MODEL ASSESSMENT WITH KOLMOGOROV–SMIRNOV STATISTICS

Petar M. Djurić [1] *and Joaquín Míguez* [2]

[1] Department of Electrical and Computer Engineering
Stony Brook University, Stony Brook, NY 11794, USA

[2] Departamento de Teoría de la Señal y Comunicaciones
Universidad Carlos III de Madrid
Avda. de la Universidad 30, Leganés, 28911 Madrid, Spain

e-mail: djuric@ece.sunysb.edu, joaquin.miguez@uc3m.es

ABSTRACT

One of the most basic problems in science and engineering is the assessment of a considered model. The model should describe a set of observed data and the objective is to find ways of deciding if the model should be rejected. It seems that this is an ill-conditioned problem because we have to test the model against all the possible alternative models. In this paper we use the Kolmogorov–Smirnov statistic to develop a test that shows if the model should be kept or it should be rejected. We explain how this testing can be implemented in the context of particle filtering. We demonstrate the performance of the proposed method by computer simulations.

Index Terms— Model assessment, particle filtering, Kolmogorov-Smirnov statistics

1. INTRODUCTION

The power of science has been recognized by the ability of the scientific method to predict the future accurately and in a consistent way. Often the accuracy is quantified by the discrepancy between future observations (once observed) and sets of predicted observations. In a general setting, a model $\mathcal{M}$ is used to predict future observations and one way of producing them is by employing the predictive distribution of the data conditioned on the model. We write the predictive distribution of the set of observations $y_{1:T} \equiv \{y_1,\ y_2 \cdots, y_T\}$ conditioned on $\mathcal{M}$ as $p(y_{1:T} \,|\, \mathcal{M})$, where

$$p(y_{1:T} \,|\, \mathcal{M}) \quad = \quad p(y_1 \,|\, \mathcal{M}) \prod_{t=1}^{T} p(y_{t+1} \,|\, y_{1:t}, \mathcal{M}) \qquad (1)$$

with the factors in (1), $p(y_{t+1} \,|\, y_{1:t}, \mathcal{M})$, being predictive distributions themselves. At time instant t, y_{t+1} is a future observation modeled by $\mathcal{M}$, and $y_{1:t} \equiv \{y_1, y_2, ..., y_t\}$ is the set of known observations.

The observations are our physical reality and are often the only ingredient that we have when we deal with the uncertainty of considered model(s). When we have more than one competing model for the observed data, we usually want to find the best of these models. This is known in the literature as the model selection problem [1]. From a Bayesian perspective, the best model is typically the model that has the maximum a posteriori probability, $p(\mathcal{M}_k \,|\, y_{1:T})$, where $\mathcal{M}_k$ signifies the $k-$th considered model and where $y_{1:T}$ is the set of data used in computing the posterior probability of $\mathcal{M}_k$. One can show that by using this criterion, one balances the goodness of fit and the complexity of the model. The implementation of the model selection is a well studied problem, and the literature on the subject is quite large. We point out that in this paper we are interested in the class of dynamic models which are nonlinear and which may contain non-Gaussian noises. Then the model selection may not be a trivial task. However, since for nonlinear dynamic models particle filtering is often the method of choice, it is useful to have approaches for model selection within the context of particle filtering. It can be shown that model selection in that case can be accomplished by following a well established theory (for example, see [2]).

In this paper, by contrast, we deal with a scenario where we have only one model, and we want to make a decision whether to keep the model or reject it. Clearly, any meaningful analysis of data requires the possibility of excluding the used model if it fails to provide satisfactory description of the data [3]. The problem of evaluating a single model is not an easy one because it seems that it is ill posed in the sense that we have to test a model $\mathcal{M}$ against unstated alternatives. If there is a true model denoted by $\mathcal{M}_0$, we have to test the hypothesis

$$\mathcal{H}: \quad \mathcal{M} \quad = \quad \mathcal{M}_0. \qquad (2)$$

In [1] this formulation of the problem is considered "rather too general to develop further in any detail." The difficulty of this "ill defined problem of model rejection" is alleviated by specifying a large set of alternative models parameterized by some conveniently chosen set of parameters where the model $\mathcal{M}_0$ is some form of parametric restriction of a more general class of models denoted by $\mathcal{M}_1$. The problem then becomes one of model selection.

Here we propose a method that is truly a method for model assessment that does not require defining alternative models. We anchor the procedure around the made observations and the model-based predicted observations. As in model selection the key role in the assessment is played by the predictive distribution of the data conditioned on the assessed model. Under certain mild assumptions,

This work has been supported by the National Science Foundation under Award CCF-0515246 and the Office of Naval Research under Award N00014-06-1-0012. The work has been carried out while the first author held the Chair of Excellence of Universidad Carlos III de Madrid-Banco de Santander.

we invoke the Kolmogorov-Smirnov statistics and use them to develop a test that can readily be used for assessing the considered model of the data. As pointed out, we are interested in dynamic nonlinear models and so we use particle filtering to generate the posterior distributions of the unknowns of the model. It turns out that the main object of interest, the predictive distribution of the model can readily be approximated by particle filtering and used for generation of future observations. The generated samples are then compared to the made observations by using the two-sample Kolmogorov-Smirnov statistics, which are finally compared to a threshold for making a decision.

The paper is organized as follows. First, in Section 2 we formulate the problem in mathematical terms. Then in Section 3, we present the proposed method, where we show how we generate samples from the predictive distribution and how we develop the test. In Section 4, we demonstrate the method with computer simulations. We conclude the paper with Section 5 by making some final remarks.

2. PROBLEM FORMULATION

Consider a state space model $\mathcal{M}$ defined by

$$\begin{aligned} \boldsymbol{x}_t &= f(\boldsymbol{x}_{t-1}, \boldsymbol{u}_t) && (3)\\ y_t &= g(\boldsymbol{x}_t, \boldsymbol{v}_t) && (4) \end{aligned}$$

where t is a discrete time index with $t \in \mathbb{N}$, $\boldsymbol{x}_t \in \mathbb{R}^{n_x}$ is the unknown state of the system, $y_t \in \mathbb{R}$ is an observation , $\boldsymbol{u}_t \in \mathbb{R}^{n_u}$ and $\boldsymbol{v}_t \in \mathbb{R}^{n_v}$ are state and observation noises, respectively, and $f(\cdot)$ and $g(\cdot)$ are known functions. The model $\mathcal{M}$, in general, is basically a set of assumptions that include

1. the definition of the state equation (3), i.e., its mathematical form,
2. the initial distribution of the state, $p(\boldsymbol{x}_0)$,
3. the assumption about the distribution of the state noise (this is usually a parametric distribution with known or unknown parameters $\boldsymbol{\psi}$ – if the parameters are unknown, then a prior $p(\boldsymbol{\psi})$ of $\boldsymbol{\psi}$ is typically adopted),
4. the definition of the observation equation (4), i.e., its mathematical form,
5. the assumption about the distribution of the observation noise (this, too, is often a parametric distribution with known or unknown parameters $\boldsymbol{\xi}$ - again, when they are unknown, one assumes a prior $p(\boldsymbol{\xi})$), and
6. independence of the noises in the state and observation equations.

More succinctly, we say that the model is defined by $f(\cdot), g(\cdot)$, $p(\boldsymbol{x}_0)$, $p(\boldsymbol{u}_t|\boldsymbol{\psi})$, $p(\boldsymbol{\psi})$, $p(\boldsymbol{v}_t|\boldsymbol{\xi})$, and $p(\boldsymbol{\xi})$.

The main question we address in this paper is the assessment of the model. How good is it? How do we test if the model is good without the need for an alternative model [1]?

3. PROPOSED METHOD

In search for an answer to the model assessment problem, we use the predictive distribution of the data, $p(y_{t+1}|y_{1:t}, \mathcal{M})$, because it encapsulates the power of the model. The main idea in assessing the model $\mathcal{M}$ is in comparing the distribution of the samples $y_{t+1}^{(k)}$ that are generated from $p(y_{t+1}|y_{1:t}, \mathcal{M})$ with the actually observed samples y_{t+1}. Before we get to explaining how we do this comparison, we show how to generate the samples $y_{t+1}^{(k)}$ from $p(y_{t+1}|y_{1:t}, \mathcal{M})$. Our intention is to apply the test to very challenging models, including dynamic nonlinear models. Therefore, it seems reasonable that we adopt a methodology that can successfully deal with such models. We chose to work with particle filtering, so we explain first how to approximate the predictive distributions with particle filtering and then how to generate samples from them.

3.1. Particle filtering

With particle filtering we process the data $y_{1:T}$ sequentially with the objective of obtaining the posterior distributions of the unknowns in the system (3)–(4) [4, 5]. Particle filtering is based on streams of particles which are recursively generated and which are appropriately weighted. Without loss of generality, let the unknowns of the model at time instant t be given by the discrete random measure $\chi_t = \{(\boldsymbol{x}_t^{(m)}, \boldsymbol{\psi}^{(m)}, \boldsymbol{\xi}^{(m)}), w_t^{(m)}\}_{m=1}^{M}$, where M is the total number of particles, and $(\boldsymbol{x}_t^{(m)}, \boldsymbol{\psi}^{(m)}, \boldsymbol{\xi}^{(m)})$ are the unknowns of the system at time instant t.

Our main interest is the generation of samples $y_{t+1}^{(k)}$ from the predictive distribution $p(y_{t+1} \,|\, y_{1:t}, \mathcal{M})$. In general, we can write,

$$\begin{aligned} p(y_{t+1} \,|\, y_{1:t}, \mathcal{M}) &= \int p(y_{t+1} \,|\, \boldsymbol{x}_{0:t+1}, \boldsymbol{\psi}, \boldsymbol{\xi}, y_{1:t}, \mathcal{M}) \\ &\times \; p(\boldsymbol{x}_{0:t+1}, \boldsymbol{\psi}, \boldsymbol{\xi} \,|\, y_{1:t}, \mathcal{M}) d\boldsymbol{x}_{0:t+1} d\boldsymbol{\psi} d\boldsymbol{\xi}. \end{aligned} \tag{5}$$

Clearly, in a general scenario, it will be very difficult to generate samples from $p(y_{t+1} \,|\, y_{1:t}, \mathcal{M})$. However, when we apply particle filtering, this may be carried out readily.

To simplify the notation and the exposition, we assume that the parameters $\boldsymbol{\psi}$ and $\boldsymbol{\xi}$ are known. Then (5) simplifies to

$$\begin{aligned} p(y_{t+1} \,|\, y_{1:t}, \mathcal{M}) &= \int p(y_{t+1} \,|\, \boldsymbol{x}_{t+1}, \mathcal{M}) \\ &\times \; p(\boldsymbol{x}_{t+1} \,|\, \boldsymbol{x}_t, \mathcal{M}) p(\boldsymbol{x}_{0:t} \,|\, , y_{1:t}, \mathcal{M}) d\boldsymbol{x}_{0:t+1}. \end{aligned} \tag{6}$$

It is clear that the last expression can also be written as

$$\begin{aligned} p(y_{t+1} \,|\, y_{1:t}, \mathcal{M}) &= \int p(y_{t+1} \,|\, \boldsymbol{x}_{t+1}, \mathcal{M}) \\ &\times \; p(\boldsymbol{x}_{t+1} \,|\, \boldsymbol{y}_{1:t}, \mathcal{M}) d\boldsymbol{x}_{t+1}. \end{aligned} \tag{7}$$

Now, if we approximate the distribution $p(\boldsymbol{x}_{t+1} \,|\, y_{1:t}, \mathcal{M})$ by

$$p(\boldsymbol{x}_{t+1} \,|\, y_{1:t}, \mathcal{M}) \simeq \sum_{m=1}^{M} w_t^{(m)} p(\boldsymbol{x}_{t+1} | \boldsymbol{x}_t^{(m)}, \mathcal{M}) \tag{8}$$

we see that we can approximately generate $y_{t+1}^{(k)}$ from $p(y_{t+1}|y_{1:t}, \mathcal{M})$ by first sampling $\boldsymbol{x}_{t+1}^{(j)}$, $j = 1, 2, \cdots, J$ from

$$\boldsymbol{x}_{t+1}^{(j)} \sim \sum_{m=1}^{M} w_t^{(m)} p(\boldsymbol{x}_{t+1} \,|\, \boldsymbol{x}_t^{(m)}, \mathcal{M}) \tag{9}$$

and then drawing $y_{t+1}^{(k)}$, $k = 1, 2, \cdots, K$ according to

$$y_{t+1}^{(k)} \sim \frac{1}{J} \sum_{j=1}^{J} p(y_{t+1} \,|\, \boldsymbol{x}_{t+1}^{(j)}, \mathcal{M}). \tag{10}$$

So, the generation of y_{t+1} is a two-step procedure. First, samples of $\boldsymbol{x}_{t+1}^{(j)}$ are drawn by using (9) and then samples of $y_{t+1}^{(k)}$ are generated according to (10).

3.2. Kolmogorov-Smirnov test

From the previous subsection we see how we can obtain samples from the predictive distribution $p(y_{t+1} \mid y_{1:t}, \mathcal{M})$. We want now to compare these samples with the actual observation y_{t+1}. To that end we propose to use the two-sample Kolmogorov-Smirnov test [6, 7].

The Kolmogorov-Smirnov test belongs to the category of non-parametric tests. If we have two sets of samples $X_1, X_2, \cdots, X_L$ and $Y_1, Y_2, \cdots, Y_K$, which respectively are independent and identically distributed according to the continuous distributions $p_x(x)$ and $p_y(y)$, then we could test for the equality of the generating distributions. Namely, under the hypothesis $\mathcal{H}_0$ we have that $F(x) = G(x)$ where $F(\cdot)$ and $G(\cdot)$ are the cumulative distribution functions of the random variables X and Y. From the available samples, we can construct the empirical distributions $\widehat{F}(\cdot)$ and $\widehat{G}(\cdot)$ and test for their agreement.

We define

$$D_{L,K} \quad = \quad \sup_x |\widehat{F}_L(x) - \widehat{G}_K(x)| \tag{11}$$

and in the sequel we refer to it as the Kolmogorov-Smirnov (KS) statistic. Then, the test based on this statistic rejects the hypothesis $\mathcal{H}_0$ at level α if

$$D_{L,K} \quad \geq \quad \gamma(\alpha, L, K) \tag{12}$$

where $\gamma(\cdot)$ is a threshold and

$$P\left(D_{L,K} \geq \gamma(\alpha, L, K)\right) \quad = \quad \alpha(L, K). \tag{13}$$

There are tables from which one can obtain the relevant thresholds (for small L and K they can be found in [8] and for large L and K one can obtain them by using limiting results from [9]).

In our problem we have an extreme situation where there is only one sample from the actual predictive distribution ($L = 1$) and as many samples as one may wish to generate from the predictive distribution of our model. First, we make the following claim:

Claim 1: The KS statistic $D_{1,K}$ satisfies

$$0.5 \leq D_{1,K} \leq 1. \tag{14}$$

□

We proceed with another claim. Without loss of generality let K be an odd positive number. Then

Claim 2: If K is an odd positive number, and if y_{t+1} comes from the same distribution as that of the samples $y_{t+1}^{(k)}$, then

$$P\left(D_{1,K} = \frac{n}{K}\right) \quad = \quad \frac{2}{K+1} \tag{15}$$

where $n = (K+1)/2, (K+3)/2, \cdots, K$.

□

Hence, we have that the KS under $\mathcal{H}_0$ is a uniformly distributed random variable on the support $\{(K+1)/2K, \cdots, (K-1)/K, 1\}$. We need two more results. They are given by the next claim.

Claim 3: The expected value and the variance of $D_{1,K}$ are given by

$$E\left(D_{1,K}\right) \quad = \quad \frac{3K+1}{4K} \tag{16}$$

$$E\left(\left(D_{1,K} - E\left(D_{1,K}\right)\right)^2\right) \quad = \quad \frac{K^3 + 3K^2 - K - 3}{48K^2(K+1)}. \tag{17}$$

□

These claims are easy to prove. It is obvious that

$$\lim_{K\to\infty} E\left(D_{1,K}\right) \quad = \quad 0.75 \tag{18}$$

$$\lim_{K\to\infty} E\left(\left(D_{1,K} - E\left(D_{1,K}\right)\right)^2\right) \quad = \quad \frac{1}{48}. \tag{19}$$

It is important to note that these results hold for *any* continuous distributions. It is also important to note that the mean and the variance in (18) and (19) are those of a uniform random variable defined on $[0.5, 1]$.

Now, as we keep processing the data and obtaining the KS statistics, we can compute the average value of the latter, that is,

$$\overline{D}_{1,K,t} \quad = \quad \frac{1}{t}\sum_{j=1}^{t} D_{1,K,j} \tag{20}$$

where we added one more subscript to the KS statistics to distinguish them as they are obtained at different time instants. If the model $\mathcal{M}$ is the correct model, and if the particle filtering correctly approximates the predictive distributions $p(y_{t+1} \mid y_{1:t}, \mathcal{M})$, then the KS statistics are all identically distributed. They are also independent, and so by central limit theorem, $\overline{D}_{1,K,t}$ is approximately normally distributed with mean 0.75 and variance $1/(48t)$. This result can then be used to develop one of the many existing tests for deciding whether to reject the model $\mathcal{M}$.

3.3. Summary of the method

In summary, the proposed method is applied as follows: for each time instant t.

1. generate samples of $y_{t+1}^{(k)}$ from the predictive distribution $p(y_{t+1}|y_{1:t}, \mathcal{M})$, which is approximated by a discrete random measure constructed by the particle filter,
2. compute the KS statistic $D_{1,K,t+1}$, and
3. update the mean $\overline{D}_{1,K,t}$ to $\overline{D}_{1,K,t+1}$.

At the end, use the obtained statistic $\overline{D}_{1,K,T}$ do decide whether to reject $\mathcal{M}$.

4. COMPUTER SIMULATIONS

In order to illustrate the validity of the proposed model assessment approach, we investigated the sensitivity of the KS statistics to parameter mismatches in the nonlinear and non-Gaussian system described by

$$x_t \quad = \quad \beta x_{t-1} + \frac{25 x_{t-1}}{1 + x_{t-1}^2} + 8\cos(\phi t) + \sigma u_t \tag{21}$$

$$y_t \quad = \quad \frac{1}{20} x_t^2 + v_t \tag{22}$$

where $u_t \sim \mathcal{N}(0,1)$ and $v_t \sim \mathcal{N}(0,1)$ were statistically independent Gaussian noise terms, the prior distribution of the state signal was Gaussian, namely $x_0 \sim \mathcal{N}(\cos(0), 1)$, and $\boldsymbol{\theta} = [\beta, \phi, \sigma]^\top$ was the vector of parameters of interest, including the signal gain β, the frequency ϕ and the standard deviation σ for the state equation. We generated observations by running the system (21)-(22) with the parameter vector $\boldsymbol{\theta}_0 = [0.5, 1.2, 1.0]^\top$, which defined the "true model" for our experiments and was denoted by $\mathcal{M}_0$. Then we used these observations in the standard particle filtering algorithm (sequential importance resampling – SIR) [4]. The algorithm was designed under each one of the following assumptions:

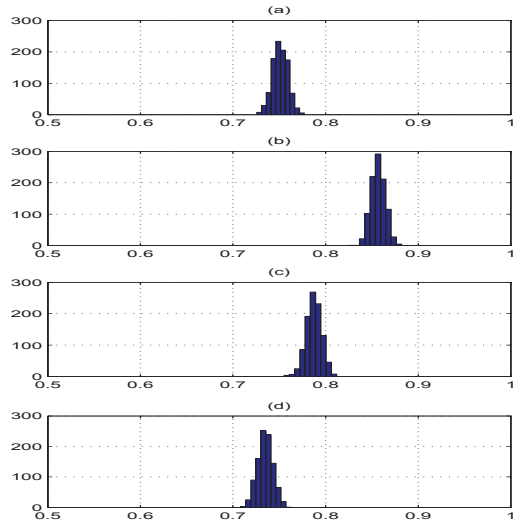

Fig. 1. Histograms of the averaged KS statistic, $\overline{D}_{1,500,300}$. Plot (a): the SIR algorithm is perfectly matched to the true model. Plot (b): mismatch of $+0.01$ in the frequency parameter, ϕ. Plot (c): mismatch of -0.1 in the signal gain, β. Plot (d): mismatch of $+0.5$ in the signal noise standard deviation, σ.

1. The parameter vector was $\boldsymbol{\theta}_1 = \boldsymbol{\theta}_0$, i.e., the SIR procedure was perfectly matched to the true model.
2. The parameter vector was $\boldsymbol{\theta}_2 = [0.5, 1.21, 1.0]^\top$, i.e., there was a mismatch of 0.01 in the frequency parameter, ϕ.
3. The parameter vector was $\boldsymbol{\theta}_3 = [0.4, 1.2, 1.0]^\top$, i.e., the signal gain, β, was decreased by 0.1.
4. The parameter vector was $\boldsymbol{\theta}_4 = [0.5, 1.2, 1.5]^\top$, i.e., the standard deviation of the noise in the state equation was increased by 0.5.

In all the cases, the SIR algorithms were run with $M = 500$ particles with a systematic resampling step at each time instant t.

We obviously had four different models, $\mathcal{M}_1, \ldots, \mathcal{M}_4$, defined by the corresponding parameter vectors, $\boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_4$. To assess the adequacy of the model $\mathcal{M}_i$, we generated a sequence of 300 observations from model $\mathcal{M}_0$, ran the SIR algorithm to obtain the approximations $p(y_{t+1}|y_{1:t}, \mathcal{M}_i)$, $t = 1, \ldots, 299$, drew $K = 500$ samples of y_{t+1} and, for each t, computed the KS statistics $D_{1,500,t+1}$. The expected value $E[D_{1,500,t}]$ was estimated by way of time averaging, $\overline{D}_{1,500,300} = \frac{1}{300}\sum_{t=1}^{300} D_{1,500,t}$. Moreover, since $\overline{D}_{1,500,300}$ is a random variable in this setting, we independently repeated the above procedure 1000 times in order to get independent draws $\overline{D}^{(j)}_{1,M,300}$, $j = 1, ..., 1000$.

Figure 1(a) shows the histogram of the averaged KS statistic obtained for model $\mathcal{M}_1 = \mathcal{M}_0$. In this case, there was no model mismatch and, as predicted by our analysis, the statistic $\overline{D}^{(j)}_{1,500,300}$, $j = 1, ..., 1000$, concentrated around the expected value $E[D_{1,500,300}] = 0.75$ for the true model. In particular, the empirical mean was 0.751 while the empirical variance was 6.858×10^{-5}.

Figure 1(b) depicts the effect on the KS statistic of assuming model $\mathcal{M}_1$, which involved a slight frequency offset of $+0.01$. The system (21)–(22) turns out to be very sensitive to changes in this parameter and this is sharply shown by the distribution of $\overline{D}_{1,500,300}$. Indeed, the histogram is clearly shifted to the right when compared with Figure 1(a). The empirical mean and variance of $\overline{D}_{1,500,300}$, assuming $\mathcal{M}_2$ was the correct model, were 0.857 and 5.783×10^{-5}, respectively. Based on the obtained statistics $\overline{D}_{1,500,300}$ in each of the trials, the model would be always rejected.

The effect of the other parameters on the KS statistics was less apparent. Figures 1(c) and 1(d) show the histograms obtained for models $\mathcal{M}_3$ and $\mathcal{M}_4$, respectively, which corresponded to mismatches in the signal gain, β, and the standard deviation of the noise in the state equation, σ. The empirical mean and variance for model $\mathcal{M}_3$ were 0.787 and 7.014×10^{-5}, respectively, while the corresponding values for $\mathcal{M}_4$ were 0.735 and 6.780×10^{-5}. From the histograms, we can conclude that the incorrect model $\mathcal{M}_4$ is the least prone to rejection.

5. CONCLUSIONS

We proposed a methodology for deciding whether to reject a considered model for a set of observed data. The test is based on the assumption that we can generate samples of future data by using the predictive distribution of the data and conditioned on the assessed model. In this paper, we approximated the predictive distribution of the data via the concept of particle filtering. The empirical distribution of the generated data is compared with the empirical distribution of the observed data by using the two-sided Kolmogorov-Smirnov test. If the model is correct, the average value of this statistic over time is approximately distributed as a Gaussian with known mean and variance. The statistics of the mean value are finally used to define a test for rejection of the model.

6. REFERENCES

[1] J. M. Bernardo and A. F. M. Smith, *Bayesian Theory*, John Wiley & Sons, New York, 1994.

[2] P. M. Djurić, "Monitoring and selection of dynamic models by Monte Carlo sampling," in *the IEEE SP Workshop on High Order Statistics*, 1999, pp. 191–194.

[3] A. S. Gelman, X.-L. Meng, and H. Stern, "Posterior predicitive assessment of model fitness via realzied discrepancies," *Statistica Sinica*, vol. 6, pp. 733–807, 1996.

[4] P. M. Djurić, J. H. Kotecha, J. Zhang, Y. Huang, T. Ghirmai, M. F. Bugallo, and J. Míguez, "Particle filtering," *IEEE Signal Processing Magazine*, vol. 20, no. 5, pp. 19–38, 2003.

[5] A. Doucet, N. de Freitas, and N. Gordon, Eds., *Sequential Monte Carlo Methods in Practice*, Springer, New York, 2001.

[6] J. P. Bickel and K. Docksum, *Mathematical Statistics*, McGraw-Hill, New York, 2001.

[7] V. K. Rohatgi and A. K. Md. E. Sales, *An Introduction to Probability and Statistics*, John Wiley & Sons, New York, 2001.

[8] F. J. Massey, "Distribution table for the deviation between two sample cumulatives," *Annals of Mathematical Statistics*, vol. 23, pp. 435–441, 1952.

[9] N. V. Smirnov, "On the estimation of the discrepancy between empirical curves of distributions for two indpendent samples," *Bull. Moscow University*, vol. 2, pp. 3–16, 1939.