The efficacy of support vector machines (SVM) in robust determination of earthquake early warning magnitudes in central Japan

Ramakrushna Reddy and Rajesh R Nair*

Department of Ocean Engineering, Indian Institute of Technology, Madras, Chennai 600 036, India.
*Corresponding author. e-mail: rajeshnair@iitm.ac.in

This work deals with a methodology applied to seismic early warning systems which are designed to provide real-time estimation of the magnitude of an event. We will reappraise the work of Simons *et al.* (2006), who on the basis of wavelet approach predicted a magnitude error of ±1. We will verify and improve upon the methodology of Simons *et al.* (2006) by applying an SVM statistical learning machine on the time-scale wavelet decomposition methods. We used the data of 108 events in central Japan with magnitude ranging from 3 to 7.4 recorded at KiK-net network stations, for a source–receiver distance of up to 150 km during the period 1998–2011. We applied a wavelet transform on the seismogram data and calculating scale-dependent threshold wavelet coefficients. These coefficients were then classified into low magnitude and high magnitude events by constructing a maximum margin hyperplane between the two classes, which forms the essence of SVMs. Further, the classified events from both the classes were picked up and linear regressions were plotted to determine the relationship between wavelet coefficient magnitude and earthquake magnitude, which in turn helped us to estimate the earthquake magnitude of an event given its threshold wavelet coefficient. At wavelet scale number 7, we predicted the earthquake magnitude of an event within 2.7 seconds. This means that a magnitude determination is available within 2.7 s after the initial onset of the P-wave. These results shed light on the application of SVM as a way to choose the optimal regression function to estimate the magnitude from a few seconds of an incoming seismogram. This would improve the approaches from Simons *et al.* (2006) which use an average of the two regression functions to estimate the magnitude.

1. Introduction

Rapid earthquake information systems provide opportunities for short-term damage and risk mitigation. For over a decade, seismic networks have been able to provide location, magnitude, and ground shaking information within the first few seconds of the onset of the P-wave (Allen and Kanamori 2003; Olivieri and Clinton 2012). However, today, many earthquake-prone regions are demanding rapid earthquake information systems to their limits in an effort to provide information within the first few seconds of the onset of the P-wave, i.e., before the damaging S-wave arrives. These warning systems provide another opportunity to reduce the costs and casualties from future earthquakes.

The effectiveness of warning systems can be predicted by using P-wave rather than S-wave energy to assess the magnitude associated with an earthquake. The S-wave travels at a slower velocity than

the P-wave, which allows the exploitation of the S–P differential travel time for issuing an alert prior to damaging ground motion (Allen and Kanamori 2003). Nakamura (1988) was the first to propose the Zero Crossing method which uses the predominant period of the first few seconds of the P-wave to estimate the magnitude of an earthquake. Allen and Kanamori (2003) demonstrated the earthquake early warning methodology in California using the predominant period of the first few seconds of the P-wave. The methodology was developed to maximize the warning time in a region where the population is collocated with the earthquake source region. Lockman and Allen (2005) used single station earthquake characterization for early warning systems as individual stations are able to deliver an accurate early warning. They suggested that utilizing the technology in regions that lack a dense seismic network but are in need of seismic hazard mitigation, is possible. They concluded that small earthquake magnitudes radiate high-frequency energy, while larger earthquakes emit lower-frequency energy. Simons *et al.* (2006) developed an alternative approach to measure the predominant period using wavelet multiscale analysis. They applied their method on data from southern California and obtained a scaling relationship between wavelet coefficients and earthquake magnitude. In this paper, we propose an algorithm using wavelet multiscale analysis with a support vector learning classifier. If there are 10 stations, it is possible to make magnitude predictions with error less than ±0.4 magnitude units. We utilized

seismograms obtained from KiK-net, Japan (http://www.kik.bosai.go.jp/kik/). The National Research Institute for Earth Science and Disaster Prevention (NIED) and the Japanese Meteorological Agency (JMA) provided the locations and earthquake magnitude. We used the data of 108 events (depth 0–70 km) of magnitude ranging from 3 to 7.4 in central Japan and Hokkaido recorded at KiK-net network (figure 1). A minimum of three and a maximum of 40 recorded waveforms per event were used for the study, and all these waveforms per event recordings are within an epicentral distance of 150 km. Waveform recorded at the station at KiK-net network (figure 2) plots time (t = 120 sec) *vs.* amplitude (cm/s^2) for two different waveforms of magnitude 5.6 (plot a) and magnitude 4 (plot b). Plots (c) and (d) are the zoomed part of the waveforms (a) and (b).

We used data of a 5.5 magnitude earthquake to demonstrate and validate the regression relations for real-time application. After performing the time-scale measurement, we obtained the wavelet coefficient at scale 7 as C7 = 1567.987. By substituting this value in the given equation, we obtain the predicted magnitude.

Ml= 1.25∗log 10(C7)+1.8
Mh= 1.41∗log 10(C7)+1.7

Ml= 1.07∗log 10(C7)+1.9
Mh= 2.40∗log 10(C7)-2.2

Ml = 5.79; Mh = 6.20 (using Simons methodology)
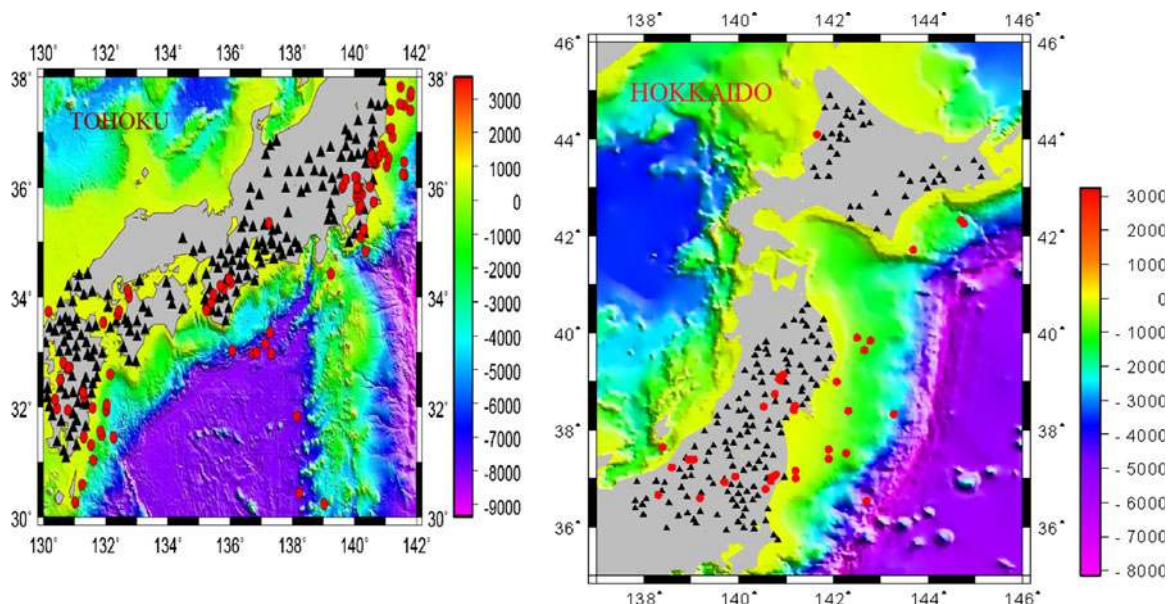Ml = 5.31; Mh = 5.46 (using SVM).

Figure 1. Bathymetry map of study area Tohoku and Hokkaido. Colours indicate seafloor depth and topography in meters. The continental region is shown in gray. The red circles represent the epicenter and the black triangles indicate the recording stations.
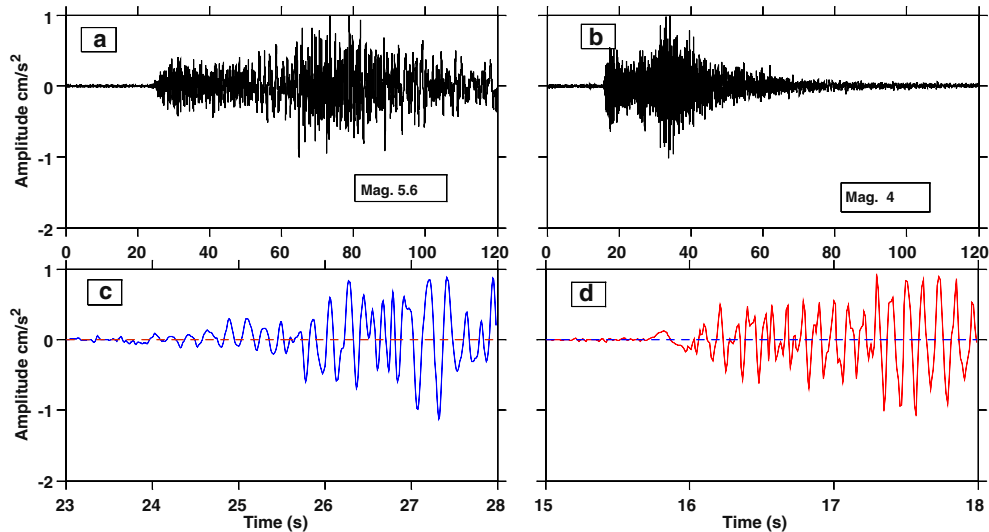
Figure 2. Waveform recorded at station at KiK-net network (**a**) and (**b**) plots time (t = 120 sec) (X-axis) *vs.* amplitude (cm/s^2) (Y-axis) for two different waveforms of magnitude 5.6 (plot a) and magnitude 4 (plot b). In plots (**a**) and (**b**), the onset of the P-wave is not clearly visible. In order to show the clear onset of P-wave, we zoomed plots (**a**) and (**b**). Plots (**c**) and (**d**), are the zoomed part of (**a**) and (**b**). In waveform (**a**), onset P-wave is at 25 sec, and the same shown in the blue plot (**c**); in waveform (**b**), the onset of P-wave is at 16 sec as clearly shown in the red plot (**d**).

2. Time–frequency analysis

For a non-stationary signal such as a seismogram, the frequency content varies with time. The amplitude spectrum of the Fourier transform (FT) indicates the presence of different frequencies but does not show their temporal distribution. In non-stationary signals, FT gives the frequency information but we do not know at what time these frequencies occur. In stationary signals, all frequencies that exist in the signal exit throughout the duration of the signal. If it is assumed that the signal through a small window of time is stationary/non-stationary, then its FT provides us with the frequency content of the signal in that time period, that is, FT can give only frequency information, it cannot give time information. By shifting this time window appropriately, the frequency content of the signal is extracted and the 2D representation of 'frequencies *vs.* time' called the Short-Time Fourier Transform (STFT) is produced. STFT will give both 'time and frequency' information but with poor resolution. To obtain better resolution, we shift from STFT to wavelet transform. An STFT can also be implemented by choosing frequency domain windows as opposed to the time domain windows (Okaya *et al.* 1992). Instead of sampling the time axis with moving windows, the frequency axis can be sampled by a set of fixed bandwidth bandpass filters whose centre frequencies are distributed uniformly along the frequency axis. During long duration, small changes in the time domain become obscured because of averaging. The opposite is true for a window function of short duration that defines short-lived variations in time but fails to detect slow frequency changes. This trade-off is called the Uncertainty Principle or Heisenberg Inequality (Claerbout 1976). Allen and Kanamori (2003) used the Zero Crossing method in order to obtain the predominant period of an incoming P-waveform. However, the STFT also has its disadvantages, such as the limit in its time–frequency resolution capability, which is due to the Uncertainty Principle. The Uncertainty Principle states that 'we cannot exactly know what frequency exists at what time instances'; we can only know what frequency bands exist at what time intervals. So the resolution does matter.

Time–frequency analysis of waveforms resampled at 50 Hz is shown in figure 3. The logarithmic amplitude of spectral density calculated over a 4 s sliding segment of the data using the Welch overlapping segment algorithm, overlaps by 50% and is windowed with Hamm taper (Welch 1967). The red line indicates the instantaneous frequency of the waveform. Due to the uncertainty arising out of poor resolution, we were unable to compute the exact predominant period for the first P break. (Using figure 3, we can estimate the predominant (instantaneous) frequency of the seismogram. The predominant period can be calculated from predominant frequency.) The reason for lower predictability of some of the previous works was due to the fact that though the signal does not exist and is completely noise free, it reaches non-zero predominant period values even before the onset of the P-wave (Allen and Kanamori 2003; Olson and Allen 2005; Wu *et al.* 2006).
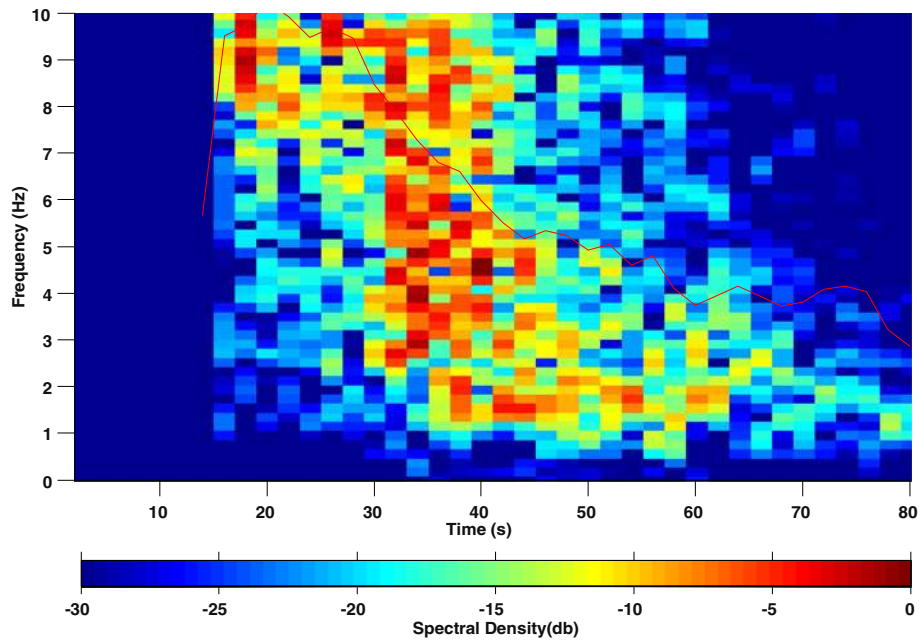
Figure 3. Time–frequency analysis of waveforms resampled at 50 Hz shown in figure 2(**b**). The logarithmic amplitude of spectral density calculated over a 4 s sliding segment of the data using the Welch overlapping segment algorithm (overlap by 50%) and windowed with Hamm taper is shown. The red line indicates the instantaneous frequency of the waveform.

3. Wavelet analysis

The shortcomings of time–frequency analysis and the difficulties in estimating the predominant frequency of a seismogram will be entirely eliminated by switching from the time–frequency domain to the time-scale (wavelets) domain. They have advantages over traditional time–frequency methods in analyzing physical situations where the signal contains discontinuities and sharp spikes. Wavelets are mathematical functions that separate data into different frequency components, and then study each component with a resolution matched to its scale. Scale is a mathematical operation that either dilates or compresses a signal. Larger scales correspond to dilated (or stretched out) signals and small scales correspond to compressed signals. Low scales (high frequencies) do not last for the entire duration of the signal. High scales (low frequencies) last for the entire duration of the signal.

In continuous wavelet transform (CWT), we have only time resolution, whereas in discrete wavelet transform (DWT), we have both time and frequency resolution. In this study, we used DWT. In a DWT, similar transformation into a 2D scale-translation space is implemented in a different manner. The DWT is implemented using quadrature mirror filter (QMF) banks (Vetterli and Cormac 1992; Vaidyanathan 1993; Simons *et al.* 2006). QMF is a set of two filters, a low-pass one and a high-pass one. During a forward transform, the original signal is filtered by a half band low-pass filter and a half band high-pass filter followed by a down-sampling by a factor of two. The output of the high-pass filter comprises the DWT coefficients for that stage (scale or level). The output of the low-pass filter is once again filtered using the two filters mentioned above. This is continued until the desired level of decomposition is achieved. Thus, at each stage of the transform, the low-pass output is examined in further detail using the high-pass filter. The QMFs are efficient, fast (NlogN operations required for an N-point DWT) and orthogonal. The wavelets used are also orthogonal, thereby ensuring a perfect reconstruction. The implementation of the DWT using QMFs and other multi-rate filter banks are discussed in detail in signal processing literature (Rioul and Vetterli 1991; Mallat and Zhang 1993; Vaidyanathan 1993; Daubechies and Sweldens 1998).

There are plenty of wavelet and scaling bases (Daubechies 1990; Mallat 1998) and computational algorithms (Strang and Nguyen 1997; Jensen 2001) available in literature. We used a biorthogonal construction with two and four vanishing moments for the primal and dual wavelets termed as Cohen–Daubechies–Feauveau (CDF(2, 4), (2, 4)) vanishing moments (Cohen *et al.* 1992). The different sets of wavelets make different trade-offs between how compactly they are localized in space and how smooth they are. The technique we used for this work is the lifting approach (Sweldens 1996). Our method of DWT is a fast, linear operation that works on the data vector. Given that the total
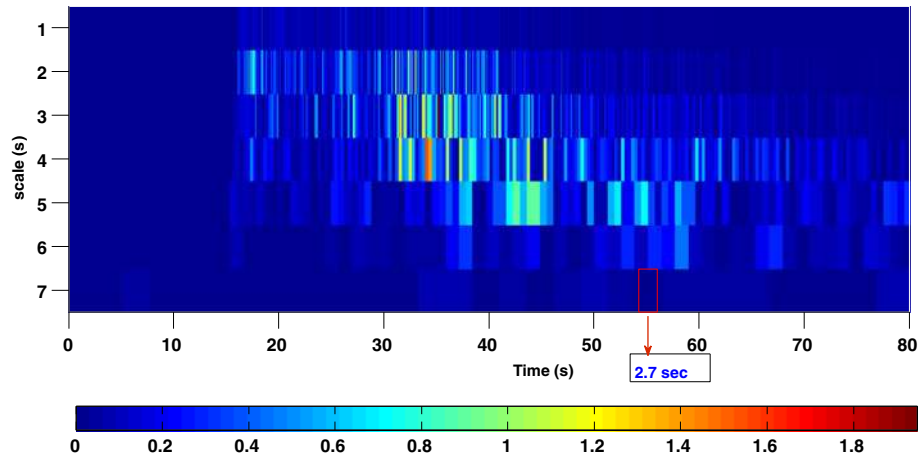
Figure 4. Time–scale analysis of waveform (resampled at 50 Hz) shown in figure 2(**b**), using CDF (2, 4) wavelet. The colour bar represents the wavelet coefficient magnitude. The 2.7 s time band can be observed at scale 7. The 2.7 sec represents the length of the dyadic grid (grid size) corresponding to scale 7. This means that a magnitude determination is available within 2.7 s after the initial onset of the P-wave. X-axis: time; Y-axis: scale 1 to 7.

number of data used is $D=2^N$ and the signal length is L, $D/2$ data were first computed at scale $L/2^{(N-1)}$, followed by $(D/2)/2$ data at scale $L/2^{(N-2)}$, and so on. This was continued till $D=2$ was obtained at scale $L/2$. The result of this algorithm is an array of the same length as the input, where the data are usually sorted from the largest scales to the smallest ones. The wavelet transform is invertible and in fact, orthogonal. The inverse transform, when viewed as a big matrix, is simply the transpose of the transform.

In order to predict the magnitude of an incoming earthquake, our foremost objective was to detect the first incoming P-wave immediately, though this was sometimes impossible owing to the noise in the signal. Even though seismograms contain a noise component, the P-wave arrival has to be detected rapidly. All the insignificant coefficients were assigned zero values by assigning a scale-dependent threshold value (Johnstone and Silverman 1997). In practice, the threshold T_j at scale j is defined in terms of a spread estimate σ_j and the number of coefficients at that scale N_j, as $T_j=\sigma_j(2\ln(N_j))^{1/2}\sigma$ represents the median absolute deviation from normally distributed coefficients. The first significant coefficient at each scale just above the threshold provides information about the P-wave, and the average of all the coefficients of the stations is used to derive the best-fit regression lines. The time-scale analysis of waveforms resampled at 50 Hz is explained by the spectrogram in figure 4. The dyadic nature of the time-scale tilting is clearly visible as blocks of constant value whose length increases with increasing scale or coarseness. In this approach, it is not necessary to assign extra memory to wavelet coefficients. This method allows integer to integer transformation. The algorithm

does not require a very high level computation to perform the transform and can do it on very low level circuiting (Simons *et al.* 2006).

We analyzed threshold wavelet coefficients from scale 1 to 7 as shown in figure 4. The magnitudes of the wavelet coefficients are shown. The dyadic nature of the time-scale tilting is clearly visible as blocks of constant value whose length increases with increasing scale or coarseness. The thickness of the block length at scale 7 shown in red colour is 2.7 s. Using this 2.7 s time window, we can predict the magnitude immediately after the initial onset of the P-wave. This means that there is no need to wait to record the whole waveform for magnitude determination and that it is possible to apply wavelet analysis to all earthquakes record at a time.

4. Support vector machines

Traditional Neural Network approaches have suffered difficulties with generalization, producing models which overfit the data as a consequence of the optimization algorithms used for data selection and the statistical measures used to select the best model. The difference between the training and the classified (test) set is that, in the training set, the data classes are known earlier. The classified set contains the data that needs to have their classes predicted. A classification task usually involves training and test (classified) sets which consist of data instances. Each instance in the training set contains one target value (class label) and several attributes (features). The goal of a classifier is to produce a model that is able to predict target values of data instances in the testing set, for which only the attributes are known.

SVM is used not only for classification, but for regression too.

SVMs were developed to solve the classification problem. However, recently, SVMs have been successfully extended to regression and density estimation problems (Refaat and Aly 2004). SVMs are gaining popularity due to their many attractive features and promising empirical performance. Once the threshold wavelet coefficients are found, the next task is to classify them into low magnitude and high magnitude events, which would enable the deduction of a suitable relationship between wavelet coefficient amplitude and earthquake magnitude. SVMs provide an ideal platform for the classification of these events as they are conceptually simple, have the capability for fast learning and often produce very accurate results. SVM is a learning machine developed from statistical theory and based on the risk minimization principle. A training data set D is given as:

$$D = \left\{ (x_i, y_i) | x_i \in R^P, \, y_i \in \{-1, 1\} \right\}_{i=1}^{n}$$

where y_i is either 1 or -1 indicating the class to which the input for the SVM is the threshold wavelet coefficients x_i belongs, in this case high and low magnitudes. Each x_i is a p-dimensional vector. A separating plane (hyperplane of dimension P) needs to be drawn for this dataset which divides the training patterns into classes, i.e., the data points with $y_i = 1$ have to be distinguished from the ones with $y_i = -1$. The hyperplane H is designed such that its distance from the nearest data points (support vectors) is the farthest. This is the so-called maximum margin principle. As shown in figure 5, the equation of the hyperplane H can be written as $w^T x + b = 0$ where w is 'weight'.

In a two-class problem, the features are weighed strongly for each of two classes, that is negative weights loading on class 1 and positive weights loading on class 2. It means that the weights are important in characterizing the class when a decision is made between class 1 and 2. If $||w||$ is chosen such that the distance between the positive and the negative hyperplanes H$^+$ and H$^-$ is at least $1/||w||$, to avoid data points from falling in between H$^+$ and H$^-$ the following condition should be met:

$$w^T x_i + b \geq +1 \quad \text{for } y_i = +1,$$
$$w^T x_i + b \leq -1 \quad \text{for } y_i = -1$$

and

$$y_i \left(w^T x_i + b \right) \geq +1$$

for all i.

To obtain the optimum hyperplane, $1/||w||$ must be maximised or $||w||$, which can be shown to be equal to the margin, must be minimized. As our study deals with only linearly separable data, this optimization problem is expressed in terms of Lagrange multipliers α_i as:

$$\min_{w,b} \max_{\alpha} \left\{ \frac{1}{2} \left\| w \right\|^2 - \sum_{i=1}^{n} \alpha_i \left[y_i \left(w^T x_i + b \right) - 1 \right] \right\}.$$

This problem is solved by standard quadratic programming techniques such as MATLAB subroutines. The solution can be expressed in terms of a linear combination of the training vectors as:
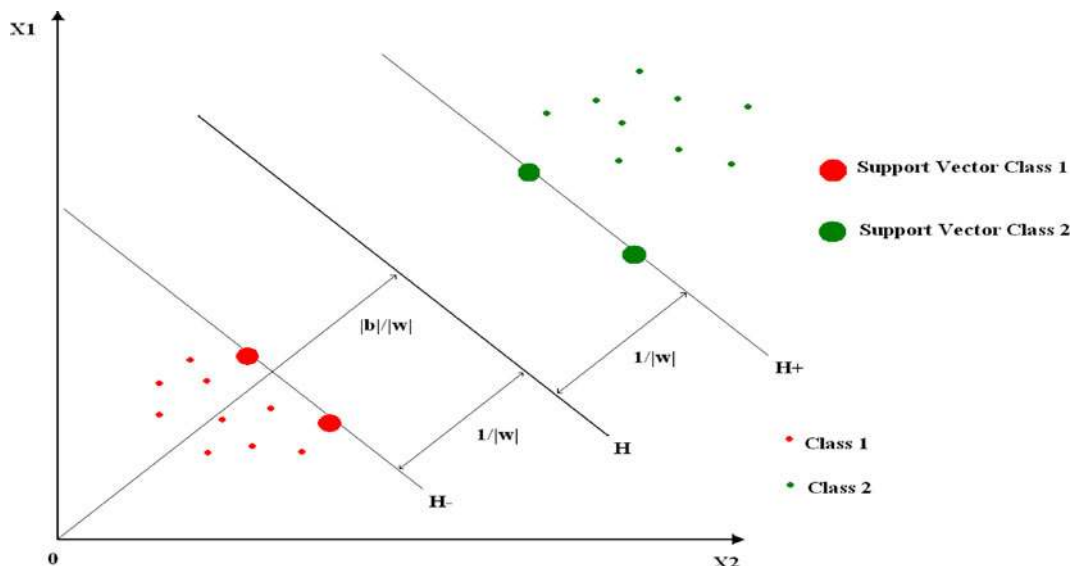
$$w = \sum_{i=1}^{n} \alpha_i x_i y_i.$$

Figure 5. Classification with SVM shown by an example with two datasets, support vectors and hyperplanes.

Only a few α_i will be greater than zero. The corresponding x_i are the support vectors which lie on the margin. Subsequently, one can also determine the offset b.

In order to train the algorithm, we performed 1500 iterations. At each iteration, a wavelet coefficient migrates from one set to another. Each migration passes through Support Set, so that at every iteration, the trained dataset changes. To minimize the misclassification error, we mapped the training data into the transformed feature space. Because this is a common and very expensive operation, another method to efficiently update the input dataset is explained in the following sections. In order to stabilize the wavelet coefficient, we trained all the data. After training and checking that the entire wavelet coefficients are in the right set, we treated them as classified coefficients. If this does not happen, each wrong sample is forgotten and trained again for an another iteration.

5. Results and discussion

The wavelet-based approach overcomes the limitations of the predominant frequency approach (high scatter and less predictive capability) used by Allen and Kanamori ([2003](#)). Instead of calculating the predominant period, we correlated scale-dependent threshold wavelet coefficients. As the wavelet-based magnitude estimation has already been done (Simons *et al.* [2006](#)) in our methodology, we demonstrate the advantage of combining the SVM approach with the wavelet-based approach for the prediction of an earthquake. These regression lines provide us with the predicted earthquake magnitude. Following this, we performed error analysis by comparing the JMA magnitude with the predicted earthquake magnitude.

The algorithm used for our methodology (results of earthquake magnitude prediction based on SVM with wavelet approach) and Simons approach (results of earthquake magnitude prediction based on DWT alone) is the forward Cohen–Daubechies–Feauveau (2, 4) biorthogonal wavelet transforms based on the fast lifting scheme. The data used for both these methods are 'wavelet coefficients'. These wavelet coefficients are obtained from 1689 seismograms.

The wavelet coefficient amplitudes obtained from the 1689 seismograms are extensively scattered. The wavelet coefficient amplitude is the key factor in predicting the magnitude of an incoming earthquake, and it is difficult to get the best regression line from the uniform regression. Therefore, to predict the accurate magnitudes, we split the entire dataset into two equal parts, namely, low (M_l) and high (M_h) magnitudes. Both M_l and M_h

produce the best estimate of magnitude. Initially, 2 s after a station triggers, M_l ($M < 5.2$) is calculated from wavelet coefficient (C_7). The estimate is updated when 2.7 s of data are available. Station-magnitude estimates are averaged to provide an event-magnitude estimate. If the event-magnitude estimate becomes greater than 5.2, then M_h is also calculated and the event-magnitude estimate is the average of both M_l and M_h from each triggered station.

$$M_l = 1.25 \log C_7 + 1.8$$
$$M_h = 1.41 \log C_7 + 1.7$$

$\Bigg\} \quad \boxed{\text{Before SVM (wavelets only)}}$

SVM, in particular, is a recently-developed, powerful, state-of-the-art technique for regression and classification. SVM seeks the best compromise between the complexities of the model and learning ability according to the limited data information available. Using an appropriate quadratic solving equation, a maximum margin hyperplane ($x - 5 = 0$) is constructed which separates the data (wavelet coefficients) with the two magnitude ranges in the best possible manner as shown in figure [6](#). This plot also shows data points as 'trained' and 'classified'.

The data is further grouped into 'classified' and 'trained' using the inherent property of the SVM. We considered only classified data. As shown in figure [7](#), significant correlations exist between the station-averaged 'classified' wavelet coefficients and the event magnitude (low and high range) at scale 7. This scale holds the most promise in making quick event magnitude determinations directly from the seismogram. We correlate the 'classified' coefficient amplitudes obtained at scale 4 to scale 7, with the best fit regression equation used for predicting the magnitude of new incoming P-wave in the studied region as shown in figure [8](#). These linear fits give the relationship between the wavelet coefficient amplitude and earthquake magnitude separately for both high as well as low magnitude events. Best fit regression lines are obtained from least-squares fitting techniques to give the relation between the average of the 'classified' threshold coefficients at a scale 4–7, C_4–C_7, and the local JMA earthquake magnitude.

$$M_l = 1.07 \log C_7 + 1.9$$
$$M_h = 2.40 \log C_7 - 2.2$$

$\Bigg\} \quad \boxed{\text{After SVM}}$

Earthquake early warning magnitude estimates made at various geological regions using different methods are presented in table [1](#). Next, we obtain the prediction error, the difference between the predicted magnitude and JMA magnitudes of an earthquake. Table [2](#) compares the errors obtained from our approach with the Simons *et al.*'s ([2006](#))
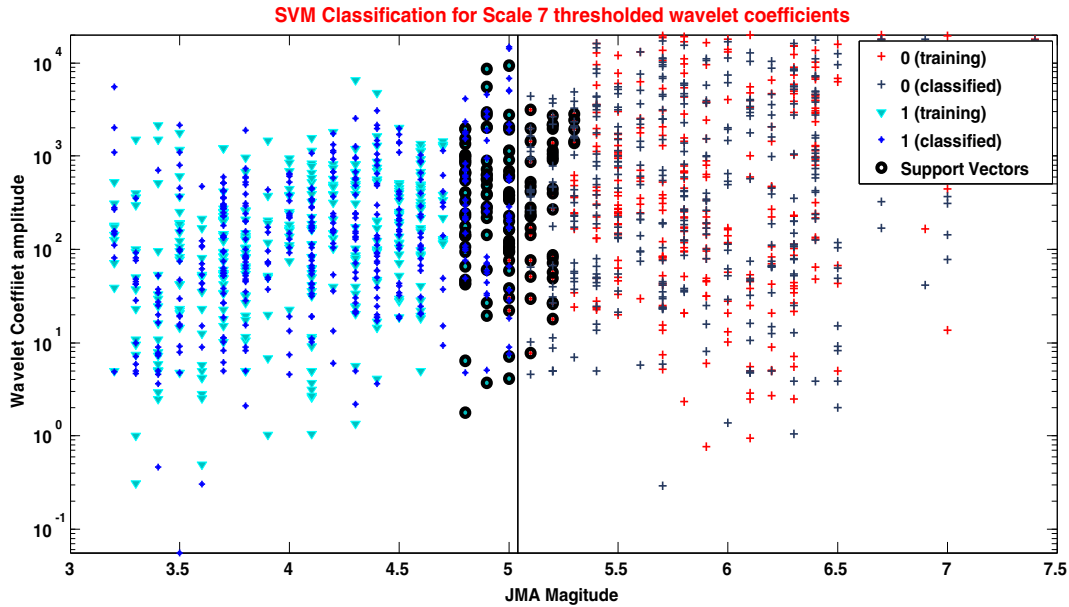
Figure 6. Plot showing the dependence of wavelet coefficient magnitude with the JMA magnitude along with the classification of the events into two classes separated by the maximum margin hyperplane (vertical line). Also shown are the vectors which have been classified and trained in the legend above. The circles denote the support vectors.
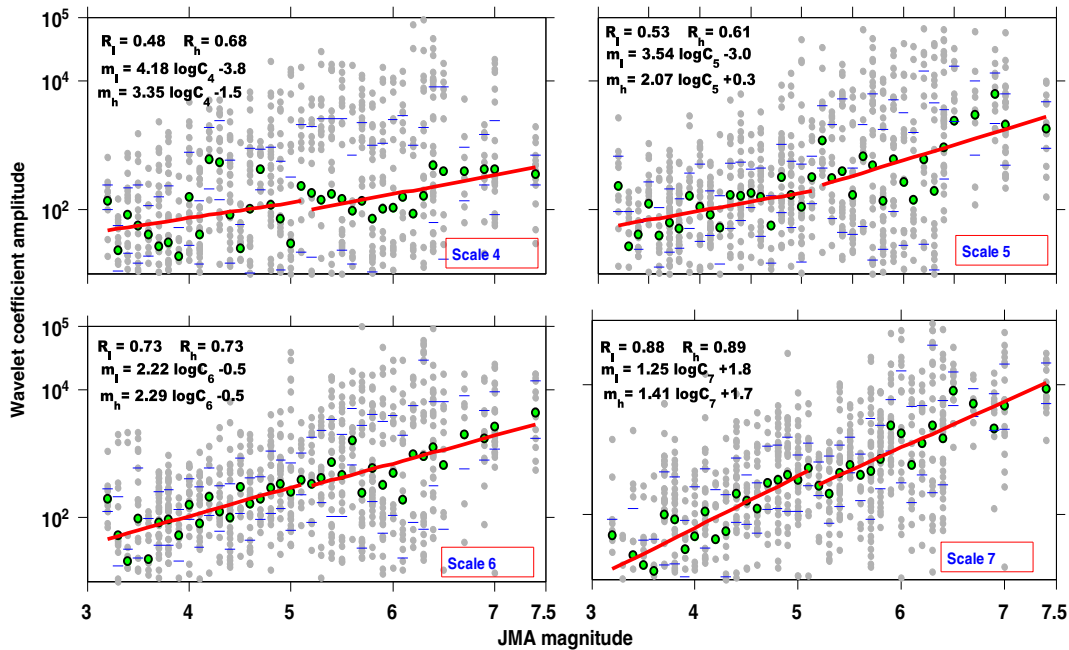

Figure 7. The correlation of threshold wavelet-scale dependent amplitudes with earthquake magnitude. Absolute magnitude of the first significant classified wavelet coefficients at scale 4 to 7 obtained from SVM analysis, on a logarithmic scale, automatically determined from 2.7 s of data symmetrically windowed around the P arrival time. Results for individual seismograms are shown as gray circles, the 25th and 75th percentile by thin blue dashes, and the averages at every distinct event magnitude by green dots. The superimposed thick red lines are least-squares best-fit lines to the average values over a low (M_l) and high (M_h) magnitude range, parameterized by the equations listed in the upper left. The correlation coefficients R_l and R_h are quoted only where they exceed the significance level. Scale number is listed at the lower right. Scales 1–6 have huge scattering; however, in scale 7 the scattering is very low. We therefore consider the scale 7 for the study.

approach that used the significant wavelet coefficients obtained from the first few seconds of the P-wave. The order of predicted earthquake magnitude errors obtained in SVM analysis are

comparatively less than the errors obtained from the wavelet-based earthquake early warning alone (Simons *et al.* 2006), (figure 9). We did further analysis in the Tohoku region and thus expanded
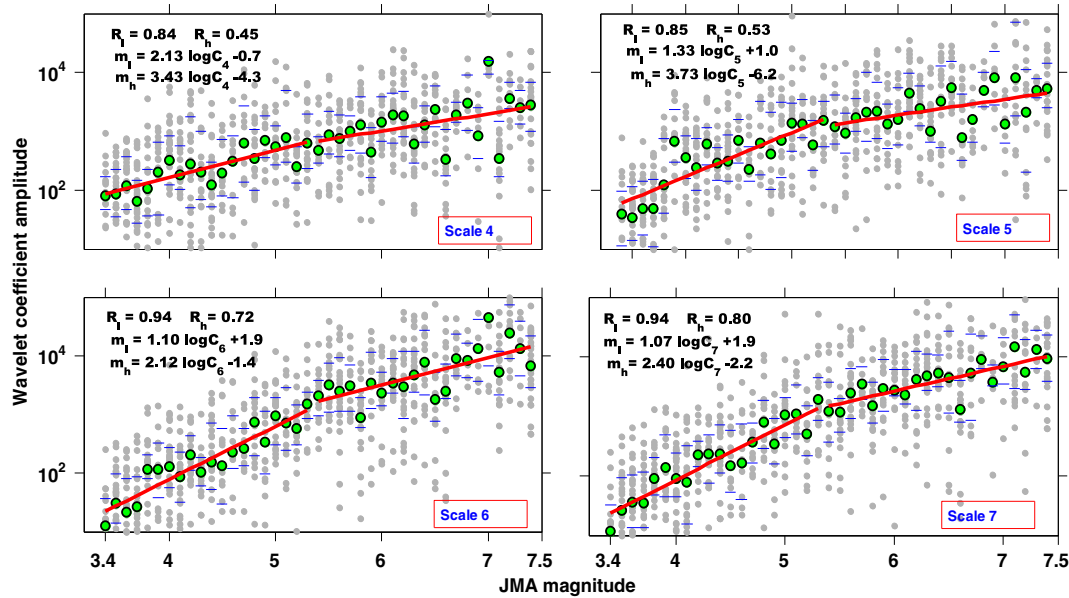
Figure 8. The correlation of threshold wavelet-scale dependent amplitudes with earthquake magnitude in Hokkaido region. Absolute magnitude of the first significant classified wavelet coefficient at scale 4 to 7 obtained from SVM analysis, on a logarithmic scale, automatically determined from 2.7 s of data symmetrically windowed onset of the P arrival time. This shows our assessment of algorithms in other geologic scenarios, especially in Tohoku.

Table 1. *Estimation of earthquake early magnitude at various regions using different methods (τ, P_d, $T\ldots$ Predominant period, A: max. amplitude of waveform, R: Epicentral distance).*

Investigator	Method	Region	Relation obtained	Magnitude prediction error
Allen and Kanamori (2003)	Zero-crossing (Z–C)	Southern California	$M_l = 6.3 \log \left(\tau_p^{\mathrm{max}}\right) + 7.1$ $M_h = 7.0 \log \left(\tau_p^{\mathrm{max}}\right) + 5.9$	±0.7
Wu and Kanamori (2005a, b)	(Z–C)	Taiwan	$\log \tau = 0.221 M - 1.113$ $M = 4.525 \log \tau - 5.036$	±0.7
Simons *et al.* (2006)	Wavelet multiscale (FFT)	Southern California	$M_l = 1.04 \log(C_5) + 0.5$ $M_h = 1.46 \log(C_5) - 1.2$	±1
Wurman *et al.* (2007)	Short time fourier transform (STFT)	Northern California	$M = 6.66 \log_{10} \left(\tau_p^{\mathrm{max}}\right) + 5.22$	±0.75
Tsang *et al.* (2007)	Z–C and peak ground displacement	Southern California	$M = 6 : 36 + 6 : 83 \log_{10} \left(\tau_p^{\mathrm{max}}\right)$ $M = 1 : 24 \log_{10} (\mathrm{P_d})$ $\quad + 1 : 65 \log_{10} (\mathrm{R}) + 5 : 07$	±0.65
Lockman and Allen (2007)	Z–C	Japan	$M_l = 6.1 \log \left(T_{\mathrm{max}}^p\right) + 6.7$ $M_h = 4.7 \log \left(T_{\mathrm{max}}^p\right) + 4.8$	±0.7
Olivieri *et al.* (2008)	Z–C	Italy	$M = 3 : 05 \log \left(\tau_p^{\mathrm{max}}\right) + 4.3$	±0.4
Lin and Wu (2010)	Peak ground acceleration	Taiwan	$M = 1.95 \log \mathrm{A} + 0.006 \mathrm{P} - 1.619$	±0.55

Table 2. *Magnitude prediction error, before and after SVM.*

Range of |error magnitude|	Percentage of seismograms
Before SVM (wavelets only)	
0–0.4	33.33%
0.4–0.8	36.36%
0.8–1.2	30.3%
After SVM	
0–0.4	57.58%
0.4–0.8	42.42%
0.8–1.2	0%

the dataset to test the usefulness of the algorithm. We thus obtained the predicted magnitude of an incoming earthquake using regression equation for the Hokkaido region as another geological scenario (figure 8). We thus used 1500 seismograms recorded at different stations in the same area and found that the method provides consistent average prediction error magnitude less than ±0.4 (figure 10). The lack of availability of a dataset with the same resolutions in other geologic scenarios limits the testing of the algorithm in these regions. The
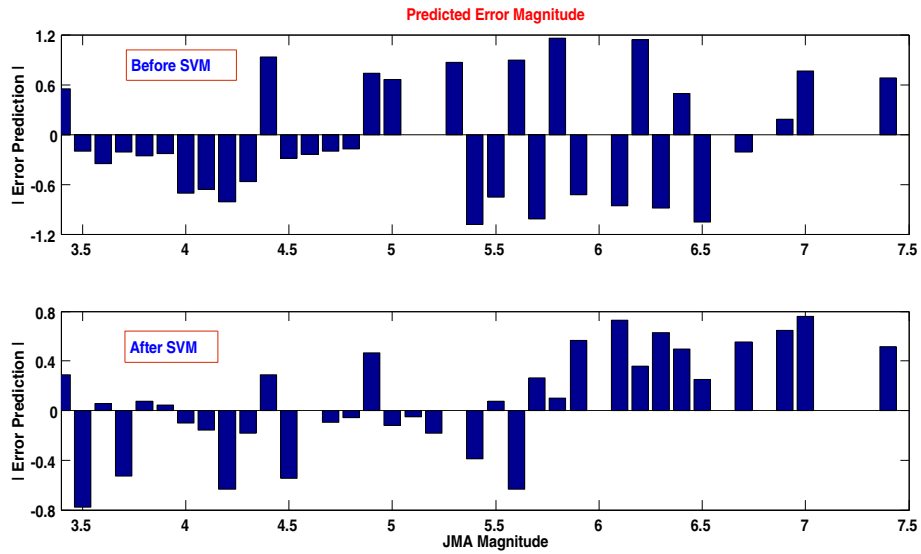
Figure 9. Errors of the prediction from thresholded wavelet coefficients at scale 7, both after and before SVM classification, i.e., the horizontal distance of their average over all reporting stations to the regression lines in figure 8.
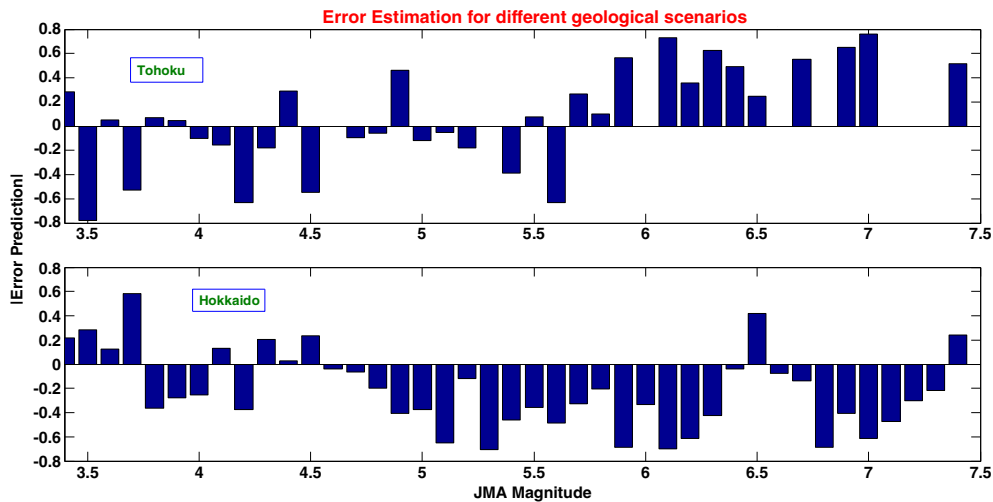

Figure 10. Comparison of predicted errors obtained from different geological scenarios. Top plot is Tohoku and the bottom plot is the Hokkaido region. In both the regions, the wavelet coefficients after the SVM are used.

predicted error magnitudes obtained from different geological scenarios are shown in table 3.

We trained and classified (test) Hokkaido and Tohoku individually using threshold wavelet coefficients. In Tohoku, we produced two results, one with wavelets alone and the second with wavelets +SVM, whereas in Hokkaido, which we consider as another geological scenario, we performed wavelet +SVM. A minimum of three and a maximum of 10 recorded waveforms per event is sufficient to predict the magnitude of an incoming earthquake. We estimated the basic statistics, mean and standard deviation of the predicted errors obtained from wavelets (Tohoku), SVM + wavelets (Tohoku) and SVM + wavelets (Hokkaido) individually (table 4).

Table 3. *Magnitude prediction error for different geological scenario.*

Range of	Percentage of seismograms	
|error magnitude|	Hokkaido	Tohoku
0–0.4	61%	57.58%
0.4–0.8	39%	42.42%
0.8–1.2	0%	0%

As shown in figure 9, after SVM, the methodology constantly overestimates earthquakes with M > 5.6, probably because in the wavelet analysis, we used all training and test data, whereas in SVM we used only classified datasets which are much

Table 4. *Mean and standard deviation of the predicted errors for different geological scenario.*

Statistics	Wavelet approach (Tohoku)			Our approach (SVM + wavelets) (Tohoku)			Hokkaido (SVM + wavelets) (other geological scenario)		
Magnitude	Low	High	ALL	Low	High	ALL	Low	High	ALL
Mean	0.45	0.80	0.62	0.26	0.44	0.35	0.29	0.38	0.33
Standard deviation	0.26	0.29	0.32	0.23	0.22	0.24	0.192	0.20	0.2

fewer compared to those in the wavelet analysis. In wavelet approach, we used least-squares estimation alone which may lead to poor generalization due to over-fitting of the given data, whereas in SVM, there is a 'Regularization Factor', which controls the complexity of the regression equation and avoids over-fitting. The flexibility for further classification into two classes of low and high magnitudes yields the refined regression equations based on the classified wavelet coefficients with substantial improvement in magnitude prediction error percentage and suggests the possibility for the future development of an SVM-based early warning scheme.

Our measurements based on SVMs represent minimum error magnitude estimates for all the events within a region in comparison to wavelet-based-approach alone. The deliberate bias introduced by SVM constraints the predicted magnitude within absolute error magnitude of 0.4. We caution that SVM-based classification led to reasonably faithful prediction of earthquake magnitude with respect to wavelet-based scheme alone and thus contain additional bias. The present procedure demonstrates the efficacy of SVM that is statistical at best for the total events in a region. Further research will have to find physical mechanisms to explain the behaviour if at all it exists.

6. Conclusions

We have presented a fully automated algorithm that has the potential to improve the reliability of magnitude determination from the first few seconds (2.7 s) of the waveform. Scale-dependent threshold amplitudes derived from the wavelet transform of the first 2–3 s of an incoming seismic P arrival are predictive of earthquake magnitude using SVM analysis, with an average error of ±0.4. However, this is not true for a single station. Usually, a single station estimate will not be accurate, although the approach has the advantage of being fast. If the number of stations is three, i.e., three waveforms per event is good enough to predict the magnitude with an error less than ±0.4. The greater the number of stations, the higher will be the accuracy in predicting magnitude. A multiple station approach would provide an accurate and precise method of autonomously detecting the incoming P-wave and predicting the magnitude of the source from the scale-dependent character of its amplitude well before the arrival of damaging ground motion.

Our results thus demonstrate a procedure that is more robust than the analysis with the forward Cohen–Daubechies–Feauveau (2, 4) biorthogonal wavelet transforms based on the lifting scheme. The flexibility for further division into two classes of low and high magnitudes, suggests the possibility for future development of SVM-based early warning schemes. Although in the present case with 1689 seismograms in the Japanese network, we demonstrate the robustness of the algorithm based on the lifting scheme using SVMs with respect to the forward Cohen–Daubechies–Feauveau (2, 4) biorthogonal wavelet transforms alone, our impression is that maximum benefit is derived in the case of a more closed cluster network of seismogram stations.

Acknowledgements

References

Allen R M and Kanamori H 2003 The potential for earthquake early warning in southern California; *Science* **300** 786–789.

Claerbout J F 1976 *Fundamentals of Geophysical Data Processing*; McGraw-Hill, New York.

Cohen A I, Daubechies I and Feauveau J 1992 Biorthogonal bases of compactly supported wavelets; *Comm. Pure Appl. Math.* **45** 485–60.

Daubechies I 1990 The wavelet transform, time–frequency localization and signal analysis; *IEEE Trans. Inf. Theory* **36(5)** 961–1005.

Daubechies I and Sweldens W 1998 Factoring wavelet transforms into lifting steps; *J. Fourier Anal. Appl.* **4(3)** 247–269.

Jensen A 2001 *la Cour-Harbo, Ripples in Mathematics*; Springer, Berlin.

Johnstone I M and Silverman B W 1997 Wavelet threshold estimators for data with correlated noise; *J. R. Stat. Soc. Ser. B* **59(2)** 319–351.

Lin T L and Wu Y M 2010 Magnitude estimation using the covered areas of strong ground motion in earthquake early warning; *Geophys. Res. Lett.* **37** L09301.

Lockman A B and Allen R M 2005 Single-station earthquake characterization for early warning; *Bull. Seismol. Soc. Am.* **95(6)** 2029–2039.

Lockman A B and Allen R M 2007 Magnitude-period scaling relations for Japan and the Pacific northwest: Implications for earthquake early warning; *Bull. Seismol. Soc. Am.* **97(1)** 140–150.

Mallat S 1998 *A wavelet tour of signal processing*; Academic Press, San Diego, Calif.

Mallat S and Zhang Z 1993 Matching pursuits with time–frequency dictionaries; *IEEE Trans. Signal Process.* **41** 3397–3415.

Nakamura Y 1988 On the urgent earthquake detection and alarm system (UrEDAS), *Proc. 9th World Conf. Earthquake Eng.* **VII** 673–678.

Okaya D E, Karageorgi T, McEvilly and Malin P 1992 Removing vibrator-induced correlation artifacts by filtering in frequency-uncorrelated time space; *Geophysics* **57** 916–926.

Olivieri M and Clinton J 2012 An almost fair comparison between Earthworm and SeisComp3; *Seismol. Res. Lett.* **83(4)** 720–727.

Olivieri M, Allen R M and Wurman G 2008 The potential for earthquake early warning in Italy using ElarmS; *Bull. Seismol. Soc. Am.* **98** 495–503.

Olson E L and Allen R M 2005 The deterministic nature of earthquake rupture; *Nature* **438** 212–214.

Refaat M Mohamed and Aly A Farag 2004 Mean field theory for density estimation using support vector machines; *7th Int. Conf. Information Fusion*, Stockholm, pp. 495–501.

Rioul O and Vetterli M 1991 Wavelets and signal processing; *IEEE Signal Process. Mag.* **11** 14–38.

Simons F J, Dando D E and Allen R M 2006 Automatic detection and rapid determination of earthquake magnitude by wavelet multiscale analysis of the primary arrival; *Earth Planet. Sci. Lett.* **250(1–2)** 214–223.

Strang G and Nguyen T 1997 *Wavelets and Filter Banks*, 2nd edn, Wellesley-Cambridge Press, Wellesley, Mass.

Sweldens W 1996 The lifting scheme: A custom-design construction of biorthogonal wavelets; *Appl. Comput. Harmon. Anal.* **3(2)** 186–200.

Tsang L, Allen R M and Wurman G 2007 Magnitude scaling relations from P-waves in southern California; *Geophys. Res. Lett.* **34** L19304.

Vaidyanathan P 1993 *Multirate systems and filter banks*; Prentice Hall.

Vetterli M and Cormac H 1992 Wavelets and filter banks: Theory and design; *IEEE Trans. Signal Process.* **40** 2207–2231.

Welch P D 1967 The use of Fast Fourier Transform for the estimation of power spectra: A method based on time averaging over short, modified periodograms; *IEEE Trans. Audio Electroacoust.* **15** 70–73.

Wu Y M and Kanamori H 2005a Rapid assessment of damaging potential of earthquakes in Taiwan from the beginning of P-waves; *Bull. Seismol. Soc. Am.* **95** 1181–1185.

Wu Y M and Kanamori H 2005b Experiment on an onsite early warning method for the Taiwan early warning system; *Bull. Seismol. Soc. Am.* **95** 347–353.

Wu Y M, Hsin-Yi Yen, Li Zhao, Bor-Shouh Huang and Wen-Tzong Liang 2006 Magnitude determination using initial P-waves: A single-station approach; *Geophys. Res. Lett.* **33** L05306.

Wurman G, Allen R M and Lombard P 2007 Toward earthquake early warning in northern California; *J. Geophys. Res.* **112** B08311.