

## Effect of Nonlinear Characteristics on Applicability of Extended Correlation Anomaly Detection to Structure-portfolio Monitoring

T. Yaoyama<sup>1</sup>, T. Hida<sup>2</sup> and T. Takada<sup>3</sup>

<sup>1</sup>The University of Tokyo, JSPS Research Fellow. Email: yaoyama@load.arch.t.u-tokyo.ac.jp

<sup>2</sup>The University of Tokyo. Email: hida@load.arch.t.u-tokyo.ac.jp

<sup>3</sup>Japan Atomic Energy Agency. Email: takada.tsuyoshi@jaea.go.jp

**Abstract:** In large-scale earthquakes, a technique for collectively and remotely monitoring a variety of structures (structure-portfolio monitoring) is expected to be an effective solution. We have proposed a novel data-driven technique for structure-portfolio monitoring based on a method called Correlation Anomaly Detection (CAD), which was originally proposed in a machine learning community and has been improved for vibration systems in our previous papers as Extended Correlation Anomaly Detection (ECAD). These studies, however, have assumed that Single-Degree-Of-Freedom (SDOF) structures linearly behave during excitation. The paper therefore investigates how the nonlinearity of structural responses influences the applicability of ECAD to structure-portfolio monitoring. Specifically, we numerically evaluate the detection performance of ECAD in different cases of hysteretic models of SDOFs.

**Keywords:** emergency responses, structure-portfolio, damage detection, anomaly detection, machine learning

### 1. Introduction

For rapid planning of emergency responses immediately after large earthquakes, such as evacuation warnings and recovery operations, it is essential to quickly evaluate the damage of structures (buildings and infrastructures) and prioritize their necessities to respond. Because of the limitation of human resources available for visual inspection, a technique for collectively and remotely monitoring a variety of structures i.e., *structure-portfolio monitoring* is expected to be an effective solution.

We have proposed a novel data-driven technique for structure-portfolio monitoring based on a method called Correlation Anomaly Detection (CAD), which was originally proposed in a machine learning community (Ide et al. 2009) and has been improved for vibration systems in our previous papers as Extended Correlation Anomaly Detection, ECAD (Yaoyama et al. 2019; Yaoyama et al. 2020). In CAD, a change of graph-structure of monitored variables evaluated as a covariance matrix, is tracked to score *correlation anomaly* for each variable. In contrast, ECAD uses a co-spectrum matrix, which is a covariance matrix expanded into the frequency domain, in order to capture correlation anomaly in the frequency content of data. In our problem, ECAD detects structural damage in a portfolio by evaluating the graph structure of seismic responses collected from structures.

Main advantages of the proposed technique are: (a) it is aimed at collectively monitoring on large portfolio of structures, which gives a unique perspective that other techniques do not have; (b) it requires neither observations of input ground motions (i.e., a kind of output-only technique) nor sensor synchronization, which allows cost-effective implementation; (c) it also requires neither prior knowledges nor assumptions on structural characteristics, which gives its wide applicability.

Our previous studies have suggested the applicability of ECAD to structure-portfolio monitoring and its superiority to CAD by performing numerical experiments on a portfolio composed of SDOF models (Yaoyama et al.

2019; Yaoyama et al. 2020). These studies, however, have assumed that SDOF structures linearly behave during excitation, and not considered the nonlinearity of structural responses. The paper therefore investigates how the nonlinearity of structures influences the applicability of ECAD to structure-portfolio monitoring. Specifically, we consider two different cases of hysteresis models: ordinary bilinear and modified Clough's, and examine the effects of the difference on detection performance.

### 2. Methodologies

#### 2.1 Correlation Anomaly Detection (CAD)

The problem of Correlation Anomaly Detection (CAD, Ide et al. 2009) is stated as follows. Consider a system composed of  $M$  variables and let  $\mathcal{D} \equiv \{\mathbf{x}(t) \in \mathbb{R}^M | t = 1, \dots, N\}$  denote a set of time-series observations for a certain period. Specifically, we define a set of data collected under normal conditions as reference data  $\mathcal{D}_{\text{ref}}$ , and one which may include abnormal conditions of some variables as test data  $\mathcal{D}_{\text{test}}$ . All variables are assumed to be standardized to zero mean and unit s.d., because we are interested only in their correlation. The aim of CAD is the quantification of *correlation anomaly* for each variable by comparing graph structures between  $\mathcal{D}_{\text{ref}}$  and  $\mathcal{D}_{\text{test}}$ .

The algorithm is described as follows. First, each data  $\mathcal{D}$  is modelled as Gaussian Graphical Model (GGM):

$$\mathcal{N}(\mathbf{x}(t) | \mathbf{0}, \Lambda^{-1}) = \frac{\det(\Lambda)^{1/2}}{(2\pi)^{M/2}} \exp\left(-\frac{1}{2}\mathbf{x}(t)^\top \Lambda \mathbf{x}(t)\right), \quad (1)$$

where  $\Lambda \in \mathbb{R}^{M \times M}$  denotes an inverse covariance matrix, i.e., a precision matrix. In order to obtain the optimal  $\Lambda^*$ , penalized maximum likelihood estimation called *sparse structure learning* is adopted;

$$\Lambda^* = \arg \max_{\Lambda} [\log \det(\Lambda) - \text{tr}(\hat{\Sigma}\Lambda) - \rho \|\Lambda\|_1], \quad (2)$$

where  $\text{tr}(\cdot)$  and  $\|\cdot\|$  respectively denote the trace and  $L_1$  norm of a matrix;  $\hat{\Sigma}$  represents the sample covariance

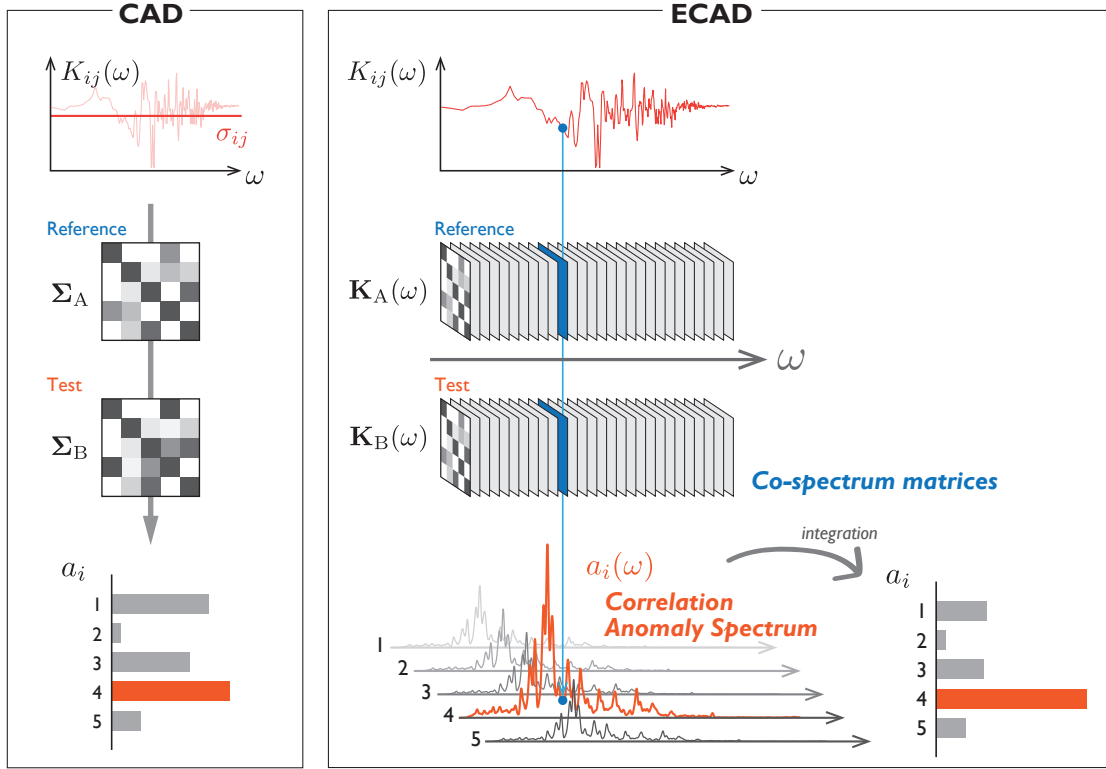


Figure 1. Comparison between CAD (Ide et al. 2009) and ECAD (Yaoyama et al. 2020)

matrix. The final term of the objective function equation (2) is the penalty term introduced to controll the sparsity of  $\Lambda^*$ , which allows the calculation stability of the matrix inversion and the noise-robust estimation of a graph structure. When the penalty coefficient  $\rho = 0$ , the solution is obtained as  $\Lambda^* = \hat{\Sigma}$ . Equation (2) can be solved by *graphical lasso* algorithm (Friedman et al. 2008).

Using the precision matrices  $\Lambda_{\text{ref}}$ ,  $\Lambda_{\text{test}}$  and their inverse  $\Sigma_{\text{ref}}$ ,  $\Sigma_{\text{test}}$ , obtained as above for both  $D_{\text{ref}}$  and  $D_{\text{test}}$ , *correlation anomaly* can be estimated using the expected Kullback-Leibler divergence:

$$d_i^{\text{ref-test}} \equiv \iint p_{\text{ref}}(\mathbf{x}_{\setminus i}) p_{\text{ref}}(x_i | \mathbf{x}_{\setminus i}) \log \frac{p_{\text{ref}}(x_i | \mathbf{x}_{\setminus i})}{p_{\text{test}}(x_i | \mathbf{x}_{\setminus i})} d\mathbf{x}_i d\mathbf{x}_{\setminus i}, \quad (3)$$

where  $\mathbf{x}_{\setminus i} \equiv \{x_1, \dots, x_{i-1}, x_{i+1}, \dots, x_M\}^\top$ , and  $p_{\text{ref}}(\mathbf{x})$  and  $p_{\text{test}}(\mathbf{x})$  represent GGMs respectively defined with  $\Lambda_{\text{ref}}$ ,  $\Lambda_{\text{test}}$ . The correlation anomaly score of each variable  $x_i$  is then defined as  $a_i \equiv \max\{d_i^{\text{ref-test}}, d_i^{\text{test-ref}}\}$ , where  $d_i^{\text{test-ref}}$  is defined by replacing ‘ref’ and ‘test’ in equation (3). The analytical solution of equation (3) is obtained in the literature (Ide et al. 2009).

## 2.2 Extended Correlation Anomaly Detection (ECAD)

Extended Correlation Anomaly Detection (ECAD, Yaoyama et al. 2020) uses covariance matrices expanded in the frequency domain, i.e., co-spectrum matrices, and

respectively applies CAD to co-spectrum matrices at each discretized frequency (Fig. 1).

First, we define the sample co-spectrum matrix  $\hat{\mathbf{K}}(\omega_k)$  for discretized circular frequency  $\omega_k = 2\pi(k-1)/N\Delta t$  ( $k = 1, \dots, N/2$ ).

$$\hat{\mathbf{K}}(\omega_k) \equiv \frac{2\pi}{N\Delta t} \text{Re} [\mathbf{X}^*(\omega_k) \mathbf{X}^\top(\omega_k)] \quad (4)$$

where  $\text{Re}[\cdot]$  denotes the real part of complex functions;  $\mathbf{X}(\omega)$  and  $\mathbf{X}^*(\omega)$  are respectively the vectors whose  $i$ -th element is the Fourier transform of  $X_i(\omega_k)$  and its conjugate  $X_i^*(\omega_k)$ .

The algorithm of ECAD is then formulated as follows. First, for  $D_{\text{ref}}$  and  $D_{\text{test}}$ , which are respectively normalized to zero mean and unit s.d., compute the co-spectrum matrices  $\hat{\mathbf{K}}_{\text{ref}}(\omega_k)$  and  $\hat{\mathbf{K}}_{\text{test}}(\omega_k)$ . Next, apply ECAD to  $\hat{\mathbf{K}}_{\text{ref}}(\omega_k)$  and  $\hat{\mathbf{K}}_{\text{test}}(\omega_k)$  for  $\omega_k$  ( $k = 1, \dots, N/2$ ) to obtain a series of correlation anomaly scores  $\{a_i(\omega_k)\}$ , which we call Correlation Anomaly SPectra (CASP). Finally, integrate  $\{a_i(\omega_k)\}$  in terms of  $\omega_k$  to obtain averaged correlation anomaly score:

$$\bar{a}_i \equiv \sum_{k=1}^{N/2} a_i(\omega_k) \Delta\omega, \quad (5)$$

where  $\Delta\omega = 2\pi/N\Delta t$ .

## 3. Problem Statement

The problem the paper tackles is stated as follows.

Let us consider a portfolio of  $M$  structures spatially distributed over a wide area. The acceleration response of each structure is monitored with a sensor installed on the top of each structure, and therefore every earthquake that strikes this area yields a set of data  $D$  composed of  $M$  variables. For evaluating structural damages in a portfolio, two experienced earthquakes are selected and thus two sets of data are determined: reference data  $D_{\text{ref}}$ , which is collected under assumably normal (undamaged) conditions, and test data  $D_{\text{test}}$ , in which some structures might have experienced damages (i.e., ductile behavior). In structure-portfolio monitoring, ECAD is applied to these  $D_{\text{ref}}$  and  $D_{\text{test}}$  to obtain correlation anomaly score for each structure as a metric for the possibility of structural damages.

In the following part, we examine the applicability of ECAD to structure-portfolio monitoring described above by some numerical experiments on a portfolio of single degree of freedom (SDOF) systems, in which especially two cases of hysteresis are assumed: ordinary bilinear models and modified Clough's models.

#### 4. Experiments

##### 4.1 Models and Conditions

Let us consider a portfolio of 18 structures modelled as SDOFs. The mass of all SDOFs is assumed to be  $10^5[\text{kg}]$ . The initial stiffnesses are given by sampling the corresponding natural frequencies from lognormal distribution with mean 1.0 [Hz] and c.o.v. 0.3. Initial stiffness proportional damping is assumed with the damping ratio 0.03 for all SDOFs. Two nonlinear response analyses are performed on the SDOFs to obtain two sets of the absolute response accelerations, which are normalized to zero mean and unit s.d. in terms of each SDOF and defined respectively as  $D_{\text{ref}}$  and  $D_{\text{test}}$ .

The assumed locations of all SDOFs, as shown in Fig. 2, correspond to the selected stations in the seismograph networks K-NET (National Research Institute for Earth Science and Disaster Resilience 2019). The SDOFs are assumed to be excited by strong motions recorded at the corresponding stations in order to represent the difference of input ground motions in a portfolio, and numbered corresponding to the codes of the K-NET stations. For example, the SDOF located on the station TKY001 is called the SDOF#001. The target earthquakes EQ. A and B specified in Table 1 are respectively used for generating  $D_{\text{ref}}$  and  $D_{\text{test}}$ . In order to generate 'anomalies' (i.e., damaged SDOFs) in  $D_{\text{test}}$ , the amplitude of the ground accelerations input to randomly selected SDOFs are increased to 1000% in order to trigger their structural damage (specifically, ductility in this study). In this study, SDOF#014, 016, 022 and 023 are selected and respectively subjected to the amplified ground accelerations with the peak of 134.2, 92.9, 149.9 and 137.9  $[\text{cm/s}^2]$ .

Two cases of hysteresis are assumed: ordinary bilinear

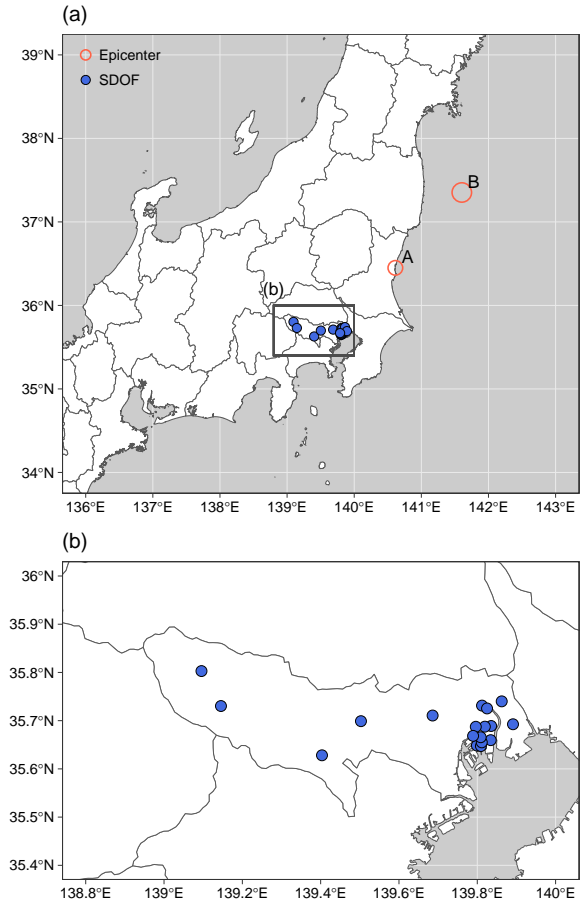


Figure 2. Assumed locations of SDOFs and the epicenters of considered earthquakes. The radius of circles for epicenters is proportional to the magnitude of earthquakes.

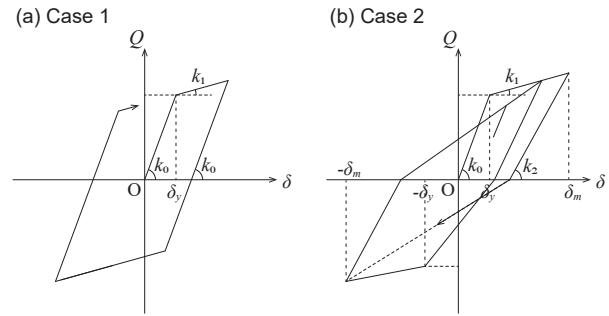


Figure 3. Diagrams of hysteresis models considered in numerical experiments: (a) ordinary bilinear and (b) modified Clough's.

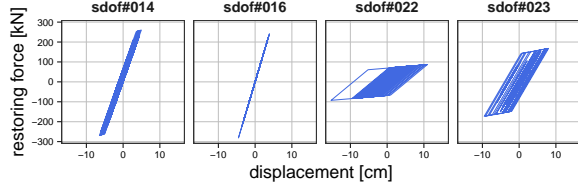
(in Case 1) and modified Clough's (in Case 2), as shown in Fig. 3. In ordinary bilinear, the post-yield stiffness is set as  $k_1 = k_0/10$ ; in modified Clough's, the post-yield stiffness is also  $k_1 = k_0/10$  and the unloading stiffness is  $k_2 = k_0(\delta_y/\delta_m)^{0.5}$ , where  $\delta_y$  and  $\delta_m$  denote a yield displacement and a maximum displacement ever experienced, respectively. The yield displacement in both models are set as  $\delta_y = 5[\text{cm}]$ .

We apply ECAD to  $D_{\text{ref}}$  and  $D_{\text{test}}$  obtained in the above procedure and evaluate its performance, where the penalty

Table 1. Earthquakes considered in numerical experiments.

EQ.	Time	Epicenter	Depth	$M_j$
A	07/27/2016 23:47	36.5°E 140.6°N	57km	5.4
B	11/22/2016 05:59	37.4°E 141.6°N	25km	7.4

(a) Case 1 (ordinary bilinear)



(b) Case 2 (modified Clough's)

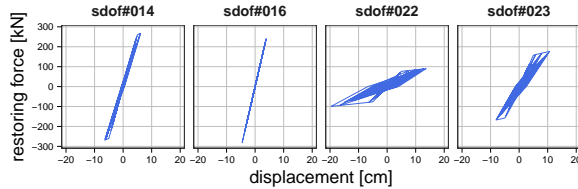


Figure 4. Hysteresis curves of all SDOFs in test data in (a) ordinary bilinear and (b) modified Clough's.

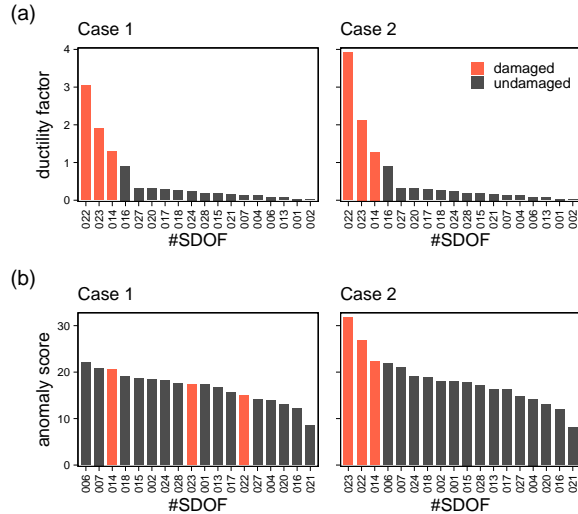


Figure 5. (a) The ductility factors of all SDOFs observed for  $D_{test}$  in both cases. (b) Correlation anomaly scores computed for all SDOFs in both cases.

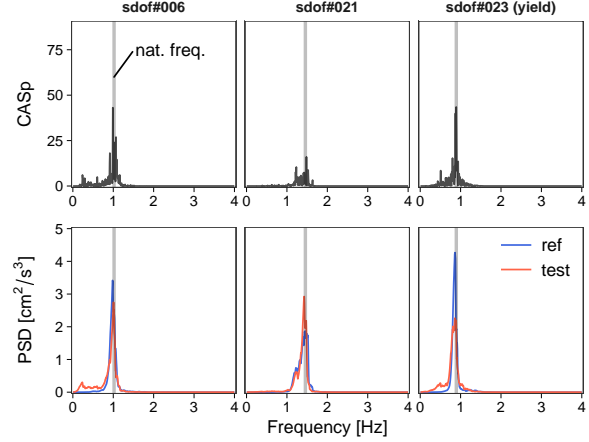
coefficient is  $\rho = 0.01$ .

#### 4.2 Results and Discussions

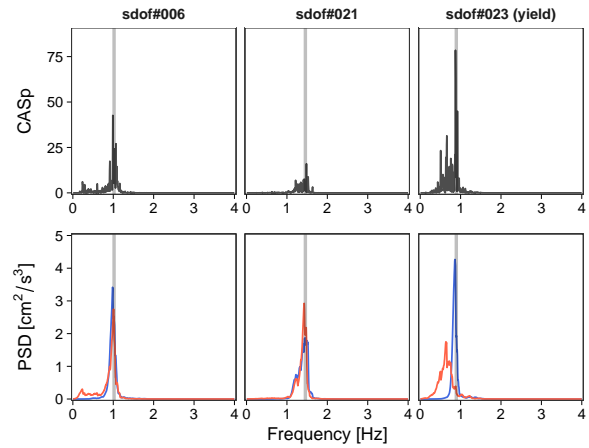
Fig. 4 shows the hysteresis curves of all SDOFs observed in the analyses for generating  $D_{test}$ . It can be seen for both cases that the SDOFs excited by intentionally amplified ground motions have actually experienced yielding except SDOF#016.

Fig. 5 shows for all SDOFs (a) the ductility factors observed and (b) the correlation anomaly scores computed by ECAD, both in the descending order. In Case 1 (ordinary bilinear), SDOF#014, one of the SDOFs which have a ductility factor larger than one, i.e., the 'damaged' SDOFs, ranks relatively higher (3rd) in the order of anomaly score, while the others, SDOF#023 and #022 rank middle or lower (9th and 13rd). This means that the latter two anomalies are overlooked. In Case 2 (modified Clough's), in contrast, all of the damaged SDOFs have higher scores than the undamaged SDOFs, which means, the anomaly detection

(a) Case 1



(b) Case 2



(c) input

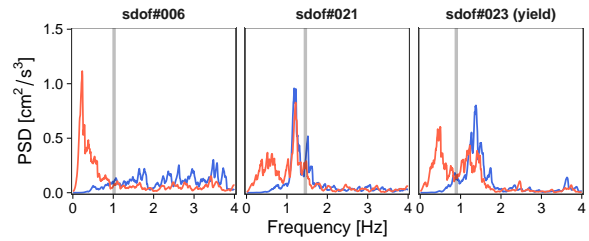


Figure 6. (a–b) Correlation Anomaly Spectra (CASP) computed for selected SDOFs and the response PSDs of the SDOFs both in  $D_{ref}$  and  $D_{test}$ , plotted respectively for (a) Case 1 and (b) Case 2. (c) The input PSDs of the SDOFs.

shows a good performance. These results imply that ECAD has a good applicability to a structure with a dynamic characteristic similar to a modified Clough's model, not to one with a characteristic similar to an ordinary bilinear model.

To discuss the results, Fig. 6(a, b) show for both cases the Correlation Anomaly Spectra (CASP) computed for selected SDOFs and also the response PSDs in both  $D_{ref}$  and  $D_{test}$ , smoothed with Parzen window of bandwidth 0.2[Hz]. Fig. 6(c) shows the PSDs of the corresponding inputs which are normalized to zero mean and unit s.d.

The natural frequencies determined by the initial stiffnesses are shown with gray lines in the figures. Three SDOFs are selected for the following reasons: #023 has a higher ductility factor and actually shows the highest score in Case 2, but ranks lower in Case 1; #021 with no structural damage shows the lowest scores in both cases; #006 shows the highest score in Case 1 and also the highest among undamaged SDOFs in Case 2, in spite of no structural damage.

The difference in the rank of SDOF#023 between both cases is apparently due to characteristics in the frequency domain. In Case 1, the CASP of SDOF#023 (damaged) has characteristics similar to that of #006 (not damaged), in terms of the height of its peaks and the width of the frequency band in which strong anomalies can be observed. In Case 2, on the other hand, the CASP of #023 has a much higher peak than #006, which is one of the reasons why the anomaly of #023 can be detected in Case 2. Moreover, it can be noticed from the PSDs of #023 in Case 2 that there is a clearer gap of the peak frequency between  $D_{\text{ref}}$  and  $D_{\text{test}}$ , compared to Case 1, which leads to strong anomalies around its natural frequency in the CASP. This difference is due to the hysteretic characteristics, that is, while in modified Clough's hysteresis yielding leads to the degradation of stiffness, in bilinear hysteresis, the unloading stiffness remains the same as the initial stiffness even after yielding and therefore the influence of yielding on the frequency content is not clear.

The higher anomaly scores of SDOF#006 can also be explained by frequency-domain characteristics. Compared to the others, the CASP of #006 in both cases has higher values in the frequency band lower than 1 [Hz]. In the same band, the response PSD of this SDOF in  $D_{\text{test}}$  has greater components than  $D_{\text{ref}}$ . The input PSD also clearly shows the change from  $D_{\text{ref}}$  to  $D_{\text{test}}$  in the corresponding band, which is considered to be the source of the higher anomaly scores of #006 in both cases. It can be stated from the above that even undamaged structures could be misdetected due to a change in the frequency content of excitations.

In summary, the following implications are obtained: (i) the damage of a structure with a hysteresis similar to an ordinary bilinear model can be overlooked by ECAD, because its ductility does not have strong influence on the frequency-domain characteristics of responses; (ii) the damage of a structure with a hysteresis that has degrading stiffnesses, like modified Clough's model, is much easier to detect, because its ductility has a clearer effect on the frequency content; (iii) the misdetection of undamaged structures can be induced by a change in the characteristics of input ground motions. Specifically, the above (i) and (ii) further imply that ECAD has a better performance for a structure with degrading stiffness, like reinforced concrete structures, than for a structure with stiffness that does not radically change even after yielding, like steel structures.

## 5. Conclusions

The paper has presented a data-driven approach for structure-portfolio performance monitoring, which adapts a machine learning method called Extended Correlation Anomaly Detection (ECAD). We have specifically

examined the effects of the nonlinearity of structural responses on the detection performance of ECAD, which had not thoroughly been studied in our previous works, by using numerical experiments on a portfolio of SDOF models. Two cases of hysteresis, ordinary bilinear models and modified Clough's models, have been considered. Main conclusions are:

- (1) In the case of ordinary bilinear models, some SDOFs does not show higher anomaly scores in spite of their high ductility factors. This means the limitation of application to such a structure that has a hysteresis similar to an ordinary bilinear model.
- (2) In the case of modified Clough's models, all of the SDOFs that experienced nonlinear behavior had higher anomaly scores than the SDOFs that behaved linearly. This implies the applicability of ECAD to such structures that have a hysteresis similar to modified Clough's model.
- (3) Although a SDOF has experienced no yielding, correlation anomalies can be misdetected due to the effect of input ground motions on the frequency-domain characteristics of responses.

The above (1) and (2) imply that ECAD is more applicable to a structure with degrading stiffness, like reinforced concrete structures, than one with stiffness that does not radically change even after yielding, like steel structures. This implication should be examined in the future works by additional numerical experiments that adopt other (more refined) hysteresis models, e.g., Takeda model, and also by physical experiments.

## Acknowledgement

The authors would like to thank National Research Institute for Earth Science and Disaster Resilience (NIED) for providing strong-motion observations, K-NET.

## References

- Friedman, J., Hastie, T. and Tibshirani, R. 2008. Sparse inverse covariance estimation with the graphical lasso. *Biostatistics*, 9(3):432–441.
- Ide, T., Lozano, A.C., Abe, N. and Liu, Y. 2009. Proximity-based anomaly detection using sparse structure learning. In *Proceedings of the 2009 SIAM international conference on data mining*, 2009. SIAM, pp. 97–108.
- Kullback, S. and Leibler, R.A. 1951. On information and sufficiency. *The annals of mathematical statistics*, 22(1):79–86.
- National Research Institute for Earth Science and Disaster Resilience. 2019. NIED K-NET, KiK-net, National Research Institute for Earth Science and Disaster Resilience, doi:10.17598/NIED.0004
- Yaoyama, T., Hida, T. and Takada, T. 2019. Correlation anomaly detection based on Fourier amplitudes of observed records for performance monitoring for large portfolio of structures. In *Proceedings of the Ninth Japan Conference on Structural Safety and Reliability*, Tokyo, 2019. JCOSAR, pp. 300–307 (in Japanese).

Yaoyama, T., Hida, T. and Takada, T. 2020. Performance monitoring on portfolio of structures in large-scale earthquake disasters. *Journal of Structural and Construction Engineering (Transactions of AIJ)*, 85(767):39–49 (in Japanese).