A BAYESIAN – DECISION THEORETIC APPROACH TO MODEL ERROR MODELING [1]

R. McVinish [*] **J.H. Braslavsky** [**] **K. Mengersen** [*]

[*] *Queensland University of Technology, Brisbane, QLD 4001, Australia*
[**] *The University of Newcastle, Callaghan, NSW 2038, Australia*

Abstract: This paper takes a Bayesian-decision theoretic approach to transfer function estimation, nominal model estimation, and quantification of the resulting model error. Consistency of the nonparametric estimate of the transfer function is proved together with a rate of convergence. The required quantities can be computed routinely using reversible jump Markov chain Monte Carlo methods. The proposed methodology has connections with set membership identification which has been extensively studied for this problem.

Keywords: Transfer functions, Decision theory, Non-parametric identification, Monte Carlo calculation, loss minimization

1. INTRODUCTION

An important problem in control engineering is the identification of a nominal model for the system from experimental data and accurate quantification of the resulting model error. One approach to this problem is set membership identification which introduces certain assumptions on the system and bounds on the observational noise to form a collection of system descriptions that are not falsified by the data (Milanese and Vicino, 1991; Garulli *et al.*, 2000). Other approaches to the problem are stochastic embedding and model error modeling. These approaches have been compared and reviewed in Reinelt *et al.* (2002).

Maximum likelihood and related prediction error approaches, of which model error modeling is a part, have been the two dominant estimation principles in system identification. These estimation principles are supported by an extensive theory and considerable success in applications. However there has also been increasing interest in the application of the Bayesian paradigm to system identification (Ninness *et al.*, 2002; Juloski *et al.*, 2004; Šmídl *et al.*, 2005). In a similar spirit to the set membership approach, the Bayesian paradigm is able to incorporate prior information about the system with observed data to form a distribution of the parameters called the posterior.

Often interest will be in infinite dimensional quantities like a transfer function or impulse response. Such problems belong to the area of Bayesian nonparametrics which has been growing rapidly in recent years (see Muller and Quintana (2004) for a review). One difficulty with Bayesian nonparametrics is the requirement to specify a prior distribution on an infinite dimensional space. Frequentist asymptotics are used to provide some guidance on how the priors should be constructed so that the resulting estimates are consistent (Ghosh and Ramamoorthi, 2003).

The objective of this paper is to demonstrate how Bayesian nonparametrics can be applied to the problem of identifying a nominal model and quantifying model error. Connections to the set membership approach are also shown.

[1] Research supported by the ARC Centre for Complex Dynamic Systems and Control CEO348165

2. PRELIMINARIES

Consider the following linear time-invariant (LTI) single-input single-output system

$$y(t) = G(z)u(t) + e(t) \tag{1}$$

where $G(z)$ is a LTI causal operator, $u(t)$ is the system input (assumed bounded), $e(t)$ is the observational noise process and $y(t)$ is the observed output. It will be useful to write the operator $G(z)$ in the form

$$G(z)u(t) = \sum_{j=0}^{\infty} g_j u_{t-j} \tag{2}$$

where g is its impulse response sequence, assumed in ℓ_1.

This paper examines two related problems in system identification: Firstly, to determine the operator $G(z)$ and quantify uncertainty about its estimation from experimental input-output data. Secondly, determine a nominal model that is suitable for robust control design.

As the Bayesian approach taken in this paper is similar in spirit to the approach taken by set membership identification (SM), it will now be briefly described. To address these problems SM introduces two assumptions about the system;

(i) G belongs to some defined set of operators $\mathcal{H}$.
(ii) For all times t, $|e(t)| < \delta$ for some known δ.

The Feasible Solution Set is then formed as

$$FSS = \mathcal{L} \cap \mathcal{H}, \tag{3}$$

where

$$\mathcal{L} = \{S : |y(t) - S(z)u(t)| < \delta, \forall t = 1, \ldots, n\}, \tag{4}$$

with S denoting a general LTI causal operator. Thus the first problem is addressed in SM by providing a set of operators which, given assumptions (i) and (ii), contains the true operator.

Now let $F(z;\xi), \xi \in \Xi$ be a parametric family of operators. The second problem is addressed by taking the nominal model parameter ξ according to

$$\xi_0 = \arg\min_{\xi \in \Xi} \sup_{S \in FSS} \|S - F(\cdot;\xi)\|, \tag{5}$$

where $\|\cdot\|$ is some norm appropriate to measuring the distance between operators. $F(z;\xi_0)$ denotes the nominal model.

3. A BAYESIAN APPROACH

In this section it will be assumed that the observational noise process $e(t)$ forms a sequence of independent and identically distributed random variables with a probability density function $p(e \mid \theta)$, θ being finite dimensional parameter. The likelihood for (g, θ) is given by

$$\pi(y \mid g, \theta, u) = \prod_{t=1}^{n} p(y(t) - G(z)u(t) \mid \theta). \tag{6}$$

Bayesian inference is based on the posterior distribution, that is the distribution of the parameters (g, θ) given the data (y, u), denoted by $\pi(\cdot \mid y, u)$. The posterior distribution is formed as the probability measure proportional to the product of the likelihood and prior distribution $\pi(\cdot)$.

3.1 The prior

For the system (1) a prior distribution on the space ℓ_1 needs to be constructed. Ideally the prior would be completely determined by the engineers' expert knowledge of the systems' dynamics. However, it is difficult for any person to adequately quantify prior belief on high-dimensional spaces and on infinite dimensional spaces the task is impossible. Motivated by this difficulty, nonparametric priors are typically constructed so that the resulting posterior distribution possesses some desirable asymptotic properties such as strong consistency. Expert knowledge is incorporated into these priors through a relatively small number of free parameters.

For this paper a "sieve" prior shall be adopted. Sieve priors reduce the difficulty in specifying a distribution on an infinite dimensional space to specifying a series of distribution of finite dimensional spaces of increasing dimension. The prior for g is now constructed.

- Let K be a random variable having support on the positive integers such that
$$\pi(\{g : K = k\}) \propto e^{-ck(\log k)^{\psi}},$$
for some positive finite constant c and $\psi \in (1, 2)$.
- Conditional on K the elements of g are independent random variables such that
$$g_j \sim \begin{cases} N(0, \sigma_j^2), & j \leq K \\ 0, & j > K \end{cases}$$
where $N(\mu, \sigma^2)$ denotes a normal distribution with mean μ and variance σ^2 and $\sum \sigma_j < \infty$. The later condition is imposed so that the mean of the ℓ_1 norm of g is finite.

The form of the prior on K is dictated by the asymptotic concerns studied in the following subsection. It is noted that the rate at which $\sigma_j \to 0$ is important as it represents *a priori* the belief of smoothness in the transfer function. The prior for g is proportional to

$$e^{-cK(\log K)^{\psi}} \prod_{j=1}^{K} \sigma_j^{-1} \exp\left(\frac{-g_j^2}{2\sigma_j^2}\right) \prod_{j=K+1}^{\infty} \delta_{\{g_j = 0\}}.$$

Realizations from this prior are just finite impulse response operators.

It is assumed *a priori* that g and θ are independent. The parameter θ is typically a finite dimensional nuisance parameter such as the variance of a normal distribution and will be given a prior with continuous density.

3.2 Consistency

Let $p_{f,t}$ denote the probability density function of the output variable from a system defined by the impulse response sequence f at time t. Let $h(p,q) = \left(\int(\sqrt{p}-\sqrt{q})^2 dx\right)^{1/2}$ denote the Hellinger distance between to probability density functions p and q. This section examines the rate at which the posterior distribution concentrates around the true system as measure by the semi-metric d_n where

$$d_n^2(f,g) = n^{-1}\sum_{t=1}^{n} h^2\left(p_{f,t}, p_{g,t}\right). \tag{7}$$

Theorem 1. Assume that the true impulse response sequence g satisfies:

(a) $\limsup_{j\to\infty} g_j/\sigma_j < \infty$,
(b) $\sum_{j>k}|g_j| < Ck^{-\beta}$ for some positive finite constants C, β.

Let $\epsilon_n = n^{-\beta/(1+2\beta)}(\log n)$ and take the prior as defined in Section 3.1. As $n\to\infty$ the posterior satisfies

$$\Pi(\{f : d_n(f,g) > R_n\epsilon_n\} \mid y,u) \to 0 \tag{8}$$

in probability, for any $R_n\to\infty$.

PROOF. For simplicity the proof is only given in the case where $e(t)$ is Gaussian with known variance θ. In this case the Hellinger distance $h^2\left(p_{f,t}, p_{g,t}\right)$ is

$$2\left[1-\exp\left(-\left|\sum(f_j-g_j)u_{t-j}\right|^2/(8\theta)\right)\right]. \tag{9}$$

It will be simpler to work with the upper bound on d_n,

$$d_n(f,g) \le C\sum_{j=0}^{\infty}\left|f_j-g_j\right|, \tag{10}$$

for some finite C, assuming the input is bounded.

To determine the rate at which the posterior converges it is sufficient to check that the conditions of Theorem 4 in Ghosal and van der Vaart (2005) hold. Let $\mathcal{F}_n$ be an increasing collection of subsets of sequences in ℓ_1. Sufficient conditions for convergence of the posterior distribution at a rate ϵ_n may be written as

$$\pi(\{g : d_n(f,g)\le\epsilon_n\}) \ge \exp(-n\epsilon_n^2) \tag{11}$$

$$\sup_{\epsilon>\epsilon_n}\log N(\epsilon/36, \mathcal{F}_n, d_n) \le n\epsilon_n^2 \tag{12}$$

$$\pi(\mathcal{F}_n^c) \le \exp(-3n\epsilon_n^2) \tag{13}$$

where $N(\epsilon,\mathcal{A},d)$ is the number of d-balls of radius ϵ required to cover the set $\mathcal{A}$.

To check (11) consider the set

$$\left\{f : K = s_n,\ \left|f_j-g_j\right| < \epsilon_n\sigma_j\right\}.$$

As $\sum\sigma_j<\infty$ then from (10) and (b) it follows that for all f in this set

$$d_n(f,g) \le C\left(\epsilon_n + s_n^{-\beta}\right)$$

for some finite constant C. Under Assumption (a)

$$e^{-cs_n(\log s_n)^\psi}\prod_{j=0}^{s_n}\left[\Phi\left(\frac{g_j+\sigma_j\epsilon_n}{\sigma_j}\right)-\Phi\left(\frac{g_j-\sigma_j\epsilon_n}{\sigma_j}\right)\right]$$
$$\ge \exp\left(-cs_n(\log(\epsilon_n^{-1})+\log(s_n)^\psi)\right)$$

where c is a finite constant and Φ is the Gaussian distribution function. Taking s_n to satisfy

$$C_1 n^{1/(1+2\beta)} < s_n < C_2 n^{1/(1+2\beta)},$$

for some constants C_1, C_2 condition (11) is satisfied.

To check (12) set $\mathcal{F}_n = \left\{f : K\le k_n, \left|f_j\right| < n\sigma_j\right\}$. The number of d_n-balls of size ϵ_n required to cover $\mathcal{F}_n$ is bounded by the number of ℓ_1 balls of size ϵ_n/B required to cover $\cup_{k=1}^{k_n}[-Cn,Cn]^{\otimes k}$, some finite C sufficiently large. It follows that

$$N(\epsilon,\mathcal{F}_n,d_n) \le k_n\left(\frac{Cn}{\epsilon_n}\right)^{k_n}$$

and so,

$$\sup_{\epsilon>\epsilon_n}\log N(\epsilon/36,\mathcal{F}_n,d_n) \le Ck_n\log\left(\frac{n}{\epsilon_n}\right) \tag{14}$$

The complement of $\mathcal{F}_n$ is the set of sequences for which at least k_n+1 elements are non-zero or at lest one of the elements is greater in absolute value than $n\sigma_j$. Thus the prior probability on $\mathcal{F}_n^c$ is bounded above by

$$\pi(\mathcal{F}_n^c) \le \pi(K>k_n) + k_n^2[1-\Phi(n)]$$
$$\le C\exp\left(-ck_n(\log(k_n)^\psi\right) \tag{15}$$

for some finite positive constants C, c.

Upon taking

$$C_1 n^{1/(1+2\beta)}(\log n) \le k_n \le C_2 n^{1/(1+2\beta)}(\log n)$$

equations (14) and (15) are seen to satisfy conditions (12) and (13). This completes the proof.

Remark 2. To demonstrate that Theorem 1 does not require the Gaussian assumption, suppose $e(t)$ has a Laplace distribution with scale parameter θ. The squared Hellinger distance $h^2(p_{f,t}, p_{g,t})$ is

$$2\left[1-\left(1+\frac{\theta\delta_t}{2}\right)e^{-\theta\delta_t/2}\right], \tag{16}$$

where $\delta_t = |\sum_{j=1}^{\infty}(f_j-g_j)u_{t-j}|$. Hence, for bounded input inequality (10) is satisfied. More generally, inequality (10) will hold for observational error with smooth probability density functions under certain conditions and so, Theorem 1 will hold for a variety of observational noise distributions.

Theorem 1 gives the convergence of the posterior distribution in terms of the semi-metric defined in (10). Being a semi-metric there may exist two impulse response sequences $f,\ g$ such that $d_n(f,g) = 0$ however $f \neq g$. It is important to note that this semi-metric is dependent on the input used to generate the data and the distance between to impulse response sequences will be zero in they produce equivalent output from the given input. If attention is restricted to a set of impulse response sequences uniformly bounded in ℓ_1 norm then on this set d_n is equivalent to

$$(f-g)'U^{(n)}(f-g), \tag{17}$$

where $U^{(n)}_{i,j} = (4\theta n)^{-1}\sum_{i=1}^{n} u_{i-j}u_{i-k}$ for Gaussian observational noise.

3.3 Decision theory

Decision theory provides a framework for the determination of estimates which are in a sense optimal. In the decision theory framework the family of nominal models $F(z;\xi),\ \xi \in \Xi$ represents a set of possible actions that can be taken. The conditional Bayes principle (Berger, 1985) states that an action should be chosen which minimizes the Bayesian expected loss

$$\rho(\pi,\xi) = \int l(S,\xi)\pi\left(dS \mid y,u\right), \tag{18}$$

where $l(S,\xi)$ is the loss incurred when the nominal model ξ is chosen and the true operator is S. It is important that the loss function be chosen so to accurately represent the estimates use and penalty for error, as incorrect loss functions can lead to poor performance of the estimate. Weighted L_2 norms on the transfer function space are commonly used to measure distance between operators and have been shown to arise in the context of other estimation methods (see Section 6 of Reinelt *et al.* (2002)). However, other norms may better reflect the ultimate objective of robust control design where relative errors in the transfer function are of greater interest than absolute errors (see for example Hildebrand and Gevers (2003) or Hjalmarsson (2005)).

3.4 Implementation

Bayesian computation typically is based on some form of Markov chain Monte Carlo (MCMC). The general methodology has been reviewed in Robert and Casella (2004). Standard MCMC techniques are unable to deal with priors like that described in Section 3.1 as the number of parameters to be sampled will depend on the value of K. For such priors the reversible jump MCMC algorithm of Green (1995) can be used. At each iteration the algorithm randomly proposes one of the following move types:

- Update g for a given K.
- Move from K to $K+1$ ('birth').
- Move from K to $K-1$ ('death').
- Update θ.

Updating g for a given K is achieved by a simple random walk Metropolis-Hastings proposal. The birth-death moves are based on the zero-th order centered proposals as defined in Brooks *et al.* (2003). Specifically, the 'birth' is made by proposing a g_{K+1} from $N(0,\sigma^2)$ where σ^2 is chosen so that the acceptance probability is 1 when $g_{K+1} = 0$ is proposed.

It is assumed that θ is given a conjugate prior so that it may be sampled using a Gibbs step. For example, suppose that the observational noise process is assumed to have a Laplace distribution with θ as the scale parameter. If it is given a Gamma prior with parameters (α,β) then θ can be sampled from a Gamma distribution with parameters

$$\alpha + N,\ \ \beta + \sum_{t=1}^{N}\left|y(t) - \sum g_j u_{t-j}\right|. \tag{19}$$

The MCMC algorithm generates a sample of operators $\{G_i\}_{i=1}^{M}$ that can be used in the calculation of the posterior expected loss by using sample averages

$$\rho\left(\pi,\xi\right) \approx \frac{1}{M}\sum_{i=1}^{M} l\left(G_i,\xi\right). \tag{20}$$

Convergence of the resulting Markov chain to its stationary distribution can be assisted by using appropriate starting values, that is by starting the chain in an area of significant probability. These starting values could be obtained using least squares estimates of the impulse response sequence given a moderate value of K .

By the law of large numbers for Markov chains this estimate will converge almost surely to the correct value as the number of samples from the posterior goes to infinity. The nominal model which minimizes the posterior expected loss can then be found by numerical optimization methods.

4. CONNECTION TO SM

The Bayesian approach described in the previous section can be shown to have connections to set membership identification. The first connection is between the FSS defined in (3) and a Bayesian credible interval. Again, π is used to denote the a prior on the space of linear time invariant causal operators. For π to be consistent with the Assumption (i) its support should coincide with the set $\mathcal{H}$. Now suppose that the observational noise $e(t),\ t = 1,\ldots,n$ has a density and, to be consistent with Assumption (ii), its support is assumed to be $[-\delta,\delta]\times\ldots\times[-\delta,\delta]$. The set $\mathcal{L}$ is precisely the set where the likelihood function is positive. It

follows that the FSS is the set of linear time invariant causal operators for which both the likelihood and the prior are positive. This is the support of the posterior distribution and so provides a 100% credible interval for G.

The second connection between the Bayesian approach and set membership identification is through the nominal model defined by (5). Some statistical properties of this estimator have been studied in Akçay *et al.* (1996). It can be seen that (5) will not, in general, satisfy the conditional Bayes principle. However, it is possible to show that it can be obtained as the limit of a sequence of decisions satisfying the conditional Bayes principle when the nominal model parameter is unique.

Theorem 3. Assume the prior and likelihood are consistent with SM assumptions (i) and (ii). Also, assume that the SM nominal model given by (5) is unique. Then, there exists a sequence of loss functions such that the conditional Bayes decisions converge to the SM nominal model.

PROOF. We give the proof for the case of Ξ being a finite set. This proof can be extended to the more general setting of Ξ being a compact subset of $\mathbb{R}^d$. Define a sequence of loss functions

$$l_m(S,\xi) = \|S - F(\cdot;\xi)\|^m. \tag{21}$$

Let ξ_0 be the unique nominal model parameter from (5) and set $\epsilon = \sup_{S\in FSS} \|S - F(\cdot;\xi_0)\|$. As the nominal model from (5) is unique then for any $\xi_1 \neq \xi_0$.

$$\sup_{S\in FSS} \|S - F(\cdot;\xi_1)\| > \epsilon. \tag{22}$$

The ratio of expected losses is bounded below by

$$\frac{\rho_m(\pi,\xi_1)}{\rho_m(\pi,\xi_0)} \geq \epsilon^{-m} \int_{\mathcal{A}} l_m(S,\xi_1)\pi(dS \mid y,u), \tag{23}$$

where

$$\mathcal{A} = \{S : \|S - F(\cdot;\xi_1)\| > \epsilon\} \cap FSS. \tag{24}$$

Assume that $\mathcal{H}$ in Assumption (i) is an open set so that FSS is also open. From (22) there exists an operator $S^* \in FSS$ and sufficiently small $\delta > 0$ such that for all $S : \|S^* - S\| < \delta$,

$$\|S - F(\cdot;\xi_1)\| > \epsilon. \tag{25}$$

As the set $\{S : \|S^* - S\| < \delta\}$ is a subset of $\mathcal{A}$ and FSS is the support of the posterior it follows that $\pi(\mathcal{A} \mid y,u) > 0$. Hence, for all $S \in \mathcal{A}$,

$$\lim_{m\to\infty} \left(\frac{\|S - F(\cdot;\xi_1)\|}{\epsilon} \right)^m \to \infty, \tag{26}$$

and so Fatou's lemma can be applied to conclude

$$\lim_{m\to\infty} \frac{\rho_m(\pi,\xi_1)}{\rho_m(\pi,\xi_0)} \to \infty. \tag{27}$$

That is, for all sufficiently large m; ξ_0 has smaller Bayesian expected loss than ξ_1. The nominal model parameter (5) can therefore be obtained through the limit of decisions satisfying the conditional Bayes principle.

5. EXAMPLE

5.1 Transfer function estimation

In this example we wish to demonstrate that having a posterior distribution constructed in the manner described in Section 3 will provide a reasonable quantification of uncertainty when the true system is LTI. For this purpose a sequence of 250 data points is simulated from the system

$$G(z) = \frac{1}{1 - 1.9z^{-1} + 1.43z^{-2} - 0.565z^{-3} + 0.1z^{-4}}.$$

The input is taken to be an iid sequence of random variables from the uniform distribution on $[-1,1]$. The observational error is simulated from a Laplace distribution with scale parameter $\theta = 1$.

The prior is taken as described in Section 3.1 with $\sigma_j = 10 \times j^{-1.5}$, and

$$\pi(\{g : K = k\}) \propto \exp(-0.25k(\log k)^{1.1}) \tag{28}$$

Simulation from the posterior is carried out using the MCMC algorithm detailed in Section 3.4. One advantage of using MCMC is that credible intervals for any quantity of interest can be formed. Figure 1 gives a 99% pointwise credible interval for the transfer function on the Bode magnitude plot.

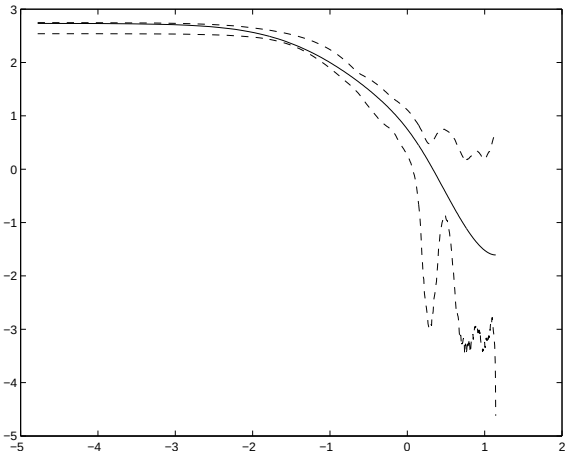

Fig. 1. Bode plot: The true transfer function (solid) and 99% pointwise credible intervals (dashed).

We note that although the intervals appear tight at the low frequencies, the true transfer function is contained within the 99% pointwise credible intervals.

5.2 Nominal model estimation

The nominal model was formed using the following set of bases:

$$F(z;\xi) = \frac{\xi_1}{1-0.7z^{-1}} + \frac{\xi_2}{1-0.4z^{-1}} + \frac{\xi_3}{1-0.2z^{-1}} \quad (29)$$

For simplicity the L_2 norm on transfer functions is used as the loss function in this example. The posterior expected loss is evaluated from the MCMC output and can be minimized by solving the normal equations for ξ. For other forms of loss and nominal models with nonlinear parameters the posterior expected loss can be minimized numerical, for example with the Nelder-Mead simplex algorithm. The estimated nominal model together with the optimal model and true transfer function are given in Figure 2.

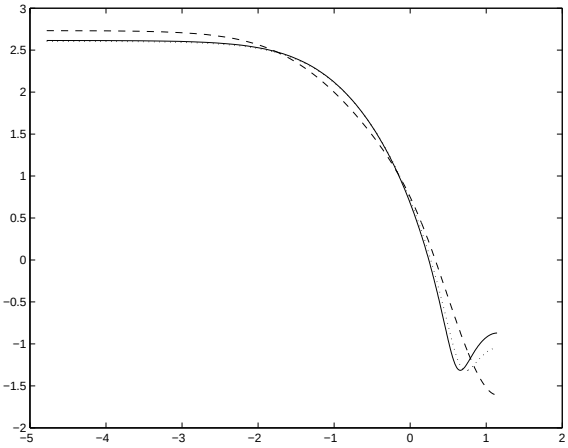

Fig. 2. Bode plot: The true transfer function (dashed), optimal nominal model (dotted), estimated nominal model

Comparing to the 99% pointwise credible interval on the Bode plot the estimated nominal model is seen to break the credible interval at $\log|\omega| \approx 0.5$, but only just.

6. CONCLUSION

This paper proposes a Bayesian-decision theoretic methodology for the quantification of uncertainty in the estimation of a transfer function and choice of nominal model. Its connection to set membership identification has been demonstrated. In the example it was demonstrated that, at least for LTI systems, the method gives a reasonable quantification of the uncertainty. There are still a number of issues to be investigated for the approach. In particular it is important to understand the behaviour of the methodology for when the true system is possibly non-linear or time-varying. Other areas requiring further investigation include sensitivity analysis, diagnostics for observational noise structure and the use of loss functions which may be better suited to control.

REFERENCES

Akçay, H., H. Hjalmarsson and L. Ljung (1996). On the choice of norms in system identification. *IEEE Trans. on Automatic Control* **41**, 1367–1372.

Berger, J.O. (1985). *Statistical Decision Theory and Bayesian Analysis, 2nd Ed.*. Springer-Verlag. New York.

Brooks, S.P., P. Guidici and G.O. Roberts (2003). Efficient construction of reversible jump Markov chain Monte Carlo proposal distributions. *J. R. Statist. Soc. B.* **65**, 3–55.

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

Ghosal, S. and A.W. van der Vaart (2005). Convergence rates of posterior distributions for non iid observations. *Ann. Statist. (accepted)*.

Ghosh, J.K. and R.V. Ramamoorthi (2003). *Bayesian nonparametrics*. Springer. New York.

Green, P. (1995). Reversible jump MCMC computation and Bayesian model determination. *Biometrika* **82**, 711–732.

Hildebrand, R. and M. Gevers (2003). Identification for control: optimal input design with respect to a worst-case ν-gap cost function. *SIAM J. Control Optim.* **41**, 1586–1608.

Hjalmarsson, H. (2005). From experiment design to closed-loop control. *Automatica* **41**, 393–438.

Juloski, A.L., S. Weiland and W.P Heemels (2004). A Bayesian approach to identification of hybrid systems. In: *Proceedings of the 43rd IEEE Conference on Decision and Control*. Atlantis, Paradise Island, Bahamas. pp. 13–19.

Milanese, M. and A. Vicino (1991). Optimal estimation theory for dynamic systems with set membership uncertainty: An overview. *Automatica* **27**, 997–1009.

Muller, P. and F.A. Quintana (2004). Nonparametric Bayesian data analysis. *Statistical Science* **19**, 95–110.

Ninness, B., S. Henriksen and T. Brinsmead (2002). System identification via a computational Bayesian approach. In: *Proceedings of the 41st IEEE Conference on Decision and Control*. Las Vegas, Nevada, USA. pp. 1820–1825.

Reinelt, W., A. Garulli and L. Ljung (2002). Comparing different approaches to model error modeling in robust identification. *Automatica* **38**, 787–803.

Robert, C.P. and G. Casella (2004). *Monte Carlo statistical methods*. Springer. New York.

Šmídl, V., A. Quinn, M. Kárny and T.V. Guy (2005). Robust estimation of autoregressive processes using a mixture-based filter-bank. *Systems & Control Letters* **54**, 315–323.