

Article

On the Uncertainty Identification for Linear Dynamic Systems Using Stochastic Embedding Approach with Gaussian Mixture Models

Rafael Orellana ^{1,2,3,*} , Rodrigo Carvajal ¹ , Pedro Escárate ⁴  and Juan C. Agüero ^{1,2} 

- ¹ Departamento Electrónica, Universidad Técnica Federico Santa María (UTFSM), Av. España 1680, Valparaíso 2390123, Chile; rodrigo.carvajal@usm.cl (R.C.); juan.aguero@usm.cl (J.C.A.)
- ² Advanced Center for Electrical and Electronic Engineering, AC3E, Av. Matta 222, Valparaíso 2580129, Chile
- ³ Departamento de Ingeniería Eléctrica, Facultad de Ingeniería, Universidad de Los Andes, Av. Alberto Carnevali, Mérida 5101, Venezuela
- ⁴ Instituto de Electricidad y Electrónica, Facultad de Ciencias de la Ingeniería, Universidad Austral de Chile (UACH), General Lagos 2086, Valdivia 5111187, Chile; pedro.escarate@uach.cl
- * Correspondence: rafael.orellanap@sansano.usm.cl

Abstract: In control and monitoring of manufacturing processes, it is key to understand model uncertainty in order to achieve the required levels of consistency, quality, and economy, among others. In aerospace applications, models need to be very precise and able to describe the entire dynamics of an aircraft. In addition, the complexity of modern real systems has turned deterministic models impractical, since they cannot adequately represent the behavior of disturbances in sensors and actuators, and tool and machine wear, to name a few. Thus, it is necessary to deal with model uncertainties in the dynamics of the plant by incorporating a stochastic behavior. These uncertainties could also affect the effectiveness of fault diagnosis methodologies used to increment the safety and reliability in real-world systems. Determining suitable dynamic system models of real processes is essential to obtain effective process control strategies and accurate fault detection and diagnosis methodologies that deliver good performance. In this paper, a maximum likelihood estimation algorithm for the uncertainty modeling in linear dynamic systems is developed utilizing a stochastic embedding approach. In this approach, system uncertainties are accounted for as a stochastic error term in a transfer function. In this paper, we model the error-model probability density function as a finite Gaussian mixture model. For the estimation of the nominal model and the probability density function of the parameters of the error-model, we develop an iterative algorithm based on the Expectation-Maximization algorithm using the data from independent experiments. The benefits of our proposal are illustrated via numerical simulations.

Keywords: expectation-maximization; gaussian mixture model; maximum likelihood; stochastic embedding; uncertainty modeling



Citation: Orellana, R.; Carvajal, R.; Escárate, P.; Agüero, J.C. On the Uncertainty Identification for Linear Dynamic Systems Using Stochastic Embedding Approach with Gaussian Mixture Models. *Sensors* **2021**, *21*, 3837. <https://doi.org/10.3390/21113837>

Academic Editor: Jongmyon Kim

Received: 7 May 2021

Accepted: 27 May 2021

Published: 1 June 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

System identification is a scientific discipline that studies techniques for modeling and estimating dynamic systems from experimental data, i.e., a number of experiments are performed on the system, and a model is then fitted to the measured data by estimating the corresponding system model parameters [1,2]. Most system identification techniques available in the literature assume that the system model that will be fitted lies in the model set [3]. However, real processes have an arbitrary complexity and using a complex model structure can lead to large variance estimation errors [4]. Moreover, the experimental data is typically obtained with the presence of sensors errors and measurement noise, modeling errors, and external disturbances that can impact on the system model accuracy; see, e.g., Reference [5]. This scenario has been considered for sensor or actuator failure diagnosis

and isolation in systems with structural uncertainties and noisy measurements, utilizing residual analysis and Bayesian frameworks to develop failure detection procedures [6–9].

In Figure 1, a typical representation of a real process and measured experimental data is shown, where u_t is the input measurements, y_t denotes the output measurements, and ω_t denotes a measurement noise. The real process (*true* system) is represented by the interaction of three components: (i) an actuator, (ii) a plant, and (iii) a sensor. This representation has been used to develop system identification methodologies for flight vehicle development in order to obtain accurate models and to validate mathematical models of the flight vehicle [10]. These system models are required, for example, to perform fault-diagnosis and adaptive control, to analyze handling qualities specification compliance, to develop high-fidelity aerodynamic databases for flight simulators, among others [11]. However, the quality of these results can be affected by the uncertainties incorporated by the (unknown) dynamics that are not modeled, the instrumentation, model simplifications, and measurement noise and sensor errors [12].

On the other hand, in Reference [13], the modeling and parameter identification in structural health monitoring was addressed. The authors considered that the multi-sensor data in a structural health monitoring can provide incomplete information due to multi-sensor faults or errors and then quantified its corresponding uncertainty with a Bayesian framework. Similarly, in Reference [14], a sensor-orientated approach for space situational awareness and traffic management uncertainty quantification was presented, analyzing the implications of the sensor uncertainties in the performance of this systems. A mathematical formulation based on the least square method was developed, providing a methodology to represent the navigation and tracking errors as an uncertainty volume that accurately depicts the size and orientation in space situational awareness evolution.

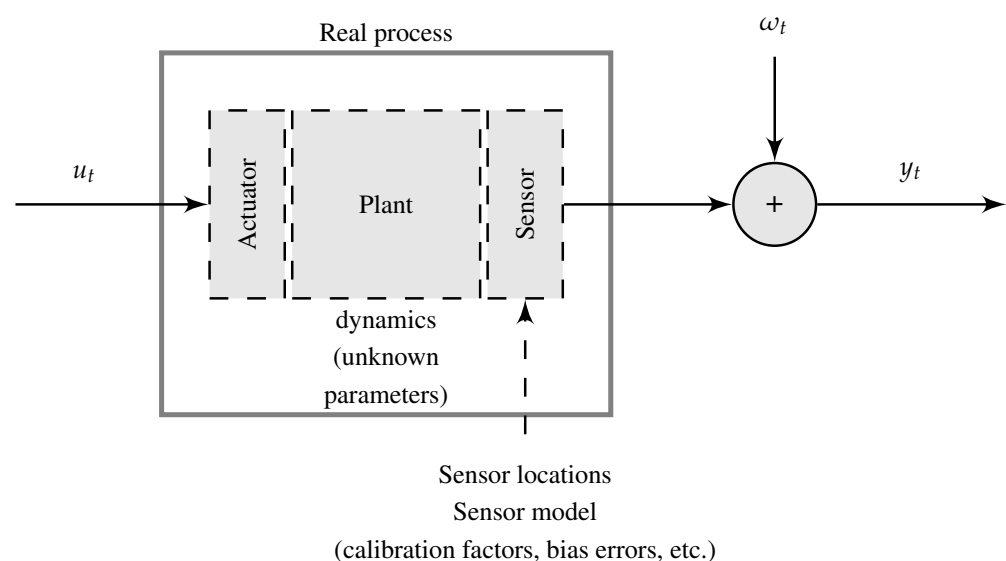


Figure 1. Representation of a real dynamic system with data measurements for system identification.

For these reasons, a suitable formulation of a model that incorporates uncertainties is an important aspect to consider in system identification methods, in order to obtain system models that represent as closely as possible the real process behavior [1,2,10]. Because of its valuable statistical properties [1–3], many identification algorithms involve the maximum likelihood (ML) principle with particular system models, such as dynamic systems [15–17], static systems [18], systems with quantized output data [19–22], and communications [23,24], to mention a few. In particular, the ML estimators are asymptotically unbiased. That is, when the number of measurements is not too large, the estimates can be far from the *true* value. This bias is typically accounted for by the measurement noise variance. However, there are scenarios in which the number of sampled-data is, in theory,

large enough, and the estimates considerably change from one experiment to another, i.e., there is an uncertainty in the system model that cannot be accounted for by the noise measurements and a large variance.

To illustrate the stochastic embedding approach, the magnitude of the frequency response of a linear dynamic system (assumed to be the *true* system) is shown in Figure 2, where $G^{[r]}(q^{-1})$ is given by

$$G^{[r]}(q^{-1}) = G_o(q^{-1}) + G_o(q^{-1})G_{\Delta}^{[r]}(q^{-1}), \quad (1)$$

where q^{-1} denotes the backward shift operator ($q^{-1}x_t = x_{t-1}$), r denotes a realization of the *true* system, $G_o(q^{-1})$ is a fixed nominal model, and $G_{\Delta}^{[r]}(q^{-1})$ denotes the r th realization of a relative error-model for the *true* system. The red-shaded region represents the area in which possible *true* systems, $G^{[r]}(q^{-1})$, i.e., realizations of the *true* system (represented by black dotted lines), can lie. The error-model can introduce model parameter uncertainties at low frequency range and model structure uncertainties at high frequency range [10]. This system model behavior can be encountered, for instance, in an aeroservoelastic system model (see, e.g., References [5,25]). Aeroservoelastic systems include dynamic coupling due to structural, control, sensor and actuator dynamics that cover both the low frequency and high frequency range. This means that, in an ML framework, the estimates can change considerable from one experiment to another, even if the number of measurements used to obtain an ML estimation is large, due to the structural and parametric uncertainties in the *true* system model. Hence, it is desirable that the models obtained from system identification techniques provide a quantification of their uncertainty, i.e., considering that the uncertainty modeling is part of the structure and estimating a nominal model with an error-model [26–28]. This approach of uncertainty modeling can be used for robust control design in which optimization methods can be used to obtain controllers in order to minimize the expected variation (variance) of the *true* system performance from a given desired behavior [29], or to obtain a probabilistic solution for robust control design [30,31].

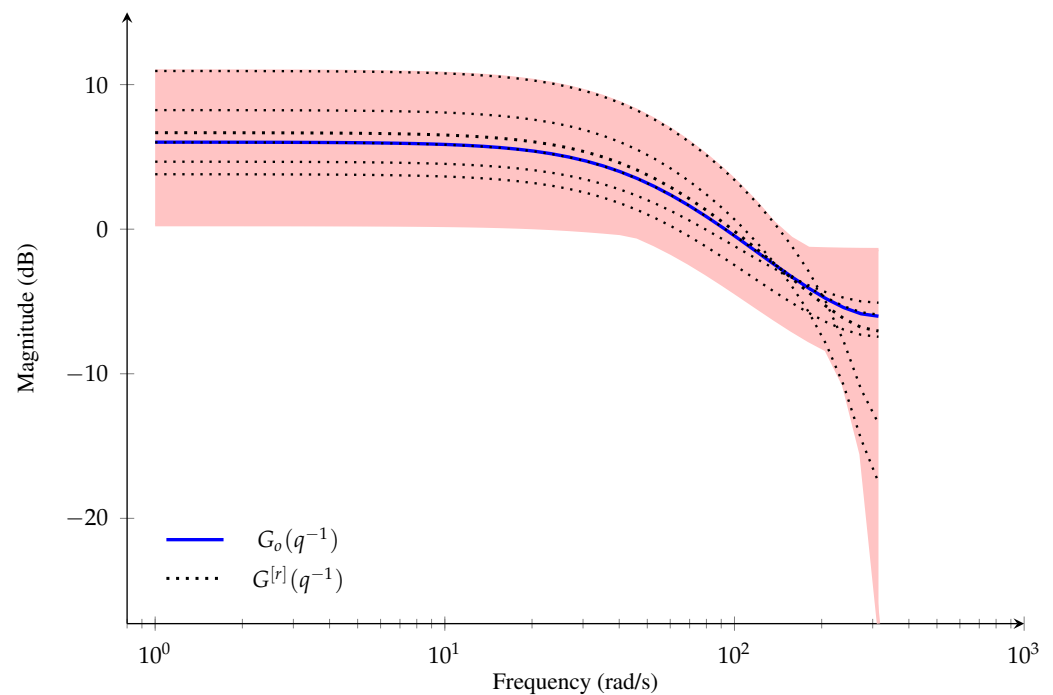


Figure 2. Magnitude of the frequency response for the linear dynamic system in Equation (1).

This alternative view of modeling error-model has been addressed in different frameworks. In Reference [32], the set membership approach was used to deal with the problem of uncertainty modeling in dynamic systems. The authors of Reference [32] considered the

error-model estimation in a deterministic framework in which it is unknown-but-bounded and obtaining a set of possible solutions. However, this method does not guarantee a small set of solutions to properly describe the system model uncertainty. In Reference [33], classical prediction error methods (PEM) [2] were used to obtain an estimation of the nominal model. Then, the unmodeled dynamics were estimated from the residuals dynamics using PEM and obtaining the corresponding model error modeling (MEM). Nevertheless, this methodology assumes that the nominal model is available. In References [26,34], a stochastic embedding (SE) approach was used to describe the model uncertainty by considering the model as a realization drawn from an underlying probability space, where the parameters that define the error-model are characterized by a probability density function (PDF). However, a flexible error-model distribution is needed to obtain a correct description of the system model uncertainty. In this approach, the uncertainty can be quantified by using ML estimation considering the parameters of the error-model as hidden variables [4,35] and solving the associated estimation problem with an expectation-maximization (EM) [36] based algorithm under Gaussian assumptions for the error-model distribution. References [4,37] also adopted a Bayesian perspective, where the parameters that define both the nominal model and the error-model can be modeled as realizations of random variables with certain prior distributions, and the posterior densities of the parameters can be estimated. There are works closely related with this framework [38,39] based on the kernel approach for system identification. Under this Bayesian framework, it is possible to obtain a unified system model of both the nominal model and error-model for the system in Equation (1) if all system models are considered finite impulse response (FIR) systems. However, this approach does not deliver the nominal model and the error-model separately when more complex model structures are involved in the *true* system.

In this paper, we propose an ML identification algorithm for modeling the uncertainty in a general class of linear dynamic systems using SE approach. That is, we aim at obtaining an estimation of the nominal model and the red-shaded region in Figure 2. We also consider that the error-model distribution is given by a Gaussian mixture model (GMM). GMMs have been utilized in non-linear filtering [40–42], identification of dynamic systems with ML framework [43,44] and a Bayesian framework [45–47]. Modeling error-model using SE approach for FIR systems have also been used with Gaussian mixtures distributions [48]. Moreover, GMMs have been used to approximate non-Gaussian distributions in the ML framework (see, e.g., References [17,47,49]) based on the Wiener approximation theorem which established that any PDF with compact support can be approximated by a finite sum of Gaussian distributions (Reference [50], Theorem 3). Combining the SE approach with Gaussian mixture distributions provides a flexible scenario for uncertainty modeling when the error-model distribution does not correspond to a GMM but can be approximated by one; the latter corresponds to the Wiener approximation theorem. The contributions of this paper can be summarized as follows:

- (i) We obtain the likelihood function for a linear dynamic system modeled with SE approach and a finite Gaussian mixture distribution. The likelihood function is computed by marginalizing the vector of parameters of the error-model as a hidden variable.
- (ii) We propose an EM algorithm to solve the associated ML estimation problem with GMMs, obtaining the estimates of the vector of parameters that define the nominal model and closed-form expressions for the GMM estimators of the error-model distribution.

The remainder of the paper is as follows: In Section 2, the system of interest for uncertainty modeling with SE framework is stated. In Section 3, the estimation problem for uncertainty modeling in a linear dynamic system is addressed using SE approach with GMMs. In Section 4, an EM algorithm is presented to solve the estimation problem. Numerical simulation examples are presented in Section 5. Finally, in Section 6, we present our conclusions.

2. Uncertainty Modeling for Linear Dynamic Systems Using Stochastic Embedding Approach

2.1. System Description

The system of interest in this paper is as follows:

$$y_t^{[r]} = G^{[r]}(q^{-1})u_t^{[r]} + H(q^{-1}, \theta)\omega_t^{[r]}, \quad (2)$$

where $r = 1, \dots, M$ denotes the r -th realization of the system, M corresponds the number of independent experiments or batches, $t = 1, \dots, N$ denotes the t -th measurement, N is the data length, $y_t^{[r]}$ is the output signal, $u_t^{[r]}$ denotes the input signal, $\omega_t^{[r]}$ is a zero-mean Gaussian white noise with variance σ^2 , and $H(q^{-1}, \theta)$ is the noise model parametrized by θ . We consider that the system $G^{[r]}(q^{-1})$ can be described as follows (see, e.g., References [26,37]):

$$G^{[r]}(q^{-1}) = G_o(q^{-1}, \theta) \left(1 + G_\Delta(q^{-1}, \eta^{[r]}) \right), \quad (3)$$

where $G_o(q^{-1}, \theta)$ is the nominal system model parametrized by θ , and $G_\Delta(q^{-1}, \eta^{[r]})$ is the error-model parametrized by $\eta^{[r]}$. Here, we consider that the PDF of the error-model in Equation (2) is a finite Gaussian mixture distribution given by:

$$p(\eta^{[r]}|\gamma) = \sum_{i=1}^{\kappa} \alpha_i \mathcal{N}(\eta^{[r]}; \mu_i, \Gamma_i), \quad (4)$$

$$\gamma = \underbrace{[\alpha_1 \ \mu_1 \ \Gamma_1]}_{\gamma_1} \cdots \underbrace{[\alpha_\kappa \ \mu_\kappa \ \Gamma_\kappa]}_{\gamma_\kappa}^T, \quad (5)$$

where $[\cdot]^T$ denotes the transpose operator, $\alpha_i > 0$, $\sum_{i=1}^{\kappa} \alpha_i = 1$, and $\mathcal{N}(\eta^{[r]}; \mu_i, \Gamma_i)$ represents a Gaussian PDF with mean value μ_i and covariance matrix Γ_i . For simplicity of the presentation, we consider that $H(q^{-1}, \theta)$ in Equation (2) is part of the nominal model, i.e., it is only parametrized by θ . The nominal system model in Equation (3) is given by

$$G_o(q^{-1}, \theta) = \frac{B(q^{-1}, \theta)}{A(q^{-1}, \theta)}, \quad (6)$$

where

$$B(q^{-1}, \theta) = b_1 q^{-1} + \dots + b_{n_b} q^{-n_b}, \quad (7)$$

$$A(q^{-1}, \theta) = 1 + a_1 q^{-1} + \dots + a_{n_a} q^{-n_a}. \quad (8)$$

Similarly, the noise system model in Equation (2) is given by

$$H(q^{-1}, \theta) = \frac{C(q^{-1}, \theta)}{D(q^{-1}, \theta)}, \quad (9)$$

with

$$C(q^{-1}, \theta) = 1 + c_1 q^{-1} + \dots + c_{n_c} q^{-n_c}, \quad (10)$$

$$D(q^{-1}, \theta) = 1 + d_1 q^{-1} + \dots + d_{n_d} q^{-n_d}. \quad (11)$$

Notice that we use the term nominal model to refer to the system model $G_o(q^{-1}, \theta)$ and the noise system model $H(q^{-1}, \theta)$. We also consider that the error-model $G_\Delta(q^{-1}, \eta^{[r]})$ in Equation (3) is a linear regression as follows:

$$G_\Delta(q^{-1}, \eta^{[r]}) = \eta_0^{[r]} + \eta_1^{[r]} q^{-1} + \dots + \eta_{n_\Delta}^{[r]} q^{-n_\Delta}, \quad (12)$$

where $\eta^{[r]} \in \mathbb{R}^{n_\Delta \times 1}$.

2.2. Standing Assumptions

The problem of interest is to estimate the vector of parameters, $\beta = [\theta^T \ \gamma^T \ \sigma^2]^T$, that defines the parameters of the nominal model, error-model and the noise variance. In addition, we consider that β_0 is the *true* vector of parameters that defines the *true* model. In order to formulate the ML estimation algorithm, we introduce the following standing assumptions:

Assumption 1. The system in Equation (2) is operating in open loop, and the input signal $u_t^{[r]}$ is an exogenous deterministic signal for each r independent experiment. In addition, the data of M independent experiments are available.

Assumption 2. The nominal model does not change from one experiment to another, whilst the error-model $G_\Delta(q^{-1}, \eta^{[r]})$ may change for each experiment, and all the realizations of $\eta = \{\eta^{[r]}\}_{r=1}^M$ are drawn from the same PDF parametrized by γ .

Assumption 3. The vector of parameters β_0 , the input signal $u_t^{[r]}$, and the noise $\omega_t^{[r]}$ in Equation (2) satisfy regularity conditions, guaranteeing that the ML estimate β_{ML} converges (in probability or a.s.) to the true solution β_0 as $N \rightarrow \infty$.

Assumption 4. The orders n_a, n_b, n_c, n_d , and n_Δ of the polynomials of system in Equation (2) and the number of components κ of the error-model distribution in Equation (4) are known.

Assumption 5. The system in Equation (2) is asymptotically stable, and its models $G^{[r]}(q^{-1})$ and $H(q^{-1}, \theta)$ have no poles-zeros on the unit circle and have no pole-zero cancellations. The noise system model $H(q^{-1}, \theta)$ is also a minimum-phase system.

Assumption 2 is needed to develop an ML methodology based on the uncertainty modeling with SE approach. Assumption 3 is necessary to develop an estimation algorithm that holds the asymptotic properties of the ML estimator. Assumption 4 can be relaxed by considering a dynamic system model that includes parametric uncertainties with a different error-model structure that we assume in Equation (3). We will address this case in Section 5. Assumption 5 is necessary to obtain an asymptotically unbiased ML estimator [51] and a system model that is controllable [52].

3. Maximum Likelihood Estimation for Uncertainty Modeling Using GMMs

In this section, we develop an ML estimation algorithm for system in Equation (2) to obtain both the nominal model and the error-model distribution. We consider that the observed data $Y^{[r]} = [y_1^{[r]} \ \dots \ y_N^{[r]}]$ is a collection of measurements for each experiment. We use capital letters to denote the vector of signals, u_t or ω_t for $t = 1, \dots, N$. Then, the system in Equation (2) can be described as follows:

$$Y^{[r]} = \mathcal{G}(\theta)U^{[r]} + \Psi^{[r]}(\theta)\eta^{[r]} + W^{[r]}, \quad (13)$$

where $Y^{[r]}, U^{[r]}, W^{[r]} \in \mathbb{R}^{N \times 1}$, $\theta \in \mathbb{R}^{n_o \times 1}$, $\eta^{[r]} \in \mathbb{R}^{n_\Delta \times 1}$, $\mathcal{G}(\theta)U^{[r]} \in \mathbb{R}^{N \times 1}$, $\Psi^{[r]}(\theta) \in \mathbb{R}^{N \times n_\Delta}$. The term $\mathcal{G}(\theta)U^{[r]}$ corresponds to the output response corresponding to the nominal model structure $\mathcal{G}(\theta)$, $\Psi^{[r]}(\theta)\eta^{[r]}$ corresponds to the output signal due to the error-model structure in Equation (3), and $W^{[r]} \sim \mathcal{N}(0, \sigma^2 I_N)$ (I_x represents the identity with dimension given by x). Notice that $\mathcal{G}(\theta)$ involves the nominal model structure utilized for $G_o(q^{-1}, \theta)$ and $H(q^{-1}, \theta)$ in Equation (2). The term $\Psi^{[r]}(\theta)$ describes the model structure given by the product $G_o(q^{-1}, \theta)G_\Delta(q^{-1}, \eta^{[r]})$ in Equation (3).

Remark 1. The system model in Equation (13) consider that the error-model structure $\Psi^{[r]}(\theta)\eta^{[r]}$ is a linear regression, and the nominal model structure $\mathcal{G}(\theta)$ can have an arbitrary structure [37]. The fact of considering the error-model as a linear regression can provide flexibility, lower compu-

tational complexity, and contain FIR model of arbitrary orders, Laguerre and Kautz models (see, e.g., Reference [53]).

We define the vector of parameters to be estimated as $\beta = [\theta^T \ \gamma^T \ \sigma^2]^T$ in order to formulate the ML estimator for the system in Equation (13):

$$\theta = [b_1 \ a_1 \ c_1 \ d_1 \ \cdots \ b_{n_b} \ a_{n_a} \ c_{n_c} \ d_{n_d}]^T, \quad (14)$$

where θ is the vector of parameters of the nominal model, γ is the vector of parameters of the error-model distribution as a GMM in Equation (4), and σ^2 is the noise variance. The corresponding ML estimation algorithm using GMMs is obtained as follows:

Lemma 1. Consider the parameters to be estimated as $\beta = [\theta^T \ \gamma^T \ \sigma^2]^T$ using Equations (5) and (14). Under the standing assumptions and using Equation (13), the ML estimator for the system model in Equation (2) is given by:

$$\hat{\beta}_{ML} = \arg \max_{\beta} \ell(\beta) \text{ s.t. } \sum_{i=1}^{\kappa} \alpha_i = 1, \ 0 \leq \alpha_i \leq 1, \quad (15)$$

where the log-likelihood function is given by

$$\ell(\beta) = \sum_{r=1}^M \log \left\{ \sum_{i=1}^{\kappa} \alpha_i \mathcal{N}(Y^{[r]}; \bar{\mu}_{ir}, \bar{\Sigma}_{ir}) \right\}, \quad (16)$$

$$\bar{\mu}_{ir} = \Psi^{[r]}(\theta) \mu_i + \mathcal{G}(\theta) U^{[r]}, \quad (17)$$

$$\bar{\Sigma}_{ir} = \sigma^2 I_N + \Psi^{[r]}(\theta) \Gamma_i [\Psi^{[r]}(\theta)]^T. \quad (18)$$

Proof. See Appendix A. \square

The result obtained from Lemma 1 differs from the classical ML formulation, where the measurement data of one experiment is used to obtain one estimate of the parameters of interest (see, e.g., References [1,2]). In contrast, from Lemma 1, the ML estimator considers the measurements of M independent experiments to obtain one estimation of the system model, i.e, the nominal model and error-model distribution as a GMM. Notice that $\bar{\mu}_{ir}$ and $\bar{\Sigma}_{ir}$ in Equations (17) and (18), respectively, depend on the vector of parameters β .

The optimization problem in Equation (15) is typically solved using gradient-based methods. However, the log-likelihood function in Equation (16) contains a matrix $\bar{\Sigma}_{ir}$ with dimension $N \times N$ that may seem cumbersome to solve for the optimization problem [37,54]. Moreover, the log-likelihood function involves a logarithm of a summation that depends of the number of components κ of the GMM, and it may be difficult to solve when the number of components in the GMM is high [55]. The EM algorithm [36] can provide a solution to deal with this difficulty since combination of an EM algorithm with GMMs typically provides closed-form estimators for the GMM parameters [56,57].

4. An EM Algorithm with GMMs for Uncertainty Modeling in Linear Dynamic Systems

The EM algorithm is a popular tool for identifying linear and non-linear dynamic systems in the time domain (see, e.g., References [58,59]) and frequency domain [60]. In particular, a common strategy used in the formulation of an EM algorithm with GMMs is to consider a hidden (latent) variable, modeled as an indicator, that determines from which GMM component an observation comes from Reference [57]. To solve the estimation problem in Equation (15), an EM algorithm with GMMs is developed from the definition of the likelihood function from the observed data of all experiments $Y^{[1:M]} = [Y^{[1]}, \dots, Y^{[M]}]$ and a hidden discrete random variable, $\zeta_{1:M}$, i.e., defining the likelihood function in

Equation (16) using the complete data, $\{Y^{[1:M]}, \zeta_{1:M}\}$. This hidden random variable ζ_r is an indicator that determines if observations $Y^{[r]}$ arise from the i th component of the GMM. Hence, the EM algorithm is given by (see, e.g., References [36,57]):

$$Q(\beta, \hat{\beta}^{(m)}) = \mathbb{E} \left\{ \log [p(Y^{[1:M]}, \zeta_{1:M} | \beta)] | Y^{[1:M]}, \hat{\beta}^{(m)} \right\}, \quad (19)$$

$$\hat{\beta}^{(m+1)} = \arg \max_{\beta} Q(\beta, \hat{\beta}^{(m)}), \quad \text{s.t.} \quad \sum_{i=1}^{\kappa} \alpha_i = 1, \quad 0 \leq \alpha_i \leq 1, \quad (20)$$

where $\mathbb{E}\{a|b\}$ and $p(a|b)$ denote the expected value and the PDF of the random variable a given the random variable b , respectively, $\hat{\beta}^{(m)}$ is the current estimate, $p(Y, \zeta_{1:M} | \beta)$ is the joint PDF of $Y^{[1:M]}$ and $\zeta_{1:M}$, and $Q(\beta, \hat{\beta}^{(m)})$ is the auxiliary function of the EM algorithm. Notice that Equations (19) and (20) correspond to the *E-step* and *M-step* of the EM algorithm, respectively [36].

In order to develop the iterative algorithm in Equations (19) and (20), we obtain the following result. This will be used to compute $Q(\beta, \hat{\beta}^{(m)})$ in Equation (19):

Lemma 2. Consider the vector of parameters $\beta = [\theta^T \ \gamma^T \ \sigma^2]^T$ defined in Equations (5) and (14). The joint log-PDF, $\log[p(Y^{[1:M]}, \zeta_{1:M} | \beta)]$, in Equation (19) is given by

$$\log[p(Y^{[1:M]}, \zeta_{1:M} | \beta)] = \sum_{r=1}^M \sum_{i=1}^{\kappa} \log[p(\zeta_r = i | \beta)] + \log[\mathcal{N}(Y^{[r]}; \bar{\mu}_{ir}, \bar{\Sigma}_{ir})], \quad (21)$$

where $\bar{\mu}_{ir}$ and $\bar{\Sigma}_{ir}$ are given by Equations (17) and (18), respectively.

Proof. See Appendix B. \square

Based on the coordinate descent algorithm [61], we obtain a new estimate in Equation (20) using the following steps:

- (1) Fixing the vector of parameters θ at its value from the current iteration $\hat{\theta}^{(m)}$ to optimize Equation (19) with respect to the GMM parameters γ and the noise variance σ^2 .
- (2) Fixing the GMM parameters and the noise variance at their values from the iteration $\hat{\gamma}^{(m+1)}$ and $\hat{\sigma}^{2(m+1)}$ and to solve the optimization problem in Equation (20) to obtain $\hat{\theta}^{(m+1)}$.

From Lemma 2, our proposed EM algorithm can be computed as follows:

Theorem 1. Consider the vector of parameters to be estimated $\beta = [\theta^T \ \gamma^T \ \sigma^2]^T$ defined in Equations (5) and (14) for the system in Equation (2). Substituting Equation (21) in Equation (19), the *E-step* in the EM algorithm is given by

$$Q(\beta, \hat{\beta}^{(m)}) = \sum_{r=1}^M \sum_{i=1}^{\kappa} \log \left[\alpha_i \mathcal{N}(Y^{[r]}; \bar{\mu}_{ir}, \bar{\Sigma}_{ir}) \right] \hat{\zeta}_{ir}^{(m)}, \quad (22)$$

$$\hat{\zeta}_{ir}^{(m)} = \frac{\hat{\alpha}_i^{(m)} \mathcal{N}(Y^{[r]}; \hat{\mu}_{ir}^{(m)}, \hat{\Sigma}_{ir}^{(m)})}{\sum_{l=1}^{\kappa} \hat{\alpha}_l^{(m)} \mathcal{N}(Y^{[r]}; \hat{\mu}_{lr}^{(m)}, \hat{\Sigma}_{lr}^{(m)})}, \quad (23)$$

where $\hat{\mu}_{ir}^{(m)}$ and $\hat{\Sigma}_{ir}^{(m)}$ are given by:

$$\hat{\mu}_{ir}^{(m)} = \Psi^{[r]}(\hat{\theta}^{(m)}) \hat{\mu}_i^{(m)} + \mathcal{G}(\hat{\theta}^{(m)}) U^{[r]}, \quad (24)$$

$$\hat{\Sigma}_{ir}^{(m)} = \hat{\sigma}^{2(m)} I_N + \Psi^{[r]}(\hat{\theta}^{(m)}) \hat{\Gamma}_i^{(m)} [\Psi^{[r]}(\hat{\theta}^{(m)})]^T. \quad (25)$$

Consider the maximization problem stated in Equation (20) using Equation (22). The M -step in the EM algorithm is carried out using the following steps:

(i) Solving Equation (20) using Equation (22) with $\theta = \hat{\theta}^{(m)}$:

$$\hat{\alpha}_i^{(m+1)} = \left(\sum_{r=1}^M \hat{\zeta}_{ir}^{(m)} \right) / M, \quad (26)$$

$$\hat{\mu}_i^{(m+1)} = \left(\sum_{r=1}^M \hat{\mu}_{ir}^{(m)} \hat{\zeta}_{ir}^{(m)} \right) / \left(\sum_{r=1}^M \hat{\zeta}_{ir}^{(m)} \right), \quad (27)$$

$$\hat{\Gamma}_i^{(m+1)} = \left(\sum_{r=1}^M \left[(\hat{\mu}_{ir}^{(m)} - \hat{\mu}_i^{(m)}) (\hat{\mu}_{ir}^{(m)} - \hat{\mu}_i^{(m)})^T + \hat{\Sigma}_{ir}^{(m)} \right] \hat{\zeta}_{ir}^{(m)} \right) / \left(\sum_{r=1}^M \hat{\zeta}_{ir}^{(m)} \right), \quad (28)$$

$$\hat{\sigma}^{2(m+1)} = \left(\sum_{r=1}^M \sum_{i=1}^{\kappa} \mathcal{R}_{ir}^{(m)} \hat{\zeta}_{ir}^{(m)} \right) / NM, \quad (29)$$

with

$$\hat{\mathcal{P}}_{ir}^{(m)} = \hat{\Gamma}_i^{(m)} \Psi^{[r]}(\hat{\theta}^{(m)})^T \left(\hat{\Sigma}_{ir}^{(m)} \right)^{-1}, \quad (30)$$

$$\hat{\mu}_{ir}^{(m)} = \hat{\mu}_i^{(m)} + \hat{\mathcal{P}}_{ir}^{(m)} \left(Y^{[r]} - \hat{\mu}_{ir}^{(m)} \right), \quad (31)$$

$$\hat{\Sigma}_{ir}^{(m)} = \left(I_{n_\Delta} - \hat{\mathcal{P}}_{ir}^{(m)} \Psi^{[r]}(\hat{\theta}^{(m)}) \right) \hat{\Gamma}_i^{(m)}, \quad (32)$$

$$\begin{aligned} \mathcal{R}_{ir}^{(m)} = & \left(Y^{[r]} - \mathcal{G}(\hat{\theta}^{(m)}) U^{[r]} - \Psi^{[r]}(\hat{\theta}^{(m)}) \hat{\mu}_{ir}^{(m)} \right)^T \\ & \left(Y^{[r]} - \mathcal{G}(\hat{\theta}^{(m)}) U^{[r]} - \Psi^{[r]}(\hat{\theta}^{(m)}) \hat{\mu}_{ir}^{(m)} \right) + \mathbf{tr} \left(\Psi^{[r]}(\hat{\theta}^{(m)})^T \hat{\Sigma}_{ir}^{(m)} \Psi^{[r]}(\hat{\theta}^{(m)}) \right), \end{aligned} \quad (33)$$

where $\mathbf{tr}(\cdot)$ denotes the trace operator.

(ii) Solving Equation (20) using Equation (22) with $\gamma = \hat{\gamma}^{(m+1)}$:

$$\hat{\theta}^{(m+1)} = \arg \min_{\theta} \sum_{r=1}^M \sum_{i=1}^{\kappa} \hat{\zeta}_{ir}^{(m)} \mathcal{B}_{ir}(\theta, \theta^{(m)}), \quad (34)$$

where

$$\begin{aligned} \mathcal{B}_{ir}(\theta, \theta^{(m)}) = & \left(Y^{[r]} - \mathcal{G}(\theta) U^{[r]} - \Psi^{[r]}(\theta) \hat{\mu}_{ir}^{(m+1)} \right)^T \\ & \left(Y^{[r]} - \mathcal{G}(\theta) U^{[r]} - \Psi^{[r]}(\theta) \hat{\mu}_{ir}^{(m+1)} \right) + \mathbf{tr} \left(\Psi^{[r]}(\theta)^T \hat{\Sigma}_{ir}^{(m+1)} \Psi^{[r]}(\theta) \right). \end{aligned} \quad (35)$$

Proof. See Appendix C. \square

From Theorem 1, we obtain closed-form expressions for the GMM and noise variance estimators. They are computed from the definition of the auxiliary function $\mathcal{Q}(\beta, \hat{\beta}^{(m)})$ using the complete log-likelihood function in Equation (21). That differs from the procedure proposed in Reference [49], where an auxiliary function is built without explicitly defining a hidden variable. However, the auxiliary function obtained in Reference [49] is equal to Equation (19) and the estimators are obtained from Equations (26)–(29) and Equation (34).

Finally, our estimation algorithm is summarized in Algorithm 1.

Algorithm 1 (EM algorithm with GMMs).

Inputs $Y^{[1:M]}, U^{[1:M]}, M, \kappa, \hat{\theta}^{(0)}, \hat{\gamma}^{(0)}$ and $\hat{\sigma}^{2(0)}$.
Outputs $\hat{\theta}, \hat{\gamma}$ and $\hat{\sigma}^2$.

- 1: $m \leftarrow 0$
- 2: **procedure** E-STEP
- 3: Compute $\hat{\mu}_{ir}^{(m)}$ and $\hat{\Sigma}_{ir}^{(m)}$ from Equations (24) and (25).
- 4: Compute $\hat{\zeta}_{ir}^{(m)}$ from Equation (23).
- 5: Compute $\hat{\mu}_{ir}^{(m)}$ and $\hat{\Sigma}_{ir}^{(m)}$ from Equations (31) and (32).
- 6: **end procedure**
- 7: **procedure** M-STEP
- 8: Estimate $\hat{\gamma}^{(m+1)}, \hat{\sigma}^{2(m+1)}$ from Equations (26)–(29).
- 9: Compute $\mathcal{B}_{ir}(\theta, \theta^{(m)})$ from Equation (35) using $\hat{\gamma}^{(m+1)}$ and $\hat{\theta}^{(m)}$.
- 10: Estimate $\hat{\theta}^{(m+1)}$ by solving Equation (34).
- 11: **end procedure**
- 12: **if** stopping criterion is not satisfied **then**
- 13: $m \leftarrow m + 1$, **return to 2**
- 14: **else**
- 15: $\hat{\theta} \leftarrow \hat{\theta}^{(m+1)}, \hat{\gamma} \leftarrow \hat{\gamma}^{(m+1)}, \hat{\sigma}^2 \leftarrow \hat{\sigma}^{2(m+1)}$
- 16: **end if**
- 17: **End**

5. Numerical Examples

In this section, we present two numerical examples with different approaches to illustrate the benefits of our proposal for modeling both the nominal model and the error-model in the system of Equation (2). This approach of numerical simulations is often used when new estimation algorithms are tested in order to evaluate the estimation performance in conditions that could reduce the safety levels or deal with problems for which experiments cannot be planned [62,63].

In the first example, we consider a variant of the example used in Reference [37] holding all standing assumptions stated in Section 2.2. The nominal system model ($G_o(q^{-1}, \theta)$ and $H(q^{-1}, \theta)$) corresponds to a Box-Jenkins model. For simplicity, we also consider that the error-model ($G_\Delta(q^{-1}, \eta^{[r]})$) corresponds to a second order FIR system in Equation (12) with an overlapped Gaussian mixture error-model distribution. We assume that there are not dynamic uncertainties at low frequencies, i.e., the error-model has zero static gain. In this case, we do not establish a comparison analysis with other approaches since the simulated data is generated using our proposed framework with SE and GMMs.

In contrast, we consider a second example to illustrate the flexibility of our proposed algorithm for modeling the error-model for the system in Equation (2). Assumption 4 is relaxed considering a model structure does not correspond to Equation (3). Instead, we consider the system in Equation (2) with an output-error model structure where the system model, $G^{[r]}(q^{-1}, \theta)$, is given by the discrete-time model of a simple resistor-capacitor continuous-time system model. Here, we consider that the time constant τ is given by $\tau = \tau_o(1 + \delta^{[r]})$, where τ_o represents the time constant computed with the nominal values of the resistor and capacitor, and $\delta^{[r]}$ denotes the tolerance (uncertainty) with a non-Gaussian distribution.

For both examples, the simulation setup is as follows:

- (1) The data length is $N = 100$.
- (2) The number of experiments is $M = 100$.
- (3) The number of Monte Carlo (MC) simulation is 25.
- (4) The stopping criterion is satisfied when:

$$\left\| \hat{\beta}^{(m)} - \hat{\beta}^{(m-1)} \right\| / \left\| \hat{\beta}^{(m)} \right\| < 10^{-6},$$

or when 2000 iterations of the EM algorithm have been reached, where $\|\cdot\|$ denotes the general vector norm operator.

Notice that each MC simulation corresponds to an estimation obtained using the data from M independent experiments.

On the other hand, for the numerical examples, the initialization of the estimation algorithm is as follows: For the nominal model initialization, we estimate a system model for each independent experiment using PEM and a sufficiently flexible (high order) structure. Then, we choose the estimated system model and the corresponding estimated noise variance with the smallest number of parameters that better described all the previously estimated models. For the error-model FIR transfer function, we adopt a similar approach for different FIR filter orders (n_Δ). Here, we choose a sufficiently complex FIR system by assuming the nominal model is equal to one ($G_o(q^{-1}, \theta) = 1$). The number of Gaussian mixture model components (κ) is defined by the user, and it is typically a large number. Once the parameters have been estimated, it is possible to discard the Gaussian component corresponding to small estimated weight (e.g., $1/(100\kappa)$). Finally, for a given κ , the initial distribution of the parameters of the error-model is chosen as follows:

- (1) From all the estimated FIR models, compute the average value of the coefficient corresponding to each tap.
- (2) Evenly space the initial mean values of the n_Δ -dimensional GMM between the estimated maximum and the minimum value of each tap.
- (3) Set the variances of the n_Δ -dimensional GMM equal to a diagonal covariance matrix with the sample variance of each tap on the diagonal.
- (4) Set the mixing weight for each GMM component equal to $1/\kappa$.

5.1. Example 1: A General System with a Gaussian Mixture Error-Model Distribution

Consider the *true* system $G^{[r]}(q^{-1})$ in Equation (2) with a nominal system model as follows:

$$G_o(q^{-1}, \theta) = \frac{b_1^0 q^{-1}}{1 + a_1^0 q^{-1}}, \quad H(q^{-1}, \theta) = \frac{1 + c_1^0 q^{-1}}{1 + d_1^0 q^{-1}}, \quad (36)$$

and the error-model in Equation (12) is given by

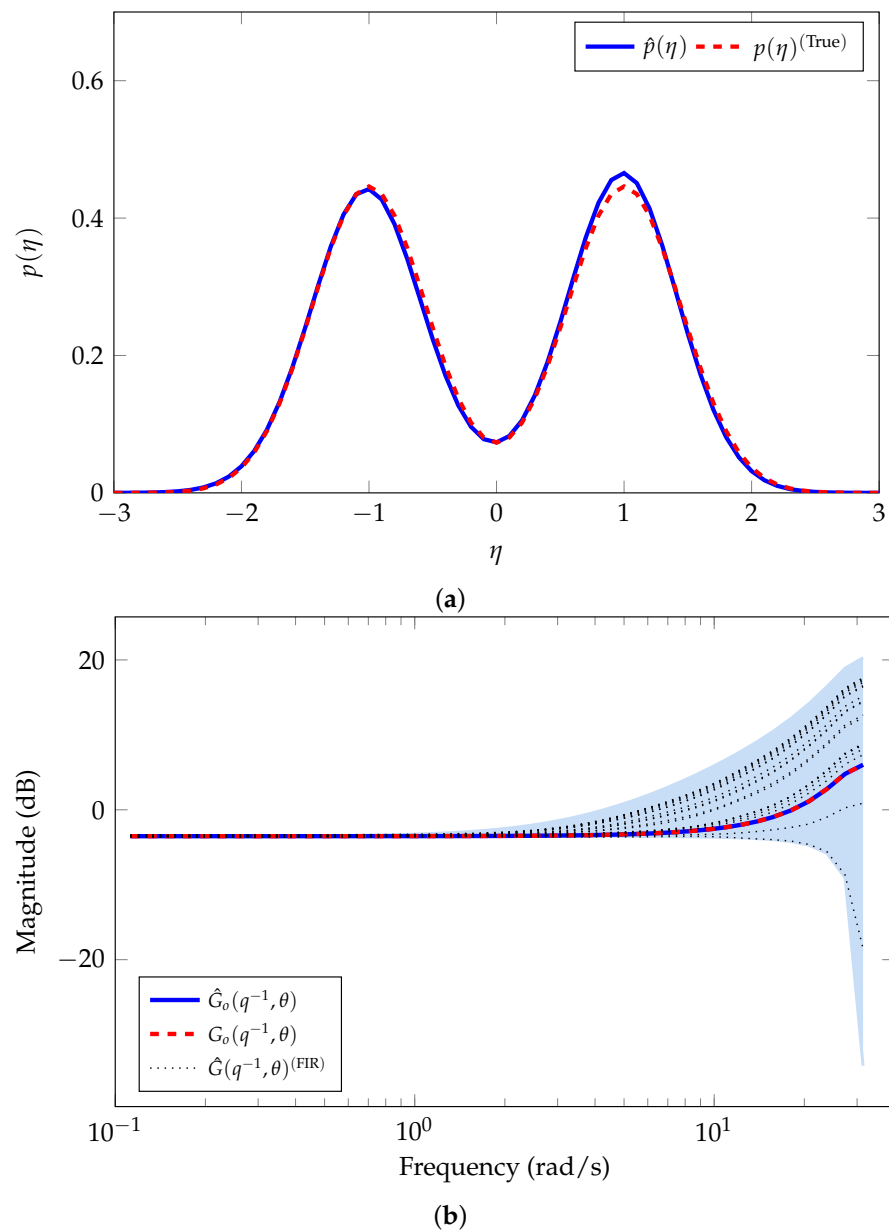
$$G_\Delta(q^{-1}, \eta^{[r]}) = \eta_0^{[r]} + \eta_1^{[r]} q^{-1}, \quad (37)$$

where $b_1^0 = 1$, $a_1^0 = 0.5$, $c_1^0 = 0.1$, and $d_1^0 = 0.8$. The input signal is $u_t^{[r]} \sim \mathcal{N}(0, \sigma_u^2)$, $\sigma_u^2 = 10$, and the noise source $\omega_t^{[r]} \sim \mathcal{N}(0, \sigma^2)$, $\sigma^2 = 0.1$. In this example, we focus on uncertainties in the system model at high frequencies, i.e., $\eta_1 = -\eta_0$. We also consider that, for each experiment, $\eta^{[r]} = [\eta_0 \ \eta_1]^T$ is drawn from a Gaussian mixture distribution in Equation (4) with $\alpha_1 = \alpha_2 = 0.5$, $\Gamma_1 = \Gamma_2 = 0.2$, $\mu_1 = -1$, and $\mu_2 = 1$. The vector of parameters to estimate is $\beta = \{\theta, \gamma, \sigma^2\}$ with $\theta = \{b_1^0, a_1^0, c_1^0, d_1^0\}$ and $\gamma = \{\alpha_i, \mu_i, \Gamma_i\}_{i=1}^\kappa$ with $\kappa = 2$. The initial values used for our proposed EM algorithm correspond to $\hat{\theta}^{(0)} = [1.2510 \ 0.4523 \ 0.0772 \ 0.7546]^T$, $\hat{\sigma}^{2(0)} = 0.0612$, $\hat{\alpha}_1^{(0)} = \hat{\alpha}_2^{(0)} = 0.5$, $\hat{\mu}_1^{(0)} = -1.1962$, $\hat{\mu}_2^{(0)} = 2.9673$, $\hat{\Gamma}_1^{(0)} = \hat{\Gamma}_2^{(0)} = 0.8651$.

Table 1 shows the estimation results of the nominal model parameters and the noise variance. Figure 3a shows the results of the estimation of the error-model distribution. The blue line corresponds to the average of all the GMMs estimated. It is clear that the estimated PDF is similar to the *true* PDF.

Table 1. Nominal model parameters and noise variance estimated for Example 1.

Parameter	True Value	Estimated Value
b_1	1	$0.9994 \pm 1.54 \times 10^{-3}$
a_1	0.5	$0.4998 \pm 4.53 \times 10^{-4}$
c_1	0.1	$0.0997 \pm 1.17 \times 10^{-2}$
d_1	0.8	$0.8017 \pm 5.97 \times 10^{-3}$
σ^2	0.1	$0.1003 \pm 1.44 \times 10^{-3}$

**Figure 3.** Nominal model and error-model distribution estimated for Example 1. (a) Estimated error-model distribution. (b) Estimated nominal model $G_o(q^{-1}, \theta)$.

On the other hand, we note that the SE method does not directly provide an uncertainty region. However, we compute M realizations of the model

$$\hat{G}_o(q^{-1}, \hat{\theta}) + \hat{G}_o(q^{-1}, \hat{\theta})G_\Delta(q^{-1}, \eta^{[r]}), \quad (38)$$

with $r = 1, \dots, M$, where $\hat{G}_o(q^{-1}, \hat{\theta})$ is the nominal model estimated and $\eta^{[r]}$ are obtained from M different realizations using the GMM in Equation (4) with the estimated parameters

$\hat{\gamma}$ obtained by using our proposed SE method. In addition, we use the PEM [1,2] to estimate the system model $G^{[r]}(q^{-1})$ with a high order FIR model. We consider the measurements of 25 independent experiments, and we estimate a 10th order FIR system model, $\hat{G}(q^{-1}, \theta)^{(FIR)}$, for each independent experiment in order to validate the uncertainty region estimated.

Figure 3b shows the magnitude of the frequency response corresponding to the average of all MC simulations for the estimated nominal model (blue line). The red-dashed line corresponds to the nominal system model in Equation (36), i.e., $G_o(q^{-1}, \theta)$. The blue-shaded region represents the area of the corresponding uncertainty region estimated of the *true* system. We observe an accurate estimation of the nominal system model. We also observe that the FIR models estimated (black-dotted lines) lie in the uncertainty region obtained with our proposed method.

5.2. Example 2: A Resistor-Capacitor System Model with Non-Gaussian-Sum Uncertainties

Consider the system in Equation (2) as follows:

$$G^{[r]}(q^{-1}, \tau^{[r]}) = \frac{q^{-1} \left(1 - e^{-T_s/\tau^{[r]}} \right)}{1 - \left(e^{-T_s/\tau^{[r]}} \right) q^{-1}}, \quad H(q^{-1}) = 1, \quad (39)$$

where $T_s = 0.01$ is the sampled period, and $\tau^{[r]} = \tau_o(1 + \delta^{[r]})$ is the time constant with $\tau_o = 10 \times 10^{-3}$, and $\delta^{[r]}$ is uniformly distributed as $\mathcal{U}[-0.4, 0.4]$. We consider that the sampled period T_s is known.

For the SE method, we consider an output-error nominal model with a FIR error-model as follows:

$$G_o(q^{-1}, \theta) = \frac{q^{-1} \left(1 - e^{-T_s/\tau_o} \right)}{1 - \left(e^{-T_s/\tau_o} \right) q^{-1}}, \quad (40)$$

$$G_{\Delta}(q^{-1}, \eta^{[r]}) = \eta_0 - \eta_0 q^{-1}. \quad (41)$$

The distribution of the parameters $\eta^{[r]}$ of the error-model is modeled using a GMM with $\gamma = \{\alpha_i, \mu_i, \Gamma_i\}_{i=1}^{\kappa}$, $\kappa = 2$ components. The vector of parameter to be estimated is $\beta = \{\tau_o, \gamma, \sigma^2\}$. As in the previous example, the uncertainty region is obtained from Equation (38) with $r = 1, \dots, M$, where $\eta^{[r]}$ is obtained from M different realizations using the GMM in Equation (4) with the estimated parameters $\hat{\gamma}$. In addition, the initial values used for our proposed EM algorithm are given by $\hat{\tau}_0^{(0)} = 9 \times 10^{-3}$, $\hat{\sigma}^{2(0)} = 0.1247$, $\hat{\alpha}_1^{(0)} = \hat{\alpha}_2^{(0)} = 0.5$, $\hat{\mu}_1^{(0)} = -0.0163$, $\hat{\mu}_2^{(0)} = 0.8025$, $\hat{\Gamma}_1^{(0)} = \hat{\Gamma}_2^{(0)} = 0.1092$.

For comparison purposes, we estimate the system uncertainty and the nominal model using MEM approach [33]. The nominal model estimated, $\hat{G}_o(q^{-1}, \theta)^{(MEM)}$, corresponds to the average of all PEM estimations obtained using the original structure given in Equation (39), i.e., $n_a = n_b = 1$ and $n_c = n_d = 0$, from the all independent experiments. In order to obtain the residuals, we compute $\varepsilon_t = y_t - \hat{G}_o(q^{-1}, \theta)^{(MEM)} u_t$ for each independent experiment. The error-model, $G_{\varepsilon}(q^{-1}, \theta_{\varepsilon})$, is obtained from the residual data as follows:

$$\varepsilon_t = G_{\varepsilon}(q^{-1}, \theta_{\varepsilon}) u_t + v_t, \quad (42)$$

where v_t is a zero-mean Gaussian noise sequence which is assumed to be uncorrelated with the input signal u_t . Thus, we consider a 10th order FIR system model for $G_{\varepsilon}(q^{-1}, \theta_{\varepsilon})$ and we use the PEM [2] to estimate the error-model parameters θ_{ε} for each experiment. The uncertainty region is computed by adding frequency by frequency the average of the all error-model estimated to the nominal model.

Finally, we consider the measurements of 25 independent experiments and we obtain an estimation of the system model in Equation (39), $\hat{G}(q^{-1}, \theta)^{(PEM)}$, using PEM with $n_a = n_b = 1$ and $n_c = n_d = 0$ for each independent experiment in order to validate the uncertainty region estimated for both SE and MEM approach.

Figure 4a shows the magnitude of the frequency response corresponding to the nominal model and the uncertainty region estimated using SE approach. The red-dashed line corresponds to the system model in Equation (39) considering the time constant without uncertainty, i.e., $G(q^{-1}, \tau_o)$. The blue line corresponds to the average of all MC simulations for the estimated nominal model. We observe an accurate estimation for the nominal model with $\hat{\tau}_o = 9.7 \times 10^{-3} \pm 0.17 \times 10^{-3}$ and noise variance $\hat{\sigma}^2 = 0.133 \pm 5.9 \times 10^{-3}$. The blue-shaded area corresponds to the uncertainty region estimated for the *true* system. Similarly, Figure 4b shows the magnitude of the frequency response corresponding to the nominal model estimated (blue line) and the uncertainty region estimated (blue-shaded area) from the residuals using MEM. We also observe an accurate estimation of the nominal model with $\hat{\tau}_o = 10.1 \times 10^{-3} \pm 2.30 \times 10^{-3}$ and noise variance $\hat{\sigma}^2 = 0.099 \pm 2.85 \times 10^{-3}$. In addition, we observe in Figure 4a,b that PEM estimations, $\hat{G}(q^{-1}, \theta)^{(PEM)}$, lie in the uncertainty region obtained using both SE approach and MEM. However, SE method describes better the uncertainty region than MEM method, specifically, the error-model at low frequencies.

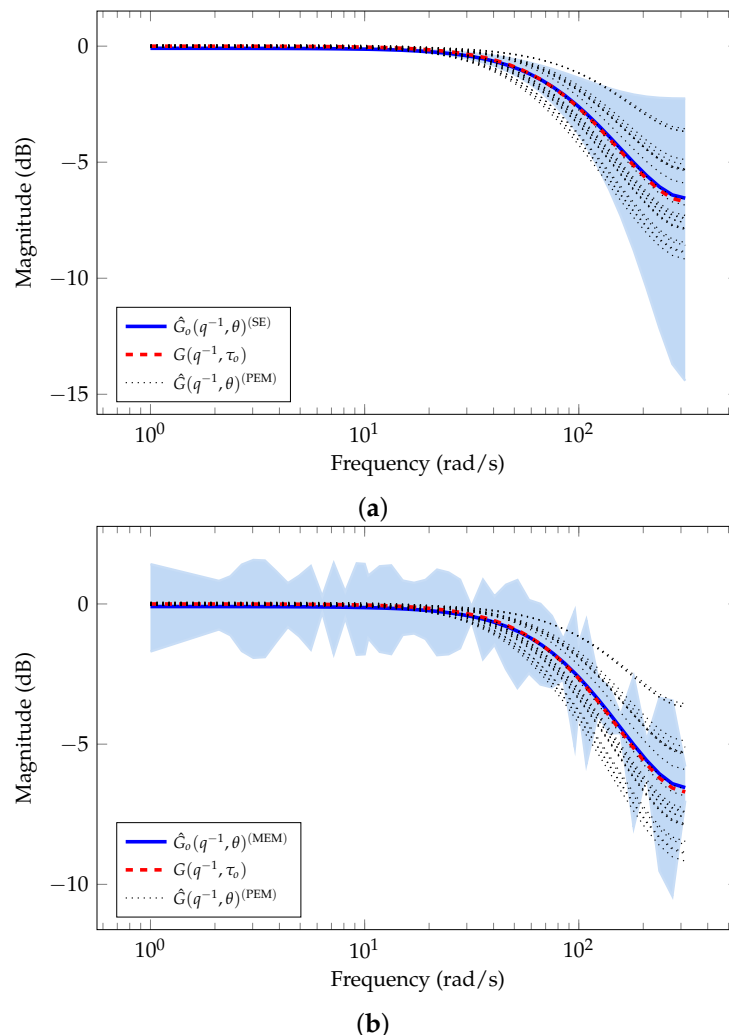


Figure 4. Nominal model and uncertainty region estimated using SE and MEM for Example 2. (a) System model estimated with SE. (b) System model estimated with MEM.

6. Conclusions

In this paper, we have addressed the uncertainty modeling problem by combining the SE approach with GMMs. We proposed an identification algorithm using ML principle to estimate both the nominal model and the distribution of the parameters of the error-model as a GMM. An EM algorithm is proposed to solve the ML estimation problem providing

closed-form expressions for the estimators of the GMM parameters and the noise variance. In the simulation examples, we considered two scenarios: (i) identification of a system model with a structure defined with SE approach using a nominal model and an error-model as a FIR system with Gaussian mixture distribution, and (ii) identification of a system model with non-Gaussian uncertainties in the parameters does not correspond to the SE framework. For both cases, we obtained accurate estimations of the nominal system model. Finally, our proposed method provided a good description of the uncertainty region in the system for the simulation examples.

Author Contributions: Conceptualization, R.O. and J.C.A.; formal analysis, R.O., R.C. and J.C.A.; funding acquisition, R.O., R.C. and J.C.A.; investigation, R.O., R.C., P.E. and J.C.A.; methodology, R.O. and J.C.A.; resources, J.C.A.; software, R.O.; validation, P.E.; visualization, P.E.; writing—original draft preparation, R.O., R.C., P.E. and J.C.A.; writing, review, and editing, R.O., R.C., P.E., J.C.A. All authors have read and agreed to the published version of the manuscript.

Funding: PIIC 05/2019 of DPP at Universidad Técnica Federico Santa María; ANID-FONDECYT 1211630 and 11201187; Advanced Center for Electrical and Electronic Engineering, AC3E, Basal Project FB0008, ANID, Chile.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: No new data were created or analyzed in this study. Data sharing is not applicable to this article.

Conflicts of Interest: The authors declare no conflict of interest.

Abbreviations

The following abbreviations are used in this manuscript:

EM:	Expectation-Maximization
FIR:	Finite impulse response
GMM:	Gaussian mixture model
MC:	Monte Carlo
MEM:	Model error modeling
ML:	Maximum likelihood
PDF:	Probability density function
PEM:	Prediction error method
SE:	Stochastic embedding

Appendix A. Maximum Likelihood Formulation with SE Approach

Let us consider that $Y^{[r]} = [y_1^{[r]} \dots y_N^{[r]}]$ is a collection of measurements for each experiment $r = 1, \dots, M$. Then, the likelihood function, $\mathcal{L}(\beta)$, for the system model in Equation (13) using the GMM in Equation (4) can be obtained by marginalizing the hidden variable, $[\eta^{[1]} \dots \eta^{[M]}]$, as follows:

$$\begin{aligned} \mathcal{L}(\beta) &= p(Y^{[1]}, \dots, Y^{[M]} | \beta) \\ &= \prod_{r=1}^M \int_{-\infty}^{\infty} p(Y^{[r]} | \eta^{[r]}, \beta) p(\eta^{[r]} | \beta) d\eta^{[r]}, \end{aligned} \quad (\text{A1})$$

where

$$p(Y^{[r]} | \eta^{[r]}, \beta) = \mathcal{N}(Y^{[r]}; \mathcal{G}(\theta)U^{[r]} + \Psi^{[r]}(\theta)\eta^{[r]}, \sigma^2 I_N), \quad (\text{A2})$$

$$p(\eta^{[r]} | \beta) = \sum_{i=1}^K \alpha_i \mathcal{N}(\eta^{[r]}; \mu_i, \Gamma_i), \quad (\text{A3})$$

with $\alpha_i > 0$, $\sum_{i=1}^{\kappa} \alpha_i = 1$. The log-likelihood function, $\ell(\beta)$, is then given by

$$\ell(\beta) = \sum_{r=1}^M \log \left\{ \sum_{i=1}^{\kappa} \alpha_i \int_{-\infty}^{\infty} p(Y^{[r]}|\eta^{[r]}, \beta) p(\eta^{[r]}|\beta) d\eta^{[r]} \right\}. \quad (\text{A4})$$

Consider the following identities:

$$\begin{bmatrix} A & B \\ C & D \end{bmatrix} = \begin{bmatrix} I & 0 \\ CA^{-1} & I \end{bmatrix} \begin{bmatrix} A & 0 \\ 0 & D - CA^{-1}B \end{bmatrix} \begin{bmatrix} I & A^{-1}B \\ 0 & I \end{bmatrix}, \quad (\text{A5})$$

$$(I + BD^{-1}C)^{-1} = I - B(D + CB)^{-1}C, \quad (\text{A6})$$

$$\det(A) \det(D - CA^{-1}B) = \det(D) \det(A - BD^{-1}C). \quad (\text{A7})$$

From Equation (A4), we define $K_i(\beta, \eta^{[r]}) = p(Y^{[r]}|\eta^{[r]}, \beta) p(\eta^{[r]}|\beta)$. Substituting Equations (A2) and (A3) in $K_i(\beta, \eta^{[r]})$, and using the identities in Equations (A5)–(A7), we obtain:

$$K_i(\beta, \eta^{[r]}) = \mathcal{N}(Y^{[r]}; \bar{\mu}_{ir}, \bar{\Sigma}_{ir}) \mathcal{N}(\eta^{[r]}; \tilde{\mu}_{ir}, \tilde{\Sigma}_{ir}), \quad (\text{A8})$$

with

$$\bar{\mu}_{ir} = \Psi^{[r]}(\theta) \mu_i + \mathcal{G}(\theta) U^{[r]}, \quad (\text{A9})$$

$$\bar{\Sigma}_{ir} = \sigma^2 I_N + \Psi^{[r]}(\theta) \Gamma_i \left[\Psi^{[r]}(\theta) \right]^T, \quad (\text{A10})$$

$$\mathcal{P}_{ir} = \Gamma_i \left[\Psi^{[r]}(\theta) \right]^T (\bar{\Sigma}_{ir})^{-1}, \quad (\text{A11})$$

$$\tilde{\mu}_{ir} = \mu_i + \mathcal{P}_{ir} (Y^{[r]} - \bar{\mu}_{ir}), \quad (\text{A12})$$

$$\tilde{\Sigma}_{ir} = \left(I_{n_\Delta} - \mathcal{P}_{ir} \Psi^{[r]}(\theta) \right) \Gamma_i. \quad (\text{A13})$$

Substituting Equation (A8) in Equation (A4), we can directly obtain the log-likelihood function in Equation (16). Then, the ML estimator corresponds to Equation (15).

Appendix B. Computing the Complete Likelihood Function

From the likelihood function in Equation (16), we have:

$$p(Y^{[r]}|\beta) = \sum_{i=1}^{\kappa} \alpha_i \mathcal{N}(Y^{[r]}; \bar{\mu}_{ir}, \bar{\Sigma}_{ir}), \quad (\text{A14})$$

where $\bar{\mu}_{ir}$ and $\bar{\Sigma}_{ir}$ are given by Equations (A9) and (A10), respectively. Let consider

$$p(\zeta_r = i) = \alpha_i, \quad i = 1, \dots, \kappa, \quad (\text{A15})$$

where $\zeta_{1:M}$ is a discrete random variable (hidden variable) such that ζ_r is an indicator that determines if the r th observation $Y^{[r]}$ arises from the i th component of the GMM in Equation (A3), that is,

$$\zeta_r \in \{1, \dots, \kappa\}, \quad r = 1, \dots, M. \quad (\text{A16})$$

Then, Equation (A14) can be obtained by marginalizing the hidden variable, ζ_r , as follows:

$$p(Y^{[r]}|\beta) = \sum_{i=1}^{\kappa} p(Y^{[r]}|\zeta_r = i, \beta) p(\zeta_r = i|\beta), \quad (\text{A17})$$

where

$$p(Y^{[r]}|\zeta_r = i, \beta) = \mathcal{N}(Y^{[r]}; \bar{\mu}_{ir}, \bar{\Sigma}_{ir}). \quad (\text{A18})$$

Defining the complete data set as $\{Y^{[1:M]}, \zeta_{1:M}\}$ and utilizing the Bayes' rule, the complete likelihood function is obtained as follows:

$$p(Y^{[1:M]}, \zeta_{1:M}|\beta) = \prod_{r=1}^M p(Y^{[r]}|\zeta_r, \beta)p(\zeta_r|\beta). \quad (\text{A19})$$

The complete log-likelihood function is given by:

$$\log [p(Y^{[1:M]}, \zeta_{1:M}|\beta)] = \sum_{r=1}^M \log [p(\zeta_r|\beta)] + \log [p(Y^{[r]}|\zeta_r, \beta)]. \quad (\text{A20})$$

Using Equations (A16) and (A17) in Equation (A20), we directly obtain Equation (21).

Appendix C. EM Algorithm with GMMs Formulation

Appendix C.1. Computing the Auxiliary Function $Q(\beta, \hat{\beta}^{(m)})$

Using Equation (21) in Equation (19), we have:

$$Q(\beta, \hat{\beta}^{(m)}) = \sum_{r=1}^M \sum_{i=1}^K \mathbb{E} \left\{ \log [p(\zeta_r = i|\beta)] | Y^{[1:M]}, \hat{\beta}^{(m)} \right\} + \mathbb{E} \left\{ \log [\mathcal{N}(Y^{[r]}; \bar{\mu}_{ir}, \bar{\Sigma}_{ir})] | Y^{[1:M]}, \hat{\beta}^{(m)} \right\}. \quad (\text{A21})$$

Solving the expected value in Equation (A21) and using Equation (A15) into Equation (A21), we have:

$$Q(\beta, \hat{\beta}^{(m)}) = \sum_{r=1}^M \sum_{i=1}^K \log [\alpha_i] \hat{\zeta}_{ir}^{(m)} + \log [\mathcal{N}(Y^{[r]}; \bar{\mu}_{ir}, \bar{\Sigma}_{ir})] \hat{\zeta}_{ir}^{(m)}, \quad (\text{A22})$$

where

$$\hat{\zeta}_{ir}^{(m)} = p(\zeta_r = i | Y^{[1:M]}, \hat{\beta}^{(m)}). \quad (\text{A23})$$

Using the Bayes' Theorem, Equations (A14), (A15), and (A17) to solve Equation (A23), we obtain:

$$\hat{\zeta}_{ir}^{(m)} = \frac{p(\zeta_r = i | \hat{\beta}^{(m)}) p(Y^{[r]} | \zeta_r = i, \hat{\beta}^{(m)})}{p(Y^{[r]} | \hat{\beta}^{(m)})}, \quad (\text{A24})$$

$$= \frac{\hat{\alpha}_i^{(m)} \mathcal{N}(Y^{[r]}; \hat{\mu}_{ir}^{(m)}, \hat{\Sigma}_{ir}^{(m)})}{\sum_{l=1}^K \hat{\alpha}_l^{(m)} \mathcal{N}(Y^{[r]}; \hat{\mu}_{lr}^{(m)}, \hat{\Sigma}_{lr}^{(m)})}, \quad (\text{A25})$$

where $\hat{\mu}_{ir}^{(m)}$ and $\hat{\Sigma}_{ir}^{(m)}$ are given by Equations (24) and (25), respectively.

Appendix C.2. Optimizing the Auxiliary Function $Q(\beta, \hat{\beta}^{(m)})$

From Equation (A8), we have the following:

$$\int_{-\infty}^{\infty} \log [K_i(\beta, \eta^{[r]})] d\eta^{[r]} = \log [\mathcal{N}(Y^{[r]}; \bar{\mu}_{ir}, \bar{\Sigma}_{ir})]. \quad (\text{A26})$$

Let consider the following identity for a random variable x :

$$\mathbb{E} \left\{ (b - Ax)^T (b - Ax) \right\} = (b - A\hat{x})^T (b - A\hat{x}) + \text{tr} \left\{ A \Sigma A^T \right\}, \quad (\text{A27})$$

where $\hat{x} = \mathbb{E}\{x\}$, and $\Sigma = \mathbb{E}\{(x - \hat{x})(x - \hat{x})^T\}$.

Using Equation (A27) to compute $\log [K_i(\beta, \eta^{[r]})]$ in Equation (A26), we obtain:

$$\begin{aligned} \log [K_i(\beta, \eta^{[r]})] &= \frac{N}{2} \log [\sigma^2] - \frac{1}{2} \log [\det(\Gamma_i)] - \frac{1}{2\sigma^2} \left[(Y^{[r]} - \mathcal{G}(\theta)U^{[r]} - \Psi^{[r]}(\theta)\hat{\mu}_{ir}^{(m)})^T \right. \\ & (Y^{[r]} - \mathcal{G}(\theta)U^{[r]} - \Psi^{[r]}(\theta)\hat{\mu}_{ir}^{(m)}) + \text{tr}(\Psi^{[r]}(\theta)\hat{\Sigma}_{ir}^{(m)}[\Psi^{[r]}(\theta)]^T) \left. \right] - \frac{1}{2} \left[\text{tr}(\Gamma_i^{-1}\hat{\Sigma}_{ir}^{(m)}) + \right. \\ & \left. (\hat{\mu}_{ir}^{(m)} - \mu_i)^T \Gamma_i^{-1} (\hat{\mu}_{ir}^{(m)} - \mu_i) \right]. \end{aligned} \quad (\text{A28})$$

Substituting Equations (A26) and (A28) in Equation (A22), and taking the derivative of Equation (A22) with respect to μ_i and equating to zero, yields:

$$\frac{\partial \mathcal{Q}(\beta, \hat{\beta}^{(m)})}{\partial \mu_i} = \sum_{r=1}^M (\hat{\mu}_{ir}^{(m)} - \hat{\mu}_i^{(m+1)}) \hat{\zeta}_{ir}^{(m)} = 0. \quad (\text{A29})$$

From Equation (A29), we directly obtain Equation (27).

Then, taking the derivative of Equation (A22) with respect to Γ_i^{-1} and equating to zero:

$$\frac{\partial \mathcal{Q}(\beta, \hat{\beta}^{(m)})}{\partial \Gamma_i^{-1}} = \sum_{r=1}^M \left((\hat{\mu}_{ir}^{(m)} - \hat{\mu}_i^{(m)}) (\hat{\mu}_{ir}^{(m)} - \hat{\mu}_i^{(m)})^T + \hat{\Sigma}_{ir}^{(m)} \right) - \hat{\Gamma}_i^{(m+1)} \hat{\zeta}_{ir}^{(m)} = 0. \quad (\text{A30})$$

From Equation (A30), we can directly obtain Equation (28).

Then, taking the derivative of Equation (A22) with respect to $\rho = 1/\sigma^2$ and equating to zero yields:

$$\begin{aligned} \frac{1}{N} \sum_{r=1}^M \sum_{i=1}^{\kappa} \left[(Y^{[r]} - \mathcal{G}(\hat{\theta}^{(m)})U^{[r]} - \Psi^{[r]}(\hat{\theta}^{(m)})\hat{\mu}_{ir}^{(m)})^T (Y^{[r]} - \mathcal{G}(\hat{\theta}^{(m)})U^{[r]} - \Psi^{[r]}(\hat{\theta}^{(m)})\hat{\mu}_{ir}^{(m)}) + \right. \\ \left. \text{tr}(\Psi^{[r]}(\hat{\theta}^{(m)})\hat{\Sigma}_{ir}^{(m)}[\Psi^{[r]}(\hat{\theta}^{(m)})]^T) \right] \hat{\zeta}_{ir}^{(m)} = \sum_{r=1}^M \sum_{i=1}^{\kappa} \hat{\rho}^{(m+1)} \hat{\zeta}_{ir}^{(m)}. \end{aligned} \quad (\text{A31})$$

Using $\hat{\sigma}^{2(m+1)} = 1/\hat{\rho}^{(m+1)}$ and Equation (33) in Equation (A31), we directly obtain Equation (29).

For the parameter α_i , we define $\mathcal{F}(\alpha_i)$ as follows:

$$\mathcal{F}(\alpha_i) = \sum_{r=1}^M \sum_{i=1}^{\kappa} \log [\alpha_i] \hat{\zeta}_{ir}^{(m)}, \quad (\text{A32})$$

subject to $\sum_{i=1}^{\kappa} \alpha_i = 1$. Then, using a Lagrange multiplier to deal with the constraint on α_i , we define:

$$\mathcal{J}(\alpha_i, \nu) = \sum_{r=1}^M \sum_{i=1}^{\kappa} \log [\alpha_i] \hat{\zeta}_{ir}^{(m)} + \nu \left(\sum_{i=1}^{\kappa} \alpha_i - 1 \right). \quad (\text{A33})$$

Taking the derivative of Equation (A33) with respect to α_i and ν and equating to zero, we obtain:

$$\frac{\partial \mathcal{J}(\alpha_i, \nu)}{\partial \alpha_i} = \sum_{r=1}^M \frac{\hat{\zeta}_{ir}^{(m)}}{\hat{\alpha}_i^{(m+1)}} - \nu = 0, \quad (\text{A34})$$

$$\frac{\partial \mathcal{J}(\alpha_i, \nu)}{\partial \nu} = \sum_{i=1}^{\kappa} \alpha_i - 1 = 0. \quad (\text{A35})$$

Then, $\hat{\alpha}_i^{(m+1)} = \sum_{r=1}^M \hat{\zeta}_{ir}^{(m)} / \nu$. Taking summation over $i = 1, \dots, \kappa$ and using Equation (A35), we have:

$$\sum_{i=1}^{\kappa} \hat{\alpha}_i^{(m+1)} = \sum_{i=1}^{\kappa} \sum_{r=1}^M \hat{\zeta}_{ir}^{(m)} / \nu, \quad (\text{A36})$$

$$v = \sum_{i=1}^{\kappa} \sum_{r=1}^M \hat{\zeta}_{ir}^{(m)}. \quad (\text{A37})$$

Notice that $\sum_{i=1}^{\kappa} \hat{\zeta}_{ir}^{(m)} = 1$, and then we have $v = M$. Then, we can obtain Equation (26). Notice that $0 \leq \hat{a}_i^{(m+1)} \leq 1$ holds even though we did not explicitly consider it in Equation (A33).

Finally, substituting Equations (A26) and (A28) in Equation (A22), we can directly obtain Equation (35) for the parameter θ .

References

- Söderström, T.; Stoica, P. *System Identification*; Prentice-Hall, Inc.: Upper Saddle River, NJ, USA, 1988.
- Ljung, L. *System Identification: Theory for the User*; Prentice Hall PTR: Upper Saddle River, NJ, USA, 1999.
- Goodwin, G.C.; Payne, R.L. *Dynamic System Identification: Experiment Design and Data Analysis*; Academic Press: New York, NY, USA, 1977.
- Ljung, L.; Goodwin, G.C.; Agüero, J.C. Stochastic Embedding revisited: A modern interpretation. In Proceedings of the 53rd IEEE Conference on Decision and Control, Los Angeles, CA, USA, 15–17 December 2014; pp. 3340–3345.
- Jategaonkar, R.V. *Flight Vehicle System Identification: A Time Domain Methodology*, 2nd ed.; Progress in Astronautics and Aeronautics; v. 245; American Institute of Aeronautics and Astronautics, Inc.: Reston, VA, USA, 2006.
- Han, Z.; Li, W.; Shah, S.L. Robust Fault Diagnosis in the Presence of Process Uncertainties. *IFAC Proc. Vol.* **2002**, *35*, 473–478. [[CrossRef](#)]
- Nikoukhah, R.; Campbell, S. A Multi-Model Approach to Failure Detection in Uncertain Sampled-Data Systems. *Eur. J. Control* **2005**, *11*, 255–265. [[CrossRef](#)]
- Patton, R.J.; Frank, P.M.; Clark, R.N. *Issues of Fault Diagnosis for Dynamic Systems*, 1st ed.; Springer Publishing Company, Inc.: Berlin/Heidelberg, Germany, 2010.
- Escobet, T.; Bregon, A.; Pulido, B.; Puig, V. *Fault Diagnosis of Dynamic Systems*; Springer: Cham, Switzerland, 2019.
- Hamel, P.G.; Jategaonkar, R.V. Evolution of flight vehicle system identification. *J. Aircr.* **1996**, *33*, 9–28. [[CrossRef](#)]
- Jategaonkar, R.; Thielecke, F. Aircraft parameter estimation—A tool for development of aerodynamic databases. *Sadhana* **2006**, *25*, 119–135. [[CrossRef](#)]
- Rosić, B.V.; Diekmann, J.H. Methods for the Uncertainty Quantification of Aircraft Simulation Models. *J. Aircr.* **2015**, *52*, 1247–1255. [[CrossRef](#)]
- Mu, H.Q.; Liu, H.T.; Shen, J.H. Copula-Based Uncertainty Quantification (Copula-UQ) for Multi-Sensor Data in Structural Health Monitoring. *Sensors* **2020**, *20*, 5692. [[CrossRef](#)]
- Hilton, S.; Cairola, F.; Gardi, A.; Sabatini, R.; Pongsakornsathien, N.; Ezer, N. Uncertainty Quantification for Space Situational Awareness and Traffic Management. *Sensors* **2019**, *19*, 4361. [[CrossRef](#)]
- Guo, J.; Zhao, Y. Recursive projection algorithm on FIR system identification with binary-valued observations. *Automatica* **2013**, *49*, 3396–3401. [[CrossRef](#)]
- Moschitta, A.; Schoukens, J.; Carbone, P. Parametric System Identification Using Quantized Data. *IEEE Trans. Instrum. Meas.* **2015**, *64*, 2312–2322. [[CrossRef](#)]
- Cedeño, A.L.; Orellana, R.; Carvajal, R.; Agüero, J.C. EM-based identification of static errors-in-variables systems utilizing Gaussian Mixture models. *IFAC-PapersOnLine* **2020**, *53*, 863–868. [[CrossRef](#)]
- Orellana, R.; Escárate, P.; Curé, M.; Christen, A.; Carvajal, R.; Agüero, J.C. A method to deconvolve stellar rotational velocities-III. The probability distribution function via maximum likelihood utilizing finite distribution mixtures. *Astron. Astrophys.* **2019**, *623*, A138. [[CrossRef](#)]
- Agüero, J.C.; Goodwin, G.C.; Yuz, J.I. System identification using quantized data. In Proceedings of the 46th IEEE Conference on Decision and Control, New Orleans, LA, USA, 12–14 December 2007; pp. 4263–4268.
- Godoy, B.I.; Goodwin, G.C.; Agüero, J.C.; Marelli, D.; Wigren, T. On identification of FIR systems having quantized output data. *Automatica* **2011**, *47*, 1905–1915. [[CrossRef](#)]
- Godoy, B.I.; Agüero, J.C.; Carvajal, R.; Goodwin, G.C.; Yuz, J.I. Identification of sparse FIR systems using a general quantisation scheme. *Int. J. Control* **2014**, *87*, 874–886. [[CrossRef](#)]
- Agüero, J.C.; González, K.; Carvajal, R. EM-based identification of ARX systems having quantized output data. *IFAC-PapersOnLine* **2017**, *50*, 8367–8372. [[CrossRef](#)]
- Carvajal, R.; Godoy, B.I.; Agüero, J.C.; Goodwin, G.C. EM-based sparse channel estimation in OFDM systems. In Proceedings of the IEEE 13th International Workshop on Signal Processing Advances in Wireless Communications, Çeşme, Turkey, 17–20 June 2012; pp. 530–534.
- Carvajal, R.; Agüero, J.C.; Godoy, B.I.; Goodwin, G.C. EM-Based Maximum-Likelihood Channel Estimation in Multicarrier Systems With Phase Distortion. *IEEE Trans. Veh. Technol.* **2013**, *62*, 152–160. [[CrossRef](#)]
- Carnduff, S. Flight Vehicle System Identification: A Time-Domain Methodology. *Aeronaut. J.* **2015**, *119*, 930–931. [[CrossRef](#)]

26. Goodwin, G.C.; Salgado, M.E. A stochastic embedding approach for quantifying uncertainty in the estimation of restricted complexity models. *Int. J. Adapt. Control Signal Process.* **1989**, *3*, 333–356. [[CrossRef](#)]
27. Van den Hof, P.; Wahlberg, B.; Heuberger, P.; Ninness, B.; Bokor, J.; Oliveira e Silva, T. Modelling and Identification with Rational Orthogonal Basis Functions. *IFAC Proc. Vol.* **2000**, *33*, 445–455. [[CrossRef](#)]
28. Peralta, P.; Ruiz, R.O.; Taflanidis, A.A. Bayesian identification of electromechanical properties in piezoelectric energy harvesters. *Mech. Syst. Signal Process.* **2020**, *141*, 106506. [[CrossRef](#)]
29. Goodwin, G.; Graebe, S.; Salgado, M.E. *Control Systems Design*; Prentice Hall: Upper Saddle River, NJ, USA, 2001.
30. Douma, S.G.; Van den Hof, P.M. Relations between uncertainty structures in identification for robust control. *Automatica* **2005**, *41*, 439–457. [[CrossRef](#)]
31. Calafiore, G.C.; Campi, M.C. The scenario approach to robust control design. *IEEE Trans. Autom. Control* **2006**, *51*, 742–753. [[CrossRef](#)]
32. Milanese, M.; Vicino, A. Optimal estimation theory for dynamic systems with set membership uncertainty: An overview. *Automatica* **1991**, *27*, 997–1009. [[CrossRef](#)]
33. Ljung, L. Model Error Modeling and Control Design. *IFAC Proc. Vol.* **2000**, *33*, 31–36. [[CrossRef](#)]
34. Goodwin, G.C.; Gevers, M.; Ninness, B. Quantifying the error in estimated transfer functions with application to model order selection. *IEEE Trans. Autom. Control* **1992**, *37*, 913–928. [[CrossRef](#)]
35. Delgado, R.A.; Goodwin, G.C.; Carvajal, R.; Agüero, J.C. A novel approach to model error modelling using the expectation-maximization algorithm. In Proceedings of the 51st IEEE Conference on Decision and Control, Maui, HI, USA, 10–13 December 2012; pp. 7327–7332.
36. Dempster, A.; Laird, N.; Rubin, D. Maximum likelihood from incomplete data via the EM algorithm. *J. R. Stat. Soc. Ser. B* **1977**, *39*, 1–38.
37. Ljung, L.; Goodwin, G.C.; Agüero, J.C.; Chen, T. Model Error Modeling and Stochastic Embedding. *IFAC-PapersOnLine* **2015**, *48*, 75–79. [[CrossRef](#)]
38. Pillonetto, G.; Nicolao, G.D. A new kernel-based approach for linear system identification. *Automatica* **2010**, *46*, 81–93. [[CrossRef](#)]
39. Ljung, L.; Chen, T.; Mu, B. A shift in paradigm for system identification. *Int. J. Control* **2020**, *93*, 173–180. [[CrossRef](#)]
40. Anderson, B.; Moore, J. *Optimal Filtering*; Prentice-Hall: Englewood Cliffs, NJ, USA, 1979.
41. Söderström, T. *Discrete-Time Stochastic Systems: Estimation and Control*, 2nd ed.; Springer: Berlin/Heidelberg, Germany, 2002.
42. Balenzuela, M.P.; Dahlin, J.; Bartlett, N.; Wills, A.G.; Renton, C.; Ninness, B. Accurate Gaussian Mixture Model Smoothing using a Two-Filter Approach. In Proceedings of the IEEE Conference on Decision and Control, Miami, FL, USA, 17–19 December 2018; pp. 694–699.
43. Bittner, G.; Orellana, R.; Carvajal, R.; Agüero, J.C. Maximum Likelihood identification for Linear Dynamic Systems with finite Gaussian mixture noise distribution. In Proceedings of the IEEE CHILEAN Conference on Electrical, Electronics Engineering, Information and Communication Technologies (CHILECON), Valparaíso, Chile, 13–27 November 2019; pp. 1–7.
44. Orellana, R.; Bittner, G.; Carvajal, R.; Agüero, J.C. Maximum Likelihood estimation for non-minimum-phase noise transfer function with Gaussian mixture noise distribution. *Automatica* **2021**, under review.
45. Sorenson, H.W.; Alspach, D.L. Recursive Bayesian estimation using Gaussian sums. *Automatica* **1971**, *7*, 465–479. [[CrossRef](#)]
46. Dahlin, J.; Wills, A.; Ninness, B. Sparse Bayesian ARX models with flexible noise distributions. *IFAC-PapersOnLine* **2018**, *51*, 25–30. [[CrossRef](#)]
47. Orellana, R.; Carvajal, R.; Agüero, J.C. Empirical Bayes estimation utilizing finite Gaussian Mixture Models. In Proceedings of the IEEE CHILEAN Conference on Electrical, Electronics Engineering, Information and Communication Technologies (CHILECON), Valparaíso, Chile, 13–27 November 2019; pp. 1–6.
48. Orellana, R.; Carvajal, R.; Agüero, J.C.; Goodwin, G.C. Model Error Modelling using a Stochastic Embedding approach with Gaussian Mixture Models for FIR system. *IFAC-PapersOnLine* **2020**, *53*, 845–850. [[CrossRef](#)]
49. Carvajal, R.; Orellana, R.; Katselis, D.; Escárate, P.; Agüero, J.C. A data augmentation approach for a class of statistical inference problems. *PLoS ONE* **2018**, *13*, e0208499. [[CrossRef](#)]
50. Lo, J. Finite-dimensional sensor orbits and optimal nonlinear filtering. *IEEE Trans. Inf. Theory* **1972**, *18*, 583–588. [[CrossRef](#)]
51. Yu, C.; Verhaegen, M. Data-driven fault estimation of non-minimum phase LTI systems. *Automatica* **2018**, *92*, 181–187. [[CrossRef](#)]
52. Campi, M.C. The problem of pole-zero cancellation in transfer function identification and application to adaptive stabilization. *Automatica* **1996**, *32*, 849–857. [[CrossRef](#)]
53. Heuberger, P.; Van den Hof, P.; Bosgra, O. A Generalized Orthonormal Basis for Linear Dynamical Systems. *IEEE Trans. Autom. Control* **1995**, *40*, 451–465. [[CrossRef](#)]
54. Chen, T.; Ljung, L. Implementation of algorithms for tuning parameters in regularized least squares problems in system identification. *Automatica* **2013**, *49*, 2213–2220. [[CrossRef](#)]
55. Jin, C.; Zhang, Y.; Balakrishnan, S.; Wainwright, M.; Jordan, M. Local Maxima in the Likelihood of Gaussian Mixture Models: Structural Results and Algorithmic Consequences. In Proceedings of the 29th Conference on Neural Information Processing Systems, Barcelona, Spain, 5–10 December 2016; pp. 4123–4131.
56. McLachlan, G.J.; Peel, D. *Finite Mixture Models*; John Wiley & Sons: New York, NY, USA, 2000.
57. Frühwirth-Schnatter, S.; Celeux, G.; Robert, C. *Handbook of Mixture Analysis*; Chapman and Hall/CRC: New York, NY, USA, 2018.

58. Gibson, S.; Ninness, B. Robust maximum-likelihood estimation of multivariable dynamic systems. *Automatica* **2005**, *41*, 1667–1682. [[CrossRef](#)]
59. Gopaluni, R.B. A particle filter approach to identification of nonlinear processes under missing observations. *Can. J. Chem. Eng.* **2008**, *86*, 1081–1092. [[CrossRef](#)]
60. Agüero, J.C.; Tang, W.; Yuz, J.I.; Delgado, R.; Goodwin, G.C. Dual time–frequency domain system identification. *Automatica* **2012**, *48*, 3031–3041. [[CrossRef](#)]
61. Wright, S. Coordinate descent algorithms. *Math. Program.* **2015**, *151*, 3–34. [[CrossRef](#)]
62. Lichota, P. Inclusion of the D-optimality in multisine manoeuvre design for aircraft parameter estimation. *J. Theor. Appl. Mech.* **2016**, *54*, 87–98. [[CrossRef](#)]
63. Pawełek, A.; Lichota, P. Arrival air traffic separations assessment using Maximum Likelihood Estimation and Fisher Information Matrix. In Proceedings of the 20th International Carpathian Control Conference, Krakow-Wieliczka, Poland, 26–29 May 2019; pp. 1–6.