

Bayesian State Estimation Method for Nonlinear Systems and Its Application to Recorded Seismic Response

Jianye Ching¹; James L. Beck²; Keith A. Porter³; and Rustem Shaikhutdinov⁴

Abstract: The focus of this paper is to demonstrate the application of a recently developed Bayesian state estimation method to the recorded seismic response of a building and to discuss the issue of model selection. The method, known as the particle filter, is based on stochastic simulation. Unlike the well-known extended Kalman filter, it is applicable to highly nonlinear systems with non-Gaussian uncertainties. The particle filter is applied to strong motion data recorded in the 1994 Northridge earthquake in a seven-story hotel whose structural system consists of nonductile reinforced-concrete moment frames, two of which were severely damaged during the earthquake. We address the issue of model selection. Two identification models are proposed: a time-varying linear model and a simplified time-varying nonlinear degradation model. The latter is derived from a nonlinear finite-element model of the building previously developed at Caltech. For the former model, the resulting performance is poor since the parameters need to vary significantly with time in order to capture the structural degradation of the building during the earthquake. The latter model performs better because it is able to characterize this degradation to a certain extent even with its parameters fixed. For this case study, the particle filter provides consistent state and parameter estimates, in contrast to the extended Kalman filter, which provides inconsistent estimates. It is concluded that for a state estimation procedure to be successful, at least two factors are essential: an appropriate estimation algorithm and a suitable identification model.

DOI: 10.1061/(ASCE)0733-9399(2006)132:4(396)

CE Database subject headings: Filters; Simulation; Seismic effects; Bayesian analysis.

Introduction

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 system state according to some representative model of it. State estimation has the potential to be widely applied in civil engineering. For instance, real-time structural health monitoring techniques that detect changes of dynamical properties of structural systems during earthquakes can be cast into a real-time state-estimation problem. More generally, a real-time state-estimation methodology can be used for system identification in order to understand better the nonlinear behavior of structures

subject to seismic loading. Also, for structural control, the ability to estimate the whole system state in real time can help to accomplish a more efficient control strategy.

Because of their wide applicability, real-time 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 a real-time identification methodology for better understanding of the degrading behavior of structures subject to dynamic loads. Ghanem and Shinozuka (1995) presented several different 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_∞ 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 identifying multi-degree-of-freedom nonlinear hysteretic systems for control and monitoring.

Development of Bayesian State-Estimation Algorithms

Among state-estimation methodologies, those founded on the Bayesian framework are powerful because of the following facts: (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

¹Assistant Professor, Dept. of Construction Engineering, National Taiwan Univ. of Science and Technology, Taiwan. E-mail: jyching@gmail.com

²Professor of Applied Mechanics and Civil Engineering, California Institute of Technology, Pasadena, CA 91125. E-mail: jimbeck@caltech.edu

³George W. Housner Senior Researcher in Civil Engineering, California Institute of Technology, Pasadena, CA 91125. E-mail: keithp@caltech.edu

⁴Graduate Student in Applied Mechanics, California Institute of Technology, Pasadena, CA 91125. E-mail: rustem@caltech.edu

Note. Associate Editor: Gerhart I. Schueller. Discussion open until September 1, 2006. Separate discussions must be submitted for individual papers. To extend the closing date by one month, a written request must be filed with the ASCE Managing Editor. The manuscript for this paper was submitted for review and possible publication on July 12, 2004; approved on June 15, 2005. This paper is part of the *Journal of Engineering Mechanics*, Vol. 132, No. 4, April 1, 2006. ©ASCE, ISSN 0733-9399/2006/4-396-410/\$25.00.

available information, which can then be used for any probability-based structural health monitoring, system identification, reliability assessment, or control technique. With the PDF available, one can estimate the state and also give a description of the associated uncertainties. The first Bayesian state-estimation algorithm was formulated by Kalman for linear systems with Gaussian uncertainties; it is well-known as the Kalman filter (KF) (Kalman 1960; Kalman and Bucy 1961). Later, the KF was modified to give the extended Kalman filter (EKF) to accommodate lightly nonlinear systems (Jazwinski 1970). The EKF has been the dominant Bayesian state-estimation algorithm for nonlinear systems and non-Gaussian uncertainties for the last 30 years.

Although the 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, and so the applicability of the Kalman filter and its extended version is questionable. These older techniques have been used by civil engineering researchers for decades (Beck 1978; Yun and Shinozuka 1980; Hoshiya and Saito 1984; Koh and See 1994) but their applicability for highly nonlinear systems and non-Gaussian uncertainties has rarely been studied.

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 algorithms 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 the civil engineering literature. 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.

Scope of This Work

In this paper, we investigate the application to recorded seismic response of a Bayesian state-estimation method called the particle filter (PF) that employs stochastic simulation. The PF technique has the following advantages: (1) it is applicable to highly nonlinear systems with non-Gaussian uncertainties; (2) it is not limited to the first two moments as in the KF and EKF; and (3) as the sample size approaches infinity, the resulting state estimates converge to their expected values. However, the simulation is usually computationally expensive. Sometimes the state estimates can be inaccurate because of insufficient samples. We introduce recent developments that address these difficulties and present new techniques that are useful to improve the convergence. We also compare the performance of the EKF with that of the PF using a case study.

We explicitly address the issue of model selection in this paper using two different identification model classes: one with time-varying linear dynamics and the other with simplified nonlinear degrading behavior. The former has been popular for system identification involving degrading structural systems. However, we show that the linear time-varying model does not perform well for the real-data case study. We discuss the drawbacks of the linear time-varying model and develop the simplified nonlinear model that works better.

This paper has the following structure. First, we define the

general problem of Bayesian state estimation for nonlinear dynamical systems and review the KF and EKF algorithms, then we introduce the particle filter techniques. Finally, we present a case study of the recorded seismic response of a building to demonstrate the application of the different methods and models.

Bayesian 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 Eq. (1) are called, from left to right, state transition (evolution or predictor) equation and observation (output or corrector) equation, respectively. In this equation, $x_k \in R^n$, $u_k \in R^p$, and $y_k \in R^q$ = system state, input (known excitation), and observed output at time k ; $w_k \in R^l$ and $v_k \in R^m$ are introduced to account for unknown disturbances, model errors, and measurement noise; f_k = prescribed state transition function at time k ; and h_k = prescribed observation function at time k . The values of the initial state x_0 and the variables w_k and v_k at each time k are uncertain and are described by probability models which, when combined with Eq. (1), prescribe the one-step ahead predictive PDFs $p(x_k | x_{k-1}, u_{k-1})$ and $p(y_k | x_k, u_k)$ in addition to $p(x_0)$. These three PDFs give a complete description of the stochastic state-space model and play the major role in the theory that follows.

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 evaluate the conditional PDF $p(x_k | \hat{Y}_k, U_k)$ for the state x_k at every time k in a Markovian fashion, i.e., to update sequentially this PDF conditional on the observed system input and output up to the current time, but by using only the information produced at the end of the previous time step and observation data from the current time step, along with the prescribed probabilistic models for w_k and v_k . From this conditional PDF, some important features of the state, such as the conditional expectation $E(x_k | \hat{Y}_k, U_k)$ and conditional covariance matrix $\text{Cov}(x_k | \hat{Y}_k, U_k)$, can be estimated. From now on, the conditioning of every PDF on U_k is left as implicit in the notation.

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

$$p(x_k | \hat{Y}_{k-1}) = \int p(x_k | x_{k-1}) p(x_{k-1} | \hat{Y}_{k-1}) dx_{k-1} \quad (2)$$

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

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.

An alternative formulation of the basic equations will prove useful in the next section. Let $X_k = \{x_0, x_1, \dots, x_k\}$ be the state history, then from Bayes Theorem, as in the second equation in Eq. (2)

$$\begin{aligned}
p(X_k|\hat{Y}_k) &= \frac{p(X_k|\hat{Y}_{k-1}) \cdot p(\hat{y}_k|X_k, \hat{Y}_{k-1})}{p(\hat{y}_k|\hat{Y}_{k-1})} \\
&= p(X_{k-1}|\hat{Y}_{k-1}) \cdot \frac{p(x_k|X_{k-1}, \hat{Y}_{k-1}) \cdot p(\hat{y}_k|X_k, \hat{Y}_{k-1})}{p(\hat{y}_k|\hat{Y}_{k-1})} \\
&= p(X_{k-1}|\hat{Y}_{k-1}) \cdot \frac{p(\hat{y}_k|x_k) \cdot p(x_k|x_{k-1})}{p(\hat{y}_k|\hat{Y}_{k-1})} \quad (3)
\end{aligned}$$

where we have used the fact that $p(\hat{y}_k|X_k, \hat{Y}_{k-1})=p(\hat{y}_k|x_k)$ and that $p(x_k|X_{k-1}, \hat{Y}_{k-1})=p(x_k|x_{k-1})$ based on Eq. (1) and the fact that the PDFs for v_k and w_k are prescribed. Evaluating the recursive equation in Eq. (3), we get

$$\begin{aligned}
p(X_k|\hat{Y}_k) &= p(x_0) \cdot \prod_{m=1}^k \frac{p(\hat{y}_m|x_m) \cdot p(x_m|x_{m-1})}{p(\hat{y}_m|\hat{Y}_{m-1})} \\
&= \frac{p(x_0)}{p(\hat{Y}_k)} \cdot \prod_{m=1}^k p(\hat{y}_m|x_m) \cdot p(x_m|x_{m-1}) \quad (4)
\end{aligned}$$

The main challenge in Bayesian state estimation for nonlinear systems is that these basic equations cannot be readily evaluated to get $p(x_k|\hat{Y}_k)$ because they involve high-dimensional integrations. When the system is linear and uncertainties are Gaussian, $p(x_k|\hat{Y}_k)$ is also Gaussian and the Kalman filter (KF) provides an exact solution for it by propagating and updating the first two moments of this PDF through time. However, many dynamical systems exhibit nonlinear behavior and so the direct use of KF is prohibited. If f_k and h_k are only slightly nonlinear, an approximation for KF can be derived by linearizing the system. The resulting filter is the well-known extended Kalman filter (Jazwinski 1970). The degree of accuracy of EKF relies on the validity of the linear approximation and EKF is not suitable for tracking multimodal or highly non-Gaussian conditional PDFs due to the fact that it only updates the first two moments.

Particle Filters

For systems with non-Gaussian uncertainties, it is often desirable to propagate and update the conditional PDF itself; however, doing so thoroughly requires 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 which can then be used to consistently estimate the conditional expectation of any function of x_k . We focus on stochastic simulation techniques and use the term particle filters (PF) to denote the resulting algorithms [following Doucet and Andrieu (2000); van der Merwe et al. (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).

Stochastic Simulation for State Estimation

Our interest is to develop a simulation algorithm for the conditional PDF $p(X_k|\hat{Y}_k)$ in Eq. (4) that is Markovian in that information is required only from time steps $k-1$ and k ; earlier state and

observation data can be forgotten. In other words, if \hat{X}_{k-1} is a sample from $p(X_{k-1}|\hat{Y}_{k-1})$, the sample 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 =new sample and \hat{X}_{k-1} =previous sample from $p(X_{k-1}|\hat{Y}_{k-1})$. However, such an algorithm is incomplete because it ignores the information \hat{y}_k , i.e., because $p(X_{k-1}|\hat{Y}_{k-1})$ is different from $p(X_{k-1}|\hat{Y}_k)=p(X_{k-1}|\hat{Y}_{k-1}, \hat{y}_k)$.

Importance-Sampling Technique

We can, however, sample from an importance-sampling PDF $q(X_k|\hat{Y}_k)$ that admits a 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 $q(X_k|Y_k)$ is 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 $q(X_k|\hat{Y}_k)$ (selected so that it is readily sampled), the expectation conditioned on \hat{Y}_k of any function of the state X_k , denoted by $r(X_k)$, can be estimated using the importance-sampling technique as follows

$$E[r(X_k)|\hat{Y}_k] \approx \frac{1}{N} \sum_{i=1}^N \hat{\beta}_k^i \cdot r(\hat{X}_k^i) \equiv \hat{r}_{k,N}^1 \quad (5)$$

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

Any quantity of interest can be estimated with the appropriate $r(\cdot)$ function in Eq. (5); for instance, if $r(X_k)=X_k$, $E[r(X_k)|\hat{Y}_k]$ is simply the conditional mean of the state history $E[X_k|\hat{Y}_k]$; if $r(X_k)=X_k X_k^T$, $E[r(X_k)|\hat{Y}_k]$ is the conditional second moment $E[X_k X_k^T|\hat{Y}_k]$. For buildings, the quantity of interest might be any performance parameter such as interstory drift, floor accelerations, repair cost, repair duration, casualties, occupancy, or operability.

Let $\{X_k^i: i=1, \dots, N\}$ denote N random samples of the state history that are to be drawn from $q(X_k|\hat{Y}_k)$. It is readily shown that the estimator $\hat{r}_{k,N}^1=(1/N)\sum_{i=1}^N \beta_k^i \cdot r(X_k^i)$ in Eq. (5) 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 $q(X_k|\hat{Y}_k)$

$$\begin{aligned}
E[\hat{r}_{k,N}^1] &= \frac{1}{N} \sum_{i=1}^N E_q[\beta_k^i \cdot r(X_k^i)] = E_q[\beta_k \cdot r(X_k)] \\
&= \int [p(X_k|\hat{Y}_k)/q(X_k|\hat{Y}_k)] r(X_k) \cdot q(X_k|\hat{Y}_k) dX_k \\
&= \int r(X_k) \cdot p(X_k|\hat{Y}_k) dX_k = E[r(X_k)|\hat{Y}_k] \quad (6)
\end{aligned}$$

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

Although $\hat{r}_{k,N}^1$ 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 Eq. (4), which is a difficult task. Nevertheless, we show that the following estimator is computable while it is also asymptotically unbiased and consistent

$$r_{k,N}^2 \equiv \left(\frac{1}{N} \sum_{i=1}^N \beta_k^i \cdot r(X_k^i) \right) / \left(\frac{1}{N} \sum_{j=1}^N \beta_k^j \right) = r_{k,N}^1 / \bar{\beta}_k^N \quad (7)$$

where

$$\bar{\beta}_k^N = \left(\sum_{j=1}^N \beta_k^j \right) / N \quad (8)$$

$$\tilde{\beta}_k^i = p(\hat{X}_k^i | \hat{Y}_k) / q(\hat{X}_k^i | \hat{Y}_k) / \left(\sum_{j=1}^N p(\hat{X}_k^j | \hat{Y}_k) / q(\hat{X}_k^j | \hat{Y}_k) \right) = \frac{p(\hat{x}_0^i)}{q(\hat{X}_k^i | \hat{Y}_k)} \cdot \prod_{m=1}^k p(\hat{y}_m | \hat{x}_m^i) \cdot p(\hat{x}_m^i | \hat{x}_{m-1}^i) / \left(\sum_{j=1}^N \frac{p(\hat{x}_0^j)}{q(\hat{X}_k^j | \hat{Y}_k)} \cdot \prod_{m=1}^k p(\hat{y}_m | \hat{x}_m^j) \cdot p(\hat{x}_m^j | \hat{x}_{m-1}^j) \right) \quad (10)$$

Therefore the factor $p(\hat{Y}_k)$ in Eq. (21) has been canceled 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 Eq. (1) uniquely specify v_m and w_m , given y_m , x_m , and x_{m-1} . We select $q(\hat{X}_k^i | \hat{Y}_k)$ so it can be readily evaluated too. The proof of the asymptotic unbiasedness and consistency of $r_{k,N}^2$ is given in Ching et al. (2004).

The selection of an importance sampling PDF that admits a real-time procedure is discussed in Ching et al. (2004). In this paper, we select the following importance sampling PDF

$$q(X_k | \hat{Y}_k) = p(x_0) \cdot \prod_{m=1}^k p(x_m | x_{m-1}) \quad (11)$$

The corresponding non-normalized importance weight is

$$\beta_k = \prod_{m=1}^k \frac{p(\hat{y}_m | x_m) \cdot p(x_m | x_{m-1})}{p(x_m | x_{m-1})} = \beta_{k-1} \cdot p(\hat{y}_k | x_k) \quad (12)$$

Because of the structure of the algorithm, at any time k , we are only required to store the sampled states and weights in the most recent two time steps, i.e., k and $k-1$, if the quantity of interest is $r(x_k)$ and so depends on the current state (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 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

$$\tilde{x}^i = \hat{x}^i, \quad \tilde{\beta}^i = \beta^i \quad (13)$$

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

$$\beta^i = \tilde{\beta}^i \cdot p(\hat{y}_k | x_k = \hat{x}^i) \quad (14)$$

3. Using the samples $i=1, \dots, N$, $E[r(x_k) | \hat{Y}_k]$ can be approximated based on Eqs. (26) and (30)

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

$$\hat{r}_{k,N}^2 = \sum_{i=1}^N \left[\beta_k^i / \left(\sum_{j=1}^N \beta_k^j \right) \right] \cdot r(\hat{X}_k^i) = \sum_{i=1}^N \tilde{\beta}_k^i \cdot r(\hat{X}_k^i) \quad (9)$$

where

$$E[r(x_k) | \hat{Y}_k] \approx \sum_{i=1}^N \left[\beta^i / \left(\sum_{j=1}^N \beta^j \right) \right] \cdot r(\hat{x}^i) \quad (15)$$

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

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

Reducing Sample Degradation: 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 Eq. (15), but they become far from uniform as k grows, which is due to the recursion in Eq. (12) 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 sample degradation can be reduced, as described in this section.

Instead of letting the N samples evolve through time independently (Algorithm 1), we can resample the samples when the importance weights become highly nonuniform (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 nonuniform. This can be done by monitoring the coefficient of variation (COV) of the importance weights. The resampling procedure is executed only when this COV 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 2: 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 $\beta^{i,j}=1/N$ for $i=1, \dots, N, j=1, \dots, L$.
2. Perform the following steps (3) and (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

$$\bar{x}^{i,j} = \hat{x}^{i,j}, \quad \tilde{\beta}^{i,j} = \beta^{i,j} \quad (16)$$

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

$$\tilde{\beta}^{i,j} = \tilde{\beta}^{i,j} \cdot p(\hat{y}_k | x_k = \bar{x}^{i,j}) \quad (17)$$

4. Compute the COV of $\{\tilde{\beta}^{i,j}: i=1, \dots, N\}$. If the COV is larger than the prescribed threshold, then execute the resampling step for $i=1, \dots, N$

$$\hat{x}^{i,j} = \bar{x}^{i,j}, \quad \text{with probability } \tilde{\beta}^{i,j} / \sum_{i=1}^N \tilde{\beta}^{i,j} \quad (18)$$

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

$$\hat{x}^{i,j} = \bar{x}^{i,j}, \quad \beta^{i,j} = \tilde{\beta}^{i,j} / \sum_{i=1}^N \tilde{\beta}^{i,j} \quad (19)$$

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

$$E[r(x_k) | \hat{Y}_k] \approx \left(\sum_{j=1}^L \hat{r}_{k,N}^j \right) / L \quad (20)$$

6. Do steps 2–5 for $k=1, \dots, T$.

After the resampling step in Algorithm 2, 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 random walk at each time step, where $p(x_k | \hat{Y}_k)$ =stationary PDF of the Markov chain. In Ching et al. (2004), we summarize the procedure for using this MCMC step.

Advantages and Disadvantages of Particle Filter 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 quantities of interest, $r(x_k)$, do not significantly change.

Case Study of Recorded Seismic Response

Building Description

The selected building for a case study is a seven-story, 66,000 square-foot (6,200 m²) hotel located in the San Fernando Valley of Los Angeles County, Calif. The building is 63 ft by 150 ft in plan, consisting of 3 bays by 8 bays, with the long direction oriented east-west. It is approximately 65 ft tall: the first story is 13 ft, 6 in.; stories 2–6 are 8 ft, 6 1/2 in.; the seventh story is 8 ft, 6 in. We refer to this building as the Van Nuys hotel. It was built in 1966 according to the 1964 Los Angeles Building Code. The lateral-force-resisting system is a perimeter nonductile reinforced-concrete moment frame in both directions; internal gravity frames also provide some resistance. The gravity system comprises two-way reinforced-concrete flat slabs supported by square columns at the interior and by the rectangular columns of the perimeter frame. The building was lightly damaged by the M6.6 1971 San Fernando event, approximately 20 km to the northeast, and severely damaged by the M6.7 1994 Northridge earthquake, whose epicenter was approximately 4.5 km to the southwest. In particular, both the south and north frames of the building were seriously damaged during the Northridge earthquake. The building has been studied extensively, e.g., by Jennings (1971); Scholl et al. (1982); Islam (1996a,b); Islam et al. (1998); Li and Jirsa (1998); and Beck et al. (2002). A detailed description for the building can be found in Beck et al. (2002).

Strong-motion data were obtained from 16 channels that recorded motions during the Northridge earthquake. Among the 16 channels, five channels measured the accelerations in the longitudinal (east-west: E-W) direction at the ground, second, third, and sixth floors and the roof. The locations of the five accelerometers were at the south-east corner of the ground floor and near the east wall at the second, third, and sixth floors and the roof. Fig. 1 shows the measured E-W acceleration time histories during the earthquake. These strong-motion data are used in this case study.

The purpose of this case study is to examine the use of Bayesian state-estimation techniques in tracking the system state and parameters of the building during the earthquake. Moreover, we examine two identification model classes: (1) a time-varying linear model; and (2) a simplified time-varying nonlinear degradation model. We show that the latter is better in tracking the system dynamics. We focus exclusively on the dynamics in the E-W direction due to the fact that this involves the south frame that was severely damaged during the Northridge earthquake.

Time-Varying Linear Identification Model

We first investigate a linear structural identification model with unknown time-varying interstory stiffnesses and dampings. This model is our first attempt because linear time-varying structural models are popular for identifying degrading civil engineering

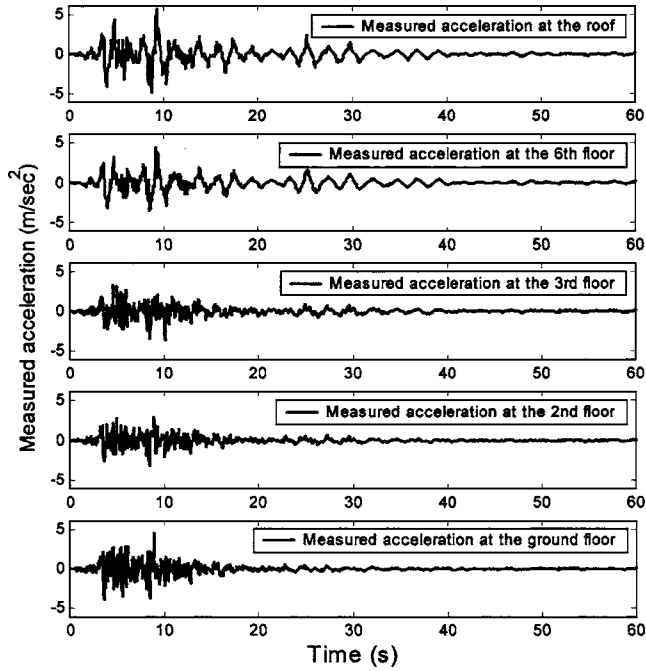


Fig. 1. Measured east-west accelerations from the Van Nuys hotel during the Northridge earthquake

systems. The model is a one-dimensional (1D) seven-degree-of-freedom lumped-mass shear-building model with the following continuous-time state-space equations

$$\frac{d}{dt} \begin{bmatrix} x_t \\ \dot{x}_t \\ \theta_t \end{bmatrix} = \begin{bmatrix} \dot{x}_t \\ -M^{-1}K(\theta_t)x_t - M^{-1}C(\theta_t)\dot{x}_t \\ 0_{18 \times 1} \end{bmatrix} - \begin{bmatrix} 0_{7 \times 1} \\ 1_{7 \times 1} \\ 0_{18 \times 1} \end{bmatrix} \cdot u_t + \begin{bmatrix} 0 \\ 0 \\ G \end{bmatrix} \cdot w_t$$

$$y_t = \begin{bmatrix} \ddot{x}_{7,t} + u_t \\ \ddot{x}_{5,t} + u_t \\ \ddot{x}_{2,t} + u_t \\ \ddot{x}_{1,t} + u_t \end{bmatrix} + H(\theta_t) \cdot v_t$$

$$= - \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix} \cdot M^{-1}[C(\theta_t)\dot{x}_t + K(\theta_t)x_t] + H(\theta_t) \cdot v_t \quad (21)$$

where $0_{m \times n}$ denotes an $(m \times n)$ matrix whose entries are all zeros (and similar for $1_{m \times n}$); $[x_t \ \dot{x}_t \ \theta_t]^T \in \mathbb{R}^{32 \times 1}$ = augmented state; $x_{i,t} \in \mathbb{R}$ denotes the displacement relative to the ground at the $(i+1)$ th floor (the eighth floor is the roof) at time t ; $u_t \in \mathbb{R}$ = measured acceleration at the ground floor of the building at time t ; $y_t \in \mathbb{R}^{4 \times 1}$ contains the E-W acceleration time histories measured at the second, third, and sixth floors and the roof at time t ; $\theta_t \in \mathbb{R}^{18 \times 1}$ = vector containing uncertain system parameters at time t , including interstory stiffness $k_{i,t}$ and damping $c_{i,t}$ of each story (index i denotes story number) plus four uncertainty parameters $h_{1,t}, \dots, h_{4,t}$, that is $\theta_t = [k_{1,t}, \dots, k_{7,t}, c_{1,t}, \dots, c_{7,t}, h_{1,t}, \dots, h_{4,t}]^T$; masses are assumed to be known perfectly and calculated according to the structural drawings: 120.02 tons for the first floor, 102.11 tons for the second to seventh floors, and 95.87 tons for the roof; $w_t \in \mathbb{R}^{18 \times 1} \sim N(0, I)$, $v_t \in \mathbb{R}^{4 \times 1} \sim N(0, I)$ and are independent at each time (Gaussian white noise); $G \in \mathbb{R}^{18 \times 18}$ = diagonal

matrix whose diagonals we specify, as described later

$$x_t = \begin{bmatrix} x_{7,t} \\ x_{6,t} \\ \vdots \\ x_{1,t} \end{bmatrix}, \quad M = \begin{bmatrix} m_7 & 0 & 0 & 0 \\ 0 & m_6 & 0 & 0 \\ 0 & 0 & \ddots & 0 \\ 0 & 0 & 0 & m_1 \end{bmatrix}, \quad F = \begin{bmatrix} -m_7 \\ -m_6 \\ \vdots \\ -m_1 \end{bmatrix}$$

$$H(\theta_t) = \begin{bmatrix} h_{1,t} & 0 & 0 & 0 \\ 0 & h_{2,t} & 0 & 0 \\ 0 & 0 & h_{3,t} & 0 \\ 0 & 0 & 0 & h_{4,t} \end{bmatrix} \quad (22)$$

$$C(\theta_t) = \begin{bmatrix} c_{7,t} & -c_{7,t} & 0 & 0 \\ -c_{7,t} & c_{7,t} + c_{6,t} & \ddots & 0 \\ 0 & \ddots & \ddots & -c_{2,t} \\ 0 & 0 & -c_{2,t} & c_{2,t} + c_{1,t} \end{bmatrix}$$

$$K(\theta_t) = \begin{bmatrix} k_{7,t} & -k_{7,t} & 0 & 0 \\ -k_{7,t} & k_{7,t} + k_{6,t} & \ddots & 0 \\ 0 & \ddots & \ddots & -k_{2,t} \\ 0 & 0 & -k_{2,t} & k_{2,t} + k_{1,t} \end{bmatrix}$$

Note that the system parameters are unknown so they are augmented into the system state in Eq. (21) and are estimated together with the other state variables. Moreover, the unknown system parameters θ_t are modeled by a random walk with time, as shown by the dynamic equation $\dot{\theta}_t = G \cdot w_t$ in Eq. (21). We do not expect that the building behaved linearly during the Northridge earthquake; instead, the rationale here is to use the time-varying linear model to track the nonlinear behavior. Also, note that the prediction-error term $H(\theta_t) \cdot v_t$ in this model is nonstationary Gaussian white noise.

The prior PDFs for x_0 and \dot{x}_0 are taken to be independent zero-mean Gaussian with zero standard deviations, that is, we are confident that the structure starts from an at-rest initial condition; the prior PDFs of $\{k_{i,0}; i=1, 2, \dots, 7\}$ and $\{c_{i,0}; i=1, 2, \dots, 7\}$ are taken to be independent Gaussian with mean equal to our best estimate of the undamaged stiffness and damping (the calibrated values, as defined later) with COV equal to 5% for the stiffnesses and 20% for the dampings. For $h_{1,0}, \dots, h_{4,0}$, the prior PDFs have means equal to 0.1 m/s² and COV equal to 20%. The diagonals of G are chosen such that in each time step (set to 0.04 s), $k_{i,t}$ and $c_{i,t}$ drift with a COV equal to 1% with a restriction that they cannot drift into negative values, and $h_{1,t}, \dots, h_{4,t}$ drift with COV equal to 20%. Note that we have allowed the uncertainty parameters to drift more freely to accommodate the possible large fluctuations of these parameters; however, for the stiffness and damping parameters, only limited fluctuations are allowed.

Results and Discussion

We first convert Eq. (21) to a discrete-time state-space system with a time step of 0.04 s by using numerical integration over a time step and only implement PF for this identification model. We employ *Matlab* routine ODE23 for this purpose. Fig. 2 shows the stiffness and damping parameter estimates and the associated 95% confidence intervals from PF (using Algorithm 2 with number of samples $N=200$ and number of parallel filters $L=10$ and the importance weight COV threshold=200%). If more samples are used in PF than the $N \cdot L=2,000$ samples, there is little change

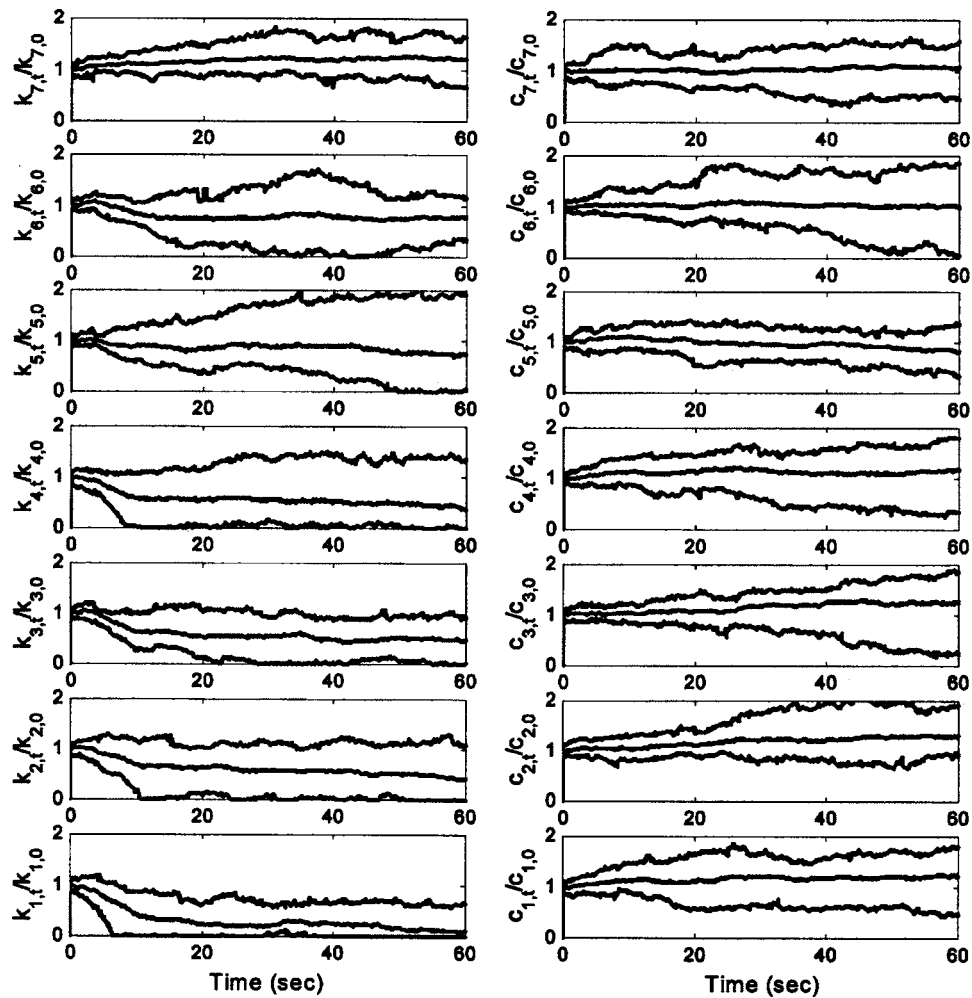


Fig. 2. Estimated stiffness (left side) and damping (right side) by the linear time-varying model; upper and lower lines are 95% confidence bounds

in the estimated means and variances of the unknown parameters, indicating that the results are nearly converged.

The predicted output accelerations [the estimated means of $p(y_k|D_k)$ where D_k denotes the strong-motion data up to time k] and interstory drifts [which can be derived from the estimated means of $p(x_k|D_k)$] from the model are illustrated in Figs. 3 and 4, in which the 95% confidence intervals are also shown. The four measured accelerations and the “measured” interstory drifts in the first and second stories (obtained from double integrating and high-pass filtering the difference between the accelerations measured at the first story and ground floor base, and second and first stories) are also plotted in the figures. The results show a significant amount of uncertainties since the confidence intervals are large, indicating that the performance of the time-varying model is not good. However, the results do predict that the largest interstory drift ratio occurs in the fourth story where the hotel was damaged most in the Northridge earthquake.

Before developing a better identification model, we discuss some issues that are essential for model selection and that caused difficulties when we implemented the time-varying linear model.

1. **Bias-variance trade-off.** Consider two extreme cases. First, suppose that the G diagonals in Eq. (21) are so large that the stiffness and damping values at different time instants have essentially no correlation. Such a model is flexible, that is, one can fit the data quite well by adjusting the time-varying

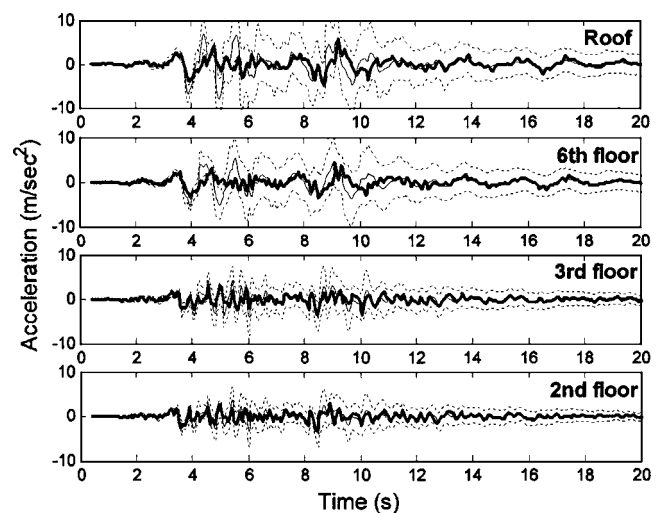


Fig. 3. Thick solid lines are measured accelerations; central thin solid lines are predicted acceleration time histories by the linear time-varying model; and upper and lower dashed lines are 95% confidence intervals

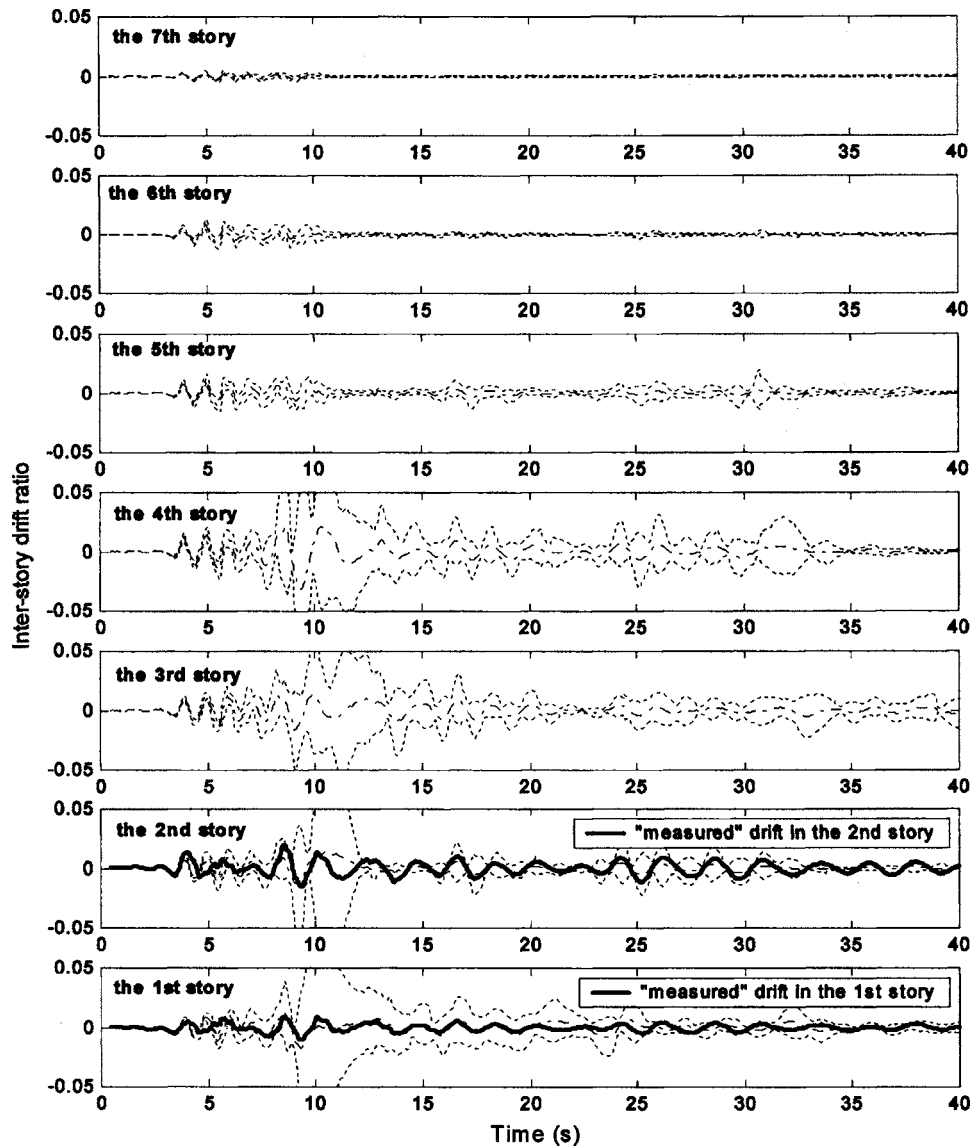


Fig. 4. Solid lines in the lower two lots are “measured” interstory drift ratios; central dashed lines are predicted interstory drift ratios by the linear time-varying model; and upper and lower dashed lines are 95% confidence intervals

parameters freely, so the model is less biased. But it is likely that there are numerous possible choices of the time-varying parameters that will also provide a comparable fit, so the model has high parameter variance. Therefore the problem of estimating the stiffnesses and dampings is ill-posed, and the estimated parameters reflect little about the actual system dynamics.

Now consider the other extreme: the G diagonals are zeros so that the stiffnesses and dampings at different time instants have perfect correlation, that is, it gives a time-invariant linear model. Such a model is rigid, since one may only at best get a moderate fit to the dataset by adjusting the time-invariant parameters, making the model more biased. But there are usually only few possible choices of the parameters that will provide the moderate fit, so the model has low parameter variance.

Unfortunately, because the Van Nuys hotel was strongly shaken, the time-invariant linear model (the rigid model) is expected to be inappropriate, giving large modeling errors. Therefore, we should choose the G diagonals so that they are

relatively large, allowing the linear model to change its parameters rapidly through time to accommodate the non-linear behavior. By doing so, the resulting model is quite flexible, and the state and parameter estimates have relatively large variances, as seen in Figs. 2–4. An ideal model is such that it is relatively rigid, so the parameters are only allowed to change slowly through time (i.e., small variance), but there still exist some choices of the system parameters that can fit the data well (i.e., small bias).

2. **Correlation in prior probability density functions.** The prior PDFs imply initially uncorrelated interstory stiffnesses and dampings. However, it would be better if some correlation is built into the prior PDF because of the following facts: First, when degradation occurs, stiffnesses usually decrease and damping usually increases; therefore they are in general correlated. Second, the stiffness and damping of adjacent stories could be correlated. This is because two adjacent stories usually have similar interstory deformation, and hence similar degradation. The linear time-varying model does not provide for this correlation. For instance, in Fig. 2, the first-story

stiffness drops, while the first-story damping is basically unchanged. To model this correlation in the prior PDFs, we must use the full G matrix and tune all of the G entries to get the desired correlation instead of only using its diagonals. However, this is not easy.

- Computational effort.** Finally, there are many (18) free parameters in the time-varying linear model, making the computations expensive.

According to the discussion of the bias-variance trade-off, the following question should be asked: Is it possible to find a model that is relatively rigid by having only a few slowly changing parameters, and yet there still exist some choices of the system parameters that can fit the data well? In the next section, we will introduce such a model.

Simplified Time-Invariant Nonlinear Degradation Model

In Ching et al. (2004), we describe in detail the procedure for developing a simplified nonlinear degradation model that is able to capture the E-W responses of the hotel building. This model is derived from a more complicated finite-element model developed at Caltech [for details of this finite-element model, see Beck et al. (2002), Appendix G]. It is a 1D seven-DOF lumped-mass shear-building model with the following governing equations

$$\begin{aligned} m_1(\ddot{x}_{1,t} + u_t) + f_{1,t} - f_{2,t} &= 0 \\ &\vdots \\ m_6(\ddot{x}_{6,t} + u_t) + f_{6,t} - f_{7,t} &= 0 \\ m_7(\ddot{x}_{7,t} + u_t) + f_{7,t} &= 0 \end{aligned} \quad (23)$$

where $f_{i,t}$ =interstory restoring force in the i th story and u_t =acceleration at the ground floor at time t . The relationship between the interstory restoring force and interstory drift is as follows

$$\begin{aligned} f_{1,t} &= c_{1,t} \cdot \dot{x}_{1,t} + k_{1,t} \cdot x_{1,t}, \\ f_{i,t} &= c_{i,t} \cdot (\dot{x}_{i,t} - \dot{x}_{i-1,t}) + k_{i,t} \cdot (x_{i,t} - x_{i-1,t}) \quad \text{for } i=2,3,\dots,7 \end{aligned} \quad (24)$$

To model the degradation, we first define the ductility index of the i th story as

$$\begin{aligned} \mu_{1,t} &= \max_{0 \leq k \leq t} (|x_{1,k}|/H_1), \quad \mu_{i,t} = \max_{0 \leq k \leq t} (|x_{i,k} - x_{i-1,k}|/H_i) \\ &\text{for } i=2,3,\dots,7 \end{aligned}$$

where H_i =story height of the i th story. The degradation (i.e., the behavior that stiffnesses decrease and dampings increase with deformation) is then modeled by the following rule

$$k_{i,t} = k_{i,0} \cdot e^{-\gamma \mu_{i,t}^\rho}, \quad c_{i,t} = \delta + \lambda \cdot \mu_{i,t} \quad (25)$$

This gives a nonlinear model for the interstory restoring forces versus drift with parameters γ , ρ , δ , λ , and $\{k_{i,0}; i=1,2,\dots,7\}$. Notice that the same parameters γ , ρ , δ , and λ are used for each of the seven stories. We further reduce the model flexibility by parameterizing $\{k_{i,0}; i=1,2,\dots,7\}$ using a single stiffness scaling parameter α such that

$$k_{i,0} = \alpha \cdot k_{i,0}^C, \quad i=1,2,\dots,7 \quad (26)$$

where $k_{i,0}^C$ =calibrated value for the interstory stiffness $k_{i,0}$ (as discussed in the following). Assuming $\{k_{i,0}; i=1,2,\dots,7\}$ to be strongly correlated up to a scaling factor α greatly reduces the

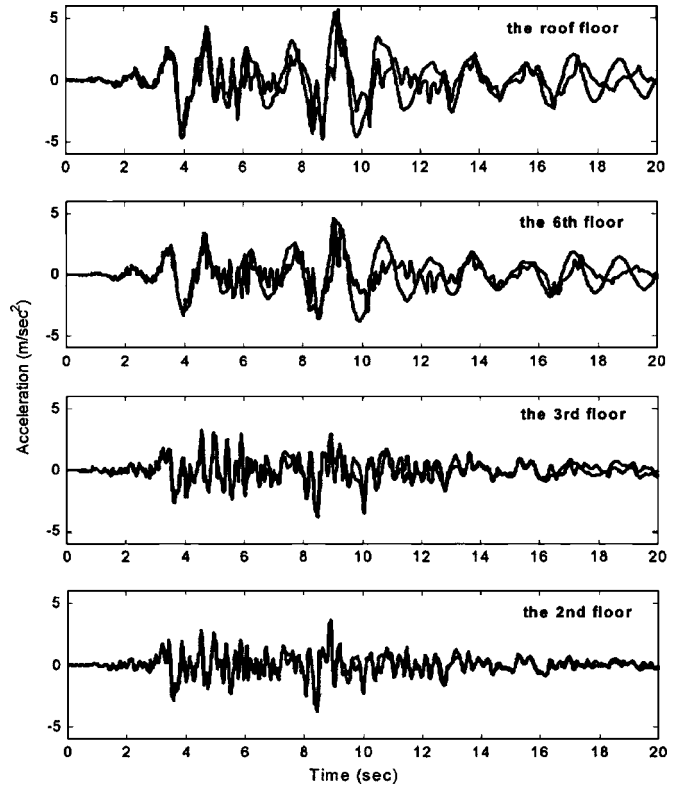


Fig. 5. Thick solid lines are measured accelerations; and thin solid lines are acceleration time histories calculated from the simplified nonlinear degradation model using recorded ground floor accelerations as input

flexibility of the model and so it is likely that this assumption will bring in more bias but reduce the variance.

A continuous-time state-space form for the simplified nonlinear degradation model can be built as follows

$$\begin{aligned} \frac{d}{dt} \begin{bmatrix} x_t \\ \dot{x}_t \end{bmatrix} &= \begin{bmatrix} \dot{x}_t \\ g(x_t, \dot{x}_t, \beta) \end{bmatrix} - \begin{bmatrix} 0_{7 \times 1} \\ 1_{7 \times 1} \end{bmatrix} \cdot u_t \\ y_t &= \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix} \cdot g(x_t, \dot{x}_t, \beta) \end{aligned} \quad (27)$$

where $\beta = [\gamma \ \rho \ \delta \ \lambda \ \alpha]^T \in R^{5 \times 1}$; and $g(x_t, \dot{x}_t, \beta) \in R^{7 \times 1}$ characterizes Eqs. (23)–(25). Using the Northridge E-W ground-floor acceleration as the input u_t , the computed output accelerations at the sensor locations using the simplified nonlinear degradation model with the following parameter setting: $\gamma=5.0$, $\rho=0.4$, $\delta=0.5$ MN/m·s, and $\lambda=50$ MN/m·s, $k_{1,0}=100$ MN/m, $k_{2,0}=120$ MN/m, $k_{3,0}=98$ MN/m, $k_{4,0}=k_{5,0}=k_{6,0}=k_{7,0}=80$ MN/m (these parameter values are called the calibrated values) are shown in Fig. 5, where we see that despite some mismatch, the output of the time-invariant model can capture the overall trend of the measured accelerations.

Time-Varying Nonlinear Degradation Identification Model

To perform state estimation, a time-varying state-space stochastic model is derived based on Eq. (27). This identification model has

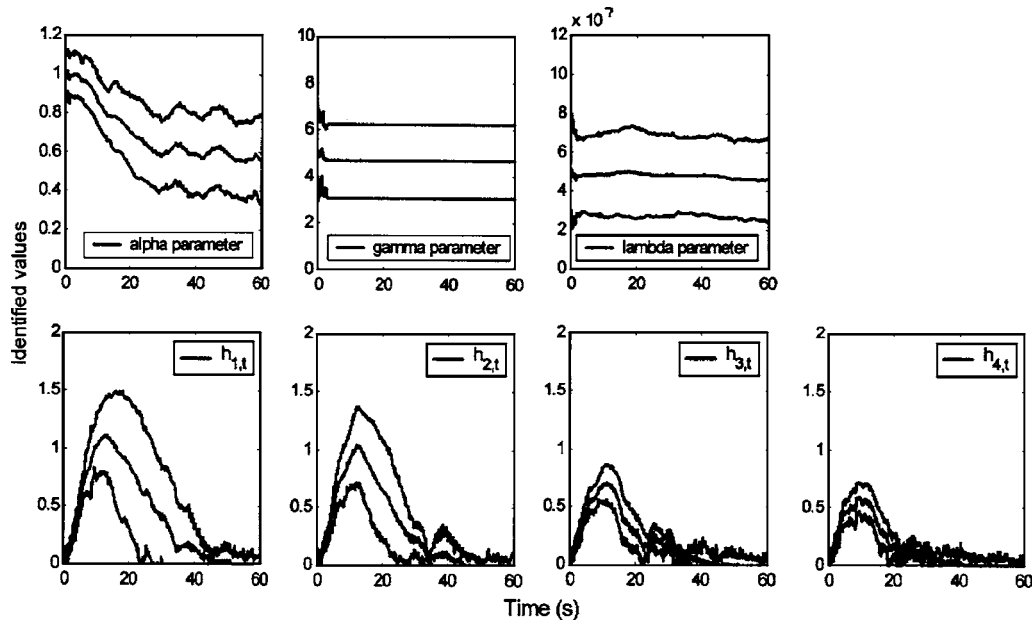


Fig. 6. Estimated system parameters by particle filter; upper and lower lines are the 95% confidence intervals (the results for the fixed parameters $\rho_i=0.4$ and $\delta_i=0.5$)

time-varying uncertainty parameters in addition to the time-varying system parameters $\beta_t = [\gamma_t, \rho_t, \delta_t, \lambda_t, \alpha_t]^T$. The rationale behind allowing the system parameters, such as β_t , to drift with time is as follows: it is very likely that the simplified nonlinear degradation model has modeling error. This modeling error can be partially compensated for by allowing the system parameters to vary with time, that is, the model is able to adaptively fit the data by slowly changing its parameter values. Nevertheless, based on the results given in the previous section (Fig. 5), we expect that the modeling error for the nonlinear model is less than that for the linear model considered previously. Therefore the amount of change in the parameters required for the nonlinear model should be less than that required for the linear model, so the time-varying nonlinear model is less flexible and should result in less parameter variance.

The resulting continuous-time identification model is as follows

$$\frac{d}{dt} \begin{bmatrix} x_t \\ \dot{x}_t \\ \theta_t \end{bmatrix} = \begin{bmatrix} \dot{x}_t \\ g(x_t, \dot{x}_t, \theta_t) \\ 0_{9 \times 1} \end{bmatrix} - \begin{bmatrix} 0_{7 \times 1} \\ 1_{7 \times 1} \\ 0_{9 \times 1} \end{bmatrix} \cdot u_t + \begin{bmatrix} 0_{7 \times 1} \\ 0_{7 \times 1} \\ G \end{bmatrix} \cdot w_t \quad (28)$$

$$y_t = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix} \cdot g(x_t, \dot{x}_t, \theta_t) + H(\theta_t) \cdot v_t$$

where $\theta_t = [\gamma_t, \rho_t, \delta_t, \lambda_t, \alpha_t, h_{1,t}, h_{2,t}, h_{3,t}, h_{4,t}]^T \in R^{9 \times 1}$; $w_t \in R^{9 \times 1} \sim N(0, I)$; $v_t \in R^{4 \times 1} \sim N(0, I)$; and matrix $H(\theta_t) \in R^{4 \times 4}$ is assumed to be diagonal as in Eq. (22), so there are four uncertainty parameters $h_{i,t}, i=1, \dots, 4$, on the diagonal entry of the $H(\theta_t)$ matrix; $g(x_t, \dot{x}_t, \theta_t)$ is the same as $g(x_t, \dot{x}_t, \beta_t)$ in Eq. (27); $G \in R^{9 \times 9}$ = prescribed matrix.

Among the system parameters, several parameters confound together; for example, there is more than one set of α_t , γ_t , and ρ_t that corresponds to the same mathematical relationship between $k_{i,t}$ and $\mu_{i,t}$; this is also the case for δ_t and λ_t . This confounding

can be resolved by holding some of the parameters constant. Alternatively, in the Bayesian framework, this can be done by constraining the prior PDFs of these parameters, as described later.

To complete the identification model, we have to specify the prior PDFs for x_0, \dot{x}_0, θ_0 and the diagonals of the G matrix. The prior PDFs of x_0, \dot{x}_0 are taken to be independent zero-mean Gaussian with zero standard deviations, since we are certain that the structure starts from an at-rest initial condition. To resolve the confounding problem, the prior PDFs for ρ_0 and δ_0 are chosen to have means equal to the calibrated values and COV equal to zero, and ρ_t and δ_t do not change with time, so they are also fixed at the optimal values: $\rho_t=0.4$ and $\delta_t=0.5$. For γ_0 and λ_0 , the prior

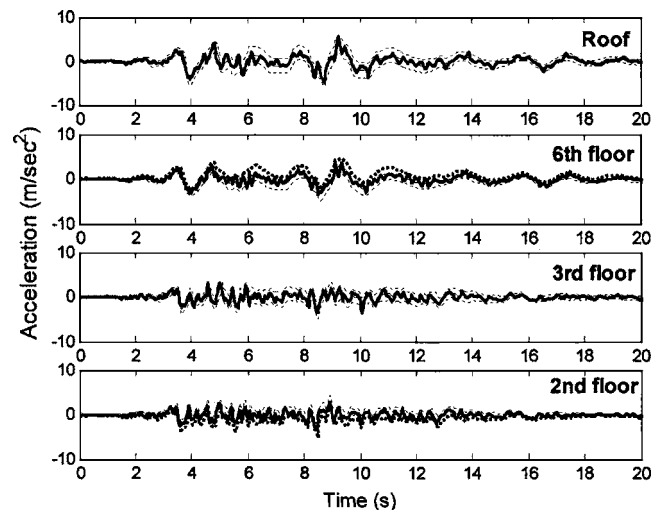


Fig. 7. Thick solid lines are measured accelerations; central thin solid lines are predicted acceleration time histories by the simplified nonlinear degradation model; and upper and lower dashed lines are 95% confidence intervals

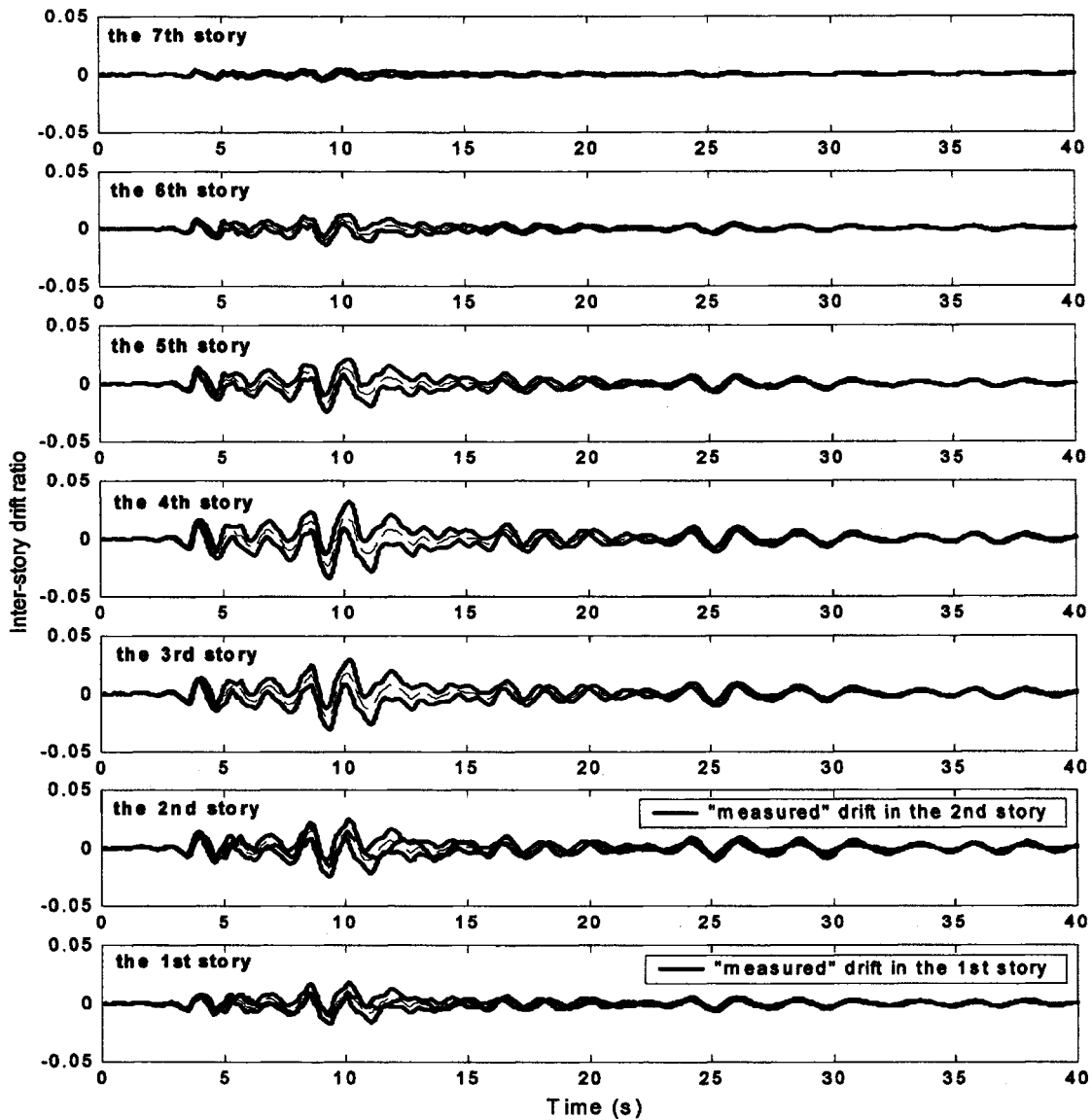


Fig. 8. Central thick solid lines in the lower two plots are “measured” interstory drift ratios; central dashed lines are predicted interstory drift ratios by the simplified nonlinear degradation model; and upper and lower solid lines are 95% confidence intervals

PDFs have means equal to the calibrated values, 5.0 and 50, respectively, each with COV equal to 20% (since they are quite uncertain), while γ_t does not change with time to avoid the confounding problem and λ_t drifts with a COV equal to 0.4%. For α_0 , the prior PDF has mean equal to 1 and COV equal to 5% (since we are usually more certain about the initial stiffness scaling parameter), and α_t drifts with a COV equal to 0.4%.

To sum up, there is only one uncertain time-varying parameter for the stiffnesses (i.e., α_t) and one for the damping (i.e., λ_t) to avoid confounding; two parameters are fixed (i.e., ρ_t and δ_t); and γ_t = uncertain time-invariant parameter. For $h_{1,0}, \dots, h_{4,0}$, the prior PDFs have means equal to 0.1 m/s² and COV equal to 20% (since they are highly uncertain), and $h_{1,t}, \dots, h_{4,t}$ fluctuate with a COV equal to 10% to allow rapid change of the prediction errors.

Preliminary Comparison of Two Identification Models

As discussed in an early section, selecting the diagonals of G is a difficult task in general. Nevertheless, we argue here that with the simplified nonlinear degradation model, selecting the diagonals of G is easier. As we have seen, the time-varying linear model must

vary its parameters (stiffnesses and dampings) rapidly to accommodate the nonlinearity (or degradation) that is present in the acceleration data; however, the simplified nonlinear degradation model is able to accommodate the degradation to a certain degree even with its parameters fixed. Therefore in order to accommodate the degradation, the simplified nonlinear degradation model may only need to vary its parameters slowly (except for the uncertainty parameters $h_{1,t}, \dots, h_{4,t}$, which need to vary rapidly). This makes it easier to decide how to choose the diagonals of G in Eq. (28), since we should use relatively small diagonal values (again, not for the diagonal values corresponding to $h_{1,t}, \dots, h_{4,t}$).

Moreover, the correlation between the interstory stiffnesses and dampings is directly considered in the simplified nonlinear degradation model since they depend on the interstory drifts, and the drifts of different stories are, in turn, correlated through the model structure. Lastly, the total number of the active parameters in the simplified nonlinear degradation model is only seven, so the flexibility of the model and the required computation resources are greatly reduced.

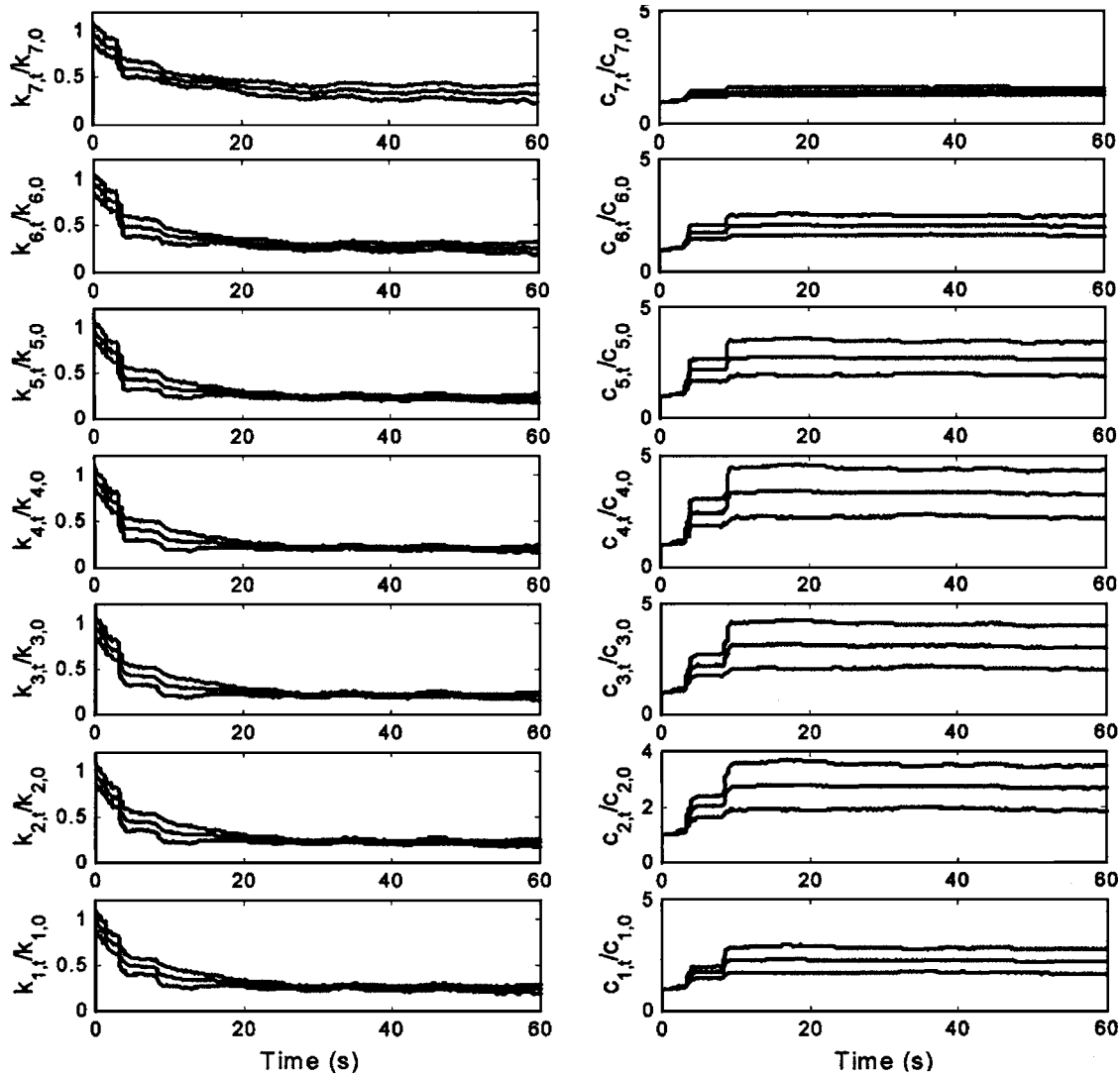


Fig. 9. Estimated stiffnesses (left) and dampings (right) by the simplified nonlinear degradation model; upper and lower lines are the 95% confidence intervals

Results and Discussion

We first convert Eq. (28) to a discrete-time system with a time step of 0.04 s. We then implement PF with $N=200$ and $L=10$ and the importance weight COV threshold=200% using Algorithm 2. If more samples are used in PF than the $N \cdot L=2,000$ samples, there is little change in the estimated means and variances of the unknown parameters, indicating that the results are nearly converged.

Fig. 6 shows the evolution of the estimated system parameters for the simplified nonlinear degradation model (where ρ_t and δ_t are not shown since they are fixed). Fig. 7 shows the predicted output accelerations, i.e., the estimated mean of $p(y_k|D_k)$, from the simplified model as well as the measured accelerations. Compared to Fig. 3, the predicted accelerations can better capture the trend of the measured accelerations, and the corresponding confidence intervals are much smaller, indicating better performance. Similar observations are found for the predicted interstory drifts (see Fig. 8 and compare it to Fig. 4).

Fig. 9 shows the estimated interstory stiffnesses and dampings from the simplified nonlinear degradation model. Compared to Fig. 2, the confidence intervals are much smaller and the correlations between stiffnesses and dampings of different stories are

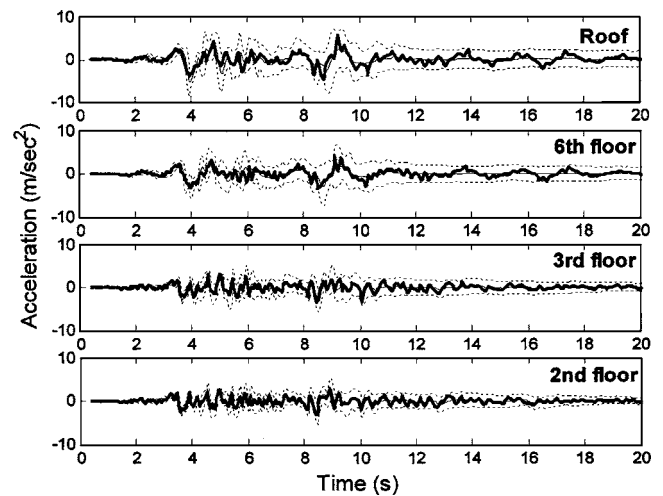


Fig. 10. Thick solid lines are measured accelerations; central thin solid lines are predicted acceleration time histories by the new linear time-varying model; and upper and lower dashed lines are 95% confidence intervals

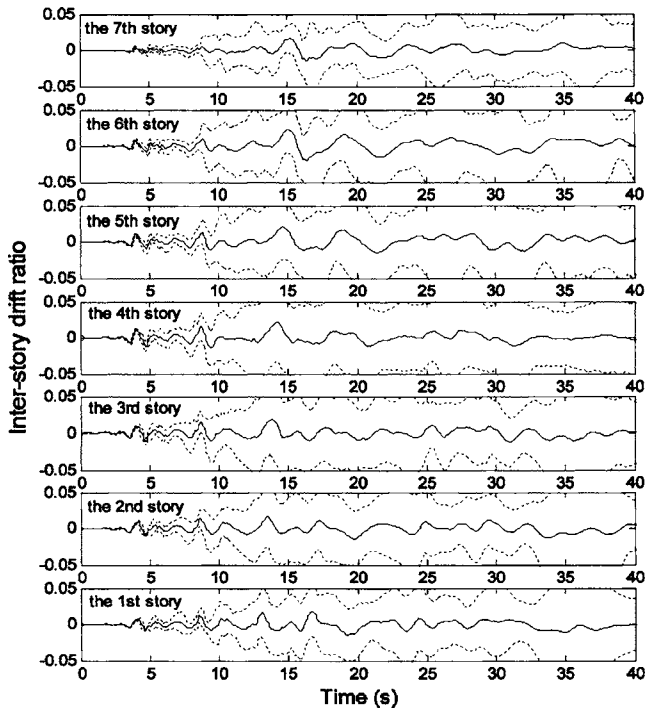


Fig. 11. Central thin solid lines are predicted interstory drift ratios by the new linear time-varying model; upper and lower dashed lines are 95% confidence intervals

more reasonable. The estimated dampings are quite different from those in Fig. 2, in which the estimated dampings almost keep constant and do not show any degradation.

It may be argued that the large confidence intervals (i.e., high uncertainties) obtained from the linear time-varying model are due to the fact that the model contains more free parameters. To examine this argument, let us consider another linear time-varying model in which the seven interstory stiffness parameters

are scaled together as in Eq. (26) so that there is only one time-varying scaling parameter for the stiffnesses. Also, we assume the time-varying damping ratios for all vibrational modes are identical through time. Thus there are only two time-varying parameters in this new linear time-varying model, less than in the simplified nonlinear degradation model. Figs. 10 and 11 show the predicted accelerations and interstory drift ratio time histories with the new linear time-varying model. It can be seen that the confidence intervals are still large although the number of time-varying parameters is less, so the large uncertainties are not completely the consequence of the number of parameters. Then what are the sources of these large uncertainties?

To answer this question, let us go back to the models: recall that there are four uncertainty parameters $h_{1,t}, \dots, h_{4,t}$ that control the amplitude of the prediction-error term $H(\theta_t) \cdot v_t$. These uncertainty parameters are critical: we identify the amplitude of prediction errors by continually updating these parameters. If the identified values of these four parameters are relatively large, then the measured accelerations \hat{Y} do not have a large influence on the state estimation results and the confidence intervals are large. The two linear time-varying models have large prediction errors, so the PF algorithm makes the four uncertainty parameters large; therefore the confidence intervals for these two linear models are also large.

We also implemented the EKF algorithm for the simplified nonlinear degradation model for this case study in Ching et al. (2004). The results from EKF are inconsistent with those from PF, as can be seen by comparing Fig. 12 with Fig. 6. We treat the results from PF as a comparison standard since it gives asymptotically consistent state estimates. Therefore we conclude that EKF does not provide consistent estimates for this real-data case. In Ching et al. (2004), we show that EKF can provide consistent state estimates using “simulated” data from a lightly nonlinear system but not a highly nonlinear one such as the chaotic Lorentz system. This suggests that for such problems, it can be misleading

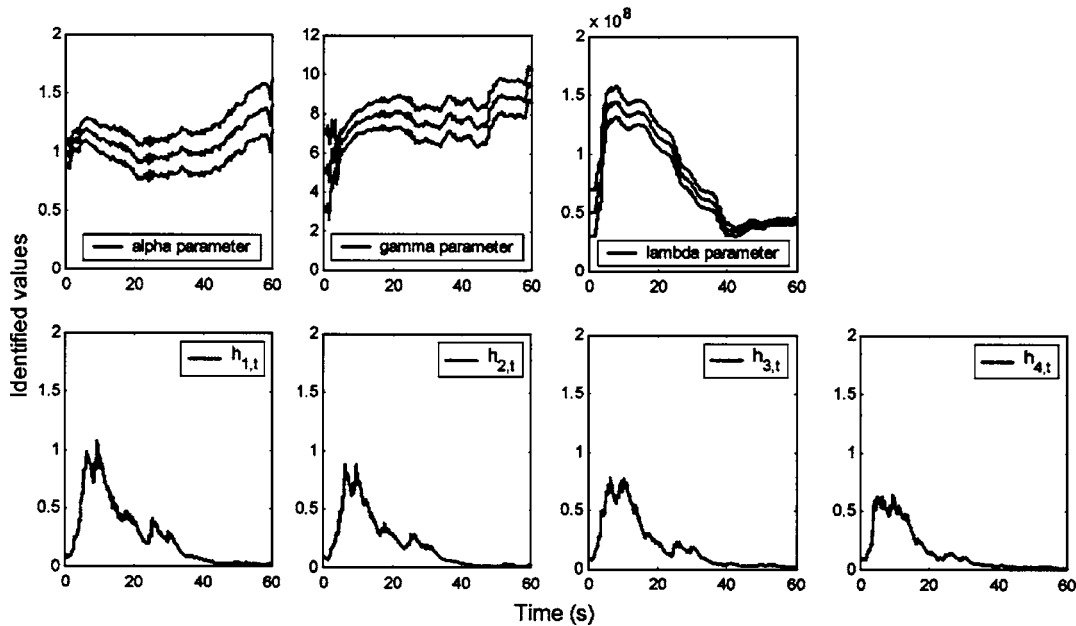


Fig. 12. Estimated system parameters by extended Kalman filter (EKF); upper and lower lines are the 95% confidence intervals (no confidence intervals are provided by EKF for the uncertainty parameters)

if we only consider the first two moments of the system state and unknown parameters; instead, it may be essential to also consider the entire posterior PDFs.

Discussion of Case Study Results

Besides the need to choose a good algorithm among different available algorithms (e.g., EKF, PF, etc.), the results from the case study suggest that selecting a good stochastic identification model is also essential. As we have seen, even with a good estimation algorithm (i.e., PF), the estimation results can be unsatisfactory if the time-varying linear model is employed. A better strategy is to take a nonlinear model that can roughly capture the degrading behavior of the building and select an identification model based on that model. In fact, this is the main idea for developing the simplified nonlinear degradation model. As a result, a successful state-estimation technique should at least consist of (1) a good model for the underlying processes and (2) a good estimation algorithm. For the Van Nuys hotel case study, it is preferable to use the simplified nonlinear degradation model with the PF algorithm.

Although the simplified nonlinear degradation model performs satisfactorily for the case study, it is not necessarily the best model to use. There may exist other degradation models that are rigid in the sense defined earlier and yet able to fit the acceleration well. To find a better model than the simplified nonlinear degradation model involves model class selection and the bias-variance trade-off, and we will not pursue this in this paper. A Bayesian approach to model class selection for dynamical systems is given by Beck and Yuen (2004).

Although not shown in this paper, the maximum interstory drift predicted by the finite element model developed at Caltech is at the third story, although the actual damage to the south frame in the 1994 Northridge earthquake was primarily to the beam-column joints at the top of the fourth-story level (Beck et al. 2002, Section 5.3.3). However, the interstory drift predicted by the simplified nonlinear degradation model reaches its maximum at the fourth story (see Fig. 8). Note that the former results do not incorporate the measured accelerations at the second, third, and sixth floors and the roof, while the latter ones do. It is likely that the interstory drift prediction from the latter has been improved because the information from these acceleration data is incorporated.

Conclusion

We have presented a real-time Bayesian state-estimation algorithm that employs a stochastic simulation approach called the particle filter (PF). Strong-motion records obtained in a seven-story Van Nuys hotel during the 1994 Northridge earthquake are studied using two different identification model classes (a time-varying linear one and a time-varying nonlinear degradation one) by applying the PF algorithm. The results show that the simplified nonlinear model significantly outperforms the time-varying linear model, which produces some unreasonable estimation results, and that the PF algorithm performs much better than EKF. It is concluded that the following two factors are keys to a successful real-time Bayesian state estimation: (1) an appropriate state-estimation algorithm and (2) an appropriate stochastic identification model.

References

- Alspach, D. L., and Sorenson, H. W. (1972). "Nonlinear Bayesian estimation using Gaussian sum approximations." *IEEE Trans. Autom. 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, Calif. (<http://resolver.caltech.edu/CaltechEERL:1978.EERL-78-01>).
- Beck, J. L., et al. (2002). "Impact of seismic risk on lifetime property values." *Final Rep., CUREe-Kajima Joint Research Program Phase IV*, Consortium of Universities for Research in Earthquake Engineering, Richmond, Calif. <http://resolver.caltech.edu/CaltechEERL:2002.EERL-2002-04>.
- Beck, J. L., and Yuen, K.-V. (2004). "Model selection using response measurements: Bayesian probabilistic approach." *J. Eng. Mech.*, 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 Rep. 2004-01*, California Institute of Technology, Pasadena, Calif. (<http://caltecheerl.library.caltech.edu/>).
- Doucet, A., and Andrieu, C. (2000). "Particle filtering for partially observed Gaussian state space models." *Rep. CUED/F-INFENG/TR393*, Dept. of Engineering, Cambridge Univ., Cambridge, U.K.
- Doucet, A., de Freitas, J. F. G., and Gordon, N. (2000). "Introduction to sequential Monte Carlo methods." *Sequential Monte Carlo methods in practice*, A. Doucet, J. F. G. de Freitas, and N. J. Gordon, eds., Springer-Verlag, Berlin.
- Doucet, A., and Godsill, S. (1998). "On sequential simulation-based methods for Bayesian filtering." *Rep. CUED/F-INFENG/TR310*, Dept. of Engineering, Cambridge Univ., Cambridge, U.K.
- Ghanem, R., and Shinozuka, M. (1995). "Structural-system identification. I: Theory." *J. Eng. Mech.*, 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." *IEE Proc. F, Radar Signal Process.*, 140(2), 107–113.
- Hoshiya, M., and Saito, E. (1984). "Structural identification by extended Kalman filter." *J. Eng. Mech.*, 110(12), 1757–1770.
- Islam, M. S. (1996a). "Analysis of the response of an instrumented 7-story nonductile concrete frame building damaged during the Northridge earthquake." *Proc., 1996 Annual Meeting of the Los Angeles Tall Buildings Structural Council*, Los Angeles.
- Islam, M. S. (1996b). *Holiday Inn, 1994 Northridge earthquake buildings case study project proposition 122: Product 3.2*, Seismic Safety Commission, Sacramento, Calif., 189–233.
- Islam, M. S., Gupta, M., and Kunnath, B. (1998). "Critical review of the state-of-the-art analytical tools and acceptance criterion in light of observed response of an instrumented nonductile concrete frame building." *Proc., Sixth U.S. National Conf. on Earthquake Engineering*, Seattle, Wash.
- Jazwinski, A. H. (1970). *Stochastic processes and filtering theory*, Academic Press, New York.
- Jennings, P. C. (1971). "Engineering features of the San Fernando Earthquake of February, 9, 1971." *EERL Rep. 71-02*, California Institute of Technology, Pasadena, Calif.
- 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 Trans. Autom. Control*, 45(3), 477–482.
- Kalman, R. E. (1960). "A new approach to linear filtering and prediction problems." *J. Basic Eng.*, 82D, 35–45.
- Kalman, R. E., and Bucy, R. S. (1961). "New results in linear filtering and prediction problems." *J. Basic Eng.*, 83D, 95–108.

- Kitagawa, G. (1996). "Monte Carlo filter and smoother for non-Gaussian nonlinear state space models." *J. Comput. Graph. Stat.*, 5, 1–25.
- Koh, C. G., and See, L. M. (1994). "Identification and uncertainty estimation of structural parameters." *J. Eng. Mech.*, 120(6), 1219–1236.
- Li, Y. R., and Jirsa, J. O. (1998). "Nonlinear analyses of an instrumented structure damaged in the 1994 Northridge earthquake." *Earthquake Spectra*, 14(2), 245–264.
- Lin, C. C., Soong, T. T., and Natke, H. G. (1990). "Real-time system identification of degrading structures." *J. Eng. Mech.*, 116(10), 2258–2274.
- Liu, J. S., and Chen, R. (1998). "Sequential Monte Carlo methods for dynamical systems." *J. Am. Stat. Assoc.*, 93, 1032–1044.
- Maruyama, O., and Hoshiya, M. (2003). "Nonlinear filters using Monte Carlo integration for conditional random fields." *Proc., 9th Int. Conf. on Applications of Statistics and Probability in Civil Engineering*, San Francisco.
- Sato, T., and Qi, K. (1998). "Adaptive H_{∞}^+ filter: Its application to structural identification." *J. Eng. Mech.*, 124(11), 1233–1240.
- Scholl, R. E., Kustu, O., Perry, C. L., and Zanetti, J. M. (1982). "Seismic damage assessment for high-rise buildings." *URS/JAB 8020*, URS/John A. Blume & Associates, Engineers, San Francisco.
- Shinozuka, M., and Ghanem, R. (1995). "Structural-system identification. II: Experimental verification." *J. Eng. Mech.*, 121(2), 265–273.
- Smyth, A. W., Masri, S. F., Chassiakos, A. G., and Caughey, T. K. (1999). "On-line parametric identification of MDOF nonlinear hysteretic systems." *J. Eng. Mech.*, 125(2), 133–142.
- van der Merwe, R., de Freitas, N., Doucet, A., and Wan, E. A. (2000). "The unscented particle filter." *Rep. CUED/F-INFENG/TR380*, Dept. of Engineering, Cambridge Univ., Cambridge, U.K.
- van der Merwe, R., and Wan, E. A. (2003). "Gaussian mixture sigma-point particle filters for sequential probabilistic inference in dynamic state-space models." *Proc., IEEE Int. Conf. on Acoustics, Speech and Signal Processing (ICASSP)*, Hong Kong.
- Wan, E. A., and van der Merwe, R. (2000). "The unscented Kalman filter for nonlinear estimation." *Proc., Symposium 2000 on Adaptive Systems for Signal Processing, Communication and Control*, IEEE, Lake Louise, Alberta, Canada.
- Yoshida, I., and Sato, T. (2002). "Health monitoring algorithm by the Monte Carlo filter based on non-Gaussian noise." *J. Nat. Disaster Sci.*, 24(2), 101–107.
- Yun, C. B., and Shinozuka, M. (1980). "Identification of nonlinear dynamic systems." *J. Struct. Mech.*, 8(2), 187–203.