



Fast estimation of earthquake arrival azimuth using a single seismological station and machine learning techniques

Luis Hernán Ochoa Gutiérrez, Carlos Alberto Vargas Jiménez, Luis Fernando Niño Vásquez
Universidad Nacional de Colombia

* Corresponding author: lhochoag@unal.edu.co

ABSTRACT

The objective of this research is to develop a new approach to estimate earthquake arrival azimuth using seismological records of the “El Rosal” station, near to the city of Bogota – Colombia, by applying support vector machines (SVMs). The algorithm was trained with time series descriptors of 863 events recorded from January 1998 to October 2008, considering only events with magnitude $\geq 2 M_L$. The earthquake signals were filtered in order to remove diverse kind of low and high-frequency noise not related to typical seismic activity in the area. During training stages of SVMs, several combinations of kernel exponent and complexity factor were applied to time series of 5, 10 and 15 seconds along with earthquake magnitudes of 2.0, 2.5, 3.0 and 3.5 M_L . The best classification of SVMs was obtained using time windows of 5 seconds and earthquake magnitudes greater than 3.0 M_L with kernel exponent of 10 and complexity factor of 2, showing an accuracy of 45.4 degrees. This research is an improvement of previous works related to earthquake arrival azimuth determination from single station data employing machine learning techniques.

Keywords: Earthquake early warning; rapid response; earthquake arrival azimuth; seismic event; Bogota-Colombia; support vector machine (SVM); seismology; earthquakes.

Estimación rápida del azimut de llegada de un terremoto utilizando registros de una sola estación sismológica y técnicas de aprendizaje de máquinas

RESUMEN

El propósito de esta investigación es desarrollar un nuevo enfoque para estimar el azimut de llegada de terremotos utilizando registros sismológicos de la estación El Rosal, cercana a la ciudad de Bogotá – Colombia, mediante la aplicación de máquinas de vectores de soporte (MVS). El algoritmo fue entrenado con descriptores de series de tiempo de 863 eventos adquiridos desde Enero 1998 hasta Octubre de 2008, considerando solamente eventos con magnitudes $\geq 2 M_L$. Las señales de los terremotos fueron filtradas para remover diversos tipos de ruidos de alta y baja frecuencia no relacionados con la actividad sísmica típica en el área. Durante las etapas de entrenamiento de la MVS fueron aplicadas varias combinaciones del exponente kernel y factor de complejidad, a series de tiempo de 5, 10 y 15 segundos junto con terremotos de magnitudes mayores a 2.0, 2.5, 3.0 y 3.5 M_L . La mejor clasificación de la MVS fue obtenida utilizando ventanas de tiempo de 5 segundos y terremotos de magnitud mayor a 3.0 M_L con exponente kernel de 10 y factor de complejidad de 2, mostrando una precisión de 45.4 grados. Esta investigación es una mejora a trabajos previos relacionados con determinación del azimut de llegada de un terremoto a partir de datos de una única estación sismológica empleando técnicas de aprendizaje de máquinas.

Palabras clave: Alerta temprana de terremoto; respuesta rápida; Azimut de llegada; evento sísmico; Bogotá-Colombia; máquina de soporte vectorial (MSV); sismología; terremotos.

Record

Manuscript received: 22/02/2018
Accepted for publication: 15/03/2019

How to cite item

Ochoa-Gutiérrez, L. H., Vargas-Jimenez, C. A., & Niño-Vásquez, L. F. (2019). Fast estimation of earthquake arrival azimuth using a single seismological station and machine learning techniques. *Earth Sciences Research Journal*, 23(2), 103-109.
DOI: <https://doi.org/10.15446/esrj.v23n2.70581>

Introduction

This study is part of a research line which proposes calculation of earthquake hypocentral parameters applying artificial intelligence methods in order to develop an early warning system for the city of Bogota. 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 earthquake arrival azimuth estimation is one of the main parameters in this system. Nearly a third of Colombia's population lives in Bogota's Savannah and surrounding which is the country's main economic center with around 40% of the gross domestic product (Ojeda et al., 2002). The density of seismological stations around the city is not high enough, this makes the required time for seismic events localization to be longer than the travel time to areas where the early warning is required. An alternative solution to this problem is employing seismological data from previous events recorded at one single station to estimate the earthquake hypocentral parameters (Ochoa et al., 2014). Automatic computation algorithms in a single broadband three-component station have been mainly developed for P and S waves onsets detection, allowing estimation of source location using the back-azimuth and the apparent surface speed measurements (Magotra et al., 1987; Roberts et al., 1989; Saita & Nakamura, 2003), or seismic moment estimation (Talandier et al., 1987; Raymond et al., 1991; Odaka et al., 2003; Horiuchi et al., 2005; Wu et al., 1998; Espinosa, 1995). Supervised machine learning techniques based on kernel methods have become a very powerful tool for mathematicians, scientist, and engineers, providing solution in areas like signal processing and pattern recognition; its implementation is quite simple and can be performed by applying mathematical functions that combine input variables as a combination of themselves, obtaining an enhanced new space with more dimensions where separation of classes can be achieved.

There are several methods to detect seismic wave and its arrival azimuth in a single three-component station (Magotra et al., 1987; Anant & Dowl, 1997); these authors employed algorithms that measure the level of linear polarization in the P wave's arrival. The methodology proposed in this research consists of applying SVMs along with a kernel function in order to estimate the arrival azimuth with minimal processing of data acquired at the station, similar to methodology applied to a fast determination of earthquake magnitude and epicenter distance using a single seismological station (Ochoa et al., 2017; Ochoa et al., 2018).

Data Sets And Methods

The dataset used in this research belongs to the "El Rosal" seismological station, located toward north-west Bogota as shows Figure 1. This station is part of the Colombian Seismic Network operated by The "Servicio Geológico Colombiano - SGC" (Colombian Geological Service).

The "El Rosal" station employs a Guralp CMG - T3E007 sensor in three components and a Nanometrics RD3-HRD24 digitizer, which provides simultaneous sampling of three channels with 24-bit of sampling rate (Bermudez & Rengifo, 2002). The data correspond to the three component raw waveforms recorded directly in this station and a seismic catalog with 2164 characterized events, selected between January 1st 1998 and October 27th 2008; all of them located less than 120 kilometers from the station.

The Colombian seismic network consists of 42 stations, with an average distance of 162 kilometers between them, which record and transmit seismic data in real time for the entire country as shown in Figure 2.

Before starting the processing related to SVMs, waveform files from "El Rosal" station were converted to the American standard code for information interchange (ASCII) format, using a Seisan package tool; earthquakes with

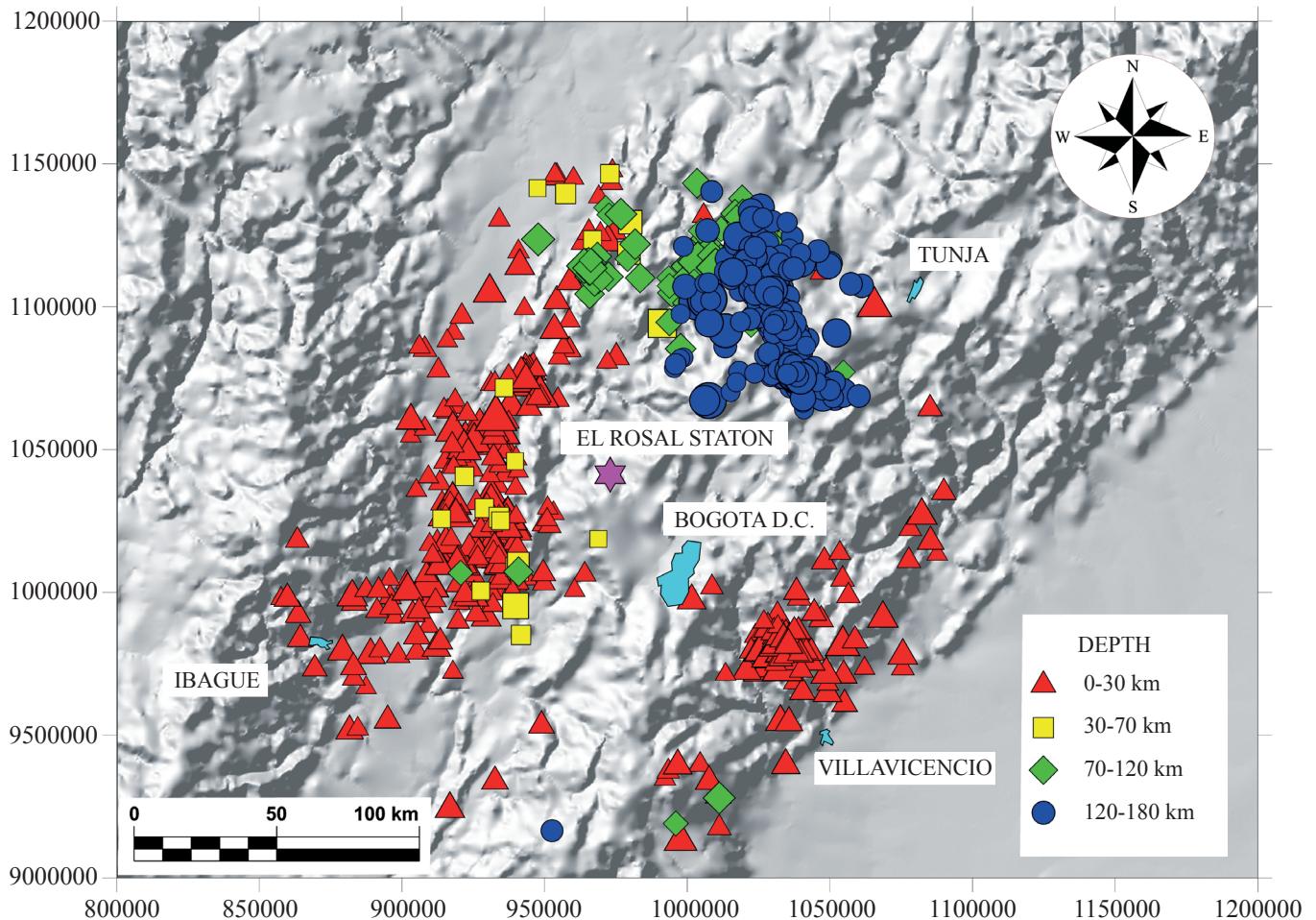


Figure 1. Location of the "El Rosal" seismological station and earthquake distribution around Bogota.

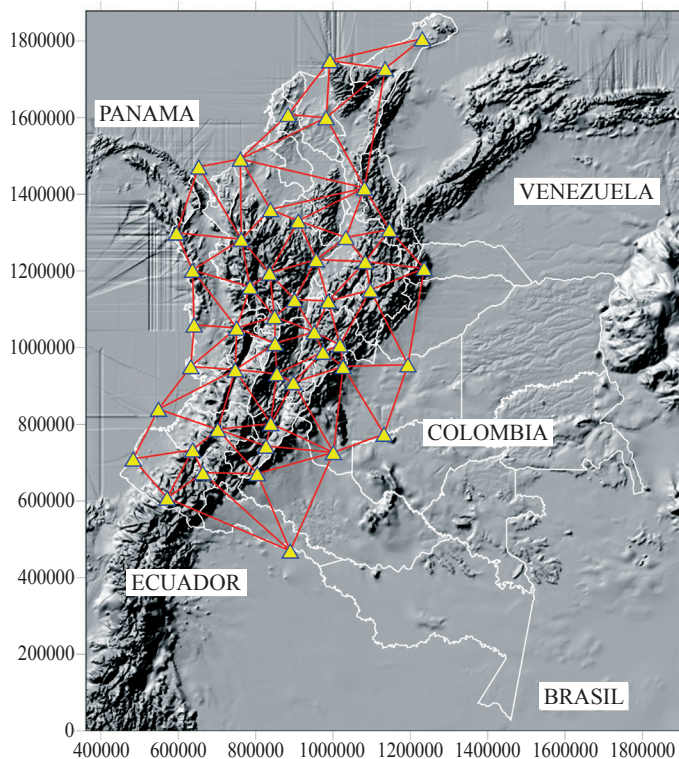


Figure 2. Colombian Seismic Network.

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 implementation of a high-pass filter with cut-off frequency of 0.075 Hz (Wu & Zhao, 2006), while high frequencies were removed with a low-pass filter with cut-off frequency of 150 Hz.

The statistical distribution of azimuth values is presented in Figure 3, where the main distribution of the whole dataset is observed. This histogram shows a high frequency in 240 degrees (azimuth), indicating an active seismogenic zone in that direction. Although this is not a homogeneous distribution, it represents the regular seismic behavior of the area.

Descriptors – Input data set of the SVM

In the first stage, parameters that have been previously used by other authors to earthquake magnitude estimation were calculated and employed as input variables or descriptors for the SVM of this research. In this sense, the relationship between maximum P wave amplitudes and local earthquake magnitudes was considered (Wu & Kanamori, 2005), where a linear regression was performed for each one of the three components of the station. Three parameters were taken from this linear regression which corresponds to the slope (M), an independent term (B) and correlation coefficient (R). The maximum amplitude values (Mx) obtained for each component’s time window were used as descriptors as well. Therefore, each event had 12 descriptors associated with this concept.

In the second place, 9 descriptors used for epicenter distance estimation were added adjusting a linear regression of an exponential function in time (t) by applying the expression “ $B_i \exp(-A_i t)$ ”; this expression belongs to the envelope of the seismic record in logarithmic scale (Odaka et al., 2003) determined also by a linear regression and its respective correlation coefficient (R), for each of component in the seismic station. The correlation coefficient (R) along with the parameters (A) and (B) were calculated for each component; where (B) represents the slope of initial part of P waves and (A) is a parameter related to amplitude variations in time.

Finally, parameters for back-azimuth determination were used to include information about sources location of the seismic events into the model.

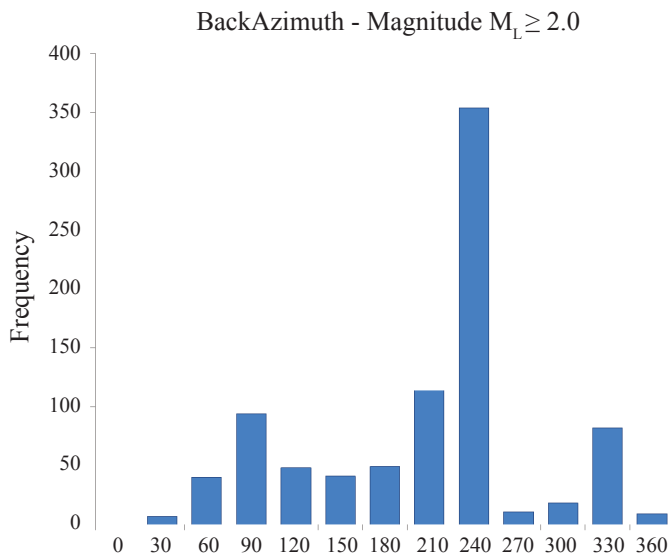


Figure 3. Statistical distribution of earthquake azimuth recorded at “El Rosal” station.

Maximum eigenvalues of two-dimensional covariance matrix were employed as input, calculated as described in Magotra et al., 1987, and Magotra et al., 1989. A windowing scheme with one second time windows was performed to obtain consecutive values for which a linear regression was calculated, also determining the slope (M), the independent term (B), the regression correlation factor (R), and this time with addition of the arithmetic mean of the eigenvalues (P). This last processing works with all components of the station at the same time, thus 4 descriptors were added as input related to this process.

In summary, the SVM of this study employs 25-time signal descriptors as input (Table 1); 12 of them related to works on magnitude calculation, 9 were associated with epicenter distance estimations and the last 4 were used in the back-azimuth determination. These descriptors were calculated for 5, 10 and 15 seconds signal of the 863 selected events.

The SVM Model

A SVM is a supervised classification technique that has its roots in statistical learning theory and has shown promising empirical results in many practical applications, from handwritten digit recognition to text categorization. SVM also works very well with high-dimensional data and avoids the curse of dimensionality problem (Tan et al., 2006). In geosciences e.g., it have been applied in earthquake characterization (Ochoa et al., 2017), automatic recognition of natural fractures (Leal et al., 2016), automatic indicator of lithologies in open hole logs (Leal et al., 2018), among other applications related to pattern recognition. The SVM model of this research was trained with the refined data set for each time window using the Waikato Environment for Knowledge Analysis WEKA 3.6 (Frank et al., 2016) and the 25 descriptors explained before. This algorithm has strong statistical support and can be easily implemented on the station by electronic processing cards.

Table 1. Summary of Descriptors Employ as Input data

Related to earthquake magnitude estimation (Wu & Kanamori, 2005)	Related to epicenter distance estimation (Odaka et al., 2003)	Related to back-azimuth estimation (Magotra et al., 1987; Magotra et al., 1989)
Slope (M) Independent Term (B) Correlation Coefficient (R) Maximum Amplitude (Mx) 12 Descriptors (4 in each component of the station).	Parameter (A) Parameter (B) Correlation Coefficient (R) 9 Descriptors (3 in each component of the station).	Slope (M) Independent Term (B) Correlation Coefficient (R) Maximum Eigenvalues (P) 4 Descriptors (all components work at the same time).

After performing several processing 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-validation process were compared. These processes were carried out testing multiple combinations of exponents and complexity factor for selected earthquake 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. The Pearson’s coefficient in this research is as high as 0.6, showing the relevance of SVM in the estimation of earthquake arrival azimuth; the lower value of this coefficient was 0.096 as shown in Table 4.

Results and discussion

Using the 25 descriptors and real magnitudes for each seismic event, a group of 12 datasets was evaluated (Table 2). Each dataset corresponds to combinations of 4 minimum magnitude filters (2.0, 2.5, 3.0 and 3.5 M_L) and 3 signal length filters (5, 10 and 15 seconds), evaluating combinations of 7 values for kernel exponent ($E = 1.5, 2, 4, 5, 10, 20$ and 50) and 6 values for complexity factor ($C = 1, 3, 5, 10, 20$ and 50); testing 504 models of SVMs in order to find the combination of parameters with the best correlation factor in arrival azimuth determination. Table 2 shows values of correlation coefficients in each combination of cut-off magnitude and time signals where kernel exponents and complexity factors were calculated. According to Table 2, the best correlation coefficient is 0.6 for a time signal of 15 seconds ($Magnitude \geq 2 M_L$).

Table 3 shows statistical summary for the best model of earthquake arrival azimuth in each combination of time signal (15, 10 and 5 seconds) and

magnitude (2, 2.5, 3 and 3.5 M_L). According to skewness values, all distributions present similar shape except samples from 5 seconds and 3.5 M_L (Skewness = -4.86), these samples show a negative skew related to low number of samples for this set (count = 33). The mean arithmetic can be affected by extremely low and high values; therefore, any interpretation from this parameter might produce a wrong perception of the data. High kurtosis values are related to over-fitting in some combinations and lower kurtosis are related to time signals of 10 and 15 seconds with cut-off of 3.5 M_L , which can be interpreted as the arrival azimuth estimation may be reliable for earthquake greater than 3.5 M_L and time signal of 15 seconds; however, the objective of this research is to develop the best model with the lowest base magnitude; in consequence and according to Table 3 the best time signal should be 5 seconds and cut-off magnitude of 3.0 M_L .

The choice of SVM final parameters is shown in Table 4, where Pearson’s correlation coefficient and mean absolute error are presented for each combination of kernel exponent “E” and complexity factor “C”, all of them for the time signal and the cut-off magnitude previously selected. These parameters were calculated using SVM algorithms in WEKA 3.6 (Frank, et al. 2016) with a standard normalized polynomial kernel and 10 fold cross-validations. According Table to 4, the earthquake arrival azimuth can be performed at “El Rosal” station using the SVM based model with a normalized polykernel of exponent 2 and complexity factor of 10 for a time signal of 5 seconds and 3.0 M_L of earthquake cut-off magnitude, showing a standard deviation of 45.4 degrees.

Figure 4 shows the cross-plot with a relationship between the real arrival azimuth (X-axis) and the azimuth calculated by the model (Y-axis), where a normal statistical pattern can be observed in the distribution of residual values (histogram). The dashed blue line represents the linear behavior of predicted data, corresponding to the locus where prediction is equal to real values. This model tends to place the prediction further south to the real location, because of seismogenic zones toward east and west of the station producing more data in those directions and lower among of information from north and south; this is a normal operational condition of “El Rosal” station and therefore, it is a behavior implicit into the model.

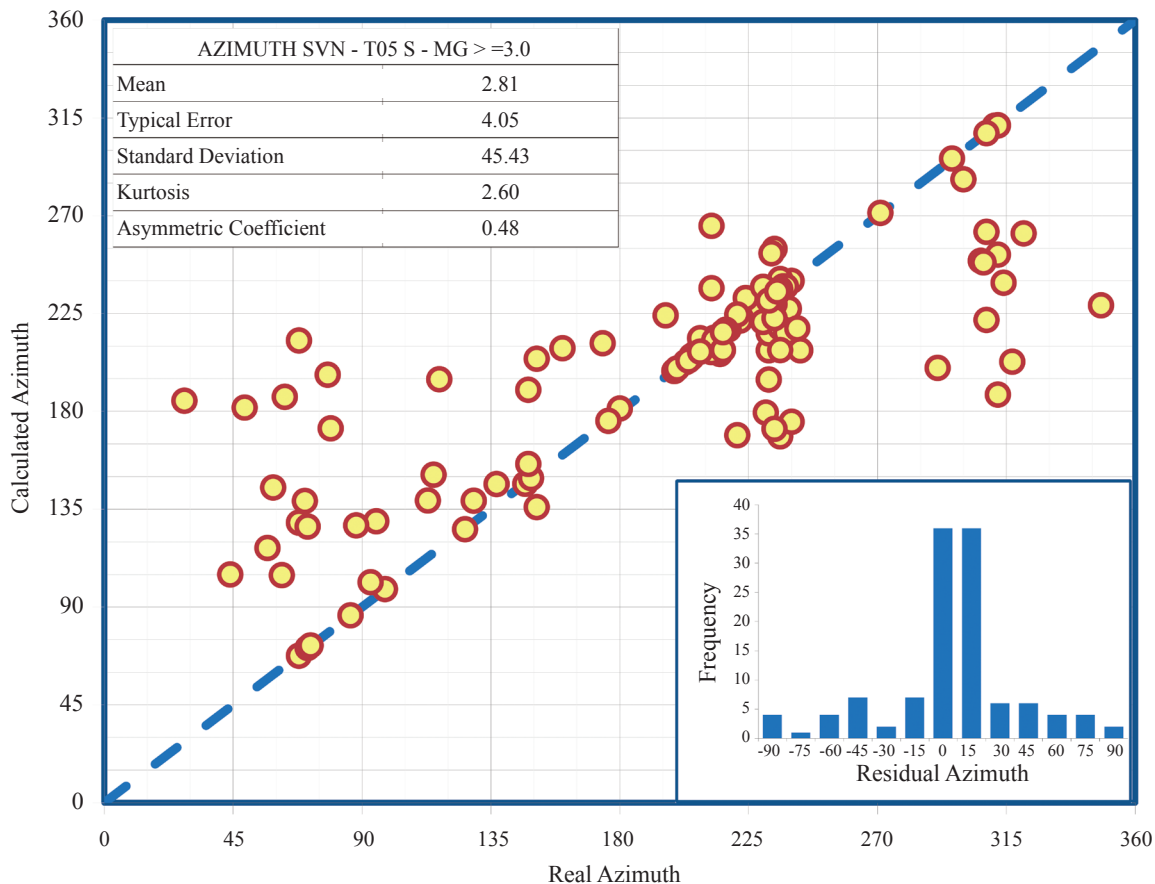


Figure 4. Correlation between real and calculated arrival azimuth with SVM.

Table 2. Cut-off Magnitude and Time Length combination

Pearson's Coefficient for Each Combination of Magnitude and Time Signal																
	Magnitude $\geq 2.0 M_L$ – “E”								Magnitude $\geq 2.5 M_L$ – “E”							
		1.5	2	4	5	10	20	50		1.5	2	4	5	10	20	50
Time Signal = 15 s – “C”	1	0.423	0.442	0.523	0.538	0.58	0.607	0.561	1	0.397	0.415	0.483	0.507	0.559	0.545	0.489
	3	0.45	0.489	0.548	0.561	0.6	0.594	0.51	3	0.447	0.466	0.523	0.543	0.578	0.518	0.443
	5	0.47	0.511	0.557	0.574	0.608	0.575	0.479	5	0.462	0.498	0.536	0.569	0.552	0.507	0.412
	10	0.496	0.525	0.575	0.588	0.603	0.54	0.437	10	0.485	0.502	0.572	0.581	0.508	0.464	0.302
	20	0.516	0.537	0.587	0.59	0.574	0.495	0.403	20	0.496	0.5	0.575	0.57	0.476	0.437	0.387
	50	0.535	0.555	0.582	0.582	0.513	0.419	0.373	50	0.49	0.531	0.553	0.525	0.439	0.378	0.387
Time Signal = 10 s – “C”	1	0.429	0.454	0.506	0.521	0.553	0.578	0.565	1	0.427	0.449	0.494	0.519	0.535	0.522	0.474
	3	0.457	0.483	0.534	0.544	0.57	0.587	0.512	3	0.456	0.47	0.432	0.539	0.525	0.489	0.419
	5	0.488	0.498	0.542	0.552	0.579	0.575	0.484	5	0.467	0.488	0.534	0.538	0.514	0.466	0.411
	10	0.485	0.521	0.554	0.56	0.587	0.546	0.459	10	0.481	0.491	0.536	0.535	0.489	0.425	0.408
	20	0.51	0.536	0.561	0.569	0.577	0.5	0.45	20	0.477	0.499	0.529	0.504	0.456	0.382	0.408
	50	0.535	0.556	0.566	0.573	0.536	0.439	0.442	50	0.479	0.513	0.481	0.471	0.394	0.351	0.408
Time Signal = 05 s – “C”	1	0.379	0.399	0.442	0.449	0.467	0.471	0.412	1	0.312	0.354	0.419	0.442	0.484	0.467	0.416
	3	0.428	0.439	0.457	0.463	0.473	0.469	0.365	3	0.41	0.419	0.458	0.475	0.491	0.469	0.373
	5	0.431	0.439	0.465	0.471	0.485	0.455	0.342	5	0.406	0.422	0.467	0.479	0.49	0.447	0.351
	10	0.434	0.447	0.466	0.464	0.474	0.726	0.328	10	0.414	0.432	0.473	0.481	0.481	0.416	0.33
	20	0.445	0.462	0.449	0.45	0.45	0.397	0.305	20	0.422	0.451	0.467	0.47	0.46	0.38	0.305
	50	0.462	0.444	0.424	0.435	0.419	0.329	0.285	50	0.45	0.44	0.44	0.456	0.41	0.334	0.265
Time Signal = 15 s – “C”	1	0.431	0.43	0.422	0.421	0.386	0.37	0.341	1	0.283	0.327	0.366	0.361	0.295	0.237	0.135
	3	0.433	0.412	0.405	0.411	0.426	0.362	0.34	3	0.364	0.419	0.475	0.37	0.348	0.241	0.135
	5	0.416	0.407	0.417	0.427	0.405	0.368	0.348	5	0.401	0.454	0.392	0.383	0.351	0.241	0.135
	10	0.415	0.413	0.441	0.457	0.382	0.388	0.34	10	0.456	0.404	0.357	0.341	0.351	0.241	0.135
	20	0.426	0.42	0.461	0.437	0.387	0.368	0.34	20	0.412	0.397	0.325	0.34	0.351	0.241	0.135
	50	0.419	0.431	0.431	0.398	0.388	0.368	0.34	50	0.379	0.323	0.325	0.34	0.351	0.241	0.135
Time Signal = 10 s – “C”	1	0.388	0.412	0.41	0.428	0.458	0.491	0.334	1	0.305	0.339	0.375	0.37	0.313	0.318	0.121
	3	0.411	0.408	0.427	0.442	0.532	0.41	0.333	3	0.324	0.33	0.317	0.334	0.326	0.32	0.121
	5	0.409	0.422	0.45	0.487	0.509	0.382	0.333	5	0.299	0.29	0.346	0.366	0.327	0.32	0.121
	10	0.42	0.409	0.486	0.513	0.435	0.382	0.333	10	0.247	0.274	0.382	0.346	0.327	0.32	0.121
	20	0.403	0.411	0.494	0.496	0.374	0.382	0.333	20	0.245	0.329	0.357	0.347	0.327	0.32	0.121
	50	0.393	0.433	0.467	0.425	0.352	0.382	0.333	50	0.318	0.357	0.357	0.347	0.327	0.32	0.121
Time Signal = 05 s – “C”	1	0.336	0.407	0.47	0.46	0.456	0.343	0.182	1	0.369	0.427	0.52	0.534	0.518	0.267	-0.34
	3	0.549	0.542	0.524	0.496	0.387	0.312	0.183	3	0.523	0.551	0.586	0.599	0.468	0.168	-0.34
	5	0.571	0.561	0.494	0.462	0.346	0.305	0.183	5	0.556	0.563	0.609	0.609	0.413	0.168	-0.34
	10	0.572	0.588	0.447	0.406	0.326	0.29	0.183	10	0.556	0.572	0.6	0.546	0.418	0.168	-0.34
	20	0.501	0.531	0.369	0.308	0.321	0.298	0.183	20	0.552	0.567	0.532	0.525	0.418	0.168	-0.34
	50	0.538	0.436	0.249	0.251	0.325	0.298	0.183	50	0.552	0.552	0.536	0.525	0.418	0.168	-0.34

Table 3. Summary for the best model of earthquake arrival azimuth in each combination

SUMMARY IN DETERMINATION OF EARTHQUAKE ARRIVAL AZIMUTH - RESIDUAL												
	Signal Time = 15 s				Signal Time = 10 s				Signal Time = 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
Skewness	0.64	1.11	1.16	0.07	1.27	1.10	1.62	0.43	0.78	1.29	0.48	-4.86
Kurtosis	4.14	4.89	7.68	1.58	6.83	3.26	16.24	2.39	1.93	4.09	2.60	27.14
Mean	2.62	2.69	-0.14	4.71	3.96	5.74	-0.09	-3.10	6.04	5.79	2.81	-1.94
Standard Deviation	46.3	40.7	31.8	35.5	37.8	52.0	27.6	34.8	53.1	46.4	45.4	17.6
Standard error	1.58	2.09	2.84	6.19	1.29	2.67	2.46	6.69	1.81	2.39	4.05	3.06
Count	863	379	126	33	863	379	126	33	864	377	126	33
Level of Confidence (95%)	3.093	4.109	5.614	12.605	2.528	5.252	4.863	13.628	3.545	4.702	8.010	6.231

Table 4. Pearson's Coefficient and Mean Absolute Error used to Determent "E" and "C"

Pearson's Coefficient		Exponent "E"									
		0.5	0.8	1.1	1.5	2	4	5	10	20	50
Complexity Factor "C"	0.1	0.096	0.12	0.125	0.137	0.159	0.209	0.23	0.272	0.28	0.187
	0.5	0.165	0.192	0.226	0.264	0.303	0.387	0.401	0.414	0.386	0.195
	0.8	0.192	0.234	0.267	0.307	0.366	0.45	0.451	0.452	0.355	0.182
	1	0.207	0.25	0.276	0.339	0.409	0.47	0.459	0.458	0.347	0.182
	2	0.225	0.287	0.365	0.472	0.522	0.51	0.506	0.417	0.327	0.172
	4	0.285	0.406	0.529	0.57	0.55	0.51	0.475	0.361	0.31	0.183
	5	0.324	0.454	0.557	0.571	0.56	0.494	0.462	0.346	0.304	0.183
	10	0.409	0.498	0.555	0.573	0.588	0.447	0.406	0.326	0.29	0.183
	20	0.438	0.507	0.553	0.6	0.532	0.368	0.308	0.321	0.298	0.183
50	0.38	0.398	0.583	0.538	0.436	0.25	0.252	0.352	0.298	0.183	
Mean Absolute Error		Exponent "E"									
		0.5	0.8	1.1	1.5	2	4	5	10	20	50
Complexity Factor "C"	0.1	53.1	52.7	52.5	52.4	52.2	51.9	51.8	51.3	51.1	52.3
	0.5	52.3	52.6	52.6	52.4	51.9	50.5	50.2	50.4	51.7	56.4
	0.8	52.6	52.7	52.5	52	51.1	49.2	49.4	49.9	54	57.7
	1	52.7	52.6	52.5	51.6	50.3	48.8	49.3	49.7	55.2	57.9
	2	53.1	52.9	51.5	48.8	47.5	48.3	48.3	53.2	58.2	58.5
	4	53.7	51.6	47.8	46.2	47.4	48.5	50.3	60.4	60.3	58.2
	5	53.9	50.6	46.8	46.4	47.3	49.6	51.9	63	60.9	58.2
	10	54.6	50.3	47	46.4	45.6	55.9	59.9	67.4	62.7	58.2
	20	56.7	52.5	47.1	45	49.9	65.8	73.2	69	62.8	58.2
50	81.3	65.4	45.1	51	61.7	85.2	85	73.1	62.8	58.2	

Conclusions & Recommendations

This model is proposed and evaluated for fast earthquake arrival azimuth determination, based on support vector machine regression through pattern recognition and characterization of earthquake signals recorded on a three components seismic station in only 5 seconds, anticipating the arrival of earthquakes in the city of Bogota. Additionally, this model can be implemented directly in the electronic devices of a seismological station, where the main mathematical process corresponds to a simple matrix product, a kernel function of exponent 2 and complexity factor of 10 with an earthquake of 3.0 M_L .

The accuracy reported in this research is lower compared with results reported by Lockman & Allen, 2005 and Eisermann et al., 2015; however it must be considered that these authors work with data from several seismological stations and not with a single station as this research does. Nevertheless, the result of 45.4 degrees in earthquake arrival azimuth estimation showed in this research is an improvement on that of Noda et al., 2012, who had standard deviation between 49.0 and 67.9 degrees working also with a single station.

The implementation of additional input variables such as predominant period, Fourier and wavelet frequency spectra should be considered in order to obtain higher correlation factors. Furthermore, the use of an updated dataset is recommended, adding information from October 27th 2008 to present; this new data along with additional input variables might improve the model performance and reaching better estimation of earthquake arrival azimuth.

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.

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

- Anant, K. S. & Dowla, F. U. (1997). Wavelet transform methods for phase identification in three-component seismograms. *Bulletin of the Seismological Society of America*, 87(6), 1598-1612. <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.871.8399&rep=rep1&type=pdf>
- Bermúdez, M. L. & Rengifo, F. (2002). *EL ROSAL: La Estación Sismológica del CTBTO en Colombia*. Bogota, Primer Simposio Colombiano de Sismología, p. 8.
- Eisermann, A. S., Ziv, A. & Wust-Bloch, G. H. (2015). Real-Time Back Azimuth for Earthquake Early Warning. *Bulletin of the Seismological Society of America*, 105(4), 2274-2285. <https://doi.org/10.1785/0120140298>
- Espinosa, J. M. (1995). Mexico City seismic alert system. *Seismological Research Letters*, 66(6), 42-53. <https://doi.org/10.1785/gssrl.66.6.42>
- Frank, E., Hall, M. & Witten, I. (2016). *The WEKA Workbench. Online Appendix for "Data Mining: Practical Machine Learning Tools and Techniques"*. Morgan Kaufmann, Fourth Edition, pp. 218-223.
- Horiuchi, S. (2005). An automatic processing system for broadcasting earthquake alarms. *Bulletin of the Seismological Society of America*, 95(2), 708-718. <https://doi.org/10.1785/0120030133>
- Leal, J., Ochoa, L. & García, J. (2016). Identification of natural fractures using resistive image logs, fractal dimension and support vector machines. *Ingeniería e Investigación*, 36(3), 125-132. <https://doi.org/10.15446/ing.investig.v36n3.56198>
- Leal, J., Ochoa, L. & Contreras, C. (2018). Automatic Identification of Calcareous Lithologies Using Support Vector Machines, Borehole Logs and Fractal Dimension of Borehole Electrical Imaging. *Earth Science Research Journal*, 22(1), 7-12. <https://doi.org/10.15446/esrj.v22n1.68320>
- Lockman, A. B. & Allen, R. M. (2005). Single-Station Earthquake Characterization for Early Warning. *Bulletin of the Seismological Society of America*, 95(6), 2029-2039. <https://doi.org/10.1785/0120040241>
- Magotra, N., Ahmed, N. & Chael, E. (1987). Seismic event detection and source location using single station (three components) data. *Bulletin of the Seismological Society of America*, 77(3), 958-971.
- Magotra, N., Ahmed, N. & Chael, E. (1989). Single-station seismic event detection and location. *IEEE Transactions on Geoscience and Remote Sensing*, 27(1), 15-23. DOI: 10.1109/36.20270
- Noda, S. (2012). Improvement of back-azimuth estimation in real-time by using a single station record. *Earth, Planets and Space*, 64(3), 305-308. <https://link.springer.com/article/10.5047/eps.2011.10.005>
- Ochoa, L. H., Niño, L. F. & Vargas, C. A. (2014). Severity Classification of a Seismic Event based on the Magnitude-Distance Ratio Using Only One Seismological Station. *Earth Sciences Research Journal*, 18(2), 115-122. <https://doi.org/10.15446/esrj.v18n2.41083>
- Ochoa, L. H., Niño, L. F. & Vargas, C. A. (2017). Fast magnitude determination using a single seismological station record implementing machine learning techniques. *Sciences Direct, Geodesy and Geodynamic*, 1-8. <https://doi.org/10.1016/j.geog.2017.03.010>
- Ochoa, L. H., Niño, L. F. & Vargas, C. A. (2018). Fast estimation of earthquake epicenter distance using a single seismological station with machine learning techniques. *DYNA*, 85(204), 161-168. <https://doi.org/10.15446/dyna.v85n204.68408>
- Odaka, T. (2003). A new method for quickly estimating epicentral distance and magnitude from a single seismic record. *Bulletin of the Seismological Society of America*, 93(1), 526-532. DOI: 10.1785/0120020008
- Ojeda, A., Martínez, S., Bermúdez, M. & Atakan, K. (2002). The new accelerograph network for Santa Fe De Bogota, Colombia. *Soil Dynamics and Earthquake engineering*, 22(9-12), 791-797. [https://doi.org/10.1016/S0267-7261\(02\)00100-8](https://doi.org/10.1016/S0267-7261(02)00100-8)
- Saita, J. & Nakamura, Y. (2003). *The early warning systems for mitigation of disasters caused by earthquakes and tsunamis*. In: J. Zschau & A. Koppers (eds). Early Warning Systems for Natural Disaster Reduction. Berlin: Springer-Verlag, pp. 453-460. https://doi.org/10.1007/978-3-642-55903-7_58
- Reymond, D., Hyvernaud, O. & Talandier, J. (1991). Automatic detection, location and quantification of earthquakes. *Pure and Applied Geophysics*, 135(3), pp. 361-382. <https://doi.org/10.1007/BF00879470>
- Roberts, R. G., Christoffersson, A. & Cassidy, F. (1989). Real-time detection, phase identification and source location estimation using single station three component seismic data. *Geophysical Journal*, 97(3), 471-480. <https://doi.org/10.1111/j.1365-246X.1989.tb00517.x>
- Talandier, J., Reymond, D. & Oka, E. A. (1987). Use of variable mantle magnitude for the rapid one-station estimation of teleseismic moments. *Geophysical Research Letters*, 14(8), 840-843. <https://doi.org/10.1029/GL014i008p00840>
- Wu, Y. M. & Zhao, L. (2006). Magnitude estimation using the first three seconds P-wave amplitude in earthquake early warning. *Geophysics Research Letter*, 33(16), L16312. <https://doi.org/10.1029/2006GL026871>
- Wu, Y. M. & Kanamori, H. (2005). Rapid assessment of damage potential of earthquakes in Taiwan from beginning of P waves. *Bulletin of the Seismological Society of America*, 93(1), 526-532. <https://doi.org/10.1785/0120040193>
- Wu, Y. M., Shin, T. C. & Tsai, Y. B. (1998). Quick and reliable determination of magnitude for seismic early warning. *Bulletin of the Seismological Society of America*, 88(5), 1254-1259. <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.462.985&rep=rep1&type=pdf>