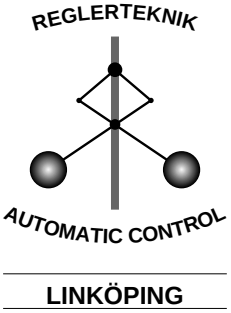
Model Error Modeling in Robust Identification (revised version)

Wolfgang Reinelt, Lennart Ljung
Andrea Garulli

Division of Automatic Control
Department of Electrical Engineering
Linköpings universitet, SE-581 83 Linköping, Sweden
WWW: http://www.control.isy.liu.se
Email: wolle@isy.liu.se, ljung@isy.liu.se

May 2001



Report No.: LiTH-ISY-R-2353
Submitted to Automatica

Comparing Different Approaches to Model Error Modeling in Robust Identification ⋆

Wolfgang Reinelt [a,1], Andrea Garulli [b] and Lennart Ljung [a]

[a] *Department of Electrical Engineering, Linköping University, 581 83 Linköping, Sweden. E-mail:* {wolle,ljung}@isy.liu.se

[b] *Dipartimento di Ingegneria dell'Informazione, Università di Siena, 53100 Siena, Italy. E-mail:* garulli@ing.unisi.it

Abstract

Identification for robust control must deliver not only a nominal model, but also a reliable estimate of the uncertainty associated with the model. This paper addresses recent approaches to robust identification, that aim at dealing with contributions from the two main uncertainty sources: unmodeled dynamics and noise affecting the data. In particular, the following methods are considered: non-stationary Stochastic Embedding, Model Error Modeling based on prediction error methods and Set Membership Identification. Moreover, we show how Set Membership Identification can be embedded into a Model Error Modeling framework. Model validation issues are easily addressed in the proposed framework. It is shown how the computation of the minimum noise bound for which a nominal model is not falsified by input-output data, can be used as a rationale for selecting an appropriate model class in the set membership setting. For all three methods, uncertainty is evaluated in terms of the frequency response, so that it can be handled by $\mathcal{H}_\infty$ control techniques. An example, where a nontrivial undermodeling is ensured by the presence of a nonlinearity in the system generating the data, is presented to compare the different methods.

Key words: identification for robust control, model error modeling, unmodeled dynamics, model validation, set membership estimation, stochastic embedding.

⋆ Preliminary versions of this work have been presented at the 1999 Conference on Decision and Control, Phoenix, AZ, USA [25] and the IFAC Symposium System Identification SYSID2000, Santa Barbara, CA, USA [7]. This paper has not been submitted to or published in another journal.

[1] Corresponding author. Fax: +46 13 282622.

1 Introduction and Motivation

One of the main objectives of control-oriented identification is to estimate models that are suitable for robust control design techniques. For this purpose, the identification procedure must deliver not only a nominal model, but also a reliable estimate of the uncertainty associated to the model. Different paradigms for the description of uncertainty have been addressed in the literature (see e.g. the special issues [14, 26] and the recent book [8]). Two main philosophies are basically adopted. The first one is based on statistical assumptions and usually leads to least squares estimation techniques and prediction error methods, see [17]. The second one relies on deterministic hypotheses, such as the identification error being unknown-but-bounded (abbreviated in the following by UBB), and has given rise to a number of techniques usually addressed as *bounded error* or *set membership* identification (see [23, 21] and the references therein).

In standard identification problems, the error originates from two different sources: a "variance" term, due to noise affecting the data, and a "bias" term, due to system dynamics which is not captured by the estimated nominal model (often addressed also as the *model error*). Clearly, the nature of these two error terms is quite different: the former is generally uncorrelated with the input signal (when the data is collected in open loop), while the latter strongly depends on the estimated nominal model and on the input used in the identification experiment. The model error is not negligible in most practical situations, especially those in which the order of the nominal model must be small (a typical requirement of robust design techniques). Moreover, while prior information on measurement noise is often available, similar hypotheses on the unmodeled dynamics seem to be less realistic.

This paper addresses three different approaches to robust identification: Stochastic Embedding, Model Error Modeling and Set Membership Identification. Each of these methods explicitly accepts the presence of bias model errors, which may motivate the term "robust identification". The first two approaches have been developed in the statistical framework, while the latter relies on UBB error assumptions. We note that the idea of explicit evaluation of bias and variance errors is used in [13] as well, and a somewhat related line of research discusses validity of a given model subject to uncertainty and noise, see [24] and the references therein.

Stochastic Embedding (SE) [11, 12] is a frequency domain method which assumes that unmodeled dynamics can be represented adequately by a non-stationary stochastic process whose variance increases with frequency. The nominal model is obtained via least-squares estimation from frequency domain data; therefore, harmonic inputs are required. The uncertainty associated with

the model is evaluated from statistical properties of the random walk process describing the unmodeled dynamics.

Model Error Modeling (MEM) [15, 16] employs standard prediction error methods to identify a nominal model from input-output time domain data [17]. Then, one can estimate the unmodeled dynamics by looking at that part of the identification residuals that originates from the input. Identification of residual dynamics (that can be performed using again prediction error methods) provides the so called *model error model*. The confidence region of a model error model allows one to evaluate the uncertainty associated to the nominal model and can be used as a model validation tool.

Set Membership Identification (SM) provides efficient algorithms for estimating the set of feasible models, compatible with the available data and the UBB error assumption. The choice of the nominal model is usually performed by minimizing a cost function related to the feasible set. The feasible set itself gives the size of the uncertainty associated with the nominal model. In the first works on set membership identification (see the survey [31] and the references therein) the contributions from unmodeled dynamics and noise were not separated (which corresponds to assuming that the plant generating the data belongs to the considered model class). More recently, the presence of model errors has been explicitly accounted for in several works and different settings (see e.g. [30, 10, 9]).

In order to deal with errors that arise both from noise and unmodeled dynamics, this paper contributes with an extension of the error modeling concept to set membership identification. In doing so, a fairly general strategy for separating error contributions is obtained and moreover, a frequency domain model validation tool for set membership identification is provided. The basic idea is to evaluate the minimum value of the error bound, which allows one to validate the estimated nominal model. It is also shown that this provides a useful rationale for the selection of an adequate model class.

As it can be seen from the above discussion, quite different approaches to robust identification are provided by the considered techniques. The main differences concern the type of data required, the selected model class, the assumptions on the noise affecting the data and the nominal model estimation criteria. The next contribution of the paper is to test the above three approaches on the same simulation setting, in order to highlight the advantages of each method and to provide, as long as it is possible, a fair comparison of the results. The most important features of each technique are emphasized. In order to ensure the presence of a nontrivial undermodeling, a nonlinearity is present in the true system considered in the test example, and identification of a low order nominal model is required. Frequency domain uncertainty regions are considered in order to provide adequate models for $\mathcal{H}_\infty$ control design.

Paper outline: Section 2 briefly reviews the technique of non-stationary stochastic embedding, while in Section 3, model error modeling ideas are summarized from a quite general point of view and Section 4 introduces the main concepts of set membership identification, including some recent material on conditional estimators. Section 5 shows how set membership identification can be embedded in a model error modeling setup, and discusses identification of both nominal model and model error, model validation and model class selection. Section 6 compares the three methods from an asymptotic point of view, in particular the construction of the model uncertainty and their relation to the true system is discussed. Section 7 reports a comparative simulation example, which illustrates the main features of non-stationary stochastic embedding and model error modeling, the latter using either prediction error methods or the proposed set membership identification approach. Benefits and advantages of the three techniques are pointed out and the link to standard robust control is discussed as well. Finally, Section 8 gives some concluding remarks and suggests directions of future work.

2 Non-Stationary Stochastic Embedding

The basic idea behind Stochastic Embedding (SE) can be described as follows. The *true* frequency response G, at a certain frequency ω, is given as

$$G(i\omega) = G_0(i\omega) + \Delta G(i\omega) \tag{1}$$

where G_0 is a "nominal system" that can be exactly represented within a parameterized family, and $\Delta G(i\omega)$ is a random variable *independent of* G_0. Notice that this is a turn-around of the conventional interpretation, where the nominal, estimated model

$$G_0(i\omega) = G(i\omega) + \Delta G(i\omega)$$

is seen as the true system plus a model error ΔG, that is *independent of* G. In contrast, the true system in (1) is a random variable as well, which is the root of the term "stochastic embedding". Now, suppose that we have noisy observations of the true system G at certain frequencies ω_k:

$$\hat{\hat{G}}_k = G(i\omega_k) + \nu_k, \quad k = 1, \ldots, m \tag{2}$$

where the noise term ν_k is independent of G and ΔG. Then, combining (1) and (2), one has

$$\hat{\hat{G}}_k = G_0(i\omega_k) + \nu_k + \Delta G(i\omega_k). \tag{3}$$

This is the basic estimation equation, that needs to be complemented with a parameterization of the model G_0 and assumptions about the variance of ν_k and $\Delta G(i\omega_k)$. In [11] it is suggested to use a linear regression parameterization of the nominal model in terms of some orthonormal basis functions, i.e. $G_0(\theta) = \sum_{i=1}^{n} \theta_i B_i$, where $\mathcal{B} := [B_1, \ldots, B_n]$ represents the selected basis and θ is the vector of parameters. Moreover, the model error ΔG is parameterized according to the same basis $\mathcal{B}$ and it is assumed that the relative model error has a variance that increases linearly or quadratically with ω. This means that ΔG can be written as $\mathcal{B}\bar{\theta}\Lambda$ where Λ is a random walk process over ω (this argument being suppressed) and $\overline{\theta}$ comes from prior knowledge. All this leads to

$$G = G_0(\theta) + G_0(\bar{\theta})\Lambda \tag{4}$$
$$= \mathcal{B}\theta + \mathcal{B}\bar{\theta}\Lambda. \tag{5}$$

The identification procedure then consists of three main steps:

(1) Pointwise least squares estimation of the transfer function for certain frequencies: therefore, the input u has to be a sum of sinusoids. This step delivers the values $\hat{\hat{G}}_k$ at certain frequencies ω_k. Additionally, statistical properties of the noise ν are calculated (i.e. an unbiased estimate of its variance), assuming Gaussian noise.
(2) Choice of a set of basis functions $\mathcal{B}$.
(3) Estimation of the parameter θ and the random walk process Λ in (5): $\hat{\hat{G}}_k = \mathcal{B}\theta + \mathcal{B}\bar{\theta}\Lambda + \nu_k$, based on the frequency function point estimates $\hat{\hat{G}}_k$. This is usually performed according to the following procedure:
   (a) An estimate $\hat{\theta}$ of θ is computed, based on the knowledge of $\hat{\hat{G}}_k$. Thus, the nominal model $G_0 = \mathcal{B}\hat{\theta}$ in (4) is the least squares estimate approximation of $\hat{\hat{G}}_k$ in the subspace spanned by basis functions $\mathcal{B}$.
   (b) As model error parameterization $\bar{\theta}$, the estimate $\hat{\theta}$ is chosen (this is a typical choice, when no a priori information on unmodeled dynamics is available). Thus, (5) becomes: $\hat{\hat{G}}_k = \mathcal{B}\hat{\theta}(1 + \Lambda) + \nu_k$.
   (c) The random walk process Λ is chosen so that the variance of the frequency response ΔG increases linearly with frequency, up to a certain roll off. The motivation is that undermodeling (of a pole) can be written as a multiplicative error. An unbiased estimate of the variance of the random walk Λ is now calculated, basically using an unbiased estimate of the variance of the noise - see [11] for closed form expressions and technical details.
   (d) Statistical properties of the model error $\mathcal{B}\bar{\theta}\Lambda$, for any frequency, are computed.

A variant of the random walk for resonant systems (integrated random walk) is described in [3]: here, the error increases quadratically with the frequency, in order to produce tighter error bounds at low frequencies in the presence of resonant poles.

Step 2 contains the crucial choice of the set of (orthonormal) basis functions. In the case of a Laguerre expansion, one can employ the minimization of the *average modeling error*, as suggested in [2, 3]. At any frequency ω_j, the total modeling error is given by $G_e(\omega_j) := \mathcal{B}(\omega_j)\hat{\theta} - G(\omega_j)$. For a closed form expression of $G_e(\omega_j)$ and its covariance $\Sigma_e(\omega_j)$, the reader is referred to [12]. We note that, for different choices of the pole p_i in the Laguerre expansion, the covariance matrix is a function of the Laguerre pole as well: $\Sigma_e(\omega_j) = \Sigma_e(\omega_j, p_i)$. Therefore, we define the average modeling error (over a certain finite frequency grid Ω) for a fixed Laguerre pole as

$$\sum_{\omega_j \in \Omega} \text{trace}(\Sigma_e(\omega_j, p_i)) \tag{6}$$

The "optimal" Laguerre pole p_i^* is the one that minimizes the mean error (6).

3 Model Error Modeling

In this section, the basic ideas of model error modeling (MEM) are described. A thorough treatment of this approach is given in [16], where prediction error methods are used to calculate nominal model as well as model error. The idea described there, however, is more general and does not depend on a certain estimation method, as we will see in the sequel. While non-linear model error models can be used, we will here adopt the general class of linear models proposed in [17].

Let (u, y) be a collection of measured data and assume that a nominal model G_0 of the system that generated the data has been estimated, according to some identification procedure, see the upper plot in Fig. 1. Then, the model error modeling strategy can be summarized as follows.

(1) Compute the residual $\epsilon = y - G_0 u$.
(2) Consider the "error" system, with input u and output ϵ, and identify a model G_e for this system. This is an estimate of the error due to under-modeling, the so-called *model error model*. If this model is not falsified by the data (ϵ, u), standard statistical properties of the estimated model lead to an *uncertainty (confidence) region of the model error model*, which will be addressed in short as *model error* (e.g., given by the 99% confidence region of the estimated model error model G_e, see lower plot in Fig. 1).

(3) From the nominal model and the model error, a nominal model along with uncertainty can be constructed in two ways:
   (a) Adding frequency by frequency the model error to the nominal model. This gives a region where true system is supposed to be found. This region can be used for the control design. It has, however, as such, nothing to do with the nominal model: the nominal model might very well be *outside* this uncertainty region (see Fig. 1, upper plot).
   (b) Create a symmetric uncertainty region in the frequency domain around the nominal model, that includes the region mentioned under (a). That may serve as the uncertainty of the nominal model as well.

(4) Model validation: the nominal model is not falsified if and only if it lies inside its own uncertainty region (as delivered in step 3a) or, equivalently, if and only if zero is an element of the model error. This can be easily checked by looking at the corresponding uncertainty bands in the Bode or Nyquist plots.

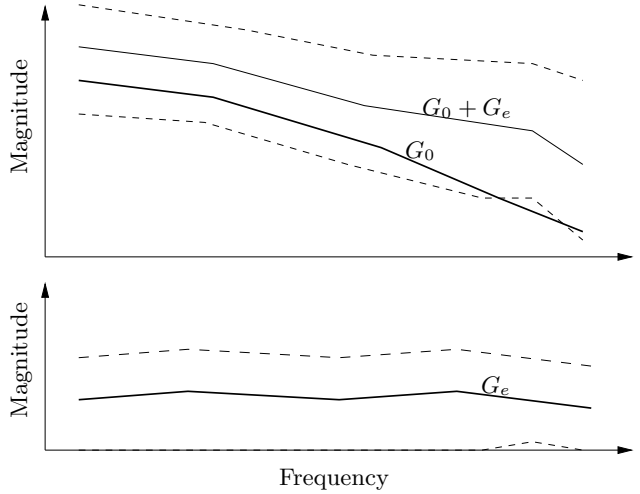


Fig. 1. The basic idea of Model Error Modeling. Upper plot: the nominal model G_0 (fat solid) along with the uncertainty region (dashed) and the center of the uncertainty region, $G_0 + G_e$ (solid). Lower plot: the model error model G_e (fat solid) and its uncertainty band (dashed). This uncertainty contains zero except for an interval at high frequencies (lower plot), where the nominal model is outside the so constructed uncertainty region (upper plot).

Identification of the model error model from residual data can be seen as a separation between noise and unmodeled dynamics. In fact, G_e is an estimate of the dynamic system ΔG, such that

$$\epsilon = \Delta G \cdot u + v, \tag{7}$$

where v is the noise term which is assumed to be uncorrelated with the input u.

A few observations can be made.

- A key question is how to choose the structure of the model error model. It should be sufficiently flexible to reveal (bias) model errors in the nominal model as well as model aspects (e.g. certain frequency regions) that perhaps have not been sufficiently excited, and therefore may give rise to significant uncertainty. It should however not be unnecessarily flexible so that it comes with considerable uncertainty regions.
- A possibility that is discussed in [16] is to start with an a priori chosen (i.e. data independent) flexible structure (like a $10th$ order FIR model). If this error model is not falsified by the data (ε, u), keep it. Otherwise increase the model error model complexity until it is unfalsified by the data. Non-parametric model error models can also be used [20, 24, 27].
- An advantage with the model error model concept is that the nominal model still can be used for control design, even if it is falsified by G_e. This is the case, for example, if the falsification takes place in "unimportant" frequency regions.
- If the nominal model as well as the model error model are linear the estimation of a model error model can be seen as a way to express the result of traditional residual analysis, [16]. Also, in principle this case could be seen as estimating a high order linear model $G_0 + G_e$, and then simplify it by some model reduction technique to the nominal model G_0.
- The comment above shows that a more important use of model error models is to let it be different from the nominal model structure, e.g. non-linear and/or time varying, and perhaps just focus on some general characteristics of it, like its gain. See, e.g. [18] and [19] for such ideas.

4 Set Membership Identification

The term Set Membership (SM) Identification has been used in recent years to indicate a wide variety of robust identification techniques, that are able to handle hard bounds on the identification error. The approach considered in this paper is the one taken in [9]. In the following, we summarize its main features.

Assume that the measured i/o data are generated by a system G, according to

$$y = Gu + v$$

where G belongs to a set $\mathcal{K}$ (the prior information on the system) and the noise v is bounded in some norm $\mathcal{Y}$, i.e. $\|v\|_{\mathcal{Y}} \leq \delta$, for given $\delta > 0$. Then it is possible to define the *feasible system set*

$$FSS = \{S \in \mathcal{K} : \ \|y - S(u)\|_{\mathcal{Y}} \leq \delta\} \tag{8}$$

which is the set of systems that are compatible with the measured data and

the *a priori assumptions*. The set FSS can be very complicated, depending on the structure of $\mathcal{K}$ and the norm $\|\cdot\|_{\mathcal{Y}}$. For example, if $\mathcal{K}$ is the class of LTI systems whose impulse response has an assigned exponential decay and noise is ℓ_∞-bounded, then FSS is an infinite dimensional polytope. Obviously, for real-world systems, including nonlinearities or time-varying drifts, FSS may be much more complex.

Since the set FSS contains all the information provided by data and a priori assumptions, it is natural to evaluate the quality of a nominal model G_0 according to its worst-case error with respect to elements of FSS. In other words, the *identification error* associated with G_0 is given by

$$E(G_0) = \sup_{S \in FSS} \|S - G_0\|_{\mathcal{S}} \tag{9}$$

where $\|\cdot\|_{\mathcal{S}}$ is a suitable norm in the system space.

In order to identify a nominal model, a model class must be chosen. A common requirement is that the nominal model must be simple (low dimensional, linearly parameterized, etc.). Hence, a typical structure for G_0 is

$$G_0(\theta) = \sum_{i=1}^{n} B_i \theta_i \tag{10}$$

where the B_i are user defined basis functions, such as FIR filters, Laguerre or Kautz filters [29], generalized orthonormal basis functions [28], etc. If we denote by $\mathcal{M}$ the set of nominal models parameterized as in (10), the problem of selecting a model in $\mathcal{M}$ according to the criterion (9) is usually addressed as *conditional* set membership identification [9]. The optimal nominal model is given by the conditional central estimate

$$G_0^* = \arg \inf_{G^* \in \mathcal{M}} \sup_{S \in FSS} \|S - G^*\|_{\mathcal{S}}. \tag{11}$$

Since in most practical situation finding an exact solution of the min-max optimization problem (11) is a prohibitive task, suboptimal estimators are considered [6]. In the case that the noise is ℓ_∞-bounded (i.e., $\|\cdot\|_{\mathcal{Y}} = \ell_\infty$) a possible choice for the nominal model is then $G_0(\theta^r)$, where

$$\theta^r = \arg \inf_{\theta \in \mathbb{R}^n} \|y - \sum_{i=1}^{n} B_i \theta_i u\|_\infty. \tag{12}$$

This is known as *restricted projection estimate* and enjoys some nice properties, including the fact that it does not depend on the actual value of the noise bound δ. Note also that it equals the *maximum likelihood estimate* under the assumption that the innovations have a symmetric, uniform distribution, with unknown bound. Its robustness properties in a stochastic setting have been

analyzed as well: in a certain sense, the ∞-norm as used in (12) appears to be less robust then the 2-norm, see [1] for a more detailed discussion. Our major motivation for this choice is the limited computational effort, as θ^r can be easily computed by linear programming. Other more sophisticated estimation algorithms can be chosen without modifying the whole robust identification strategy.

5 Set Membership Estimation in a Model Error Modeling framework

5.1 Estimation of Model Error Model

The key idea in the Model Error Modeling concept is that, whatever nominal model has been identified, the actual size of the unmodeled dynamics can be evaluated from available data by analyzing the residuals as described in Sec. 3, (7). Once the nominal model has been estimated (e.g, as in (12)), the identification of the error system can be performed in the set membership identification framework as well, exploiting the noise bound on v. If the structure of the model error model is chosen as

$$G_e(\bar{\theta}) = \sum_{i=1}^{\bar{n}} \bar{B}_i \bar{\theta}_i \tag{13}$$

the set of all feasible parameters for the model error model is given by

$$FES = \{\bar{\theta} \in \mathbb{R}^{\bar{n}} : \ \|\epsilon - \sum_{i=1}^{\bar{n}} \bar{B}_i \bar{\theta}_i u\|_{\mathcal{Y}} \leq \delta\}. \tag{14}$$

Then, one may select the worst-case optimal model error model by computing the Chebyshev center of FES

$$\bar{\theta}^* = \arg \inf_{\bar{\theta} \in \mathbb{R}^{\bar{n}}} \sup_{\tilde{\theta} \in FES} \|\tilde{\theta} - \bar{\theta}\| \tag{15}$$

where $\|\cdot\|$ denotes the Euclidean norm (of course, other norms can be chosen).

As for (11), problem (15) may be computationally infeasible, and suboptimal solutions are sought. For example, since for ℓ_∞-bounded noise FES is a polytope in $\mathbb{R}^{\bar{n}}$, it can be recursively approximated by simpler regions, like ellipsoids or parallelotopes [5, 4], and the center of these approximating sets may be chosen as an estimate of $\bar{\theta}^*$. More sophisticated set approximation strategies, based on inner and outer bounding via polytopes, have been recently proposed in [22]. They provide a viable trade-off between conservativeness of

the approximation and computational complexity, which is a key issue in SM identification.

If robust identification is oriented to $\mathcal{H}_\infty$ control, an uncertainty band associated to the nominal model frequency response must be delivered to the control designer. In this framework, the frequency plot of the model error model uncertainty region can be obtained by mapping the set FES onto the complex plane for each frequency of interest. This leads to the frequency domain uncertainty set

$$\mathcal{V}(FES) = \bigcup_{\omega} \mathcal{V}_\omega(FES)$$

where

$$\mathcal{V}_\omega(FES) = \{z \in \mathbb{C} : \ z = \mathcal{B}_\omega \bar{\theta}, \ \bar{\theta} \in FES\} \tag{16}$$

and $\mathcal{B}_\omega = [\bar{B}_1(e^{i\omega}) \dots \bar{B}_{\bar{n}}(e^{i\omega})] \in \mathbb{C}^{1\times\bar{n}}$. Once again, the computation of $\mathcal{V}(FES)$ may be a formidable task if the exact FES defined by (14) is considered. Hence, set approximations are useful also in this respect. Examples of the resulting plots for nominal model and model error will be shown in Sec. 7.

The overall set membership identification strategy can be summarized as follows.

(1) Identify the nominal model $G_0(\theta)$ (for example using (12)).
(2) Compute the residual $\epsilon = y - G_0(\theta)u$.
(3) Select a model error model structure, choose the noise bound δ and compute (or approximate) the model error (14).
(4) Identify a nominal model error model $G_e(q;\bar{\theta})$, using optimal or suboptimal estimators based on FES in (14).
(5) Map the nominal model plus the model error onto the frequency domain.

Remark. The use of residual data in set membership identification for evaluating the worst-case $\mathcal{H}_\infty$ norm of the unmodeled dynamics has been introduced in [10], for standard least-squares nominal models. The above set membership model error modeling strategy can be seen as a general framework in which the structures of the nominal model and model error model, and the corresponding identification algorithms must be chosen by the user according to the specific problem (a priori knowledge, noise bound, error norm, etc.).

In the next subsection, it will be shown how it is possible to exploit the above framework to obtain a useful measure of the size of the unmodeled dynamics. This will also provide useful information about the selection of the nominal model class.

5.2 Noise Bound and Model Class Selection

In the Set Membership procedure previously described, a key parameter is the noise bound δ in (14). Indeed, the size of the model error clearly depends on δ.

If δ is known a priori, the standard SM approach presented in Sec. 4 can be applied. The frequency domain uncertainty region, computing according item 3b of the MEM scheme in Sec. 3, turns out to be

$$G_0(e^{i\omega}) + \Delta G(e^{i\omega}) \ , \quad \text{where} \quad |\Delta G(e^{i\omega})| < \sup_{z \in \mathcal{V}_\omega(FES)} |z|,$$

and can be used as such for robust control design.

However, in several practical problems, only a rough estimate of δ is a priori known (denoted by $\hat{\delta}$ in the following). Clearly, if $\hat{\delta}$ is an overestimate of the true bound δ, a conservative FES will be obtained; conversely, a too small $\hat{\delta}$ may lead to an empty FES. Hence, one may try to refine the choice of δ according to the size of the resulting frequency-domain uncertainty region. In other words, δ can be used as a tuning parameter, in order to evaluate the "distance to validation" of the selected nominal model, as suggested in the following.

Let us consider the set $\mathcal{V}_\omega(FES)$ in (16) and define

$$d(\omega) = \min_{z \in \mathcal{V}_\omega(FES)} |z|, \tag{17}$$

which is, at frequency ω, the minimal distance of the uncertainty band to the nominal error. If we set

$$d = \sup_{\omega} d(\omega) \tag{18}$$

then it is clear that the nominal model can be deemed to be unfalsified if $d = 0$. As in standard MEM, the sup in (18) is computed over the frequency range which is important for control design. If ellipsoidal approximations of FES are considered, the sets $\mathcal{V}_\omega(FES)$ are ellipses in the complex plane and the optimization problem (17) can be easily solved. Then, d in (18) can be approximated by taking the maximum over a finite number of frequencies in the range of interest.

It is worth observing that d is a function of δ, since FES (and its frequency image $\mathcal{V}_\omega(FES)$) depend on the noise bound. Then, one can look for the minimum value of δ for which the zero model is included in $\mathcal{V}(FES)$, i.e.

$$\delta^* = \min_{\{\delta:\ d=0\}} \delta. \tag{19}$$

Since d is an increasing function of δ, the computation of δ^* can be easily performed within the desired precision, using a standard bisection on δ. Obviously, δ^* will not be larger than the following value, see (12):

$$\delta^{nom} = \|y - \sum_{i=1}^{n} B_i \theta_i^r u\|_\infty, \tag{20}$$

which is the minimum noise bound for the nominal model when using no model error model.

It is believed that the quantity δ^* is important in several respects. First, it provides a way to select the structure of the model error model. In fact, δ^* depends on the model class selected for identification of the model error model: the "richer" this class, the smaller δ^*. On the other hand, a too simple model class may result in a large value of δ^*, which would contradict the a priori assumption on the noise size. Hence, δ^* can be used to "validate" either the noise bound or the model error model class, according to the available a priori information. If a reliable a priori estimate $\hat{\delta}$ of the noise size is available, the class for model error model identification could be chosen so that δ^* turns out to be as close as possible to $\hat{\delta}$. Conversely, if the model error model class is suggested by a priori information about the system dynamics, δ^* provides a validation tool for the noise bound (for example, if δ^* is much smaller than $\hat{\delta}$, one may conclude that the latter was overestimated).

Another useful property of δ^* is that it gives a rationale for selecting an appropriate model class for nominal model identification. Let us denote by p a vector of parameters that defines a model class for G_0 (for example, the number and/or the pole locations of the family of basis functions B_i). For each p, one can perform set membership identification of nominal model and model error model as suggested in Section 4, and compute $\delta^*(p)$. Then, at least in principle, one can choose the optimal model structure by solving

$$\inf_p \delta^*(p). \tag{21}$$

In practice, the optimization in (21) can be performed over a finite set $\mathcal{P} = \{p_1 \dots p_m\}$ (see the example in Section 7, where p is the pole of a Laguerre expansion of fixed order).

6 Asymptotic Behavior of the Approaches

Before turning to the simulations, it is motivated to pause and consider how the three different approaches "think" when constructing the uncertainty region. We shall pinpoint a few aspects by considering the limiting case that

the number of observed data tends to infinity. We shall adopt the following assumptions and notations

(1) The true system may be quite general, non-linear, time-varying etc.
(2) The input is a sum of sinusoids, as required by the SE approach. Clearly, the asymptotic properties derived in the following do only hold provided that the experiment is informative. In the situation of estimating a model with degree n of numerator and denominator this means that the input has to consist of at least $n+1$ sinusoids to be persistently exciting. For a throughout discussion of this topic see [17, Sec.14.2].
(3) Let $G_\dagger$ denote the second order LTI equivalent to the true system, under this input. A more detailed discussion on this is contained in [19], we just state the basic definition for reading convenience.

Stack the input-output data in a new signal: $\tilde{z} := \begin{bmatrix} y \\ u \end{bmatrix}$, assume that the data are quasistationary and that the spectral function [17, Sec.2], $\Phi_{\tilde{z}}(z) = \begin{bmatrix} \Phi_y(z) & \Phi_{yu}(z) \\ \Phi_{uy}(z) & \Phi_u(z) \end{bmatrix}$, is well defined. Consider a spectral factorization $\Phi_{\tilde{z}}(z) = L(z)L^T(1/z)$, so that $L(z)$ and $L^{-1}(z)$ are stable, causal 2-by-2 transfer function matrices. Then define

$$P(z) = \begin{bmatrix} \Phi_{yu}(z) & \Phi_y(z) \end{bmatrix} L^T(1/z)^{-1} = \sum_{k=-\infty}^{0} p_k z^{-k} + \sum_{k=1}^{\infty} p_k z^{-k} = P_-(z) + P_+(z)$$

where $P_+(z)$ is the strictly causal part of the left hand side. Next define W_u and W_y by $P_+(z)L^{-1}(z) = \begin{bmatrix} W_u(z) & W_y(z) \end{bmatrix}$. By construction W_u and W_y will be strictly causal and

$$\hat{y}(t|t-1) = W_u(q)u(t) + W_y(q)y(t) \tag{22}$$

is the Wiener filter for estimating (predicting) $y(t)$ from $u(s), y(s); s \leq t-1$. Let $e_\dagger(t) = y(t) - \hat{y}(t|t-1)$, then (22) can be rearranged as

$$y(t) = G_\dagger(q)u(t) + H_\dagger(q)e_\dagger(t) \tag{23}$$

where $H_\dagger(z) = (I - W_y(z))^{-1}, G_\dagger(z) = H_\dagger(z)W_u(z)$.

The bottom line of this discussion is that any quasistationary input-output data set can be seen as being produced by (23), with a signal $e_\dagger$ which has a constant spectrum ("white noise") and such that $e_\dagger(t)$ is uncorrelated with past $u(s), s < t$.This means that considering just second order properties (i.e. the spectra) of the signals y and u, we cannot disprove that they have been generated by (23). Note also, that $G_\dagger$ equals the true system, in case it is linear [17, p.264].

(4) Let G_* denote the best linear L_2-approximation (in the norm of the input spectrum) of the true system (and hence of $G_\dagger$) within the set of nominal models, that is:

$$G_* = \arg\min_{\tilde{G}} \int_{-\pi}^{\pi} \left| G_\dagger(e^{i\omega}) - \tilde{G}(e^{i\omega}) \right|^2 \Phi_u(e^{i\omega}) d\omega,$$

where $\{\tilde{G}\}$ belongs to the set of nominal models (according to some parameterization, $G_\dagger$ is the second order equivalent and $\Phi_u(e^{i\omega})$ is the input spectrum (see above).

Remark. SE and SM employ, at least as described above, a model parameterization using basis functions. This, purely technical and not fundamental, fact will be neglected in this section. Whenever speaking of model classes of a given order that contain a certain (linear) model, we will implicitly assume the correct choice of basis functions. Note, however, that this assumption does not change the character of the questions discussed here.

6.1 The SE approach

In the limit as the number of data tends to infinity, the estimates $\hat{\hat{G}}_k$ will equal $G_\dagger$ at the frequency points in question (cf. Fig. 2). This can be interpreted as $\nu_k = 0$. The second step is to fit these points in the complex plane to a linear nominal model of fixed order. This will lead to a nominal estimate G_0 that not necessary equals G_* but in spirit is an approximation of it. The uncertainty region is now constructed based on the distances between G_0 and $G_\dagger$ at the frequency points of excitation (the arrows in Fig. 2). Each distance is viewed as a realization of a random variable with zero mean and variance $\Lambda * \omega$ (or $\Lambda * \omega^2$), from which Λ is estimated. This leads to the uncertainty region around G_0 as described in Sec. 2 and depicted in Fig. 2. Note that, as the point estimates will very likely be inside the uncertainty region, $G_\dagger$ will be inside as well.

Note also that if the order of the nominal model is as large as that of $G_\dagger$, then, in the limit, $G_\dagger$ is delivered as the nominal model, and the uncertainty region shrinks to zero.

6.2 The MEM approach

If an output error model structure is used for the nominal model, this will converge to G_* as the number of data tends to infinity. If the model error model is chosen to be linear and sufficiently flexible, it will converge to a limit

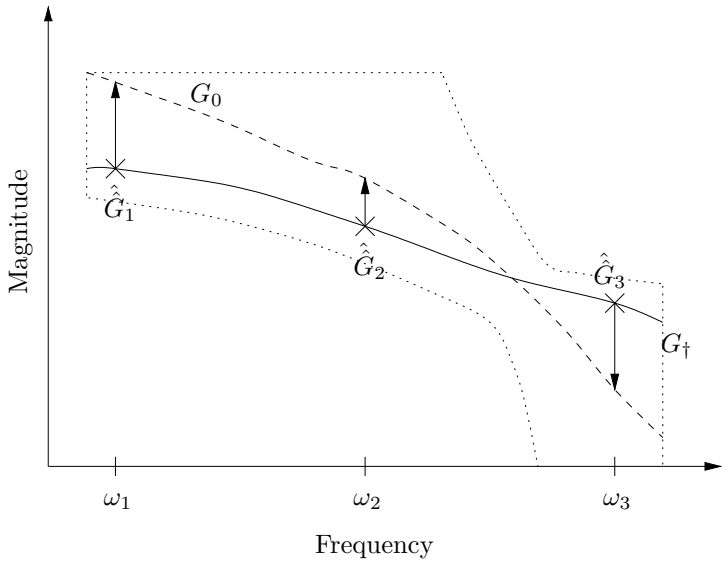


Fig. 2. SE and MEM: Frequency responses of nominal model G_0 (dashed) and second order linear equivalent $G_\dagger$ (solid), along with the frequencies ω_i contained in the input signal and the point estimates $\hat{\hat{G}}_i$ at those frequencies (crosses). A possible (symmetric) uncertainty region around the nominal model is given by the dotted area.

such that the nominal model plus the model error model equals $G_\dagger$. Also, the uncertainty region will shrink to zero, using conventional (linear, second order) techniques. The reason is that the true system cannot be distinguished from $G_\dagger$ using only linear, second order statistics. (This is really the meaning of it being a second order linear equivalent.)

So, following the approach 3a in the model order modeling scheme in Sec. 3, the delivered nominal model is $G_0 = G_*$ and the uncertainty set is just $G_\dagger$, regarded as certain. Using approach 3b, a symmetrical uncertainty set is formed around G_* so that $G_\dagger$ is included. Note that this is very much in the spirit of how the SE uncertainty set is constructed (cf. Fig. 2 and the discussion above).

In case the order of the nominal model is increased up to that of $G_\dagger$, the delivered uncertainty set is this model, considered as certain, just as in the SE case.

6.3 The SM approach

The nominal model $G_0(\theta^r)$, with θ^r given by (12), will be well defined in the limit, if the actual noise/innovations sequence is bounded. It will typically not be equal to G_*, since another norm is used, but it is an approximation in a similar spirit. The construction of the uncertainty set is more difficult to

analyze in this case.

Consider first the case where the true system is linear, with innovations independent of the input. Assume that these innovations have a symmetrical and bounded distribution $|\nu| \leq d^*$ (not necessarily uniform). Assume first also that the nominal model set is so large that $G_\dagger$ is included in this set. Then θ^r, in the limit, will describe this true system. That also means the smallest value of δ that gives a nonempty FSS will be $\delta = d^*$, with $FES = \{0\}$ in (14). This means that in this special case SM, as described in Section 5.2 will deliver the single system $G_\dagger$ as the uncertainty set. This is the same result as for the MEM and SE approaches. Note, however, that for SM, we had to assume that $G_\dagger$ is the true system, not only the second order linear equivalent to achieve this. The reason is that SM employs not only second-order techniques in the model estimation.

Now, what happens when $G_\dagger$ is more complicated than the nominal model? In general, FSS does not shrink to a singleton since SM is not trying to approximate the true system, but the *entire set* FSS in (8). If, however, FSS shrinks to a singleton (the true system!) then G_0^* in (11) is the same as G_* (if the S-norm in (11) is L_2 weighted by the input spectrum). Hence, G_0^* can be seen as a counterpart of G_*, which gives a further perspective in the comparison of the different methods.

7 Simulation Example

7.1 Experimental Setup

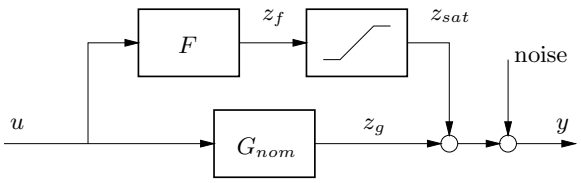


Fig. 3. The experimental setup: A linear system G_{nom} in parallel with a linear filter F and a saturation nonlinearity. The output signal y is corrupted by noise.

Consider the experimental setup as shown in Fig. 3. Its core consists of the following fifth order linear system

$$G_{nom}(s) = \frac{24.15(s+1)}{0.012s^5 + 0.25s^4 + 1.80s^3 + 5.86s^2 + 7.33s + 1}.$$

Its dynamics are perturbed by a nonlinearity in the medium frequency range. This is due to a parallel connection with a linear second order filter $F(s) =$

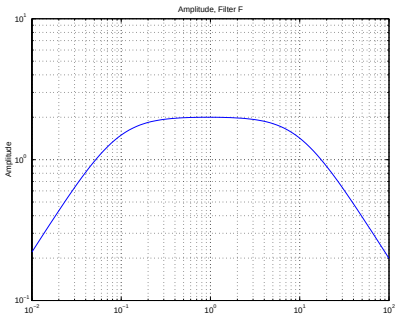


Fig. 4. Amplitude plot of the second order linear filter F.

$\frac{2s}{0.1s^2+s+0.09}$ and a saturation nonlinearity, that saturates at level $l = 1$. According to the amplitude curve of the filter, which is given in Fig. 4, the nonlinearity will be "active" in the frequency range $\omega = [0.1; 10]$, dependent on the amplitude of the input signal. The system will behave linearly, when the output of the filter z_f is smaller than 1 in amplitude; then $y = z_g + z_{sat} = (G_{nom} + F)u$ holds.

In order to be able to apply Stochastic Embedding, the input signal we use is a linear combination of 53 sinusoids (the contained frequencies are marked as crosses in Fig. 6). We use a sample time of $T_s = 0.04s$ and 3000 samples. We will use four different data sets to examine the behavior of the *linear* identification methods to this *nonlinear* plant, which are shown in the following table:

Dataset	excitation of saturation	input level (first half)	input level (sec. half)
1	none (linear behavior)	1	1
2	low	1	2
3	medium	1	6
4	high	10	10

Experiments 1 and 4 excite the system with an input signal as described above with an amplitude of about 1 and 10 respectively. Experiments 2 and 3 switch from a low amplitude input signal to a higher one. We note that the results obtained later on will not change significantly when switching form a higher to a lower amplitude. For experiment 3, these signals are reported in Fig. 5.

First, we perform spectral analysis using the *noiseless* output signal y and the corresponding input signal u. The result is depicted in Fig. 6 (left), together with the Bode plot of the linear system $G_{nom} + F$, which will be "the true plant" for low input level (right). All frequency profiles are quite similar for low frequencies, while they are significantly different at higher frequencies. Clearly, the frequency measurements obtained in this case may change for different

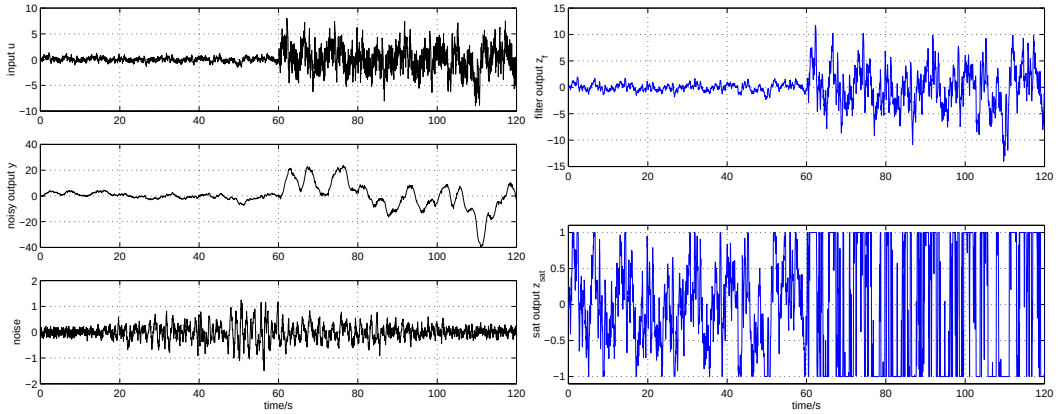


Fig. 5. Selected signals from dataset 3. Left: Test input u, noisy output y and output error noise (top-down). Right: Output of the filter F, z_f, and of the saturation nonlinearity, z_{sat}.

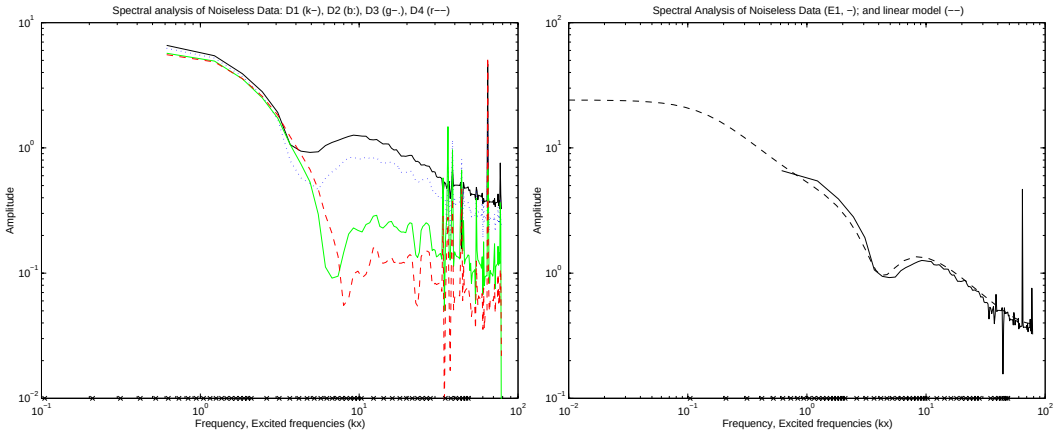


Fig. 6. Left: Spectral analysis (employing a small smoothing window) of the four noiseless datasets: dataset 1 (solid), dataset 2 (dotted), dataset 3 (dash-dotted) and dataset 4 (dashed). The crosses on the frequency axis mark the 53 frequencies, contained in the multi-sinusoid input signal. Right: Bode plot of the linear "mode" $G_{nom} + F$ (dashed) in comparison to the spectral analysis of dataset 1 (solid).

input signals, especially for an input signal with a completely different power spectrum. Nevertheless, we keep the obtained frequency profiles as a reference in the comparison of the different methods, for the specific inputs selected.

Additionally, the output is corrupted by noise, which is also shown in Fig. 5 (left). The noise is a normally distributed one with variance 0.0102, added to a scaled version (to a maximum amplitude of about 1.4) of the $200Hz$ signal registered during the Loma Prieta earthquake in the Santa Cruz Mountains in October 1989 (which has been measured at the Charles F. Richter Seismological Laboratory and made been available by The MathWorks Inc.). This noise signal has been chosen in order to avoid usage of standard stationary stochastic processes or boundary visiting signals as noise models.

In the following, datasets 1-4 will be used for identification purposes. Validation datasets will be employed within those methods that allow an explicit validation procedure (i.e. Model Error Modeling and Set Membership Identification).

The goal of the identification experiments is to obtain a nominal model of order not larger than 4 with a reliable uncertainty description in the frequency domain, which is of suitable size for robust controller design using standard $\mathcal{H}_\infty$ methods.

7.2 Identification via Non-Stationary Stochastic Embedding

The first step within non-stationary stochastic embedding is the estimation of the transfer function at those frequencies contained in the input signal. The second step is then the estimation of the nominal model, based on these transfer function point estimates. By problem definition we are restricted to models of order 4, thus we choose such a Laguerre expansion for our identification. In order to obtain comparable results, we use a confidence level of 99% here as well as for the prediction error methods in the Model Error Modeling approach. We remark, however, that the shape of the uncertainty band does not change paramountly at low and medium frequencies, when decreasing the confidence level. Moreover, preliminary test runs show that the integrated random walk will produce uncertainty bands that are quite tight for lower frequencies and useless large for higher frequencies. This is a typical situation reported in [2] when having a true system without any resonances and applying the random walk, that is "designed" for resonant poles (i.e. using the "wrong" random walk model). In contrast, the random walk (increasing linearly with frequency) produces reasonable uncertainty bands. From these facts we conclude that the system does not contain important resonances, and continue with error-propagation by random walk strategy. We are left with the choice of the pole in the Laguerre basis. Therefore, we employ the minimization of the average modeling error, as suggested in Sec. 2. A typical plot of mean error vs. pole location for fixed order is reported in Fig. 7 (dataset 4).

Choosing the pole for the Laguerre expansion to $p = -0.2895$ (dataset 1,2) and $p = -0.5737$ (dataset 3,4) respectively, least squares estimation of the nominal model and error propagation leads to the results as depicted in Fig. 8. Note that choice of error propagation and confidence level do not influence the nominal model. In all four cases, the uncertainty band includes the frequency profile obtained by spectral analysis quite well.

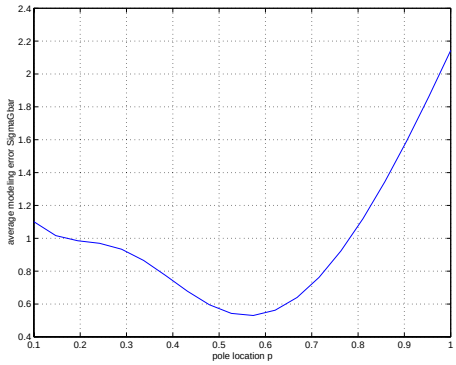


Fig. 7. SE: (absolute value of the) Laguerre pole vs. mean error over all frequencies, in order to choose the "optimal" Laguerre pole, that guarantees minimum error. Result is based on dataset 4.

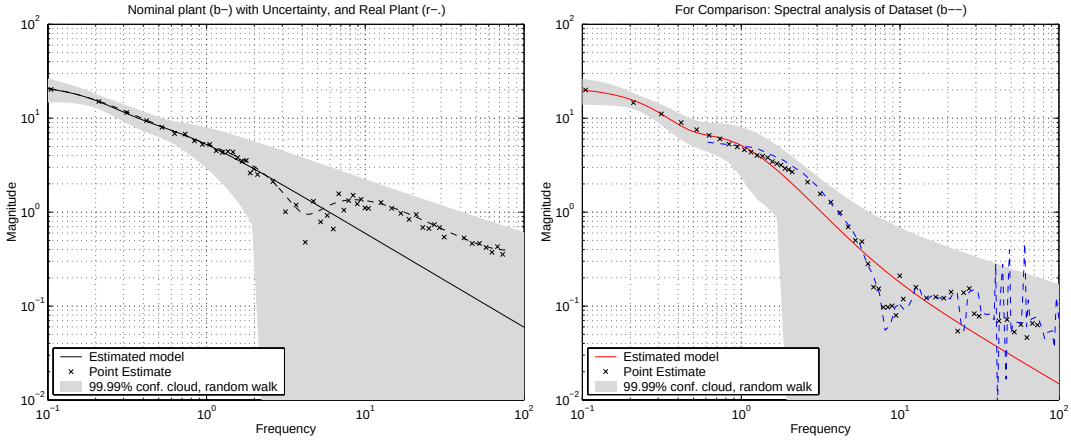


Fig. 8. SE: Identification based on datasets 1 (left) and 4 (right) respectively. Both approaches use a $4th$ order Laguerre expansion with random walk as error propagation and pole located at $p = -0.2895$ and $p = -0.5737$ respectively. The 99% confidence clouds are drawn (shaded), as well as transfer function point estimates (crosses), estimated nominal model (solid) and spectral analysis of noiseless data (dashed).

7.3 Identification via Prediction Error Methods

Prediction Error Methods allow to estimate quite flexible Output Error models. As both other methods used in this comparison use Laguerre models as model classes, we pick this model class (for the estimation of the nominal model) as well in this section. How to choose the pole of the Laguerre expansion? It turns out, that the SE as well as the SM approach come up with a $4th$ order Laguerre basis at $p = -0.2895$, hence, we pick the same Laguerre basis here (note that it would be possible to estimate the system poles via Prediction Error Methods and pass this information to the SE and SM approach). For dataset 1, the least squares estimate of the nominal model, together with the standard residual test, is shown in Fig. 9 (hence the nominal models used

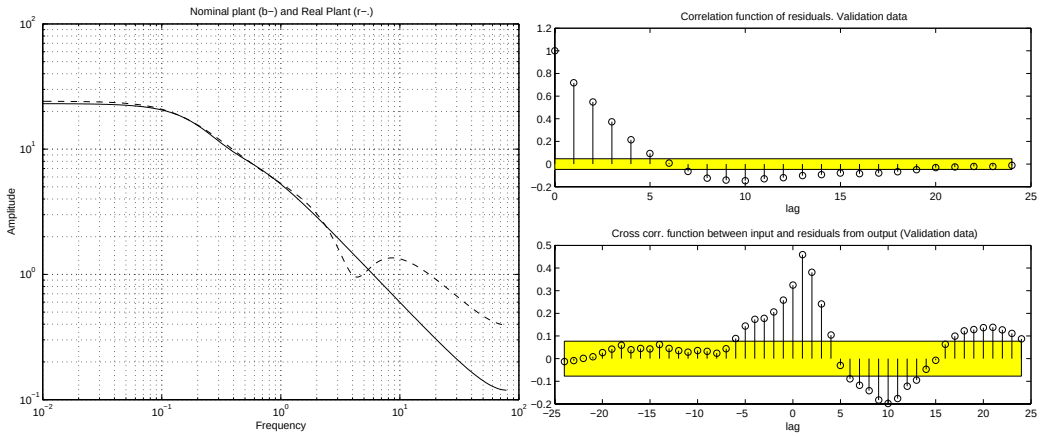


Fig. 9. PEM: Left: $4th$ order Laguerre model, using dataset 1 for least squares identification. Right: Correlation function of the residuals form the nominal model (upper plot) and cross-correlation function (lower plot) using the validation dataset.

here are the same as in the SE approach). We observe, that the nominal model does not pass the classical residual test. A similar statement holds for the remaining three datasets. Hence, we are in need of more detailed information about the accuracy of the identified model. Based on the obtained nominal model, we therefore proceed with estimation of a model error model based on the residual data (calculated from the validation dataset).

We choose a quite flexible model error model of the form $\epsilon(t) = \frac{B(q)}{F(q)}u(t) + \frac{C(q)}{D(q)}v(t)$ with polynomial orders $n_B = n_F = 20, n_C = n_D = 10$ for all datasets in order to pass the residual test (of the model error model). It is worth noting that, in the case of dataset 1, the order of the model error model can be reduced so that the order of nominal and model error model add up to the "real" order, which is quite reasonable as we have "linear data". The result is reported in Fig. 10, showing the non-symmetric version of the model error. The extra information gained is that the nominal model is not that accurate, as it is outside its own uncertainty region. We note that the real plant is very well *inside* this uncertainty region. The symmetric version of the model error model will therefore contain nominal model *and* real plant and is therefore suited for a robust controller design; we will come back to this issue later.

7.4 Identification via Set Membership Estimation

To obtain a $4th$ order nominal model we use a $4th$ order Laguerre expansion. The poles of the Laguerre expansion are chosen according to the strategy suggested in Section 5.2. Hence we will use the hard bound on the noise as additional tuning parameter for the selection of the model class. A typical picture of Laguerre pole location vs. minimum, non-falsifying noise bound is

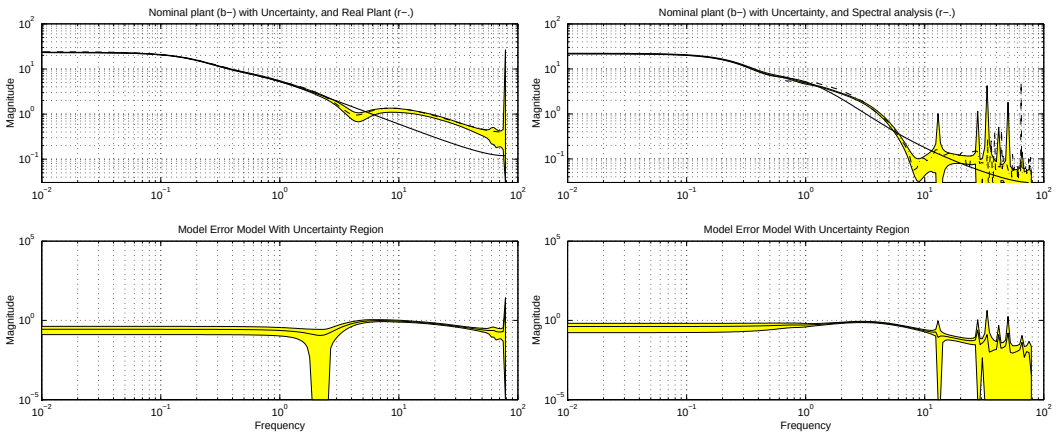


Fig. 10. PEM: Left: dataset 1, right: dataset 4. Upper plots: $4th$ order Laguerre (nominal) model (solid) along with the "non-symmetric" version of the uncertainty region according to step 3a in the MEM scheme (shaded) and spectral analysis of noiseless data (dashed). Lower plots: model error models (solid) with uncertainty regions.

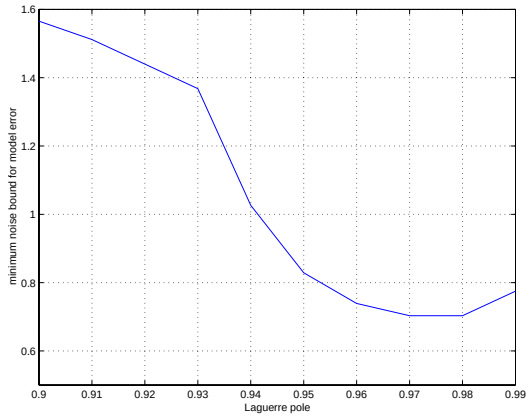


Fig. 11. SM: Location of Laguerre pole versus minimum, non-falsifying noise bound of the model error (dataset 1).

reported in Fig. 11.

Using dataset 1,3 and 2,4, we end up with a Laguerre pole at $p = 0.97$ and $p = 0.98$ respectively. Due to the increasing amount of nonlinearity, the noise bound computed as in (19) increase as $\delta^* = 0.70, 0.83, 1.29, 2.63$ for datasets 1-4 respectively. The estimated noise bounds for datasets 1-3 match pretty well the size of the actual noise, which is about 1, see Fig. 5. Clearly, the higher value for dataset 4 is due to a significant contribution from the nonlinearity.

In order to obtain a frequency domain uncertainty representation, we follow the procedure suggested in Section 5.2. To reduce the computational burden, we approximate the FES in (14) via ellipsoids [5]. Fig. 12 reports the final result for datasets 1 and 4. We observe, that the nominal model matches the linear model quite well in the low and middle frequency ranges. Moreover,

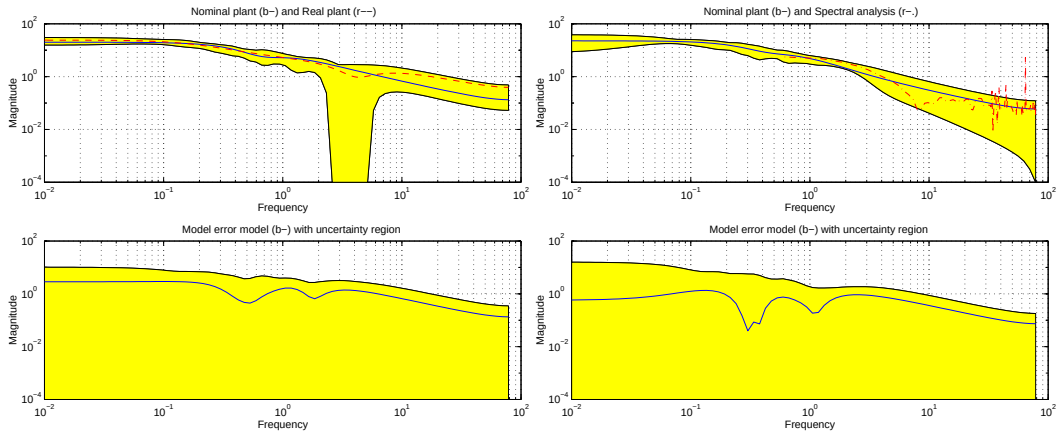


Fig. 12. SM: For datasets 1 (left) and 4 (right): The upper plot shows restricted projection estimate (solid) inside the unfalsified uncertainty region (shaded) compared to linear plant (left plot) and spectral analysis of noiseless data (right plot) respectively. The lower plots show the model error model inside its unfalsified uncertainty region.

the uncertainty band is reasonably small and includes the frequency profiles provided by spectral analysis.

7.5 What is a "Good" Uncertainty Set?

What should we expect from the produced uncertainty sets? They all describe the uncertainty in terms of areas around Nyquist plots. They are thus all part of a linearity paradigm, while the simulated system had an essential non-linearity. Is it at all possible to describe the model error in the linear model by linear techniques? A detailed discussion of this topic is carried out in [19], and the answer is: Yes, for the purpose of robust-stability control design the true system can be redrawn as in Fig. 13. The upper path can be described as the linear filter F, followed by an unstructured nonlinear block $\Delta(\cdot) = -\frac{1}{2}(\cdot) + \mathrm{sat}(\cdot)$, that is bounded in gain by $1/2$. A direct consequence from the small gain theorem is that if we design a controller K such that

$$\|F\frac{K}{1-G_cK}\|_\infty \cdot \frac{1}{2} < 1, \quad G_c := G_{nom} + \frac{1}{2}F \tag{24}$$

holds, then the closed loop will remain stable for all blocks Δ with gain less than $1/2$. Notice that this is the same as stabilizing all systems with frequency functions in the region

$$G_c \pm \frac{1}{2}F. \tag{25}$$

So, in this sense a good uncertainty region is given by expression (25). Note in particular that this region will not decrease if more data are observed. This is a

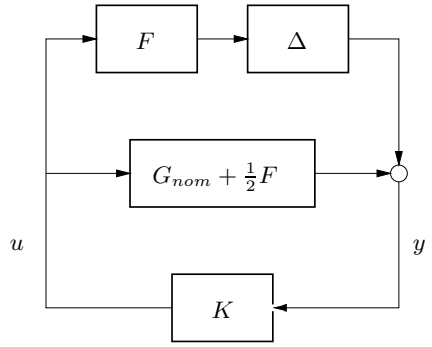


Fig. 13. The experimental setup, redrawn: the saturation nonlinearity is now replaced by a symmetric (in gain) nonlinearity $\Delta(\cdot) = -\frac{1}{2}(\cdot) + \mathrm{sat}(\cdot)$, that is fed with the filtered control signal, given by the controller K.

"true" remaining uncertainty after linear approximation of the true non-linear system.

Continuing the discussion in Sec. 6, can we expect the three approaches to deliver this uncertainty region? Both MEM (with linear model error models) and SE work with second order techniques and only "see" the second order equivalent $G_\dagger$. With more and more data, these two approaches will merely (in different ways) describe the distance between G_* and $G_\dagger$, see discussion in Sec. 6. The SM approach may behave somewhat differently, but as long as no explicit non-linear error model is used in (14) it is difficult to see how a correct description of the model error can be achieved.

This observation points to the need of more sophisticated (i.e. nonlinear) model error models to improve the identified uncertainty sets.

In the case of finite data, however, the situation may be different. Moreover, all *real* plants are nonlinear, and in practice often *linear* methods are used to derived a model for this plant. Hence, a comparison of the achieved uncertainty regions (as carried out in the next section), is interesting.

7.6 Comparison

Following the discussion in the previous subsection, we will compare the uncertainty regions, delivered by robust identification techniques with the magnitude envelope given by G_{nom} and $(G_{nom} + F)$ as lower and upper bound respectively. For ease of notation, we will refer to the latter region as the "true uncertainty region". Figures 14 and 15 reports the result for datasets 2 and 4. The identified uncertainty regions should cover parts of the true uncertainty region, which obviously depends of the excitation of the nonlinearity by the input signal.

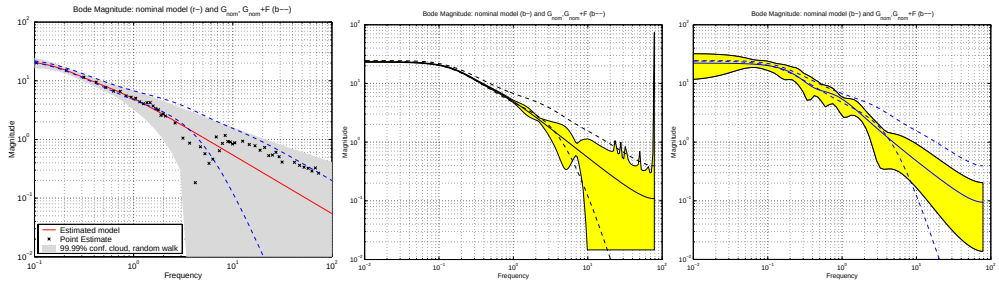


Fig. 14. Dataset 2: Nominal model (solid) and uncertainty region (shaded) in comparison to the amplitudes of $(G_{nom} + \frac{1}{2}F) \pm \frac{1}{2}F$ (dashed). From left to right for SE (continuous time), MEM and SM (discrete time). The uncertainty region for MEM is the symmetric version, as in step 3b of the Model Error Modeling scheme.

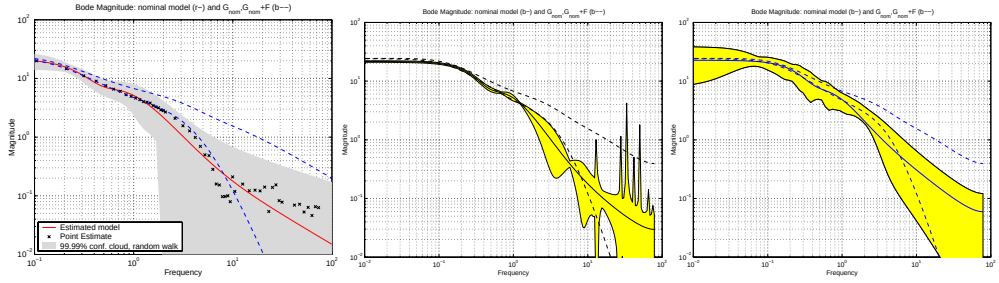


Fig. 15. Dataset 4: Nominal model (solid) and uncertainty region (shaded) in comparison to the amplitudes of $(G_{nom} + \frac{1}{2}F) \pm \frac{1}{2}F$ (dashed). From left to right for SE (continuous time), MEM and SM (discrete time). The uncertainty region for MEM is the symmetric version, as in step 3b of the Model Error Modeling scheme.

We observe, that for both datasets the non-symmetric uncertainty produced by prediction error methods is quite tight and misleading, because the nominal model is *falsified* by the model error, see Fig. 10. Therfore, we now show the *symmetric* uncertainty region according to step 3b of the MEM scheme in Fig. 14

We shall discuss the result obtained for dataset 2 (Fig. 14) and dataset 4 (Fig. 15) in the light of the asymptotic behavior of the SE and MEM approach, cf. Sec. 6. Asymptotically, the point estimates (the crosses in Fig. 14-left) will equal the second order linear equivalent $G_\dagger$ at the frequencies in question. This is, apart from the outlier at $\omega \approx 4$ the case (compared to the spectral analysis in Fig. 6). The nominal model, delivered by SE does not match $G_\dagger$, which is due to the parameterization by Laguerre functions in the first place. The uncertainty set produced by prediction error methods, used in the MEM approach (Fig. 14-middle) does not contain this outlier, as the statement in Sec. 6 is *asymptotic* and, moreover, in probability: at a certain probability level (here: 99%), all point estimates are inside the uncertainty region. We note that only slightly increasing the probability level will "blow up" the uncertainty region, as the logarithmic scale may be somewhat misleading when discussing uncertainty in terms of the frequency response.

A simple method to design a robust controller is the following: we calculate the minimal error bound $\gamma(i\omega)$, so that the uncertainty band shown in the Bode plots is covered by the frequency image $\{G_0(i\omega) + \Delta_a(i\omega), |\Delta_a(i\omega)| \leq \gamma(i\omega)\}$ (this is exactly the step 3b as described in the MEM scheme). Doing so, we create an additive error Δ_a for the nominal model. Application of the Small Gain Theorem yields that all controllers K fulfilling $|\frac{K}{1+KG_0}| < \gamma^{-1}$ at each frequency will guarantee robust stability. Hence, controller design boils down to the task to press $S = \frac{K}{1+KG_0}$ below the inverse additive error. For dataset 2, the inverse errors are shown in Fig. 16. Additionally, we design all controllers such that the nominal open loops have about the same DC gain, similar cross-over frequency and the same roll-off at high frequencies. In the MEM and SM case it is sufficient to design a simple proportional controller to achieve this. In the SE case, the technique we use is open loop shaping, i.e. weighting of the nominal plant and then computation of the $\mathcal{H}_\infty$ optimal controller. The resulting open loops are reported in Fig. 16. We observe, that in the SE case we have a sightly lower cross-over frequency, which will result in a slower system. This seems to be unavoidable because we have to take care of the higher uncertainty in the mid-frequency range. We note that, despite the similar nominal open loop, the actual controllers are quite different, due to the different nominal models, see Fig. 17. The overall look at Fig. 16 now promises robust stability (with respect to the identified uncertainty band) and about the same performance, which is supported by simulate studies of the closed loop with reference signals so that nonlinearity is active; a sample is reported in Fig. 17, showing a slightly slower response in the SE case.

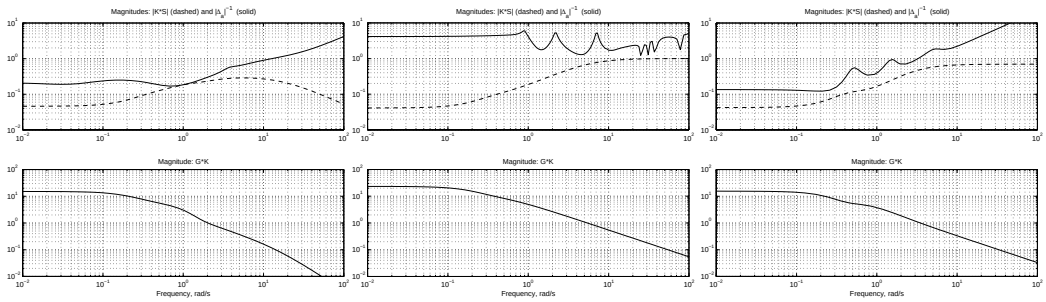


Fig. 16. Dataset 2. Upper plots: Inverse additive uncertainty (solid), as gained from the identification experiment and resulting gain $K \cdot S$ for the designed controller (dashed). Lower plots: Open loop gain with the designed controller: $K \cdot G_0$. From left to right for SE, MEM and SM.

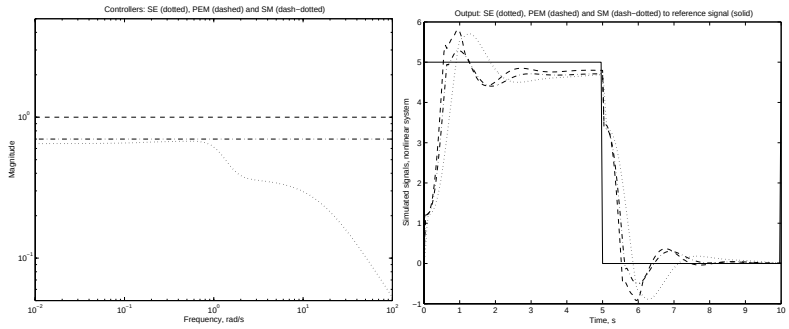


Fig. 17. Left: Magnitude of the respective controllers: dotted: SE, dashed: PEM, dash-dotted: SM. Right: response of the closed nonlinear loop: Solid: reference signal, dotted: SE, dashed: PEM, dash-dotted: SM.

7.8 Properties and benefits of the three methods

Our comparison addresses three different approaches to robust identification: Stochastic Embedding, Model Error Modeling and Set Membership Identification. The first two approaches have been developed in the statistical framework, while the latter relies on unknown-but-bounded error assumptions. Due to the different nature of the three methods, an overall comparison might become unfair at a certain stage. Due to restrictions on harmonic input signal for example, one might not always be able to apply Stochastic Embedding to a given dataset. We therefore concentrate ourselves on highlighting the most important properties and benefits of each method.

Model Error Modeling and Set Membership Identification allow an explicit validation step (on fresh validation data). As no further knowledge than the nominal i/o behavior is needed, these frameworks also allow validation of an already existing model for the process. Even when invalidated, valuable information on the underlying error can be obtained, for instance in which frequency range the model is sufficiently reliable.

In contrast, Stochastic Embedding allows the choice between different random walk models (the variance may increase in different ways with frequency) to perform the error propagation. Although this is no explicit validation step, it is a quite reasonable concept to produce reliable confidence regions.

The Model Error Modeling setup enjoys flexibility in the choice of the structure of the nominal model, to end up with for instance OE or ARX models (although we did not exploit this feature in here). Stochastic Embedding and the presented Set Membership Identification are (the latter at least in this context) tied to a parameterization of orthonormal basis functions. In principle, one could formulate both methods in terms of freely parameterized OE models for instance. This, however, would lead us to a *nonlinear* optimization. For easier computation, the *linear* parameterization in terms of basis functions

was chosen. Clearly, we trade a computational issue for a parameterizational one, which is supported by the fact that both schemes come along with a rationale for choice of the pole in a Laguerre expansion. It is worth to note that both methods (SE and SM) end up with basically the same recommendation (in our case at $s = -0.29, -0.57$, corresponding to $z = 0.98, 0.97$ in discrete time).

We observe, that all methods deliver an estimated nominal model, along with an uncertainty region, which is certainly suited for robust controller design, for instance using a simple open loop shaping approach, based on the estimated uncertainty band in the frequency domain.

8 Conclusions and Future Works

We compared identification methods, working in the time and frequency domain and following statistical or deterministic philosophy. Due to these different natures of the approaches, a fair comparison is quite difficult. Moreover, application of one of the methods will heavily depend on the a priori knowledge (for example, if harmonic input signals are possible or a priori information on the noise amplitude or statistics are available). However, we showed the main features of the methods and obtained reliable nominal models and acceptable related uncertainties in all three cases.

Moreover, the ideas of model error modeling and their application in the context of set membership identification have been analyzed. In particular, it has been shown that the separation of noise and unmodeled dynamics is quite natural in this framework, and that the minimum noise bound for which a nominal model is not falsified by the data can be easily computed and used as a tool for model class selection.

The analysis and illustrations of this paper have some shortcomings:

- The models and methods have essentially been confined to linear techniques, and they cannot, as such, capture the non-linear test system. Rather, as pointed out, these linear models will aim for the second order equivalent (that depends on the input). Still, the comparisons made here are meaningful: the second order equivalent in this case is infinite dimensional. Therefore the resulting, robustly identified models reflect what can be extracted about this complex system from finite, noisy data. In a sense, the resulting uncertainty could very well reflect the actual model error. In fact, as Fig. 14 shows, this is not quite unrealistic.
- One might ask what is the difference between a linear nominal model in conjunction with a linear model error model on the one hand, and a high

order estimated model on the other. Well, the answer is "essentially none", as pointed out in [16]. The model error model allows more complex and also more superficial models. We have confined ourselves to linear error models in order to allow comparisons with the SE approach.

This points to the natural continuation of this work. It will be necessary to pay more attention to non-linear model errors, since this in practice is much more common than linear undermodeling. A step in this direction is taken in [19].

Acknowledgements

The first author was supported by the European Commission through the program Training and Mobility of Researchers - Research Networks and through the project System Identification (FMRX CT98 0206); contacts with the participants in the European Research Network System Identification (ERNSI) are gratefully acknowledged by all authors.

References

[1] H. Akcay, H. Hjalmarsson, and L. Ljung. On the choice of norms in system identification. *IEEE Trans. on Automatic Control*, 41(9):1367–1372, Sept. 1996.

[2] J. H. Braslavsky. Model error quantification via non-stationary stochastic embedding: Hairdryer example. CIDAC, Dept of Electrical Engineering, Univ of Newcastle, Callaghan, Australia. Manual for NSSE-package, Mar. 1999.

[3] J. H. Braslavsky and G. C. Goodwin. A note on non-stationary stochastic embedding for modelling error quantification in the estimation of resonant systems. Technical Report EE99014, CIDAC, Dept of Electrical Engineering, Univ of Newcastle, Callaghan, Australia, Jan. 1999.

[4] L. Chisci, A. Garulli, A. Vicino, and G. Zappa. Block recursive parallelotopic bounding in set membership identification. *Automatica*, 34:15–22, 1998.

[5] E. Fogel and F. Huang. On the value of information in system identification – bounded noise case. *Automatica*, 18(12):229–238, Dec. 1982.

[6] A. Garulli, B. Z. Kacewicz, A. Vicino, and G. Zappa. Error bounds for conditional algorithms in restricted complexity set membership identification. *IEEE Trans. on Automatic Control*, 45(1):160–164, 2000.

[7] A. Garulli and W. Reinelt. Model Error Modeling in Set Membership

Identification. In *Proc. of the System Identification Symposium SYSID*, pages WeMD1–3, Santa Barbara, CA, USA, June 2000.

[8] A. Garulli, A. Tesi, and A. Vicino, editors. *Robustness in Identification and Control*. Number 245 in Lecture Notes in Control and Information Sciences. Springer-Verlag, 1999.

[9] A. Garulli, A. Vicino, and G. Zappa. Conditional central algorithms for worst-case set membership identification and filtering. *IEEE Trans. on Automatic Control*, 45(1):14–23, Jan. 2000.

[10] L. Giarrè, M. Milanese, and M. Taragna. $\mathcal{H}_\infty$ identification and model quality evaluation. *IEEE Trans. on Automatic Control*, 42(2):188–199, 1997.

[11] G. C. Goodwin. Identification and robust control: Bridging the gap. In *Proc. of the 7th IEEE Mediterranian Conference on Control and Automation*, Haifa, Israel, June 1999.

[12] G. C. Goodwin, J. H. Braslavsky, and M. M. Seron. Non-stationary stochastic embedding for transfer function estimation. In *Proc. of the 14th IFAC World Congress*, Bejing, China, July 1999.

[13] R. G. Hakvoort and P. M. J. Van den Hof. Identification of probabilistic system uncertainty regions by explicit evaluation of bias and variance errors. *IEEE Trans. on Automatic Control*, 42(11):1516–1528, Nov. 1997.

[14] R. L. Kosut, G. C. Goodwin, and M. P. Polis. Special issue on system identification for robust control design. *IEEE Trans. on Automatic Control*, 37(7), July 1992.

[15] L. Ljung. Comments on model validation as set membership identification. In A. Garulli, A. Tesi, and A. Vicino, editors, *Robustness in Identification and Control*, pages 7–16. Springer, 1999.

[16] L. Ljung. Model validation and model error modeling. In B. Wittenmark and A. Rantzer, editors, *Proc. of the Åström Symposium on Control*, pages 15–42, Lund, Sweden, Aug. 1999. Studentliteratur, Lund, Sweden.

[17] L. Ljung. *System Identification – Theory For the User*. Prentice Hall, Upper Saddle River, NJ, USA, 2nd edition, 1999.

[18] L. Ljung. Model error modeling and control design. In *Proc. of the IFAC Symposium SYSID*, pages WeAM1–3, Santa Barbara, CA, USA, June 2000.

[19] L. Ljung. Estimating linear time-invariant models of nonlinear time-varying systems. *European Journal of Control*, 2001. Special issue for ECC 2001.

[20] M. Milanese. Learning models from data: the set membership approach. In *Proc. of the American Control Conference*, Philadelphia, PA, USA, June 1998.

[21] M. Milanese, J. P. Norton, H. Piet-Lahanier, and E. Walter, editors. *Bounding Approaches to System Identification*. Plenum Press, New York, NY, USA, 1996.

[22] M. Milanese and M. Taragna. Suboptimality evaluation of approximated model in H_∞ Identification. In *Proc. of the 38th IEEE Conference on*

Decision and Control, pages 1494–1499, Phoenix, AZ, USA, Dec. 1999.

[23] M. Milanese and A. Vicino. Optimal estimation theory for dynamic systems with unknown but bounded uncertainty: an overview. *Automatica*, 27(6):997–1009, June 1991.

[24] K. Poola, P. Khargonekar, A. Tikku, J. Krause, and K. Nagpal. A time domain approach to model validation. *IEEE Trans. on Automatic Control*, 39(5):951–959, May 1994.

[25] W. Reinelt, A. Garulli, L. Ljung, J. H. Braslavsky, and A. Vicino. Model Error Concepts in Identification for Control. In *Proc. of the 38th IEEE Conference on Decision and Control*, pages 1488–1493, Phoenix, AZ, USA, Dec. 1999.

[26] T. Söderström and K. J. Åström. Special issue on trends in system identification. *Automatica*, 31(17), Dec. 1995.

[27] A. Stenman and F. Tjärnström. A nonparametric approach to model error modeling. In *Proc of the IFAC Symposium on System Identification SYSID*, pages WeMD1–1, Santa Barbara, CA, USA, June 2000.

[28] P. M. J. Van den Hof, P. S. C. Heuberger, and J. Bokor. System identification with generalized orthonormal basis functions. *Automatica*, 31(12):1821–1834, 1995.

[29] B. Wahlberg. System identification using Laguerre models. *IEEE Trans. on Automatic Control*, 39:1276–1282, 1991.

[30] B. Wahlberg and L. Ljung. Hard frequency-domain model error bounds from least-squares like identification techniques. *IEEE Trans. on Automatic Control*, 37(7):900–912, 1992.

[31] E. Walter and H. Piet-Lahanier. Estimation of parameter bounds from bounded-error data: a survey. *Mathematics in Computer and Simulation*, 32:449–468, 1990.