





Fast estimation of earthquake epicenter distance using a single seismological station with machine learning techniques

Luis H. Ochoa, Luis F. Niño & Carlos A. Vargas

Universidad Nacional de Colombia, Bogotá, Colombia. lhochoag@unal.edu.co, lfninov@unal.edu.co, cavargasj@unal.edu.co

Received: October 19th, de 2017. Received in revised form: November 28th, 2017. Accepted: December 15th. 2017

Abstract

A Support Vector Machine Regression (SVMR) algorithm was applied to calculate the epicenter distance using a ten seconds signal, after primary waves arrive at a seismological station near to Bogota - Colombia. This algorithm was tested with 863 records of earthquakes, where the input parameters were an exponential function of waveform envelope estimated by least squares and maximum value of recorded waveforms for each component of the seismic station. Cross validation was applied to normalized polynomial kernel functions, obtaining mean absolute error for different exponents and complexity parameters. The epicenter distance was estimated with 10.3 kilometers of absolute error, improving the results previously obtained for this hypocentral parameter. The proposed algorithm is easy to implement in hardware and can be employed directly in the field, generating fast decisions at seismological control centers increasing the possibilities of effective reactions.

Keywords: earthquake early warning; support vector machine regression; earthquake; rapid response; epicenter distance; seismic event; seismology; Bogota - Colombia.

Estimación rápida de la distancia epicentral de un terremoto utilizando registros de una sola estación sismológica, mediante técnicas de aprendizaje de máquinas

Resumen

Se aplicó un algoritmo de máquinas de vector de soporte para calcular la distancia epicentral utilizando una señal de diez segundos, después del arribo de ondas primarias a una estación sismológica cercana a Bogotá - Colombia. Este algoritmo fue probado con 863 registros de terremotos donde los parámetros de entrada fueron una función exponencial de la envolvente estimada para los mínimos cuadrados y el valor máximo de las formas de ondas registradas en cada componente de la estación sísmica. Validación cruzada fue aplicada a funciones kernel polinomiales normalizadas, obteniendo la media del error absoluto para diferentes exponentes y parámetros de complejidad. La distancia epicentral se estimó con 10.3 kilómetros de error absoluto, mejorando los resultados previamente obtenidos para este parámetro hipocentral. El algoritmo propuesto es fácil de implementar y puede ser empleado directamente en campo, generando decisiones rápidas en centros de control sismológico incrementado posibilidades de tener reacciones efectivas.

Palabras clave: alerta temprana de terremoto; máquinas de soporte vectorial; terremoto; respuesta rápida; distancia epicentral; evento sísmico; sismología; Bogotá - Colombia.

1. Introduction

Bogota's Savannah and surrounding areas are home to nearly a third of Colombia's population and are the country's main economic center with around 40% of the gross domestic product [1]. In case of a destructive seismic event in this area, the entire country would face many harmful social and economic effects; this is why a seismic early warning system around Bogota is important and the epicenter distance estimation is one of the main parameters in this system. The

© The author; licensee Universidad Nacional de Colombia. Revista DYNA, 85(204), pp. 161-168, March, 2018, ISSN 0012-7353 DOI: http://dx.doi.org/10.15446/dyna.v85n204.68408

How to cite: Ochoa, L.H., Niño, L.F. and Vargas, C.A., Fast estimation of earthquake epicenter distance using a single seismological station with machine learning techniques. DYNA, 85(204), pp. 161-168, March, 2018.

epicenter distance represents the length between the earthquake epicenter and the seismological station, and epicenter is the surface area vertically above the earthquake focus [2]. The density of seismological stations around Bogota is not high enough, making the time used for the localization of seismic events longer than the travel time to areas where the early warning is required. In this case, an alternative solution may be implemented, by using seismological data of previous events recorded at one single station in order to calculate earthquake hypocentral parameters. A seismic early warning system emits an alert from a few seconds to a few tens of seconds before the stronger shaking movement arrives; it can be based on one three-component station, implementing methods of bioinspired computing, natural computation or computational intelligence. These methods have been successfully applied in multiple areas of knowledge; in seismology, these methods allow the estimation of hypocentral parameters using just few seconds of a signal registered at a single seismological station, achieving acceptable accuracy and generating reliable alerts. This approach is very useful in areas with sparse seismic networks [3,4]. Automatic computation algorithms in a single broadband three-component station have been mainly developed for P and S waves onsets detection, allowing the estimation of source location using the back-azimuth and the apparent surface speed measurements [5-7], or seismic moment estimation [8-13]. On the other hand, kernel-based methods have become a very powerful tool for mathematicians, scientists and engineers, providing a very rich and surprising solution in areas such as signal processing and pattern recognition [14]. Its implementation is quite simple by applying a function that combines input variables as a combination of themselves using a function of dot products, obtaining an enhanced new space with more dimensions, mapping variables in a hyperspace where separation of classes (in the case of classification) by a linear function or hyperplane can be achieved.

The study area corresponds to the Bogota's Savannah and its surroundings, where some important fault systems are present, such as Piedemonte Llanero, La Salina, Bogota Savannah and even the Ibague fault system Fig. 1.

Also, the Bogota city has been built on soft lacustrine soil [1], which is a natural seismic wave amplifier producing high damage to the infrastructures, similarly to damages occurred in Mexico City in the past [15]. This area suffers high seismicity that can affect Bogota, the country's capital and most important social and economic populated center.

2. Data set used and methods applied

The data set used in this research belongs to *El Rosal* seismological station, located toward north-west Bogota as shows Fig.2. This station is part of the Colombian Seismic Network administrated by The "*Servicio Geológico Colombiano - SGC*" (Colombian Geological Service).

The Colombian Geological Survey has a main network composed by 42 stations transmitting in real time, recording seismic activity for the entire country as shows Fig.3, with an average distance between stations of 162 kilometers. *El Rosal*



Figure 1. Sketch of Bogota's Savannah and fault systems. Source: The authors.



Figure 2. Distribution of seismic events around Bogota. Source: The authors.

station employs a brand Guralp CMG - T3E007 sensor in the three components and a nanometrics RD3-HRD24 digitizer, which provides simultaneous sampling of three channels with a 24-bit resolution [16]. The data correspond to the three component raw waveforms recorded directly in this station and a seismic catalogue with 2164 characterized events, selected between January 1st 1998 and October 27th 2008; all these events were located less than 120 kilometers from the station.

2.1. Data pre-processing

Before starting the processing related to SVMR, waveform files from *El Rosal* station were converted to the



Figure 3. Seismological Colombian Network. Source: Modified from Google Earth.

American standard code for information interchange (ASCII) format, using a Seisan package tool [17]; earthquakes with magnitudes lower than 2.0 M_L were ignored; therefore the followed processes were applied on the remaining 1011 events. Since the selected seismic records present variable levels of noise, it was necessary to filter them

out with both high and low frequency filters. Low frequencies correspond to instrumental noise that can be easily eliminated through the implementation of a high-pass filter with a cut-off frequency of 0.075 Hz [18], while high frequencies were removed with a low-pass filter with a cut-off frequency of 150 Hz.



Figure 4. Statistical distribution of epicenter distance. Source: The authors.

The statistical distribution of epicenter distance values is presented in Fig.4, where main distribution of whole data set is observed. This bar chart shows that the seismicity surrounding *El Rosal* station is presented from 40 kilometers and beyond, with great amount of events close to 90 kilometers. Although this is not a homogeneous distribution, it corresponds to local conditions and regular behavior for this variable and therefore the model has to work properly under this condition.

2.2. Descriptors – Input data set of SVMR

In this study, some parameters that have been previously used for other authors to magnitude estimation were calculated; then, they were employed as input variables or descriptors for the SVMR algorithm. In the first stage, the relationship between maximum amplitude of the wave in a short period of time was selected, along with local magnitude of the earthquake [19]. Consecutive, maximum peaks were highlighted and a linear regression was performed for each of the three components, correlating not only the maximum peak, but also the way it changes while energy reaches the sensors. Three basic parameters were chosen from the linear regressions such as, slope (M), independent term (B) and correlation coefficient (R), for each of the three components. The maximum amplitude values (Mx) obtained for the time of each component were used as descriptors as well. Thus, each event had 12 descriptors associated with this concept.

Furthermore, nine descriptors used previously for epicenter distance estimation were added. These descriptors were performed by adjusting a linear regression of an exponential in eq. (1):

$$Bt \exp(-AT) \tag{1}$$

This equation belongs to the envelope of the seismic record in a logarithmic scale determined by linear regression

and its respective correlation coefficient (R) for each component [10].

Similarly, some parameters used for previous backazimuth determination were used to include information about the source location of the seismic event into the model. Maximum eigenvalues of the two-dimensional covariance matrix were used as input descriptors, calculated as described in [5,20]. A windowing scheme with one second time windows was performed, obtaining consecutive values for which a linear regression was calculated, in a similar way as described above, determining the slope (M), the independent term (B) and the correlation factor (R) of the regression, as well as the arithmetic mean of the eigenvalues (P). Despite this parameter involving three components, only four descriptors were added to the process.

In sum, a total of 25 descriptors were used in the SVMR model for local epicenter distance determination, 12 of them related to previous works on magnitude estimation, 9 were associated with previous epicenter distance estimations and the last 4 descriptors were used in back-azimuth determination. These descriptors were calculated for 5, 10 and 15 seconds signal windows for 1011 selected events with magnitudes greater than 2.0. All extreme or anomalous values were eliminated reducing the data set to 863 events.

2.3. The SVMR model

The model was trained with the refined data set for each time window, using the Weka 3.6 software [21]. This algorithm has a strong statistical support and is easily implemented on the station by electronic processing cards. After performing several tests, a standard normalized polynomial kernel was selected. In order to choose the kernel exponent and the complexity factor, correlation factors and minimum absolute error obtained by a 10 fold cross correlation process were compared. These processes were carried out testing multiple combinations of exponents and complexity factor for deferments magnitudes and time signals. The correlation coefficient calculated for each partition corresponds to the Pearson's Coefficient, which measure the linear relationship between two variables independently of their scales. This coefficient takes values between 1 and -1; a value of zero means that a linear relationship between two variables could not be found. A positive value of this relation means that two variables change in the same way, i.e. high values of one variable correspond to high values of the other and vice versa. The closer this value is to one, the greater certainty that two variables have a linear relation.

3. Results

Using the descriptors explained above and real magnitude for each considered seismic event, a group of 12 datasets were evaluated. Each dataset corresponds to a combination of 4 minimum magnitude filters (2.0, 2.5, 3.0 and 3.5) and 3 signal length filters (5, 10 and 15 s), evaluating combinations of 7 values for kernel exponent and 6 values for complexity factor, completing a total of 504 tested models of SVMR in order to find the combination of parameters with the best correlation factor in epicenter distance determination.

Magnitude >= 3.0

Magnitude >= 3.5

Magnitude >= 2.5

		-								1	_																1 1								
					EX	PON	ENT							EXI	ON	ENT						EXF	ONE	ENT							EXP	ON	ENT		
			1.5	ы	4	'n	đ	8	8			1.5	ы	4	'n	đ	8	8		1.5	ы	4	ы	10	ล	ß			1.5	ы	4	'n	8	ล	6
s		1	0.508	0.517	0.544	0.556	0.573	0.547	0.500		1	0.556	0.565	0.573	0.578	0.594	0.581	0.563	1	0.621	0.648	0.687	0.695	0.677	0.648	0.549		1	0.413	0.453	0.507	0.514	0.494	0.446	0.342
= 15	TOR	3	0.513	0.525	0.559	0.566	0.556	0.503	0.462		3	0.564	0.562	0.574	0.581	0.586	0.540	0.536	3	0.653	0.657	0.697	0.695	0.678	0.642	0.549		3	0.459	0.486	0.491	0.480	0.461	0.446	0.342
lime	ПҮЕА	5	0.517	0.533	0.561	0.565	0.540	0.471	0.449		5	0.564	0.558	0.575	0.575	0.563	0.517	0.522	5	0.638	0.646	0.691	0.689	0.663	0.640	0.549		5	0.473	0.491	0.466	0.455	0.461	0.446	0.342
gnal 7	MPLEX	10	0.523	0.537	0.560	0.560	0.510	0.424	0.434		10	0.553	0.555	0.570	0.577	0.512	0.490	0.514	10	0.629	0.660	0.683	0.677	0.650	0.640	0.549		10	0.486	0.484	0.434	0.451	0.461	0.446	0.342
Sig	8	20	0.524	0.534	0.559	0.552	0.473	0.391	0.420		20	0.544	0.552	0.569	0.563	0.459	0.468	0.514	20	0.639	0.666	0.662	0.640	0.646	0.640	0.549		20	0.474	0.446	0.440	0.451	0.461	0.446	0.342
		50	0.516	0.538	0.544	0.511	0.405	0.354	0.409		50	0.541	0.553	0.542	0.503	0.403	0.428	0.514	50	0.642	0.647	0.553	0.549	0.646	0.640	0.549		50	0.429	0.385	0.440	0.451	0.461	0.446	0.342
		<u> </u>								1																									
		-	1.5	ы	4	IJ	10	2	8			1.5	ы	4	IJ	10	20	8		1.5	ы	4	ы	10	20	ន			1.5	ы	4	ы	10	20	ន
		1	0.539	0.560	0.587	0.590	0.587	0.577	0.550		1	0.558	0.558	0.564	0.563	0.566	0.589	0.613	1	0.621	0.634	0.675	0.679	0.677	0.665	0.597		1	0.518	0.499	0.411	0.417	0.463	0.488	0.334
= 10	ä	3	.559 (.573 (.584 (988.	.570 (.560 (.515 (3	.542 (.541 (.554 (.558	.562 (.587	.583	3	.645 (.653 (.680	0880	.633 (.650 (.597 (3	.459 (.418 (.407 (.429 (.466 (.488	.334 (
ue.	YFACT	5	.563 0	.575 0	583 0	583	554 0	536 0	494 0		5	537 0	550 0	.558 0	561 0	.551 0	.570 0	0 085	5	.642 0	.650 0	.672 0	649 0	.627 0	.651 0	597 0		5	.435 0	.423 0	424 0	.433 0	.466 0	488 0	.334 0
ial Ti	ргехи	10	569 0	575 0	579 0	568 0	540 0	499 0	479 0		10	542 0	556 0	560 0	559 0	540 0	550 0	0 085	10	636 0	645 0	633 0	611 0	602 0	651 0	597 0		10	423 0	401 0	426 0	433 0	466 0	488 0	33.4 0
Sig	COM	20	567 0.	575 0.	560 0.	549 0.	516 0.	466 0.	471 0.		20	550 0.	548 0.	551 0.	535 0.	508 0.	516 0.	580 0.	20	623 0.	644 0.	598 0.	594 0.	599 0.	651 0.	597 0.		20	412 0.	406 0.	426 0.	433 0.	466 0.	488 0.	334 0.
		50	569 0.	564 0.	533 0.	525 0.	45.4 0.	404 0.	461 0.		50	542 0.	558 0.	518 0.	491 0.	460 0.	482 0.	880 0.	50	623 0.	610 0.	565 0.	544 0.	599 0.	651 0.	597 0.		50	405 0.	415 0.	426 0.	433 0.	466 0.	\$88 0.	33.4 0.
			0.1	0	0.1	0.1	0	0	0			0.1	0.1	0.1	0	0	0	0.1	~	0.1	0.1	0.1	0.1	0.1	0	0.1			0.	0	0	0	0	0	0
			ы				_	_]		ы				_	_	_		IJ				_	_	_	1		ы					_	_
			1.	N N	4	л С	1 10	4 20	1 20			8 1.	2	10 14	ю 00	0 10	9 20	5		0 1.	80	4	7 5	5 10	4 20	S S			6 1.	N 20	4	n m	7 10	20	1
5 s		1	0.45	3 0.47	4 0.48	1 0.48	1 0.48	2 0.47	3 0.38		1	8 0.49	3 0.49	7 0.49	3 0.49	7 0.49	1 0.46	1 0.43	1	9 0.52	1 0.51	0.53	9 0.53	1 0.51	6 0.46	0.28		1	4 0.39	9 0.37	9 0.37	6 0.39	8 0.39	3 0.14	1 -0.2
e = 0	CLOR	3	0.470	0.475	0.484	0.483	0.473	0.442	0.315		3	0.483	0.495	0.507	0.495	0.457	0.461	0.383	3	0.489	0.503	0.540	0.539	0.483	0.356	0.279		3	0.374	0.405	0.455	0.455	0.303	0.145	-0.23
Tim	КПҮ FA	5	0.465	0.478	0.478	0.471	0.472	0.419	0.305		5	0.489	0.500	0.498	0.480	0.440	0.440	0.358	5	0.487	0.522	0.547	0.533	0.418	0.330	0.279		5	0.420	0.458	0.454	0.422	0.307	0.148	-0.21
ignal	OMPLE	10	0.465	0.479	0.464	0.457	0.464	0.372	0.293		10	0.493	0.510	0.471	0.455	0.426	0.411	0.357	10	0.507	0.545	0.527	0.481	0.341	0.318	0.279		10	0.493	0.461	0.398	0.390	0.307	0.148	-0.21
S	8	20	0.472	0.475	0.446	0.444	0.439	0.323	0.289		20	0.501	0.495	0.444	0.429	0.419	0.362	0.349	20	0.538	0.548	0.458	0.410	0.295	0.318	0.279		20	0.478	0.453	0.402	0.391	0.307	0.148	-0.21
		50	0.469	0.448	0.426	0.439	0.378	0.258	0.282		50	0.475	0.445	0.419	0.398	0.375	0.288	0.336	50	0.540	0.520	0.365	0.317	0.272	0.318	0.279		50	0.439	0.406	0.402	0.391	0.307	0.148	-0.21

Figure 5. Correlation coefficients for each combination of Kernel. Source: The authors.

Magnitude >= 2.0

Fig. 5 shows values of correlation coefficients in each combination of cut-off magnitude and time signal where kernel exponents and complexity factors were calculated for each combination, where values in green squares represent better correlation, while those in red squares represent values of lower correlation. Afterward, those with the highest value of correlation coefficient were chosen. Fig. 5 also shows optimal values of correlation coefficients for time signals of 10 and 15 seconds, the average being of 0.7.

This value is an acceptable correlation, indicating that the model is predicting the epicenter distance with good accuracy. For a time signal of 5 seconds, correlation

			SUMMA	RY OF EP	ICENTER	DISTANC	E DETER	MINATIO	N - RESID	UAL		
		Signal Ti	me = 15 s			Signal Ti	ime = 10 s			Signal Tin	ne = 05 s	
Minimum Magnitude	2.0	2.5	3.0	3.5	2.0	2.5	3.0	3.5	2.0	2.5	3.0	3.5
Mean	0.16	0.13	1.03	-0.46	0.21	0.56	0.45	-1.06	1.05	-0.16	-0.96	-0.39
Typical Error	0.51	0.63	0.80	1.14	0.51	0.70	0.80	1.74	0.56	0.73	0.95	0.65
Standard Deviation	15.1	12.3	8.9	6.6	15.0	13.6	9.0	10.0	16.6	14.2	10.6	3.7
Kurtosis	1.34	1.36	5.13	10.85	1.48	0.57	3.30	1.97	0.82	0.92	3.57	18.48
Asymmetric Coefficient	0.18	-0.10	1.51	-2.50	0.10	-0.05	0.14	-1.05	0.35	0.10	0.01	-3.10
Count	863	379	126	33	863	379	126	33	864	377	126	33
Level of Confidence (95%)	1.009	1.245	1.574	2.323	1.00	1.377	1.578	3.540	1.107	1.437	1.873	1.317

Summary for the best epicenter distance models in each combination.

Source: The authors.

Table 1

coefficient values are just above 0.5, pointing out that this time signal is not long enough to estimate an acceptable epicenter distance.

Although a correlation coefficient is more suited to a time signal of 15 seconds, the 10-signal seconds was selected as the best model in the shortest period of time; moreover, a cut-off magnitude must be established for the 10 seconds signal allowing accuracy in the final value of epicenter distance.

The choice of parameters is shown in Fig.6, where correlation coefficient and mean absolute error are presented for each combination of kernel exponent and complexity factor, all of them for the time signal and the cut-off magnitude selected.

CORRELATION COEFFICIENT		EXPONENT													
		0.5	0.8	1.1	1.5	2	4	5	10	20	50				
0.1		0.568	0.571	0.590	0.606	0.613	0.633	0.638	0.659	0.648	0.596				
	0.5	0.604	0.599	0.602	0.608	0.619	0.661	0.675	0.679	0.663	0.596				
ĸ	0.8	0.597	0.604	0.608	0.618	0.631	0.674	0.679	0.684	0.660	0.596				
ğ	1	0.598	0.606	0.612	0.621	0.635	0.675	0.679	0.677	0.665	0.597				
ΣE	2	0.607	0.618	0.618	0.634	0.653	0.677	0.686	0.638	0.651	0.597				
EX.	4	0.611	0.614	0.629	0.645	0.653	0.677	0.667	0.631	0.651	0.597				
Mo	5	0.616	0.610	0.632	0.642	0.650	0.672	0.649	0.627	0.651	0.597				
Ŭ	10	0.613	0.625	0.625	0.636	0.645	0.633	0.612	0.602	0.651	0.597				
	20	0.428	0.584	0.612	0.623	0.644	0.599	0.594	0.599	0.651	0.597				
	50	0.241	0.277	0.596	0.624	0.611	0.565	0.543	0.599	0.651	0.597				
MEAN A	MEAN ABSOLUTE		EXPONENT												
ERROR															
Liu	IOR	0.5	0.8	1.1	1.5	2	4	5	10	20	50				
	OR 0.1	0.5 12.58	0.8 12.35	1.1 12.13	1.5 11.88	2 11.66	4 11.16	5 10.97	10 10.43	20 10.48	50 11.48				
	OR 0.1 0.5	0.5 12.58 11.37	0.8 12.35 11.28	1.1 12.13 11.25	1.5 11.88 11.18	2 11.66 11.02	4 11.16 10.45	5 10.97 10.19	10 10.43 10.20	20 10.48 10.15	50 11.48 10.62				
	0.1 0.5 0.8	0.5 12.58 11.37 11.29	0.8 12.35 11.28 11.30	1.1 12.13 11.25 11.25	15 11.88 11.18 11.09	2 11.66 11.02 10.94	4 11.16 10.45 10.29	5 10.97 10.19 10.38	10 10.43 10.20 10.30	20 10.48 10.15 10.39	50 11.48 10.62 10.58				
ACTOR	0.1 0.5 0.8 1	0.5 12.58 11.37 11.29 11.31	0.8 12.35 11.28 11.30 11.29	1.1 12.13 11.25 11.25 11.19	1.5 11.88 11.18 11.09 11.05	2 11.66 11.02 10.94 10.93	4 11.16 10.45 10.29 10.39	5 10.97 10.19 10.38 10.52	10 10.43 10.20 10.30 10.44	20 10.48 10.15 10.39 10.39	50 11.48 10.62 10.58 10.56				
TY FACTOR	0.1 0.5 0.8 1 2	0.5 12.58 11.37 11.29 11.31 11.32	0.8 12.35 11.28 11.30 11.29 11.09	1.1 12.13 11.25 11.25 11.19 11.14	1.5 11.88 11.18 11.09 11.05 10.96	2 11.66 11.02 10.94 10.93 10.65	4 11.16 10.45 10.29 10.39 10.74	5 10.97 10.19 10.38 10.52 10.69	10 10.43 10.20 10.30 10.44 11.28	20 10.48 10.15 10.39 10.39 10.76	50 11.48 10.62 10.58 10.56				
LEXITY FACTOR	0.1 0.5 0.8 1 2 4	0.5 12.58 11.37 11.29 11.31 11.32 11.13	0.8 12.35 11.28 11.30 11.29 11.09 11.16	1.1 12.13 11.25 11.25 11.19 11.14 11.13	1.5 11.88 11.18 11.09 11.05 10.96 10.91	2 11.66 11.02 10.94 10.65 10.93	4 11.16 10.45 10.29 10.39 10.74 11.04	5 10.97 10.19 10.38 10.52 10.69 11.18	10 10.43 10.20 10.30 10.44 11.28 11.75	20 10.48 10.15 10.39 10.76 10.81	50 11.48 10.62 10.58 10.56 10.56				
OMPLEXITY FACTOR	0.1 0.5 0.8 1 2 4 5	0.5 12.58 11.37 11.29 11.31 11.32 11.13 11.10	0.8 12.35 11.28 11.30 11.29 11.09 11.16 11.22	1.1 12.13 11.25 11.25 11.19 11.14 11.13 11.09	15 11.88 11.18 11.09 11.05 10.96 10.91 11.01	2 11.66 11.02 10.93 10.65 10.93 11.20	4 11.16 10.45 10.29 10.39 10.74 11.04 11.19	5 10.97 10.38 10.52 10.69 11.18 11.49	10 10.43 10.20 10.30 10.44 11.28 11.75 11.96	20 10.48 10.15 10.39 10.39 10.76 10.81 10.81	50 11.48 10.62 10.58 10.56 10.56				
COMPLEXITY FACTOR	OR 0.1 0.5 0.8 1 2 4 5 10	0.5 12.58 11.37 11.29 11.31 11.32 11.13 11.10 11.50	0.8 12.35 11.28 11.30 11.29 11.09 11.16 11.22 11.13	1.1 12.13 11.25 11.25 11.19 11.14 11.13 11.09 11.44	1.5 11.88 11.18 11.09 11.05 10.96 10.91 11.01 11.43	2 11.66 11.02 10.94 10.93 10.65 10.93 11.20 11.80	4 11.16 10.45 10.29 10.74 11.04 11.04 11.94	5 10.97 10.19 10.38 10.52 10.69 11.18 11.49 12.44	10 10.43 10.20 10.30 10.44 11.28 11.75 11.96 12.68	20 10.48 10.15 10.39 10.76 10.81 10.81 10.81	50 11.48 10.62 10.56 10.56 10.56 10.56				
COMPLEXITY FACTOR	00R 0.1 0.5 0.8 1 2 4 5 10 20	0.5 12.58 11.37 11.29 11.31 11.32 11.13 11.10 11.50 14.85	0.8 12.35 11.28 11.30 11.29 11.09 11.16 11.22 11.13 12.00	1.1 12.13 11.25 11.25 11.19 11.14 11.09 11.44 11.73	11.88 11.18 11.09 11.05 10.96 10.91 11.01 11.43 12.12	2 11.66 11.02 10.94 10.65 10.93 11.20 11.80 12.08	4 11.16 10.45 10.29 10.39 10.74 11.04 11.19 11.94 12.93	5 10.97 10.19 10.38 10.69 11.18 11.49 12.44 13.36	10 10.43 10.20 10.30 11.44 11.28 11.75 11.96 12.68 12.70	20 10.48 10.15 10.39 10.76 10.81 10.81 10.81 10.81	50 11.48 10.62 10.56 10.56 10.56 10.56				

Figure 6. Parameters selection for each dataset. Source: The authors.

Table 1 shows, for a time signal of 10 seconds and 3.0 of cut-off magnitude, a mean value of 0.45 kilometers with standard deviation of 9 kilometers, this cut-off being the one finally implemented in the model.

These parameters were calculated using SVMR algorithm in Weka 3.6 with a standard normalized polynomial kernel and 10 fold cross-validations. We can see that quality factors enable an accuracy of 10.9 kilometers in epicenter distance, considering the cross-validation, and also allowing verification of the model at the same time.

Based on these results, a support vector machine of normalized polynomial kernel can be implemented using an exponent of 10 and a complexity factor of 0.8. The standard deviation was of 10.3 kilometers, which is good enough for an early warning generation, considering that most of the seismic events are located farther than 40 kilometers from *El Rosal* station.

Fig. 7 shows the cross-plot with relationship between the real epicenter distance (X axis) and the distance calculated by the model (Y axis). A normal statistical pattern can be observed in the distribution of residuals, also confirming that the calculated distance is of around 10.9 kilometers. The dashed blue line represents the linear behavior of predicted data, corresponding to the locus where prediction is equal to real values.

From these results, we can establish that the model works properly, allowing prediction of the distance where earthquakes occur with good accuracy, considering that the values are obtained with a 10 seconds signal, which is good enough for an early warning system.

4. Conclusions

The SVMR model proposed here is an important step towards the implementation of an early warning system of earthquakes for the city of Bogota - Colombia, and for other populated centers in the world.

The result showed in this study is an improvement on that of [22] and [23], who were as accurate as ± 15 and $\pm 16-19$ kilometers respectively.



Figure 7. Correlation between real and calculated epicenter distance with SVMR.

Source: The authors.

This model is proposed and evaluated for fast epicenter distance determination, based on support vector machine regression through pattern recognition and characterization of earthquake signals recorded on a three components seismic station in only ten seconds, anticipating the arrival of earthquakes in the city of Bogota. An earthquake travels the distance to main seismic alignments in at least 30 seconds, allowing an early warning generation, which must be in less than 10 seconds. Additionally, this model can be implemented directly in the seismological station embedded in electronic devices, where the main mathematical process corresponds to a simple matrix product involving the given kernel for the epicenter distance and a vector which contains the calculated descriptors of the current event.

5. Recommendations

Despite the good results obtained in terms of epicenter distance determination, it is important to continue developing the model introduced here, in order to calculate other hypocentral parameters.

It is important to find ways to improve the prediction accuracy based on further research, supported by computational intelligence and geophysics research groups as well as the seismological network in Bogota's Savannah and its surroundings managed by the Universidad Nacional de Colombia.

The use of other descriptors such as predominant period, Fourier and wavelet frequency spectra should be considered to obtain higher correlation factors and better estimation values for local magnitude and other hypocentral parameters estimations, required to generate a reliable and fast earthquake early warning system.

Datasets should be complemented with recent seismic events, specifically from October 27th of 2008 to the present, as this period has presented larger set of earthquakes with magnitudes greater than 3.0.

Acknowledgments

The authors are grateful to the Servicio Geológico Colombiano (SGC) for providing the dataset used in this study and to Universidad Nacional de Colombia for supporting our efforts to achieve a fast and reliable early warning system for Bogota D.C. – Colombia.

References

- [1] Ojeda, A., Martinez, S., Bermudez, M. and Atakan, K., The new accelerograph network for Santa Fé de Bogotá, Colombia. Soil Dynamics and Earthquake engineering, October - December, 22(9-12), pp. 791-797, 2002. DOI: 10.1016/S0267-7261(02)00100-8
- [2] Don, L. and Judson, S., Fundamentos de geología física. México: Editorial Limusa, 2000.
- [3] Ochoa, L.H., Niño, L.F. and Vargas, C.A., Severity classification of a seismic event based on the magnitude-distance ratio using only one seismological station. Earth Sciences Research Journal, December, 18(2), pp. 115-122, 2014. DOI: 10.15446/esrj.v18n2.41083
- [4] Ochoa, L.H., Niño, L.F. and Vargas, C.A., Fast magnitude determination using a single seismological station record implementing machine learning techniques. Sciences Direct, Geodesy and Geodynamic, March, pp. 1-8, 2017. DOI: 10.1016/j.geog.2017.03.010
- [5] Magotra, N., Ahmed, N. and Chael, E., Seismic event detection and source location using single station (three components) data. Bull. Seism. Asoc. Am., June, 77(3), pp. 958-971, 1987.
- [6] Roberts, R.G., Christoffersson, A. and Cassidy, F., Real-time detection, phase identification and source location estimation using single station three component seismic data. Geophysical Journal, June, 97(3), pp. 471-480, 1989. DOI: 10.1111/j.1365-246X.1989.tb00517.x
- [7] Saita, J. and Nakamura, Y., The early warning systems for mitigation of disasters caused by earthquakes and tsunamis. In: Zschau, J. and Kuppers, A., eds. Early warning systems for natural disaster reduction. Berlin: Springer-Verlag, pp. 453-460, 2003. DOI: 10.1007/978-3-642-55903-7_58
- [8] Talandier, J., Reymond, D. and Oka, E.A., Use of variable mantle magnitude for the rapid one-station estimation of teleseismic moments. Geophysical Research Letters, August, 14(8), pp. 840-843, 1987. DOI: 10.1029/GL014i008p00840
- [9] Reymond, D., Hyvernaud, O. and Talandier, J., Automatic detection, location and quantification of earthquakes. Pure and Applied Geophysics, March, 135(3), pp. 361-382, 1991. DOI: 10.1007/BF00879470
- [10] Odaka, T. et al., A new method for quickly estimating epicentral distance and magnitude from a single seismic record. Bull. Seism. Soc. Am., February, 93(1), pp. 526-532, 2003. DOI: 10.1785/0120020008
- [11] Horiuchi, S. et al., An automatic processing system for broadcasting earthquake alarms. Bull. Seism. Asoc. Am., April, 95(2), pp. 708-718, 2005. DOI: 10.1785/0120030133
- [12] Wu, Y.M., Shin, T.C. and Tsai, Y.B., Quick and reliable determination of magnitude for seismic early warning. Bull. Seism. Soc. Am., October, 88(5), pp. 1254-1259, 1998.
- [13] Espinosa-Aranda, J.M. et al., Mexico city seismic alert system. Seismol. Res. Lett., November, 66(6), pp. 42-53, 1995. DOI: 10.1785/gssrl.66.6.42
- [14] Taylor, J.S. and Cristianini, N., Kernel methods for pattern recognition. First ed. Cambridge, United Kingdom: Cambridge University Press, 2004.
- [15] Rodríguez, J.A., Ramirez, F. and Escallon, J.P., Geotechnical seismic characterization for the microzonation of Bogotá. Thessaloniki-Greece, 4 International Conference on Earthquake Geotechnical Engineering, 2007.
- [16] Bermúdez, M.L. and Rengifo, F., EL ROSAL: La Estación Sismológica del CTBTO en Colombia. Bogota, Primer Simposio Colombiano de Sismología, 2002, 8 P.

- [17] Ottemoller, L., Voss, P. and Havskov, J., Seisan earthquake analysis software for windows, solaris, Linux and Macosx, s.l.: s.n., 2016.
- [18] Wu, Y.M. and Zhao, L., Magnitude estimation using the first three seconds P-wave amplitude in earthquake early warning. Geophys. Res. Lett., August, 33(16), pp. L16312, 2006. DOI: 10.1029/2006GL026871
- [19] Wu, Y.M. and Kanamori, H., Rapid assessment of damage potential of earthquakes in Taiwan from beginning of P waves. Bull. Seism. Soc. Am., February, 93(1), pp. 526-532, 2005. DOI: 10.1785/0120040193
- [20] Magotra, N., Ahmed, N. and Chael, E., Single-station seismic event detection and location. IEEE Transactions on Geoscience and Remote Sensing, January, 27(1), pp. 15-23, 1989. DOI: 10.1109/36.20270
- [21] Eibe, F., Hall, M.A. and Witten, I.H., The WEKA workbench. Online appendix for "Data mining: Practical machine learning tools and techniques. Fourth Edition ed. s.l.: Morgan Kaufmann, 2016.
- [22] Lockman, A.B. and Allen, R.M., Single-station earthquake characterization for early warning. Bull. Seism. Soc. Am., December, 95(6), pp. 2029-2039, 2005. DOI: 10.1785/0120040241
- [23] Böse, M., Heaton, T.H. and Hauksson, E., Rapid estimation of earthquake source and ground-motion parameters for earthquake early warning using data from a single three-component broadband or strong-motion sensor. Bulletin of the Seismological Society of America, 102(2), pp. 738-750, 2012. DOI: 10.1785/0120110152.

L.H. Ochoa, received his BSc. Eng in Civil Engineering in 1988, a MSc. degree in Geophysics in 2003, a MSc. degree in Geomatics in 2007, and a PhD. in System Engineering in 2017, all of them from the Universidad Nacional de Colombia, Bogotá. He dedicates himself to research in intelligent systems, geophysics and geodesy. His main interests include earthquake early warning using intelligent systems. ORCID: 0000-0002-3607-7339

L.F. Niño, received his BSc. Eng in System Engineering (1990) and a MSc. degree in Mathematics in 1995 from Universidad Nacional de Colombia, Bogota – Colombia. A MSc. degree in Computer Science (1999) and a PhD. degree in Computer Science at The University of Memphis – United States of America. He dedicates himself to teaching and research in computer science, and his main academic interests include computational intelligence and its applications, particularly in the life science. ORCID: 0000-0003-4703-0007

C.A. Vargas, received his BSc. in Geology (1993) from Universidad de Caldas, Colombia. A MSc. degree in Seismic Engineering and Structural Dynamic (2000) from Universidad Politécnica de Catalunya, Spain. A MSc. degree in Physics Instrumentation (2006) from Universidad Tecnológica de Pereira - Colombia. His PhD. in Seismic Engineering and Structural Dynamic (2003) from Universidad Politécnica de Catalunya, Spain. His postdoctoral research (2010) from University of Texas, Institute for Geophysics, United States of America. He dedicates himself to research in geodynamic and basin analysis. His main interest include subduction systems and their response in sedimentary basins. ORCID: 0000-0002-5027-9519



UNIVERSIDAD NACIONAL DE COLOMBIA SEDE MEDELLÍN FACULTAD DE MINAS

Área Curricular de Ingeniería Civil

Oferta de Posgrados

Especialización en Vías y Transportes Especialización en Estructuras Maestría en Ingeniería - Infraestructura y Sistemas de Transporte Maestría en Ingeniería – Geotecnia Doctorado en Ingeniería - Ingeniería Civil

Mayor información:

E-mail: asisacic_med@unal.edu.co Teléfono: (57-4) 425 5172