurements, as seen in Fig.1 (a) and (b). A change in $\bar{g}_i$ is detected based on measurements and the calculated PDFs as further explained below.

Level-2 algorithm – A maximum likelihood based FDD algorithm is applied to detect the average of the unknown input variable $\boldsymbol{g}$ over a pre-defined window of time. The likelihood function is based on the error in mean and variance between a set of measurements and predictions calculated with a gPC model in (4). To reduce computational effort, this step is only launched when large deviations from an input mean are detected with the *level-1 algorithm* as explained

above. However, in principle this second level algorithm *(level-2 algorithm)* can be executed at each time interval as an online dynamic parameter estimator but at the cost of increased computation. Details on the two algorithms are given below.

3.2 Level-1: generation of PDFs of measured variables when the system is operated around a particular mean ($\bar{g}_i$)

For the purpose of calculating the PDFs profile, it is assumed that measurements of the measured variables ($\boldsymbol{x}$) around each mean value $\bar{g}_i$ are available. It is also assumed in this step that $\bar{g}_i$ remains constant but its exact value is not known. The constancy of $\bar{g}_i$ can be experimentally inferred from the constancy of measured and/or controlled variables.

In the absence of measurement noise, if the inputs would be known the PDF profiles of output variables ($\boldsymbol{x}$), that can be measured and used for detection such as manipulated and controlled variables of the closed-loop system, could be calculated with the analytical expressions of a gPC as per the method shown in Section 2. In reality, due to noise, model error (*e.g.,* gPC truncation error) and lack of exact knowledge about the input, *i.e.,* it is only assumed that the mean is constant but its exact value is not available, the PDFs of $\boldsymbol{x}$ have to be calibrated using actual process measurements. To this purpose, the mean and variance of the unknown input variable are found from an optimization problem around each mean (operating mode) of the input shown in Fig.1 (a) as:

$$\min_{\lambda_{lower}} J = \sum_{i=1}^{n} (\vartheta_{1,i} - \upsilon_{1,i})^2 + \sum_{i=1}^{n} (\vartheta_{2,i} - \upsilon_{2,i})^2 + \sum_{i=1}^{n} \sigma_{n,i}^2 \quad (11)$$

, where $\vartheta_{1,i}$ and $\vartheta_{2,i}$ are the predicted mean and variance of a particular measured variable ($\boldsymbol{x}$) of the problem to be used for detection. These predicted means and variances are given explicitly by (8) and (9) using the gPCs representations of $\boldsymbol{x}$ given in Section 2, and are functions of the stochastic input as shown in Fig.1 (a). The terms $\upsilon_{1,i}$ and $\upsilon_{2,i}$ are the measured mean and variance of $\boldsymbol{x}$ in (1). The last term $\sigma_{n,i}$ is utilized to represent the standard deviation of noise that is also expressed by a gPC expansion of the following form:

$$\sigma_n(t,\xi) = \sum_{k=1}^{\infty} \sigma_{n,k}(t) \Phi_k(\xi) \quad (12)$$

, where $\sigma_{n,k}$ is the gPC coefficients of noise at time instant t, $\Phi_k(\xi)$ is the multi-dimensional polynomials in terms of ξ, and the variance of noise is calculated by (9).

The decision variable λ_{lower} in (11) is a vector consisting of the mean and variance of the unknown input ($\boldsymbol{g}$) and noise σ_n, and n is the number of the measured variables used to calibrate the gPC model. Due to noise and truncation error introduced by the gPC approximation, the mean and variance of the input variable ($\boldsymbol{g}$) defining λ_{lower} calculated from (11) deviate from the actual values entering the process. After obtaining λ_{lower}, it is possible to calculate the actual gPC coefficients for the state variables $\boldsymbol{x}$. Using these coefficients, the PDF profiles for $\boldsymbol{x}$*'s* around each constant mean value (operating mode) are approximated by substituting samples ($\boldsymbol{\xi}$) from its a priori known distribution into the resulting gPC expansions given in (4), and the PDFs are calculated as an histogram composed of bins each corresponding to different ranges of values of $\boldsymbol{x}$ (Du, et al., 2014).

Then, for each of the mean values considered in (10), a PDF can be calculated. If the system is operated around a constant mean $\bar{g}_i$, the corresponding i (i = *1,…, k*) in (10) is detected from the PDF profiles for a given measurement as follows:

$$i = arg\ max\ \{P_i\} \quad (13)$$

, where i is the operating mode as defined in (10). P_i means the probability of being operated around a particular mean $\bar{g}_i$ for a given measurement, each referred heretofore as '*Mode*'. The solution of this problem is depicted in Fig.2 showing 3 PDFs that correspond to 3 different operating modes (input mean values). For example, three probabilities (red dots) can be found for a given measured output shown in Fig.2, where the maximum probability is used to indicate that the system is operating around the mean value corresponding to '*Mode j*'.

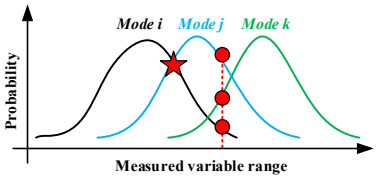

Fig. 2 Visual interpretation of FDD

3.3 Level-2: Maximum likelihood based input estimation and classification

The *Level-1 algorithm* presented in Section 3.2 assumes that the system is operated around a specific mean value $\bar{g}_i$ in (10), but it does not take into account transient responses resulting from the step changes between $\bar{g}_i$ shown in Fig.1 (a). This section explains the problem of estimating these input changes based on the maximization of a likelihood function. Define $f_x^{'}$ as the PDF of a measured output of interest, which can be estimated by a Gaussian kernel density function (Wand & Jones, 1995) as:

$$f_x^{'}(x', \pi(\xi)) = \frac{1}{n_i} \sum_{i=1}^{n_i} G_k \left[x^{'} - \pi(t, x(\xi), u; g(\xi)) \right] \quad (14)$$

, where G_k denotes the Gaussian kernel, $x^{'}$ is the measured variable, n_i is the number of samples drawn from an a priori known distribution of $\boldsymbol{\xi}$. The π operator is defined as the gPC model conditioned on the gPC coefficients and is obtained as explained in Section 2. Then, a likelihood function of output variables $\boldsymbol{x}$ over a moving time window of m measurements can be calculated as:

$$\ell(\pi, x') = \prod_{j=1}^{m} f_x^{'}(x_j^{'}; \pi) \quad (15)$$

Then, for the moving window of length m, an estimate of an average value of the input $\boldsymbol{g}$ is obtained by maximizing the likelihood function as follows:

$$\max_{\lambda_{upper}} \ell(\pi, x') = \prod_{j=1}^{m} f_x^{'}(x_j^{'}; \pi) \quad (16)$$

, where the decision variable λ_{upper} in (16) is the average of the dynamic value of the unknown time varying input $\boldsymbol{g}$ and/or its confidence interval over the moving window of m measurements. It is worth noting that the same set of $\boldsymbol{\xi}$ used in (15) is also used to maximize (16). For classification purposes, the average value of the input $\boldsymbol{g}$ resulting from (16) is compared to the values $\bar{g}_i$ to identify which operating mode is active. Although the *level-2 algorithm* could be executed at each time interval, it is only used after a large deviation is detected by the *level-1 algorithm* to reduced computation.

3.4 FDD by combining level-1 and level-2 algorithms

The *two-level* fault detection algorithm proceeds as per the following steps:

Step 1 – The PDF profiles of the measured variable $\boldsymbol{x}$ in (4) operating around each one of the mean value $\bar{g}_i$ in (10) is approximated using *level-1 algorithm* in Section 3.1;

Step 2 – When a sample of measurements is available, the probabilities P_i in (13) are assessed ($i = 1,...,k$). The maximum probability is used to infer a particular mean value $\bar{g}_i$ (operating mode) as shown in Fig.2.

Step 3 – A potential change in the operating mode (mean value $\bar{g}_i$) is detected when the probability of a given output measurement switches across a limit between two adjacent PDF, as depicted in Fig.2 (red star), corresponding to, $P_i = P_j$;

Step 4 – If a change in operating mode is detected in Step 3, the maximum likelihood based input estimation (*level-2 algorithm*) in Section 3.3 is executed and a value and/or a confidence interval of the input $\boldsymbol{g}$ averaged over a moving window of m measurements is estimated;

Step 5 – The value obtained from the *level-2 algorithm* in Step 4 is compared based on a minimum distance criterion to the set of mean values $\bar{g}_i$ (operating modes) obtained in (11) and the corresponding operating mode (mean value) is diagnosed.

To evaluate the performance of the proposed algorithm, the fault detection rate (*FDR*) (Yin, et al., 2012) is defined as:

$$FDR = n_d/n_{total} \tag{17}$$

, where n_{total} is the total number of tested samples and n_d is the number of samples that have been correctly classified.

4. EXAMPLE: REACTOR-SEPARATOR PROCESS

Simulation studies of a nonlinear chemical process consisting of two reactors and a separator are used to demonstrate the efficacy of the proposed two levels' methodology and for comparison with a Particle Filter (PF) (see Fig.3).

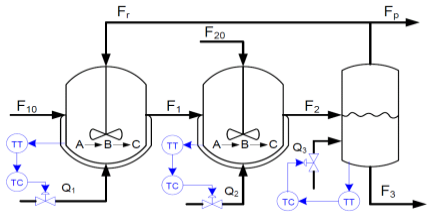

Fig. 3 Two reactors in series with separator and recycle

A stream of reactant A is added to each reactor and converted to the product B by the first order reaction. The feed mass fraction of reactant A (x_{A0}) is the unknown input fault ($\boldsymbol{g}$) and is corrupted by normally distributed perturbations around three mean values (operating modes) as described in (10) (Fig. 1). The first principles' model for the plant is:

$$\dot{H}_1 = (1/\rho A_1)(F_{f1} + F_R - F_1)$$
$$\dot{x}_{A1} = (1/\rho A_1 H_1)(F_{f1}x_{A0} + F_R x_{AR} - F_1 x_{A1}) - k_{A1}x_{A1}$$
$$\dot{x}_{B1} = (1/\rho A_1 H_1)(F_R x_{BR} - F_1 x_{B1}) + k_{A1}x_{A1} - k_{B1}x_{B1}$$
$$\dot{T}_1 = (1/\rho A_1 H_1)(F_{f1}T_0 + F_R T_R - F_1 T_1) - (1/C_p)(k_{A1}x_{A1}\Delta H_A + k_{B1}x_{B1}\Delta H_B) + Q_1/\rho A_1 C_p H_1$$
$$\dot{H}_2 = (1/\rho A_2)(F_{f2} + F_1 - F_2)$$
$$\dot{x}_{A2} = (1/\rho A_2 H_2)(F_{f2}x_{A0} + F_1 x_{A1} - F_2 x_{A2}) - k_{A2}x_{A2}$$
$$\dot{x}_{B1} = (1/\rho A_2 H_2)(F_1 x_{B1} - F_2 x_{B2}) + k_{A2}x_{A2} - k_{B2}x_{B2}$$
$$\dot{T}_2 = (1/\rho A_2 H_2)(F_{f2}T_0 + F_1 T_1 - F_2 T_2) - (1/C_p)(k_{A2}x_{A2}\Delta H_A + k_{B2}x_{B2}\Delta H_B) + Q_2/\rho A_2 C_p H_2$$
$$\dot{H}_3 = (1/\rho A_3)(F_2 - F_D - F_R - F_3)$$
$$\dot{x}_{A2} = (1/\rho A_3 H_3)(F_2 x_{A2} - (F_D + F_R)x_{AR} - F_3 x_{A3})$$
$$\dot{x}_{B2} = (1/\rho A_3 H_3)(F_2 x_{B2} - (F_D + F_R)x_{BR} - F_3 x_{B3})$$
$$\dot{T}_3 = (1/\rho A_3 H_3)(F_2 T_2 - (F_D + F_R)T_R - F_3 T_3) + Q_3/\rho A_3 C_p H_3$$

, where the subscript *'i'* (*i.e.,* 1, 2, 3) refers to the number of vessel, x is the mass fraction of A or B, T is temperature, H is the level, F is the flow rate and the reaction terms are:

$$F_i = k_{vi}H_i, k_{Ai} = k_A \exp(-E_A/RT_i)$$
$$k_{Bi} = k_B \exp(-E_B/RT_i)$$

The recycle flow and the three weight percent factors satisfy:

$$F_i = k_{vi}H_i, x_{AR} = \alpha_A x_{A3}/x_3, x_{BR} = \alpha_B x_{B3}/x_3$$
$$x_3 = \alpha_A x_{A3} + \alpha_B x_{B3} + \alpha_C x_{C3}, x_{C3} = (1 - x_{A3} - x_{B3})$$

Each of the tanks in the plant has an external heat input and is manipulated by a *PI* controller as:

$$Q_i(t) = Q_{(ss)i}(t) + K_{pi}(T_{(set)i} - T_i) + K_{pi}/\tau_i \int_0^t (T_{(set)i} - T_i(t^*))dt^*$$

The parameters used for the simulation are given in (Stewart, et al., 2010).

5. RESULTS AND DISCUSSIONS

5.1 Model formulation for Reactor-Separator case study

The FDD problem consists of detecting the mean values of unknown feed mass fraction x_{A0} based on measurements of the heat input Q_1. Since the solution of the gPC coefficients involved in the gPC expansions of each one of the states ($\boldsymbol{x}$ in (1)) given in Section 2 requires the application of Galerkin projection, it is only applicable to polynomial terms. Hence, the use of Galerkin projection for non-polynomial terms such as the Arrhenius energy function components k_{Ai}, requires approximation by a 2nd order Taylor expansion around each mean value on input x_{A0}. Since the random variable ξ is normally distributed, the basis polynomial functions for the gPC are chosen as Hermite to maintain orthogonality (Xiu, 2010).

5.2 Level-1: calibration of PDFs of the measured variable Q_1 for operation around constant values of x_{A0}.

For simplicity, 3 mean values on feed mass fraction (x_{A0}) are studied, *i.e.,* 0.65, 0.75 and 0.85 (k = 3 in (10)). Stochastic perturbations are added around these mean values, which are assumed to be normally distributed with zero mean and a standard deviation of 0.1. The changes in the mean follow a ML-PRS signal as shown in Fig.1 (a). The number of step changes of the unknown input (x_{A0}) among the 3 selected mean values (operating mode) in the ML-PRS is 242 and the maximum number of measurements between two consecutive step changes in input is limited to 1000.

Table 1 shows the model calibration results with the *level-1 algorithm* as described in (11) for a 1% noise level based on measurements of Q_1. The highest order of the polynomials used for the gPC models is 2 (p = 2 in (6)).

Table 1. Model calibration result for *level-1 algorithm*

x_{A0}	x'_{A0}	σ_{A0}	σ_n	*Time*(*s*)
0.65	0.6370	0.0937	0.0188	992
0.75	0.7364	0.0979	0.0199	788
0.85	0.8319	0.0933	0.0201	871

In Table 1, x_{A0} is the known mean value used for simulations, x'_{A0} and σ_{A0} are the mean and standard deviation calculated from problem (11), σ_n is the standard deviation of noise. As explained before, the mean and standard deviation of the input resulting from (11) (x'_{A0}, σ_{A0}) are not identical to the actual simulated values used for calibration (x_{A0}, *0.1*), since the sensor data are corrupted by noise and uncertainty (see the differences between the values of the 1st and 2nd column in Table 1).

Once the gPC model is constructed, the PDF profiles of the measured variable (Q_1), estimated at the 3 mean values of the feed mass fraction (x'_{A0}), can be obtained. Fig.4 shows the PDF profiles for the external heat in vessel 1, Q_1, where the horizontal axis signifies the range of Q_1, and the vertical axis is the normalized probability.

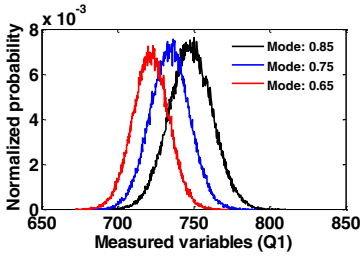

Fig. 4 PDFs of 3 modes on measured variable (Q_1)

Although the *level-1 algorithm* has been proposed only as a preliminary step of the *level-2 methodology*, the efficiency of the *level-1 algorithm* in detecting a mode is tested first to justify the need for the *level-2 algorithm*. Table 2 shows the result of Fault Detection Rate (*FDR*) with different noise levels using the PDF each calculated based on the assumption of constant values of x_{A0}. To comply with the assumption that the system is operated around a fixed input, the detection efficiency is first studied by using collected measurements of Q_1 before a switch between means occurred (see inset Fig.1 (b)-A).

Table 2. *FDR* with *level-1 algorithm* (fixed means)

x_{A0}	*Noise level*		
	1%	*2%*	*3%*
0.65	94%	91%	92%
0.75	92%	90%	86%
0.85	96%	94%	91%
Average	94%	92%	90%

In Table 2, there are 1000 test samples for each feed mass fraction, and the average of *FDR* decrease as expected when the noise level increases. It is worth noting that the model calibration as per the optimization in (11) would be prohibitive if Monte Carlo (MC) simulations were to be used instead of a gPC. For instance, the processor time required for one cost evaluation with MC (5000 samples) is ~15465 seconds. The search for the optimum in (11) for each mean value requires 40~60 iterations and takes approximately 171 ~ 257 hours on average. However, the proposed method takes ~15 minutes to calculate the optimum in (11) for all mean values, as can be seen in Table 1. Also, 5000 samples with MC was found to be inaccurate as compared to the gPC method in terms of *FDR*, thus an even larger number of samples are required to obtain comparable *FDR* as with gPC, which would further increase the computational burden.

As mentioned above, the *level-1 algorithm* is only suitable when the system is operating for long periods around a fixed mean. Thus, it is expected to be less accurate during periods where changes among mean values occur. To demonstrate this point, Table 3 shows the *FDR* with the *level-1 algorithm* by using measurements collected during the transition periods, *i.e.,* immediately after the occurrence of a step change in mean value of x_{A0} (see inset Fig.1 (b)-B). Windows of 50 measurements of Q_1 are used for testing and the detection is based on the average of the probabilities of these 50 measurements with respect to the PDF profiles generated in the *level-1 algorithm*.

Table 3. FDR with *level-1 algorithm* (transition)

x_{A0}	*0.65*	*0.75*	*0.85*	*Average*
FDR	*68%*	*72%*	*67%*	*69%*

As seen in Table 3, the *FDR* for switching periods between mean values on average is very low at 59%, thus justifying the necessity for the *level-2 algorithm* that does not assume that the mean of input x_{A0} is known as in the *level-1 algorithm*. However, it should be noted that the *level-1 algorithm* is proposed only to evaluate the necessity for executing the *level-2 algorithm* so as avoid executing the *level-2 algorithm* too frequently which would require excessive computation time.

5.3 FDD results using the level-2 algorithm

The *Level-2 algorithm* is only launched after the *level-1 algorithm* has indicated a potential change in the input mean value of x_{A0}.Table 4 shows the results by using the *level-2 algorithm* for three case studies to evaluate the efficacy and

computation time. In the first case study, windows of 50 measurements of Q_1 (m = 50 in (16)) are used to compare the results obtained by the *level-1 algorithm* as shown in Table 3. For the first case study only the average value of x_{A0} over the windows is chosen as a decision variable for λ_{upper} in (16). For the other two case studies, 100 measurements of Q_1 are used (m = 100 in (16)). The decision variable of the second case study in (16) is the average value of x_{A0}, while in the third case study both the average value and confidence interval of x_{A0} are used. It should be noted that only average values of the input x_{A0} are obtained in (16) over a time window of m measurements.

Table 4. *FDR* with *level-2 algorithm*

Case studies	*FDR*	*Time(s)*
1 (*m = 50*)	70%	225
2 (*m = 100*)	85%	498
3 (*m = 100*)	80%	1133

It can be seen by comparing Table 3 and Table 4 that the *level-2 algorithm* shows better *FDR* performance as compared to the *level-1 algorithm* alone. For instance, the *FDR* is 70% for the first case study, which has been increased by ~ 10 percent point as compared with the *level-1 algorithm* for the same number of measurements thus confirming the necessity for the *level-2 algorithm* to detect transitions among mean values of x_{A0}.

Finally, comparison studies are conducted between the proposed algorithm and a particle filter (PF). The noise for the PF algorithm is assumed to be equal to the one used for the gPC based *level-2 algorithm*. Fig.5 shows the dynamic value and posterior standard deviation (*s.t.d.*) calculated by the PF for one of the tested input values, *i.e.,* x_{A0} = 0.8388, where three different initial states and 100 particles are used.

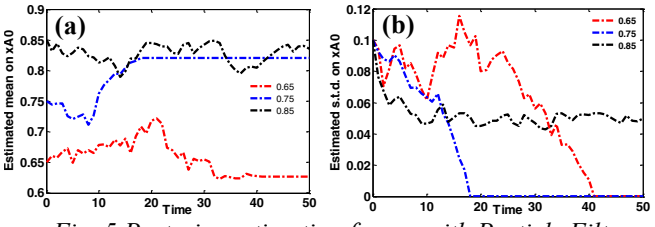

Fig. 5 Posterior estimation for x_{A0} with Particle Filter

The PF is executed for a duration of *time* = 50 to permit a fair comparison with the gPC based algorithm that uses a window of m=50 measurements. As seen in Fig.5, three sets of the initial state in PF provide very different posterior estimates on the average value of x_{A0}. For example, the PF method stabilizes at ~0.6321 with an initial state 0.65, which would indicate that the closest feed concentration, out of the 3 mean values considered in the case study, is 0.65 whereas the actual input value is 0.8388. As expected, the PF approach cannot capture the posterior variance on x_{A0} as shown in Fig.5 (b), since the standard deviation theoretically goes to 0 as time progresses.

Additional studies are conducted to investigate the processor time with the PF. As done for the gPC method, 50 and 100 measurements are tested respectively. The PF requires ~6800 seconds to run for 50 time intervals and ~13780 seconds for 100 time intervals. As shown in Table 4, the gPC based *level-2 algorithm* requires significantly less computational effort, while it is not sensitive to the user choice of initial guesses as the PF. In order to reduce the sensitivity to initial guesses one can execute the PF algorithm for a larger set of initial guesses and then average the results. However based on the computation times discussed above, such approach will be prohibitive especially for real time operation.

6. CONCLUSIONS

This paper proposes a *two-level* fault detection and diagnosis approach based on generalized polynomial chaos modelling and maximum likelihood based estimator for faults of a stochastic nature. The proposed method is demonstrated using a simulation of a nonlinear chemical plant with two continuously stirred tank reactors and a flash tank separator. The results show that the proposed methodology is highly computationally efficient as compared to simulation based approaches such as MC and PF and it is not sensitive to the user selected tuning parameters such as PF.

ACKNOWLEDGEMENTS

NSERC is acknowledged for the financial support.

REFERENCES

Chen-Charpentier, B. & Stanescu, D., 2014. Parameter estimation using polynomial chaos and maximum likelihood. International Journal of Computer Mathematics, 91(2), pp. 336-346.

Du, Y., Budman, H. & Duever, T., 2014. Integration of fault diagnosis and control by finding a trade-off between the observability of stochastic fault and economics. Cape Town, South Africa, The 19th World Congress of the IFAC.

Elsheikh, A. H., Hoteit, I. & Wheeler, M. F., 2014. Efficient Bayesian inference of subsurface flow models using nested sampling and sparse polynomial chaos surrogates. Computer Methods in Applied Mechanics and Engineering, Volume 269, pp. 515-537.

Nagy, Z. & Braatz, R. D., 2007. Distributional uncertainty analysis using power series and polynomial chaos. Journal of Process Control, Volume 17, pp. 229-240.

Pence, B. L., Fathy, H. K. & Stein, J. L., 2011. Recursive maximum likelohood parameter estimation for state space systems using polynomial chaos theory. Automatica, Volume 47, pp. 2420-2424.

Stewart, B. T. et al., 2010. Cooperative distributed model predictive control. Systems and Control Letters, Volume 59, pp. 460-469.

Xiu, D., 2010. Numerical Methods for Stochastic Computation: A Spectral Method Approach. s.l.:Princeton University Press.

Yin, S. et al., 2012. A comparison study of basis data driven fault diagnosis and process monitoring methods on teh benchmark Tennessee Eastman process. Journal of Process Control, Volume 22, pp. 1567-1581.