SECOND EUROPEAN CONFERENCE ON EARTHQUAKE ENGINEERING AND SEISMOLOGY, ISTANBUL AUG. 25-29, 2014

A NAÏVE BAYESIAN APPROACH TO STRONG GROUND MOTION SELECTION FOR NON-LINEAR SEISMIC ANALYSIS OF STRUCTURES

Maria LANCIERI[1], Marine RENAULT[2] Louis LECOMTE[3], Matthieu PERRAULT[4], David BAUMONT[5], Philippe GUEGUEN[6], Catherine BERGE-THIERRY [7]

ABSTRACT

Selecting and scaling strong ground motion for seismic performance assessment using nonlinear response-history analysis is still matter of debate in the earthquake engineering community. The common practice of using time series matching the design spectrum could be not suitable when the scope of the performance analysis is to assess the probability of shear failure. In that case the variability of the strong-ground motion should be taken into account. Recent works on conditional spectrum give clear responses on how to select spectrum-compatible data with conditional variability (Jayaram et al., 2011). However such techniques could lead to select hundreds of signals (Renault and Kurman, 2013).
In this paper we investigate the possible role of ground motion intensity measures (IM) in constraining data selection. To this aim, we propose a naïve Bayesian inference scheme to predict the response of a single degree of freedom (SDOF) system in function of a set of IM. The SDOF response is reduced to a description of three possible statuses: elastic, if the induced drift is lower than the yield displacement, plastic if the drift ranges between the yield and the ultimate drift values, and fragile if the drift reaches the ultimate drift. Our goal is to evaluate the conditional probability of observing a given status of the SDOF system in function of the IM array. The results of the Bayes classification on the training dataset are promising. Indeed, to validate the presented methodology and evaluate its prediction capability, we performed a blind test on a second dataset, completely disjointed from the training one, composed of 7000 waveforms recorded in Japan. To better understand what are the seismic features related with the plastic and fragile behavior, we performed a time-frequency analysis using the Stockwell transform. This latter part of the work is devoted to interpret the elasto-plastic-fragile system behavior making a link between the engineering and seismological features.

INTRODUCTION

The intensity measures (IM) are parameters describing some features of the seismic ground shaking, such as peak ground acceleration (PGA) or spectral acceleration (SA), which can be related with its effects on the structure. They represent the link between the seismic hazard assessment provided by

[1] PhD, IRSN, Fontenay aux Roses (France), maria.lancieri@irsn.fr
[2] Jr engineering, IRSN, Fontenay aux Roses (France),
[3] Jr engineering, IRSN, Fontenay aux Roses (France),
[4] PhD, UJF, Grenoble (France), matthieu.perrault@ujf-grenoble.fr
[5] PhD, IRSN, Fontenay aux Roses (France), david.baumont@irsn.fr
[6] PhD, UJF, Grenoble (France), philippe.gueguen@ujf-grenoble.fr
[7] PhD, CEA, Saclay (France), catherine.berge-thierry@cea.fr

seismologists and structural analysis conducted by engineers. How to select an appropriate IM (or set of IM), is today still matter of debate; an appropriate IM must be "efficient" and "sufficient" (Giovenale et al. 2004, Luco and Cornell, 2007) and "predictable" trough a seismic hazard analysis. An IM is efficient if it exhibits a small variability with respect to demand measure (DM) and it is sufficient if it is conditionally independent of the event magnitude M and distance R. It is important to notice that those conditions not only depend on ground motion nature but also on the characteristics of the analyzed structure. If the IM is even the demand measure or a specified limit state can be expressed as follow:

$$\lambda_{DM}(s) \cong \sum_{all\ s} P(DM > s_i | IM = y_j) \Delta\lambda_{IM}(y_j) \quad (1)$$

Where $P(DM > s_i | IM = y_i)$ express the probability of exceeding a given DM level s conditioned to a IM value y_i, and $\Delta\lambda_{IM}(y_i)$ is the annual frequency of IM=y_i.

It is interesting to observe that the best IM is the structural demand measure itself. Traditionally the most ground motion IM is the spectral acceleration Sa at a frequency very close to the fundamental period of the studied structure with a damping of 5%, or the peak ground acceleration (PGA). The correlations of DM of several typologies of structures to a given IM (or set of IMs) have been discussed by several authors. Among them Bommer and Martinez-Pereira (2000) focused their study on the importance of strong ground motion parameters as PGA, peak ground velocity (PGV) or duration. Lestuzzi et al. (2004) investigate the correlation between the responses of single degree of freedom, non-linear systems (hereinafter SDOF-s) and several IM (Arias intensity, Housner intensity and a custom spectral intensity based on the fundamental frequency of the SDOF models). Hancock and Bommer (2006) reviewed the influence of the ground motion duration on damage, concluding that it strongly depends on the metric used to quantify the damage and that the duration and the signal amplitude both affect the damage level, and the contribution of the duration cannot be easily decoupled. Luco et al. (2005) studied the sufficiency and the efficiency of a set of ground motion parameters with respect to the damage of elastic, ductile and fragile steel moment-resisting frame. In that case the sufficient IM vector includes the first mode spectral acceleration, the higher mode spectral acceleration and the first mode inelastic spectral acceleration.

In this paper, we will not only focus on the correlation between several IMs and the response of given systems, but we will move a step forward by proposing a technique to select strong ground motion based on the naïve Bayesian classifier. The basic idea is to evaluate for each waveform the conditional probability of driving the SDOF-s in a given state s_i, given an array of classical ground motion parameters ($\overline{IM}$). In the first part of the work we will exploit a training dataset composed by 6373 waveforms a non-linear SDOF. The correlation between each IM and the system demand will be used to define the probability density function of a given IM in function of the system limit state (s), that will be combined via a likelihood product. Since the degree of correlation between a given IM and the response of the non-linear SDOF-s strongly depends on the system characteristics, one of the advantages of the proposed algorithm is the combination of different IM describing the amplitude, frequency content and energy release. In the second part of the work the proposed scheme will be applied to a target dataset completely disjointed from the first one. Seven thousand waveforms recorded in Japan have been classified with respect the behavior of the defined SDOF-s, using the probability density functions issued from the training data-set.

THE NAÏVE BAYESIAN ALGORITHM

Suppose that the demand measure of a structural system can be described by a variable s that is categorical and depends on some other variable IM=$\{IM_1, \ldots, IM_N\}$. A variable is categorical if it has a discrete instantiation. In the presented application the DM can be reduced to the ultimate state (s) of the SDOF-s: elastic, plastic or fragile. The problem we are facing is to predict the conditional probability that a given structure be driven in the ultimate state s_i given a waveform characterized by an array of intensity measure $\{im_{1p}, \dots, im_{np}\}$ (conditional distribution of the state s given $\overline{IM}$). This is the typical classification problem that can be address using a naïve Bayesian classifier:

$$P(s|\overline{IM}) = \frac{P(s)P(\overline{IM}|s)}{P(\overline{IM})} \quad (2)$$

To implement a Bayesian procedure we need:

- To choose a probability density $P(s)$ called prior distribution, that express our belief on the parameters s before to see any data
- To determine a statistical model $P(\overline{IM}|s)$ that express our beliefs about data IM given the parameter s
- To update the beliefs after observing data $\{IM_1, \dots, IM_N\}$ and calculate the posterior distribution $P(s|\overline{IM})$.

We are interested in estimating the $P(s = s_i|IM_{1p}, \dots, IM_{Np})$ that can be expressed as:

$$P(s = s_i|\overline{IM} = im_{1p}, \dots, im_{Np}) = \frac{P(s=s_i)P(im_{1p},\dots,im_{Np}|s=s_i)}{\sum_j P(s=s_j)P(im_{1p},\dots,im_{np}|s=s_j)} \quad (3)$$

where the summation over j covers the whole event space, and s_j is a partition of the event space. Assuming that the IM variables are conditionally independent on the state s the conditional probability $P(im_{1p}, \dots, im_{np}|s = s_i)$ can be expressed as the likelihood product:

$$P(IM_{1p}, \dots, IM_{Np}|s = s_i) = \prod_{j=1}^{N} P(IM_{jp} = im_p|s_i) \quad (4)$$

where $P(IM_{jp} = im_p|s_i)$ is the conditional probability of observing the im_p value of the j-th IM on the state s_i.

The prior distribution P(s) is assumed to be uniform and equal to one:

$$P(s) = 1 \quad (5)$$

The probability $P(IM_{jp} = im_p|s_i)$ can be estimated from the relative frequency of data, measured on the training dataset :

$$P(IM_{jp} = im_p|s = s_i) = \frac{\#(IM_{jp}=im_p \wedge s=s_i)}{\#(IM_{1p}=im_p)} \quad (6)$$

where $\wedge$ is the logic "and" operator. Equation (6) express the ratio between the simultaneous observation of the limit state s_i and the p-*th* value of the j-th considered IM, and the number of observation of the p-th value of the j-th considered IM. A consequence of the formulation in eq. 6 is that the sum of the $P(IM_{jp} = im_p|s_i)$ is equal to 1:

$$\sum_s P(IM_{jp} = im_p|s_i) = 1 \quad (7)$$

This propriety is "transferred" to the $P(s|\overline{IM}) = \frac{P(s)P(\overline{IM}|s)}{P(\overline{IM})}$ leading to define complementary probability level for each state, conditioned to the observation $\overline{IM}$.

THE RELATIVE FREQUENCY CURVES

The relative frequency of data described in equation (6) has been obtained by injecting a training dataset of accelerometric waveforms in a non-linear SDOF-s. The SDOF system is modeled using the Takeda et al. (1970) hysteretic model. This model is generally used to simulate the reinforced concrete behavior since it includes some realistic conditions to describe the reloading curves and it takes into account the stiffness degradation caused by the excursions in the plastic domain.

The initial fundamental frequency of the SODF-s is equal to 2 Hz, the strength reduction factor (R) is equal to 3 and the yield drift is of 1.85 cm. The used DM is the SDOF-s maximum drift.

We used as strong ground motion inputs 6373 accelerometric records, all with peak ground acceleration (PGA) > 0.025g recorded worldwide (Europe, New Zealand, Taiwan and US West Coast). Each waveform is described by an array of classical intensity measures (IM) listed in table 1.

Table 1. Intensity Measures

Abbreviation	Parameter description	Equation
PGA	Pic ground acceleration	$max\|a(t)\|$
PGV	Pic ground velocity	$max\|\dot{a}(t)\|$
PGD	Pic ground displacement	$max\| \ddot{a}(t)\|$
Arias	Arias Intensity	$\frac{\pi}{2g}\int a^2(t)\,dt$
Housner	Housner Intensity	$\int_{0.1s}^{2.5s} PSV(\tau)d\tau$
SpInt	Spectral Intensity	$\int_{0.1s}^{0.5s} PSA(\tau)d\tau$
SCAV	Standardized cumulative absolute velocity	$\int_{t_i}^{t_i+1} \|a(t)dt\| \quad \forall t_i \; where \; a(t_i) \geq 0.025\, g$
PSA 1Hz	-	Pseudo-acceleration spectra estimated at 1Hz -
PSA 2Hz	-	Pseudo-acceleration spectra estimated at 2Hz -
PSA 5Hz	-	Pseudo-acceleration spectra estimated at 5Hz -

Table 1. Investigated IM, definitions and equations: a(t) is the acceleration, PSV the pseudo-velocity spectrum, PSA the pseudo-acceleration spectrum and g the standard gravity.

In Figure 1 we plot the distribution for each IM in function of the final state of the SDOF system for the studied system (right column), and the relative frequency curves (left column). An IM can be considered a good proxy when:

- the elastic (gray) and fragile (blue) curves keep a monotonic behaviour, while the plastic state frequency curve (red) has a bell shape (increasing – stable – decreasing probability);
- the maximum probabilities over each state are well separated in terms of IM values.

For each IM we identify two “crossing ranges”: the range of values where the fragile state becomes more frequent (and probable) than the plastic one, and the range of values where the fragile state becomes more frequent of the plastic one. The width of the “crossing range” depends on the scatter of the IM respect the final drift, a further indicator of scattering is the overlapping of the distributions on the right column of Figure 1.

The PSA is the best IM, indeed it coincides with the SDOF-s response until the yield drift. Also the PGV and the Housner, Arias and spectral intensities can be considered good IMs. The PGA and the SCAV are IM with a large overlapping of the state: a large range of values has the same level of probabilities to be observed over the three states, but outside this range the IM is a good discriminator. On the contrary the PGD is an example of bad IM (for this specific SDOF-s) since from $\log_{10}$(PGD) > 0.5 we are not able to discriminate the three states.

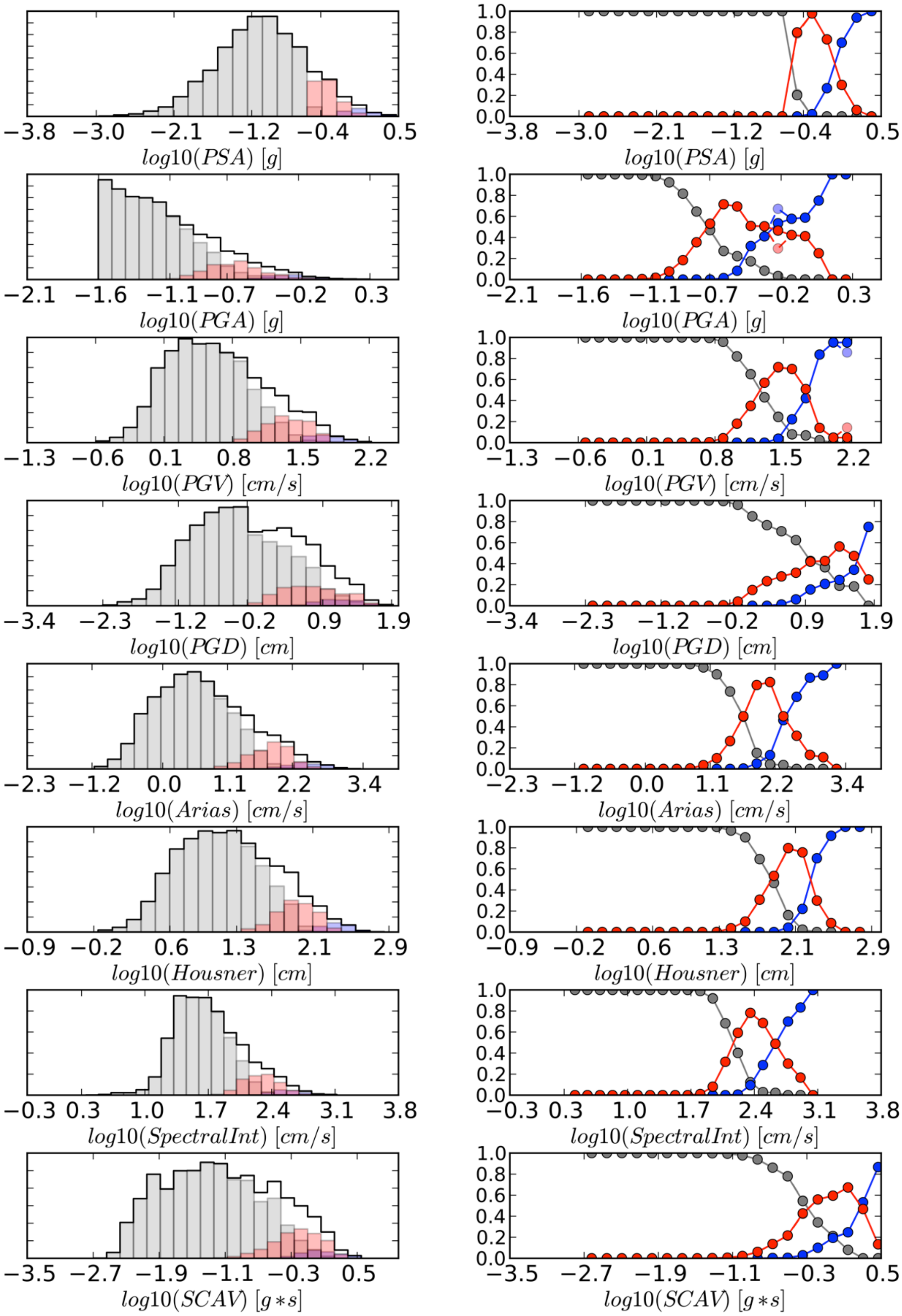

Figure 1. Left Colum: IM distributions in function of the final state, the bold black line is the IM distribution, gray bins distribution for the elastic state, red bins distribution for the plastic state, blue distribution for the plastic state. Right Colum: probability curves for the IM conditioned to the final state (in grey the elastic, in red the plastic, in blue the fragile state) issued from the IM relative frequency. In some cases (i.e. PGA and PGV curves) we slightly adjust the trends to keep a monotonic behavior compatible with the physic of the process.

THE CLASSIFIER PERFORMANCE

In figure 2 we show the naïve Bayesian classifier (NBC) performances on the training dataset. On the left column we compare the distribution of each IM for the three final states with that obtained via the NBC. The matrix on the right column gives a glance on the classification errors; it features on the x-axis the observed status and on the y-axis the predicted one, the elements distributed along the diagonal are the correctly labelled data, the elements outside the diagonal are misclassified data. Data lying over the diagonal are classified within a final state that is "more critical" than the observed one, whereas data placed below the diagonal are classified in a "lighter" state. We remark that data driving the system in linear state are the best labelled (les that 0.2% of data classified as "plastic"). The 26% of plastic-data are misclassified, mostly as elastic. The algorithm fails in distinguishing between plastic and fragile data. This is related to different factors: the used elasto-plastic model does not vary the system resistance when the critical drift (plasticity limit) is reached; the plastic and fragile distribution are not well sampled because the lack of observations; the accelerometric input has a complex signature that drives the system to fragile state. This latter aspect will be discussed later on the paper.

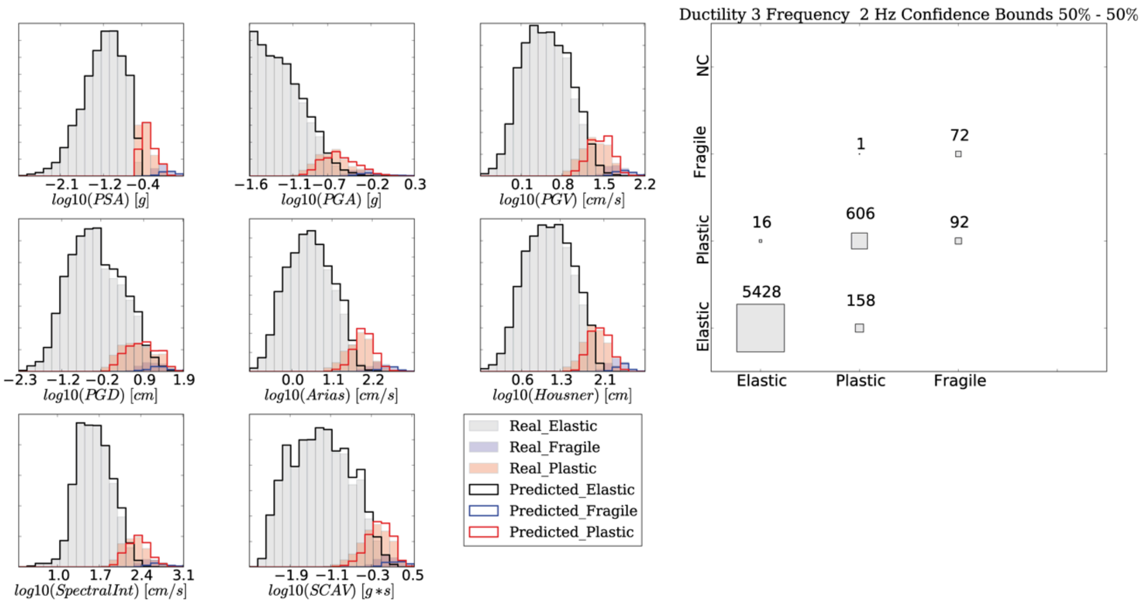

Figure 2 Performance of the naïve Bayesian classifier on the training dataset. On the left the observed distributions (filled bin the colours scale is described on the figure), and the distributions retrieved using the classification method. On the right column the "classification matrix" displaying on the X the observed state and on the Y the predicted ones.

The Bayesian classifier scheme has been applied to a second dataset completely disjointed from the first one. Indeed, this is the kind of application that will give insights on the efficiency of the proposed classifier. The idea is to use the probability distributions in Figure 1 to classify the systems responses, when solicited with the new data, in the final three states. The results will be then compared with those measured directly by injecting the new accelerograms in the SDOF system. The new database is composed by 7000 waveforms recorded at K-net Japanese network from 1994 to 2013 with PGA > 0.1 g. In Figure 3 we show the classifier results on the Japanese dataset. Given the blind character of the application the results are encouraging: the 99% of elastic-input are correctly labeled, as the 80% of plastic and fragile data. The classifier has very good performances on the fragile-plastic state discrimination, even better than those observed on training dataset. This is probably due to a better sampling of those states in this dataset.

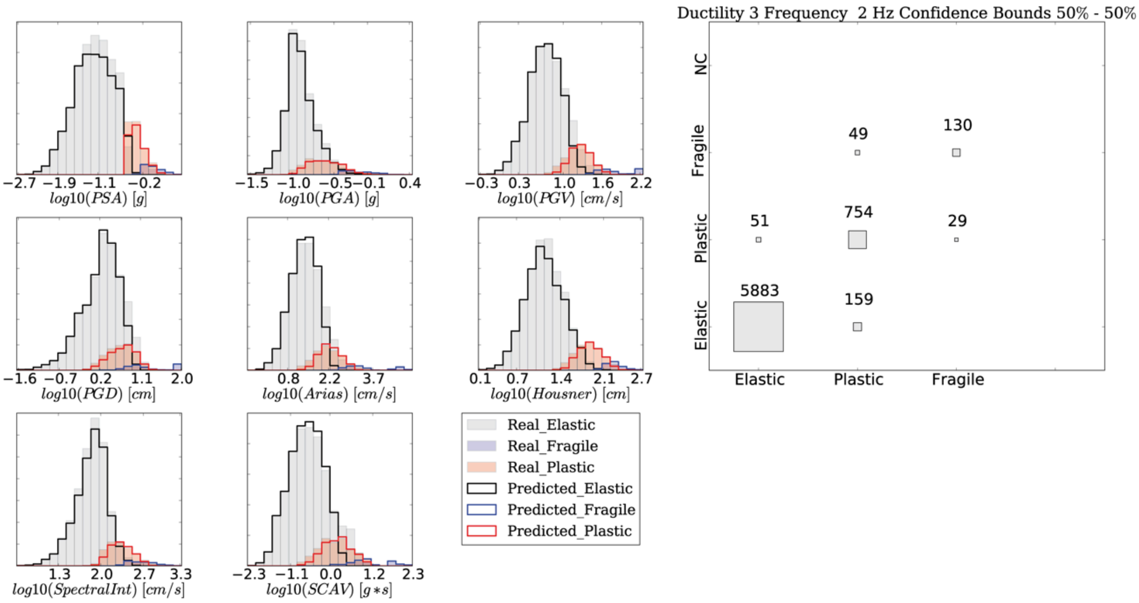

Figure 3 Performance of the naïve Bayesian classifier on the Japanese dataset. On the left the observed distributions (filled bin the colours scale is described on the figure), and the distributions retrieved using the classification method. On the right column the "classification matrix" displaying on the X the observed state and on the Y the predicted ones.

TIME-FREQUENCY ANALYSIS

The main advantage of the proposed classification method is the jointly use of several IM both related to pick, frequency and energy content. The importance of looking at the three families of indicator, instead to focus on a single one, is supported by the results of the Stockwell transform (Stockwell et al. 1996) analysis done on signals driving the system in fragile state. In Figure 4 we display the seismic input injected in the SDOF-s, the system response and the time/frequency representation of the two signals, the blue and the red bars indicate the time instants where the system enters in plastic state and reaches the fragile state.

The pulse inducing the plasticization of the system does not have an important amplitude, but the energy is mainly concentrated around 2 Hz (the system F_0). We can see on the fourth panel that it induces a resonance effect on the structure; the signals is then characterized by a sequence of spots of energy released around the initial fundamental frequency; finally the system reaches the critical drift. In this example the fragile state can be mainly explained looking at frequency and integral IM. In Figure 5 we illustrate a different case where the fragile state is directly induced by a big energy spot, centered on the system F_0 directly driving the system in the fragile state. In that case the system behavior can be mainly explained looking at frequency and pick quantities.

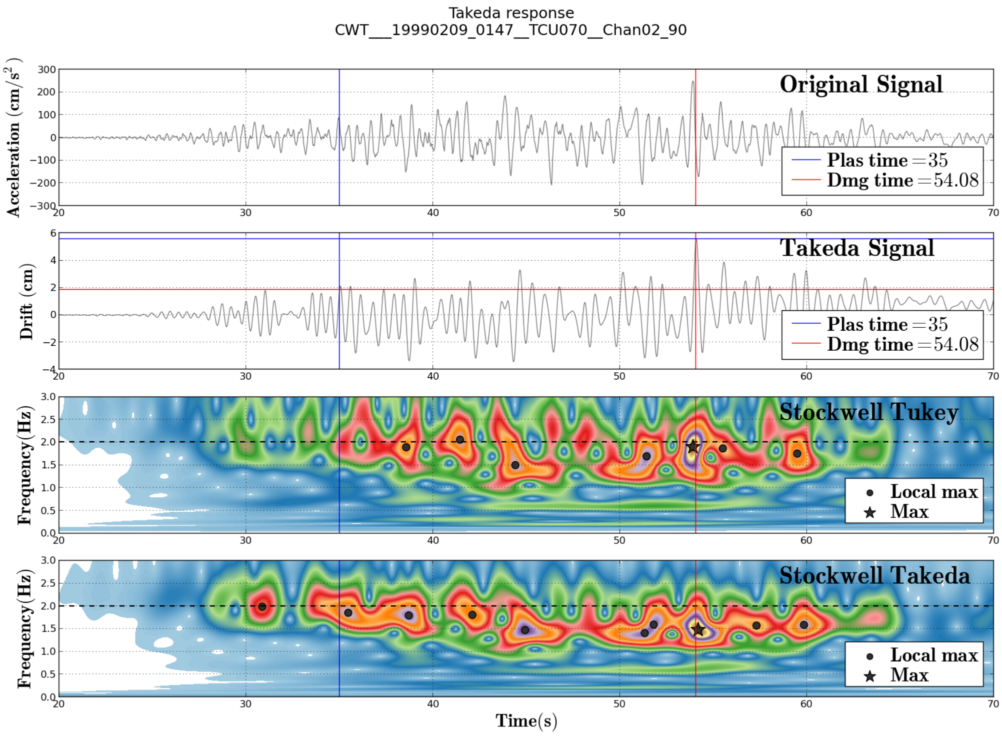

Figure 4 From the top: acceleration time history inputted in the SDOF-s, drift measured on the top of the SDOF system, Stockwell transform of the original signal and Stockwell transform of the drift. The vertical lines indicate the time at which the system enters in plastic (blue) and in fragile (red) domain. The dots on the time/frequency plots are the relative maxima of amplitude. The horizontal lines on the drift refer to the yield and ultimate drifts.

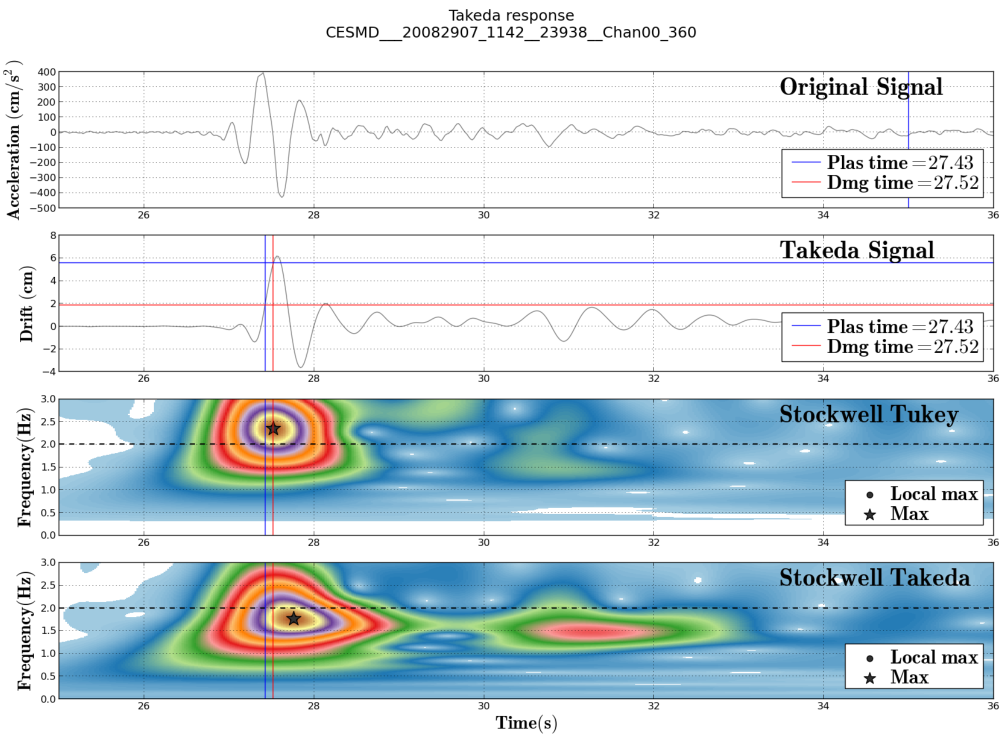

Figure 5: From the top: acceleration time history inputted in the SDOF-s, drift measured on the top of the SDOF system, Stockwell transform of the original signal and Stockwell transform of the drift. The vertical lines indicate the time at which the system enters in plastic (blue) and in fragile (red) domain. The dots on the

time/frequency plots are the relative maxima of amplitude. The horizontal lines on the drift refer to the yield and ultimate drifts.

CONCLUSIONS

In this work we introduced a naïve Bayesian algorithm to select seismic inputs for non-linear response history analysis. The basic idea is to develop a tool to classify waveforms in function of the final state (elastic, plastic or fragile) induced on a given structure, and select only the suited inputs. We implemented a classifier based on a single-degree-of-freedom system governed by the Takeda hysteretic model. The probability curves issued by the training phase have been used for a blind classification of a new dataset of seismic inputs, completely disjointed from the training dataset. The algorithm performance is very satisfying. The strength of the Bayesian approach is the possibility to mix information coming from different IM, leading to a more complete description of the waveforms driving the system into plastic or fragile states. To determine how exportable are the presented results, a possible perspective of this work is to test the classification method for complex structures with characteristics (frequency, stiffness, ductility) similar to the simple oscillator. Indeed, it is important to remark that the proposed Bayesian scheme does not depend on the particular structural model neither on the selected IM array, so it can be customized following specific needs. Of course more complex structural models will demand massive computational effort in the training phase.

REFERENCES

Jayaram N., Lin T., and Baker J. W. (2011) "A computationally efficient ground-motion selection algorithm for matching a target response spectrum mean and variance." *Earthquake Spectra,* 27(3), 797-815

Renault, P. and Kurmann D. (2013) "Comparison of uniform hazard spectra and conditional spectra approach in the Framework Of Fragility Curve Development" *Transactions SMiRT-22* San Francisco, California USA 18-23 August.

Giovenale P., Cornell A. and Esteva L. (2004) "Comparing the adequacy of alternative ground motion intensity measures for the estimation of structural responses" *Earthquake Engng Struct. Dyn.,* 3, 951–979 *(DOI: 10.1002/eqe.386)*

Luco N and Cornell A. (2007) "Structure-Specific Scalar intensity Measures for Near-Source and Ordinary Earthquake Ground Motions" *Earthquake Spectra* 23(2), 357-392

Bommer and Martinez-Pereira (2000) "The Effective Duration of Earthquake Strong Motion", *Journal Of Earthquake Engineering*, 3, 127-172

Lestuzzi P., Schawab P., Koller M. and Lacave C. (2004) "How to choose Earthquake Recordings for Non-linear Seismic Analysis of Structures", *Proceedings of 13th World Conference on Earthquake Engineering*

Hancock and Bommer (2006) "A state of knowledge review of the influence of strong-motion duration on structural damage" *Earthquake Spectra*, 22, 827-845

Luco N., Manuel L., Baldava S. and Bazzurro P. (2005) "Correlation of Damage of Steel Moment Resisting Frames to a Vector-Valued Set of Ground Motion Parameters" *Proceedings of 9th International Conference on Structural Safety and Reliability*, Rome, Italy

Takeda T., Sozen M. A., and Nielsen N. N. (1970). ''Reinforced concrete response to simulated earthquakes.'' *J. Struct. Div., ASCE,* 96(12), 2557 – 2573.

Stockwell R, Mansinha L and Lowe RP (1996) "Localization of the complex spectrum: the S transform Signal" Processing, *IEEE Transactions on,* 44 (4)