

# BAYESIAN STATE AND PARAMETER ESTIMATION OF UNCERTAIN DYNAMICAL SYSTEMS

JIANYE CHING<sup>1</sup>, JAMES L. BECK<sup>2</sup> AND KEITH A. PORTER<sup>3</sup>

## ABSTRACT

The focus of this paper is Bayesian state and parameter estimation using nonlinear models. A recently developed method, the particle filter, is studied that is based on stochastic simulation. Unlike the well-known extended Kalman filter, the particle filter is applicable to highly nonlinear systems with non-Gaussian uncertainties. Recently developed techniques that improve the convergence of the particle filter simulations are introduced and discussed. Comparisons between the particle filter and the extended Kalman filter are made using several numerical examples of nonlinear systems. The results indicate that the particle filter provides consistent state and parameter estimates for highly nonlinear systems, while the extended Kalman filter does not.

Key words: Bayesian analysis, State estimation, Parameter estimation, Dynamical systems, Monte Carlo simulation, Importance sampling, Particle filters, Stochastic simulation

## 1. INTRODUCTION

### 1.1 Applications of state estimation in civil engineering

State estimation is the process of using dynamic data from a system to estimate quantities that give a complete description of the state of the system according to some representative model of it. State estimation has the potential to be widely applied in civil engineering. For instance, structural health monitoring techniques that detect changes of dynamical properties of

---

<sup>1</sup> George W. Housner Postdoctoral Fellow, Dept. of Civil Engineering, California Institute of Technology, Pasadena, CA 91125, USA. Email: jyching@netzero.net

<sup>2</sup> Professor of Applied Mechanics and Civil Engineering, Dept. of Civil Engineering, California Institute of Technology, Pasadena, CA 91125, USA. Email: jimbeck@caltech.edu

<sup>3</sup> George W. Housner Senior Researcher, Dept. of Civil Engineering, California Institute of Technology, Pasadena, CA 91125, USA. Email: keithp@caltech.edu

structural systems during earthquakes can be cast into a state estimation problem. More generally, system identification is useful for better understanding the nonlinear behavior of structures subject to seismic loading. A state estimation methodology can be used for this purpose. For structural control, the ability to estimate system states in real time may help to accomplish an efficient control strategy. For performance-based earthquake engineering, state estimation can provide crucial information to assess seismic performance of an instrumented building or bridge in terms of repair costs, casualties and repair duration (dollars, death and downtime) shortly after the cessation of strong motion.

Due to their wide applicability, state estimation and identification methods have been studied in civil engineering for various purposes: Beck (1978) used an invariant-embedding filter for modal identification; Yun and Shinozuka (1980) used an extended Kalman filter to study nonlinear fluid-structure interaction; Hoshiya and Saito (1984) used the extended Kalman filter for structural system identification; Lin *et al.* (1990) developed an identification methodology for better understanding of the degrading behavior of structures subject to dynamic loads; Koh and See (1994) developed an adaptive filter algorithm that also updates uncertainty estimates; Ghanem and Shinozuka (1995) presented several adaptive estimation techniques (e.g. extended Kalman filter, recursive least squares, recursive prediction error methods) and verified them using experimental data (Shinozuka and Ghanem 1995); Glaser (1996) used the Kalman filter to identify the time-varying natural frequency and damping of a liquefied soil to get insight into the liquefaction phenomenon; Sato and Qi (1998) derived an adaptive  $H_\infty$  filter and applied it to time-varying linear and nonlinear structural systems in which displacements and velocities of the floors are measured; Smyth *et al.* (1999) formulated an adaptive least squares algorithm for iden-

tifying multi-degree of freedom nonlinear hysteretic systems for the purpose of control and monitoring.

## **1.2 Development of Bayesian state-estimation algorithms**

Among state estimation methodologies, those founded on the Bayesian framework are powerful because: (1) they are rigorously based on the probability axioms and therefore preserve information; and (2) they give the probability density function (PDF) of the system state conditioned on the available information, which may then be used for any probability-based structural health monitoring, system identification, reliability assessment and control techniques. With the PDF available, we can not only estimate the state but also give a description of the associated uncertainties. For the Bayesian state-estimation algorithms, Kalman formulated the well-known Kalman filter (KF) (Kalman 1960; Kalman and Bucy 1961) for linear systems with Gaussian uncertainties. Later, KF was modified to give the extended Kalman filter (EKF) (Jazwinski 1970) to accommodate lightly nonlinear systems, and this is basically the dominant Bayesian state-estimation algorithm for nonlinear systems and non-Gaussian uncertainties for the last 30 years.

Although EKF has been widely used, it is only reliable for systems that are almost linear on the time scale of the updating intervals (Julier *et al.* 2000; Wan and van der Merwe 2000). However, civil-engineering systems are often highly nonlinear when subject to severe loading events, in this case, the applicability of the KF and EKF is often questionable. These older techniques have been used by civil engineering researchers for decades (Beck 1978; Yun and Shinokawa 1980; Hoshiya and Saito 1984; Koh and See 1994) although their applicability for nonlinear systems and non-Gaussian uncertainties is seldom verified either empirically or theoretically.

Several important breakthroughs (Alspach and Sorenson 1972; Gordon *et al.* 1993; Kitagawa 1996; Doucet *et al.* 2000; Julier *et al.* 2000) have produced Bayesian state-estimation algo-

rithms that are applicable to highly nonlinear systems. State estimation for general nonlinear dynamical systems is still an active research area, and novel techniques (e.g.: van der Merwe *et al.* 2000; van der Merwe and Wan 2003) can be found in the most recent signal-processing literature. Although these breakthroughs have had significant impact in the area of signal processing, they are rarely seen in civil engineering. Exceptions include Yoshida and Sato (2002) and Maruyama and Hoshiya (2003), who have implemented an improved version of Kitagawa's approach for system identification and damage detection.

### **1.3 Scope of this paper**

In this paper, we introduce some recent developments in Bayesian state estimation that use stochastic simulation. The technique called particle filter (PF) is presented and discussed. These simulation techniques have the following advantages: (1) they are applicable to highly nonlinear systems with non-Gaussian uncertainties; (2) they are not limited to the first two moments as in the KF and EKF; and (3) as the sample size approaches infinity, the resulting estimates of the state, or any function of the state, converge to their expected values conditional on the dynamic data up to the present time. However, the simulation is usually computationally expensive and sometimes the state estimates can be inaccurate due to insufficient samples. We introduce several developments more recent than Kitagawa (1996) that address the aforementioned difficulties and present new techniques that improve convergence. The performance of the EKF and PF methods is compared through several numerical examples.

This paper has the following structure: In Section 2, we define the general problem of Bayesian state estimation for nonlinear dynamical systems. In Section 3, we review the KF and EKF algorithms. In Section 4, we introduce importance-sampling filter techniques, including the

particle filter. In Section 5, several numerical examples are presented to compare the older and newer techniques.

## 2. STATE ESTIMATION

Consider the following discrete-time state-space model of a dynamical system:

$$x_k = f_{k-1}(x_{k-1}, u_{k-1}, w_k) \quad y_k = h_k(x_k, u_k, v_k) \quad k = 1, 2 \dots T \quad (1)$$

The two equations in (1) are called, from left to right, state transition (or evolution) and observation (or output) equations, respectively. In this equation,  $x_k \in R^n$ ,  $u_k \in R^p$  and  $y_k \in R^q$  are the system state, input (known excitation) and output at time  $k$ , respectively;  $w_k \in R^l$  and  $v_k \in R^m$  are introduced to account for unknown disturbances, model errors and measurement noise;  $f_k$  is the prescribed state transition function at time  $k$ ; and  $h_k$  is the prescribed observation function at time  $k$ . The values of the variables  $x_k$ ,  $y_k$ ,  $w_k$  and  $v_k$  are uncertain and so are modeled as random variables, while  $u_k$  is considered to be a known excitation.

For each time  $k$ , the dynamical system input  $u_k$  and output  $\hat{y}_k$  are measured. (In order to avoid confusion, we denote the observed output value by  $\hat{y}_k$ ). We denote  $\{\hat{y}_1, \hat{y}_2, \dots, \hat{y}_k\}$  and  $\{u_1, u_2, \dots, u_k\}$  by  $\hat{Y}_k$  and  $U_k$ , respectively. Our goal is to sequentially evaluate the conditional probability density function (PDF)  $p(x_k | \hat{Y}_k)$  for the state  $x_k$  at every time  $k$ , i.e. to sequentially update this conditional PDF using the observed system input and output up to the current time, based on prescribed probabilistic models for  $w_k$  and  $v_k$ . From this conditional PDF, some important features of the state, such as the conditional expectation  $E(x_k | \hat{Y}_k)$  and conditional covariance matrix  $Cov(x_k | \hat{Y}_k)$ , can be estimated. Note that the conditioning of every PDF on  $U_k$  is left implicit.

The basic equations for updating  $p(x_{k-1} | \hat{Y}_{k-1})$  to  $p(x_k | \hat{Y}_k)$  are the predictor and updater (or corrector) equations that follow from the Theorem of Total Probability and Bayes Theorem, respectively:

$$p(x_k | \hat{Y}_{k-1}) = \int p(x_k | x_{k-1}) p(x_{k-1} | \hat{Y}_{k-1}) dx_{k-1}$$

$$p(x_k | \hat{Y}_k) = \frac{p(\hat{y}_k | x_k) p(x_k | \hat{Y}_{k-1})}{\int p(\hat{y}_k | x_k) p(x_k | \hat{Y}_{k-1}) dx_k} = \frac{p(\hat{y}_k | x_k) p(x_k | \hat{Y}_{k-1})}{p(\hat{y}_k | \hat{Y}_{k-1})} \quad (2)$$

where  $\hat{Y}_{k-1}$  is dropped in  $p(x_k | x_{k-1})$  and  $p(\hat{y}_k | x_k)$  because the models for the state transition and observation PDFs make it irrelevant. The main challenge in Bayesian state estimation for nonlinear systems is that these basic equations cannot be readily evaluated because they involve high-dimensional integrations.

### 3. KALMAN FILTER

When  $f_k$  and  $h_k$  in (1) are both linear in  $u_k$ ,  $x_k$ ,  $w_k$  and  $v_k$ , i.e.

$$f_k(x_k, u_k, w_k) = A_k x_k + B_k u_k + G_k w_k \quad h_k(x_k, u_k, v_k) = C_k x_k + D_k u_k + H_k v_k \quad (3)$$

and  $w_k$  and  $v_k$  are zero-mean independent Gaussian random variables with identity covariance matrices, then the conditional PDF is also Gaussian and can be updated analytically. Furthermore, it is sufficient to update the first two moments because they completely specify this conditional PDF. The updating algorithm is the well-known Kalman filter (KF). It comprises two steps, the predictor (uncertainty propagation) and updater (estimation) steps.

In the uncertainty propagation step, the goal is to compute  $p(x_k, y_k | \hat{Y}_{k-1})$  from  $p(x_{k-1} | \hat{Y}_{k-1})$ . First,  $p(x_k | \hat{Y}_{k-1})$  is computed based on  $p(x_{k-1} | \hat{Y}_{k-1})$  using the following moment equations:

$$\begin{aligned}
x_{k|k-1} &\equiv E(x_k | \hat{Y}_{k-1}) = A_{k-1}x_{k-1|k-1} + B_{k-1}u_{k-1} \\
P_{k|k-1} &\equiv Cov(x_k | \hat{Y}_{k-1}) = A_{k-1}P_{k-1|k-1}A_{k-1}^T + G_{k-1}G_{k-1}^T
\end{aligned} \tag{4}$$

Note that the values  $x_{0|0}$  and  $P_{0|0}$  have to be given prior to the initialization of the algorithm.

Second,  $p(y_k | \hat{Y}_{k-1})$  is computed based on  $p(x_k | \hat{Y}_{k-1})$  and  $u_k$  using the following moment equations:

$$\begin{aligned}
y_{k|k-1} &\equiv E(y_k | \hat{Y}_{k-1}) = C_k x_{k|k-1} + D_k u_k \\
P_{k|k-1}^y &\equiv Cov(y_k | \hat{Y}_{k-1}) = C_k P_{k|k-1} C_k^T + H_k H_k^T
\end{aligned} \tag{5}$$

and finally, the conditional covariance between  $x_k$  and  $y_k$  is the  $n \times q$  matrix computed as follows:

$$P_{k|k-1}^{xy} \equiv Cov(x_k, y_k | \hat{Y}_{k-1}) = P_{k|k-1} C_k^T \tag{6}$$

This completes the computation of all the moments needed to specify the Gaussian PDF  $p(x_k, y_k | \hat{Y}_{k-1})$ .

In the estimation step,  $p(x_k | \hat{Y}_k)$  is updated based on (2):

$$\begin{aligned}
p(x_k | \hat{Y}_k) &= \frac{p(\hat{y}_k | x_k) p(x_k | \hat{Y}_{k-1})}{p(\hat{y}_k | \hat{Y}_{k-1})} \\
&= \text{const} \cdot e^{-\frac{1}{2} [(\hat{y}_k - C_k x_k - D_k u_k)^T (H_k H_k^T)^{-1} (\hat{y}_k - C_k x_k - D_k u_k)]} \cdot e^{-\frac{1}{2} [(x_k - x_{k|k-1})^T (P_{k|k-1})^{-1} (x_k - x_{k|k-1})]}
\end{aligned} \tag{7}$$

where  $\text{const}$  is a quantity not depending on  $x_k$ . Because  $p(x_k | \hat{Y}_k)$  is Gaussian, differentiating

$p(x_k | \hat{Y}_k)$  with respect to  $x_k$  and solving for zero, we obtain  $x_{k|k}$ ; on the other hand,  $P_{k|k}$  is equal

to the negative of the inverse of the Hessian of  $\log[p(x_k | \hat{Y}_k)]$ :

$$\begin{aligned}
x_{k|k} &= x_{k|k-1} + \left[ I + P_{k|k-1} C_k^T (H_k H_k^T)^{-1} C_k \right]^{-1} \cdot P_{k|k-1} C_k^T (H_k H_k^T)^{-1} \cdot (\hat{y}_k - y_{k|k-1}) \\
P_{k|k} &= \left[ P_{k|k-1}^{-1} + C_k^T (H_k H_k^T)^{-1} C_k \right]^{-1}
\end{aligned} \tag{8}$$

Making use of the following lemmas

$$(I + PQ)^{-1} P = P(I + QP)^{-1} \quad (A^{-1} + VC^{-1}V^T)^{-1} = A - AV(C + V^T AV)^{-1} V^T A \quad (9)$$

where  $P$  and  $Q$  are conformable matrices,  $A$  and  $C$  are positive definite, we conclude with the following equations for the estimation step:

$$\begin{aligned} x_{k|k} &= x_{k|k-1} + P_{k|k-1}^{xy} \cdot (P_{k|k-1}^y)^{-1} \cdot (\hat{y}_k - y_{k|k-1}) \\ P_{k|k} &= P_{k|k-1} - P_{k|k-1}^{xy} \cdot (P_{k|k-1}^y)^{-1} \cdot P_{k|k-1}^{xy T} \end{aligned} \quad (10)$$

### 3.1 Extended Kalman filter

Many dynamical systems exhibit nonlinear behavior, and the direct use of KF is prohibited. However, if  $f_k$  and  $h_k$  are only slightly nonlinear, an approximation for KF can be derived by linearizing the uncertainty propagation and estimation steps. The resulting filter is the well-known extended Kalman filter (EKF). To explain the linearization (LN) technique for the uncertainty propagation step, we consider the following general uncertainty propagation problem:

$$Y = f(X) \quad (11)$$

where  $X \in R^n$  and  $Y \in R^m$  are uncertain vectors. Using Taylor series expansion around  $X = EX$ , we have

$$f(X) = f(EX) + \nabla_x f \cdot (X - EX) + HOT \quad (12)$$

where  $\nabla_x f$  is the Jacobian matrix evaluated at  $x = EX$ ;  $HOT$  denotes the higher order terms.

As a result, the first two moments of  $Y$  are

$$EY = f(EX) + HOT \quad (13)$$

and

$$Cov(Y) = (\nabla_x f) \cdot Cov(X) \cdot (\nabla_x f)^T + HOT \quad (14)$$

Under the assumption that  $f(x)$  is nearly linear near  $x = EX$ , all higher order terms vanish; therefore,  $EY$  and  $Cov(Y)$  are approximated by

$$EY_{LN} = f(EX) \quad Cov(Y)_{LN} = (\nabla_x f) \cdot Cov(X) \cdot (\nabla_x f)^T \quad (15)$$

where  $EY_{LN}$  and  $Cov(Y)_{LN}$  denote the approximations of the LN technique for  $EY$  and  $Cov(Y)$ .

The approximations  $EY_{LN}$  and  $Cov(Y)_{LN}$  are accurate estimates of  $EY$  and  $Cov(Y)$  if  $f(x)$  is almost linear on the support region of the PDF of  $X$  and become exact when  $f(x)$  is linear in  $x$ . On the other hand, the approximations are poor if  $f(x)$  is highly nonlinear on the support region of the PDF of  $X$ .

For the uncertainty propagation step in EKF, the goal is to find the LN approximations of  $x_{k|k-1}$ ,  $P_{k|k-1}$ ,  $y_{k|k-1}$ ,  $P_{k|k-1}^y$  and  $P_{k|k-1}^{xy}$  based on  $x_{k-1|k-1}$  and  $P_{k-1|k-1}$ . To simplify the notation, we define  $z_k = [x_k^T \quad w_k^T \quad v_k^T]^T$  and  $z_{k|k} = [x_{k|k}^T \quad 0^T \quad 0^T]^T \in R^{n+l+m}$ , so  $f_k(x_k, u_k, w_k) = f_k(z_k, u_k)$ .

When propagating from  $[x_{k-1|k-1}, P_{k-1|k-1}]$  to  $[x_{k|k-1}, P_{k|k-1}]$ , we expand  $f_{k-1}(z_{k-1}, u_{k-1})$  in the neighborhood of  $z_{k-1|k-1}$ . With the LN approximation, we get

$$x_{k|k-1} = E[f_{k-1}(z_{k-1}, u_{k-1})] \approx f_{k-1}(z_{k-1|k-1}, u_{k-1}) \equiv x_{k|k-1}^{LN} \quad (16)$$

and

$$\begin{aligned} P_{k|k-1} &\approx (\nabla_z f_{k-1}) \cdot E\left\{(z_{k-1} - z_{k-1|k-1})(z_{k-1} - z_{k-1|k-1})^T \mid D_{k-1}\right\} \cdot (\nabla_z f_{k-1})^T \\ &= (\nabla_z f_{k-1}) \cdot Cov\{z_{k-1} \mid D_{k-1}\} \cdot (\nabla_z f_{k-1})^T \equiv P_{k|k-1}^{LN} \end{aligned} \quad (17)$$

where  $\nabla_z f_{k-1} \in R^{n \times (n+l+m)}$  is the Jacobian matrix evaluated at  $z_{k-1} = z_{k-1|k-1}$ . It can be seen that

$$P_{k|k-1}^{LN} = (A_{k-1}^{LN}) \cdot P_{k-1|k-1} \cdot (A_{k-1}^{LN})^T + (G_{k-1}^{LN}) \cdot (G_{k-1}^{LN})^T \quad (18)$$

where  $A_{k-1}^{LN} \equiv \nabla_x f_{k-1} \Big|_{z_{k-1}=z_{k-1|k-1}} \in \mathbb{R}^{n \times n}$  and  $G_{k-1}^{LN} \equiv \nabla_w f_{k-1} \Big|_{z_{k-1}=z_{k-1|k-1}} \in \mathbb{R}^{n \times l}$  are the Jacobian matrices.

Similarly,  $h_k$  is also linearized to get the approximations for  $y_{k|k-1}$ ,  $P_{k|k-1}^y$  and  $P_{k|k-1}^{xy}$ :

$$\begin{aligned} y_{k|k-1} &\approx h_{k-1}(x_{k|k-1}^{LN}, u_k, 0) \equiv y_{k|k-1}^{LN} \\ P_{k|k-1}^y &\approx (C_k^{LN}) P_{k|k-1}^{LN} (C_k^{LN})^T + (H_k^{LN})(H_k^{LN})^T \equiv P_{k|k-1}^{y, LN} \\ P_{k|k-1}^{xy} &\approx P_{k|k-1}^{LN} (C_k^{LN})^T \equiv P_{k|k-1}^{xy, LN} \end{aligned} \quad (19)$$

where  $C_k^{LN} \equiv \nabla_x h_k \Big|_{x_k=x_{k|k-1}, v_k=0} \in \mathbb{R}^{q \times n}$  and  $H_k^{LN} \equiv \nabla_v h_k \Big|_{x_k=x_{k|k-1}, v_k=0} \in \mathbb{R}^{q \times m}$ .

For the estimation step, (10) can still be used as an approximation. If  $f_k$  and  $h_k$  are indeed linear, EKF is identical to KF. The degree of accuracy of EKF relies on the validity of the linear approximation. EKF is not suitable to track multi-modal or highly non-Gaussian conditional PDFs due to the fact that it only updates the first two moments.

When the system parameters are unknown, it is important to also estimate them. Uncertain parameters can be augmented into system states and estimated using EKF (even if the original dynamical system is linear, the dynamical system for the augmented state is nonlinear and so KF is not applicable). However, the EKF algorithm is not suitable for estimating unknown parameters used to parameterize the amplitudes of the uncertainty terms  $w_k$  and  $v_k$ . We discuss this issue and provide new solutions in Appendix I.

#### 4. PARTICLE FILTERS

We have seen that EKF can only propagate and estimate the first two moments of the conditional PDF. For systems with non-Gaussian uncertainties, it is often desirable to propagate and estimate the conditional PDF itself; however, doing so requires, in effect, an infinite number of parameters to represent the functional form of the conditional PDF. An alternative is to conduct stochastic simulation by drawing samples from the conditional PDF so that the conditional

expectation of any function of  $x_k$  can be consistently estimated. We focus on stochastic simulation techniques for state estimation in this section and use the term particle filters (PF) to denote the resulting algorithms (following van der Merwe *et al.* 2000; Doucet and Andrieu 2000). Similar PF algorithms have been called Monte Carlo filters by Kitagawa (1996) and sequential Monte Carlo Bayesian filters by Doucet and Godsill (1998) and Doucet *et al.* (2000).

#### 4.1 Basic equations

We first present some basic equations that are useful throughout this section. Let  $X_k = \{x_0, x_1, \dots, x_k\}$ , then according to Bayes rule,

$$\begin{aligned}
p(X_k | \hat{Y}_k) &= \frac{p(X_k, \hat{Y}_k)}{p(\hat{Y}_k)} = \frac{p(X_{k-1}, x_k, \hat{Y}_{k-1}, \hat{y}_k)}{p(\hat{Y}_k)} \\
&= \frac{p(X_{k-1}, \hat{Y}_{k-1}) \cdot p(x_k, \hat{y}_k | X_{k-1}, \hat{Y}_{k-1})}{p(\hat{Y}_k)} = \frac{p(X_{k-1} | \hat{Y}_{k-1}) \cdot p(x_k, \hat{y}_k | X_{k-1}, \hat{Y}_{k-1})}{p(\hat{y}_k | \hat{Y}_{k-1})} \quad (20) \\
&= p(X_{k-1} | \hat{Y}_{k-1}) \cdot \frac{p(\hat{y}_k | x_k, X_{k-1}, \hat{Y}_{k-1}) \cdot p(x_k | X_{k-1}, \hat{Y}_{k-1})}{p(\hat{y}_k | \hat{Y}_{k-1})} = p(X_{k-1} | \hat{Y}_{k-1}) \cdot \frac{p(\hat{y}_k | x_k) \cdot p(x_k | x_{k-1})}{p(\hat{y}_k | \hat{Y}_{k-1})}
\end{aligned}$$

where we have used the fact that  $p(\hat{y}_k | x_k, X_{k-1}, \hat{Y}_{k-1}) = p(\hat{y}_k | x_k)$  and that  $p(x_k | X_{k-1}, \hat{Y}_{k-1}) = p(x_k | x_{k-1})$  based on (1) and the fact that the PDFs for  $v_k$  and  $w_k$  are prescribed. Evaluating the recursive equation in (20), we get

$$p(X_k | \hat{Y}_k) = p(x_0) \cdot \prod_{m=1}^k \frac{p(\hat{y}_m | x_m) \cdot p(x_m | x_{m-1})}{p(\hat{y}_m | \hat{Y}_{m-1})} = \frac{p(x_0)}{p(\hat{Y}_k)} \cdot \prod_{m=1}^k p(\hat{y}_m | x_m) \cdot p(x_m | x_{m-1}) \quad (21)$$

#### 4.2 Importance sampling for state estimation

Our interest is to develop a stochastic simulation algorithm for the conditional PDF  $p(X_k | \hat{Y}_k)$  that is Markovian in that information is required only from time steps  $k-1$  and  $k$ , and the earlier state information and observation data can be forgotten. In other words, the sam-

ple from  $p(X_k | \hat{Y}_k)$  must have the form  $\hat{X}_k = \{\hat{X}_{k-1}, \hat{x}_k\}$ , where  $\hat{x}_k$  is the new sample and  $\hat{X}_{k-1}$  is the previous sample from  $p(X_{k-1} | \hat{Y}_{k-1})$ . However, such a stochastic simulation algorithm cannot be directly implemented. This is because  $p(X_{k-1} | \hat{Y}_{k-1})$  is different from  $p(X_{k-1} | \hat{Y}_k) = p(X_{k-1} | \hat{Y}_{k-1}, \hat{y}_k)$ .

We can, however, sample from an importance sampling PDF  $q(X_k | \hat{Y}_k)$  that admits a Markovian sampling procedure by choosing  $q(X_{k-1} | \hat{Y}_{k-1})$  so that it is identical to  $q(X_{k-1} | \hat{Y}_k)$ . In other words, the structure of  $q(X_k | \hat{Y}_k)$  must be such that  $X_{k-1}$  is independent of  $y_k$  conditioned on  $Y_{k-1}$ . Drawing  $N$  samples  $\{\hat{X}_k^i : i = 1, \dots, N\}$  randomly from  $q(X_k | \hat{Y}_k)$  (we choose  $q(X_k | \hat{Y}_k)$  so that it can be readily sampled), the expectation of any function  $r(X_k)$  conditioned on  $\hat{Y}_k$  can be estimated using the importance sampling technique as follows:

$$E[r(X_k) | \hat{Y}_k] \approx \frac{1}{N} \sum_{i=1}^N \hat{\beta}_k^i \cdot r(\hat{X}_k^i) \equiv \hat{r}_{k,N}^1 \quad (22)$$

where  $\hat{\beta}_k^i = p(\hat{X}_k^i | \hat{Y}_k) / q(\hat{X}_k^i | \hat{Y}_k)$  is the non-normalized importance weight of the  $i$ -th sample.

Any quantity of interest can be estimated with the appropriate  $r(\cdot)$  function in (22), such as engineering response parameters, economic performance parameters or structural model parameters. Note that the conditional variance of any quantity  $Z_k$  can be computed by first setting  $r(Z_k) = Z_k$  so that  $E[r(Z_k) | \hat{Y}_k]$  is simply the conditional expectation of  $Z_k$  and then setting  $r(Z_k) = Z_k Z_k^T$  so that  $E[r(Z_k) | \hat{Y}_k]$  is the conditional second moment of  $Z_k$ .

Let  $\{X_k^i : i = 1, \dots, N\}$  denote the state variables corresponding to  $N$  random samples from  $q(X_k | \hat{Y}_k)$  (before drawing the actual samples). It is readily shown that the estimator

$r_{k,N}^1 = \frac{1}{N} \sum_{i=1}^N \beta_k^i \cdot r(X_k^i)$  is an unbiased estimator of  $E[r(X_k) | \hat{Y}_k]$  if the support region for

$p(X_k | \hat{Y}_k)$  is a subset of that for  $q(X_k | \hat{Y}_k)$ :

$$\begin{aligned} E[r_{k,N}^1] &= \frac{1}{N} \sum_{i=1}^N E_q[\beta_k^i \cdot r(X_k^i)] = E_q[\beta_k \cdot r(X_k)] \\ &= \int \left[ \frac{p(X_k | \hat{Y}_k)}{q(X_k | \hat{Y}_k)} \right] r(X_k) \cdot q(X_k | \hat{Y}_k) dX_k \\ &= \int r(X_k) \cdot p(X_k | \hat{Y}_k) dX_k = E[r(X_k) | \hat{Y}_k] \end{aligned} \quad (23)$$

According to the Central Limit Theorem,  $r_{k,N}^1$  converges (as  $N$  approaches infinity) to a Gaussian random variable with mean equal to  $E[r(X_k) | \hat{Y}_k]$  and with variance that decays as  $1/N$ .

Therefore,  $r_{k,N}^1$  is a consistent estimator of  $E[r(X_k) | \hat{Y}_k]$ .

Although  $r_{k,N}^1$  is unbiased and consistent, it is not a feasible estimator because the non-normalized importance weights  $\beta_k^i = p(X_k^i | \hat{Y}_k) / q(X_k^i | \hat{Y}_k)$  depend on  $p(X_k^i | \hat{Y}_k)$ , which cannot be computed easily since in order to evaluate  $p(X_k^i | \hat{Y}_k)$ , we have to evaluate  $p(\hat{Y}_k)$ , as shown by (21), which is a difficult task. Nevertheless, we show that the following estimator is computable while it is asymptotically unbiased and consistent:

$$r_{k,N}^2 \equiv \left( \frac{1}{N} \sum_{i=1}^N \beta_k^i \cdot r(X_k^i) \right) / \left( \frac{1}{N} \sum_{j=1}^N \beta_k^j \right) = r_{k,N}^1 / \bar{\beta}_k^N \quad (24)$$

where

$$\bar{\beta}_k^N = \left( \sum_{j=1}^N \beta_k^j \right) / N \quad (25)$$

Note that  $\hat{r}_{k,N}^2$ , unlike  $\hat{r}_{k,N}^1$ , can be computed conveniently from samples  $\{\hat{X}_k^i : i = 1, \dots, N\}$ :

$$\hat{r}_{k,N}^2 = \sum_{i=1}^N \left[ \beta_k^i / \left( \sum_{j=1}^N \beta_k^j \right) \right] \cdot r(\hat{X}_k^i) = \sum_{i=1}^N \tilde{\beta}_k^i \cdot r(\hat{X}_k^i) \quad (26)$$

where

$$\begin{aligned}\tilde{\beta}_k^i &= p(\hat{X}_k^i | \hat{Y}_k) / q(\hat{X}_k^i | \hat{Y}_k) \Big/ \left( \sum_{j=1}^N p(\hat{X}_k^j | \hat{Y}_k) / q(\hat{X}_k^j | \hat{Y}_k) \right) \\ &= \frac{p(\hat{x}_0^i)}{q(\hat{X}_k^i | \hat{Y}_k)} \cdot \prod_{m=1}^k p(\hat{y}_m | \hat{x}_m^i) \cdot p(\hat{x}_m^i | \hat{x}_{m-1}^i) \Big/ \left( \sum_{j=1}^N \frac{p(\hat{x}_0^j)}{q(\hat{X}_k^j | \hat{Y}_k)} \cdot \prod_{m=1}^k p(\hat{y}_m | \hat{x}_m^j) \cdot p(\hat{x}_m^j | \hat{x}_{m-1}^j) \right)\end{aligned}\quad (27)$$

Note that the troublesome factor  $p(\hat{Y}_k)$  in (21) has been cancelled due to the use of the normalized importance weights  $\{\tilde{\beta}_k^i : i=1, \dots, N\}$ , i.e.  $\sum_{i=1}^N \tilde{\beta}_k^i = 1$ . Also, the likelihood functions  $p(\hat{y}_m | \hat{x}_m^i)$  and  $p(\hat{x}_m^i | \hat{x}_{m-1}^i)$  can be readily evaluated using the prescribed PDFs for  $v_m$  and  $w_m$  if the mappings in (1) uniquely specify  $v_m$  and  $w_m$ , given  $y_m, x_m$  and  $x_{m-1}$ . Our choice of  $q(\hat{X}_k^i | \hat{Y}_k)$  is such that it can be readily evaluated too.

To sketch the proof for the asymptotic unbiasedness and consistency of  $r_{k,N}^2$ , note that

$$E_q(\beta_k^i) = \int \left[ p(X_k | \hat{Y}_k) / q(X_k | \hat{Y}_k) \right] \cdot q(X_k | \hat{Y}_k) \cdot dX_k = 1 \quad (28)$$

Therefore,  $\bar{\beta}_k^N$  in (25) converges (as  $N$  approaches infinity) to a Gaussian random variable with mean equal to 1 and with a variance that decays as  $1/N$ . As a consequence,

$$\lim_{N \rightarrow \infty} (\bar{\beta}_k^N) = 1 \quad w.p.1 \quad (29)$$

where w.p. stands for “with probability”, and

$$\begin{aligned}\lim_{N \rightarrow \infty} [r_{k,N}^2] &= \lim_{N \rightarrow \infty} \left[ \left( \frac{1}{N} \sum_{i=1}^N \beta_k^i \cdot r(X_k^i) \right) \Big/ \bar{\beta}_k^N \right] = \lim_{N \rightarrow \infty} \left[ \frac{1}{N} \sum_{i=1}^N \beta_k^i \cdot r(X_k^i) \right] = \lim_{N \rightarrow \infty} [r_{k,N}^1] \\ &= E[r(X_k) | \hat{Y}_k] \quad w.p.1\end{aligned}\quad (30)$$

which shows that  $r_{k,N}^2$  is asymptotically unbiased and consistent as  $N$  approaches infinity because  $r_{k,N}^1$  is.

The selection of an importance sampling PDF that admits a Markovian procedure is discussed in Appendix II. The conclusion is that the following importance sampling PDF performs well:

$$q(X_k | \hat{Y}_k) = p(x_0) \cdot \prod_{m=1}^k p(x_m | x_{m-1}, \hat{y}_m) \quad (31)$$

The corresponding modified non-normalized importance weight is

$$\beta_k = \prod_{m=1}^k \frac{p(\hat{y}_m | x_m) \cdot p(x_m | x_{m-1})}{p(x_m | x_{m-1}, \hat{y}_m)} = \beta_{k-1} \cdot \frac{p(\hat{y}_k | x_k) \cdot p(x_k | x_{k-1})}{p(x_k | x_{k-1}, \hat{y}_k)} \quad (32)$$

Doucet and Godsill (1998) and Liu and Chen (1998) discuss the optimality of this importance sampling PDF.

Because of the structure of the algorithm, at any time  $k$ , we are only required to store the sampled states and weights in time steps  $k$  and  $k-1$ , if the quantity of interest is  $r(x_k)$  and so depends only on the current state (although, clearly, additional dependence on the previous state  $x_{k-1}$  can also be treated). As a result, the following recursive algorithm can be used:

**Algorithm 4.1: Basic PF algorithm**

- (1) Initialize the  $N$  samples: Draw  $\hat{x}^i$  from  $p(x_0)$  and set  $\beta^i = 1/N$ ,  $i = 1, \dots, N$ .
- (2) At time  $k$ , store the previous samples and weights

$$\tilde{x}^i = \hat{x}^i \quad \tilde{\beta}^i = \beta^i \quad (33)$$

For  $i = 1, \dots, N$ , draw  $\hat{x}^i$  from  $p(x_k | x_{k-1} = \tilde{x}^i, \hat{y}_k)$  and update the importance weight

$$\beta^i = \tilde{\beta}^i \cdot \frac{p(\hat{y}_k | x_k = \hat{x}^i) \cdot p(x_k = \hat{x}^i | x_{k-1} = \tilde{x}^i)}{p(x_k = \hat{x}^i | x_{k-1} = \tilde{x}^i, \hat{y}_k)} \quad (34)$$

- (3) For  $i = 1, \dots, N$ ,  $E[r(x_k) | \hat{Y}_k]$  can be approximated based on (26) and (30):

$$E[r(x_k) | \hat{Y}_k] \approx \sum_{i=1}^N \left[ \beta^i / \left( \sum_{j=1}^N \beta^j \right) \right] \cdot r(\hat{x}^i) \quad (35)$$

where  $r(\cdot)$  is a function that maps from  $x_k$  to the quantity of interest.

(4) Do Steps (2) and (3) for time steps  $k = 1, \dots, T$ .

Usually,  $p(x_k | x_{k-1} = \tilde{x}^i, \hat{y}_k)$  in Step 2 is difficult to sample. Note that estimating the first two moments of  $p(x_k | x_{k-1} = \tilde{x}^i, \hat{y}_k)$  is a problem that can be solved using a single-time-step EKF algorithm. The least-informative PDF (i.e. the maximum entropy PDF; see Jaynes 1957) given the estimated two moments, which is a Gaussian PDF (denoted by  $p_{LI}(x_k | x_{k-1} = \tilde{x}^i, \hat{y}_k)$ ;  $LI$  subscript means ‘least-informative’), can be used in the importance sampling PDF in (31) and hence in Step 2 of Algorithm 4.1. The use of  $p_{LI}(x_k | x_{k-1} = \tilde{x}^i, \hat{y}_k)$  is discussed in Doucet and Godsill (1998) and van der Merwe *et al.* (2000).

**Algorithm 4.2: Determining**  $p_{LI}(x_k | x_{k-1} = \tilde{x}^i, \hat{y}_k)$

(1) Uncertainty propagation: compute from (16) and (17)

$$\begin{aligned} E_{LN}(x_k | x_{k-1} = \tilde{x}^i) &= f_{k-1}(\tilde{x}^i, u_{k-1}, 0) \equiv \tilde{x}_{k|k-1}^i \\ Cov_{LN}(x_k | x_{k-1} = \tilde{x}^i) &= (G_{k-1}^{LN}) \cdot Cov\{w_{k-1}\} \cdot (G_{k-1}^{LN})^T \equiv \tilde{P}_{k|k-1}^i \end{aligned} \quad (36)$$

where  $G_{k-1}^{LN} \equiv \nabla_w f_{k-1} \big|_{x_{k-1}=\tilde{x}^i, w_{k-1}=0}$  is the Jacobian matrix, and from (19)

$$\begin{aligned} E_{LN}(y_k | x_{k-1} = \tilde{x}^i) &= h_{k-1}(\tilde{x}_{k|k-1}^i, u_k, 0) \equiv \tilde{y}_{k|k-1}^i \\ Cov_{LN}(y_k | x_{k-1} = \tilde{x}^i) &= (C_k^{LN}) \tilde{P}_{k|k-1}^i (C_k^{LN})^T + (H_k^{LN})(H_k^{LN})^T \equiv \tilde{P}_{k|k-1}^{y,i} \\ Cov_{LN}(x_k, y_k | x_{k-1} = \tilde{x}^i) &= \tilde{P}_{k|k-1}^i (C_k^{LN})^T \equiv \tilde{P}_{k|k-1}^{xy,i} \end{aligned} \quad (37)$$

where  $C_k^{LN} \equiv \nabla_x h_k \big|_{x_k=\tilde{x}_{k|k-1}^i, v_k=0}$  and  $H_k^{LN} \equiv \nabla_v h_k \big|_{x_k=\tilde{x}_{k|k-1}^i, v_k=0}$ .

(2) Estimation: compute from (10)

$$\begin{aligned}
E_{LI}(x_k | x_{k-1} = \tilde{x}^i, \hat{y}_k) &= \tilde{x}_{k|k-1}^i + \tilde{P}_{k|k-1}^{xy,i} \cdot (\tilde{P}_{k|k-1}^{y,i})^{-1} \cdot (\hat{y}_k - \tilde{y}_{k|k-1}^i) \\
Cov_{LI}(x_k | x_{k-1} = \tilde{x}^i, \hat{y}_k) &= \tilde{P}_{k|k-1}^i - \tilde{P}_{k|k-1}^{xy,i} \cdot (\tilde{P}_{k|k-1}^{y,i})^{-1} \cdot \tilde{P}_{k|k-1}^{xy,i T}
\end{aligned} \tag{38}$$

Finally,  $p_{LI}(x_k | x_{k-1} = \tilde{x}^i, \hat{y}_k)$  is the Gaussian PDF with the two moments in (38).

### 4.3 Reducing degradation in performance: recursive resampling and parallel particle filters

Note that it is desirable to have the importance weights  $\{\beta^i : i = 1, 2, \dots, N\}$  be approximately uniform so that all samples contribute significantly in (35), but they become far from uniform as  $k$  grows, which is due to the recursion in (32) and the fact that  $q(X_k | \hat{Y}_k) \neq p(X_k | \hat{Y}_k)$ . Ultimately, a few weights become much larger than the rest, so the effective number of samples is small. Nevertheless, this degradation can be reduced, as described in this section and the next.

Instead of letting the  $N$  samples evolve through time independently (Algorithm 4.1), we can resample the samples when the importance weights become highly non-uniform (Kitagawa 1996; Doucet and Godsill 1998; Liu and Chen 1998; Doucet and Andrieu 2000). After the resampling, the importance weights become uniform, therefore the degradation problem is alleviated. The resampling step tends to terminate small-weight samples and duplicate large-weight samples and, therefore, forces the  $N$  samples to concentrate in the high probability region of  $p(x_k | \hat{Y}_k)$ .

Although the resampling step sets the weights back to uniform, the price to pay is that the samples become dependent and therefore collectively carry less information about the state. As a result, the resampling procedure should only be executed when the importance weights become highly non-uniform. This can be done by monitoring the coefficient of variation (c.o.v.) of the importance weights. The resampling procedure is executed only when this c.o.v. exceeds a certain threshold, indicating that the variability in the importance weights is large.

Another way to alleviate the dependency induced by the resampling step is to conduct several independent PF algorithms and combine all of the obtained samples. Although the samples obtained in a single algorithm can be highly dependent, the samples from different algorithms are completely independent. The resulting algorithm is as follows:

**Algorithm 4.3: Parallel PF algorithm with resampling**

(1) Initialize  $N$  samples for each of the  $L$  parallel PFs: Draw  $\hat{x}^{i,j}$  from  $p(x_0)$  and set

$$\beta^{i,j} = 1/N \text{ for } i = 1, \dots, N, j = 1, \dots, L.$$

(2) Perform the following steps (3)-(4) for  $j = 1, \dots, L$  independently. Since the processes are completely independent, they can be conducted in parallel.

(3) At time  $k$ , store the previous samples and weights

$$\tilde{x}^{i,j} = \hat{x}^{i,j} \quad \tilde{\beta}^{i,j} = \beta^{i,j} \tag{39}$$

For  $i = 1, \dots, N$ , draw  $\bar{x}^{i,j}$  from  $p_{Ll}(x_k | x_{k-1} = \tilde{x}^{i,j}, \hat{y}_k)$  and update the importance weight

$$\bar{\beta}^{i,j} = \tilde{\beta}^{i,j} \cdot \frac{p(\hat{y}_k | x_k = \bar{x}^{i,j}) \cdot p(x_k = \bar{x}^{i,j} | x_{k-1} = \tilde{x}^{i,j})}{p_{Ll}(x_k = \bar{x}^{i,j} | x_{k-1} = \tilde{x}^{i,j}, \hat{y}_k)} \tag{40}$$

(4) Compute the c.o.v. of  $\{\bar{\beta}^{i,j} : i = 1, \dots, N\}$ .

If the c.o.v. is larger than the prescribed threshold, then execute the resampling step for  $i = 1, \dots, N$ :

$$\hat{x}^{i,j} = \bar{x}^{i,j} \quad w.p. \quad \bar{\beta}^{i,j} / \sum_{i=1}^N \bar{\beta}^{i,j} \tag{41}$$

and set  $\beta^{i,j} = 1/N$  for  $i = 1, \dots, N$ . Otherwise, for  $i = 1, \dots, N$ :

$$\hat{x}^{i,j} = \bar{x}^{i,j} \quad \beta^{i,j} = \bar{\beta}^{i,j} / \sum_{i=1}^N \bar{\beta}^{i,j} \tag{42}$$

$$\text{Store } \hat{r}_{k,N}^j = \sum_{i=1}^N r(\hat{x}^{i,j}) \cdot \beta^{i,j}$$

(5)  $E[r(x_k) | \hat{Y}_k]$  can be then approximated by

$$E[r(x_k) | \hat{Y}_k] \approx \left( \sum_{j=1}^L \hat{r}_{k,N}^j \right) / L \quad (43)$$

(6) Do Steps 2 to 5 for  $k = 1, \dots, T$ .

#### 4.4 Reducing degradation in performance: local random walk

After the resampling step (Algorithm 4.3), large-weight samples are duplicated; therefore, some samples are the same samples of  $p(x_k | \hat{Y}_k)$ , which is not desirable from the point of view of preventing degradation. Andrieu *et al.* (1999) use the Markov chain Monte Carlo (MCMC) technique to force the duplicated samples to take a local random walk at each time step, where  $p(x_k | \hat{Y}_k)$  is the stationary PDF of the Markov chain. The detailed procedure for the MCMC step is also given in Ching *et al.* (2004).

#### 4.5 Advantages and disadvantages of the PF technique

The advantages of the PF technique include (1) as  $N$  (the number of samples per algorithm) approaches infinity, the value of any function of the state  $x_k$  estimated by PF converges to its expected value; therefore, the PF technique can be used to validate other methodologies; and (2) parallel computations are possible for PF algorithms. A disadvantage of the PF technique is that it is computationally expensive, especially when the degradation is severe so that we need large  $N$  and  $L$  to have the algorithm converge. In general, the required  $N$  and  $L$  grow with the size of the effective support region of  $p(x_k | \hat{Y}_k)$ . A simple test for convergence is to add parallel particle filters until the estimated quantity of interest,  $r(x_k)$ , does not significantly change. For

linear systems with time-varying unknown parameters, a more efficient PF algorithm is derived in Ching *et al.* (2004).

## 5. NUMERICAL EXAMPLES

We present three examples in this section. We generate data that is contaminated by noise for three simulated dynamical systems. With the simulated data, we use identification models that are derived from these dynamical systems to conduct EKF and PF and we compare their performance. The goal of these examples is to see if these techniques produce consistent results.

### 5.1 Planar four-story shear building with time-varying system parameters

#### *Data generation*

We first describe the system that generates the simulated data. Consider an idealized planar four degrees of freedom (DOF) shear building system with known time-invariant masses equal to  $m_1 = m_2 = m_3 = m_4 = 250,000kg$  (subscript denotes the story/floor number). The inter-story stiffnesses  $k_1, k_2, k_3$  and  $k_4$  change through time as shown in Figure 1 ( $k_1, k_3$  and  $k_4$  drift around certain values, while  $k_2$  significantly decreases and then partially recovers). The inter-story viscous damping coefficients are  $c_1, c_2, c_3$  and  $c_4$  and are also time-varying (Figure 1). The time evolutions of  $k_1, k_3, k_4, c_1, c_2, c_3$  and  $c_4$  are Brownian motions with standard deviation of the drift equal to 2% of their mean values during each sampling interval described later.

The governing equation of this system subject to base excitation is

$$M\ddot{x}_t + C_t\dot{x}_t + K_t x_t = F u_t \quad (44)$$

where

$$\begin{aligned}
x_t &= \begin{bmatrix} x_{4,t} \\ x_{3,t} \\ x_{2,t} \\ x_{1,t} \end{bmatrix} & M &= \begin{bmatrix} m_4 & 0 & 0 & 0 \\ 0 & m_3 & 0 & 0 \\ 0 & 0 & m_2 & 0 \\ 0 & 0 & 0 & m_1 \end{bmatrix} & F &= \begin{bmatrix} -m_4 \\ -m_3 \\ -m_2 \\ -m_1 \end{bmatrix} \\
C_t &= \begin{bmatrix} c_{4,t} & -c_{4,t} & 0 & 0 \\ -c_{4,t} & c_{4,t} + c_{3,t} & -c_{3,t} & 0 \\ 0 & -c_{3,t} & c_{3,t} + c_{2,t} & -c_{2,t} \\ 0 & 0 & -c_{2,t} & c_{2,t} + c_{1,t} \end{bmatrix} & K_t &= \begin{bmatrix} k_{4,t} & -k_{4,t} & 0 & 0 \\ -k_{4,t} & k_{4,t} + k_{3,t} & -k_{3,t} & 0 \\ 0 & -k_{3,t} & k_{3,t} + k_{2,t} & -k_{2,t} \\ 0 & 0 & -k_{2,t} & k_{2,t} + k_{1,t} \end{bmatrix} \quad (45)
\end{aligned}$$

$x_{i,t}$  is the displacement of the  $(i+1)$ -th floor (fifth floor is the roof) relative to the ground at time  $t$ ;  $u_t$  is the acceleration at the base (first floor of the building); and  $c_{i,t}$  and  $k_{i,t}$  are the inter-story damping coefficient and stiffness of the  $i$ -th story, respectively, at time  $t$ .

We generate the data using white-noise for the excitation  $u_t$ . The observed data  $\hat{y}_t$  is absolute acceleration time histories at the four stories:

$$y_t = \begin{bmatrix} \ddot{x}_{1,t} + u_t \\ \ddot{x}_{2,t} + u_t \\ \ddot{x}_{3,t} + u_t \\ \ddot{x}_{4,t} + u_t \end{bmatrix} + \Gamma \cdot v_t = -M^{-1}[C_t \dot{x}_t + K_t x_t] + \Gamma \cdot v_t \quad (46)$$

where  $v_t \in R^4 \sim N(0, I)$  are the (stationary) measurement uncertainties for  $y_t$ ;  $\Gamma = \text{diag}(\gamma_1, \dots, \gamma_4)$  is such that the overall signal/noise rms amplitude ratios for each channel is roughly equal to 10. Both the excitation and observation are sampled at a sampling interval of 0.02 second and are shown in Figure 2. With the excitation and observation of the system, i.e.  $\{u_t : t = 1, \dots, T\}$  and  $\{\hat{y}_t : t = 1, \dots, T\}$ , the goal is to estimate the system states (displacements and velocities) as well as the system parameters (dampings, stiffnesses and the uncertainty parameters) in real time.

### Identification model

Now we describe the identification model. We use the following time-varying linear state-space model as our identification model:

$$\frac{d}{dt} \begin{bmatrix} x_t \\ \dot{x}_t \\ \theta_t \end{bmatrix} = \begin{bmatrix} \dot{x}_t \\ -M^{-1}K_t x_t - M^{-1}C_t \dot{x}_t \\ 0 \end{bmatrix} + \begin{bmatrix} 0 \\ M^{-1}F \\ 0 \end{bmatrix} \cdot u_t + \begin{bmatrix} 0 \\ 0 \\ G \end{bmatrix} \cdot w_t \quad (47)$$

$$y_t = -M^{-1}K_t x_t - M^{-1}C_t \dot{x}_t + H_t \cdot v_t$$

where  $w_t \sim N(0, I)$ ;  $v_t \sim N(0, I)$ ;  $G \in R^{12 \times 12}$  is a diagonal matrix whose diagonals we have to specify (the reason that we have to specify it will be explained later);  $H_t = \text{diag}(h_{1,t}, h_{2,t}, h_{3,t}, h_{4,t})$  where the diagonals are unknown parameters (i.e. the four uncertainty parameters); and  $\theta_t \in R^{12}$  is the vector containing system parameters (including four stiffnesses, four dampings and four uncertainty parameters). The dimension of the state of the identification model is twenty (four displacements, four velocities, four stiffnesses, four dampings and four uncertainty parameters) although the dimension of the state in (44) is only eight.

To complete the probabilistic identification model, we must also specify the prior PDF for the entire (augmented) state trajectory  $\left\{ \begin{bmatrix} x_t^T & \dot{x}_t^T & \theta_t^T \end{bmatrix}^T : t = 0, \dots, T \right\}$ . More specifically, for the model in (47), we must specify the following: the prior PDF of  $x_0$  and  $\dot{x}_0$ , the prior PDF of  $\theta_0$ , and the diagonals of the  $G$  matrix. Note that the identification model in (47) uses a Brownian-motion prior PDF for the parameter evolution  $\{\theta_t : t = 0, \dots, T\}$  due to the following dynamical equation in (47):

$$\dot{\theta}_t = G \cdot w_t \quad (48)$$

A diagonal  $G$  matrix means that all system parameters are known a priori to drift independently. The effect of the  $G$  matrix is similar to the forgetting factor often used in adaptive filtering (Ljung and Gunnarsson 1990). When the entries of  $G$  are large, the system parameters are allowed to drift more freely, relaxing the dependency between parameter values of adjacent time steps; therefore, the identified parameters only reflect most recent data. The converse is true when the entries of  $G$  are small; in this case, the identified parameters can reflect data from the remote past.

In this example, the prior PDF for  $x_0$  and  $\dot{x}_0$  is taken to be zero-mean Gaussian with large variances; the prior PDF of  $\theta_0$  is taken to be Gaussian with mean equal to the actual value of  $\theta_0$  and large variances; the diagonal entries of  $G$  are chosen such that in each time step, each parameter drifts with a coefficient of variation (c.o.v., defined by the standard deviation divided by the mean value) equal to 2%. Recall that for  $k_1, k_3, k_4, c_1, c_2, c_3$  and  $c_4$ , their actual evolutions (see Figure 1) are Brownian motions with the same 2% drift c.o.v., i.e. there is no modeling error for the evolutions of  $k_1, k_3, k_4, c_1, c_2, c_3$  and  $c_4$ . But for  $k_2$  and the four uncertainty parameters, the actual evolutions are not Brownian motions (the actual evolution of  $k_2$  is shown in Figure 1; the four uncertainty parameters are actually constant), while the identification model uses a Brownian-motion prior on their evolutions.

Before we can proceed, we first convert (47) to the following discrete-time system using numerical integration (integrate over time step):

$$\begin{bmatrix} x_k \\ \dot{x}_k \\ \theta_k \end{bmatrix} = f_{k-1} \left( \begin{bmatrix} x_{k-1} \\ \dot{x}_{k-1} \\ \theta_{k-1} \end{bmatrix}, u_{k-1}, w_{k-1} \right) \quad y_k = h_k \left( \begin{bmatrix} x_k \\ \dot{x}_k \\ \theta_k \end{bmatrix}, u_k, v_k \right) \quad (49)$$

where  $f_{k-1}$  is evaluated using *Matlab* command ODE23. With the above discrete-time identification model, EKF can be conducted. Since EKF is not suitable for directly tracking the uncertainty parameters (i.e. matrix  $H$  in Appendix I), we implement Algorithm A.2 with (64) and  $\gamma = 0.95$  for the estimation of the uncertainty parameters. For PF, the uncertainty parameters are not separately treated because they are included in the augmented state vector in (47).

### Results

The stiffness, damping and uncertainty parameter estimates, and the associated 95% confidence intervals from EKF and PF (with  $N = 200$  samples for each of  $L = 10$  parallel PF and the importance weight c.o.v. threshold = 200% using Algorithms 4.3 with the local random walk step described in Ching *et al.* (2004)) are shown in Figures 3-8 (there is no confidence interval available for the EKF uncertainty parameter estimates). For this example, using more samples in PF than  $N \cdot L = 2000$  gives little improvement in the convergence of PF, indicating that the results are close to convergence. We treat the results from PF as a comparison standard since it asymptotically gives consistent estimates for the conditional means and variances.

In Figures 3-8, the thick lines indicate the actual parameter evolutions while the thin dashed lines are the conditional means of the identified system parameters and the thin dotted lines indicate the 95% confidence intervals. The results from EKF are similar to those of PF. Both algorithms successfully track the system parameters; for most parameters, the actual parameter evolutions lie within the 95% confidence bounds. Notice that although the Brownian motion prior for  $k_2$  and the uncertainty parameters does not exactly match their actual evolutions, both Bayesian algorithms can still appropriately track  $k_2$  and the uncertainty parameters. Compared to the accuracy of the stiffnesses, the estimates of the damping and uncertainty parameters are worse and the associated uncertainties are larger. Although EKF and PF perform

roughly equally in this example, there is a noticeable difference in the variances of the identified damping from PF, which are slightly larger than those from EKF.

## 5.2 Nonlinear hysteretic damping system with unknown system parameters

### *Data generation*

The previous example is a time-varying linear system. In the current example, we consider a time-varying nonlinear system consisting of a single DOF (SDOF) Bouc-Wen hysteretic damping system (Wen 1980). The purpose of this example is to compare the performances of different methods for tracking the state and unknown parameters of a nonlinear system. The system that generates the data can be described by the following governing equation:

$$\frac{d}{dt} \begin{bmatrix} x_t \\ \dot{x}_t \\ r_t \end{bmatrix} = \begin{bmatrix} \dot{x}_t \\ -1/m \cdot r_t + 1/m \cdot u_t \\ \theta_{1,t} \cdot \dot{x}_t - \theta_{2,t} \cdot |\dot{x}_t| |r_t|^{\theta_{3,t}-1} r_t + \theta_{3,t} \cdot \dot{x}_t |r_t|^{\theta_{4,t}} \end{bmatrix} \quad (50)$$

$$y_t = -1/m \cdot r_t + 1/m \cdot u_t + v_t$$

where  $r_t$  is the restoring force of the SDOF system;  $m$  is the mass, which is set to unity during the data generation;  $u_t$  is a white-noise excitation force on the mass;  $y_t$  is the acceleration measured on the mass;  $v_t$  is stationary such that the overall signal/noise amplitude ratio is 10;  $\theta_{1,t}, \theta_{2,t}, \theta_{3,t}, \theta_{4,t}$  are time-varying system parameters (their actual fluctuations are shown in Figures 10-11, and they are Brownian motions with drift c.o.v. equal to 2% during each sampling interval):  $\theta_{1,t}$  is the stiffness,  $\theta_{2,t}$ ,  $\theta_{3,t}$  and  $\theta_{4,t}$  are parameters that fine tune the shape of the hysteretic loop. Note that Bouc-Wen hysteretic damping system is Markovian in the sense that we can define a system state such that the current system status is completely characterized by the

state. Both the excitation  $u_t$  and observation  $\hat{y}_t$  (shown in Figure 9) are sampled at a sampling interval of 0.5 second (roughly five sample points per oscillation cycle of the system).

### Identification model

Given the data  $u_t$  and  $\hat{y}_t$ , the goal is to estimate the means and variances of the identified system state and system parameters using the following identification model:

$$\frac{d}{dt} \begin{bmatrix} x_t \\ \dot{x}_t \\ r_t \\ \theta_{1,t} \\ \theta_{2,t} \\ \theta_{3,t} \\ \theta_{4,t} \\ h_t \end{bmatrix} = \begin{bmatrix} \dot{x}_t \\ -1/m \cdot r_t + 1/m \cdot u_t \\ \theta_{1,t} \cdot \dot{x}_t - \theta_{2,t} \cdot |\dot{x}_t| |r_t|^{\theta_{4,t}-1} r_t + \theta_{3,t} \cdot \dot{x}_t |r_t|^{\theta_{4,t}} \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix} + \begin{bmatrix} 0 \\ 0 \\ 0 \\ G \\ 0 \end{bmatrix} \cdot w_t \quad (51)$$

$$y_t = -1/m \cdot r_t + 1/m \cdot u_t + h_t \cdot v_t$$

where  $w_t \in R^5 \sim N(0, I)$  ;  $v_t \in R \sim N(0, 1)$  ; we assume that  $m$  is known and so it is not considered as one of the uncertain parameters.

The prior PDF for  $x_0$  and  $\dot{x}_0$  is taken to be zero-mean Gaussian with large variances; the prior PDF of  $\theta_{1,0}, \theta_{2,0}, \theta_{3,0}, \theta_{4,0}, h_0$  is taken to be Gaussian with mean equal to their actual value at time zero and large variances; the  $G$  matrix in (51) is taken to be diagonal. The diagonals of  $G$  are chosen such that in each time step, each parameter is allowed to drift with a c.o.v. equal to 2%, i.e. no modeling error for the evolutions of  $\theta_{1,t}$ ,  $\theta_{2,t}$ ,  $\theta_{3,t}$  and  $\theta_{4,t}$ ; but modeling error exists for the evolution of  $h_t$  in (51) (the actual  $h_t$  is constant instead of a Brownian motion).

For EKF, we implement Algorithm A.2 with (64) and  $\gamma = 0.95$  for the estimation of the uncertainty parameters  $h_t$ . As before, for PF, there is no need for the uncertainty parameters to

be separately treated. The continuous-time model is numerically integrated to get the discrete-time version of this model similar to (49) with sampling rate equal to 0.5 second.

### *Results*

Figures 10-15 show the results of identification of EKF and PF (as before,  $N = 200$  and  $L = 10$  and the importance weight c.o.v. threshold = 200% using Algorithms 4.3 with the local random walk step). For this example, using more samples in PF than  $N \cdot L = 2000$  gives little improvement in the convergence of PF, indicating that the results are close to convergence. We treat the results from PF as a comparison standard. As before, the 95% confidence intervals on the parameters and states are indicated by thin dotted lines in Figures 10-15, except for Figure 15.

We find that EKF performs less effectively than PF: at some time instants, the EKF estimates of the stiffness parameter  $\theta_{1,t}$  oscillate around the actual evolution (Figure 10), while this is not seen for PF (Figure 11). Also, the EKF estimates for  $\theta_{2,t}$  (Figure 10) significantly deviate from those of PF (Figure 11). For the estimation of displacement, velocity and restoring force, the performances from the three methods are similar (Figures 12 and 13). However, PF estimates the uncertainty parameters  $h_i$  much better than EKF (Figures 15 and 16).

### **5.3 Lorenz chaotic system**

The Lorenz system is a chaotic system discovered by Lorenz (1963) when he solved a simplified Rayleigh-Bernard problem regarding two-dimensional fluid motion driven by buoyancy due to a temperature difference across its height. The resulting simplified set of differential equations that consider the first few modes of the system is

$$\begin{aligned}
\dot{x}_{1,t} &= -\sigma(x_{1,t} - x_{2,t}) \\
\dot{x}_{2,t} &= r \cdot x_{1,t} - x_{2,t} - x_{1,t}x_{3,t} \\
\dot{x}_{3,t} &= x_{1,t}x_{2,t} - bx_{3,t}
\end{aligned} \tag{52}$$

where  $x_{1,t}$  specifies the time evolution of the stream function of the first mode, whose contours are the streamlines;  $x_{2,t}$  and  $x_{3,t}$  specify the time evolutions of the temperature of the first two modes of the system; the parameter  $\sigma$  depends on the properties of the fluid (for water the value is typically between 1 and 4); the number  $b$  depends on the scales of the modes;  $r$  is the temperature difference: for small  $r$ , the system is asymptotically stable, i.e.  $x_{1,t \rightarrow \infty} = x_{2,t \rightarrow \infty} = x_{3,t \rightarrow \infty} = 0$ . For large  $r$ , chaos occurs with the so-called butterfly attractor where the ultimate fate of a trajectory of the system is to wander around two unstable equilibrium points and the trajectory is extremely sensitive to its initial condition.

### *Data generation*

In this example, the values of  $\sigma$ ,  $b$  and  $r$  are set to be 3, 1 and 26 ( $r$  is large so that the butterfly attractor occurs), and we observe  $x_{1,t}$  (contaminated by noise) with sampling interval of 0.5 second:

$$y_t = x_{1,t} + h \cdot v_t \tag{53}$$

where  $h$  is chosen such that the overall signal/noise ratio is 10. Figure 16 shows the observed value  $\hat{y}_t$ , which clearly shows that the trajectory of  $x_{1,t}$  switches several times between the two unstable equilibrium points at  $x_1 = -5$  and  $x_1 = 5$  (especially during 0-25 second).

### *Identification model*

The goal is to estimate the trajectory of the three system states based on  $\hat{y}_t$  using EKF and PF. When applying EKF and PF, we assume that we are very uncertain about the position of

the initial state (i.e. large variances for the prior PDF of the three states); we also assume that  $\sigma$ ,  $b$ ,  $r$  and  $h$  are known, and their actual values are used during the identification. Equations (52) and (53) are directly used in the identification model.

### *Results*

Figures 17-18 show the estimates made by EKF and PF (with  $N = 200$  and  $L = 10$  and the importance weight c.o.v. threshold = 200% using Algorithms 4.3 with the local random walk step). For this example, using  $N \cdot L = 2000$  samples in PF is found to be sufficient for convergence, so once again, we treat the results from PF as a comparison standard. Also, as before, the 95% confidence intervals on the states are indicated by thin dotted lines in Figures 17-18.

It is clear that PF can successfully track all three system states, while EKF can only reliably track the observed state variable  $x_{1,t}$  (since  $\hat{y}_t$  directly measures  $x_{1,t}$ , it is possible that an inappropriate filtering algorithm can still track  $x_{1,t}$  perfectly). EKF can track some parts of  $x_{2,t}$  and  $x_{3,t}$  (but performs poorly in other parts, especially in the beginning portions of  $x_{2,t}$  and  $x_{3,t}$  where the system switches between the two equilibrium points).

## **6 CONCLUSION**

We have presented two Bayesian state-estimation algorithms in detail, including the older extended Kalman filter (EKF) and the newer particle filter (PF), which is a stochastic simulation approach. Their performance is examined using three numerical examples, which show that PF is the best one to use, while EKF can sometimes create misleading results; the examples represent three different classes of dynamical systems: a linear model with time-varying system parameters (Section 5.1), the nonlinear hysteretic model (Section 5.2) that can be considered to give moderately nonlinear behavior and the Lorenz chaotic model (Section 5.3) that gives highly nonlinear behavior.

We believe that PF has performed satisfactorily for all examples since it always provides estimates for the system state and unknown parameters with associated confidence intervals that are consistent with their actual values. In theory, PF should provide estimates that asymptotically converge to the expected values. It turns out that EKF can only track the system state and unknown parameters for the first example, its performance for the second example is only fair, and it performs poorly for the Lorenz chaotic example. This is consistent with the expectation that EKF is not suitable for highly nonlinear models.

## REFERENCES

- Alspach, D.L. and Sorenson, H.W. (1972). Nonlinear Bayesian estimation using Gaussian sum approximations, *IEEE Transactions on Automatic Control*, **17**(4), 439-448.
- Andrieu, C., de Freitas, J. F. G. and Doucet, A. (1999). Sequential MCMC for Bayesian model selection, *IEEE Higher Order Statistics Workshop*, Ceasarea, Israel, 130-134.
- Beck, J. L. (1978). *Determining Models of Structures from Earthquake Records*, EERL Report 78-01, Earthquake Engineering Research Laboratory, California Institute of Technology, Pasadena, California. <<http://resolver.caltech.edu/CaltechEERL:1978.EERL-78-01>>
- Beck, J.L. and Yuen, K.-V. (2004). Model selection using response measurements: Bayesian probabilistic approach, *Journal of Engineering Mechanics*, **130**(2), 192-203.
- Ching, J., Beck, J.L., Porter, K.A., and Shaikhutdinov, R. (2004) *Real-time Bayesian State Estimation of Uncertain Dynamical Systems*. EERL Report 2004-01, California Institute of Technology, Pasadena, CA. <<http://caltecheerl.library.caltech.edu/>>
- Doucet, A. and Godsill, S. (1998). *On Sequential Simulation-based Methods for Bayesian Filtering*, Report CUED/F-INFENG/TR310, Department of Engineering, Cambridge University.
- Doucet, A. and Andrieu, C. (2000). *Particle Filtering for Partially Observed Gaussian State Space Models*, Report CUED/F-INFENG /TR393, Department of Engineering, Cambridge University.

- Doucet, A., de Freitas, J.F.G. and Gordon, N. (2000). Introduction to sequential Monte Carlo methods, in *Sequential Monte Carlo Methods in Practice*, Doucet, A., de Freitas, J.F.G., and Gordon, N.J. (eds), Springer-Verlag.
- Ghanem, R. and Shinozuka, M. (1995). Structural-system identification. I: Theory, *Journal of Engineering Mechanics*, **121**(2), 255-264.
- Glaser, S.D. (1996). Insight into liquefaction by system identification, *Geotechnique*, 46(4), 641-655.
- Gordon, N.J., Salmond, D.J. and Smith, A.F.M. (1993). Novel approach to nonlinear/non-Gaussian Bayesian state estimation, *IEEE Proceedings-F*, **140**(2): 107-113.
- Hoshiya, M and Saito, E. (1984). Structural identification by extended Kalman filter. *Journal of Engineering Mechanics*, **110**(12), p. 1757.
- Jaynes, E. T. (1957). Information theory and statistical mechanics, *Physical Review*, **106**, 620-630. Also in Rosenkrantz, R. D. (Ed) (1982). *E. T. Jaynes, Papers on Probability, Statistics and Statistical Physics*, Dordrecht, Holland: D. Reidel Publishing Co.
- Jazwinski, A.H. (1970). *Stochastic Processes and Filtering Theory*, Academic Press, New York.
- Julier, S.J., Uhlmann, J.K., and Durrant-Whyte, H.F. (2000). A new method for the nonlinear transformation of means and covariances in filters and estimators, *IEEE Transactions on Automatic Control*, **45**(3), 477-482.
- Kalman, R.E. (1960). A new approach to linear filtering and prediction problems, *J. Basic Engr.*, **82D**, 35-45.
- Kalman, R.E. and Bucy, R.S. (1961). New results in linear filtering and prediction problems, *J. Basic Engr.*, **83D**.
- Kitagawa, G. (1996). Monte Carlo filter and smoother for non-Gaussian nonlinear state space models, *Journal of Computational and Graphical Statistics*, **5**, 1-25.
- Koh, C. G. and See, L. M. (1994). Identification and uncertainty estimation of structural parameters. *J. Engng. Mech.*, **120**(6), p. 1219.

- Lin, C.C., Soong, T.T., and Natke, H.G. (1990). Real-time system identification of degrading structures, *Journal of Engineering Mechanics*, **116**(10), 2258-2274.
- Liu, J.S. and Chen, R. (1998). Sequential Monte Carlo methods for dynamical systems, *Journal of the American Statistical Association*, **93**, 1032-1044.
- Ljung, L. and Gunnarsson, S. (1990). Adaptation and tracking in system identification – a survey, *Automatica*, **26**(1), 7-21.
- Lorenz, E.N. (1963). Deterministic nonperiodic flow, *Journal of Atmospheric Sciences*, **20**, 130-141.
- Maruyama, O. and Hoshiya, M. (2003). Nonlinear filters using Monte Carlo integration for conditional random fields, *Proceedings of the 9<sup>th</sup> International Conference on Applications of Statistics and Probability in Civil Engineering*, San Francisco, July.
- Sato, T. and Qi, K. (1998). Adaptive  $H_\infty$  filter: Its application to structural identification, *Journal of Engineering Mechanics*, **124**(11), 1233-1240.
- Shinozuka, M. and Ghanem, R. (1995). Structural-system identification. II: Experimental verification, *Journal of Engineering Mechanics*, **121**(2), 265-273.
- Shumway, R. H. and Stoffer, D. S. (1982). An approach to time series smoothing and forecasting using the EM algorithm, *Journal of Time Series Analysis*, **4**, 253-263.
- Smyth, A.W., Masri, S.F., Chassiakos, A.G., and Caughey, T.K. (1999). On-line parametric identification of MDOF nonlinear hysteretic systems, *Journal of Engineering Mechanics*, **125**(2), 133-142.
- van der Merwe, R. and Wan, E.A. (2003). Gaussian mixture sigma-point particle filters for sequential probabilistic inference in dynamic state-space models, *Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Hong Kong.
- van der Merwe, R., de Freitas, N., Doucet, A. and Wan, E.A. (2000). *The Unscented Particle Filter*, Report CUED/F-INFENG/TR380, Department of Engineering, Cambridge University.
- Wan, E.A. and van der Merwe, R. (2000). The unscented Kalman filter for nonlinear estimation, *Proceedings of Symposium 2000 on Adaptive Systems for Signal Processing, Communication and Control*, IEEE, Lake Louise, Alberta, Canada.

- Wen, Y. K. (1980). Equivalent linearization for hysteretic systems under random excitations. *J. Appl. Mech.*, **47**(1), 150–154.
- Yoshida, I. and Sato, T. (2002). Health monitoring algorithm by the Monte Carlo filter based on non-Gaussian noise, *Journal of Natural Disaster Science*, **24**(2), 101-107.
- Yun, C.B. and Shinozuka, M. (1980). Identification of nonlinear dynamic systems, *Journal of Structural Engineering*, **8**(2), 187-203.

## APPENDICES

### Appendix I. Estimating uncertainty parameters in EKF

When the amplitudes of the uncertainty terms are unknown, it is important to estimate them. This can usually be done by parameterizing the uncertainty amplitudes using some uncertainty parameters, e.g. the covariance matrices of the uncertainties. Uncertainty parameters can be augmented into system states along with other system parameters and estimated using EKF. Estimating uncertainty parameters is essential since  $p(x_k | \hat{Y}_k)$  usually strongly depends on them. Throughout this section, we consider the following state transition and observation equations:

$$x_k = f_{k-1}(x_{k-1}, u_{k-1}, w_{k-1}) \quad y_k = h_k(x_k, u_k) + H_k(\rho_k) \cdot v_k \quad (54)$$

where  $\rho_k$  parameterizes the covariance matrix of  $H_k v_k$ , denoted by  $\Lambda_k$  (i.e.  $\Lambda_k(\rho_k) = H_k(\rho_k) H_k^T(\rho_k)$ ). We focus on the case that the joint PDF of  $\{w_k : k = 1, \dots, T\}$  is known, while  $\Lambda_k(\rho_k)$  matrix is unknown since this is the way we model the uncertainties for the numerical examples in Section 5. In practice, we cannot assume that both the PDFs of the process and measurement uncertainty terms to be unknown. It is impossible to estimate both of them merely using the observations  $\hat{Y}_k$ .

We show that  $\{\rho_k : k = 1, \dots, T\}$  is not identifiable using EKF, as follows. Suppose we would like to estimate the state and  $\rho_k$  at the same time by augmenting the state vector into  $x_k^a = \begin{bmatrix} x_k^T & \rho_k^T \end{bmatrix}^T$  and conducting EKF. Notice that in the estimation step for EKF (see (10)), the main driving force for updating the state from  $x_{k|k-1}$  to  $x_{k|k}$  is the prediction error  $\hat{y}_k - y_{k|k-1}$ . An element of the state vector will be effectively updated if  $y_{k|k-1}$  is sensitive to the change of that element, and vice versa. For the system in (54),  $y_{k|k-1}$  obtained from EKF is insensitive to  $\rho_k$ . Therefore, the uncertainty parameters  $\{\rho_k : k = 1, \dots, T\}$  are not identifiable using EKF.

The identifiability issue for EKF can be handled by an Expectation-Maximization (EM) algorithm (Shumway and Stoffer 1982), described as follows. First, consider a more restrictive uncertainty model:  $\rho_1 = \dots = \rho_T = \rho$  (i.e. uncertainty amplitudes are time-invariant). We will return to the general time-varying case later. The following EM algorithm can be used to estimate  $\rho$ :

**Algorithm A.1: EM algorithm**

(1) Initialize  $\rho^{(0)}$ .

(2) At the  $r$ -th iteration:

$$\text{(E-step) Evaluate } L^{(r)}(\rho) \equiv E \left[ \log p \left( \hat{Y}_k, X_k \mid \rho \right) \middle| \hat{Y}_k, \rho^{(r)} \right] \quad (55)$$

$$\text{(M-step) Compute } \left\{ \rho^{(r+1)} \right\} = \arg \max_{\rho} L^{(r)}(\rho) \quad (56)$$

(3) Go back to step (2) for the  $(r+1)$ -th iteration and continue cycling until  $\rho^{(r)}$  converge.

The term  $\log p \left( \hat{Y}_k, X_k \mid \rho \right)$  in (55) has the following expression:

$$\begin{aligned}
& \log p(\hat{Y}_k, X_k | \rho) \\
&= \text{const} + \sum_{t=1}^k \left( -\frac{1}{2} \log(\det[\Lambda_t(\rho)]) - \frac{1}{2} [\hat{y}_t - h_t(x_t, u_t)]^T [\Lambda_t(\rho)]^{-1} [\hat{y}_t - h_t(x_t, u_t)] \right)
\end{aligned} \tag{57}$$

where *const* denotes a quantity that does not depend on  $\rho$ . Therefore,

$$\begin{aligned}
L^{(r)}(\rho) &= E \left[ \log p(\hat{Y}_k, X_k | \rho) \middle| \hat{Y}_k, \rho^{(r)} \right] \\
&= \text{const} + \sum_{t=1}^k \left( -\frac{1}{2} \log(\det[\Lambda_t(\rho)]) \right) \\
&+ \sum_{t=1}^k E \left( -\frac{1}{2} [\hat{y}_t - h_t(x_t, u_t)]^T [\Lambda_t(\rho)]^{-1} [\hat{y}_t - h_t(x_t, u_t)] \middle| \hat{Y}_k, \rho^{(r)} \right) \\
&= \text{const} + \sum_{t=1}^k \left( -\frac{1}{2} \log(\det[\Lambda_t(\rho)]) \right) \\
&+ \sum_{t=1}^k E \left( -\frac{1}{2} \text{TR} \left\{ [\Lambda_t(\rho)]^{-1} [\hat{y}_t - h_t(x_t, u_t)] \cdot [\hat{y}_t - h_t(x_t, u_t)]^T \right\} \middle| \hat{Y}_k, \rho^{(r)} \right) \\
&= \text{const} + \sum_{t=1}^k \left( -\frac{1}{2} \log(\det[\Lambda_t(\rho)]) \right) \\
&+ \sum_{t=1}^k E \left( -\frac{1}{2} \text{TR} \left\{ [\Lambda_t(\rho)]^{-1} [\hat{y}_t \cdot \hat{y}_t^T - 2h_t(x_t, u_t) \cdot \hat{y}_t^T + h_t(x_t, u_t) \cdot h_t(x_t, u_t)^T] \right\} \middle| \hat{Y}_k, \rho^{(r)} \right) \\
&= \text{const} + \sum_{t=1}^k \left( -\frac{1}{2} \log(\det[\Lambda_t(\rho)]) \right) \\
&+ \sum_{t=1}^k \left( -\frac{1}{2} \text{TR} \left\{ [\Lambda_t(\rho)]^{-1} [\hat{y}_t \cdot \hat{y}_t^T - 2h_{t|k}^{(r)} \cdot \hat{y}_t^T + P_{t|k}^{h,(r)} + h_{t|k}^{(r)} \cdot h_{t|k}^{(r)T}] \right\} \right)
\end{aligned} \tag{58}$$

where  $\text{TR}[\cdot]$  denotes matrix trace;  $h_{t|k}^{(r)}$  and  $P_{t|k}^{h,(r)}$  denote  $E[h_t(x_t, u_t) | \hat{Y}_k, \rho^{(r)}]$  and

$\text{Cov}[h_t(x_t, u_t) | \hat{Y}_k, \rho^{(r)}]$ , respectively. Unfortunately, Algorithm A.1 cannot be implemented on-

line because evaluating  $h_{t|k}^{(r)}$  and  $P_{t|k}^{h,(r)}$  for  $t < k$  requires an offline methodology. Now let us

consider an on-line version of  $L^{(r)}(\rho)$ , denoted by  $l^{(r)}(\rho)$ :

$$\begin{aligned}
l^{(r)}(\rho) &= \text{const} + \sum_{t=1}^k \left( -\frac{1}{2} \log(\det[\Lambda_t(\rho)]) \right) \\
&+ \sum_{t=1}^k \left( -\frac{1}{2} \text{TR} \left\{ [\Lambda_t(\rho)]^{-1} [\hat{y}_t \cdot \hat{y}_t^T - 2h_{t|t}^{(r)} \cdot \hat{y}_t^T + P_{t|t}^{h,(r)} + h_{t|t}^{(r)} \cdot h_{t|t}^{(r)T}] \right\} \right)
\end{aligned} \tag{59}$$

One can verify that  $E[L^{(r)}(\rho)] = E[l^{(r)}(\rho)]$ , i.e. they have the same expected value. In Algorithm A.1, instead of evaluating and maximizing  $L^{(r)}(\rho)$  in the E-step and M-step, we evaluate and maximize  $l^{(r)}(\rho)$ , and the resulting algorithm is on-line.

Now let us consider a general uncertainty model:  $\rho_k$  varies with time. So we have

$$l^{(r)}(\rho_1 \cdots \rho_k) = \text{const} + \sum_{t=1}^k \left( -\frac{1}{2} \log(\det[\Lambda_t(\rho_t)]) \right) + \sum_{t=1}^k \left( -\frac{1}{2} \text{TR} \left\{ [\Lambda_t(\rho_t)]^{-1} \left[ \hat{y}_t \cdot \hat{y}_t^T - 2h_{t|t}^{(r)} \cdot \hat{y}_t^T + P_{t|t}^{h,(r)} + h_{t|t}^{(r)} \cdot h_{t|t}^{(r)T} \right] \right\} \right) \quad (60)$$

For an on-line methodology, at time instant  $k$  we are only interested in estimating  $\rho_k$ . In the case where the parameterization of  $\Lambda_k$  is such that it is a full matrix without any specific structure (i.e.  $\rho_k$  contains all of the entries in the upper triangle region of  $\Lambda_k$ ), maximizing  $l^{(r)}(\rho_1 \cdots \rho_k)$  with respect to  $\rho_k$  yields

$$\Lambda_k(\rho_k^{(r+1)}) = \hat{y}_k \cdot \hat{y}_k^T - 2h_{k|k}^{(r)} \cdot \hat{y}_k^T + P_{k|k}^{h,(r)} + h_{k|k}^{(r)} \cdot h_{k|k}^{(r)T} \quad (61)$$

In the case where  $\Lambda_k$  is constrained to be diagonal (i.e.  $\rho_k$  contains all the diagonal entries of  $\Lambda_k$ ), maximizing  $l^{(r)}(\rho_1 \cdots \rho_k)$  with respect to  $\rho_k$  yields

$$\Lambda_k(\rho_k^{(r+1)}) = \text{diag} \left( \hat{y}_k \cdot \hat{y}_k^T - 2h_{k|k}^{(r)} \cdot \hat{y}_k^T + P_{k|k}^{h,(r)} + h_{k|k}^{(r)} \cdot h_{k|k}^{(r)T} \right) \quad (62)$$

where  $\text{diag}(A)$  is a diagonal matrix whose diagonal entries are identical to those of  $A$ .

The EKF algorithm with improved ability of estimating uncertainty parameters is as follows:

**Algorithm A.2: EM algorithm for estimating uncertainty parameters in EKF**

- (1) Initialize EKF (i.e. specify  $x_{0|0}$  and  $P_{0|0}$ , and so on).

(2) At the beginning of time step  $k$ , assign a reasonable value for  $\rho_k^{(0)}$ . Then do the iterations in Algorithm A.1:

- a. Evaluate  $h_{k|k}^{(r)} = E[h_k(x_k, u_k) | \hat{Y}_k, \rho_1^c \cdots \rho_{k-1}^c, \rho_k^{(r)}]$  and  $P_{k|k}^{h,(r)} = \text{Cov}[h_k(x_k, u_k) | \hat{Y}_k, \rho_1^c \cdots \rho_{k-1}^c, \rho_k^{(r)}]$  ( $\rho_i^c$  is the final estimate of  $\rho_i$ , see step c. below) using EKF.
- b. Maximize  $l^{(r)}(\rho_1 \cdots \rho_k)$  with respect to  $\rho_k$  to obtain  $\rho_k^{(r+1)}$  as well as  $\Lambda_k(\rho_k^{(r+1)})$  (e.g. (61) or (62)).
- c. Go back to step a. for the  $(r+1)$ -th iteration and continue cycling until  $\rho_k^{(r)}$  converges. Denote the converged estimate of  $\rho_k$  by  $\rho_k^c$ . The final estimate of  $x_{k|k}$  and  $P_{k|k}$  from EKF are  $E(x_k | \hat{Y}_k, \rho_1^c \cdots \rho_k^c)$  and  $\text{Cov}(x_k | \hat{Y}_k, \rho_1^c \cdots \rho_k^c)$ .

(3) Go back to step (2) for the  $(k+1)$ -th iteration.

A simple technique that incorporates a forgetting factor  $\gamma$  ( $0 \leq \gamma \leq 1$ ) can also implemented to control the degree of oscillation in  $\{\rho_k : k = 1, \dots, T\}$  through time. This is done by letting the actual estimate of  $\Lambda_k$  be the convex combination of  $\Lambda_{k-1}(\rho_{k-1}^c)$  and the new estimate of  $\Lambda_k$ . Using (61) and (62) as examples:

$$\Lambda_k(\rho_k^{(r+1)}) = \gamma \cdot \Lambda_{k-1}(\rho_{k-1}^c) + (1-\gamma) \cdot \left[ \hat{y}_k \cdot \hat{y}_k^T - 2h_{k|k}^{(r)} \cdot \hat{y}_k^T + P_{k|k}^{h,(r)} + h_{k|k}^{(r)} \cdot h_{k|k}^{(r)T} \right] \quad (63)$$

for (61), and

$$\Lambda_k(\rho_k^{(r+1)}) = \gamma \cdot \Lambda_{k-1}(\rho_{k-1}^c) + (1-\gamma) \cdot \text{diag} \left[ \hat{y}_k \cdot \hat{y}_k^T - 2h_{k|k}^{(r)} \cdot \hat{y}_k^T + P_{k|k}^{h,(r)} + h_{k|k}^{(r)} \cdot h_{k|k}^{(r)T} \right] \quad (64)$$

for (62).

The identifiability issue for uncertainty parameters does not exist for PF due to the fact that stochastic simulation provides consistent estimates for any system parameters.

## Appendix II. Choice of a good importance sampling PDF

In this appendix, we discuss the selection of the importance sampling PDF admitting a Markovian sampling procedure using the following two examples:

$$q_1(X_k | \hat{Y}_k) = p(x_0) \cdot \prod_{m=1}^k p(x_m | x_{m-1}) \quad (65)$$

and

$$q_2(X_k | \hat{Y}_k) = p(x_0) \cdot \prod_{m=1}^k p(x_m | x_{m-1}, \hat{y}_m) \quad (66)$$

Their corresponding non-normalized importance weights are

$$w_k^{(1)} \equiv \frac{p(X_k | \hat{Y}_k)}{q_1(X_k | \hat{Y}_k)} = \prod_{m=1}^k \frac{p(\hat{y}_m | x_m)}{p(\hat{y}_m | \hat{Y}_{m-1})} = w_{k-1}^{(1)} \cdot \frac{p(\hat{y}_k | x_k)}{p(\hat{y}_k | \hat{Y}_{k-1})} \quad (67)$$

and

$$w_k^{(2)} \equiv \frac{p(X_k | \hat{Y}_k)}{q_2(X_k | \hat{Y}_k)} = \prod_{m=1}^k \frac{p(\hat{y}_m | x_m) \cdot p(x_m | x_{m-1})}{p(x_m | x_{m-1}, \hat{y}_m) \cdot p(\hat{y}_m | \hat{Y}_{m-1})} = w_{k-1}^{(2)} \cdot \frac{p(\hat{y}_k | x_k) \cdot p(x_k | x_{k-1})}{p(x_k | x_{k-1}, \hat{y}_k) \cdot p(\hat{y}_k | \hat{Y}_{k-1})} \quad (68)$$

respectively. Both choices result in inevitable degradation, i.e. the corresponding importance weights become highly non-uniform through time.

We use the following example to show that  $q_1(X_k | \hat{Y}_k)$  results in more severe degradation. Consider the following 1-D state-space system where  $w_k$  and  $v_k$  are zero-mean Gaussian variables with unit variance:

$$x_{k+1} = 0.5 \cdot x_k + u_k + 2 \cdot w_k \quad y_k = x_k + 0.01 \cdot v_k \quad (69)$$

Note that the observation  $\hat{y}_k$  is accurate because of the small uncertainty (0.01 standard deviation). For  $q_1(X_k | \hat{Y}_k)$ , at time  $k$ ,  $\hat{X}_{k-1}^i$  is updated by  $\hat{X}_k^i = \{\hat{X}_{k-1}^i, \hat{x}_k^i\}$ , where

$$\hat{x}_k^i \sim N(0.5 \cdot \hat{x}_{k-1}^i + u_{k-1}, 2^2) \quad (70)$$

which is a wide bandwidth Gaussian function. However, such  $\hat{x}_k^i$  is very likely to reside in a low likelihood region of  $p(\hat{y}_k | x_k)$ , i.e.  $p(\hat{y}_k | \hat{x}_k^i)$  is small, due to the fact that  $p(\hat{y}_k | x_k)$  is a sharp Gaussian likelihood function whose maximum is close to  $\hat{y}_k$  owing to the high observation accuracy. Nevertheless, as  $N$  grows, some samples of  $x_k$  will have chance to reside in the high likelihood (sharp) region of  $p(\hat{y}_k | x_k)$ ; therefore, the set  $\{p(\hat{y}_k | \hat{x}_k^i) : i = 1, \dots, N\}$  becomes quite non-uniform. Recall that the weight that associated with  $q_1(X_k | \hat{Y}_k)$  is

$$w_k^i = \prod_{m=1}^k p(\hat{y}_m | \hat{x}_m^i) \quad (71)$$

So the set  $\{w_k^i : i = 1, \dots, N\}$  becomes highly non-uniform, and this degradation gets aggravated as  $k$  grows. On the other hand, for  $q_2(X_k | \hat{Y}_k)$ , at time  $k$ ,  $\hat{X}_{k-1}^i$  is updated by  $\hat{X}_k^i = \{\hat{X}_{k-1}^i, \hat{x}_k^i\}$ ,

where

$$\hat{x}_k^i \sim p(x_k | \hat{x}_{k-1}^i, \hat{y}_k) \quad (72)$$

which is a PDF whose center is close to  $\hat{y}_k$ . Therefore, non-uniformity of  $\{w_k^i : i = 1, \dots, N\}$  is less severe.

Nevertheless,  $q_1(X_k | \hat{Y}_k)$  remains a popular importance sampling PDF (e.g. it was adopted in Kitagawa (1996)) due to the fact that if the uncertainty term  $w_{k-1}$  is additive, i.e.

$$x_k = f_{k-1}(x_{k-1}, u_{k-1}) + G_{k-1} \cdot w_{k-1} \quad (73)$$

then the sample at each time  $k$  is simply

$$\hat{x}_k^i \sim N[f_{k-1}(\hat{x}_{k-1}^i, u_{k-1}), G_{k-1} G_{k-1}^T] \quad (74)$$

which is easy to sample.

## FIGURE CAPTION

Figure 1. Time evolutions of the actual inter-story stiffnesses and dampings

Figure 2. The simulated excitation and observation data

Figure 3. The EKF estimates of inter-story stiffness

Figure 4. The PF estimates of inter-story stiffness

Figure 5. The EKF estimates of inter-story damping

Figure 6. The PF estimates of inter-story damping

Figure 7. The EKF estimates of uncertainty parameters  $h_{1,t}, h_{2,t}, h_{3,t}, h_{4,t}$

Figure 8. The PF estimates of uncertainty parameters  $h_{1,t}, h_{2,t}, h_{3,t}, h_{4,t}$

Figure 9. The excitation force  $u_t$  and the observed acceleration  $\hat{y}_t$

Figure 10. The EKF estimates of the system parameters

Figure 11. The PF estimates of the system parameters

Figure 12. The EKF estimates of the system states

Figure 13. The PF estimates of the system states

Figure 14. The EKF estimates of the uncertainty parameter

Figure 15. The PF estimates of the uncertainty parameter

Figure 16. The plot for  $\hat{y}_t$

Figure 17. The EKF estimates of the system states

Figure 18. The PF estimates of the system states

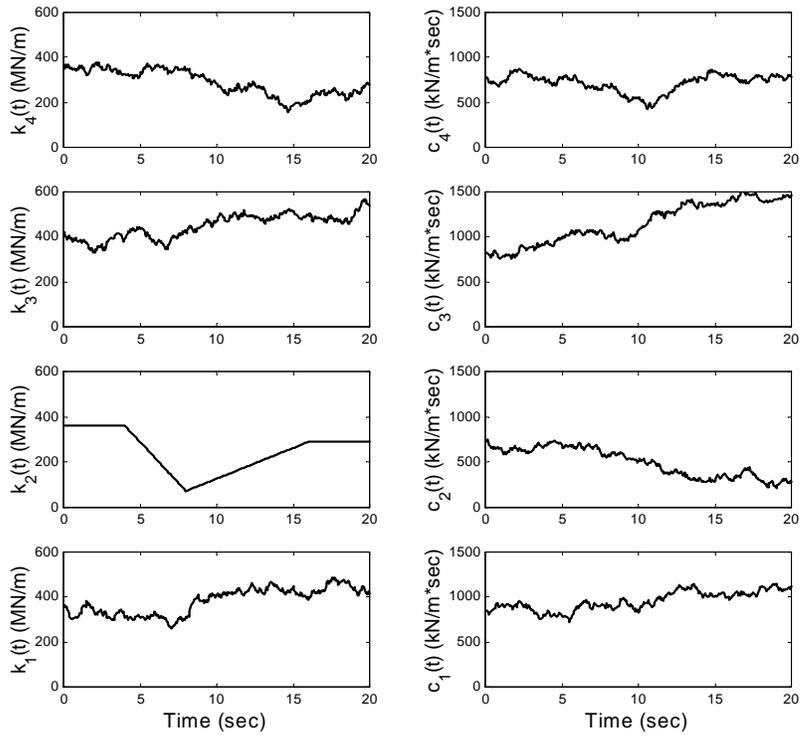


Figure 1. Time evolutions of the actual inter-story stiffnesses and dampings

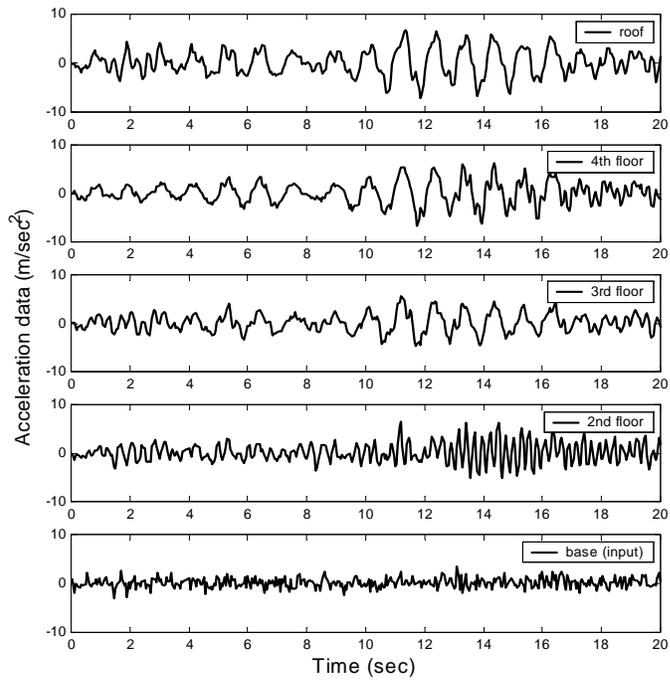


Figure 2. The simulated excitation and observation data

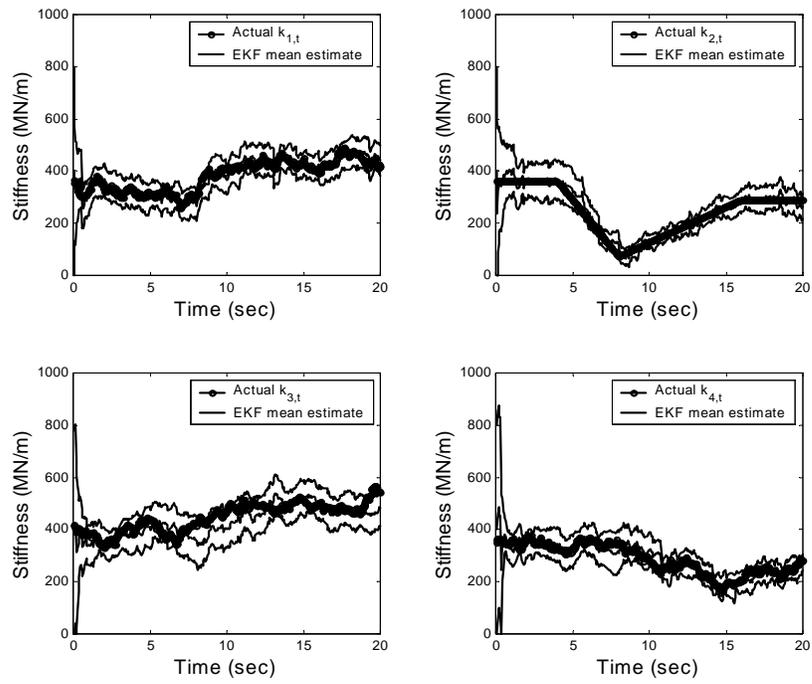


Figure 3. The EKF estimates of inter-story stiffness

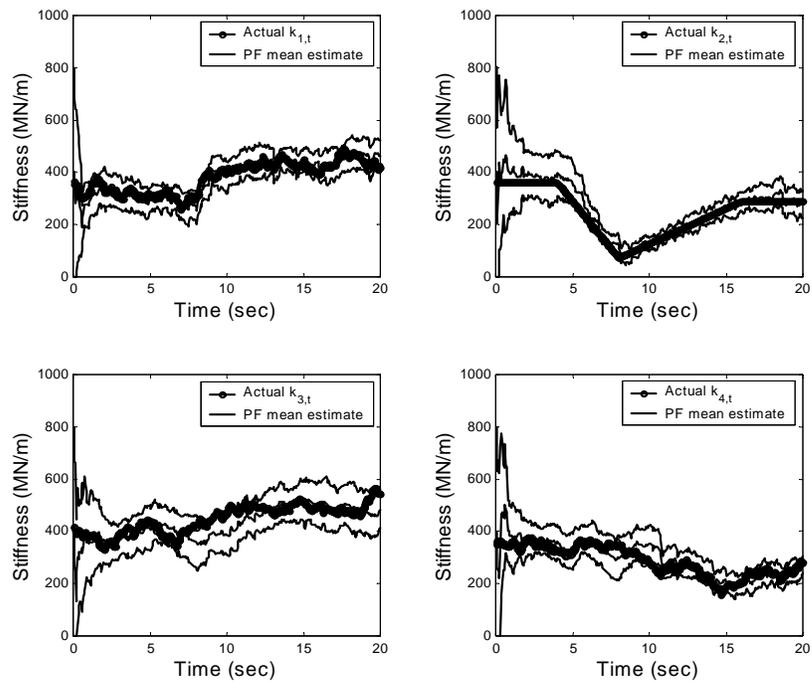


Figure 4. The PF estimates of inter-story stiffness

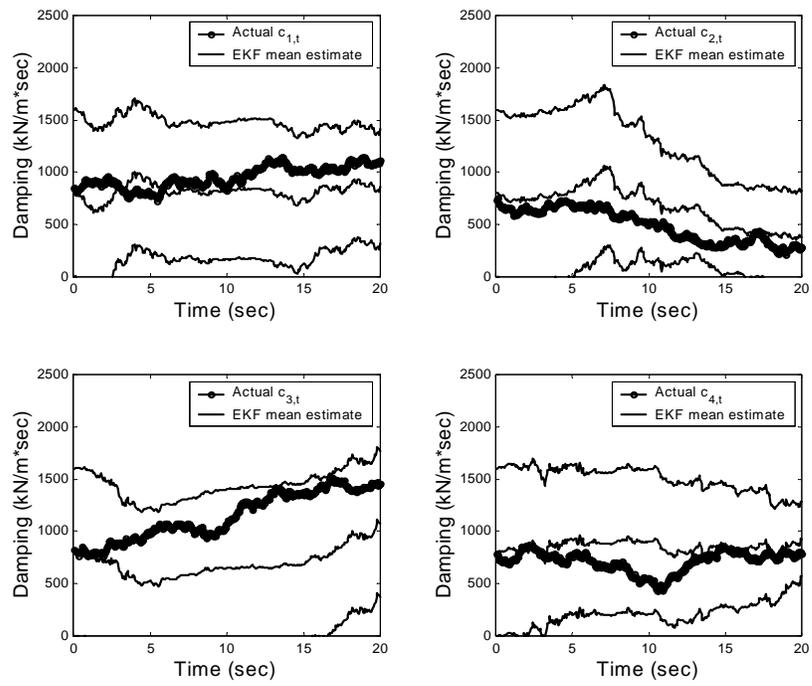


Figure 5. The EKF estimates of inter-story damping

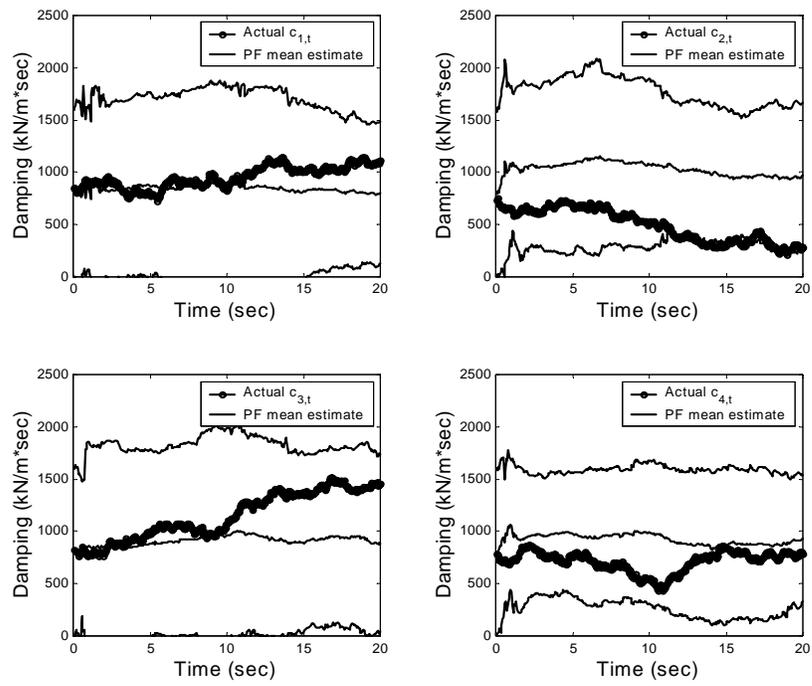


Figure 6. The PF estimates of inter-story damping

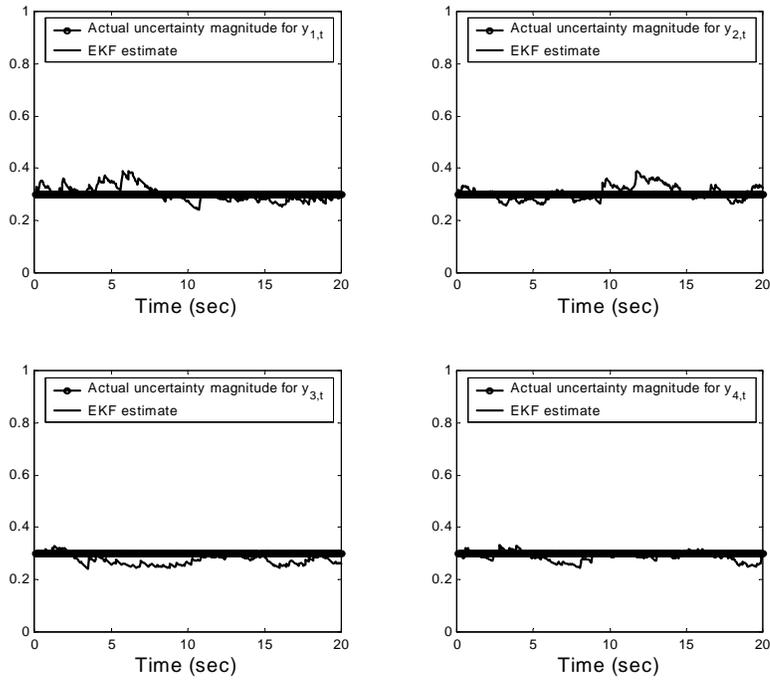


Figure 7. The EKF estimates of uncertainty parameters  $h_{1,t}, h_{2,t}, h_{3,t}, h_{4,t}$

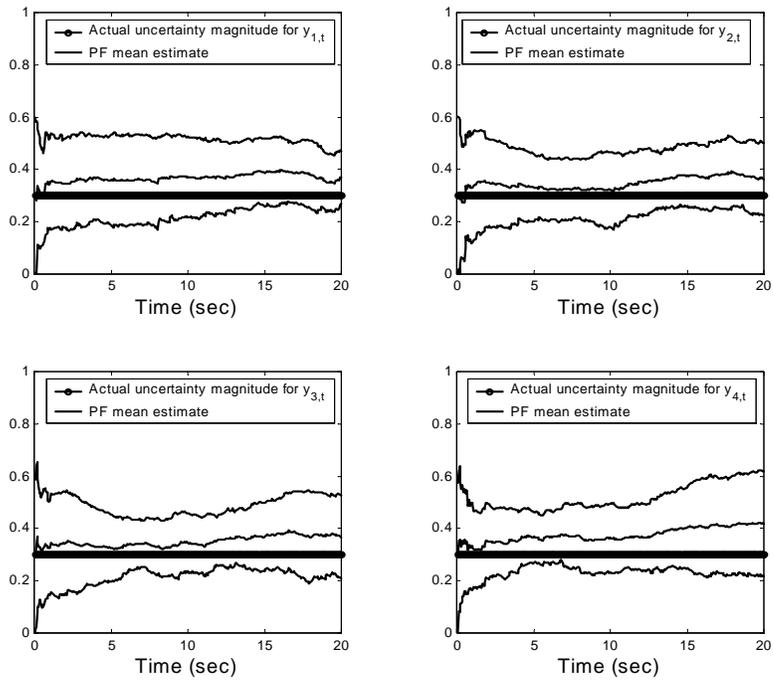


Figure 8. The PF estimates of uncertainty parameters  $h_{1,t}, h_{2,t}, h_{3,t}, h_{4,t}$

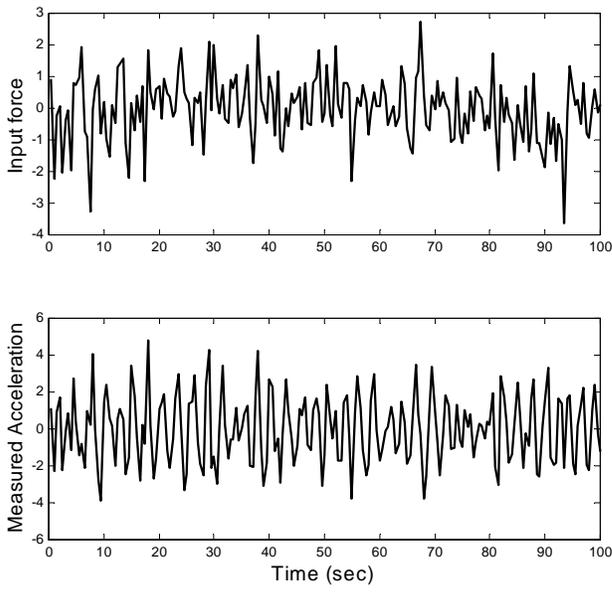


Figure 9. The excitation force  $u_t$  and the observed acceleration  $\hat{y}_t$

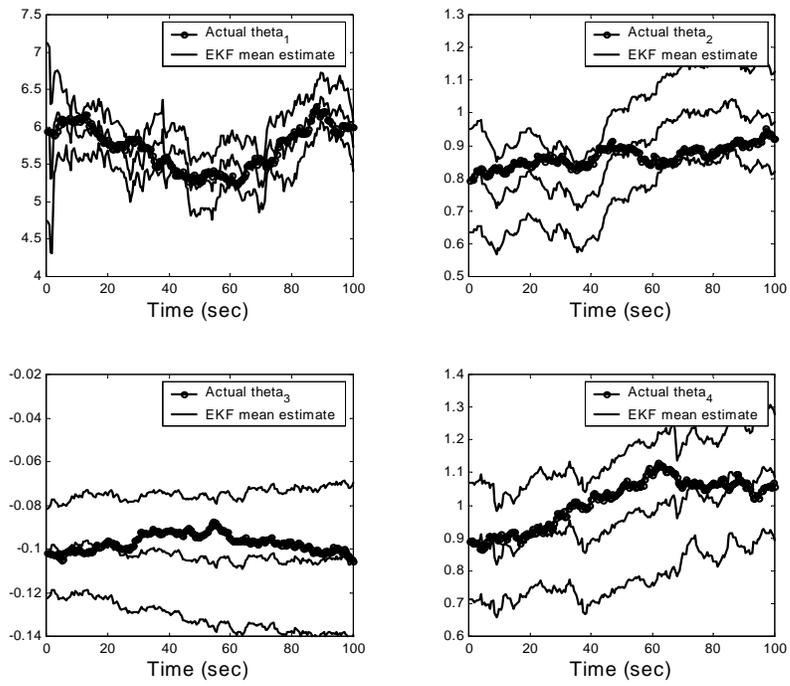


Figure 10. The EKF estimates of the system parameters

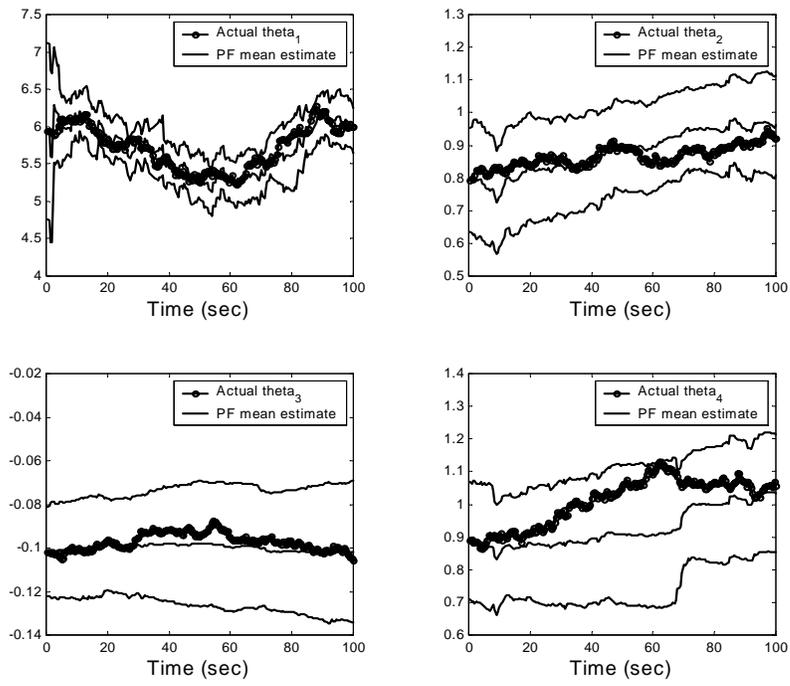


Figure 11. The PF estimates of the system parameters

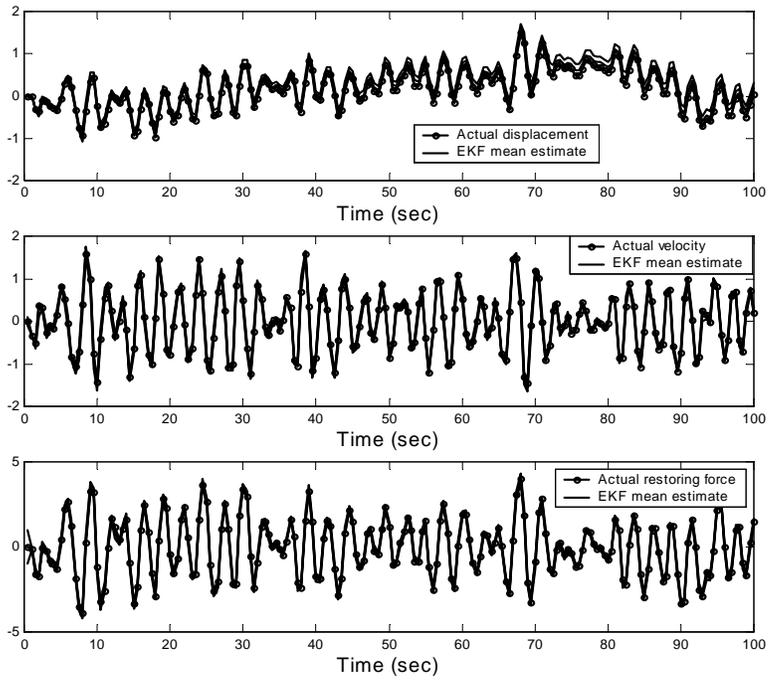


Figure 12. The EKF estimates of the system states

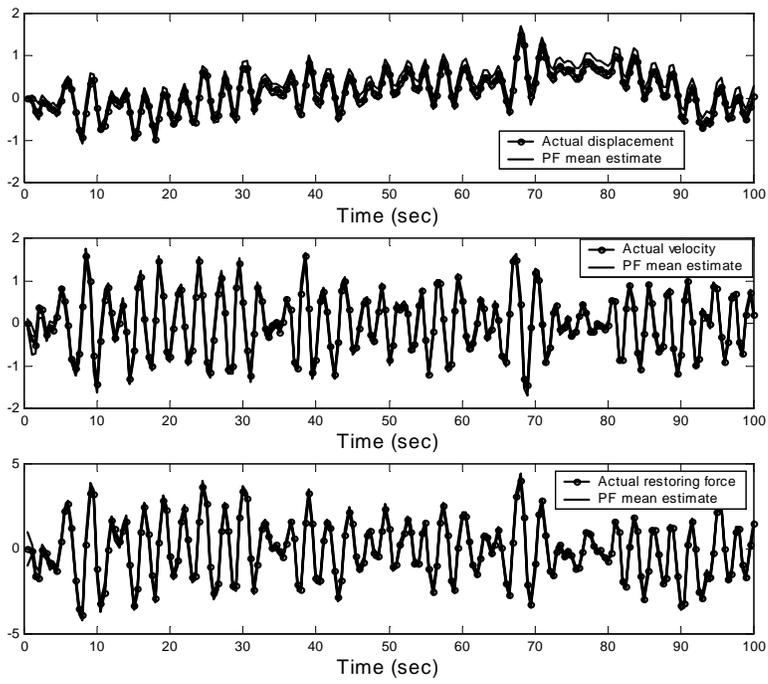


Figure 13. The PF estimates of the system states

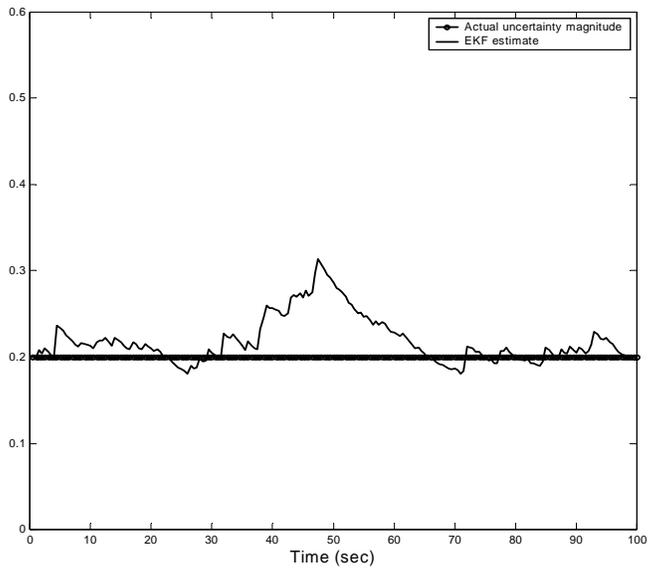


Figure 14. The EKF estimates of the uncertainty parameter

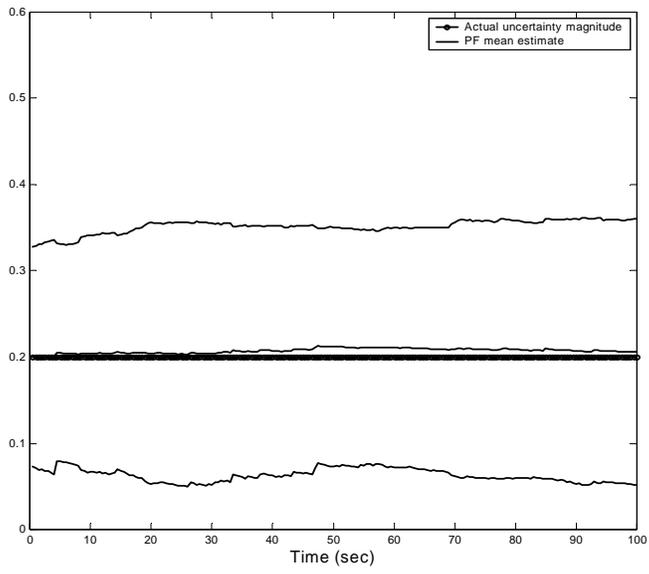


Figure 15. The PF estimates of the uncertainty parameter

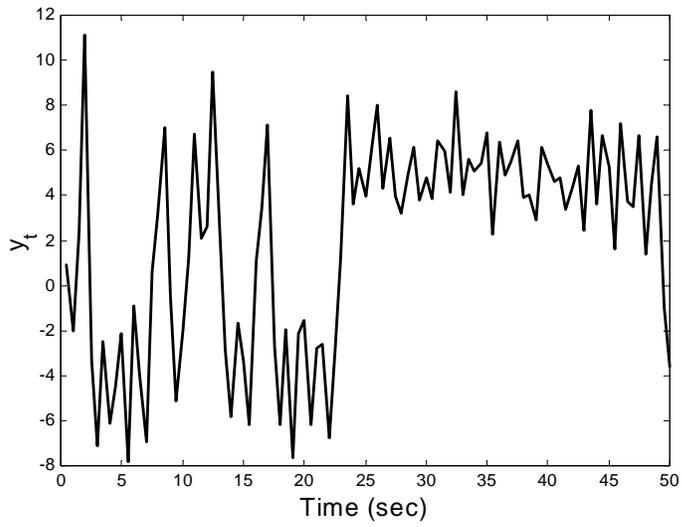


Figure 16. The plot for  $\hat{y}_t$

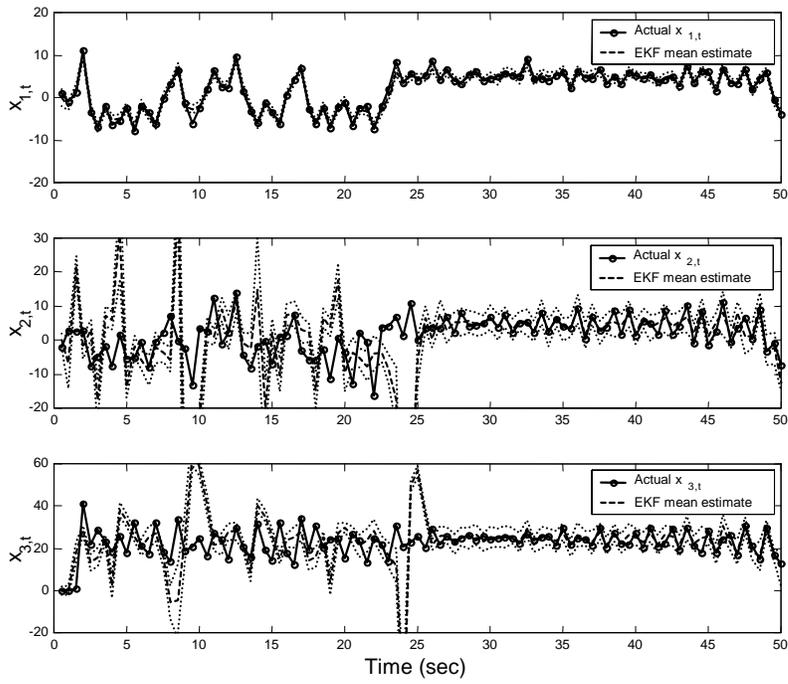


Figure 17. The EKF estimates of the system states

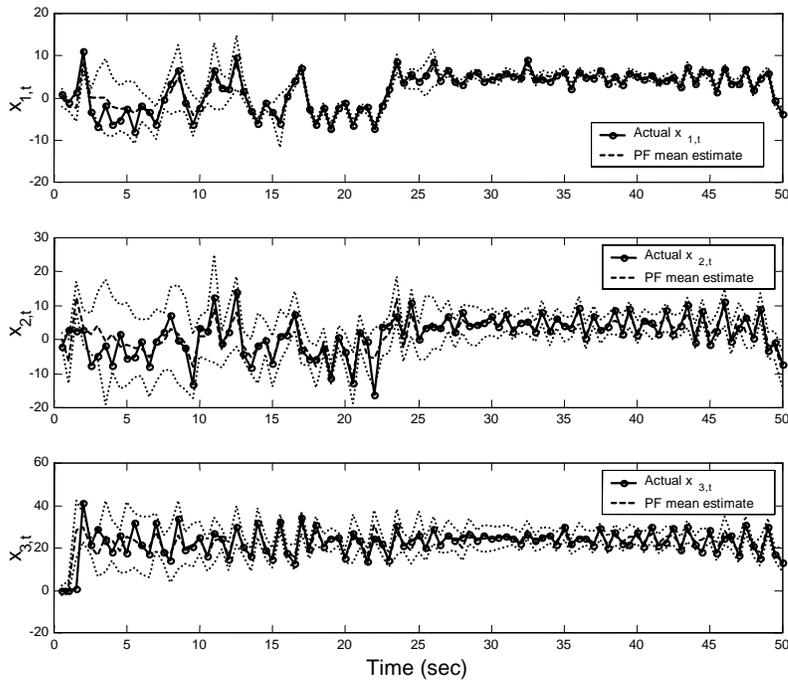


Figure 18. The PF estimates of the system states