Wind and Structures, Vol. 6, No. 2 (2003) 107-126
DOI: http://dx.doi.org/10.12989/was.2003.6.2.107

The use of linear stochastic estimation for the reduction of data in the NIST aerodynamic database

Y. Chen[†], G. A. Kopp[‡] and D. Surry[‡†]

Alan G. Davenport Wind Engineering Group, Boundary Layer Wind Tunnel Laboratory, The University of Western Ontario, London, Ontario, N6A 5B9, Canada

(Received August 15, 2002, Accepted February 10, 2003)

Abstract. This paper describes a simple and practical approach through the application of Linear Stochastic Estimation (LSE) to reconstruct wind-induced pressure time series from the covariance matrix for structural load analyses on a low building roof. The main application of this work would be the reduction of the data storage requirements for the NIST aerodynamic database. The approach is based on the assumption that a random pressure field can be estimated as a linear combination of some other known pressure time series by truncating nonlinear terms of a Taylor series expansion. Covariances between pressure time series to be simulated and reference time series are used to calculate the estimation coefficients. The performance using different LSE schemes with selected reference time series is demonstrated by the reconstruction of structural load time series in a corner bay for three typical wind directions. It is shown that LSE can simulate structural load time series accurately, given a handful of reference pressure taps (or even a single tap). The performance of LSE depends on the choice of the reference time series, which should be determined by considering the balance between the accuracy, data-storage requirements and the complexity of the approach. The approach should only be used for the determination of structural loads, since individual reconstructed pressure time series (for local load analyses) will have larger errors associated with them.

Key words: aerodynamic database; data reduction; Linear Stochastic Estimation; low buildings; pressure time series; reconstruction; structural loads.

1. Introduction

Due to the advances of electronically-scanned pressure measurement technology, information storage and computational capacities, increasing attention has been paid to *database-assisted design* (DAD) for wind effects and wind loading on low-rise buildings (e.g., Simiu and Stathopoulos 1997, Whalen *et al*. 1998, Rigato *et al*. 2001, Chen *et al*. 2002a, 2002b, 2003). In recognition of the fact that DAD would be more realistic and more risk-consistent than conventional wind codes, and thus lead to more economical designs, it has been proposed as a future code alternative (Rigato *et al*. 2001). However, there are some concerns associated with the use of aerodynamic databases, such as data-storage problems, on one hand, and a limited number of building configurations, on the other. Because aerodynamic databases consist of long-term time series of pressures acquired at densely-

† Post-doctoral Fellow
‡ Associate Professor and Canada Research Chair in Enivornmental Fluid Mechanics
‡† Professor and Research Director

distributed tap locations covering the entire building envelope for many different conditions (e.g., wind direction, building geometry, and flow conditions), the demand on storage space is extreme. It is estimated that the database used for codification will result in about 6 Gigabits of storage if it contains the same building configurations as the original 1970s data (i.e., three building heights with three roof slopes in one terrain) (Stathopoulos 1979). Our previous work (Chen *et al*. 2002b, 2003) has focused on methods to interpolate the basic data to other building configurations for which no data exist. Although the huge quantity of pressure time series data can be stored on CDs, this will have a negative impact on the practicality and feasibility of DAD. Therefore, efforts are needed for exploring accurate techniques which can be used to effectively reduce the size of the aerodynamic database. Linear Stochastic Estimation (LSE), based on the covariance matrix, offers such an approach.

Stochastic estimation, generally speaking, is an approximation of a random variable in terms of some other random variables by truncating higher-order terms of a Taylor series expansion. For a surface pressure field, $P(x, y; t)$, the stochastic estimate of the pressure field is the best mean square estimate of $P(x, y; t)$ given some reference data, $P_{ref}(t)$. This is called the *conditional average* of $P(x, y; t)$ given $P_{ref}(t)$, denoted as $\mathrm{E}(P|P_{ref})$ (Adrian 1975). Often, linear stochastic estimation is used to estimate a random variable as a linear combination of some known reference variables.

One advantage of this approach is that it has the ability to reconstruct the whole random field by using only a few reference variables in conjunction with the covariance matrix. Clearly, this indicates that LSE has the potential to reduce the data-storage requirements for an aerodynamic database by the storing covariance matrix and some selected reference time series. In addition, storing covariances could allow simpler usage by designers whose needs are less sophisticated than those using the database for time series analysis. In other words, the reconstructed database associated with covariance matrix could be more flexible. For example, the stored covariance matrices could be directly used with Load-Response-Correlation (LRC) method developed by Kasperski (1992) and Holmes and Syme (1994) to calculate expected effective static load distributions corresponding to a peak structural load effect (e.g., bending moment, support reaction). Furthermore, this could remove the need to estimate the peak factors required with the LRC method and would, therefore, yield the true advantage of having the aerodynamic database.

As a simple and practical approach, LSE was first applied to the field of turbulent flows by Adrian (1975) in order to estimate the large-scale organized (i.e., coherent) turbulent structures. Since its introduction in turbulence, LSE has been successfully applied to simulate a variety of turbulent flows, such as wakes, pipe flows, shear layers, and jets. These examples include the estimation of conditional eddies (i.e., the flow field described by conditional averages) (Adrian 1975, Tung and Adrian 1980, Adrian and Moin 1988, Adrian *et al*. 1989), and the reconstruction of velocity fields (Druault *et al*. 1999, Delville *et al*. 2000, Péneau *et al*. 2000) by using the "conditional information" specified at one or more reference locations. A review of LSE in the analysis of turbulent velocity fields has been presented by Adrian (1994).

The main objective of the present work is to develop a simple and practical approach which is capable of reconstructing wind-induced pressure time series for structural load analyses on the envelope of low buildings, given only a handful of reference pressure taps. In other words, the LSE approach can be viewed as a means to predict the whole pressure field by knowing the state of the pressure field at a limited number of tap locations based on the two-point spatial covariance. This approach has the potential to reduce the size of aerodynamic databases, providing a flexible database for numerical structural analysis.

2. Background

2.1. Review of linear stochastic estimation

LSE is described in (Adrian *et al*. 1975, 1988, 1989, 1994), but for convenience, the derivation is presented here. LSE is based on the assumption that a random field can be considered as a linear combination of some other random variables that are known and correlated. Applying this to a surface pressure field (herein, non-dimensionalized pressure coefficients are used), the pressure time series at a tap location (assuming the ergodicity of the time series) can be estimated with pressure time series at some reference tap locations, $C_{p_{ref}}(x_i, y_i;\ t)$ (also denoted as $C_{p_{i,ref}}(t)$), as:

$$\hat{C}_p(x, y;\ t) \ = \ \sum_{i\,=\,1}^{N} b_i(x, y) C_{p_{ref}}(x_i, y_i;\ t) + b_0(x, y) \tag{1}$$

where $\hat{C}_p(x, y;\ t)$ is the linear estimate of pressure time series $C_p(x, y;\ t)$ at a location (x, y), $b_i(x, y)$ is the estimation coefficient associated with the ith reference pressure time series, $b_0(x, y)$ is a bias term which is required in the present work since pressure time series with non-zero mean values are considered directly (if only fluctuations are considered, b_0 is equal to zero), and N is the total number of the reference pressure time series. In LSE, the most important step is to find the estimation coefficients, b_i, i=0, 1, …, N, by using two-point correlations, as shown below.

LSE is developed from the linear estimate of the conditional average of a random field. Accordingly, given the state of the pressure field at some reference pressure taps, $C_{p_{i,ref}}(t)$, i=1, 2, …, N (represented by an event data vector $\mathbf{C}_{\mathbf{p_{ref}}}$), the pressure field can be estimated by its conditional average,

$$\begin{aligned} \hat{C}_p(x, y;\ t) \ &= \text{linear estimate of } \ E(C_p|\mathbf{C}_{\mathbf{p_{ref}}}) \\ &= \text{linear estimate of } \ f(\mathbf{C}_{\mathbf{p_{ref}}}) \end{aligned} \tag{2}$$

where $E(C_p|\mathbf{C}_{\mathbf{p_{ref}}})$ is the conditional average of the pressure field given a set of event data $\mathbf{C}_{\mathbf{p_{ref}}}$, and can be viewed as a nonlinear function for $\mathbf{C}_{\mathbf{p_{ref}}}$, denoted as $f(\mathbf{C}_{\mathbf{p_{ref}}})$. By expanding $f(\mathbf{C}_{\mathbf{p_{ref}}})$ in a Taylor series about $\mathbf{C}_{\mathbf{p_{ref}}}=\overline{C}_{p_{ref}}$ as:

$$E(C_p|\mathbf{C}_{\mathbf{p_{ref}}}) \ = \ f(\mathbf{C}_{\mathbf{p_{ref}}}) \ = \ f(\overline{C}_{p_{ref}}) + f'(\overline{C}_{p_{ref}})(\mathbf{C}_{\mathbf{p_{ref}}} - \overline{C}_{p_{ref}}) + ... \tag{3}$$

or in a tensor form,

$$E(C_p|\mathbf{C}_{\mathbf{p_{ref}}}) \ = \ b_0 + b_i C_{p_{i,ref}} + d_{ij} C_{p_{i,ref}} C_{p_{j,ref}} + ... \tag{4}$$

where $f'(\bullet)$ is a derivative function, b_i, d_{ij} are estimation coefficients. By truncating the nonlinear terms of the Taylor series expansion (i.e., only using the constant and first-order terms), the linear stochastic estimate of the conditional average of the pressure field is derived, as expressed in Eq. (1). In this case, LSE is numerically equivalent to linear nonhomogeneous mean-square estimation method, described by Papoulis (1984) or the multi-variable linear regression model discussed by Walpole and Myers (1985), as briefly reviewed below.

The estimation coefficients, b_i, $i=0, 1, \ldots, N$, are determined by minimizing the mean square value (ε) of the resulting error (e). The mean square value is defined as:

$$\varepsilon = \overline{e^2} = \overline{[\hat{C}_p(x, y; t) - C_p(x, y; t)]^2}$$
$$= \frac{1}{M}\sum_{t=1}^{M}\left\{\left[b_0 + \sum_{i=1}^{N} b_i C_{p_{ref}}(x_i, y_i; t)\right] - C_p(x, y; t)\right\}^2 \tag{5}$$

where $e = \hat{C}_p(x, y; t) - C_p(x, y; t)$, and M is the length of pressure time series record.

According to the orthogonality principle (Papoulis 1984), the mean square error ε is minimum if $b_i (i=0, 1, \ldots, N)$ are such that the resulting error e is statistically orthogonal to each of the data. In other words, ε is minimum if the partial derivative of ε with respect to each b_i is equal to zero,

$$\frac{\partial \varepsilon}{\partial b_i} = 0 \qquad i = 0, 1, \ldots, N \tag{6}$$

Thus, expanding Eq. (4) yields:

$$\frac{\partial \varepsilon}{\partial b_0} = \frac{2}{M}\sum_{t=1}^{M}\left\{\left[b_0 + \sum_{i=1}^{N} b_i C_{p_{ref}}(x_i, y_i; t)\right] - C_p(x, y; t)\right\} = 0 \tag{7}$$

and

$$\frac{\partial \varepsilon}{\partial b_k} = \frac{2}{M}\sum_{t=1}^{M}\left\{\left[\left(b_0 + \sum_{i=1}^{N} b_i C_{p_{ref}}(x_i, y_i; t)\right) - C_p(x, y; t)\right] C_{p_{ref}}(x_k, y_k; t)\right\} = 0 \tag{8}$$

By setting $k=1, 2, \ldots, N$, Eqs. (7) and (8) can be written in a $(N+1)\times(N+1)$ system of linear algebraic equations as follows:

$$\begin{bmatrix} 1 & \overline{C}_{p_{1,ref}} & \overline{C}_{p_{2,ref}} & \cdots & \overline{C}_{p_{N,ref}} \\ \overline{C}_{p_{1,ref}} & \overline{C_{p_{1,ref}} C_{p_{1,ref}}} & \overline{C_{p_{1,ref}} C_{p_{2,ref}}} & \cdots & \overline{C_{p_{1,ref}} C_{p_{N,ref}}} \\ \overline{C}_{p_{2,ref}} & \overline{C_{p_{2,ref}} C_{p_{1,ref}}} & \overline{C_{p_{2,ref}} C_{p_{2,ref}}} & \cdots & \overline{C_{p_{2,ref}} C_{p_{N,ref}}} \\ \cdots & \cdots & \cdots & \overline{C_{p_{i,ref}} C_{p_{j,ref}}} & \cdots \\ \overline{C}_{p_{N,ref}} & \overline{C_{p_{N,ref}} C_{p_{1,ref}}} & \overline{C_{p_{N,ref}} C_{p_{2,ref}}} & \cdots & \overline{C_{p_{N,ref}} C_{p_{N,ref}}} \end{bmatrix} \begin{Bmatrix} b_0 \\ b_1 \\ b_2 \\ \cdots \\ b_N \end{Bmatrix} = \begin{Bmatrix} \overline{C_p} \\ \overline{C_p C_{p_{1,ref}}} \\ \overline{C_p C_{p_{2,ref}}} \\ \cdots \\ \overline{C_p C_{p_{N,ref}}} \end{Bmatrix} \tag{9}$$

Eq. (9) can be transformed into a $N\times N$ system of equations based on the covariance matrix by multiplying the first row by $\overline{C}_{p_{1,ref}}$, $\overline{C}_{p_{2,ref}}$, …, $\overline{C}_{p_{N,ref}}$, respectively, and then subtracting the remaining equations, respectively. This yields:

$$\begin{bmatrix} c_{11} & c_{12} & \dots & c_{1N} \\ c_{21} & c_{22} & \dots & c_{2N} \\ \dots & \dots & c_{ij} & \dots \\ c_{N1} & c_{N2} & \dots & c_{NN} \end{bmatrix} \begin{Bmatrix} b_1 \\ b_2 \\ \dots \\ b_N \end{Bmatrix} = \begin{Bmatrix} c_{10} \\ c_{20} \\ \dots \\ c_{N0} \end{Bmatrix} \tag{10}$$

or,

$$[\mathbf{C}]\{\mathbf{B}\} = \{\mathbf{C_0}\} \tag{11}$$

(Eq. (10) is called the Yule-Walker equation (Papoulis 1984)) where each element, c_{ij} in $[\mathbf{C}]$, $i, j=1, 2, ..., N$, is the spatial covariance (zero time lag) between the reference time series at locations i and j, which is calculated with their fluctuations as follows:

$$\begin{aligned} c_{ij} &= \overline{C'_{p_{ref}}(x_i, y_i; t) C'_{p_{ref}}(x_j, y_j; t)} \\ &= \overline{C_{p_{ref}}(x_i, y_i; t) C_{p_{ref}}(x_j, y_j; t)} - \overline{C}_{p_{ref}}(x_i, y_i) \overline{C}_{p_{ref}}(x_j, y_j) \end{aligned} \tag{12}$$

(if $i=j$, it is the variance) where the pressure fluctuation is defined as:

$$C'_p(x, y; t) = C_p(x, y; t) - \overline{C}_p(x, y) \tag{13}$$

c_{i0} in $\{\mathbf{C_0}\}$, $i=1, 2, ..., N$, is the covariance between the i^{th} reference pressure time series and the pressure time series to be simulated, which is given by:

$$\begin{aligned} c_{i0} &= \overline{C'_{p_{ref}}(x_i, y_i; t) C'_p(x, y; t)} \\ &= \overline{C_{p_{ref}}(x_i, y_i; t) C_p(x, y; t)} - \overline{C}_{p_{ref}}(x_i, y_i) \overline{C}_p(x, y) \end{aligned} \tag{14}$$

By solving Eq. (11), the estimation coefficients $b_i(x, y)$ for simulating $C_p(x, y; t)$, can be easily computed:

$$\{\mathbf{B}\} = [\mathbf{C}]^{-1}\{\mathbf{C_0}\} \tag{15}$$

Then, by substituting $b_i (i=1, 2, ..., N)$ into the first row of Eq. (9), b_0 is obtained:

$$b_0 = \overline{C}_p - \sum_{i=1}^{N} \overline{C}_{p_{ref}}(x_i, y_i) b_i \tag{16}$$

Finally, by substituting the calculated estimation coefficients b_i (from Eqs. (15) and (16)) into Eq. (1), the pressure field can be estimated.

It is noted that although LSE and the linear mean-square estimation method or the multi-variable linear regression method are numerically equivalent, there still exists the difference which lies in the interpretation. Adrian (1994) states that "the mean square error of the linear estimate of $C_p(x, y)$ must be large when $C_p(x, y)$ is uncorrelated with $\mathbf{C_{p_{ref}}}$, due, for example, to a large separation

between (x, y) and the location of the event data, the error of the linear stochastic estimate of $E(C_p|\mathbf{C}_{\mathbf{p_{ref}}})$ may be small because $E(C_p|\mathbf{C}_{\mathbf{p_{ref}}})$ also vanishes as the separation becomes large".

2.2. Reconstruction methodology

Fig. 1 describes the framework for reconstructing individual pressure time series for the analysis of structural loads. It is shown that once the estimation coefficients b_i, i=0, 1, …, N, have been determined by using the covariance information in the pressure field (according to Eqs. (15) and (16)), pressure time series at the required location can be estimated according to Eq. (1), given the reference pressure taps.

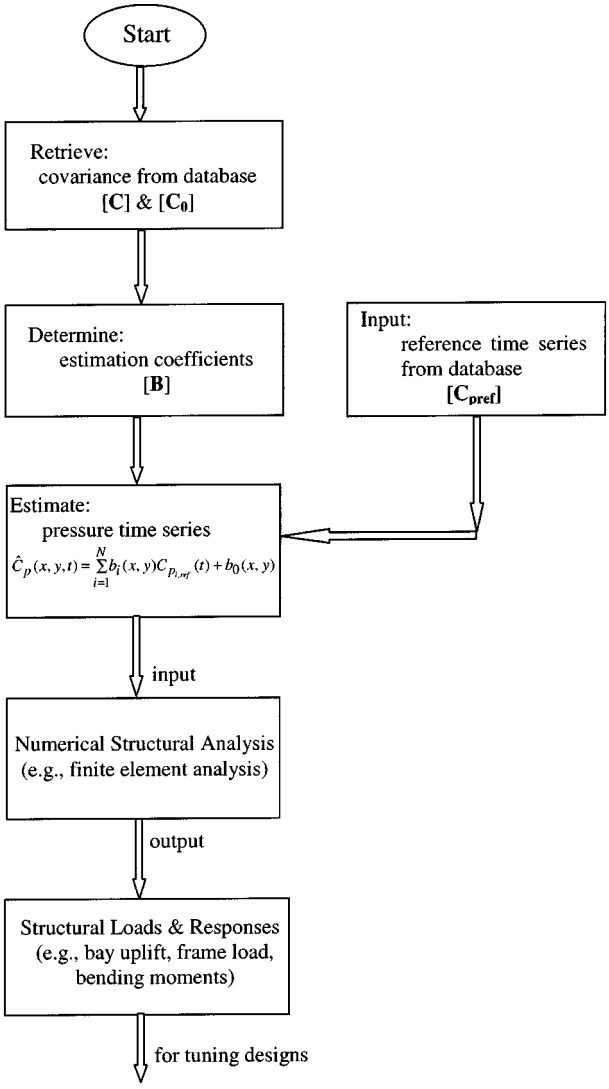

Fig. 1 Reconstruction methodology using the LSE

These reconstructed pressure time series can then be used to calculate structural load time series on a desired region, or they can be input to a structural analysis program (e.g., Whalen *et al*. 1998). Bay uplift, frame uplift and bending moments are typical examples of the structural loads of concern in the design of low buildings for wind. They are calculated from the pressure time series as, for example,

$$\hat{F}_1(t) = \frac{1}{A}\sum_i (\hat{C}_p(x_i, y_i;\ t) \times \Delta A_i) \tag{17}$$

$$\hat{F}_2(t) = \frac{1}{A}\sum_i (\hat{C}_p(x_i, y_i;\ t) \times \Delta A_i \times I_i) \tag{18}$$

$$\hat{F}_3(t) = \frac{1}{W \times A}\sum_i (\hat{C}_p(x_i, y_i;\ t) \times \Delta A_i \times L_i) \tag{19}$$

where $\hat{F}_1(t)$, $\hat{F}_2(t)$, $\hat{F}_3(t)$ denote the bay uplift, frame uplift, and bending moment time series, respectively. $\hat{C}_p(x_i, y_i;\ t)$ is the linear estimate of pressure time series at location i, A is the total area of a roof zone, W is the building width, ΔA_i is the tributary area of tap i, I_i is the corresponding influence line coefficient, L_i is the bending moment arm of the simulated pressure time series at location i, based on the distance from the ridge or eave edges. It is noted that the structural loads considered in this study are induced only from wind pressure time series. The bending moments have been non-dimensionalized by building width.

3. Wind tunnel experimental data

The experimental database of pressure time series was acquired on a 1:100 scale gable-roofed generic building model in the Boundary Layer Wind Tunnel II at the University of Western Ontario as a contribution to the NIST aerodynamic database. The building has a rectangular plan area of 80 ft (24.4 m) by 125 ft (38.1 m) with a roof slope of 1 in 12 and a ridge parallel to the long wall. Four eave heights, 16 ft, 24 ft, 32 ft and 40 ft (4.9 m, 7.3 m, 9.8 m and 12.2 m), were employed in the wind tunnel tests. The roof dimensions, pressure tap layout and the definition of wind direction are shown in Fig. 2. A total of 665 pressure taps were instrumented over the entire surface of the building model with 335 pressure taps on the roof.

Pressure time series were measured for four roof heights, 37 approaching wind angles, and two upstream terrains with a high-speed solid-state pressure scanning system. The 37 wind angles were between 180° and 360° in increments of 5°. Two target upstream terrains (open country and suburban) were modeled in the wind tunnel model tests. The characteristics of the wind tunnel boundary layer flow matched the Exposure C (open country) and Exposure B (suburban) described in ASCE 7-98 (2000). The simulated exposures have equivalent roughness lengths, z_0, of 0.03 m and 0.3 m, respectively. The pressure signals were sampled at 500 Hz for 100 seconds and were measured essentially simultaneously. Assuming that the wind tunnel/full-scale velocity ratio is 1:3, the corresponding full-scale sampling frequency is 15 Hz. Each time series record (of 50000 data points) is then equivalent to 56 minutes in full-scale. Pressure time series in the aerodynamic database were corrected for residual non-simultaneity and were digitally low-pass filtered at 200 Hz. The reference wind tunnel speed for the measurements was 45 ft/s (13.7 m/s). Typically, mean eave

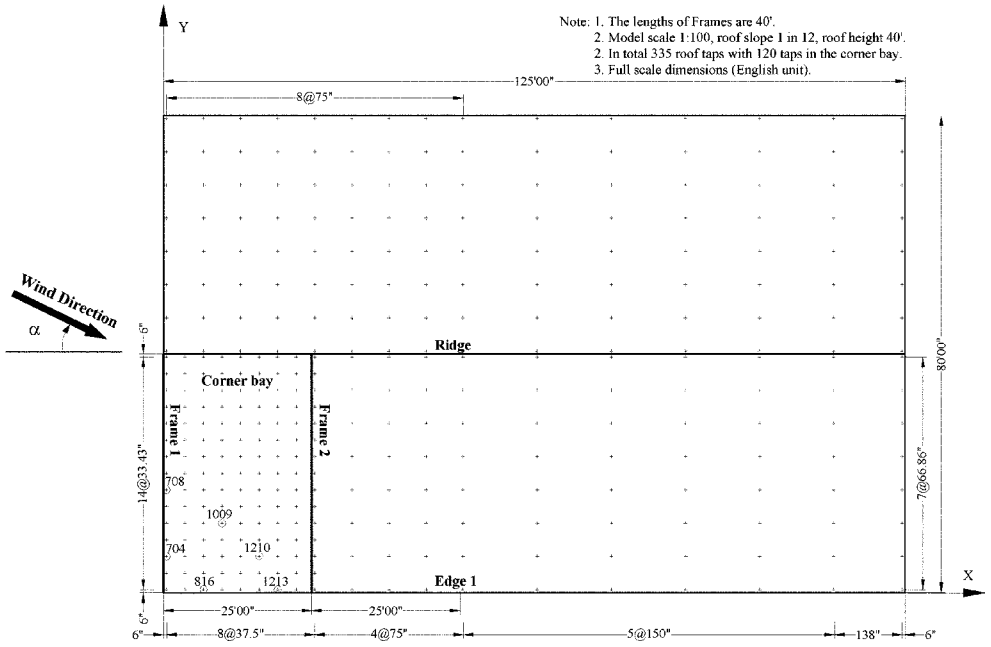

Fig. 2 Roof surface of the low building model (full-scale dimensions)

height wind speeds were about 64% of this. All pressure data used in this study were expressed in non-dimensional pressure coefficients referenced to the mean dynamic pressure in the uniform flow at a reference height above the wind tunnel boundary layer as:

$$C_p(t) = \frac{p(t) - \bar{p}_\infty}{\frac{1}{2}\rho \bar{V}_{ref}^2} \tag{20}$$

where $p(t)$ is the pressure measured on the building surface, ρ is the air density, and $\bar{V}_{ref}$ is the approaching mean wind speed recorded at the reference height. The pressure coefficients can be easily transformed into full-scale wind pressures, given the upstream dynamic pressures at the reference height. Further details about the wind tunnel experiments can be obtained in Ho *et al*. (2002).

In this study, the set of experimental pressure time series on a corner bay of the building model with a full-scale eave height of 40 ft (12.2 m) and three typical wind directions (270°, 320° and 360°) were employed as example studies to evaluate the performance for the simulation of structural load time series by using the LSE. The corner bay was chosen since this is the most interesting region because of the strong vortices that often occur due to flow separation and turbulence.

4. Results and discussion

4.1. Cases studied

In LSE, there can be different choices for selecting reference time series in order to reconstruct the pressure field. On one hand, the greater the number of reference time series that are used, the

Table 1 Reference pressure tap locations for the four LSE schemes

Case	Reference Pressure Taps
LSE-1	Tap #704
LSE-2	Taps #704 & #816
LSE-4	Taps #704, #816, #708 & #1213
LSE-5	Taps #704, #816, #708, #1213 & #1009

better the reconstruction performance LSE will achieve; on the other hand, the approach will become more complex and there will be less data reduction. Thus, one of the objectives is to make the number of the reference time series as small as possible. This will find its usefulness when considering the data-storage problems in the development of the aerodynamic database. The choice of reference time series should be determined by considering a balance between accuracy, data-storage and complexity. This is problem-dependent and done by trial-and-error since there are a large variety of parameters involved in wind tunnel tests. The general guideline is to select those which are strongly correlated with the target pressure field, and can dominate or characterize fluctuating features of the target pressure field.

In this study, four schemes of reference time series were investigated, as listed in Table 1. Herein, the notation LSE-*i* represents a set of *i* reference time series (see Fig. 2 for the locations of the selected taps). It is expected that the reference time series chosen from the separated flow regions can simulate the adjacent individual pressures more accurately than the far field pressure taps due to larger correlations. Fig. 3 gives an example of the distribution of the correlation coefficient relative to one reference tap (#704) for a cornering wind. It can be seen that the reference tap is generally more highly correlated with the adjacent pressures which characterize the fluctuating features of structural load time series.

Following the computational steps illustrated in Fig. 1, individual pressure time series at any required locations and the corresponding structural loads can be reconstructed using different reference time series. Good reconstruction requires that the simulated structural load time series preserve the key statistical and probabilistical features of the experimental data with an acceptable accuracy (e.g., usually a prediction error less than, say, 10%). Among these statistics, mean values, rms values, peaks, correlations, autospectra are the more important measures of the basic quality of time series reconstruction. Skewness and kurtosis coefficients which are the third and fourth statistical moments about the mean value respectively, are used to measure the non-Gaussian features of the time series.

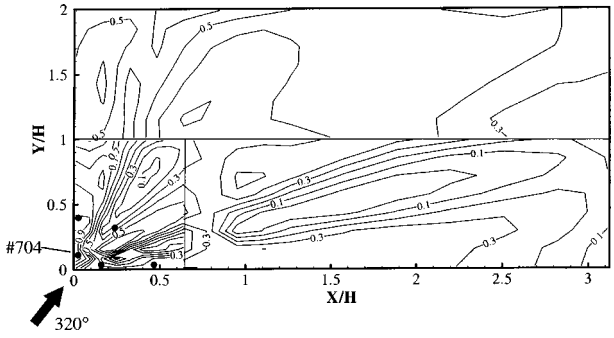

Fig. 3 Contours of the correlation coefficient relative to tap #704 in open country terrain for $\alpha = 320^{o}$

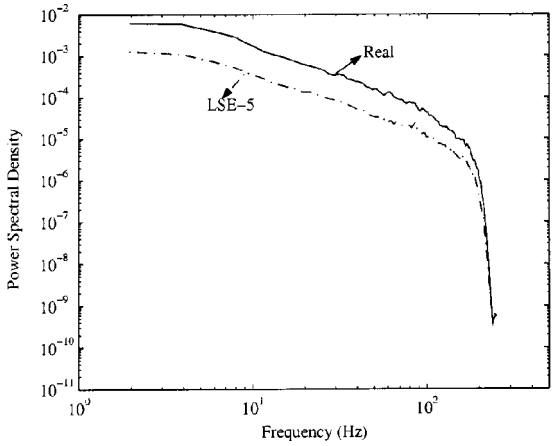

Fig. 4 Autospectra of the measured and reconstructed pressure time series of tap #1210 in open country terrain for $\alpha=320^{o}$

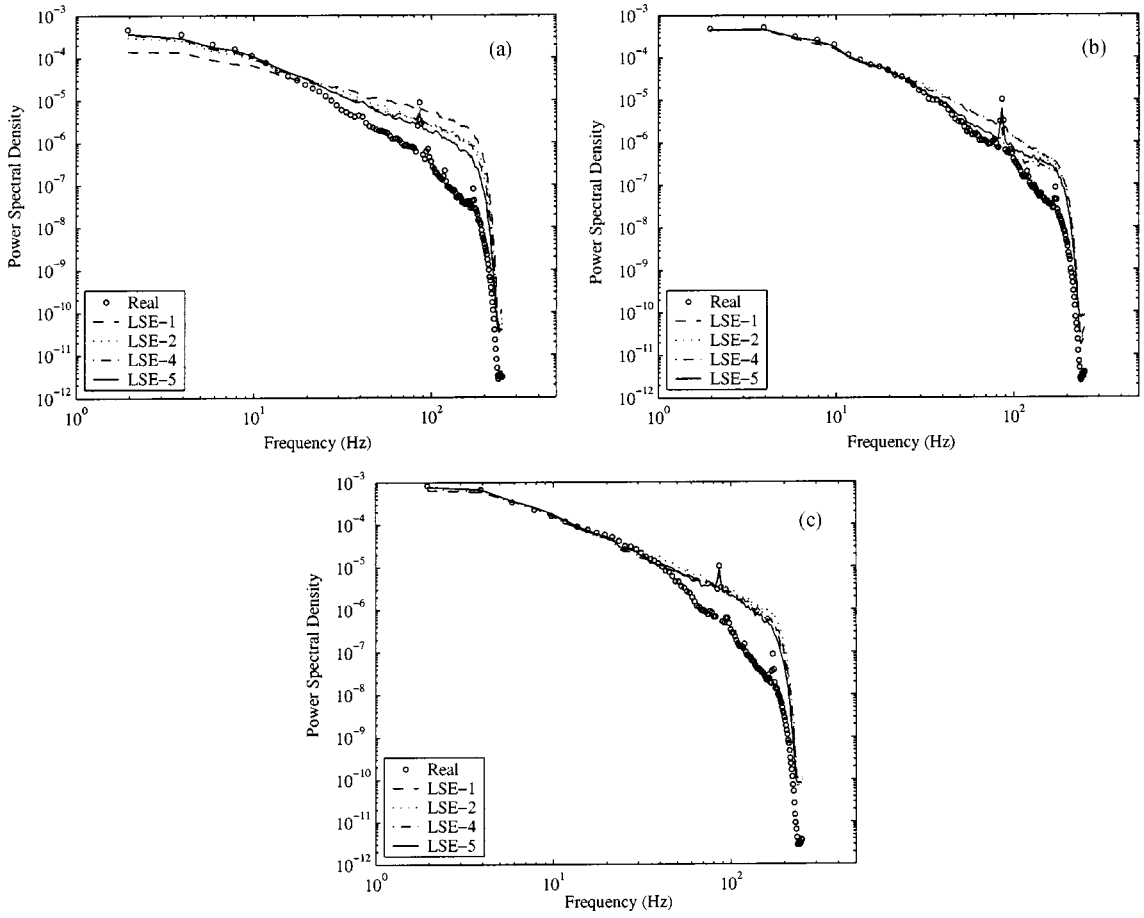

Fig. 5 Autospectra of the measured and reconstructed uplift time series in a corner bay in open country terrain for: (a) $\alpha=270^{o}$, (b) $\alpha=320^{o}$ and (c) $\alpha=360^{o}$

4.2. Reconstruction of individual pressure time series

As an example of the performance of an individual pressure time series, the reconstruction of tap #1210 is presented using the LSE-5 scheme. Fig. 4 shows the reconstructed and actual autospectra for a cornering wind. As can be seen, the reconstructed autospectrum deviates significantly from that of the experimental data, especially at the energetic frequencies. This indicates that the LSE is not appropriate for capturing the fluctuation energy of individual pressure time series (at least as currently formulated with only a few taps).

4.3. Reconstruction of bay uplift

Fig. 5 shows the comparison of autospectra of the experimental and the reconstructed bay uplift time series using LSE-1 through LSE-5 for three wind directions. As expected, the autospectra obtained with LSE-5 resemble most closely those of the experimental data. For LSE-1, except for the 270^o azimuth, the simulated autospectra match well with the original autospectra, especially in the low frequency domain (which accounts for the largest fraction of the fluctuation energy). The poorer performance of LSE-1 for 270^o azimuth indicates in this case, that more reference time series are required to capture the fluctuations (e.g., using LSE-2), or possibly a different tap should be used. Generally, the more reference time series that are used, the better the agreement with the real data since more "conditional information" is known. It is also observed that almost all of the schemes failed to reproduce the sudden spike in the autospectra at roughly 90 Hz, which is due to the blade passage frequency. This can be attributed to the limitations of the linear assumption in the methodology.

4.4. Reconstruction of frame uplift

Table 2 summarizes the statistics from the experimental and the reconstructed frame uplift time series (Frame 1). Similar findings for the bay uplift reconstruction were obtained (with further details in Chen, 2002). As shown, the LSE approach achieves good performance in reconstructing the frame uplift time series, including using only one reference tap. In particular, the mean value of the frame uplift is reconstructed perfectly. Among the schemes considered, LSE-5 exhibits the best performance. For example, by using LSE-5 for 320^o azimuth, the maximum prediction errors of the other statistics (rms, peak suction, skewness and kurtosis) are below 3%. The large correlation coefficient (0.97) between the reconstruction and the experimental data indicates an accurate estimation of the fluctuating features. The (weakly) non-Gaussian features of the experimental frame uplift in this corner bay are also preserved well, as confirmed by the good prediction of skewness and kurtosis coefficients (Note that the absolute values of skewness and kurtosis coefficients of an ideal Gaussian time series are 0.0 and 3.0, respectively). For this direction, LSE-1 also performs well in estimating the rms and peak suction with a maximum absolute error of 6%.

Fig. 6 shows the probability density functions (PDF) and cumulative distribution functions (CDF) of the measured and reconstructed frame uplift time series for a cornering wind. It can be seen that both the PDF and CDF of the original time series including the upper and lower tails were preserved well in the time series reconstructed by using LSE-1 through LSE-5. Considering that the observed peaks (i.e., the single worst values) are not commonly used in engineering practice due to the significant stochastic variability in time series samples (Sadek and Simiu 2002), more reliable peaks are estimated through an extreme value analysis. The commonly used Type I (Gumbel)

Table 2 Statistics of the reconstructed and experimental uplift time series for Frame 1 with wind directions of 270°, 320° and 360° in open country terrain

α	Case	Mean	Rms	Min.	Max.	Skewness	Kurtosis	CorrCoef
270°	Real	-0.1791	0.0467	-0.4333	-0.0567	-0.61	3.39	
	LSE-1	-0.1791 (0.0%)	0.0340 (-33.7%)	-0.4148 (-4.3%)	-0.0778 (37.1%)	-0.75 (23.0%)	4.01 (18.3%)	0.66
	LSE-2	-0.1791 (0.0%)	0.0373 (-20.1%)	-0.4179 (-3.5%)	-0.0827 (45.8%)	-0.63 (3.3%)	3.52 (3.8%)	0.80
	LSE-4	-0.1791 (0.0%)	0.0428 (-8.3%)	-0.4762 (9.9%)	-0.0712 (25.6%)	-0.65 (6.6%)	3.56 (5.0%)	0.92
	LSE-5	-0.1791 (0.0%)	0.0436 (-6.6%)	-0.4643 (7.2%)	-0.0700 (23.4%)	-0.67 (9.8%)	3.61 (6.5%)	0.93
320°	Real	-0.2441	0.0654	-0.5872	-0.0140	-0.68	3.83	
	LSE-1	-0.2441 (0.0%)	0.0615 (-6.0%)	-0.5798 (-1.3%)	-0.0494 (253.1%)	-0.75 (10.3%)	3.94 (2.9%)	0.94
	LSE-2	-0.2441 (0.0%)	0.0621 (-5.0%)	-0.5728 (-2.5%)	-0.0411 (193.5%)	-0.73 (7.3%)	3.91 (2.1%)	0.95
	LSE-4	-0.2441 (0.0%)	0.0630 (-3.6%)	-0.5707 (-2.8%)	-0.0329 (135.1%)	-0.72 (5.9%)	3.82 (-0.3%)	0.96
	LSE-5	-0.2441 (0.0%)	0.0634 (-3.0)	-0.5770 (-1.7%)	-0.0190 (-35.9%)	-0.70 (-2.9%)	3.82 (-0.3%)	0.97
360°	Real	-0.2093	0.0652	-0.5546	-0.0577	-0.74	3.63	
	LSE-1	-0.2093 (0.0%)	0.0567 (-13.0%)	-0.6846 (23.4%)	-0.0997 (72.7%)	-1.29 (74.3%)	5.94 (63.6%)	0.87
	LSE-2	-0.2093 (0.0%)	0.0591 (-9.4)	-0.6344 (14.4%)	-0.0869 (50.6%)	-0.96 (29.7%)	4.32 (19.0%)	0.91
	LSE-4	-0.2093 (0.0%)	0.0627 (-3.9%)	-0.5497 (-0.9%)	-0.0703 (21.7%)	-0.81 (9.5%)	3.82 (5.2%)	0.96
	LSE-5	-0.2093 (0.0%)	0.0632 (-3.2%)	-0.5604 (1.0%)	-0.0691 (19.6%)	-0.81 (9.5%)	3.82 (5.2%)	0.97

Note: 1. The bracketed terms are the errors relative to the actual values.
2. CorrCoef represents correlation coefficient relative to the experimental time series.

extreme value distribution (Cook 1985) was chosen for the current analysis of the largest peaks (i.e., peak suctions multiplied by −1). The Gumbel distribution function is defined as:

$$-\ln[-\ln(-P)]=a(x-u) \tag{21}$$

where x denotes a peak value, P is the probability of no exceedance of the peak value x, a is a measure of dispersion, and u is a measure of location. In this study, the distribution parameters were determined by fitting the peaks on Gumbel plot paper using the least-mean-square method. The peaks were taken from 50 segments of frame uplift time series. Fig. 7 compares the distribution of the largest peaks of the experimental and reconstructed frame uplift time series using LSE-5. Table 3 summarizes the comparison of the estimated peaks corresponding to the probability of no exceedance, as well as the fitted Gumbel parameters. It can be seen that the reconstructed extreme value distribution exhibits good agreement with the experimental data. This is also indicated by well-matched Gumbel parameters. The peaks corresponding to the probability of no exceedance of

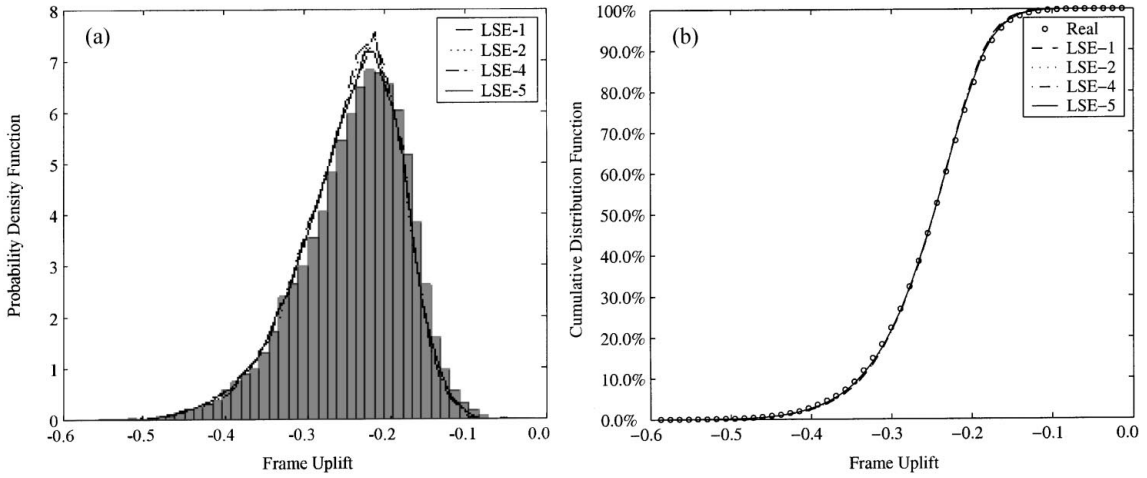

Fig. 6 (a) Probability distribution functions, and (b) cumulative distribution functions of the measured and reconstructed uplift time series for Frame 1 with $\alpha=320^{o}$ in open country terrain

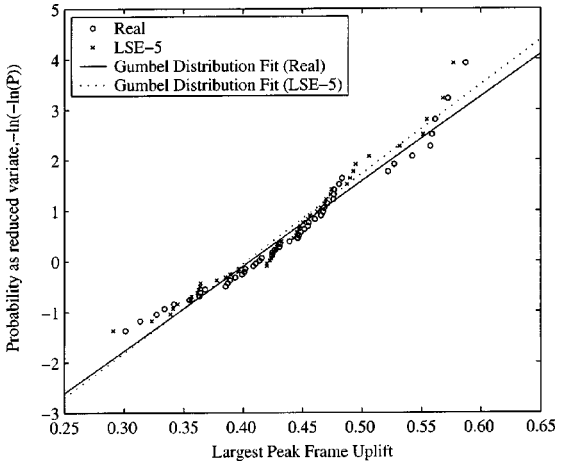

Fig. 7 Distribution of the minima of the uplift time series for Frame 1 (multiplied by −1) with $\alpha=320^{o}$ in open country terrain

Table 3 The observed and estimated worst peaks for frame uplift (Frame 1) and ridge bending moment

Structural Load Effects	Case	Observed Worst Peak	Estimated Worst Peaks			Gumbel Distribution Parameters	
			P(probability of no exceedance)			$y=-\ln[-\ln(P)]=a(x-u)$	
			99%	90%	80%	a	u
Frame Uplift	Real	-0.5872	-0.6793	-0.5395	-0.4948	16.80	0.4055
	LSE-5	-0.5770	-0.6624	-0.5298	-0.4874	17.72	0.4028
	Error%	-1.7%	-2.5%	-1.8%	-1.5%		
Ridge Bending Moment	Real	-0.2290	-0.2625	-0.2133	-0.1975	47.70	0.1661
	LSE-5	-0.2257	-0.2588	-0.2095	-0.1938	47.69	0.1623
	Error%	-1.5%	-1.4%	-1.8%	-1.9%		

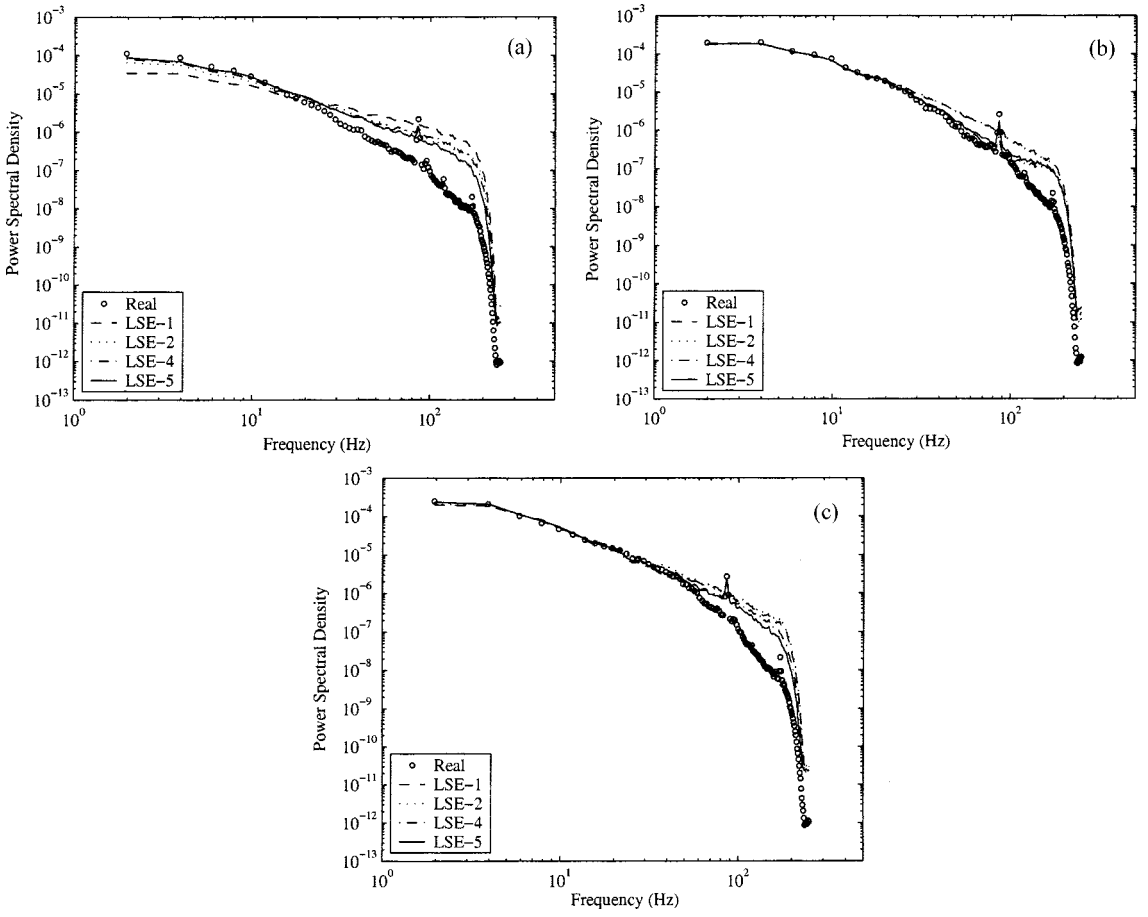

Fig. 8 Autospectra of the measured and reconstructed uplift time series for Frame 1 in open country terrain with: (a) α=270° (b) α=320° and (c) α=360°

99%, 90%, and 80% can be reconstructed accurately with a maximum error less than 3%.

The good performance of the LSE approach is also confirmed by the comparison of the corresponding autospectra between the LSE reconstructions and the experimental data. As shown in Fig. 8, all the simulated power spectra by using various LSE schemes match well with the original data, except for LSE-1 for 270° azimuth.

4.5. Reconstruction of ridge bending moment

Table 4 gives the statistics of the reconstructed and experimental bending moment calculated about ridge for three wind directions. As shown, LSE-5 achieves the best performance for reconstructing the ridge bending moment time series, where the maximum estimation error for all the key statistics (except for the most positive peak) is less than 10%, and all the corresponding correlation coefficients are larger than 0.93. It appears that two pressure taps (LSE-2) are sufficient to reconstruct the bending moment in terms of mean, rms and minimum peak. Even given one reference tap, the LSE still captures the minimum peak value accurately with an absolute error less

Table 4 Statistics of the reconstructed and experimental ridge bending moment for the corner bay with wind directions of 270°, 320° and 360° in open country terrain

α	Case	Mean	Rms	Min.	Max.	Skewness	Kurtosis	CorrCoef
270°	Real	-0.1129	0.0299	-0.2814	-0.0332	-0.59	3.35	
	LSE-1	-0.1129 (0.0%)	0.0208 (-30.4%)	-0.2709 (-3.7%)	-0.0449 (35.2%)	-0.75 (27.1%)	4.01 (19.7%)	0.70
	LSE-2	-0.1129 (0.0%)	0.0268 (-10.4%)	-0.2835 (0.7%)	-0.0449 (35.2%)	-0.63 (6.8%)	3.49 (4.2%)	0.90
	LSE-4	-0.1129 (0.0%)	0.0286 (-4.4%)	-0.2949 (4.8%)	-0.0443 (33.3%)	-0.61 (3.4%)	3.40 (1.5%)	0.96
	LSE-5	-0.1129 (0.0%)	0.0291 (-2.7%)	-0.2960 (5.2%)	-0.0421 (26.7%)	-0.63 (6.8%)	3.46 (3.3%)	0.97
320°	Real	-0.0995	0.0263	-0.2290	-0.0009	-0.55	3.72	
	LSE-1	-0.0995 (0.0%)	0.0211 (-19.6%)	-0.2150 (-6.1%)	-0.0326 (3450%)	-0.75 (36.4%)	3.94 (5.9%)	0.80
	LSE-2	-0.0995 (0.0%)	0.0247 (-6.1%)	-0.2209 (-3.6%)	-0.0203 (2112%)	-0.70 (27.3%)	3.86 (3.8%)	0.94
	LSE-4	-0.0995 (0.0%)	0.0248 (-5.5%)	-0.2217 (-3.2%)	-0.0197 (2046%)	-0.68 (23.6%)	3.78 (1.6%)	0.95
	LSE-5	-0.0995 (0.0%)	0.0252 (-4.0%)	-0.2257 (-1.5%)	-0.0111 (1105%)	-0.62 (12.7%)	3.78 (1.6%)	0.96
360°	Real	-0.1015	0.0309	-0.2552	-0.0300	-0.73	3.62	
	LSE-1	-0.1015 (0.0%)	0.0252 (-18.6%)	-0.3126 (22.5%)	-0.0527 (76.1%)	-1.29 (76.7%)	5.94 (64.1%)	0.81
	LSE-2	-0.1015 (0.0%)	0.0272 (-12.0%)	-0.2832 (11.0%)	-0.0448 (49.5%)	-0.94 (28.8%)	4.23 (16.9%)	0.88
	LSE-4	-0.1015 (0.0%)	0.0296 (-4.3%)	-0.2548 (-0.2%)	-0.0327 (9.1%)	-0.77 (5.5%)	3.75 (3.6%)	0.96
	LSE-5	-0.1015 (0.0%)	0.0299 (-3.3%)	-0.2630 (3.1%)	-0.0327 (9.2%)	-0.77 (5.5%)	3.74 (3.3%)	0.97

Note: 1. The bracketed terms are the errors relative to the actual values.
2. CorrCoef represents correlation coefficient relative to the experimental time series.

than about 6% for the wind directions of 270° and 320°.

The comparisons in terms of PDF and CDF for a wind direction of 320° are shown in Fig. 9. All the LSE schemes (except for LSE-1) have good performance to preserve the statistical features of the original data. For the distribution of the largest peaks, using five reference taps also can reproduce the extreme value distribution well, as shown in Fig. 10. The estimated peaks of the bending moment time series for three probabilities of no exceedance are given in Table 3. As shown, the LSE reconstruction accurately reproduces not only the observed single worst peak but also the estimated peaks through the Gumbel distribution, with a maximum absolute error less than 2%.

Fig. 11 shows the corresponding autospectra of the LSE reconstructed and experimental ridge bending moment. As shown, two reference taps (LSE-2) are sufficient to reconstruct the majority of the fluctuation energy of the original time series. This is also in accordance with the above comparison in terms of statistical features.

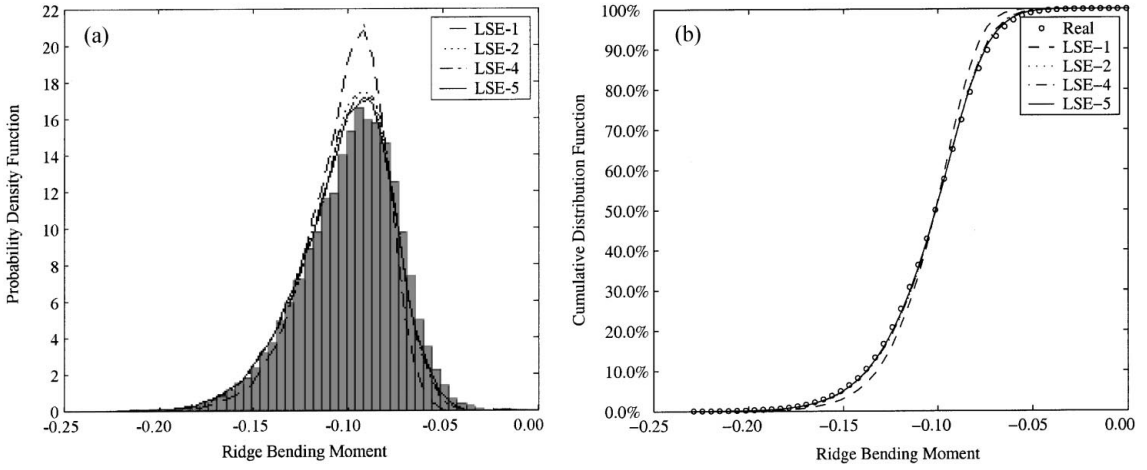

Fig. 9 (a) Probability distribution functions and (b) cumulative distribution functions of the measured and reconstructed ridge bending moment time series for the corner bay with α=320° in open country terrain

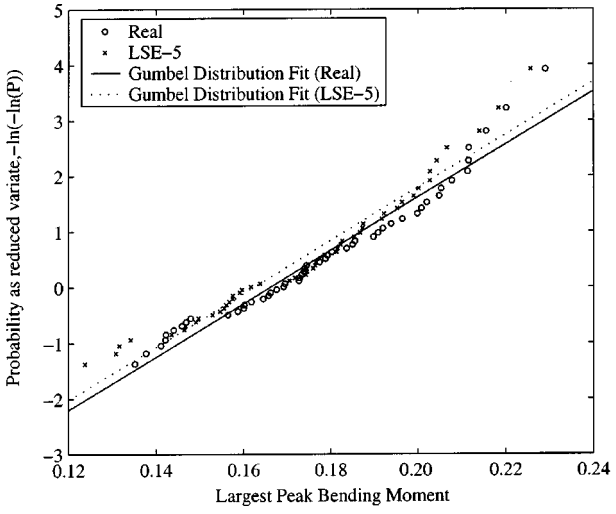

Fig. 10 Distribution of the minima of the ridge bending moment time series (multiplied by −1) for the corner bay for α=320° in open country terrain

4.6. Discussion

The LSE approach for estimating structural loads with the individual reconstructed pressure time series is simple in idea and implementation. As a stochastic approach, LSE is essentially a linear estimate of the conditional average of a random variable by neglecting the nonlinear terms of a Taylor series expansion, given some reference variables. This assumption is open to question, especially when applied to strongly non-Gaussian time series (e.g., pressure time series in the separated flow regions) because it is believed that linear functions are unlikely to be able to capture non-linearity. Of course, stochastic estimation of a random variable given some reference variables and also the higher-order terms can improve the accuracy of estimation (Adrian 1994), but it will

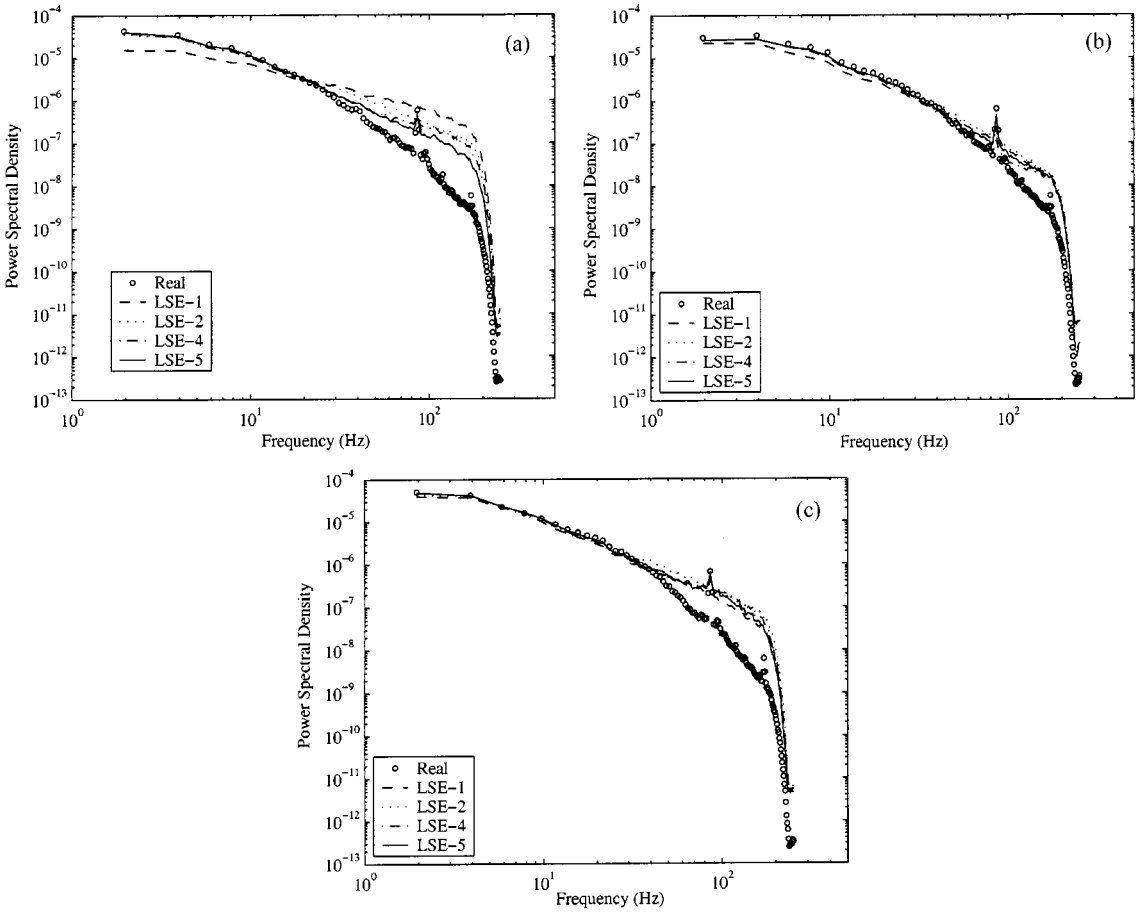

Fig. 11 Autospectra of the measured and reconstructed ridge bending moment time series for the corner bay in open country terrain with: (a) α=270° (b) α=320° and (c) α=360°

increase complexity and reduce practicality. Furthermore, it was shown by Tung and Adrian (1980) that the higher-order terms of Taylor series expansion would not have a significant effect on the estimation of the large-scale coherent structures in a turbulent velocity field, which verifies the use of LSE with the assumption of linear approximation. As well, the purpose of using LSE in this study is for approximate estimation of structural load time series on large-scale loads, not for local reconstruction of individual pressure time series. As shown in this study, for the reconstruction of structural load time series, LSE is efficient since the target time series are generally weakly non-Gaussian or Gaussian. However, in the case of simulating strongly non-Gaussian time series, it is not realistically expected that the LSE based on the assumption of linear approximation can accurately capture the fluctuating features, although it still outperforms linear interpolation (Chen 2002). With this concern, it is suggested that nonlinear models be used for local reconstruction of pressure time series (e.g., time-delay ANN approaches (Giralt *et al.* 2000, Chen *et al.* 2002a)). This would need further investigation.

Regarding time series reconstruction, Proper Orthogonal Decomposition (POD) (e.g., Bienkiewicz *et al.* 1995), another standard stochastic approach based on using two-point correlations, also has

potential. The difference between them lies in the interpretation of the covariance information. POD uses the covariance matrix of the whole surface field to estimate modes by solving an eigenvalue problem, and the reconstruction of pressure time series can be performed with the first few dominant modes. In contrast, LSE uses the covariance matrix of only the selected reference time series and the target time series to determine estimation coefficients. They both have their advantages and disadvantages. Generally, LSE is simpler, and can reproduce the mean value of pressure time series completely, but the reference time series need to be optimally determined in order to achieve the best possible reconstruction performance. POD reduces data by eliminating higher-order modes which contain "unimportant" information.

Like for POD, LSE is applicable only if covariance matrices are known (since the covariance matrix will be used to calculate the estimation coefficients). The covariance information can be obtained from either wind tunnel or full-scale experiments and should be stored with the database. As shown in this study, the LSE can save at least 96% data storage if five properly selected reference taps are sufficient to reconstruct the pressure field accurately in the corner bay (where 120 pressure taps were instrumented). Clearly, this indicates that aerodynamic databases which contain a large number of pressure time series on the entire surface of low buildings can be reduced to a much smaller data set by storing some optimal reference time series and the covariance matrices. The reconstructed pressure time series can be used as inputs to the finite element analysis programs, provided that the reconstruction accuracy of LSE with these stored reference time series is sufficiently acceptable. However, several taps would need to be stored from each bay and the walls in order to achieve better performance. Besides the potential for data reduction associated with DAD, the covariance matrix can also be directly used by LRC method, offering a more flexible database.

5. Conclusions

A simple and practical approach through the application of LSE has been developed to reconstruct wind-induced pressure time series for the analysis of structural loads in a corner bay of a low building where the covariance matrix is known. The reconstruction is based on the assumption that pressure time series at any location can be estimated as a linear function of some other pressure time series, and structural loads in a desired region can be calculated with these individual simulated pressures. The effect of using different reference time series has been investigated in order to achieve the best performance. The comparison between the simulation and the actual measured data was at the level of statistical features including mean, rms, peak, skewness, kurtosis, correlations, autospectra, PDF, and CDF. Some major conclusions based on the present results can be made as follows:

- LSE is efficient for reconstructing pressure time series for use in the analysis of structural loads, given a handful of reference pressure taps (even only one tap).
- LSE can reconstruct the mean value of pressure time series perfectly; for other statistical features, the performance depends on the number and location of reference time series.
- In order to achieve the best performance, reference time series need to be optimally determined. The objective is to keep the number of reference pressure taps as small as possible, providing that the simulation accuracy is still acceptable.
- The LSE can save about 96% data storage when simulating the pressure field of a corner bay with five taps. This indicates that LSE has potential to be used for reducing data storage requirements in the development of the aerodynamic database.

Acknowledgements

This work presented was part of the Ph.D. program of the first author under the supervision of the second and third authors. The financial support of NIST/TTU and NSERC is gratefully acknowledged. One of the authors (GAK) acknowledges the support provided by the Canada Research Chairs Program. The authors would also like to thank Dr. T.C.E. Ho for his help with the data handling.

References

Adrian, R.J. (1975), "On the role of conditional averages in turbulence theory", *Turbulence in Liquids: Proc. Fourth Biennial Symp. on Turb. in Liquids*, September. J. Zakin and G. Patterson (Eds), Science Press, Princeton, 323-332, 1977.

Adrian, R.J. and Moin, P. (1988), "Stochastic estimation of organized turbulent structure: homogeneous shear flow", *J. Fluid Mech.*, **190**, 531-559.

Adrian, R.J., Jones, B.G., Chung, M.K., Hassan, Y., Nithianandan, C.K., and Tung, A.T.C. (1989), "Approximation of turbulent conditional averages by stochastic estimation", *Phys. Fluids*, **1**(6).

Adrian, R.J. (1994), "Stochastic estimation of conditional structure: a review", *App. Sci. Res.*, **53**, 291-303.

ASCE Standard 7-98 (2000), *Minimum Design Loads for Buildings and Other Structures, Revision of ANSI/ ASCE 7-95*, American Society of Civil Engineers, Reston, Virginia.

Bienkiewicz, B., Tamura, Y., Ham, H.J., Ueda, H. and Hibi, K. (1995), "Proper orthogonal decomposition and reconstruction of multi-channel roof pressure", *J. Wind Eng. Ind. Aerod.*, **54/55**, 369-381.

Chen, Y. (2002). "Time series simulation of wind-induced pressures on low buildings", Ph.D. thesis, Department of Civil and Environmental Engineering, The University of Western Ontario, London, Ontario, Canada.

Chen, Y., Kopp, G.A. and Surry, D. (2002a), "Interpolation of wind-induced pressure time series with an artificial neural network", *J. Wind Eng. Ind. Aerod.*, **90**, 589-615.

Chen, Y., Kopp, G.A. and Surry, D. (2002b), "Interpolation of pressure time series in an aerodynamic database for low buildings", submitted to *J. Wind Eng. Ind. Aerod.*

Chen, Y., Kopp, G.A. and Surry, D. (2003), "Prediction of (to appear) pressure coefficients on roofs of low buildings using artificial neural networks", *J. Wind Eng. Ind. Aerod.*, **91**, 423-441.

Cook, N.J. (1985), *The designer's guide to wind loading of building structure: Part 1*, Butterworths, Building Research Establishment Report.

Delville, J., Lamballais, E. and Bonnet, J.-P. (2000), "POD, LODS and LSE : their links to control and simulations of mixing layers", *ERCOFTAC Bulletin*, **46**, 29-38.

Druault, P., Lamballais, E., Delville, J. and Bonnet, J.P. (1999), "Development of experiment/simulation interfaces for hybrid turbulent results analysis via the use of DNS", *Proc. 1st Int. Symp. Turb. Shear Flow Phenomena*, Santa Barbara, California, 779-784.

Giralt, F., Arenas, A., Ferre-Giné, J., Rallo, R. and Kopp, G.A. (2000), "The simulation and interpretation of turbulence with a cognitive neural system", *Phys. Fluids*, **12**, 1826-1835.

Ho, T.C.E, Surry, D., Morrish, D., and Kopp, G.A. (2002), "The NIST aerodynamic database for wind loads on low buildings: Part 1. Basic aerodynamic data and archiving", *under preparation.*

Holmes, J.D. and Syme, M.J. (1994), "Wind loads on steel-framed low-rise buildings", *Steel Construction*, **28**(4), 2-12.

Kasperski, M. (1992), "Extreme wind load distributions for linear and nonlinear design", *Eng. Struct.*, **14**, 27-34.

Papoulis, A. (1984). *Probability, Random Variables and Stochastic Theory* (2nd Edn.), McGraw-Hill, New York.

Péneau, F., Faghani, D. and Boisson, H.C. (2000), "Linear stochastic estimation of velocity entrance signals for a L.E.S. of a turbulent flat plate boundary layer", *ERCOFTAC Bulletin*, **46**, 39-43.

Rigato, A., Chang, P. and Simiu, E. (2001), "Database-assisted design, standardization and wind direction effects", *ASCE J. Struct. Eng.*, **127**(8), 855-860.

Sadek, F. and Simiu, E. (2002), "Peak non-Gaussian wind effects for database-assisted low-rise building design", *ASCE J. Eng. Mech.*, **128**(5), 530-539.

Simiu, E. and Stathopoulos, T. (1997), "Codification of wind loads on buildings using bluff body aerodynamics and climatological data base", *J. Wind Eng. Ind. Aerod.*, **69-71**, 497-506.

Stathopoulos, T. (1979). "Turbulent wind action on low-rise buildings", Ph.D Dissertation, Department of Civil

and Environmental Engineering, The University of Western Ontario, London, Ontario, Canada.
Tung, T.C. and Adrian, R.J. (1980), "Higher-order estimates of conditional eddies in isotropic turbulence", *Phys. Fluids*, **23**(7), 1469-1470.
Walpole, R.E. and Myers, R.H. (1985), *Probability and Statistics for Engineering and Scientists* (3rd Edn.), Macmillan Publishing Company, New York.
Whalen, T., Simiu, E., Harris, G., Lin, J. and Surry, D. (1998), "The use of aerodynamic databases for the effective estimation of wind effects in main wind-force resisting systems: application to low buildings", *J. Wind Eng. Ind. Aerod.*, **77-78**, 685-693.

Notation

A	area of a roof zone
ΔA_i	tributary area of tap i
$\{\mathbf{B}\}$	estimation coefficient vector
$b_i(x, y)$	ith estimation coefficient for simulating pressure time series at (x, y)
$[\mathbf{C}]$	spatial covariance matrix between reference time series
c_{ij}	spatial covariance between reference time series at location i and j
$\{\mathbf{C_0}\}$	spatial covariance vector between the pressure time series to be reconstructed and reference time series
C_p	non-dimensional wind pressure coefficient
$C_p(x, y; t)$	pressure time series at location (x, y)
$C_{p_{i,ref}}(t)$	reference pressure time series at location i, also denoted as $C_{p_{ref}}(x_i, y_i; t)$
$C_p'(x, y; t)$	fluctuation of pressure time series at location (x, y)
$\bar{C}_p(x, y)$	mean value of pressure time series at location (x, y)
$\tilde{C}_p(x, y)$	rms value of pressure time series at location (x, y)
$\hat{C}_p(x, y; t)$	linear stochastic estimate of pressure time series at location (x, y)
$\hat{C}_{p_i}(t)$	linear stochastic estimate of pressure time series at location i
$F_1(t)$	uplift time series
$F_2(t)$	frame uplift time series
$F_3(t)$	bending moment time series
I_i	influence coefficient of tap i
L_i	bending moment arm of tap i
t	time instant
(x, y)	spatial coordinate of a tap location
W	building width
$\mathrm{E}(\bullet\|\bullet)$	conditional average operator
$[\bullet]^{-1}$	inverse of matrix
$\overline{(\bullet)}$	average operator
α	wind direction

Abbreviations

CorrCoef	correlation coefficient
DAD	database-assisted design
LRC	load-response-correlation
LSE	linear stochastic estimation
LSE-i	linear stochastic estimation which uses i reference time series
NIST	National Institute of Standards and Technology
POD	proper orthogonal decomposition
rms	root-mean-square