

Normalized proper orthogonal decomposition (NPOD) for building pressure data compression

Dan Ruan^{a,*}, Hua He^c, David A. Castañón^b, Kishor C. Mehta^c

^a*Department of Electrical Engineering and Computer Science, University of Michigan, Ann Arbor, MI 48109*

^b*Department of Electrical and Computer Engineering, Boston University, Boston, MA 02215*

^c*Department of Civil Engineering, Texas Tech University, Lubbock, TX 79406*

Received 18 April 2004; received in revised form 5 July 2005; accepted 10 January 2006

Available online 28 February 2006

Abstract

Proper orthogonal decomposition (POD) has been widely realized as a tool for compressing fluctuating building pressure data in wind-engineering area. Modes are determined by eigen decomposition of data covariance matrix, and then truncation is applied to retain only the modes with the highest energy. However, as observed by S. Kho, C. Baker, R. Hoxey [Pod/arma reconstruction of the surface pressure field around a low rise structure, *J. Wind Eng. Ind. Aerodyn.* (90) (2002) 1831–1842] and analyzed in the authors' previous paper D. Ruan, H. He, D. Smith, K.C. Mehta [A Semi-optimal Mode Selection Scheme for Pod Based Compression of Wind Field Data, Seoul, Korea, 2004], reconstruction performance varies a lot among individual taps. In D. Ruan, H. He, D. Smith, K.C. Mehta [A Semi-optimal Mode Selection Scheme for Pod Based Compression of Wind Field Data, Seoul, Korea, 2004], a semi-optimal mode choosing scheme was proposed in an integer programming (IP) framework and numerical experiments have shown sound results. Nevertheless, computation cost for solving IP optimization could be an issue as the total number of taps grows. In this paper, we address the problem with normalized proper orthogonal decomposition which is from a completely different perspective: we first normalize the data to force every tap to have similar contribution to the total energy; the standard POD is then applied to get a compact representation. During the reconstruction phase, normalized data is first restored and a scaling procedure which is exactly the inverse to the normalization is applied to finally reconstitute the data

*Corresponding author. Tel.: +1 7346440476; fax: +1 7347638041.

E-mail addresses: druan@umich.edu (D. Ruan), hua.he@ttu.edu (H. He).

in its original form. Feasibility test with experimental data from Texas Tech University yields good results. Computation cost is almost the same as the standard POD method.

© 2006 Elsevier Ltd. All rights reserved.

Keywords: Data compression; Proper orthogonal decomposition (POD); Principal component analysis (PCA); Eigendecomposition; Normalized cumulative fluctuation energy (NCFE); Integer programming (IP); Normalized proper orthogonal decomposition (NPOD)

1. Introduction

Principal component analysis (PCA) is a powerful tool in compact data representation/data compression. Beginning with a data set in a high-dimensional space, PCA tries to find a lower dimensional hyperplane that best represents the data points. In wind engineering field, it is generally recognized as proper orthogonal decomposition (POD) and is widely studied to describe fluctuating building surface pressure data [3–8]. Some authors have tried to fit autoregressive (AR) model to extracted POD modes [9,10] in order to achieve further compression. While some investigators [4,3] claimed that this decomposition helps in identifying the hidden systematic structure in the pressure fluctuation, other parties argued [11], that the most useful aspect of POD is its economy in describing the spatial/temporal variation of wind pressure field [12–14].

In traditional POD-based compression methods, the covariance matrix of the observation data is first eigendecomposed to get the set of eigenvectors (modes). The modes are then arranged in an energy non-increasing fashion and truncation is applied to retain only a subset of them. However, as Kho et al. [1] observed, the exclusion of higher (less energetic) modes leads to significant inaccuracy of reconstituted time series at some of the tapping points.

In the authors' previous work [2], contributions of each mode to the fluctuation energy at individual tapping points are analyzed, and an IP framework is proposed to choose a subset of the eigenvectors after eigendecomposition for reconstruction purpose. However, the solutions to IP optimization problems are not trivial and the computation cost grows with the number of taps included.

In this work, we take a completely different perspective by taking advantage of proper normalization. We avoid the unbalanced performance in the traditional POD-based methods by first normalizing the observation data to make each tap contribute similarly to the total energy. This results in a decomposition that assigns similar weight to each dimension of data (each tap). In the compression step, only the first few normalized modes are retained as in traditional POD. During the reconstitution process, normalized data are first restored according to the traditional POD approach, then a scaling procedure which plays the role of inverse normalization is applied and transforms the normalized data to its original space.

Feasibility test and verification is conducted using experimental data collected at wind engineering research field laboratory (WERFL) at Texas Tech University.

2. Proper orthogonal decomposition (POD)

POD, which is also known as PCA in statistics and signal processing is a well-established technique for dimension reduction. The most common way to look at POD is in terms of

an orthogonal projection which maximizes the variance in the projected space. Given an dyadic N -dimensional observation vector $x(t)$, POD aims at finding an orthogonal linear projection $y = W^T x^1$ such that the variance of $y \in \mathfrak{R}^M$ ($M < N$) is maximized. The i th element of y is called the i th principal component.

From an optimization point of view, POD minimizes the reconstruction error in the mean square sense, i.e.,

$$W^* = \arg \min_W E_t \|x(t) - \hat{x}(t)\|^2, \quad (1)$$

where $\hat{x}(t) = Wy$ is the reconstruction (back projection) in \mathfrak{R}^N from the projected data y . It could be shown that the i th column vector of W^{*2} corresponds to the normalized eigenvector associated with the i th largest eigenvalue of the covariance matrix $R_x = E\{xx^T\}$.

In the scenario of building pressure data, observation data in original (high) dimension is acquired by stacking the simultaneous measurements from all taps at each time instant. Technically, let $x(i, t)$ be the measurement at tap i at time t with $i = 1, 2, \dots, N$ where N is the total number of taps. The original data lies in a space whose dimension is determined by the number of taps.

3. Normalized proper orthogonal decomposition

The motivation of normalized proper orthogonal decomposition (NPOD) comes from the desire to give each dimension of observed variable (in our case, each tap) similar weight in the decomposition process. In traditional POD method, the principal modes come from eigendecomposition of the variance matrix, which is scale variant, i.e., principal modes change when some dimension of the original observation is scaled. A direct result of this in the application of POD to surface field pressure data is that data from windward taps receive more attention than leeward taps as pressure data from windward taps in general have larger variance than leeward ones. From a data processing point of view, this means different dimensions of pressure data are heterogeneous with respect to their variance and this leads to the unfair reconstruction performance observed in [1,2].

In NPOD method proposed in this paper, we first normalize the data by subtracting the mean and then dividing by square-root of the variance at each dimension so that each dimension is now standardized with zero mean and unit variance. Using this normalized data as the input to traditional POD system guarantees each dimension to receive similar weight in the decomposition. In the reconstruction process, the normalized data is first reconstructed from lower dimension data and then approximation of raw data is acquired by inverse scaling and proper shifting as illustrated in Fig. 1.

4. Experimental setup and data acquisition for feasibility test

Experimental data from the WERFL at Texas Tech University are used to carry out both feasibility test and verification examination in this paper. WERFL is used to conduct full-scale measurements of wind and its effects on buildings. It consists of a 160-ft high meteorological tower, a rotatable $30 \times 45 \times 13$ ft test building, and a $10 \times 10 \times 8$ ft data acquisition room. The test building is anchored to a rigid frame undercarriage, which is

¹The superscript (T) in this paper denotes transpose of matrix.

²We use superscript asterisk (*) to denote the optimal value of the objective in this paper.

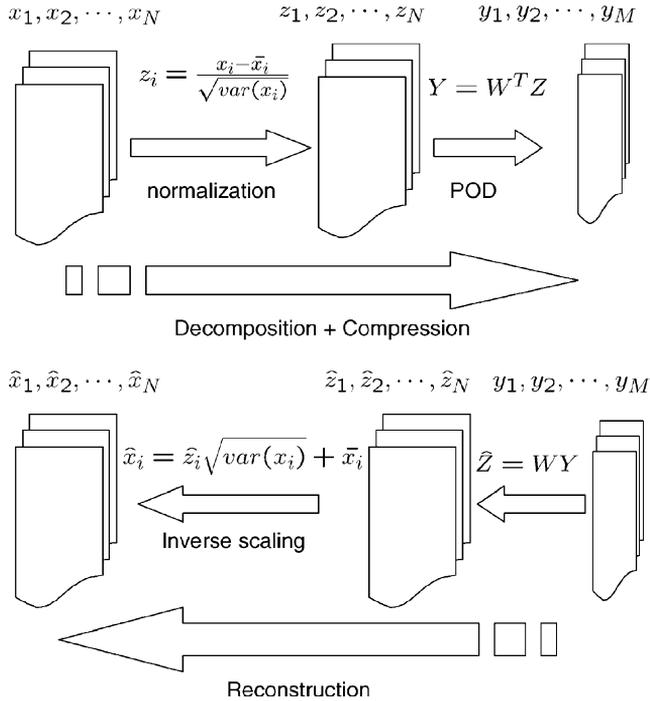


Fig. 1. Data flow of NPOD.

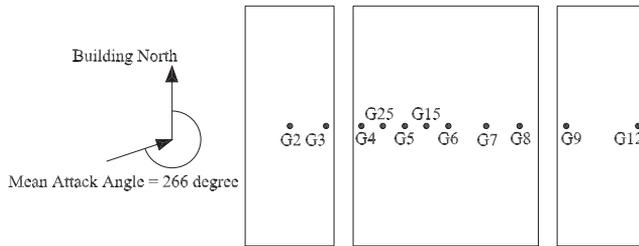


Fig. 2. Layout of tapping points. In later discussions, we may use eastward indexing of the tapping points for expression convenience where no confusion is expected. ($G2 \rightarrow \text{tap1}, G3 \rightarrow \text{tap2}, \dots$).

supported by hydraulic jacks at each corner. The jacks are mounted on wheels which ride on a circular rail. This setup allows researchers to raise and rotate the building to any direction as desired [15,16].

204 pressure taps are installed outside of the test building on the roof and walls (Fig. 9). In the feasibility test, we use the observation data from 11 tapping points located across the central frame of the building (Fig. 2) during test Run M15N541, measured at 40 Hz for 15 min, i.e., $15 \text{ min} \times 60 \text{ s/min} \times 40 \text{ Hz} = 36,000$ wind pressure data are recorded for each tapping point.

5. Numerical results

5.1. Heterogeneous data w.r.t. variance

The heterogeneous nature of the wind pressure field data is shown in Fig. 3. In fact, the ratio of the largest variance versus the smallest variance among taps in this demonstration case is more than 31, which means a “stretching factor” of about 5.57 in some dimension with respect to some other.

5.2. Results from POD

Here we present the computation results of POD. Fig. 4 shows the magnitude of all eigenvalues calculated from POD. It demonstrates the fast decaying behavior of the eigenvalues, which suggests the concentration of energy amongst the first few modes. Fig. 5 offers an even more straightforward observation. We computed the normalized cumulative fluctuation energy (NCFE) up to the n th mode by $NCFE(n) = \sum_{i=1}^n \lambda_i / \sum_{i=1}^N \lambda_i$. As a function of n which is the number of ordered eigenvalues incorporated, NCFE converges fast which shows the concentration of energy amongst the first few modes. In particular, $\sum_{n=1}^5 NCFE(n) = 0.9083$, i.e., modes 1–5 cumulate 90.83% of the total fluctuation energy, which in the traditional way, suggests that using modes 1–5 for reconstruction should be

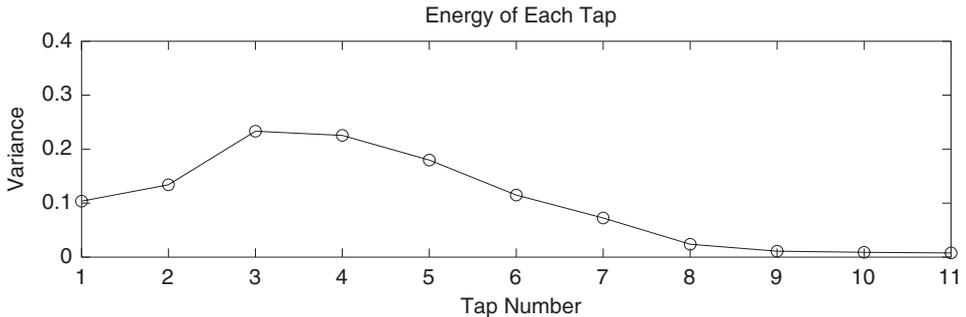


Fig. 3. Energy of each tapping point indicated by variance.

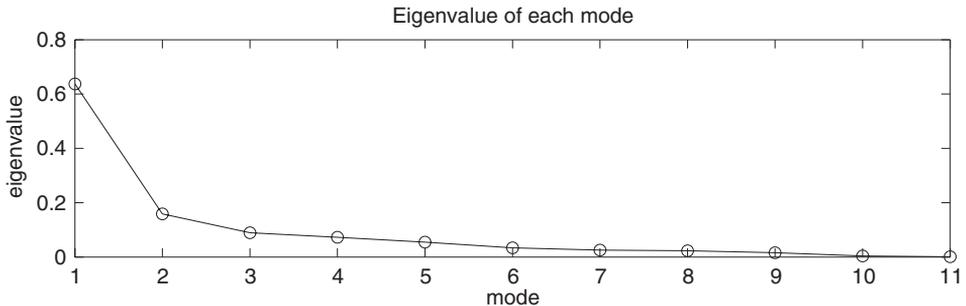


Fig. 4. Eigenvalues of each mode.

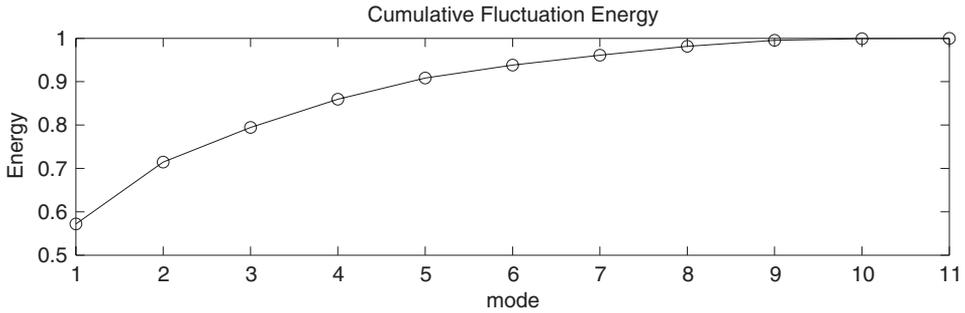


Fig. 5. Normalized cumulative fluctuation energy (NCFE).

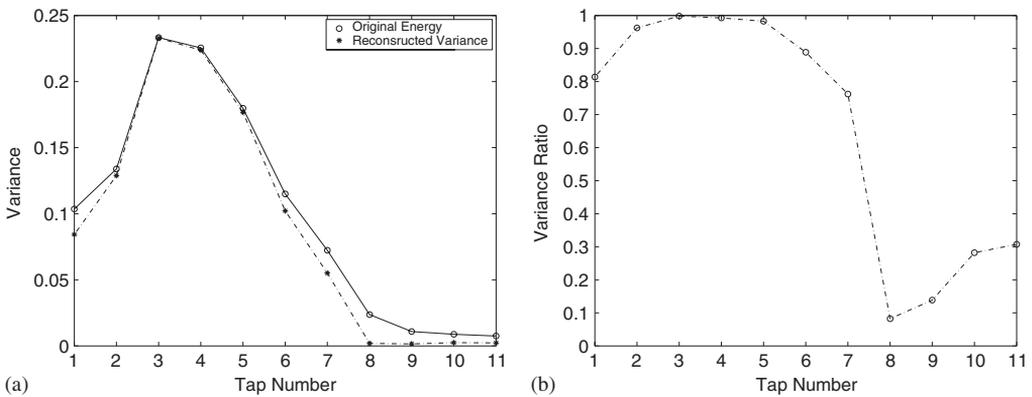


Fig. 6. (a) Variance from reconstructed fluctuation pressure coefficient versus variance from original pressure coefficient, and (b) reconstruction performance measured by variance ratio r_i .

satisfactory. In the following comparisons, we fix the number of modes retained after compression M to be 5.

It could be observed from Figs. 7(a), (b) that although POD has performed its role of coordinate transform well, the degree that the spectra of reconstructed pressure coefficient resemble original data varies among taps.

As discussed in detail in Ref. [2], the unbalanced reconstruction performance at individual taps is due to the fact that POD is completely energy concentration oriented. As a direct result, the first few principal modes favor directions dominated by taps with larger variance. In return, those modes hardly represent the information that is carried by other taps, which results in poor reconstruction performance for these “inferior” taps. We show the quantitative performance evaluation in Fig. 6 below. The reconstruction performance at each tap is measured quantitatively by the ratio between variance of reconstructed signal and original signal, i.e., $r_i = \text{var}(\hat{x}_i)/\text{var}(x_i)$ with i being the tap number, x_i and \hat{x}_i being the original pressure data sequence and reconstructed pressure sequence respectively.

Indeed, the traditional POD method does capture the main energy concentration, but fails to offer a fair representative description for each tap point.

5.3. Results from NPOD

Discussion from previous sections has revealed the reason for unbalanced reconstruction performance at different taps. Taps with higher energy play dominant roles in the selection of modes and taps with relative low energy tend to be neglected in the mode selection process. Since the pressure coefficients at all taps should be reconstructed similarly well, we apply normalization at the first place to compensate for this biased behavior.

Normalization of the raw observation is done by removing the mean and scale the data into unit variance at each dimension (each tap).

$$z_i(t) = \frac{x_i(t) - E(x_i(t))}{\sqrt{E(x_i(t)^2)}} \quad (2)$$

for $i = 1, 2, \dots, N$. Traditional POD is then applied to the modified data and only the first 5 principal components are retained for reconstruction. Let W be the $N \times M$ matrix (in this case $N = 11$, $M = 5$) such that each column of W is the eigenvector of the covariance matrix of Z .

The compression is done by

$$Y = W^T Z \quad (3)$$

or equivalently for $j = 1, 2, \dots, M$,

$$y_j(t) = \sum_{i=1}^N W_{i,j} z_i(t). \quad (4)$$

In the reconstruction process, the modified data is first estimated via

$$\hat{Z} = WY \quad (5)$$

or $\hat{z}_i(t) = \sum_{j=1}^M W_{i,j} y_j(t)$ for $i = 1, 2, \dots, N$. The mapping of modified data Z back into raw data space is done by simply inverting the previous normalization process. Mathematically,

$$\hat{x}_i(t) = \hat{z}_i(t) \times \sqrt{E(x_i(t)^2)} + E(x_i(t)), \quad (6)$$

where the first- and second-order statistics of x_i are previous stored quantities from compression step.

Reconstruction results from proposed NPOD method are shown below together with that from traditional POD-based method, both with the first (corresponding to the highest energy) 5 modes, in Fig. 7. We also demonstrate the reconstruction results quantitatively in Fig. 8 by comparing the variance of the reconstructed data versus the original data. Comparison between Figs. 6 and 8 clearly shows that NPOD offers a much better trade-off performance among taps.

6. Verification with more general data source

In previous section, we have shown that NPOD outperforms traditional POD method in the sense of fair reconstruction with the same pre-chosen number of modes. However, it is not obvious how sensitive this superiority behavior is with respect to the number of modes we use, i.e., if we use more/fewer number of modes, will the relative performance change?

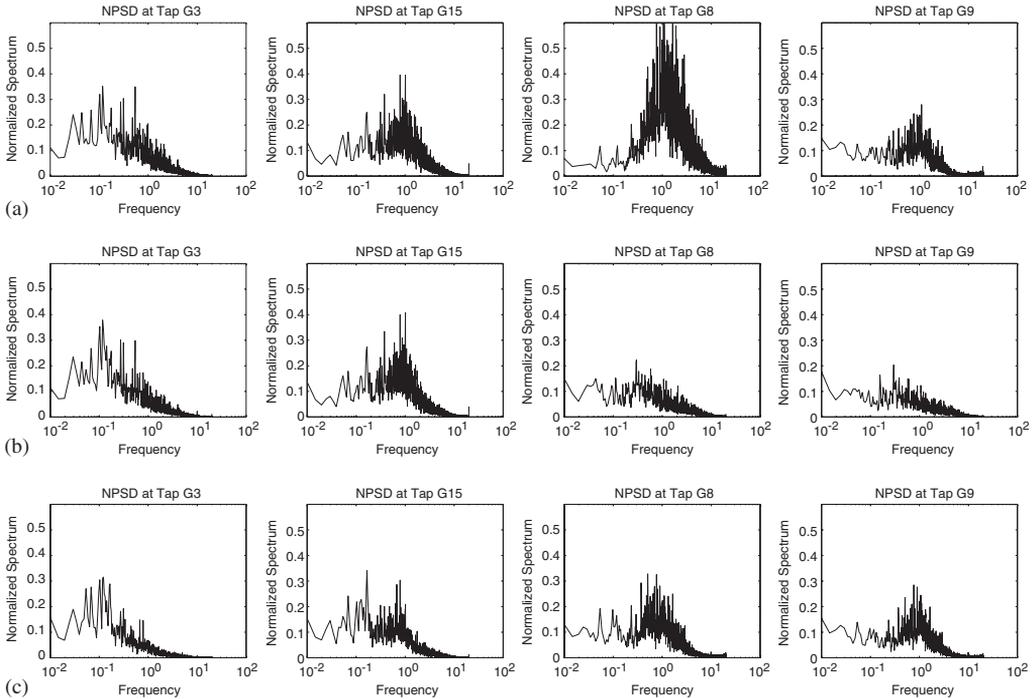


Fig. 7. Fluctuating pressure coefficient spectra at taps G3, G15, G8, G9 (labeled as tap 2, 6, 9, 10, respectively, in numerical analysis): (a) original data, (b) POD reconstructed fluctuating data and (c) NPOD reconstructed data.

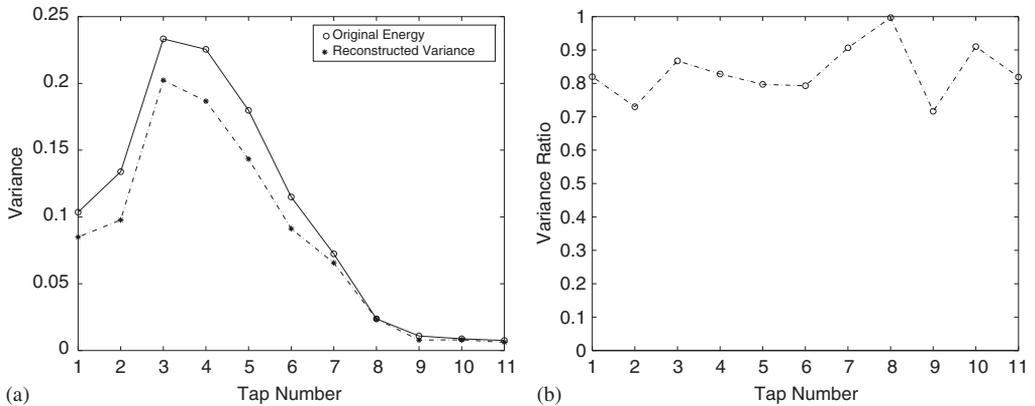


Fig. 8. (a) Variance of reconstructed fluctuation pressure coefficient vs. variance of original pressure coefficient and (b) reconstruction performance measured by variance ratio r_f .

In this section, we exhaustively operate with each possible number of modes and show that NPOD method almost always outperforms POD approach. In other words, it is almost always true that NPOD dominates POD.

6.1. Verification experiment setup

For verification purpose, wind pressure field data from all 204 taps, which are closer to data structure in real applications are used. The allocation of tapping points are indicated in Fig. 9. Shaded points are where further comparison of power spectra reconstruction levels are to be conducted. The mean attacking angle of incident wind in this run is 270.4° with respect to building north. Other setups are identical to the feasibility test situation.

6.2. Numerical results

It can be observed from Fig. 10 that the energy distribution with respect to the tap location confirms the physical fact that taps on the windward slope have in general higher energy than leeward taps.

As before, we arrange eigenvalues in a non-increasing fashion and their corresponding eigenvectors yield modes in traditional POD method with decreasing significance (Fig. 11).

For each possible choice of total mode number for reconstruction, we apply both traditional POD and proposed NPOD. The metric we use to measure performance is the average fluctuation energy reconstruction ratio, i.e., $\bar{r}(M) = \left(\sum_{i=1}^N r_i(M)\right)/N$. Notice that reconstruction procedure are repeated for $M = 1, 2, \dots, N$ number of modes, and

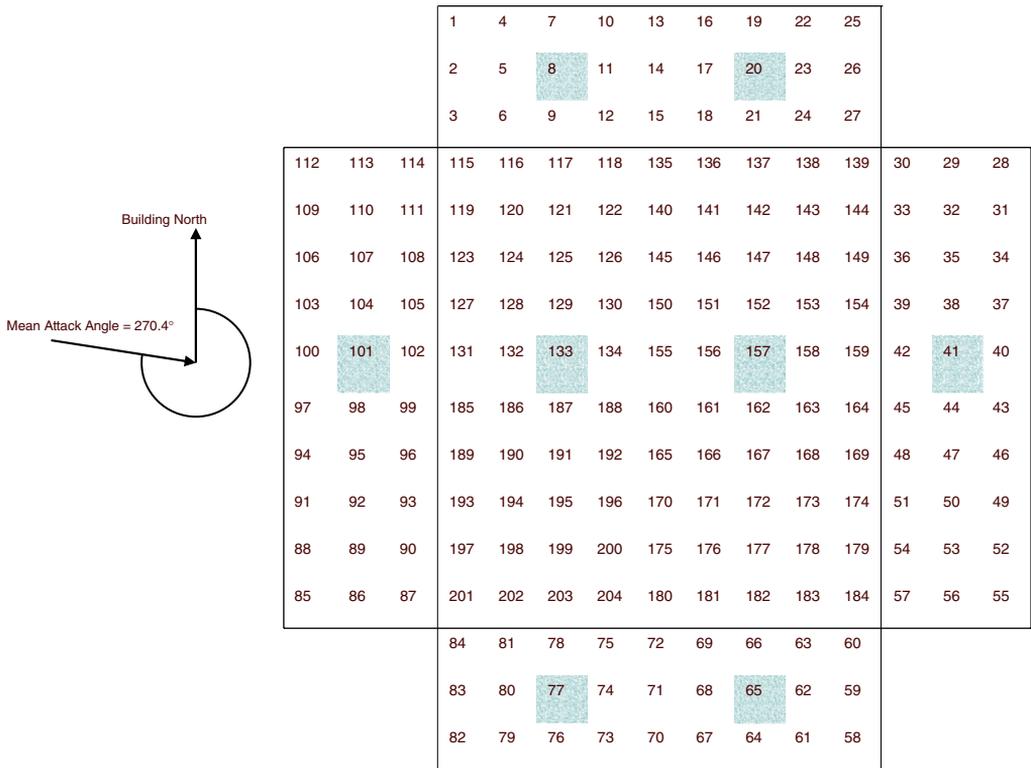


Fig. 9. Tap layout in the verification experiment.

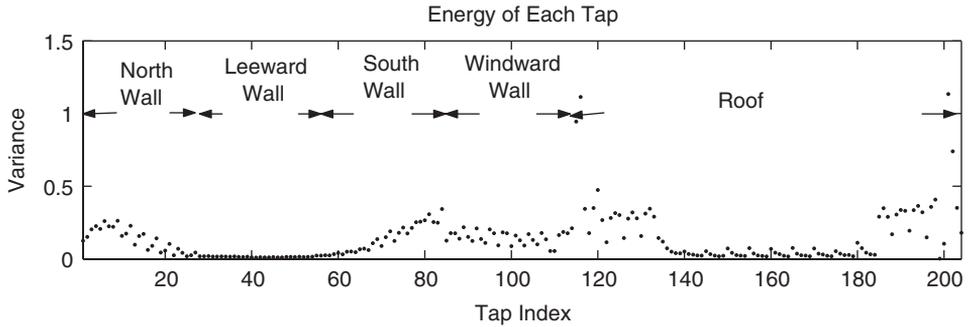


Fig. 10. Energy of each tapping point indicated by variance.

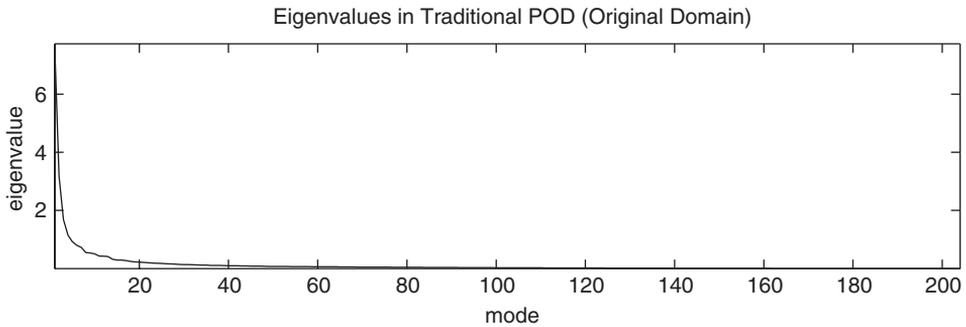


Fig. 11. Eigenvalues of each mode in traditional POD method using raw data.

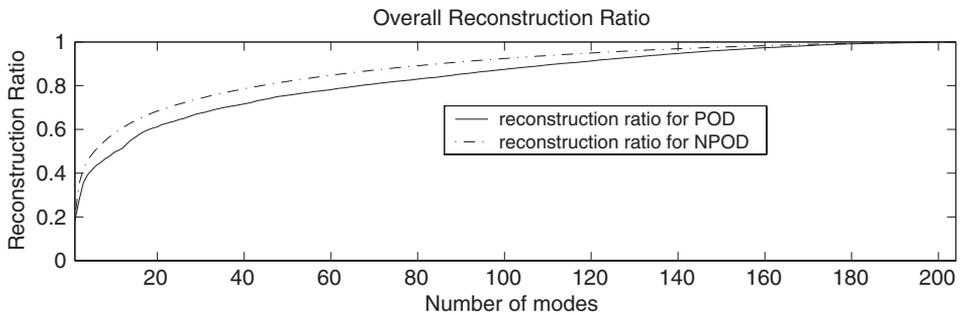


Fig. 12. Overall reconstruction level with different number of modes.

reconstruction performance is a function of M . It is obvious from Fig. 12 that the proposed NPOD method almost always outperforms the traditional POD approach in the sense of average energy reconstruction ratio.

To justify that the average energy reconstruction ratio is indeed a reasonable metric and thus the proposed NPOD approach is a sound one, we take a closer look at the

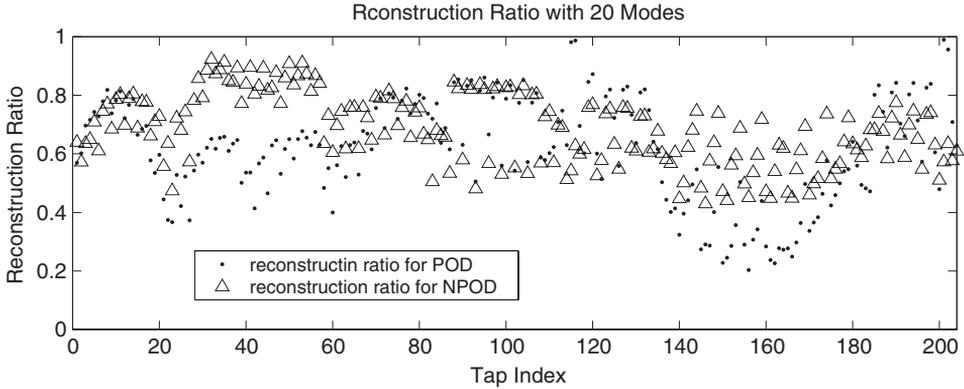


Fig. 13. Comparison of reconstruction level between traditional POD (solid dots) and proposed NPOD (up triangles) with 20 modes.

reconstructed power spectra in case where a total of $M = 20$ modes were used for reconstruction. In fact, the choice of M could be any number between 1 and N . We choose $M = 20$ as it is a sound one in traditional POD setup.

In Fig. 13, we show the energy reconstruction ratio $r_i(20)$ for each individual tap $i = 1, 2, \dots, N$. We can see that NPOD-based reconstruction not only has a more balanced (flat) distribution of reconstruction ratio among different taps, but also higher than POD-based method in general. Particularly, NPOD-based reconstruction has a higher lower bound (worst case performance) than POD-based method. Another interesting observation we can make by looking at Fig. 10 (which we reproduced in the bottom of Fig. 13 for convenience) and POD reconstruction ratio (solid dots) alone in Fig. 13 is that the POD-based method shows a strong-biased performance favoring higher energy taps and tends to ignore taps with relative lower energy, which confirms to our earlier analysis about the blind energy-concentration oriented property of POD.

Reconstruction results for fluctuation pressure coefficient spectra also demonstrate the soundness of NPOD approach as in Fig. 15.

6.3. Robustness test

By balancing the reconstruction ratio among all taps, NPOD does compromise the performance to some degree at the taps with higher energy (windward taps in general). One may wonder what would happen if there are relatively more windward taps used in the decomposition.

Notice that actual physical locations of the taps are transparent to the POD method. The decomposition only looks for the coordinate that best concentrate the energy. Similarly, NPOD only cares about the *relative* energy distribution and makes corresponding adjustment by proper scaling. In other words, NPOD favors reconstructions at taps with relative lower energy, but not necessarily deteriorates performances for taps with absolute high energy. To justify this argument, we exclude taps on the north, south and east walls (tap numbered 1 to 84 in Fig. 9) from The experiment data set and carry out the POD/NPOD compression/reconstruction with 20 modes as before. Note that

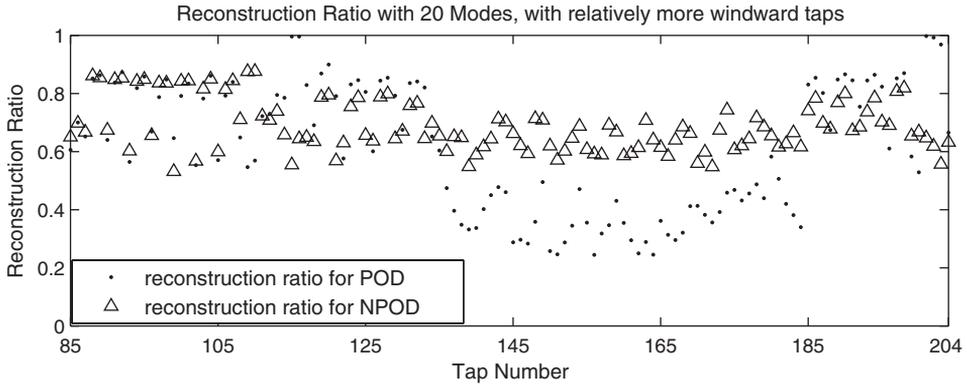


Fig. 14. Comparison of reconstruction level with mostly windward taps between traditional POD (solid dots) and proposed NPOD (up triangle) with 20 modes.

we expect the reconstruction ratio with the same method (either POD or NPOD) to be higher than the original test using data from all taps since we are asking for a lower compression rate at this time. The comparison is shown in Fig. 14.

The experiment results agree with our analysis in that NPOD outperforms POD in general, and in particular, shows an obvious advantage in reconstructing taps 135–184 which corresponds to the far side of roof (Fig. 9) and less energetic taps shown in Fig. 10. Reconstruction performance on other taps are comparable between POD and NPOD (Fig. 15).

7. Interpretation of NPOD

It can be shown that POD is the optimal solution to the optimization problem $\min E_t \|x(t) - \hat{x}(t)\|^2$. If a perfectly balanced reconstruction quality at all taps, i.e., $r_i = r_j \forall i, j \in (1, N)$ is desired, a scaled optimization problem $\min \sum_{i=1}^N ((x_i - \hat{x}_i)/x_i)^2$ needs to be solved. The proposed NPOD method could be looked on as an effort to approach an approximation to such a solution.

One may wonder whether there exist physical interpretations of NPOD that could possibly associate the NPOD pressure modes with the underlying flow phenomena. As NPOD is a normalized variant of the classical POD method, whose physical interpretation is yet open to wide discussion [3,4,6,11], we would like to pose NPOD in a similar way: we are not excluding the potential of possible physical interpretations of NPOD pressure modes but we do believe any conclusive statement requires extensive justification and careful analysis.

In particular, we agree with Holmes et. al [11] in that orthogonality condition is the most dominating factor in shaping modes in POD-based approaches, and any analysis associating the modes to physical causes needs careful justification with extensive flow measurements as well as deep understanding of the underlying flow dynamics.

On the other hand, we would expect the modes to exhibit similar behaviors as those of POD. Of course, due to proper normalization, NPOD modes have a more balanced distribution across the taps than original POD modes. Given these properties, we believe

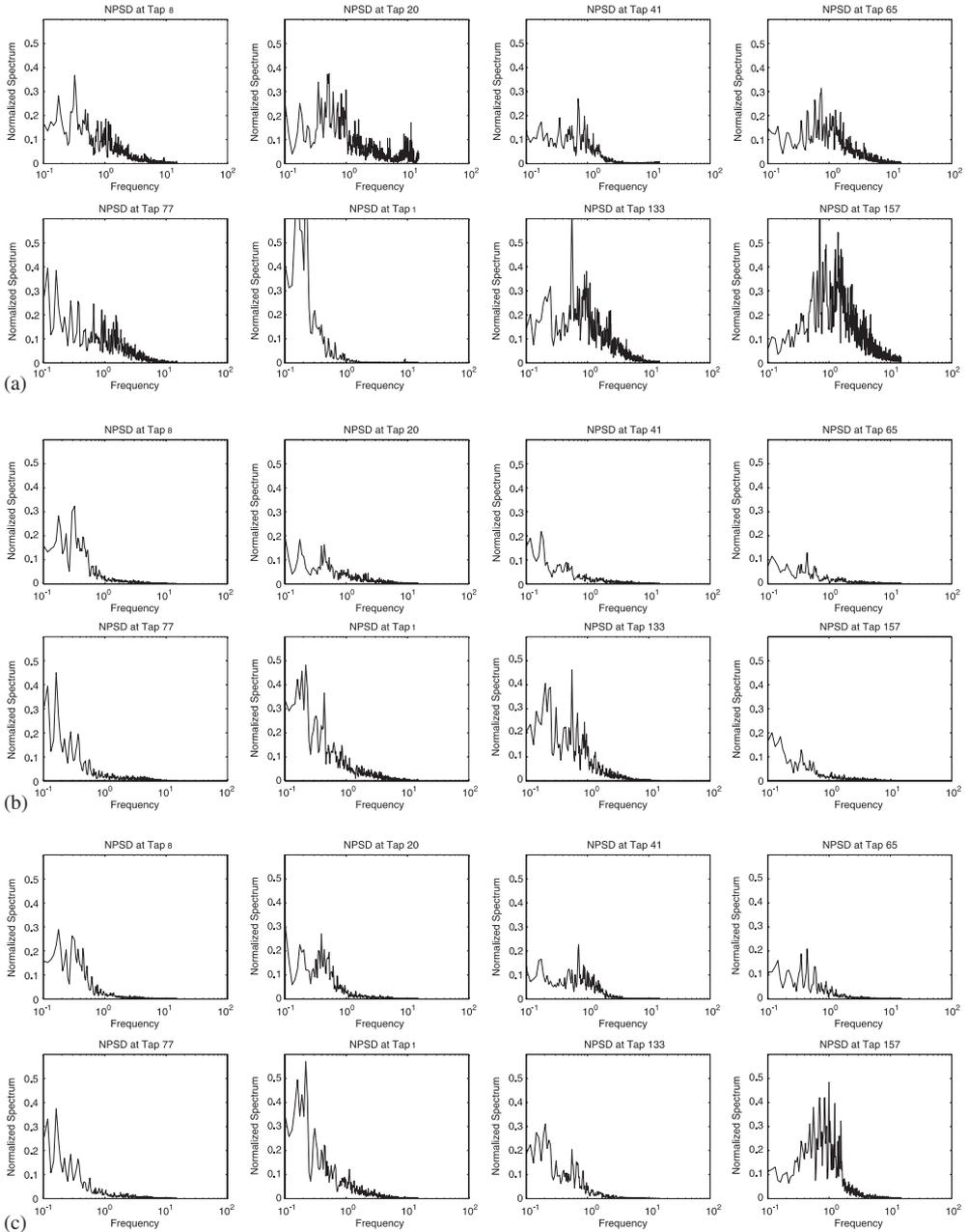


Fig. 15. Fluctuating pressure coefficient spectra at taps 8, 20, 41, 65, 77, 101, 133, 157, respectively: (a) original spectra, (b) POD-based reconstructed spectra and (c) NPOD-based reconstructed spectra.

that any sound association of the POD modes to the underlying flow phenomena (if any) would conveniently export to that of NPOD, even though we are not claiming any particular benefit of using NPOD in this perspective.

8. Conclusions and future work

- (1) Although the most energetic POD modes account for the majority of fluctuating energy, tapping points with lower variance are very likely to be under-represented by only the first few most energetic POD modes. Reconstruction from those modes only leads to substantial error at those tapping points.
- (2) The NPOD method proposed here provides a more balanced reconstruction performance among all taps. It also provides higher lower bound on the reconstruction ratio of all taps. Computation cost is comparable to traditional POD-based method as computing scaling factor is trivial (and actually could be done iteratively when data set is extremely large).
- (3) NPOD could be interpreted as an approximation to a scaled optimization problem.
- (4) Planned future work will be in the temporal compression of time functions ($y_i(t)$'s). More systematic construction of ARMA models (joint structural and parametric optimization) will be investigated. Wavelet transform facilitated approaches will be further explored.

Acknowledgment

The wind pressure coefficient data used in this paper come from the full size wind pressure measurements conducted under the Department of Commerce NIST/TTU Cooperative Agreement Award 70NANB3H5003, which is highly appreciated by the authors. The authors would like specifically thank Dr. Douglas A. Smith and Mr. Stephen Morse at Texas Tech University for their generously providing these data. The authors thank the anonymous reviewers for their constructive and illuminating suggestions.

References

- [1] S. Kho, C. Baker, R. Hoxey, Pod/arma reconstruction of the surface pressure field around a low rise structure, *J. Wind Eng. Ind. Aerodyn.* (90) (2002) 1831–1842.
- [2] D. Ruan, H. He, D. Smith, K.C. Mehta, A Semi-optimal Mode Selection Scheme for Pod Based Compression of Wind Field Data, Seoul, Korea, 2004.
- [3] C.J. Baker, Aspects of the use of proper orthogonal decomposition of surface pressure fields, *Wind Struct.* 3 (2) (2000) 97–115.
- [4] Y. Tamura, S. Suganuma, H. Kikuchi, K. Hibi, Proper orthogonal decomposition of random wind pressure field, *J. Fluids. Struct.* 13 (7–8) (1999) 1069–1095.
- [5] C.W. Letchford, K.C. Mehta, The distribution and correlation of fluctuating pressures on the texas tech building, *J. Wind Eng. Aerodyn.* 50 (1993) 225–234.
- [6] B. Bienkiewicz, H.J. Ham, Y. Sun, Proper orthogonal decomposition of roof pressure, *J. Wind Eng. Ind. Aerodyn.* 50 (1993) 193–202.
- [7] G. Solari, L. Carassale, Modal transformation tools in structural dynamics and wind engineering, *Wind Struct.* 3 (4) (2000) 221–241.
- [8] X. Gilliam, J.P. Duniak, D.A. Smith, F. Wu, Using projection pursuit and proper orthogonal decomposition to identify independent flow mechanisms, *J. Wind Eng. Ind. Aerodyn.* 92 (1) (2004) 53–69.
- [9] S.H. Jeong, B. Bienkiewicz, Application of autoregressive modeling in proper orthogonal decomposition of building wind pressure, *J. Wind Eng. Ind. Aerodyn.* 69–71 (1997) 685–695.
- [10] M.D. Paola, I. Gullo, Digital generation of multivariate wind field processes, *Probab. Eng. Mech.* 16 (2001) 1–10.
- [11] J.D. Holmes, R. Sankaran, K.C. Kwok, M.J. Syme, Eigenvector modes of fluctuating pressures on low-rise building models, *J. Wind Eng. Ind. Aerodyn.* 69–71 (1997) 697–707.

- [12] R.J. Best, J.D. Holmes, Use of eigenvalues in the covariance integration method for determination of wind load effects, *J. Wind Eng. Ind. Aerodyn.* (13) (1983) 371–382.
- [13] J.D. Holmes, Analysis and synthesis of pressure fluctuation on bluff bodies using eigenvectors, *J. Wind Eng. Ind. Aerodyn.* (33) (1990) 219–230.
- [14] J.D. Holmes, Optimised peak load distributions, *J. Wind Eng. Ind. Aerodyn.* (41) (1992) 267–276.
- [15] M.L. Levitan, K.C. Mehta, Texas tech field experiments for wind loads. Part I: building and pressure measurement system, *J. Wind Eng. Ind. Aerodyn.* (43) (1992) 1565–1576.
- [16] M.L. Levitan, K.C. Mehta, Texas tech field experiments for wind loads. Part II: meteorological instrumentation and terrain parameters, *J. Wind Eng. Ind. Aerodyn.* (43) (1992) 1577–1588.