Journal of Process Control 66 (2018) 68–83

Contents lists available at ScienceDirect

Journal of Process Control

journal homepage: www.elsevier.com/locate/jprocont

Review

# Approaches to robust process identification: A review and tutorial of probabilistic methods

Hariprasad Kodamana, Biao Huang*, Rishik Ranjan, Yujia Zhao, Ruomu Tan, Nima Sammaknejad

*Department of Chemical and Materials Engineering, University of Alberta, Edmonton, Alberta T6G 2G6, Canada*

ARTICLE INFO

*Article history:*
Received 15 December 2016
Received in revised form 18 February 2018
Accepted 28 February 2018

*Keywords:*
Robust identification
Outliers
*t*-Distribution
Laplace distribution
Mixture of Gaussian distribution
Flat-topped distribution
Bayesian methods
EM algorithm
Variational Bayesian approach
Expectation Propagation
Monte Carlo methods
ARX models

ABSTRACT

Industrial data sets are often contaminated with outliers due to sensor malfunctions, signal interference, and other disturbances as well as interplay of various other factors. The effect of data abnormalities due to the outliers has to be systematically accounted while developing models that are resistant towards unforeseen effects of the outliers. The spectrum of methods that account for irregularities in process data while modeling are collectively known as robust identification methods. Even though, there are various non-probabilistic methods to tackle robust identification, few of them have considered the effect of outliers explicitly. In contrast to that, probabilistic identification methods ensure that these effects are given due attention. Despite these advantages, the probabilistic robust identification strategies have hardly been explored by practitioners. This review paper provides a general introduction to the probabilistic methods for robust identification, illustrates the main steps involved in the development of models, and reviews the related literature. Further, the paper contains two tutorial examples, including an industrial case study, to highlight various steps involved in the robust identification process.



**Contents**



* Corresponding author.
*E-mail address:* biao.huang@ualberta.ca (B. Huang).

https://doi.org/10.1016/j.jprocont.2018.02.011
0959-1524/

## 1. Introduction

In current industrial settings, advanced process control is an integral aspect of process operations for achieving safe operation of the complex industrial units, while attaining desired operational goals and economic benefits. Almost all prevailing advanced process control strategies are model based, requiring an accurate and compact mathematical description of the process. These models could be either (i) first-principles based, using physical knowledge of the process such as mass and momentum balances; (ii) completely data-driven black-box models; (iii) grey-box models that are a combination of both of the formers. In reality, it is acutely difficult to synthesize first-principles models of industrial units due to their complexity. Moreover, their validity remains questionable if there are a number of assumptions used, while modeling, for the sake of simplicity. Hence, the data driven system identification techniques are proven to be viable alternatives in this scenario. For comprehensive treatment of various system identification techniques, the reader is referred to stalwart references like [1–3].

Even though data driven system identification methods are proven to be boon in developing models for complex industrial processes, there are some caveats. Sometimes routine operating data tends to observe large deviation from its normal operating range. These outlying measurements or "outliers" are one of the common factors that may affect the quality of operational and laboratory data [4–6]. The outliers in operational data are mostly due to irregular and isolated disturbances, instrument failures, wrong instrument readings, potential human errors, and transmission problems. Moreover, almost all prevailing process systems are multi-variable and interacting among its variables. In such cases identification is anticipated to be highly sensitive to outliers. If we employ operational data that contain outliers in system identification, it would lead to poor fitting of the model parameters. Although, the screening of outliers is helpful, it can result in biased estimation of the parameters [7], thereby adversely affecting the identification process.

Robust system identification techniques are the host of methods that address the above mentioned problems, resulting from the unforeseen effects of noise in the data set, systematically and holistically. There have been earlier attempts to deal with the problem of robust identification without hinging on the framework of probabilistic inference, for example, employing, Huber's regression [8,9], wavelet based M-step estimators [10], simultaneous identification of parameters together with uncertainty bounds [11], worst case uncertainties [12–15], instrumental variables [16,17], subspace methods [18], co-prime factorizations [19], radial basis networks [20], invalidation methods [11], quantified estimations [21], relay and sinusoidal tests [22,23], optimization [24], pseudo singular values [25], fast-Fourier transforms [26], kernel based correntropy [27], and set membership identification methods [28–30]. Robust algorithms for the identification of output error model are also proposed in literature considering Huber's statistics [31], Hermite polynomials [32], projection operators [33], support vector machines [34], and Bayesian Kernals [35,36]. There have also been efforts in literature considering outliers in identification, for instances, identification with occasional outliers [37,38], piece wise affine systems [39,40], multi variable ARX models [41], ARMA models [42], and switched regression models [43]. These classical techniques use deterministic approaches to handle outliers while assuming noise to follow Gaussian distribution. Whilst, in probabilistic approaches of robust identification, the noise model is selected to account for the outlying data, thereby statistically compensating their unforeseen effects. Further, these methods fruitfully use the available information in terms of priors, embedding more trust in the model [44]. In addition, with the advent of active research happening in the area of stochastic model predictive control [45], there is also a need to accurately model the process dynamics accounting the outliers.

This paper reviews and introduces various robust probabilistic identification techniques that could be employed for modeling outlier contaminated data. Pointers to relevant literature are also provided for various robust identification techniques, so as to give a glimpse of the active research happening in the area, but we do not claim it to be an exhaustive review. The paper reports the following probabilistic methods to model outliers in the identification, namely, (i) mixture of Gaussian distribution; (ii) $t$-distribution; (iii) Laplace distribution; (iv) Flat-topped $t$- distribution. The article also revisits various methods for parameter estimations under probabilistic framework and introduces various techniques such as Expectation-Maximization algorithm, Hierarchical methods for Bayesian parameter estimation, Variational Bayesian methods, Expectation Propagation approaches and Monte Carlo Sampling based methods as tools for carrying out the robust identification exercises. Two tutorial examples, that include an industrial case study, are also provided to illustrate some of these methods in detail. The industrial case study focuses on oil sands processes, a major source of oil in Canada. Sensors used in oil sands processing are often subject to outliers due to the existence of harsh operating conditions, which necessitates robust model identification techniques. In such scenarios, delays and discontinuities associated with sensor sampling are of prime concern and will be addressed in the case studies.

The remainder of this article is organized as follows: Section 2 introduces the model structure considered. Section 3 briefly reviews various deterministic methods for robust identification. Section 4 introduces the probabilistic approach for robust identification, gives detailed accounts of various approaches to outlier modeling for robust identification and also presents methods for parameter estimation. Section 5 discusses detailed steps of robust identification problems with necessary examples. Section 6 draws concluding remarks from the study and proposes possible future research directions.

## 2. The model

Let us consider the following model structure for identification,

$$y_k = f(\mathbf{x}_k, \theta) + \epsilon_k \tag{1}$$