NOTE:

- Please attach the signed copyright release form at the end of your paper and upload as a single ‘pdf’ file
- This coversheet is intended for you to list your article title and author(s) name only
- This page will not appear in the book or on the CD-ROM

Title: Understanding Environmental Effect on Building Vibration for Structural Health Monitoring Using Event Detection and Causal Analysis

Authors: Yixin Zheng
         Lin Zhang
         Pei Zhang
         Hae Young Noh

PAPER DEADLINE: **May 15, 2013**

PAPER LENGTH: **8 PAGES MAXIMUM**

SEND PAPER TO: Professor Fu-Kuo Chang
                Organizing Chairman
                Aeronautics and Astronautics Department
                Stanford University
                Durand Building
                Stanford, California 94305, U.S.A.

                Tel: 650/723-3466
                Fax: 650/725-3377
                E-mail: fkchang@stanford.edu

Please submit your paper in Microsoft Word® format or PDF if prepared in a program other than MSWord. We encourage you to read attached Guidelines prior to preparing your paper—this will ensure your paper is consistent with the format of the articles in the CD-ROM.

NOTE: Sample guidelines are shown with the correct margins. Follow the style from these guidelines for your page format.

Hardcopy submission: Pages can be output on a high-grade white bond paper with adherence to the specified margins (8.5 x 11 inch paper. Adjust outside margins if using A4 paper). Please number your pages in light pencil or non-photo blue pencil at the bottom.

Electronic file submission: When making your final PDF for submission make sure the box at “Printed Optimized PDF” is checked. Also—in Distiller—make certain all fonts are embedded in the document before making the final PDF.
ABSTRACT
Structural damage diagnosis algorithms often involve analyzing vibration responses to extract information about dynamic characteristics closely related to integrity. The vibration data are, however, often influenced by various environmental effects, which may degrade diagnosis performance and result in erroneous decisions. This paper focuses on understanding the consequence of significant environmental effects, such as trains passing nearby, on building vibration data using event detection methods and causal analysis. The event detection methods identify significant events using wavelet analysis, which is effective for decomposing non-stationary signals. The causal analysis allows us to investigate wave propagation patterns in structures by quantifying causal dependencies between measurements collected from different locations using directed information. These methods are applied to acceleration data collected from 40 accelerometers deployed to an 11-story office building located next to railway. The results show clear patterns of causal dependencies among vibration data from different locations in the building, and their patterns change under different environmental conditions. It provides insights about environmental loading effects on building vibration that can be helpful for improving the accuracy of damage diagnosis.

INTRODUCTION
In recent years, there has been an increasing interest in the adoption of sensor network and data processing techniques for Structural Health Monitoring (SHM) to ensure safety and functionality. One of major approaches in this field is vibration-based monitoring using accelerometers, which identify damage by detecting changes in dynamic characteristics of structures (1, 2). While vibration data acquisition systems have been significantly developed, tools to analyze and understand vibration data are

Yixin Zheng and Lin Zhang, Dept. of Electronic Engineering, Tsinghua University, Beijing, China
Pei Zhang, Dept. of Electrical and Computer Engineering, Carnegie Mellon University, Moffett Field, CA, 94035, USA
Hae Young Noh, Dept. of Civil and Environmental Engineering, Carnegie Mellon University, Pittsburgh, PA, 15213, USA
still far from being powerful. Applying modal analysis to the vibration data and comparing to design parameters is a common practice (1, 2), which help verify the overall performance of the structure, but it is not sufficient to understand how and why the vibration characteristics of the structure have changed (3). Moreover, these methods are often vulnerable to varying environmental and operational conditions. When structures are subjected to extreme conditions, such as strong wind, earthquakes, and severe environmental conditions, more sophisticated tools are required to identify those conditions, analyze their consequences on structural vibration, and finally diagnose the structure reliably.

Recent advances in information theory concerning directed information (DI) provides a new perspective for vibration data analysis. DI is a measure that quantifies the causal relationship between two stochastic processes (4, 5). DI was first introduced to characterize feedback channels and has been extensively applied beyond information theory to biology, portfolio theory, data compression, and hypothesis testing (6). While its theoretical properties have been extensively studied in the last few decades, it is not until recently that practical estimators of the DI were developed and proven to be effective in analyzing various data sets (7).

In this paper, we introduce a novel method to detect unusual significant events and characterize their influence on structural vibration using DI estimations. Specifically, our contributions are as follows: (i) We deploy a monitoring system in an 11-story office building next to a railway that continuously collects acceleration responses of the building from multiple sensors. Due to the unique location, we can provide an example of how significantly environmental conditions can affect building vibrations; (ii) We propose an event detection method to identify the signals influenced by relevant events. Signal processing techniques, such as wavelet analysis, are utilized to extract features, after which supervised learning algorithms are applied to classify patterns that correspond to different events. This method is validated with the data set collected from the monitoring system described above to demonstrate that precise detection can be achieved with proper algorithms; (iii) We introduce a causal analysis using DI estimators to characterize building vibrations under the influence of train passings. Using their ability to quantify the causal dependency between stochastic processes, we characterize the relationship among vibrations at different locations to study the propagation of waves stimulated by trains. Results reveal obvious patterns that provide insights for monitoring, assessing, and securing structural health.

**BUILDING MONITORING SYSTEM**

A sensor network that consists of 40 accelerometers is deployed in an 11-story office building to monitor vibration under various environmental conditions. In order to collect vibration under significant events, a building that is 80 – 140m away from railway, as shown in Figure1a, is selected. More than ten trains with heavy loads including cargos and hundreds of passengers pass by the building daily. The network of accelerometers is distributed over the building to record its vibration at different locations and directions. The accelerometers are divided into 16 groups and deployed on the 1st, 5th, 8th and 11th (top) floor of the building, with 4 groups on each floor, whose horizontal positions are illustrated in the vertical view in Figure1a. On the
1st and 11th floor, each group consists of 3 co-located single-axis sensors to monitor the acceleration in west-east, north-south, and vertical directions. On the 5th and 8th floor, however, each group consists of only 2 co-located single-axis sensors to monitor the acceleration in west-east and north-south directions. Coaxial wires are used to connect sensors to a data collector at each floor, which sends the vibration data to a central server through local area networks.

ALGORITHM

Our main objective is to characterize the causal relationship between each pair of vibration data at different locations, under the influence of significant events. Before the causal analysis is applied to the vibration data, an event detection method is required to identify the event occurrences. The event detection involves some challenges. One is that it is difficult to separate the vibration stimulated by our detection target (i.e., train passing) from interferences and noise because the signals caused by environmental events of our interest are often masked by noise, and while the excitation wave travels through complex media including ground and the building structure to the sensor it is attenuated and distorted. Figure 2 shows a fragment of the signals that lasts for 60s at a sample rate of 200Hz, which includes a part influenced by trains and demonstrates the difficulty to identify such influence in our scenario. This is especially the case for signals shown in Figures 2c and 2d, because their locations are further away from the railway (Figure 1a). A camera is installed near the railway to obtain reference data, but the time delay between the appearances of a train in the camera and its influence on the vibration signal makes the event detection non-trivial. This happens because the detection targets and the sensors are not co-located. The goal of the detection method is to train a classification model to detect train passings from vibration data without the information from the camera, but time delay, that is often too significant to ignore, makes it hard to label vibration data as influenced or not. This presents a significant challenge to the application of supervised learning algorithms for event detection. We address this challenge by using time-frequency
domain analysis to extract features that are otherwise opaque in the time domain, determining the delay between the observed events and their influence on sensors using cross-correlation analysis, and then providing accurate labeling of training data to apply supervised learning techniques to discover patterns. Then the causal analysis based on the concept of directed information (DI) is applied for data characterization. The algorithm is summarized in Figure 1b.

**Event Detection Method**

**Feature Extraction:** Various signal processing techniques are applied to vibration data to extract features for event detection. In this paper, we describe two of these techniques and leave others to future work. Fourier analysis is recognized as the most widely-adopted signal processing method, which has computationally efficiency and explicit physical interpretation. We used Fourier analysis to investigate the power spectral density (PSD) of the vibration signal and found that, while the PSD reveals some signatures from the train, such as larger peaks around 10Hz, they were not consistent and reliable enough to benefit from classification techniques. On the other hand, wavelet analysis is another signal processing tool that linearly decomposes data into time-localized waveforms of various frequencies, referred to as wavelets.

Unlike Fourier analysis, wavelet analysis represents time evolution of frequency contents, and thus appropriate for analyzing the influence of a train that changes structural vibration in both time and frequency aspects. Application of wavelet analysis in SHM is summarized by Noh et al. Figures 3a and 3b present the wavelet analysis results of two vibration data collected from sensor B and C, respectively, using Continuous Wavelet Transform (CWT). The lighter the color is, the larger the coefficients are, indicating higher vibration energy at the frequency and time. We can clearly observe lighter areas in both figures, around 20 – 40 in time and 12 – 24 in scale, which reflects the influence of the passing train. Note that for Daubechies3 wavelet we use, a scaling factor of 16 corresponds to 10Hz. Fourier-based features are defined as the frequency and the corresponding PSD for three largest peaks, while wavelet-based features are average absolute coefficients at each scale from 1 to 32.
**Event Labeling:** To apply supervised learning algorithms to the vibration data to detect train passing events, each signal needs to be accurately labeled in order to train a classifier. As mentioned above, simply matching camera observations with corresponding signals does not work in our scenario because of the delay that generally exists between the camera observation of events and their influence on the sensor measurements. We therefore introduce a method to determine this delay using cross-correlation analysis. First, measurements from a representative sensor are segmented into small time fragments, and a corresponding binary ‘observation sequence’ is obtained from camera observations. Trains are either north or south bound, and separate sequences are created for each direction. The next step identifies an ‘influence sequence’, which indicates whether each signal fragment is influenced by a train. This influence sequence is the label we aim to obtain for training a classifier. This indicator can be obtained from either Fourier or wavelet analysis and a simple threshold technique. After obtaining observation and influence sequences, their cross-correlation is computed to determine the time delay. The time lag with the maximum cross-correlation value is regarded as the most probable delay. The result of cross-correlation analysis is shown in Figures 3c and 3d, which demonstrate a significant delay between observations of events and their influences on the measurements. We can observe that the time shift is +5 and -6 seconds for north and south bounds, respectively. Combining this delay computed from a representative sensor and the ‘observation sequence’ from the camera, we can label all training data.

**Classification:** The final step for event detection is to apply supervised learning algorithms using the labeled data for training. Due to the space limit, only two classifiers are considered for our application, the Random Forest (RF) (10) and the Bayesian Network (BN) (11). RF is an ensemble algorithm that uses randomized subsets to train a forest of decision trees and choose the mode of their outputs as the classification result. BN learns the conditional dependencies of the random input variables to form a probabilistic graph, and produce outputs accordingly.

**Causal Analysis**

Causal analysis provides causal dependencies among measurements from different locations using DI, which is a measure to quantify causal influence between two stochastic processes. Using the event detection method, we can analyze how the causal relationship changes when a train is present. Let uppercase letters $X, Y, \ldots$ refer to random variables, while lowercase letters $x, y, \ldots$ refer to their values. Taking $n$ as the length of a signal, $X^n$ denotes the n-tuple random variables $(X_1, X_2, \ldots, X_n)$. The alphabets of $X$ is in its calligraphic form, $\mathcal{X}$. In information theory, the Shannon entropy $H(X)$ quantifies the uncertainty in $X$ and is defined as (12)

$$H(X) \triangleq -\sum_{x \in \mathcal{X}} p_X(x) \log(p_X(x))$$

where $p_X(x)$ is the probability distribution of $X$. The mutual information (MI) of $X$ and $Y$, $I(X; Y)$, is a measure of mutual dependence or the information that $X$ (or $Y$)
contains about $Y$ (or $X$). $I(X; Y)$ is defined as

$$I(X; Y) \triangleq \sum_{x \in X} \sum_{y \in Y} p_{X,Y}(x, y) \log \frac{p_{X,Y}(x, y)}{p_X(x)p_Y(y)} = H(Y^n) - H(Y^n | X^n)$$

The DI and DI rate (13) from $X^n$ to $Y^n$ represents the information flow and are given as

$$I(X^n \rightarrow Y^n) \triangleq \sum_{i=1}^{n} I(X^i; Y^i | Y^{i-1}) = H(Y^n) - H(Y^n \parallel X^n)$$

$$\bar{I}(X^n \rightarrow Y^n) \triangleq \lim_{n \rightarrow +\infty} \frac{1}{n} I(X^n \rightarrow Y^n)$$

where $H(Y^n \parallel X^n)$ denotes causally conditional entropy (13):

$$H(Y) \triangleq -\sum_y p(y) \log(p(y))H(Y^n \parallel X^n) \triangleq \sum_{i=1}^{n} H(Y_i | Y^{i-1}, X^i)$$

Note that $I(X^n \rightarrow Y^n) \neq I(Y^n \rightarrow X^n)$. For the proof of the existence of the limit and its relationship with entropy rate and causally conditional entropy rate, please refer to (7, 13) for detail.

Despite extensive research in theoretical properties of DI, practical estimation algorithms for $\bar{I}(X^n \rightarrow Y^n)$ has been developed only recently. To achieve practical estimations, the Context Tree Weighting (CTW) algorithm (14) offers an effective and reliable method to derive universal probability assignment from data with unknown probabilistic model and parameters, and thus enables the estimation of various information measures, including DI. Moreover, CTW has linear computational and storage complexity with respect to data length and directly provides probability assignment. We applied four CTW-based DI estimators that are introduced by Jiao et al.(7).

**APPLICATION**

We used the 10-fold cross validation method to evaluate the classification accuracy. The data set contains 440 seconds of vibration data sampled at 200Hz from sensor A in Figure 1a, on the 11th floor. Among the data set, 140 seconds of data are labeled ‘Train’ and 300 seconds labeled ‘No Train’. We first divided the vibration data...
into fragments (windows) and extracted Fourier- and wavelet-based features from each fragment. Detection accuracy was evaluated for various combinations of window length, feature, and classification techniques, as shown in Figure. The detection accuracy was above 85% for all combinations, and even accuracy above 97% was achievable. For wavelet-based features, the accuracy improved as the window length increases, approaching 100% in the end, which was not the case for Fourier-based feature. The comparison result reveals that wavelet analysis performs better, while the RF and the BN perform similarly with this data set. Note that the detection accuracy is measured fragment-wise, not event-wise, because our objective is to determine whether each fragment is influenced by the train and characterize the influence.

DI-based causal analysis was applied to the vibration data that are labelled using the event detection method. We divided the data into two parts depending on whether they are influenced by the train or not. Then the DIs are estimated to investigate causal influences among each pair of sensor measurements and how they change under train loading. The results from the four sensors recording vertical vibration on the 11th floor are discussed in this section. Since the estimation of DI requires the measurements to be discrete and relatively small, the sensor readings are discretized into 5 levels, considering computational constraints (7). Figure 4b shows the DI from sensor B to C under the influence of train passings. The inverse DI (inv DI), which represents the DI from sensor C to B, is significant, while the DI approaches zero, implying that there exists strong causal dependency from sensor C to B when train passes. Note that the MI is the sum of DI and inv DI.

We conducted similar analysis for all combinations of four sensors to obtain a causal graph, and the results are illustrated in Figure 5. The patterns of causal graphs show that the train load affects the causal dependencies among different locations. In particular the following observations are made: (i) for free vibration (no train), the vibration at the edges of the building leads the center of the building; (ii) for forced vibration (train), the vibration at the south wing (perpendicular to the railway) of the building leads the north wing (parallel to the railway); and (iii) causal dependencies among vibrations at different locations show clear patterns, and they change as the environmental load changes. Observation (i) agrees with a physical intuition that the center of the building is stiffer than the edges. As a result, the edges are more likely to be triggered to vibrate from random ambient noise, and then the vibration propagates into the center. On the other hand, when there is a specific directionality in the

![Figure 5. Directed information analysis result](image-url)
loading, such as the train pass, observation (ii) implies that the wing perpendicular to the direction of the load is more likely to be triggered to vibrate, which propagates to the other wing. Further investigation is necessary, however, to validate whether these intuitions are true using expertise in structural engineering, soil-structure interaction, and geomechanics. Proper modal analysis combined with numerical modeling may help to address these questions. The results are, however, promising that DI-based causal graph can be used for characterizing building vibrations and their changes due to environmental effects and potentially damage.

CONCLUSION

This paper introduces an event detection method and causal analysis to identify and characterize the influence of significant environmental loads to structural vibration. For event detection, features are extracted using Fourier and wavelet analysis from acceleration measurements, and then supervised learning algorithms are applied for classification. The causal analysis characterizes how the loads influence the causal relationships of structural vibrations using DI. We deployed a vibration monitoring system in an office building near railway to validate the algorithm. The results show that there are clear causal dependencies among measurements collected from different locations, and the pattern of the causal relationship changes as the environmental condition changes due to train passing by, showing a strong potential for its use in SHM. Future research need to be conducted to physically interpret causal graphs.

REFERENCES