The Detection of Multiple Faults in a Bayesian Setting using Dynamic Programming Approaches

Hamed Habibi[1], Ian Howard[1], and Reza Habibi[2]
[1]Faculty of Science and Engineering, School of Civil and Mechanical Engineering, Curtin University, Australia
[2]Iran Banking Institute, Central Bank of Iran, Tehran, Iran
Corresponding author: Hamed Habibi, email: hamed.habibi@postgrad.curtin.edu.au

Abstract. Inspired by the need for improving the reliability and safety of complex dynamic systems, this paper tackles the multiple faults detection problem using Dynamic Programming (DP) based methods under the Bayesian framework. These methods include (*i*) Maximum-A-Posteriori (MAP) estimator approach, (*ii*) Monte Carlo Markov Chain (MCMC) posteriors, (*iii*) Set Membership (SM) approach, (*iv*) probability of fault and (*v*) alternative methods. Using Bernoulli and Poisson priors, the Bayesian DP-type MAP estimate of all unknown parameters is presented. To derive the posterior distributions of Bayesian point estimations, the MCMC method is applied. For the SM approach, the Bayesian feasible parameter space is derived, as Bayesian confidence interval. The SM criteria are proposed to detect multiple faults which also reduces the Bayesian complexity of MAP estimator. For online fault detection, using the Bayesian model selection technique and the MAP estimator, the DP-based probability of faults is given, serving as a Bayesian early warning system. Since running DP algorithms is a time-consuming, alternative methods are also proposed using the modified MAP estimator. These methods use iterative approximations of MAP estimates, via the application of an iterative Expectation–Maximization algorithm technique. Numerical simulations are conducted and analysed to evaluate the performance of the proposed methods.

Keywords: *Bayesian posterior; Bayesian Model Selection; Dynamic Programming; Early Warning System; Iterative MAP; MCMC; Probability of Fault; Set Membership.*

1 Introduction.

Industrial dynamic plants, especially those that operate in harsh environments, are prone to fault occurrence, which in turn leads to performance degradation, shut down, or catastrophic breakdown [1]. Accordingly, the automated Fault Detection and Identification (FDI) techniques have been widely exploited in recent decades and applied to industrial plants, such as wind turbines [2, 3], induction motors [4] and aircraft landing systems [5]. It is worth noting that FDI plays a crucial role in preventive maintenance, in which the estimated fault information can be used to plan the maintenance procedures. Consequently, the required downtime and maintenance costs are reduced. On the other hand, timely and accurate FDI may prevent catastrophic failure events, so improving the overall system safety and reliability. The main problem still to be addressed for FDI application on complex nonlinear systems is the presence of uncertainty and disturbance on the nominal model, which might cause false/missed detections [6]. Accordingly, several FDI approaches have been proposed to cope with this issue, such as fuzzy FDI [7], residual filter , robust sliding mode observer [8] and neural networks .

From a mathematical point of view, the FDI problem can be addressed as change-point analysis, which can be outlined as the estimation of parameter changes in statistical distribution functions of the observed data series [9]. In this regard, FDI and change-point analysis are used interchangeably. In the signal processing context, the problem of multiple change-points is of interest, in which the number of changes (segments) and locations of changes are unknown parameters. For this purpose, the most widely adopted technique is the maximum likelihood estimate. The change-point modelling involves three main steps; choice of parametric forms, segment boundaries and number of changes [10].

The widely-adopted approach of change-point analysis is binary splitting in which the data set is split into two or more subsequences considering a predictor. This procedure is continued recursively until the subsequences are no longer separable. This is a greedy approach such that it usually fails to optimally split the data set. The multiple change-points problem can be considered as an optimal identification of segment neighborhood in the system identification context [11]. In [12], an optimal multi-way splitting algorithm for an exponential family model was proposed using the Dynamic Programming (DP) algorithm. The DP algorithm has an execution time that is linear with the number of segments and quadratic for the number of potential change-points. DP methods for detection of multiple change-points in multivariate data were considered in [13]. In [14] the DP method was applied to compute the critical values of multiple change-points tested in a regression setting. A pruned DP for optimal multiple change-points detection was proposed in [15]. A fast segmentation procedure based on DP algorithm for multiple change-points detection was studied in [16].

In addition to deterministic model characterization, which leads to the worst-case scenario and hard bound on the uncertainty region, stochastic models lead to probabilistic uncertainty regions. Particularly, the Bayesian inferential setting in FDI has been increasingly adopted which enables the implementation of expert's knowledge [17]. The Bayesian binary segmentation procedure for a Poisson process with multiple change-points was studied in [18]. The exact and efficient Bayesian estimators for the number of change-points and their locations was also evaluated in [19]. A Bayesian setting in [20] was proposed giving an optimal FDI with a

linear computational cost. The main unique characteristics of the Bayesian setting is that the belief degree is augmented into the likelihood. On the other hand, it can deal with the dynamics that are nonlinear in parameters. Also, this setting is suitable for small amounts of data, as prior distribution is obtained using an expert's knowledge. It is worth noting that there are considerable works on the application of the Bayesian inferential setting in system identification problems. However, the difficulty in the calculation of the posterior distribution is the main problem still to be addressed. The Markov Chain Monte Carlo (MCMC) method has been proposed as an advanced simulation technique to eliminate this difficulty [21]. The application of MCMC in multiple change-point detection was considered in [22]. The CpInGLM, a new R-package, to evaluate the performance of the DP method using MCMC was proposed in [23]. The EM algorithm could be applied to approximate posteriors and overcome the DP time consuming difficulties (see [24]).

The Set Membership (SM) technique is another approach to reduce the Bayesian FDI computational complexity and is widely used in the system identification field. Computation of variance and bias is referred to as the error modeling approach which is applied mostly in the SM method [25]. An advantage of the SM method is that it does not need to identify the functional form of the underlying model, for example, the regression model, and it permits one to specify the optimal system identification (see [26] and references therein). When performing the SM method, it is hypothesized that the error of identification is unknown but bounded. This hypothesis leads to a convex polytope or in some cases a much more complicated shape called the Feasible Parameter Set (FPS). For computational purposes, some recursive set-based formulas for the FPS are derived. The optimal functional form of, for example, the regression model is derived by minimizing the identification error throughout the FPS [27]. As mentioned in [28], the SM method is applicable in FDI. A fault is detected if at some points the FPS becomes empty, that is $\mathrm{FPS} = \emptyset$. In [27], FDI using the SM approach under the Bayesian setting was studied, in which the Bayesian credible set (Bayesian confidence interval) played the role of the FPS and some necessary and sufficient conditions were derived for checking the hypothesis of no change-point. These methods are offline manner of multiple faults detection. Some measures such as average run lengths and probability of faults compute the delay in fault detection and the location of possible future faults of dynamical system, respectively. The second measure plays the role of an early warning system (EWS) [29]. Indeed, the locations of faults specifies structure of mathematical model of dynamical system and some techniques such as Bayesian model section (BMS) may be applied to compute the probability of faults [30]. The probability of fault as the main indicator, is studied in [2, 9].

Motivated by the aforementioned considerations, the main contribution of this paper can be outlined as follows.

1- The FDI problem is considered in a Bayesian setting and the DP approach is adopted for finding MAP estimates of unknown parameters using prior's information.
2- In contrast to [12], where the exponential distribution family was considered, in this paper a general distribution family is considered. This assumption allows us to consider the change point analysis in the presence of heavy tailed observations, which appears frequently in the FDI context [31].
3- To reduce the overall computational complexity, the DP method is applied in the MCMC framework.
4- SM technique is adopted as the Bayesian's belief set, into which DP is implemented. Accordingly, SM criteria are obtained.
5- The BMS is utilized to compute the fault probability as an EWS, as a useful online tool to detect the gradual faults in uncertain dynamic systems.
6- To compare the numerical results, the alternative methods, based on iterative Maximum-A-Posteriori (MAP) estimators are proposed. Consequently, different practical systems are numerically considered, and the performance of the proposed method is investigated.

The rest of paper is organized as follows. In the Section 2, the Bayesian DP FDI technique is derived under two different types of priors. Also, MCMC is used to approximate the posterior density function. To this end, the DP method is applied. The DP-based SM FDI in the Bayesian setting is given, and corresponding criteria are derived. Using the BMS technique, probabilities of faults as EWS are computed, in two cases of abrupt and gradual types of change point models. In Section 3, an alternative method is described on which basis the results are compared. For running the iterative MAP algorithm, the Expectation–Maximization (EM) algorithm is used. In Section 4, the numerical simulations are conducted which includes illustrative examples, practical systems and comparisons. Finally, the conclusions are given in Section 5.

2 DP-based methods for multiple FDI

In this section, the multiple FDI problems are studied under the Bayesian inference setting and using the DP method. DP methods are divided in three categories. The first category contains methods which use the Bayesian point estimations. These methods are MAP estimators and MCMC posterior. Indeed, the MAP estimates of the fault locations are derived, and it is seen that they are the penalized contrast estimators (see [13]). Also, the posterior approximations of the MCMC method are derived using the DP method. Indeed, the

DP posterior density functions, as exact solutions for posterior density functions, are derived. The second category contains the SM technique, as a Bayesian FPS, according to [27]. Then, SM methods are developed in which the FPS can be interpreted as a Bayesian credible (confidence) interval, which is an alternative method to the Bayesian point estimations, considered in the second category. Finally, using the BMS, the online version of multiple fault detection is studied, by deriving the probability of possible future faults, which can be used as an EWS.

Suppose that the density function of vector valued variable $\boldsymbol{X}$ denoted by $f(\boldsymbol{x}, \boldsymbol{\theta})$ and a random sample of size n, $\boldsymbol{X}_1, \dots, \boldsymbol{X}_n$ from $f(\boldsymbol{x}, \boldsymbol{\theta})$ are given. The parameter $\boldsymbol{\theta}$ is vector valued. A typical multiple change-points problem is proposed as follows.

There are $R-1$ *number of change-points* $\tau_1, \dots, \tau_{R-1}$ *such that variables* $\boldsymbol{X}_j$ *come from* $f(.\,, \boldsymbol{\theta}_{i-1})$, *for* $\tau_{i-1}+1 \le j \le \tau_i$, $i = 2, \dots, R$. *Let* $\tau_1 = 1, \tau_R = n$ *and* $l_i = \tau_i - \tau_{i-1}$.

2.1 DP method in MAP estimators. Here, the non-Bayesian DP method of Hawkins [12] for multiple change-points is extended to the Bayesian setting of DP method. Bayesian methods have advantages such as updating parameters, a useful technique for involving expert opinion to the problem at hand, as well as for finding a practical solution when there is no enough available data [30]. Also, in the mathematical framework, the exponential family of distributions, considered in [14], is replaced by a general class of distributions, since some other types of distributions such as Weibull, Rayleigh and logistic distributions with fat tails appear as disturbance distributions in FDI [31]. To compare the performances of the DP method with other proposed methods, the normal distribution is applied; however, a comparison for non-normal distribution is also given.

The likelihood function of parameters $\boldsymbol{\theta}_1, \dots, \boldsymbol{\theta}_R, \tau_1, \dots, \tau_{R-1}, R$ is given by

$$\begin{aligned} L(\boldsymbol{\theta}_1, \dots, \boldsymbol{\theta}_R, \tau_1, \dots, \tau_{R-1}, R) &= \prod_{i=1}^{R} \prod_{j=\tau_{i-1}+1}^{\tau_i} f(\boldsymbol{x}_j, \boldsymbol{\theta}_i) \\ &= \prod_{i=1}^{R} \exp\{\sum_{j=\tau_{i-1}+1}^{\tau_i} \log(f(\boldsymbol{x}_j, \boldsymbol{\theta}_i))\}. \end{aligned} \tag{1}$$

Here, the notation of *log* means the logarithm in natural base. Let $\pi_i(\boldsymbol{\theta}_i)$ be the prior information about $\boldsymbol{\theta}_i$. Thus, given $\tau_1, \dots, \tau_{R-1}$, the posterior $\boldsymbol{\theta}_i$ is proportional to

$$\prod_{i=1}^{R} \exp\{\log(\pi_i(\boldsymbol{\theta}_i)) + \sum_{j=\tau_{i-1}+1}^{\tau_i} \log(f(\boldsymbol{x}_j, \boldsymbol{\theta}_i))\}. \tag{2}$$

Let $\widehat{\boldsymbol{\theta}}_i$ be the MAP estimate of $\boldsymbol{\theta}_i$, given $\tau_1, \dots, \tau_{R-1}$ based on observations $\boldsymbol{X}_j$ of i-th segment $\tau_{i-1}+1 \le j \le \tau_i$, assuming τ_{i-1} and τ_i are known (see [24]). Therefore, the profile pseudo-likelihood of $\tau_1, \dots, \tau_{R-1}$ based on the Bayesian posterior in (2) is given as follows [12].

$$\begin{aligned} L(\tau_1, \dots, \tau_{R-1}) &= \prod_{i=1}^{R} \exp\{\log\left(\pi_i(\widehat{\boldsymbol{\theta}}_i)\right) + \sum_{j=\tau_{i-1}+1}^{\tau_i} \log(f(x_j, \widehat{\boldsymbol{\theta}}_i))\} = \\ &\prod_{i=1}^{R} \exp\{\log\left(\pi_i(\widehat{\boldsymbol{\theta}}_i)\right) + \sum_{j=\tau_{i-1}+1}^{\tau_i} \boldsymbol{y}_{ij}\} = L(l_1, \dots, l_R, R), \end{aligned} \tag{3}$$

where $l_i = \tau_i - \tau_{i-1}$, $y_{ij} = \log(f(\boldsymbol{x}_j, \widehat{\boldsymbol{\theta}}_i))$, and $\overline{\boldsymbol{y}}_{i+} = \left(\sum_{j=\tau_{i-1}+1}^{\tau_i} y_{ij}\right)/l_i$. Omitting useless terms in (3), Hawkins [12] proposed a DP recursive algorithm for maximum $-2\ln(L(l_1, \dots, l_R, R))$ over $\tau_1, \dots, \tau_{R-1}, R$. The full posterior function is obtained by considering appropriate priors $p(\tau_1, \dots, \tau_{R-1})$. Notice that any prior $p(\tau_1, \dots, \tau_{R-1})$ can be represented as prior $p(l_1, \dots, l_R)$. Two important classes of priors are

Class (i). The Poisson process for $\tau_i, i = 1, \dots, R$ of [30] at which it is assumed that l_i's are independent with common Poisson distribution with intensity parameter $\gamma > 0$.

Class (ii). The Bernoulli process for $\mathrm{r_t}, \mathrm{t} = 1, \dots, \mathrm{n}$ of [32] where $\mathrm{r_t}$ is one if a change has occurred at time t and zero otherwise. Assuming $\mathrm{P}(\mathrm{r_t} = 1) = \pi$, then the prior $\mathrm{P}(\mathrm{l_i} = \mathrm{x})$ is proportional to $(1-\pi)^{\mathrm{x}}, \mathrm{x} = 0,1,2, \dots$.

Hence, the posterior function of $l_1, \dots, l_R$ or $\tau_1, \dots, \tau_{R-1}$ is proportional to

$$\begin{aligned} p(l_1, \dots, l_R) \exp\left\{\log\left(\pi_i(\widehat{\boldsymbol{\theta}}_i)\right) + \sum_{i=1}^{R} l_i \overline{\boldsymbol{y}}_{i+}\right\} = \\ \exp\left\{\sum_{i=1}^{R} \left\{\{l_i \overline{\boldsymbol{y}}_{i+} + \log\left(\pi_i(\widehat{\boldsymbol{\theta}}_i)\right)\right\} + \log(p(l_1, \dots, l_R))\right\}. \end{aligned} \tag{4}$$

The MAP estimates of $l_1, \dots, l_R$ are derived by maximizing (4). When, the l_i's are assumed to be independent having density functions $p_i(l_i)$ for each i's, then (4) is reduced to

$$\exp\{\sum_{i=1}^{R} l_i \overline{\boldsymbol{y}}_{i+} + \log p_i(l_i) + \log\left(\pi_i(\widehat{\boldsymbol{\theta}}_i)\right)\}. \tag{5}$$

Eq. (5) is a penalized contrast approach for estimating the number of faults in a multiple faults problem. Similar works are [33] with the Schwartz criterion, [32] with normal distribution and [34] with the non-Bayesian penalized contrasts approach. For example, assuming the normal distribution for observations and Bernoulli process for fault points τ_i's, the results of [32] are derived. Under non-informative priors and under the exponential family of distributions, the work of [12] is derived. For example, assuming uninformative prior π_i, consider the symmetric location family of distributions represented by density function

$$f(x, \theta) = g(x - \theta), \tag{6}$$

where g is a symmetric density function about zero. A natural estimator $\hat{\theta}_i$ is $\bar{x}_i = \left(\sum_{j=\tau_{i-1}+1}^{\tau_i} x_j\right)/l_i$. Then,

$$y_{ij} = \log\left(g\left(x_j - \bar{x}_i\right)\right) + \log\, p_i(l_i). \tag{7}$$

For the scale family distributions, where

$$f(x,\theta) = \frac{1}{\theta} g\left(\frac{x}{\theta}\right), \tag{8}$$

then

$$y_{ij} = -\log\left(\hat{\theta}_i\right) + \log p_i(l_i), \tag{9}$$

at which a proxy for $\hat{\theta}_i$ is $\left(\sum_{j=\tau_{i-1}+1}^{\tau_i} x_j\right)/l_i$. For the exponential family of distributions,

$$f(x,\theta) = \exp\{-\theta x + d(\theta) - c(x)\}, \tag{10}$$

omitting the useless term $c(x)$, the first term of posterior density function corresponds to the results of [12].

Here, for finding the MAP estimators, the DP algorithm is performed under two formulations for prior's classes (*i*) and (*ii*). Following [12], let $G(r,m)$ be the negative of maximized log posterior for a r-segments configuration fitted to sequence $\boldsymbol{X}_1, \dots, \boldsymbol{X}_m$. Then, following [12], it is seen that

$$G(R,n) = -max_{\{\tau_i\}} \sum_{k=1}^{R} Q(\tau_{k-1}, \tau_k). \tag{11}$$

For the case of class (*i*), with varying intensity parameter γ_i from one segment to the next, then

$$Q(\tau_{i-1}, \tau_i) = \{l_i \bar{\boldsymbol{y}}_{i+} + \log\left(\pi_i\left(\widehat{\boldsymbol{\theta}}_i\right)\right) - \gamma_i + l_i \log\left(\gamma_i\right) - \log\left(l_i!\right)\}. \tag{12}$$

Removing the unwanted second term γ_i from the Q function expression, it is seen that

$$Q(\tau_{i-1}, \tau_i) = \{l_i \bar{\boldsymbol{y}}_{i+} + \log\left(\pi_i\left(\widehat{\boldsymbol{\theta}}_i\right)\right) + l_i \log\left(\gamma_i\right) - \log\left(l_i!\right)\}. \tag{13}$$

For the case of class (*ii*), the Q function is given by

$$Q(\tau_{i-1}, \tau_i) = l_i \bar{\boldsymbol{y}}_{i+} + \log\left(\pi_i\left(\widehat{\boldsymbol{\theta}}_i\right)\right) - (l_i - 1)\log(1-\pi). \tag{14}$$

Proposition 1. *(a)* and *(b)* are true.

(a) $G(1,m) = Q(0,m)$,

(b) $G(r,m) = min_{0<h<m} G(r-1,h) + Q(h,m)$.

Since the same result of [12] is derived, then Hawkins method ([12], page 326) is applicable, here. Therefore, using a backward approach (see [12]) and assuming $\tau_R = n$ all fault times τ_k's and the number of faults R are estimated. As follows, $G(R,n)$ is compared to $F(R,n)$ defined by Hawkins [12]. To this end, following [12], let $F(R,n)$ be the (-1) times to the maximized log likelihood. Then, Proposition 2 is presented.

Proposition 2. Assuming non-informative prior for $\boldsymbol{\theta}_i$, then (*a*)-(*c*) are true.

(*a*) $F(R,n) - G(R,n) = \sum_{i=1}^{R} \log\, p_i(l_i)$,

(*b*) Under the Lavielle and Lebarbier's prior [32], assuming $p_i = p = p_n > 0.5$ and $(n-1)\log\left(p_n/1-p_n\right)$ goes to zero for large sample sizes n, then $|F(R,n) - G(R,n)| \to 0$, as $n \to \infty$.

(*c*) Under the Djafari and Ferron's prior [30], assuming $\gamma_i = \gamma_n$ and $nlog(\gamma_n e^{-\gamma_n}) \to 0$, then

$$|F(R,n) - G(R,n)| \to 0, \text{ as } n \to \infty.$$

2.2 DP in MCMC posterior. Here, like [32] in the case of normal distribution, the MCMC is presented to approximate the posterior function in the general class of distributions. A difference between our method and the work of [32] is that the DP method is applied to find the families of full conditional distributions which are necessary to perform the MCMC.

The full posterior is proportional to

$$\prod_{i=1}^{R} \prod_{j=\tau_{i-1}+1}^{\tau_i} f\left(\boldsymbol{x}_j, \boldsymbol{\theta}_i\right) \pi_i(\boldsymbol{\theta}_i) p_i(l_i). \tag{15}$$

The families of full conditional distributions are $\{\pi(\boldsymbol{\theta}_1, \dots, \boldsymbol{\theta}_R | l_i, i = 1, \dots, R, \boldsymbol{x}_j, j = 1,2, \dots, n\}$ and $\{\pi(l_i, i = 1, \dots, R, |\boldsymbol{\theta}_1, \dots, \boldsymbol{\theta}_R \boldsymbol{x}_j, j = 1,2, \dots, n\}$. In the case of scalar parameter θ_I, assume the j-th observation of the i-th segment has $N(\theta_I, \sigma^2)$ distribution and instead of estimating θ_I, suppose that it has normal $N(\mu_I, v_i^2)$ distribution, then the posterior of θ_I given l_i is given by $N(\alpha_i \mu_i + (1-\alpha_i)\bar{x}_{i.}, \alpha_i v_i^2)$ where $\alpha_i = \sigma^2/(\sigma^2 + l_i v_i^2)$. Finally, a MCMC method is applied to approximate the joint posterior of $\boldsymbol{\theta}_i$ and l_i given observations $\boldsymbol{x}_j, j = 1,2, \dots, n$.

Here, the DP method is suggested to derive the family of distributions of $\{\pi(l_i, i = 1, \dots, R, | \boldsymbol{\theta}_1, \dots, \boldsymbol{\theta}_R \boldsymbol{x}_j, j = 1,2, \dots, n\}$. The full conditional densities are

$$\begin{aligned} \pi\left(\boldsymbol{\theta}_1, \dots, \boldsymbol{\theta}_R \middle| {l_i}^k, i = 1, \dots, R\right) &= \prod_{i=1}^{R} \prod_{j=\tau_{i-1}^k+1}^{\tau_i^k} f\left(\boldsymbol{x}_j, \boldsymbol{\theta}_i\right) \pi_i(\boldsymbol{\theta}_i), \\ \pi\left(l_i, i = 1, \dots, R \middle| {\boldsymbol{\theta}_1}^k, \dots, {\boldsymbol{\theta}_R}^k\right) &= \prod_{i=1}^{R} \prod_{j=\tau_{i-1}+1}^{\tau_i} f\left(\boldsymbol{x}_j, {\boldsymbol{\theta}_i}^k\right) p_i(l_i). \end{aligned} \tag{16}$$

Starting with initial values $l_i^0, i = 1,2, \dots, R$, the MAP estimates of $\boldsymbol{\theta}_i^0, i = 1,2, \dots, R$ are obtained. Then, the MAP estimates of $l_i^1, i = 1,2, \dots, R$ are obtained by maximizing $\pi\left(l_i, i = 1, \dots, R \middle| {\boldsymbol{\theta}_1}^k, \dots, {\boldsymbol{\theta}_R}^k\right)$ with respect to $l_i, i = 1, \dots, R$. This procedure is repeated until it converges. A DP algorithm for obtaining the MAP estimates of

$l_i, i = 1, \dots, R$ is presented as follows. To this end, similar to Proposition 1, let $H(R,n)$ be the maximized *log* posterior of $\pi\left(l_i, i = 1, \dots, R \middle| {\boldsymbol{\theta}_1}^k, \dots, {\boldsymbol{\theta}_R}^k\right)$ for a R-segments configuration.

Proposition 3. *(a^*)* and *(b^*)* are true.

(a^*) $H(1,m) = Q^*(0,m)$,

(b^*) $H(r,m) = min_{0<h<m} H(r-1,h) + Q^*(h,m)$,

where $Q^*(\tau_{i-1},\tau_i) = \sum_{j=\tau_{i-1}+1}^{\tau_i} \log\left(f\left(\boldsymbol{x}_j, {\boldsymbol{\theta}_i}^k\right) p_i(l_i)\right)$.

2.3 SM FDI. The introduced model of Section 2 for FDI purposes contains the uncertainty. Interval sets are usually the simple way to define uncertainties. To this end, it is enough to enclose a lower and upper bound for error estimation and then find an FPS for the parameter that should be estimated. The Bayesian method gives a similar result using the credible set based on confidence interval using posterior distribution. In this section, following [27], the Bayesian SM method is proposed for the multiple faults detection problem. To this end, first, some SM criteria are given under the null hypothesis of no change-point. Then, these criteria are extended to detect the fault in at most one change-point (AMOC) models. Finally, the Bayesian SM methods are given for multiple change-points detection.

2.3.1 Deriving SM criteria. A modified version of the least square method for estimating unknown parameters to fit the generalized beta distribution is proposed in [35]. To describe more, suppose that F_θ is the distribution function of random variables $X_1, \dots, X_n$ under the null hypothesis of no change-point. Let

$$Z_j = F_\theta\left(X_{(j)}\right), j = 1,2, \dots, n, \tag{17}$$

where $X_{(j)}$ is the j-th ordered statistic. It is easy to see that Z_j has the beta distribution with parameters $j, n-j+1$. Hence, $E(Z_j) = j/(n+1)$. Therefore, in a form of regression, write

$$\frac{j}{n+1} = Z_j + \varepsilon_j = F_\theta\left(X_{(j)}\right) + \varepsilon_j, \tag{18}$$

where $E\left(\varepsilon_j\right) = 0$, for $j = 1,2, \dots, n$. The least square estimate of θ is derived by minimizing the object function $\sum_{j=1}^{n}(Z_j - j/(n+1))^2$ with respect to θ. AbouRizk [35] used the weighted least square method to estimate θ by minimizing the object function $\sum_{j=1}^{n} w_j(Z_j - j/(n+1))^2$, where $w_j = (n+1)^2(n+2)/j(n-j+1)$. Two other object functions are $T_n = max_{1\le t\le n}|\varepsilon_t|$ and $T_n^* = max_{1\le t\le n}|\varepsilon_t/\sqrt{w_t}|$.

In [27], variables $j/(n+1)$'s are data and Z_j are the unknown regression functional form. Let Θ_0 be the initial parameter space of θ. Notice that the set of all θ's in Θ_0 for which the error terms $\varepsilon_t, t = 1, \dots, n$ are bounded, constitutes the FPS. That is,

$$FPS(n) = \{\theta \in \Theta_0, T_n < \sigma\}, \tag{19}$$

for some thresholds $\sigma > 0$. Notice that, $T_n = \max(T_{n-1}, |\varepsilon_n|)$, and hence,

$$FPS(n) = FPS(n-1) \cap S(n), \tag{20}$$

where

$$S(n) = \left\{\theta \in \Theta_0, \left|Z_s - F_\theta\left(X_{(s)}\right)\right| < \sigma, s = n\right\}. \tag{21}$$

The weighted SM approach may be used to construct the modified FPS which is given by

$$FPS^*(n) = \{\theta \in \Theta_0, T_n^* < \sigma\}. \tag{22}$$

To consider the correlations between T_n's (or, T_n^*'s) the Autoregressive to Anything (ARTA) process technique of [36] is applied. These models are useful to study stationary time series having arbitrary marginal distributions and autocorrelation structures. An ARTA(1) model proposed for $\mathrm{T_n}$ may be given by

$$\mathrm{T_n} = \rho \mathrm{T_{n-1}} + \zeta_\mathrm{n}, \tag{23}$$

for some suitable white noise sequence ζ_n.

Here, the parameter ρ is estimated by the least square method as follows

$$\hat{\rho}_\mathrm{n} = \frac{\sum_{\mathrm{j=1}}^{\mathrm{n}} \mathrm{T_j T_{j-1}}}{\sum_{\mathrm{j=2}}^{\mathrm{n}} \mathrm{T_{j-1}^2}}. \tag{24}$$

The ARTA(1) model extracts the recursive relation between $FPS(n)$ and $FPS(n-1)$.

The following proposition summarizes the above discussion for estimated parameters. To this end, let $\hat{\theta}$ be the least square estimate of θ under the null hypothesis of no faults obtained by minimizing objective function $\sum_{j=1}^{n}\left(\frac{j}{n+1} - F_\theta\left(X_{(j)}\right)\right)^2$, where $X_{(j)}$ is j-th ordered statistics and F_θ be the common distribution function of $X_j, 1 \le j \le n$. Assume that $\hat{\varepsilon}_j = \frac{j}{n+1} - F_{\hat{\theta}}\left(X_{(j)}\right)$, $\hat{T}_n = max_{1\le t\le n}|\hat{\varepsilon}_t|$ and $\widehat{FPS}(n) = \left\{\theta \in \Theta_0, \hat{T}_n < \sigma\right\}$, for some thresholds $\sigma > 0$.

Proposition 4. (*a*)-(*c*) are correct.

(*a*) There are some faults among the first n observations if and only if $\widehat{FPS}(n) = \emptyset$.

(*b*) Suppose that $\hat{T}_n^* = max_{1\le t\le n}|\hat{\varepsilon}_t/\sqrt{w_t}|$, $w_j = \frac{(n+1)^2(n+2)}{j(n-j+1)}$ and $\widehat{FPS}^*(n) = \{\theta \in \Theta_0, \hat{T}_n^* < \sigma\}$ be the weighted SM criteria. There are some faults among the first observations if and only if $\widehat{FPS}^*(n) = \{\theta \in \Theta_0, \hat{T}_n^* < \sigma\}$.

(*c*) Let $\hat{\rho}_n = \frac{\sum_{j=1}^{n} T_j T_{j-1}}{\sum_{j=2}^{n} T_{j-1}^2}$. Then, the ARTA(1) model $\hat{T}_n = \hat{\rho}_n \hat{T}_{n-1} + \zeta_n$ verifies the nested relation $\widehat{FPS}(n) \subseteq \widehat{FPS}(n-1)$. Similar result is correct for $\widehat{FPS}^*(n)$ by replacing $\hat{T}_n$ with $\hat{T}_n^*$.

So far, the ordered statistics are used to construct the FPS's. However, an alternative simple method exists and is given as follows. Define $\delta_i = F_\theta(X_i), i = 1,2,\dots,n$, where F_θ is the initial distribution function of X_i's before any changes. Then, δ_i's are independent and identically uniformly distributed on (0,1) random variables under the null hypothesis of no change-point. Thus, $\delta_i = 0.5 + e_i$. Let $T_n^{**} = max_{1\le i\le n}|\delta_i - 0.5|$. Here, the FPS is $\{\theta \in \Theta_0, T_n^{**} < \sigma\}$. Let $\delta_i^* = \Phi^{-1}(\delta_i), i = 1,2,\dots,n$, where Φ^{-1} is the inverse of a distribution function of standard normal distribution. Then, under the null hypothesis of no change-point, δ_i^*'s are independent and standard normally distributed random variables. Therefore, there is no change-point if $max_{1\le i\le n}|\delta_i^*| < \sigma$.

The following remark summarizes the above discussion in estimated parameters version.

Remark 1. Let $\hat{\delta}_i = F_{\hat{\theta}}(X_i), i = 1,2,\dots,n$. Then, there are some faults among n observations if and only if the feasible parameter space $\{\theta \in \Theta_0, \hat{T}_n^{**} < \sigma\}$ is empty where $\hat{T}_n^{**} = max_{1\le i\le n}|\hat{\delta}_i - 0.5|$. Let $\hat{\delta}_i^* = \Phi^{-1}(\hat{\delta}_i), i = 1,2,\dots,n$. Therefore, there is no change-point if $max_{1\le i\le n}|\hat{\delta}_i^*| < \sigma$.

Here, the above criteria are extended to AMOC models. Suppose that there is a change at unknown time point t^* and parameters before and after the change are θ_1 and θ_2, respectively. Following [14], assuming θ_1 and θ_2 are known, the least square estimate of t^* is the minimizer of

$$\sum_{i=1}^{t}\left(F_{\theta_1}(X_{(i)}) - \frac{i}{t+1}\right)^2 + \sum_{i=t+1}^{n}\left(F_{\theta_2}(X^*_{(i)}) - \frac{i}{n-t+1}\right)^2, \tag{25}$$

where $X_{(i)}$ is the i-th ordered statistic of $X_1,\dots,X_t$ and $X^*_{(i)}$ is the i-th ordered statistic of $X_{t+1},\dots,X_n$. When θ_1 and θ_2 are unknown, they are replaced with their estimates based on observations of the corresponding segments. Similarly, the SM criterion is the minimizer of

$$SM_t = \max\left(max_{1\le i\le t}|\varepsilon_{i,t}|, max_{t+1\le i\le n}|\varepsilon^*_{i,t}|\right), \tag{26}$$

where $Z_{i,t} = F_{\theta_1}(X_{(i)}) - i/(t+1)$ and $Z^*_{i,t} = F_{\theta_2}(X^*_{(i)}) - i/(n-t+1)$.

The following proposition summarizes the above discussion.

Proposition 5. At AMOC model, the location of fault is estimated by minimizer of SM_t defined by formula (26) where where $Z_{i,t} = F_{\theta_1}(X_{(i)}) - i/(t+1)$ and $Z^*_{i,t} = F_{\theta_2}(X^*_{(i)}) - i/(n-t+1)$.

Here, the Bayesian SM criteria based on credible sets are studied. The recursive relation for FPS constitutes a set-based adaptive filter for online FDI. Indeed, as soon as $FPS(t^*)$ is empty, then there is a fault at t^*. In this way, multiple faults can be detected sequentially.

Under their three simplifier conditions including (i), (ii) and (iii), given as below, Fernández-Cantí *et al* [27] conclude that the Bayesian FPS is the set of all θ's for which the likelihood function is not zero for every x in the support of underlying distribution and as soon as the likelihood function is zero a fault is detected. The simplifier conditions are

(i) non-informative multivariate prior distribution for θ,

(ii) uniform distribution $U(-\sigma,\sigma)$ for ε_t and

(iii) equation-error assumption.

Under conditions (i), (ii) and (iii), a change is detected as soon as the likelihood of $(\varepsilon_1/\sqrt{w_1},\dots,\varepsilon_t/\sqrt{w_t})$ is zero. Indeed, the fault occurs at time t, where the likelihood

$$\prod_{i=1}^{t} U_{(-\sigma,\sigma)}\left(\frac{\varepsilon_i}{\sqrt{w_i}}\right) = U_{(0,\sigma)}(|T_t^*|), \tag{27}$$

is zero where $U_{(-\sigma,\sigma)}(x) = 1/2\sigma$ for $x \in (-\sigma,\sigma)$ and zero otherwise. A similar criterion is based on δ_i's. Then, there is a fault at time t if the likelihood $\prod_{i=1}^{t} U_{(-\sigma,\sigma)}(\delta_i - 0.5) = U_{(0,\sigma)}(|T_t^{**}|)$ is zero.

It is seen that the multiple change-points detection using a SM method is a sequential procedure. To detect all change-points, simultaneously, one can see that the change-point estimates are the optimizers of the following objective function

$$\sum_{i=1}^{R}\sum_{j=\tau_{i-1}+1}^{\tau_i}\left(F_{\hat{\theta}_i}(x_{(j)}) - \frac{j}{l_i+1}\right)^2, \tag{28}$$

where $l_i = \tau_i - \tau_{i-1}$, $\hat{\theta}_i$ is the least squares of θ_i based on observations $x_k, k = \tau_{i-1}+1,\dots,\tau_i$ and $x_{(j)}$ is the j-th ordered statistics among observations $x_k, k = \tau_{i-1}+1,\dots,\tau_i$ (i.e., i-th segment), $i = 1,2,\dots,R$. To find $\tau_i's$

for $i = 1,2,\dots,R$, it is enough to apply the time consuming DP method. The following proposition summarizes the above discussion.

Proposition 6. (*a*)-(*b*) are correct.

(*a*) Under above mentioned Bayesian SM conditions (*i*)-(*iii*), the fault occurs at time t, if and only if the likelihood $U_{(0,\sigma)}(|T_t^*|)$ is zero where $U_{(-\sigma,\sigma)}(x) = 1/2\sigma$ for $x \in (-\sigma,\sigma)$ and $T_n^* = max_{1\le t\le n}|\varepsilon_t/\sqrt{w_t}|$.

(*b*) Using method of (*a*), the multiple faults are detected, sequentially. To estimate all faults, simultaneously, minimize objective function of (28) using a dynamic programming.

2.3.2 DP-based SM. In proposition 6, the DP method presented to solve the multiple faults detection problem. Here, an alternative type of SM approach based on likelihood ratio and applying the DP algorithm is presented. As it was stated in Section 2.3.1, according to [27] and the three simplifier conditions, the likelihood process $\prod_{j=1}^{t} f(x_j,\theta_1), t = 1,2,\dots,n$ becomes zero at $t = t^*$ if t^* is a change-point, at which f_{θ_1} is the density of observations before any change. As soon as a change is detected at k^*, then the next change is searched by $\prod_{j=t^*+1}^{t} f(x_j,\theta_2), t = t^*+1,\dots,n$. This procedure is continued until all changes and change number are estimated. The likelihood for all observations under the null hypothesis of no change-point is $\prod_{j=1}^{n} f(x_j,\theta_1)$ which is zero when there is at least one fault. The likelihood for multiple change-points is $\prod_{i=1}^{R}\prod_{j=\tau_{i-1}+1}^{\tau_i} f(x_j,\theta_i)$ which is applied to detect all faults. To integrate both methods, the likelihood ratio criterion is given by

$$\Lambda = \frac{\prod_{i=1}^{R}\prod_{j=\tau_{i-1}+1}^{\tau_i} f(x_j,\theta_i)}{\prod_{j=1}^{n} f(x_j,\theta_1)} = \prod_{i=1}^{R}\prod_{j=\tau_{i-1}+1}^{\tau_i} \frac{f(x_j,\theta_i)}{f(x_j,\theta_1)}, \tag{29}$$

which is large if the null hypothesis of no fault is wrong and it is small if the null hypothesis is correct. The logarithm of Λ, i.e., $\delta = \log(\Lambda)$ is given by

$$\delta = \sum_{i=1}^{R}\sum_{j=\tau_{i-1}+1}^{\tau_i}\{\log(f(x_j,\theta_i)) - \log(f(x_j,\theta_1))\}. \tag{30}$$

Here, $\theta_i, i = 1,2,\dots,R$ are estimated by their maximum likelihood estimates (MLE). When the priors are uninformative, then the MAP estimates are MLEs. Replacing $\log\left(f(x_j,\theta_i)\right)$ with the $\left\{logf\left((x_j,\theta_i)\right) - \log\left(f(x_j,\theta_1)\right)\right\}$ in Proposition 1, a DP based method is obtained for δ, the log-likelihood ratio test statistic. Next, the limiting local behavior of δ, under the null hypothesis of no change-point, is studied. To this end, suppose that $\theta_i = \theta_1 + w_i\gamma$ and $\gamma \to 0$ for some $w_i > 0$. Then, δ is approximated by

$$\delta \approx \sum_{i=1}^{R} w_i \sum_{j=\tau_{i-1}+1}^{\tau_i} \frac{\partial}{\partial\theta_1}\left(\log\left(f(x_j,\theta_1)\right)\right). \tag{31}$$

For some values like $w_i = 1/R$, as $n \to \infty$, using the central limit theorem, it is seen that

$$n^{-0.5}R\delta \to^{d} N(0, I(\theta_1)), \tag{32}$$

where the $I(\theta_1)$ is the Fisher information computed in θ_1 (see [24]). Using the Skutsky theorem (see [24]), it is seen that

$$n^{-0.5} I(\hat{\theta}_1)^{-0.5} R\delta \to^{d} N(0,1). \tag{33}$$

The above approach can be extended, using marginal likelihood function [24]. One can see that in the presence of multiple change-points, then the marginal log-likelihood is partitioned as follows:

$$\sum_{i=1}^{R}\sum_{j=\tau_{i-1}+1}^{\tau_i}\int \log f_{\theta_i}(x_j)\pi_i(\theta_i)d\theta_i. \tag{34}$$

Let $Q^*(\tau_{i-1},\tau_i) = \sum_{j=\tau_{i-1}+1}^{\tau_i}\int \log f_{\theta_i}(x_j)\pi_i(\theta_i)d\theta_i$. Then, (34) reduces to a DP problem as,

$$\max_{\{\tau_i\}_{i=1,\dots,R}} \sum_{i=1}^{R} Q^*(\tau_{i-1},\tau_i). \tag{35}$$

Thus, again, the DP algorithm is applied to derive the true change-points estimates.

The following proposition summarizes the above discussion.

Proposition 7. (*a*)-(*d*) are correct.

(*a*) The null hypothesis of no fault is rejected if and only if values of δ of exact formula (30) or approximated formula (31) is large.

(*b*) Replacing $\log\left(f(x_j,\theta_i)\right)$ with the $\left\{logf\left((x_j,\theta_i)\right) - \log\left(f(x_j,\theta_1)\right)\right\}$ in Proposition 1, a DP estimate of δ is obtained.

(*c*) Under the null hypothesis of no fault, then $n^{-0.5}I(\hat{\theta}_1)^{-0.5}R\delta \to^{d} N(0,1)$, where the $I(\theta_1)$ is the Fisher information computed in θ_1.

(d) The DP algorithm of Proposition 1 may be applied by replacing $Q(\tau_{i-1},\tau_i)$ with $Q^*(\tau_{i-1},\tau_i)$ of formula (34).

2.4 Probability of fault. The probability of a fault is a useful indicator for alarming the existence of a change. This tool can be used as an EWS [37], as studied in Section 4.3.4. In this section, using the BMS and DP approaches, the probability of the fault is derived for the multiple change-points case. Then, the DP approach is applied to extend the proposed method to multiple change-points cases. The BMS applies the Bayesian inferential framework for model selection. This method can be integrated to Bayesian model averaging and applied in cases where there are uncertainties in models and parameters, both of which should be estimated. Traditionally, the BMS method addresses model uncertainty regression problems by considering different subsets of independent variables. Then, the most probable subset of independent variables (models) based on posterior probabilities assigned to each model is selected. Indeed, this method is a type of Bayesian variable selection. There are many practical situations where choosing a unique model is a risky work. As a motivating example, consider the data concerning cancer of the esophagus. Some modelling work should be done to make sure the existence of cancer is correctly detected. As another example, consider that to specify the correct model to be developed for political science research, then, the uncertainties can be high. However, this paper applies the BMS method in a different way. Here, it is not interested to find the best subset of independent variables, instead, different locations of change-points (for example, in AMOC case) defines different potential volunteers for the correct model. Indeed, in this section, the change-point detection is interpreted as selection of a piecewise model. In this way, the Bayesian posterior probability of the fault is derived. The fault probabilities are given in three types of faults, including sharp change and gradual change in an offline manner, and the online case.

(a) Sharp change. Suppose that $\boldsymbol{X}_i, i = 1,2, \dots, \tau_1$ come from density $f_{\boldsymbol{\theta}_1}$ and the remaining observations $\boldsymbol{X}_i, i = \tau_1 + 1, \dots, n$ have common density $f_{\boldsymbol{\theta}_2}$ where $\boldsymbol{\theta}_1 \neq \boldsymbol{\theta}_2$. Consider the two priors $\pi_i(\boldsymbol{\theta}_i)$ for $\boldsymbol{\theta}_i, i = 1,2$. Next, consider $n-1$ models $M_j: \tau_1 = j, j = 1,2, \dots, n-1$. Under the j-th model, allocating equal prior probability $1/(n-1)$ to every model, the marginal log-likelihood based on BMS technique is given $l_j = \sum_{i=1}^{j} \int \pi_1(\boldsymbol{\theta}_1)\, log f_{\boldsymbol{\theta}_1}(\boldsymbol{x}_i)\, d\boldsymbol{\theta}_1 + \sum_{i=j+1}^{n} \int \pi_2(\boldsymbol{\theta}_2)\, log f_{\boldsymbol{\theta}_2}(\boldsymbol{x}_i)\, d\boldsymbol{\theta}_2$. Then, the normalized l_j gives the probability of fault $p_j = l_j / \sum_{h=1}^{n-1} l_h$. Practically, the transformation $\mathrm{p}_\mathrm{j}^* = (p_j - \min(p_j))/(\max(p_j) - \min(p_j))$ gives a better representation for the probability of the fault.

Next suppose that there are $R-1$ change-points. The log-posterior function is proportional to

$$\sum_{i=1}^{R} \sum_{j=\tau_{i-1}+1}^{\tau_i} \log\{f(\boldsymbol{x}_j, \boldsymbol{\theta}_i)\pi_i(\boldsymbol{\theta}_i)\}. \tag{36}$$

Under models $M_j: \tau_k = j$, (that is τ_k is known value of j) for some fixed k, the maximized format of (36) is

$$\sum_{i=1}^{R} \sum_{j=\hat{\tau}_{i-1}+1}^{\hat{\tau}_i} \log\{f(\boldsymbol{x}_j, \widehat{\boldsymbol{\theta}}_i)\pi_i(\widehat{\boldsymbol{\theta}}_i)\}, \tag{37}$$

at which $\hat{\tau}_i, i = 1, \dots, k-1, k+1, \dots, R$ are MAP estimates of τ_i and $\hat{\tau}_k = \tau_k$ is known. Therefore, the DP algorithm of Section 2.1 is applied to find R-2 change-points and again the posterior fault probability of $\tau_k = j$ is given by p_j. Similarly, the joint probability of faults $\tau_{k_1} = j_1 \,\&\, \tau_{k_2} = j_2$ is obtainable under the model $M_{j_1 j_2}: \tau_{k_1} = j_1 \,\&\, \tau_{k_2} = j_2$. Alternative approaches to compute the probability of the fault can be found in [2, 29].

(b) Gradual change. All the previous FDI models are abrupt change type at which a parameter (some parameters) of interest changes suddenly at some unknown time points. For example, an abrupt change may occur in the coefficients of the system sensors. However, the gradual change type may happen slowly, for example, as the change in mean or variance of statistical distribution of a specified parameter. Indeed, the mean or variance is a function of the considered system parameter which is not fixed. As time goes on, the system parameter gradually changes which leads to a gradual change of the mean or variance. For the sake of problem simplicity, random variables are assumed to be univariate. Multivariate cases are argued similarly. Following [38], suppose that $x_i, i = 1,2, \dots, n$ are independent random variables where x_i is distributed as $N(\theta_i, \sigma^2)$ where $\theta_i = \theta + d_i\delta$ at which $d_i = c_i, i = 1,2, \dots, \tau$ and is one for $i = \tau + 1, \dots, n$. Here, c_i is an increasing sequence of positive numbers such that $c_i \in (0,1)$. This is a gradual type of change point where θ_i starts to get large with slope δ and intercept θ before the change time τ. Here, c_i can be interpreted as time varying characteristic measures of a mechanical component. However, after τ, it changes to the higher level $\theta + \delta$. Here, with loss of generality, it is assumed that the magnitude of change δ is positive. In these cases, the effects of change on the system performance are not considerable at the early stages. However, it may lead to a catastrophic effect on the system as it gradually increases. Accordingly, the fault probability can be used as an EWS, to alarm a gradual change on the system, while the traditional FDI techniques fail to detect it as early as possible, until the change effect is obvious. These types of EWS are widely used in detecting financial crises as well as in the quality control field.

(c) Online fault probability. Here, the online fault probability algorithm, based on a new format of BMS, is proposed for EWS purposes. Suppose that $\mathrm{x}_1, \dots, \mathrm{x}_\mathrm{n}, \mathrm{n} \geq 1$ is observed and there is no change among them.

They come from $N(\theta_1, \sigma^2)$ and observation x_{n+1} is suspected to change its mean such that it has $N(\theta_2, \sigma^2)$ where $\theta_2 \neq \theta_1$. To run the BMS, the proposal model is $M_n: \tau_1 = n + 1$. Thus, the log-likelihood is given by $l_n = \sum_{i=1}^{n} f_{\theta_1,\sigma^2}(x_i)$. Assuming equal priors for all models, then, the posterior probability of the probable model, which can be interpreted as the probability of existence of the change at the n^{th} time point is proportional to l_n. For the sake of distinct visualization, the transformation $(l_n - \min_{1\leq i\leq n} l_i)/(\max_{1\leq i\leq n} l_i - \min_{1\leq i\leq n} l_i)$ is given as the probability of the fault. The unknown parameters θ_1, σ^2 are replaced with their sample estimates.

3 Alternative methods.

The high computational expense and time consumption are the main reported downsides of the DP-based methods, regardless of accuracy in performance. To overcome this difficulty, an approximate method is proposed in this section. Using the notation of Section 2.1, four iterative procedures are presented here to obtain the MAP estimates of location of faults under the above-mentioned types of prior distributions. These methods are referred to as iterative methods in the next sections.

(a) Poisson prior. Suppose that $l_1, \dots, l_R$ have Poisson distributions with time varying intensity parameter γ_i and consider an uninformative prior for $\boldsymbol{\theta}_i$, the posterior of $l_1, \dots, l_R$ is proportional to

$$\begin{aligned} &\propto \prod_{i=1}^{R} (e^{\bar{y}_{i+}})^{l_i} \prod_{i=1}^{R} \frac{e^{-\gamma_i} \gamma_i^{l_i}}{l_i!}, \\ &\propto \prod_{i=1}^{R} \frac{(\gamma_i e^{\bar{y}_{i+}})^{l_i}}{l_i!}, \end{aligned} \tag{38}$$

which has the Poisson distribution with parameter $(\gamma_i e^{\bar{y}_{i+}})$. The MAP estimate of l_i is the integer value of $\gamma_i e^{\bar{y}_{i+}}$, i.e., $[\gamma_i e^{\bar{y}_{i+}}]$ (the mode of Poisson posterior distribution). However, since the length of mean $\bar{\boldsymbol{y}}_{i+}$ depends on l_i, thus, the $\bar{\boldsymbol{y}}_{i+}$ plays the role of a missing value. To remove this difficulty, three different iterated procedures are proposed as follows.

Procedure1) In iteration k, the posterior l_i^k has posterior Poisson law with intensity parameter $(\gamma_i e^{\mathbf{y}_{i+}^{k}})$, where $\mathbf{y}_{i+}^{k}$ is calculated in iteration $k-1$. By repeating this procedure K times, the average of $l_i^k, k = 1,2, \dots, K$ is considered as a Bayesian estimate of actual l_i. The simulation results show that by choosing suitable hyper-parameters γ, the change-point estimates are accurate. In a simulation study, $\gamma_i's$ are obtained using the trial and error method. However, in practical situations, a response surface methodology may be used to derive a specified formula for γ_i as a function of l_i.

Procedure 2) As it is stated, the posterior is a Poisson distribution with parameter $(\gamma_i e^{\bar{y}_{i+}})$. Assuming all γ_i's be equal to γ. Then, one can show that the maximum likelihood estimation of γ equals to

$$\hat{\gamma} = \frac{1}{n} \sum_{i=1}^{R} e^{\bar{y}_{i+}}. \tag{39}$$

First, using a pre-determined segmentation, $\hat{\gamma}$ is estimated and then the change-point locations are derived. Again, $\hat{\gamma}$ is re-estimated and locations of change-points are derived again. This iterated procedure is repeated until it converges. As well as, one can show that the EM algorithm estimate of γ is given by (40), as follows.

$$\hat{\gamma}_{t+1} = \frac{n}{\sum_{i=1}^{R} E_{\hat{\gamma}_t}(e^{\bar{y}_{i+}})}, \tag{40}$$

where $E_{\hat{\gamma}_t}(e^{\bar{y}_{i+}})$ is the expectation of $e^{\bar{y}_{i+}}$at where l_i has a Poisson distribution with parameter $\hat{\gamma}_t$ and the observations y_{ij}'s of the i-th partitions (i.e. $j = \tau_{i-1} + 1, \dots, \tau_i$) are kept fixed.

Procedure 3) For implementing the EM algorithm, the logarithm of posterior is given by

$$\sum_{i=1}^{R} l_i \log(\gamma_i) + l_i \bar{\boldsymbol{y}}_{i+} - \log(l_i!). \tag{41}$$

Take the expectation of (41) where $l_i = l_i^k$. Then,

$$\sum_{i=1}^{R} l_i \log(\gamma_i) + l_i E_{l_i^k}(\bar{\boldsymbol{y}}_{i+}) - \log(l_i!). \tag{42}$$

Optimizing (42) with respect to l_i yields an update estimate of l_i as follows

$$l_i^{k+1} = \left[\gamma_i e^{E_{l_i^k}(\bar{y}_{i+})}\right]. \tag{43}$$

To perform this iterative method, the hyper-parameters γ_i are substituted by its empirical estimate which is the MAP estimates l_i^k.

(b) Bernoulli prior. Under uninformative prior for $\boldsymbol{\theta}_i$, the posterior of l_i is proportional to

$$\begin{aligned} &\propto \prod_{i=1}^{R} (e^{\bar{y}_{i+}})^{l_i} \prod_{i=1}^{R} (1-\pi)^{l_i}, \\ &\propto \prod_{i=1}^{R} e^{-\varphi l_i}, \end{aligned} \tag{44}$$

where $\varphi = -\ln(1-\pi) - \bar{\boldsymbol{y}}_{i+}$. To make sure that φ is positive, it is enough to suppose that $\pi > 1 - e^{\bar{y}_{i+}}$. If $\bar{\boldsymbol{y}}_{i+}$ is small, then $\pi > \bar{\boldsymbol{y}}_{i+}$. Then, the posterior has exponential distribution with parameter $1/\varphi$. This iterated procedure is presented in another way, as follows.

Procedure 4) The posterior is proportional to the

$$\prod_{i=1}^{R} \exp\{l_i \bar{\boldsymbol{y}}_{i+} + \log p_i(l_i)\}. \tag{45}$$

The MAP estimation of l_i is found by

$$\bar{\boldsymbol{y}}_{i+} + \frac{\partial}{\partial l_i} \log p_i(l_i) = 0. \tag{46}$$

For the prior of class (ii), then (46) is reduced to the

$$\bar{\boldsymbol{y}}_{i+} + \log(1 - \pi_i) = 0. \tag{47}$$

Hence,

$$\sum_{j=\tau_{i-1}+1}^{\tau_i} y_{ij} + (\tau_i - \tau_{i-1}) \log(1 - \pi_i) = 0. \tag{48}$$

Solving iteratively, (47) leads to estimators for $\tau_i, i = 1,2, \dots, R$ and R. When, π_i is small, then $\log(1 - \pi_i) \approx \pi_i$ and (48) is reduced to

$$\sum_{j=\tau_{i-1}+1}^{\tau_i} y_{ij} + (\tau_i - \tau_{i-1})\pi_i = 0. \tag{49}$$

Notice that, when $\pi_i \to 0$, then the Bernoulli prior is close to the Poisson prior with parameter π_i (see [39]).

4 Simulations.

In this section, using numerical simulations, change-point detectability of the proposed methods, including DP-MAP, iterative and SM, are verified, and their performances are compared. Then, three practical dynamic systems, including first order time delay, second order uncertainty, and integration, with different fault scenarios are given and it is aimed to accurately detect the mentioned faults, using the proposed methods. To summarize the proposed multiple FDI algorithm, the flowchart is given in Figure 1, through which the proposed methods are numerically analyzed.

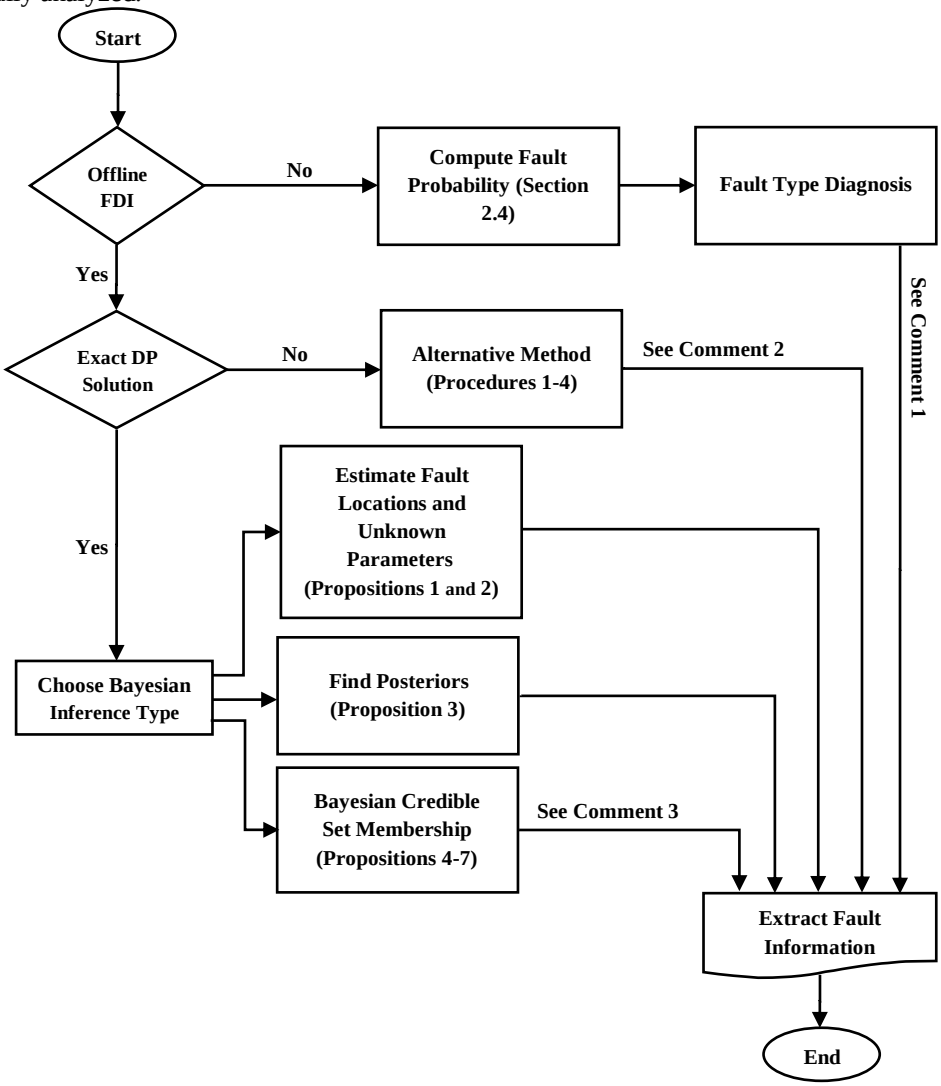


Figure 1: Illustrative flowchart of overall DP-based methodologies in Bayesian multiple FDI. Comment 1: The fault type includes abrupt and gradual faults which are characterized as sudden and slow changes, respectively, in the observed/estimated parameters. Comment 2: Procedure 1 is used for Poisson prior-iteration of posterior-

averaging, Procedure 2 is used for Poisson prior-EM approximation of posterior with simplified assumption, Procedure 3 is used for Poisson prior-EM approximation of posterior using iteration, Procedure 4 is used for approximating posterior under the Bernoulli prior. Comment 3: Proposition 4 computes $\widehat{\text{FPS}}$, $\widehat{\text{FPS}}^*$(n) and ARTA(1), Proposition 5 is applicable in AMOC model and estimates the fault location, Proposition 6 is applicable in AMOC model and detects sequentially multiple faults, Proposition 7 is applicable for simultaneous multiple faults.

4.1 Illustrative examples. In this section, four different examples are given in which, for the sake of explanation, the state is generated randomly and is assumed to be known. On the other hand, the unknown random noise content is added to the measured state.

Example 1: DP-MAP method. Here, the DP-based method of proposition 1 is applied to two cases of epidemic and multiple change-point problems.

(*a*) Epidemic change-point patterns occur in many fields such as epidemiology and medicine studies of influenza disease mortality. This pattern is a special case of multiple change-point problems. In epidemic change-point patterns there are three segments of observations. The distribution of observations of the middle segment differs from the first and the last segments. When the length of the middle segment is small, then the detection of change may be difficult, and methods of change-point analysis can be misleading [40]. To describe more, let $y_t = \beta_t x_t + \varepsilon_t, t = 1, \dots, n$ represents the measurement equation, where ε_t is a sequence of independent random variables, having common distribution $N(0, \sigma^2 = 0.01)$ and β_t has the following epidemic pattern

$$\beta_t = \begin{cases} b_1, & 0 \le t \le \tau_1 = 40 \\ b_2, & 41 \le t \le \tau_2 = 65, \\ b_1, & 66 \le t \le 100 \end{cases}$$

and x_t are the sorted sequence of 100 uniform observations on (0,1). Here, the DP method for change detection in β_t based on y_t, is studied. The hyper-parameter $\lambda = 0.2$ is fixed. Table 1 gives various estimates of τ_1 and τ_2 based on different values of b_1, b_2 and σ. It is seen that as $\sigma \to 0$ or $b_2 - b_1$ gets large then the method performs well.

Table 1: Various estimates of τ_1 and τ_2.

b_1	b_2	σ	$\hat{\tau}_1$	$\hat{\tau}_2$
0.2	-1	0.1	41	66
0.2	-1	1	49	66
1	3	0.01	41	65
1	3	0.1	42	66
1	3	1	43	64
0.1	0.2	0.01	43	68
0.1	0.2	0.1	44	71
0.1	0.2	1	49	72

It is obvious that more noise content on data and small change size, lead to less accurate change-point detection. Significantly, as $\sigma \to 0$ or $b_2 - b_1$ gets large, then the method detects the change more accurately. It is worth noting that, after detection of the change time, the suitable statistical methods such as the least square algorithm can be adopted in the diagnosed segment to estimate change size, which is not considered in this example.

(b) Multiple change-point patterns occur in observed data in many fields, such as genomics and econometrics. In contrast to epidemic patterns, in the multiple one, there are several observation segments and each segment has a different parameter distribution [41]. This makes the change-point detection more challenging. Here, suppose that the multiple fault patterns have occurred in β_t as follows

$$\beta_t = \begin{cases} b_1, & 0 \le t \le \tau_1 = 30 \\ b_2, & 31 \le t \le \tau_2 = 48 \\ b_3, & 49 \le t \le \tau_3 = 70 \\ b_4, & 71 \le t \le \tau_3 = 100 \end{cases}$$

Indeed, there are three change-points and four different segments. Using proposition 2, Table 2 gives the estimated values of $\tau_i, i = 1,2,3$.

Table 2: Various estimates of $\tau_i, i = 1,2,3$.

b_1	b_2	b_3	b_4	σ	$\hat{\tau}_1$	$\hat{\tau}_2$	$\hat{\tau}_3$
0	-1	0.2	0.6	0.1	31	49	70
0	-1	0.2	0.6	1	30	53	74
-1	3	0.1	0.15	0.01	31	46	72

-1	3	0.1	0.15	0.1	34	49	71
1	2	3	4	0.01	31	49	71
-0.1	-0.2	-0.3	-0.4	0.01	35	48	70
-0.1	-0.2	-0.3	-0.4	0.1	38	50	72

Similar to the epidemic change-point case, again, it is shown that when the locations of change-points are far from each other or when the σ is small, then the method performs well. However, clearly, in most circumstances, τ_2 and τ_3 are slightly estimated better than τ_1. This phenomenon is natural, because of the delay in diagnosing the new segments. As suggested in [29], it is better to apply a two-way FDI method to resolve this problem. That is, in addition to the start-to-end data analysis, conducting the change-point detection from the end of the data sequence to the starting point and finally combine the results from both ways. It is worth mentioning that this correction is applicable in an offline manner. However, it is obvious that the change-points are almost accurately estimated considering different signs of changes with small and large sizes and noise contents, using the proposed DP method.

Example 2: SM method. In this example, the SM method is applied to consider three different cases including null hypothesis checking, AMOC and multiple faults. In the null hypothesis case, let $X_1, \dots, X_{150}$ be a sequence of independent and identically normally distributed values with mean $\theta = 0.35$ and standard deviation σ. Next, suppose that θ is unknown and let $\Theta_0 = \{0.05(0.01)0.5\}$. For both cases $\sigma = 0.1$ and $\sigma = 0.3$, then $\widehat{FPS} = \widehat{FPS}^*$ of proposition 4 (part a, b) are $\{0.35(0.01)0.38\}$ and $\{0.29(0.01)0.44\}$, respectively, which are not empty. This fact indicates there is no fault among $n = 150$ observations. For another example, let $X_1, \dots, X_{877}$ be a sequence of independent variables where the first 354 observations are distributed as $N(0,0.25)$, the next 523 observations come from $N(1,0.25)$. The time series of $SM_t, t = 1, \dots, 877$ of proposition 5 are plotted in Figure 2. It is seen that SM_t is minimized at $t = 354$. Time series plot of $\hat{\rho}_t$, the coefficient of regression of SM_t on SM_{t-1} (see Proposition 4 part c) is given in Figure 3. It also shows the change-point, accurately.

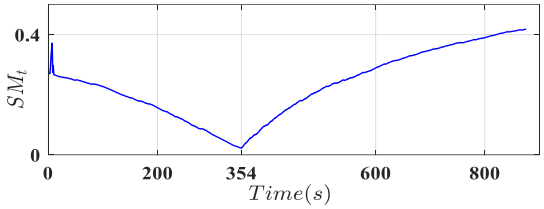


Figure 2: SM_t times series.

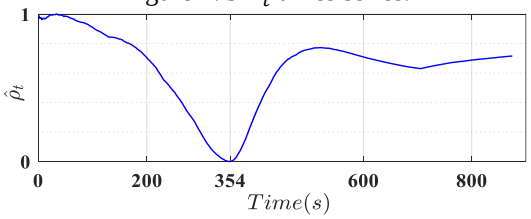


Figure 3: Time series $\hat{\rho}_t$.

As other example in the multiple change-points case, suppose that there exist 599 observations such that the first 111 observations come from normal distribution with zero mean and standard deviation 0.1. The 123 middle observations are normally distributed with mean 1 and standard deviation 0.1 and the remaining come from normal distribution with the same standard deviation and mean 2.5. Using Propositions 6 and 7, time series of T_t is plotted in Figure 4 which shows the second change-point at $t = 111 + 123 = 234$. Similarly, the first and third change-points will be detected repeatedly. Here, the multiple change-points were detected by binary-segmentation (find a change points, segment the data stream to two parts and search for other change-points until, no change point is remained) which is an alternative method to DP based method. This type of change point detection is useful for online FDI problem by using Proposition 6.

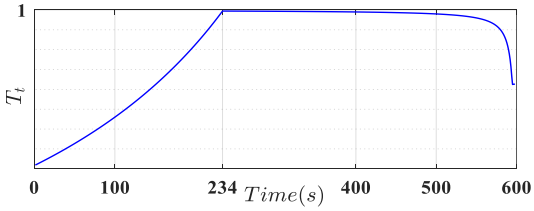


Figure 4: Time series plot of T_t.

Example 3: Alternative method. Here, the alternative (iterative) method ability to detect the change-point is analyzed in a time series setting. Consider the first order autoregressive AR(1) as follows $x_t = \beta_t x_{t-1} + \varepsilon_t, t = 1, \dots, n$. Here, $\beta_t = b_1$ for $t = 1,2, \dots, \tau$ and $\beta_t = b_2$ for $t = \tau + 1, \dots, n$ and $n = 100$. Also, ε_t's are independent variables having common distribution $N(0, \sigma^2)$. Under Poisson prior, Procedure 1 of Section 3 is applied and by trial and error effort, it is seen that the suitable hyper-parameter of Poisson distribution λ is selected as 10. Number of iterations is 100. Using procedure 1, Table 3 gives the values of τ estimation. It is seen that results of Procedure 2 are the same.

Table 3: Iterative τ-estimates.

τ	b_1	b_2	σ	$\hat{\tau}$
10	1	3	0.01	10
10	1	3	0.1	12
50	1.2	1	0.01	51
50	1.2	1	0.1	54
90	0.5	0.75	0.01	91
90	0.5	0.75	0.1	92

Evidently, the iterative method is able to detect change-point in all cases, including small and big change size with different noise content. Also, the change-point has been set at the beginning, middle and end of the time series to accurately evaluate the proposed method. It should be noted that as the noise content on the time series is increased, the change-point detection accuracy is slightly decreased. As stated before, an advantage of this method is its simplicity in implementation.

Example 4: Probability of fault. Here, the fault probability using the relations in Section 2.4 is computed for offline and online settings.

(a) Abrupt change. Here, following (37), consider the case that f_{θ_1} is $N(0,0.01)$, $f_{\theta_2} = N(0.1,0.01)$. The actual location of change is 250 among 1000 observations. The priors of θ_1 and θ_2 are flat improper uniform distributions on the real line. Figure 5 shows the probability of the fault $\mathrm{p}_j^*, j = 1, \dots, 999$ derived by the BMS approach.

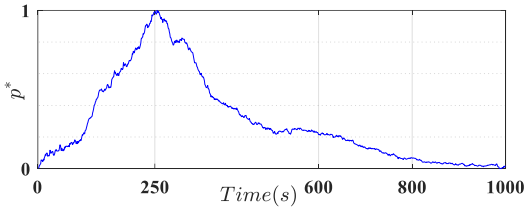


Figure 5: BMS probability p^* for abrupt change.

Figure 5 shows an inverse V-shaped plot. Considering the BMS procedure, the V-shape is justified as follows. Although the actual change-point is 250, the BMS procedure checks, for instance, if the change-point is 100, in its 100th stage. Based on this assumption, the overall likelihood of the total sample (1000 observations) is computed. Then, the first 100 observations come from f_{θ_1} and the remaining 900 observations come from f_{θ_2}. By the way, the actual density for the 101st to 250th observations is f_{θ_1}, which is mistakenly assumed to be f_{θ_2}. This makes the overall likelihood become small. At the points close to the actual change-point the overall likelihood becomes larger. This is the case similarly after the actual change-point occurs. Accordingly, the fault probability is inverse V-shaped, and the highest probability is computed at the change-point, which provides suitable visual interpretation. Some remarks worth noting are addressed as follows.

(*i*) The increasing part of Figure 5 provides indication that a probable change will occur in the near future. This is the EWS property of the probability of a fault which avoids the surprising consequences of abrupt changes.

(*ii*) From the engineering perspective, it is a proper monitoring tool, on which basis the safety scheme may be designed to avoid, tolerate, transfer or reduce the future risks and losses.

(*iii*) The probability after the change-point decreases as the observations after this moment have no influence.

(*iv*) As soon as a change is detected, all changes will be detected, sequentially and repeatedly.

(*vi*) In Figure 5, the left part of the plot indicates that there is a possible change in future; however, a natural question is when it will occur. The answer is; as soon as the increasing part starts to decrease and behave conversely.

(b) Gradual change. Here, following Section 2.4, suppose that, the actual change is 250 among 1000 observations. Let $\theta = 0, \delta = 0.1, \sigma = 0.1$ and c_i's are the 250 sorted observations of standard uniform

distribution defined on (0,1). Again, the priors of all unknown parameters are an improper uniform distribution on (0,1). Figure 6 shows the EWS probability of change.

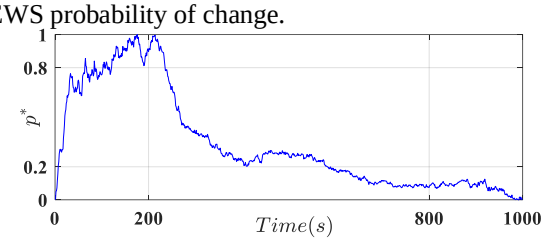


Figure 6: BMS probability p^* of gradual change.

The maximum point of Figure 6 indicates that the location of change occurs around the time step 250, although in this case, diagnosing the true change is more challenging than for the abrupt change case.

(*c*) *Online probability of fault*. In this section, considering $\theta_1 = 0, \theta_2 = 0.5$ and $\sigma = 0.1$, the fault probability is computed in an online framework, illustrated in Figure 7. It is seen that after 250, the probability of the fault starts to increase indicating a change occurrence at 250.

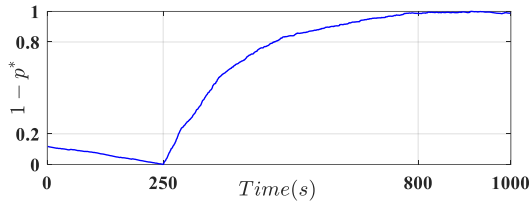


Figure 7: BMS probability (p^*) of online change.

4.2 Performance comparisons for Non-normal distributions. In previous sections, the performances of the Bayesian DP-MAP method, iterative and SM were compared in normal distributions in different types of change-point scenarios. However, sometimes, other distributions (which do not belong to the exponential family of distributions) such as the Rayleigh distribution, outperforms the normal distributions and are better fitted to variables that exist in the applied problem at hand, e.g. the FDI problem of offshore wind turbine analysis [31]. The Kurtosis measure has had an influential effect in change-point detection. This is why, the FDI in a general class of distributions is considered in this paper. In this section, the performances of the three Bayesian methods are compared, in Table 4, following Proposition 3, within leptokurtic (including log-normal, Laplace, logistic Rayleigh, Weibull, Pareto) and platykurtic distributions involving uniform law. These types of distributions do not belong to the exponential family and appear in several fields of engineering, e.g. complex power systems [42], offshore wind turbine [31] and software reliability [43]. Here, it is assumed that $X_1, \dots, X_{250}$, $X_{251}, \dots, X_{375}$, $X_{376}, \dots, X_{500}$ are four segments with different distributions. For the iterative method, the best choice for λ of the Poisson distribution is 10.

After estimating the change points, estimation of θ_i in each segment can be easily done using the maximum likelihood method, which is ignored here. The estimate errors of each method are given by $e_1 = |\hat{\tau}_1 - 250|$ and $e_2 = |\hat{\tau}_2 - 375|$. These values are given in Table 5 for each distribution.

The series of various e_1 and e_2 based types of distributions are illustrated in Figures 8 and 9, respectively. It is seen that the worst results, for the three methods, (except for the Iterative method) belong to the Pareto distribution. Again, the three methods work well, in most of the cases.

Table 4: Performances of three methods: non-normal cases

Distribution	Notation	θ_i, i: 1,2,3	$\hat{\tau}_i, i$: 1,2		
			DP-MAP	Iterative	SM
Log-normal	$LN(\theta, 0.5)$	$\begin{bmatrix}-1\\0.5\\2\end{bmatrix}$	$\begin{bmatrix}250\\375\end{bmatrix}$	$\begin{bmatrix}252\\375\end{bmatrix}$	$\begin{bmatrix}251\\375\end{bmatrix}$
Laplace	$Laplace(\theta, 2)$	$\begin{bmatrix}-2\\0.25\\-0.5\end{bmatrix}$	$\begin{bmatrix}254\\380\end{bmatrix}$	$\begin{bmatrix}255\\381\end{bmatrix}$	$\begin{bmatrix}253\\377\end{bmatrix}$
logistic	$L(\theta, 3)$	$\begin{bmatrix}1\\0.25\\0.5\end{bmatrix}$	$\begin{bmatrix}251\\380\end{bmatrix}$	$\begin{bmatrix}255\\382\end{bmatrix}$	$\begin{bmatrix}255\\376\end{bmatrix}$

Rayleigh	$R(\theta,2)$	$\begin{bmatrix}1\\0.25\\0.75\end{bmatrix}$	$\begin{bmatrix}253\\379\end{bmatrix}$	$\begin{bmatrix}254\\381\end{bmatrix}$	$\begin{bmatrix}253\\378\end{bmatrix}$
Weibull	$Weibull(2,\theta)$	$\begin{bmatrix}0.25\\0.9\\0.5\end{bmatrix}$	$\begin{bmatrix}252\\378\end{bmatrix}$	$\begin{bmatrix}256\\371\end{bmatrix}$	$\begin{bmatrix}253\\381\end{bmatrix}$
Pareto	$P(2,\theta)$	$\begin{bmatrix}0.5\\0.75\\1.25\end{bmatrix}$	$\begin{bmatrix}257\\380\end{bmatrix}$	$\begin{bmatrix}254\\376\end{bmatrix}$	$\begin{bmatrix}255\\383\end{bmatrix}$

Table 5: Comparisons of three methods: non-normal cases

		Distribution					
	method	Log-Normal	Laplace	Logistic	Rayleigh	Weibull	Pareto
e_1	DP-MAP	0	4	1	3	2	7
e_2		0	2	4	4	3	5
e_1	Iterative	2	5	5	4	3	4
e_2		0	3	3	3	1	3
e_1	SM	1	3	5	3	3	5
e_2		0	2	2	3	4	8

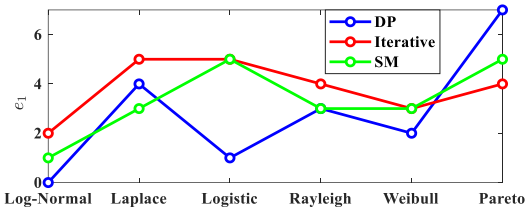


Figure 8: Values of e_1.

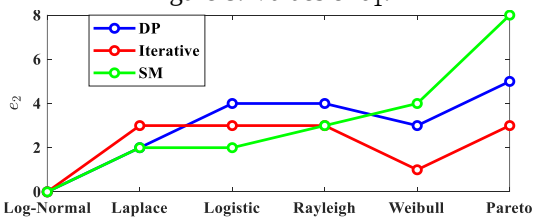


Figure 9: Values of e_2.

4.3 Dynamic system FDI examples. In this section, to practically evaluate the proposed FDI methods, first and second order dynamic systems are presented, and it is aimed to detect the injected sensor faults, whose measurements are contaminated with noise contents. The dynamic systems are the wind turbine generator and pitch mechanism, which are very prone to be faulty, due to harsh offshore operation. Also, the proposed methods are evaluated on an integrated system, in which it is challenging to detect sensor faults [9]. Finally, the probability of the fault is investigated as an early warning alarm for the existence of gradual faults in a specified dynamic system.

4.3.1 First order time delayed system. The wind turbine generator is an electrical system that is used to produce electrical power and adjust the generator load torque to maximize captured power. The electrical generator torque is measured and fed into the controller that controls the wind turbine [3]. So, occurrence of any sensor fault will lead to undesirable control and, consequently, less produced power. This provides motivation that the presence of faults on the generator torque sensor must be detected. The generator is modelled as a first order dynamic system with time delay as,

$$\dot{T}_g(t) = -a_g T_g(t) + a_g T_{g,ref}(t - t_{g,d}),$$

where, T_g is generator torque, $T_{g,ref}$ is the reference generator torque load, which is requested from the generator by controlling its current, and $t_{g,d} = 10\,ms$ is the communication delay. Also, $a_g = 1/\tau_g$, where $\tau_g = 20\,ms$ is the system time constant. Also, the generator torque sensor is modelled as,

$$T_{g,s} = \alpha_{T_g} T_g + \nu_{T_g},$$

where, $\nu_{T_g} \sim N(0, \sigma_{T_g}^2)$ is measurement noise and σ_{T_g} is standard deviation. $\alpha_{T_g} = 1$ and $\alpha_{T_g} \neq 1$ are measurement coefficients in fault free and faulty situations, respectively. Also, it is assumed that $\alpha_{T_g} \neq 0$, which means total loss of sensor outputs is avoided. Although, if the physical redundancy is provided, this

assumption can be relaxed. It should be noted that the presence of input time delay may lead to inaccurate FDI using residual-based model based FDI methods [44]. The input generator torque is selected as, $T_{g,ref} = 200(\sin(10t) + 1)\ Nm$. In Figure 10, the effects of sensor fault and noise are illustrated in which, $\alpha_{T_g} = 0.5$ for $0.5 \leq t \leq 0.8\ s$ and $\alpha_{T_g} = 1$ for other time steps, and $\sigma_{T_g} = 9\ Nm$.

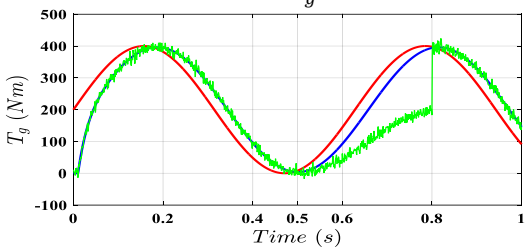


Figure 10: Generator reference torque $T_{g,ref}$ (red line), produced generator torque T_g (blue line), and measured generator torque $T_{g,s}$ (green line).

Now, in Table 6, for different fault scenarios, the detected faults are summarized using the proposed methods to evaluate their fault detectability.

Table 6: Generator sensor FDI results.

			FDI method								
		Noise	DP-MAP			Iterative			SM		
	Fault scenario	σ_{T_g}	$\hat{\tau}_1$			$\hat{\tau}_1$			$\hat{\tau}_1$		
AMOC	$\alpha_{T_g} = 0.1$ for $\tau_1 = 0.8 \leq t$	0.09	0.84			0.83			0.81		
		0.9	0.86			0.84			0.81		
		9	0.86			0.84			0.83		
	$\alpha_{T_g} = 0.5$ for $\tau_1 = 0.5 \leq t$	0.09	0.50			0.50			0.50		
		0.9	0.50			0.50			0.50		
		9	0.54			0.52			0.51		
	$\alpha_{T_g} = 0.9$ for $\tau_1 = 0.2 \leq t$	0.09	0.22			0.22			0.20		
		0.9	0.23			0.22			0.21		
		9	0.25			0.24			0.23		
			$\hat{\tau}_1$	$\hat{\tau}_2$		$\hat{\tau}_1$	$\hat{\tau}_2$		$\hat{\tau}_1$	$\hat{\tau}_2$	
Epidemic	$\alpha_{T_g} = 0.9$ for $\tau_1 = 0.3 \leq t \leq 0.9 = \tau_2$	0.09	0.33	0.93		0.33	0.91		0.31	0.90	
		0.9	0.35	0.98		0.34	0.95		0.31	0.91	
		9	0.39	1.00		0.37	0.97		0.33	0.94	
	$\alpha_{T_g} = 0.5$ for $\tau_1 = 0.4 \leq t \leq 0.8 = \tau_2$	0.09	0.40	0.81		0.40	0.80		0.40	0.80	
		0.9	0.43	0.81		0.42	0.81		0.41	0.81	
		9	0.48	0.86		0.45	0.86		0.43	0.84	
	$\alpha_{T_g} = 0.1$ for $\tau_1 = 0.5 \leq t \leq 0.6 = \tau_2$	0.09	0.51	0.61		0.51	0.60		0.50	0.60	
		0.9	0.52	0.63		0.51	0.62		0.51	0.62	
		9	0.55	0.67		0.53	0.64		0.51	0.63	
			$\hat{\tau}_1$	$\hat{\tau}_2$	$\hat{\tau}_3$	$\hat{\tau}_1$	$\hat{\tau}_2$	$\hat{\tau}_3$	$\hat{\tau}_1$	$\hat{\tau}_2$	$\hat{\tau}_3$
Multiple	$\alpha_{T_g} = 0.2$ for $\tau_1 = 0.2 \leq t \leq 0.4 = \tau_2$ and $\alpha_{T_g} = 0.9$ for $\tau_3 = 0.9 \leq t$	0.09	0.21	0.41	0.92	0.20	0.41	0.91	0.20	0.40	0.91
		0.9	0.23	0.42	0.96	0.21	0.42	0.93	0.20	0.41	0.91
		9	0.24	0.45	0.98	0.23	0.43	0.95	0.21	0.41	0.93
	$\alpha_{T_g} = 0.3$ for $\tau_1 = 0.3 \leq t \leq 0.6 = \tau_2$ and $\alpha_{T_g} = 0.7$ for $\tau_3 = 0.8 \leq t$	0.09	0.32	0.61	0.81	0.31	0.60	0.80	0.30	0.60	0.80
		0.9	0.32	0.61	0.80	0.32	0.61	0.81	0.30	0.61	0.80
		9	0.33	0.64	0.82	0.32	0.62	0.81	0.31	0.61	0.81
	$\alpha_{T_g} = 0.4$ for $\tau_1 = 0.4 \leq t \leq 0.5 = \tau_2$ and $\alpha_{T_g} = 0.5$ for $\tau_3 = 0.6 \leq t$	0.09	0.40	0.51	0.61	0.40	0.50	0.60	0.40	0.50	0.60
		0.9	0.41	0.51	0.61	0.40	0.51	0.61	0.40	0.50	0.60
		9	0.43	0.55	0.64	0.42	0.53	0.62	0.41	0.50	0.61

In the AMOC case, it is seen that, for various values of τ_1, as σ_{T_g} gets large, then all of three methods are not accurate, however, for all selected values of σ_{T_g}, SM has better performance. Also, the iterative method performs better than the DP method. However, it is worth mentioning that almost all of the three methods are accurate in most of the cases. Also, as naturally expected, it is seen that when the change-point is in the middle

of the data sequence all methods perform very well, for all values of σ_{T_g} and α_{T_g}. From these interpretations, it can be concluded that σ_{T_g} and τ_1 are the most important parameters affecting the performance of change-point detection methods, compared to the parameter α_{T_g}. The same observation from the results can be made about σ_{T_g} and α_{T_g} for the epidemic and multiple change cases. It is also seen that for all locations of change-points and for small σ_{T_g}, all three methods work well. Similar to the results of the previous section, it is seen that as the noise of the data gets large, the performance of the three methods becomes inaccurate. However, in most of the cells, all three methods detect the change locations accurately. The performance of SM is higher than any other method. Also, DP-MAP has the worst performance. It is interesting that the differences of the actual change-points and their estimates are negligible for all three patterns of change-point, i.e., AMOC, epidemic and multiple cases and for various values of parameters α_{T_g}, σ_{T_g} and location of changes. It should be noted that if the states are not available, their estimation via a Kalman filter can be used [9].

4.3.2 Second order uncertain system. The wind turbine pitch mechanism is considered in this section which is a hydraulic mechanism. The pitch angle of wind turbine blades is measured and fed back into a controller structure to fulfill tracking of the reference pitch angle β_{ref} [3]. The pitch angle control is designed to regulate the generated power and to avoid over speeding. So, the faults in the pitch angle sensor may lead to catastrophic operation and it is an important task to detect it. On the other hand, due to the harsh operational environment, some unknown dynamic changes deviate the pitch mechanism operation from its normal one. The most often reported dynamic changes are pump gear wear, hydraulic leakage and high air content in hydraulic oil, whose effects on the pitch mechanism are changes to the natural frequency and damping ratio of the mechanism from its normal situation, as summarized in Table 7. The effect of these dynamic changes is considered as dynamic uncertainty on the systems, which make the pitch mechanism slower. So, it is a challenging issue to distinguish any sensor faults from the mentioned uncertainties, as well as, sensor noise. The pitch mechanism is modelled as a second order system with uncertainty as [45],

$$\ddot{\beta} = -{\omega_{n,N}}^2\beta - 2\omega_{n,N}\xi_N\dot{\beta} + {\omega_{n,N}}^2\beta_{ref} + \Delta\tilde{f}_{PAD}$$

where, β, ω_n, ξ are pitch angle, natural frequency and damping ratio, respectively. $\Delta\tilde{f}_{PAD}$ represents the model uncertainty due to dynamic changes, which are modelled as a convex function of natural frequencies and damping ratios, as, $\Delta\tilde{f}_{PAD} = -\alpha_{f_1}\Delta({\tilde{\omega}_n}^2)\beta - 2\alpha_{f_2}\Delta(\tilde{\omega}_n\tilde{\xi})\dot{\beta} + \alpha_{f_1}\Delta({\tilde{\omega}_n}^2)\beta_{ref}$, $\Delta({\tilde{\omega}_n}^2) = {\omega_{n,HL}}^2 - {\omega_{n,N}}^2$ and $\Delta(\tilde{\omega}_n\tilde{\xi}) = \omega_{n,HAC}\xi_{HAC} - \omega_{n,N}\xi_N$.

Table 7: Frequency and damping ratio of pitch mechanism in normal as well as dynamic change.

	Natural Frequency(rad/s)	Damping Ratio	Indicator
Normal	$\omega_{n,N} = 11.11$	$\xi_N = 0.6$	$\alpha_{f_1} = \alpha_{f_2} = 0$
Pump Wear	$\omega_{n,PW} = 7.27$	$\xi_{PW} = 0.75$	$\alpha_{f_1} = 0.6316, \alpha_{f_2} = 0.29688$
Hydraulic Leak	$\omega_{n,HL} = 3.42$	$\xi_{HL} = 0.9$	$\alpha_{f_1} = 1, \alpha_{f_2} = 0.87853$
High Air Content	$\omega_{n,HAC} = 5.73$	$\xi_{HAC} = 0.45$	$\alpha_{f_1} = 0.81083, \alpha_{f_2} = 1$

Also, the pitch angle sensor is modelled as, $\beta_s = \alpha_\beta\beta + \nu_\beta$, where, $\nu_\beta \sim N(0, \sigma_\beta^2)$ is measurement noise and σ_β is standard deviation. $\alpha_\beta = 1$ and $\alpha_\beta \neq 1$ are measurement coefficients in fault free and faulty situations, respectively. Also, it is assumed that $\alpha_\beta \neq 0$, which means total loss of sensor outputs is avoided. Although, if the physical redundancy is provided, this assumption can be relaxed. It should be noted that the presence of system uncertainty $\Delta\tilde{f}_{PAD}$ leads to inaccurate detection using residual-based FDI methods [46, 47]. The reference pitch angle is selected as, $\beta_{ref} = 10\sin(t)$ (°). In Figure 11 the effect of pump wear, hydraulic leak and high air content are illustrated, where it is obvious that the changed dynamic response is slower than the normal one.

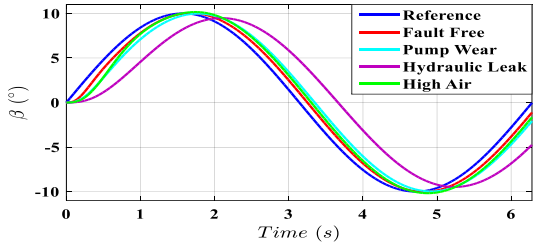


Figure 11: The effect of dynamic changes on pitch actuator response.

In Figure 12, the effects of sensor fault and noise in the hydraulic leak situation is illustrated in which, $\alpha_\beta = 0.5$ for $3 \leq t \leq 5\ s$ and $\alpha_\beta = 1$ for other time steps, and $\sigma_\beta = 0.2$ (°).

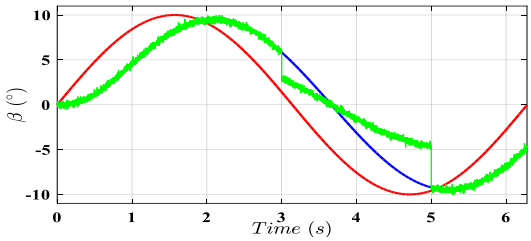


Figure 12: Reference pitch angle β_{ref} (red line), actual pitch angle β (blue line), and measured pitch angle β_s (green line).

Now, in Table 8, for two different fault scenarios and considering all given model uncertainties, the detected faults are summarized using the proposed methods to evaluate the fault detectability.

Table 8: Pitch angle sensor FDI results

			FDI method					
		Noise	DP-MAP		Iterative		SM	
	Fault scenario	σ_β	$\hat{\tau}_1$	$\hat{\tau}_2$	$\hat{\tau}_1$	$\hat{\tau}_2$	$\hat{\tau}_1$	$\hat{\tau}_2$
Pump Wear	$\alpha_\beta = 0.9$ for $\tau_1 = 1 \le t \le 3 = \tau_2$	0.2	1.03	3.02	1.02	3.02	1.01	3.00
		2	1.06	3.05	1.05	3.05	1.01	3.01
	$\alpha_\beta = 0.5$ for $\tau_1 = 3 \le t \le 5 = \tau_2$	0.2	3.00	5.01	3.00	5.00	3.00	5.00
		2	3.01	5.01	3.01	5.01	3.00	5.00
High Air Content	$\alpha_\beta = 0.9$ for $\tau_1 = 1 \le t \le 3 = \tau_2$	0.2	1.05	3.04	1.02	3.04	1.02	3.01
		2	1.08	3.06	1.03	3.04	1.02	3.02
	$\alpha_\beta = 0.5$ for $\tau_1 = 3 \le t \le 5 = \tau_2$	0.2	3.02	5.03	3.01	5.01	3.01	5.01
		2	3.05	5.07	3.03	5.04	3.01	5.02
Hydraulic Leak	$\alpha_\beta = 0.9$ for $\tau_1 = 1 \le t \le 3 = \tau_2$	0.2	1.11	3.12	1.07	3.09	1.06	3.04
		2	1.19	3.25	1.17	3.18	1.07	3.08
	$\alpha_\beta = 0.5$ for $\tau_1 = 3 \le t \le 5 = \tau_2$	0.2	3.15	5.19	3.09	5.11	3.05	5.05
		2	3.28	5.34	3.19	5.27	3.09	5.15

The results given in Table 8, shows that all faults are detected well by all three methods, despite the presence of model uncertainty. It is obvious that the performance of the SM method is more accurate.

4.3.3 Integrated system. For a regression analysis including one independent variable, the dependent variable can be non-stationary. When we are utilizing the integrated variables, it is necessary to make them stationary. In fact, sometimes the available information from the dependent variable is an integrated level and it is needed to take the first time derivation to convert it to a nonintegrated variable [9]. Integrated systems can be seen in many cases when the data are not available in direct observations, instead they are samples of the integral of a stochastic time series. This case happens in many fields such as with ice-core data on oxygen isotopes or with integrated volatility of financial problems [48]. In many practical situations, the integrated systems are suspected to produce spurious inference, because the dependent variable has an integrated form. A practical solution is to differentiate the integrated variables to get the stationary data series [9].

Accordingly, in this example, FDI of an integrated system is studied. Consider a type 1 servo system given by

$$G(s) = \frac{1}{s(s+1)(s+2)},$$

where the reference input is a step function. Utilizing a state feedback controller K, the closed loop system is as

$$\dot{x} = Ax + Bu,$$
$$y = x_1 + v_I,$$
$$u = -Kx,$$
$$y^*(\mathrm{t}) = \mathrm{x}_1^*(\mathrm{t}) = \int_0^{\mathrm{t}} \mathrm{x}_1(s)\,\mathrm{ds}.$$

where, $A = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & -2 & -3 \end{bmatrix}$, $B = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}$, $C = [1 \quad 0 \quad 0]$, $K = [160 \quad 54 \quad 11]$, to have closed-loop poles at $s = -2 \pm j2\sqrt{3}$ and $s = -10$, and $v_I \sim N(0, \sigma_I^2)$. The system output is x_1. However, it is assumed that $\mathrm{x}_1^*(\mathrm{t}) = \int_0^{\mathrm{t}} \mathrm{x}_1(s)\,\mathrm{ds}$ and $y^*(\mathrm{t}) = \mathrm{x}_1^*(\mathrm{t})$. On the other hand, after the fault occurs, the output is $y = \alpha_I x_1 + v_I$, and accordingly, $x_{1f}^*(t) = \int_0^t x_1(s)\,ds$ and $y^*(t) = x_{1f}^*(t)$. It is aimed to detect the fault in this integrated system, i.e. the partial sum of state is considered [9]. In Figure 13, the effects of sensor fault and noise for the integrated system is illustrated in which, $\alpha_I = 0.25$ for $1.7 \le t \le 5\ s$ and $\alpha_I = 1$ for other time steps, and $\sigma_I = 0.05$.

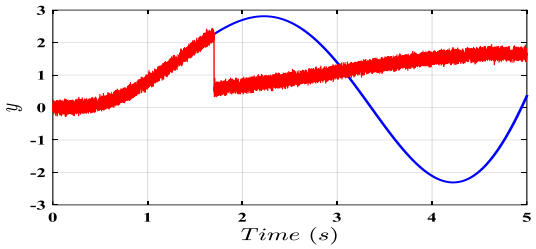


Figure 13: Input (blue line) and measured (red line) signals of the integrated system.

Now, in Table 9, for different fault scenarios for the integrated system, the detected faults are summarized using the proposed methods to evaluate their fault detectability.

Table 9: Integrated system FDI results

		FDI method					
	Noise	DP-MAP		Iterative		SM	
Fault scenario	σ_I	$\hat{\tau}_1$	$\hat{\tau}_2$	$\hat{\tau}_1$	$\hat{\tau}_2$	$\hat{\tau}_1$	$\hat{\tau}_2$
$\alpha_I = 0.9$ for $\tau_1 = 3 \leq t \leq 4 = \tau_2$	0.05	3.06	4.04	3.03	4.01	3.01	4.00
	0.5	3.24	4.15	3.19	4.08	3.08	4.05
$\alpha_I = 0.7$ for $\tau_1 = 2 \leq t \leq 3 = \tau_2$	0.05	2.04	3.05	2.02	3.04	2.00	3.01
	0.5	2.09	3.10	2.05	3.08	2.02	3.05
$\alpha_I = 0.5$ for $\tau_1 = 1 \leq t \leq 2 = \tau_2$	0.05	1.05	2.06	1.03	2.04	1.01	2.01
	0.5	1.07	2.12	1.06	2.08	1.03	2.04
$\alpha_I = 0.1$ for $\tau_1 = 4 \leq t \leq 4.5 = \tau_2$	0.05	4.05	4.55	4.03	4.53	4.01	4.51
	0.5	4.09	4.63	4.08	4.59	4.04	4.54

The results of performance analysis for the different fault scenarios are the same as for other previous examples. Again, SM has the best performance, although the differences here between the three methods is negligible and the methods again work well for integrated systems.

4.3.4 EWS. In this section the online fault probability is computed as an EWS. For this aim, the pitch actuator dynamics, given in Section 4.3.2, are considered. However, to accurately investigate EWS capability of the computed fault probability, it is assumed that the hydraulic leak is occurring gradually between $2 \leq t \leq 4\ s$, such that,

$$\begin{cases} \alpha_{f_1} = \alpha_{f_2} = 0, & 0 \leq t < 2 \\ \alpha_{f_1} = 0.5t - 1, \ \alpha_{f_2} = 0.87853\alpha_{f_1}, & 2 \leq t < 4. \\ \alpha_{f_1} = 1, \alpha_{f_2} = 0.87853, & 4 \leq t \end{cases}$$

Also, $\alpha_\beta = 1$ and $\sigma_\beta = 0.2$ (°). It is aimed to only detect the gradual hydraulic change, whose effect might be negligible at the time of occurrence, but considerable over longer-term operation. So, it is desirable to detect the pitch actuator hydraulic leak as soon as possible. In this manner, the EWS feature is achieved. Figure 14 gives the online probability of the fault on the time span (0,1). Comparing points before and after 0.2, it clearly shows that point 0.2 is suspected as a starting point of the fault in the system. Clearly, the probability is not too high, however, this early warning alarm motivates to suspect the time of 0.2 s as the start of the potential fault.

To check this stylized fact properly and to avoid detecting spurious changes, the probability of fault is plotted for time span (0,0.3) in Figure 15. It is seen that after 0.6, there is a high probability to have the fault. The point 0.6 in this graph is equivalent for 0.2 in the previous Figure 14, since $0.6 \approx 0.2/0.3$.

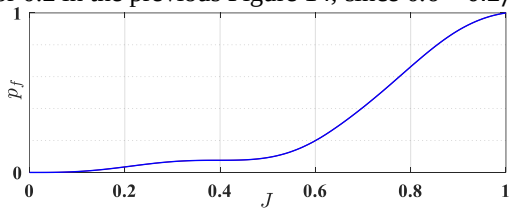


Figure 14: The overall probability of fault on (0,1)

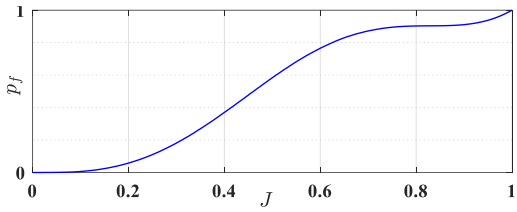


Figure 15: The probability of fault on (0,0.3)

As soon as this change (0.2) is detected, other changes are detected, sequentially. To this end, the probability of the fault is plotted for the time span (0.2,1). Figure 16 shows the possibility of a change at 0.4. It is seen that 0.2, in this new graph, is a starting point for a change which is equivalent to $0.2 + 0.2 = 0.4$, which is the actual change point. Other possible changes are also detected sequentially.

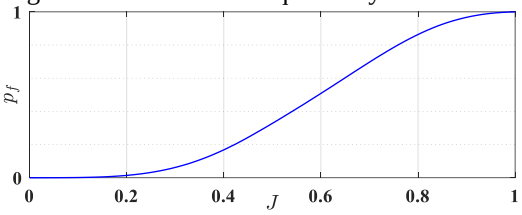
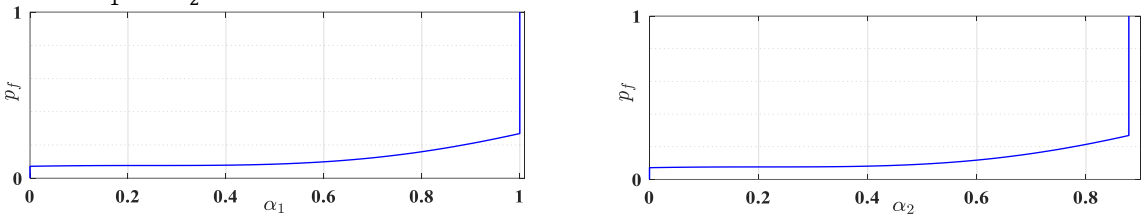


Figure 16: The probability of fault on (0.2,1)

EWS is derived via investigation of the fault probability plot versus parameters α_1, α_2. As illustrated in Figure 17, there is an early warning message about the faults occurring in the pitch actuator system due to the change in parameters α_1 and α_2.



(a) (b)

Figure 17: Probability of fault vs. α_1 (a) and α_2 (b).

5 Conclusions.

In this paper, the Bayesian multiple faults detection problem was considered, and DP-based methods were proposed. These methods consisted of the DP-based MAP, DP-based MCMC, the SM, iterative MAP and fault probability methods. First, the penalized MAP procedure based on the DP method was obtained. The DP method was applied to find the full conditionals of the MCMC procedures. The feasibility parameter space of the SM approach was derived using the DP approach. The iterative procedure was presented since the DP-based procedure was time-consuming. So, the recursive EM algorithm estimation method was adopted. Also, the DP method was used to find the fault probability by applying the BMS procedure. The fault detectability of the proposed schemes was evaluated and compared via the illustrative studies for randomly generated states. Also, to examine the proposed schemes, several practical examples were considered. It was shown that for the disturbed and uncertain dynamic systems with significant measurement noise, in which situation the model-based fault detection schemes fail to detect the fault appropriately, the proposed schemes were able to detect the fault moment accurately. However, for almost all of the various fault scenarios, the SM method had the best performance, followed by the iterative method and the DP-MAP method. However, it was worth mentioning that all three methods worked well in most of the cases.

References

[1] J. Harmouche, C. Delpha, and D. Diallo, "Incipient fault amplitude estimation using KL divergence with a probabilistic approach," *Signal Processing*, vol. 120, pp. 1-7, 2016.

[2] H. Habibi, I. Howard, and R. Habibi, "Bayesian Fault Probability Estimation: Application in Wind Turbine Drivetrain Sensor Fault Detection," *Asian Journal Control*, 2018 https://doi.org/10.1002/asjc.1973.

[3] H. Habibi, I. Howard, and S. Simani, "Reliability improvement of wind turbine power generation using model-based fault detection and fault tolerant control: A review," *Renewable Energy*, vol. 135, pp. 877-896, 2019.

[4] J. L. C. Hernandez, D. L. Almanza-Ojeda, S. Ledesma, A. Garcia-Perez, R. de Jesus Romero-Troncoso, and M.-A. Ibarra-Manzano, "Quaternion Signal Analysis Algorithm for Induction Motor Fault Detection," *IEEE Transactions on Industrial Electronics*, 2019.
[5] Y. Qiu, X. Liang, and Z. Dai, "Backstepping dynamic surface control for an anti-skid braking system," *Control Engineering Practice*, vol. 42, pp. 140-152, 2015.
[6] A. Youssef, C. Delpha, and D. Diallo, "An optimal fault detection threshold for early detection using Kullback–Leibler Divergence for unknown distribution data," *Signal Processing*, vol. 120, pp. 266-279, 2016/03/01/ 2016.
[7] A. Chibani, M. Chadli, P. Shi, and N. B. Braiek, "Fuzzy fault detection filter design for t–s fuzzy systems in the finite-frequency domain," *IEEE Transactions on Fuzzy Systems*, vol. 25, no. 5, pp. 1051-1061, 2017.
[8] A. Rastegari, M. M. Arefi, and M. H. Asemani, "Robust H∞ sliding mode observer-based fault-tolerant control for One-sided Lipschitz nonlinear systems," *Asian Journal of Control*, vol. 21, no. 1, pp. 114-129, 2019.
[9] H. Habibi, I. Howard, and R. Habibi, "Bayesian Sensor Fault Detection in a Markov Jump System," *Asian Journal Control*, vol. 19, no. 4, pp. 1465–1481, 2017.
[10] W. J. Faithfull, J. J. Rodríguez, and L. I. Kuncheva, "Combining univariate approaches for ensemble change detection in multivariate data," *Information Fusion*, vol. 45, pp. 202-214, 2019.
[11] Y. Guédon, "Exploring the latent segmentation space for the assessment of multiple change-point models," *Computational Statistics*, vol. 28, no. 6, pp. 2641-2678, 2013.
[12] D. M. Hawkins, "Fitting multiple change-point models to data," *Computational Statistics & Data Analysis*, vol. 37, no. 3, pp. 323-341, 2001.
[13] E. M. Maboudou-Tchao and D. M. Hawkins, "Detection of multiple change-points in multivariate data," *Journal of Applied Statistics*, vol. 40, no. 9, pp. 1979-1995, 2013.
[14] J. Bai and P. Perron, "Computation and analysis of multiple structural change models," *Journal of applied econometrics*, vol. 18, no. 1, pp. 1-22, 2003.
[15] G. Rigaill, "Pruned dynamic programming for optimal multiple change-point detection," *arXiv preprint arXiv:1004.0887*, 2010.
[16] P. Fryzlewicz, "Wild binary segmentation for multiple change-point detection," *The Annals of Statistics*, vol. 42, no. 6, pp. 2243-2281, 2014.
[17] B. Ait-El-Fquih, J. F. Giovannelli, N. Paul, A. Girard, and I. Hoteit, "Parametric Bayesian estimation of point-like pollution sources of groundwater layers," *Signal Processing*, p. 107339, 2019/10/10/ 2019.
[18] T. Young Yang and L. Kuo, "Bayesian binary segmentation procedure for a Poisson process with multiple changepoints," *Journal of Computational and Graphical Statistics*, vol. 10, no. 4, pp. 772-785, 2001.
[19] P. Fearnhead, "Exact and efficient Bayesian inference for multiple changepoint problems," *Statistics and computing*, vol. 16, no. 2, pp. 203-213, 2006.
[20] R. Killick, P. Fearnhead, and I. A. Eckley, "Optimal detection of changepoints with a linear computational cost," *Journal of the American Statistical Association*, vol. 107, no. 500, pp. 1590-1598, 2012.
[21] B. Ninness and S. Henriksen, "Bayesian system identification via Markov chain Monte Carlo techniques," *Automatica*, vol. 46, no. 1, pp. 40-51, 2010.
[22] A. Benson and N. Friel, "Adaptive MCMC for multiple changepoint analysis with applications to large datasets," *Electronic Journal of Statistics*, vol. 12, no. 2, pp. 3365-3396, 2018.
[23] J. Hofrichter, "Change point detection in generalized linear models," PhD, Technische Universitat Graz, Austria, 2007.
[24] Y. Pawitan, *In all likelihood: statistical modelling and inference using likelihood*. Oxford University Press, 2001.
[25] M. Milanese and M. Taragna, "H∞ set membership identification: A survey," *Automatica*, vol. 41, no. 12, pp. 2019-2032, 2005.
[26] J. Blesa, V. Puig, and J. Saludes, "Identification for passive robust fault detection using zonotope-based set-membership approaches," *International Journal of Adaptive Control and Signal Processing*, vol. 25, no. 9, pp. 788-812, 2011.
[27] R. M. Fernández-Cantí, J. Blesa, V. Puig, and S. Tornil-Sin, "Set-membership identification and fault detection using a bayesian framework," *International Journal of Systems Science*, vol. 47, no. 7, pp. 1710-1724, 2016.
[28] V. Reppa and A. Tzes, "Fault detection and diagnosis based on parameter set estimation," *IET control theory & applications*, vol. 5, no. 1, pp. 69-83, 2011.
[29] H. Habibi, I. Howard, and R. Habibi, "Sensor fault detection and isolation: a game theoretic approach," *International Journal of Systems Science*, vol. 49, no. 13, pp. 2826-2846, 2018.

[30] A. Mohammad-Djafari and O. Féron, "Bayesian approach to change points detection in time series," *International Journal of Imaging Systems and Technology,* vol. 16, no. 5, pp. 215-221, 2006.
[31] W. Zhao, D. Siegel, J. Lee, and L. Su, "An integrated framework of drivetrain degradation assessment and fault localization for offshore wind turbines," *IJPHM Special Issue on Wind Turbine PHM,* vol. 4, p. 46, 2013.
[32] M. Lavielle and E. Lebarbier, "An application of MCMC methods for the multiple change-points problem," *Signal Processing,* vol. 81, no. 1, pp. 39-53, 2001.
[33] Y.-C. Yao, "Estimating the number of change-points via Schwarz'criterion," *Statistics & Probability Letters,* vol. 6, no. 3, pp. 181-189, 1988.
[34] M. Lavielle, "Using penalized contrasts for the change-point problem," *Signal processing,* vol. 85, no. 8, pp. 1501-1510, 2005.
[35] S. M. AbouRizk, D. W. Halpin, and J. R. Wilson, "Fitting beta distributions based on sample data," *Journal of Construction Engineering and Management,* vol. 120, no. 2, pp. 288-305, 1994.
[36] M. C. Cario and B. L. Nelson, "Autoregressive to anything: Time-series input processes for simulation," *Operations Research Letters,* vol. 19, no. 2, pp. 51-58, 1996.
[37] J. J. Dabrowski, C. Beyers, and J. P. de Villiers, "Systemic banking crisis early warning systems using dynamic Bayesian networks," *Expert systems with applications,* vol. 62, pp. 225-242, 2016.
[38] A. K. Gupta and A. Ramanayake, "Change points with linear trend for the exponential distribution," *Journal of statistical planning and inference,* vol. 93, no. 1-2, pp. 181-195, 2001.
[39] P. Congdon, *Bayesian statistical modelling*. John Wiley & Sons, 2007.
[40] Z. Chen, Z. Li, and M. Zhou, "Detecting Change-Points in Epidemic Models," *Journal of Advanced Statistics,* vol. 1, no. 4, p. 181, 2016.
[41] Y. S. Niu, N. Hao, and H. Zhang, "Multiple change-point detection: A selective overview," *Statistical Science,* vol. 31, no. 4, pp. 611-623, 2016.
[42] Y. Zhang and Z. Wang, "A novel approach to fault detection in complex electric power systems," *Advances in electrical and Computer engineering,* vol. 14, no. 3, pp. 27-32, 2014.
[43] S. Inoue and S. Yamada, "Generalized discrete software reliability modeling with effect of program size," *IEEE Transactions on Systems, Man, and Cybernetics-Part A: Systems and Humans,* vol. 37, no. 2, pp. 170-179, 2007.
[44] J. Li, X. Wang, F. Han, and G. Wei, "Fault detection for discrete piecewise linear systems with infinite distributed time-delays," *International Journal of Systems Science,* pp. 1-9, 2017.
[45] H. Habibi, H. R. Nohooji, I. Howard, and S. Simani, "Fault-Tolerant Neuro Adaptive Constrained Control of Wind Turbines for Power Regulation with Uncertain Wind Speed Variation," *Energies,* vol. 12, no. 24, p. 4712, 2019.
[46] K. Zhao, Y. Song, and C. Wen, "Computationally inexpensive fault tolerant control of uncertain non-linear systems with non-smooth asymmetric input saturation and undetectable actuation failures," *IET Control Theory Appl,* vol. 10, no. 15, pp. 1866-1873, 2016.
[47] J. Yang and F. Zhu, "FDI design for uncertain nonlinear systems with both actuator and sensor faults," *Asian Journal Control,* vol. 17, no. 1, pp. 213-224, 2015.
[48] F. Baltazar-Larios and M. Sørensen, "Maximum likelihood estimation for integrated diffusion processes," in *Contemporary Quantitative Finance,* C. Chiarella, Ed. 1 ed.: Springer, 2010, pp. 407-423.