

Reconocimiento de patrones espaciales sísmicos en el suroccidente colombiano

Hernán Darío Benítez Restrepo

17 de diciembre de 2010

0.1. Resumen

En el catálogo de sismos del occidente colombiano se recopilan los eventos sísmicos ocurridos en la región. La presentación de terremotos en mapas regionales permite una identificación visual no automatizada (humana) de patrones espaciales. Sin embargo, este enfoque es limitado dada la cantidad de datos que se deben analizar y la valoración subjetiva (no automatizada) de los mismos. De esta problemática surge la pregunta: ¿Cómo reconocer patrones espaciales sísmicos y evaluar de manera no-subjetiva el catálogo de sismos del sur occidente colombiano?. El objetivo de la investigación es establecer una metodología de reconocimiento y clasificación de patrones espaciales sísmicos por análisis cluster y validar los resultados obtenidos con respecto a las provincias sismotectónicas modeladas en las evaluaciones de amenaza sísmica del sur occidente colombiano. El análisis cluster es un instrumento eficiente para extraer información y pautas ocultas en una cantidad grande de datos sismológicos ya que este tipo de análisis agrupa observaciones de modo que aquellas que pertenecen a un cluster son más similares entre sí que observaciones que pertenecen a otro cluster. Para alcanzar los objetivos se hará un análisis exploratorio de datos espaciales, se usará análisis cluster y se validará con el criterio de un sismólogo experto para crear una metodología de análisis de datos sísmicos espaciales que sirva como referencia a otras investigaciones y soporte evaluaciones de amenaza sísmica en el sur occidente colombiano. Se espera obtener un catálogo de sismos depurado, un análisis cluster que entregue provincias sismotectónicas y un manual que contenga la metodología establecida para el preprocesamiento y análisis de datos espaciales sísmicos.

0.2. Planteamiento del problema

El sur occidente colombiano está sujeto a movimientos tectónicos debido al choque de las placas Nazca, Suramérica y Caribe, por ello está fuertemente amenazado por fallas sismogénicas que en su mayor parte tienen comportamientos hasta ahora desconocidos. Es así como Cali se ha visto afectada por diversos sismos a lo largo de su historia, como los ocurridos en marzo de 1566, julio 9 de 1766, mayo 15 de 1885, enero 31 de 1906, enero 19 de 1958, julio 30 de 1962, febrero 9 de 1967, noviembre 23 de 1979, noviembre 19 de 1991, febrero 8 de 1995 y noviembre 15 de 2004. La amenaza a la que está sometido el sur occidente colombiano se cuantifica por el valor esperado de futuras acciones sísmicas y se expresa en términos de sus probabilidades de ocurrencia. Esta cuantificación se puede mejorar si se determinan provincias sismotectónicas o regiones con características sismológicas similares. La determinación de las provincias sismotectónicas es fundamental para obtener la amenaza sísmica que junto con la vulnerabilidad de las personas y las estructuras civiles permiten estimar el riesgo al que están sometidas la población y la infraestructura económica.

La delineación de las provincias sismotectónicas se hace comúnmente de forma no cuantitativa basada en información geológica y neotectónica apoyada en el catálogo de sismos, que es el registro en superficie de la dinámica de las provincias, donde se presentan los eventos sísmicos en una región. Tradicionalmente el análisis espacial

de dichas provincias inicia con un modelo tectónico *a priori* basado en la inspección visual no automatizada de la distribución espacial de los sismos en superficie usando procedimientos gráficos ligados en general a un mapa.

Los mapas geológicos y de fallas sismogénicas del Suroccidente Colombiano [1], [2], [3] son insumos muy valiosos en la investigación de amenazas y de riesgo sísmico. Otra fuente muy importante de información sismológica en el sur occidente colombiano es el catálogo de sismos del occidente colombiano CATOCC [4] donde se recopilan los eventos ocurridos en la región organizados por localización espacial (epicentros), profundidad, fecha de ocurrencia y magnitud. Estos datos de terremotos expuestos en mapas regionales permiten una identificación visual no automatizada (humana) de patrones espaciales. Sin embargo, este enfoque es limitado dada la cantidad de datos que se deben analizar y la valoración subjetiva (no automatizada) de los mismos. Por estas razones se hace necesario usar herramientas de minería de datos que permitan reconocer patrones espaciales para soportar la determinación de las provincias sismotectónicas de la región. De esta problemática surge la pregunta: ¿Cómo reconocer patrones espaciales sísmicos y evaluar de manera no-subjetiva el catálogo de sismos del suroccidente colombiano?.

Un catálogo de sismos depurado y agrupado por provincias sismo tectónicas sería el insumo principal para el cálculo de la amenaza que a su vez permita tomar decisiones de orden gubernamental en prevención y atención de desastres, hacer cálculos de amenazas, determinar el riesgo sísmico y proponer planes de ordenamiento territorial.

0.3. Situación actual del conocimiento en el área de la investigación

0.3.1. Minería de datos

La minería de datos (data mining) es la búsqueda de conocimiento a partir de bases de datos extensas y se ocupa de la extracción no trivial de información implícita anteriormente desconocida y potencialmente útil. El papel de la minería de datos en el análisis sismológico es cada vez más destacado debido a que los avances en las tecnologías de captura de datos proporcionan un número inmenso de estos. Además, permiten encontrar a través de métodos cuantitativos estadísticos, patrones que no pueden verse con la sola inspección visual y a partir de estos patrones ocultos predecir distribuciones espaciales nuevas. El análisis cluster es un instrumento eficiente para extraer información y pautas ocultas en una cantidad grande de datos sismológicos [5] ya que este tipo de análisis agrupa observaciones de modo que aquellas que pertenecen a un cluster son más similares entre sí que observaciones que pertenecen a otro [6], [7].

0.3.2. Análisis Estadístico Exploratorio de Datos

Como etapa previa al análisis cluster es fundamental hacer un análisis exploratorio de los datos que ayude a definir respecto a los datos el tipo de distribuciones de frecuencia no normales; las relaciones no lineales entre las variables, las observaciones espacialmente autocorreladas; las estructuras de las variaciones espaciales según grado de precisión, ruido y niveles de error; las variables de interés y las relaciones locales entre los datos [8]. En el análisis estadístico exploratorio de una serie de datos una de las primeras etapas es la descripción de la distribución y el cálculo de estadísticas básicas, de manera que se pueda verificar si ellos cumplen con las condiciones de base para la aplicación de un determinado modelo estadístico orientado a verificar los supuestos asociados a una distribución normal, para luego proceder al cálculo de estadísticas descriptivas como media, varianza, desviación estándar, coeficiente de variación, cuantiles, mínimo y máximo entre otras. Posteriormente, es común ajustar un modelo y estimar valores para toda la población en estudio tomando como hipótesis de partida que los datos tiene una distribución estadística normal. Sin embargo, cuando se analizan datos espaciales se debe usar análisis exploratorio de datos que comúnmente está dividido en dos partes: a) un análisis univariado y bivariado, usando la estadística tradicional, análisis exploratorio de datos (AED) y b) un análisis exploratorio de datos espaciales (AEDE) de las relaciones espaciales utilizando geoestadística entre otros [9]. El problema fundamental en el análisis de datos espaciales es la presencia de auto-correlación espacial que genera parámetros estadísticos sesgados debido a la dependencia espacial de los datos. Por otra parte, en el análisis de datos espaciales la variabilidad de los datos cambia de acuerdo con la escala. La variación surge de tres fuentes fundamentales: errores estocásticos, errores en las medidas y errores en el muestreo. En el contexto de las bases de datos espaciales es urgente utilizar herramientas exploratorias de los datos para encontrar patrones y otros eventos interesantes que tengan algún grado de precisión y permitan definir **donde** mirar en el espacio **cuando** mirar en tiempo y **que** ver con respecto a los atributos.

0.3.3. Clustering de datos sismológicos

La causa principal de los terremotos es la ruptura de la corteza de la tierra llamada falla. Si varios eventos sísmicos se concentran a lo largo de un segmento de línea indica la existencia de una falla [10]. Por otra parte, muchos eventos sísmicos dispersos en una región indican un sistema de fallas que componen las provincias sismotectónicas. La identificación y clasificación de las provincias sismotectónicas con características similares en una región de interés es el fundamento de los estudios de amenaza sísmica; lo cual se realiza mediante la zonificación tectónica a partir de análisis no cuantitativos y subjetivos o por inspección visual de los catálogos de sismos que recogen datos sobre la distribución regional de sismos [11]. El reconocimiento de patrones de manera no subjetiva de datos sísmicos históricos e instrumentales proporciona resultados más robustos y es una herramienta más conveniente para extraer conocimiento útil de un conjunto de datos. Como trabajo previo se puede citar un análisis espacio temporal con transformada wavelet del catálogo de sismos instrumental de los Alpes Occidentales [12] encontraron que los patrones

se ajustaban muy bien a los elementos geológicos/tectónicos y por tanto las zonas sismogénicas se pueden delimitar mejor con este método que por la simple observación de un mapa de eventos sísmicos. En Irán se utilizó un algoritmo de clustering difuso no supervisado para detectar patrones espaciales en el catálogo de sismos histórico e instrumental lo cual reveló patrones ocultos que permiten clasificar de manera diferente los epicentros espacialmente distribuidos de los terremotos [13]. La comparación entre los resultados de este análisis y las provincias sismotectónicas de Irán mostraron que las provincias encontradas confirmaron las provincias conocidas y evidenciaron rasgos nuevos para la interpretación sismológica y por consiguiente para la evaluación de la amenaza sísmica. Finalmente, es importante resaltar que el análisis exploratorio de datos espaciales no se reporta como un paso seguido durante la investigación en [12] y [13] mientras que en esta propuesta se presenta como el primer paso a seguir dada la naturaleza espacial de los datos y el objetivo de establecer una metodología de análisis de datos espaciales.

Con base en el planteamiento del problema y el estado del arte de la investigación se plantea la hipótesis: el reconocimiento de patrones espaciales ayudará a encontrar nuevas provincias sismotectónicas y a confirmar las modeladas en las evaluaciones de amenaza sísmica del suroccidente colombiano.

0.4. Objetivo de la investigación

0.4.1. Objetivo general

Establecer una metodología de reconocimiento y clasificación de patrones espaciales sísmicos y validar los resultados obtenidos con respecto a las provincias sismotectónicas modeladas en las evaluaciones de amenaza sísmica del sur occidente colombiano.

0.4.2. Objetivos específicos

- Aplicar el AED (Análisis Exploratorio de Datos) y el AEDE (Análisis Exploratorio de Datos Espaciales) al catálogo de sismos del sur occidente colombiano para obtener una versión depurada.
- Analizar y validar mediante aprendizaje no supervisado (clustering) los datos del catálogo depurado de sismos del sur occidente colombiano para obtener patrones espaciales que correspondan a provincias sismotectónicas modeladas en las evaluaciones de amenaza sísmica del sur occidente colombiano o a provincias desconocidas.
- Validar las provincias sismotectónicas obtenidas en el sur occidente colombiano mediante el análisis clustering basándose en el criterio de un experto (sismólogo) para ajustar los parámetros del método clustering obtenido.
- Documentar los pasos seguidos para el análisis exploratorio de los datos y la determinación por clustering de las provincias sismotectónicas utilizando e

identificando aquellos que tuvieron los mejores resultados para establecer la metodología.

0.5. Metodología propuesta

En la metodología se muestran los pasos a seguir para alcanzar cada objetivo específico, la escogencia del tamaño de la muestra, el catálogo de datos y los pasos a seguir para procesar, analizar y presentar los datos.

1. Aplicar el AED (Análisis Exploratorio de Datos) y el AEDE (Análisis Exploratorio de Datos Espaciales) al catálogo de sismos del sur occidente colombiano

- **Establecimiento del tamaño de muestra**

El catálogo de sismos del sur occidente colombiano que se va a utilizar es una ventana del CATOCC entre 4.5° y 5.5° Norte y 75° a 80° Oeste que comprende el un total de 4151 eventos históricos e instrumentales. Cada uno de los eventos tiene como atributo epicentros (latitud, longitud), magnitud, profundidad, tiempo. Inicialmente sólo se encontrarán los cluster en el espacio de los epicentros y una vez estos se definan podrán cruzarse con cluster en el tiempo para separar los eventos y luego comparar.

- **Análisis exploratorio de los datos**

- Examinar la distribución de los datos.
- Detectar valores extremos locales y globales.
- Buscar tendencias globales.
- Examinar la auto-correlación espacial y variación direccional de los datos.
- Entender la co-variación entre múltiples variables.
- Realizar transformaciones de los datos si son requeridas.
- Realizar un análisis geoestadístico de los datos.

- **Análisis exploratorio espacial de los datos**

Realizar un análisis exploratorio de datos espacial es describir el comportamiento o estructura de un atributo sobre el espacio, para lo cual se determina el grado de dependencia espacial que presentan las observaciones en relación con sus vecinos y la continuidad espacial o rugosidad de la superficie. Los estimadores mayormente utilizados son las gráficas h-scatterplots, la covarianza, semivarianza y autocorrelación espacial. Una diferencia con respecto al análisis univariado y bivariado es que estos indicadores pueden ser ajustados de diferentes maneras sobre el espacio (isotrópico y anisotrópico), cambiando las distancias y ángulos sobre los cuales se hace el análisis. Finalmente, por medio del análisis exploratorio es posible detectar problemas de representatividad en las muestras, es decir, sectores con menor densidad de puntos respecto a otros, valores

anómalos o existencia de grupos o subpoblaciones y tendencias en los datos.

2. Analizar y validar mediante aprendizaje no supervisado (clustering) los datos del catálogo depurado de sismos del suroccidente colombiano para obtener patrones espaciales que correspondan a provincias sismotectónicas modeladas en las evaluaciones de amenaza sísmica del sur occidente colombiano o a provincias desconocidas.
 - En la etapa de clasificación de los datos se usarán las técnicas de clustering por jerarquía y particionamiento. Para este último se implementarán las técnicas de fuzzy clustering y hard clustering con el fin de escoger la que menor error de clasificación entregue. Para validar las técnicas de clustering se usará validación cruzada dividiendo los datos en conjunto de validación cruzada y prueba (70 % y 30 %, respectivamente) dividiendo en $k=10$ el conjunto de validación.
 - Una vez se determinen los cluster con los epicentros de los sismos se hará un estudio a la luz de otras características como intensidad y profundidad del sismo con el fin de conocer el efecto que estas variables tienen sobre los clusters encontrados. Además, se hará un análisis temporal de los cluster con el fin de saber cuales datos equivalen a un sólo evento sísmico.
3. Validar las provincias sismotectónicas obtenidas en el sur occidente colombiano mediante el análisis clustering basándose en el criterio de un experto (sismólogo) para ajustar los parámetros del método clustering obtenido.
 - Los clusters encontrados serán visualizados en una mapa del sur occidente colombiano en donde se confrontarán con aquellas fallas encontradas en la zona para determinar si el enfoque usado descubrió provincias sismotectónicas desconocidas hasta el momento y para confirmar las existentes.
 - Una vez se encuentren los agrupamientos de datos que indican las fallas estos serán confrontados con el criterio de sismólogos para establecer los lineamientos de las provincias sismotectónicas y hacer ajustes sobre los parámetros del análisis clustering.
4. Documentar los pasos seguidos para el análisis exploratorio de los datos y la determinación por clustering de las provincias sismotectónicas utilizando e identificando aquellos que tuvieron los mejores resultados para establecer la metodología.
 - Se creará un manual completo sobre los pasos seguidos y resultados obtenidos con el fin de establecer una metodología que pueda servir de referencia a otros investigadores que analicen datos espaciales sísmicos.

5. **Problemas que se pueden presentar:** La base de datos ha sido construida con la unión de diferentes tipos de datos adquiridos con diferentes instrumentos a lo largo de varios años. El catálogo existente no dispone de información sobre los instrumentos utilizados para la captura de registros sísmológicos lo que puede significar la presencia de datos atípicos. Por esta razón el análisis exploratorio de datos se hará una depuración removiendo los datos atípicos para homogenizar la base del catálogo.

0.6. Cronograma

Ver anexo. En el cronograma presentado en el anexo se plantean las siguientes tareas:

- T0 Aplicar análisis exploratorio de datos y análisis exploratorio de datos espaciales
 1. T0.1 Recolección y análisis de la información relacionada con análisis exploratorio de datos sísmicos y análisis exploratorio de datos espaciales.
 2. T0.2 Depuración de la base de datos mediante análisis mediante representación gráfica y remoción de datos atípicos a través del análisis exploratorio de datos y análisis de la autocorrelación espacial de los mismos.
 3. T0.3 Documentación de los resultados obtenidos
- T1 Clasificación de patrones por aprendizaje no supervisado (clustering)
 1. T1.1 Aplicar agrupamiento de datos (clustering) a la base de datos depurada por particionamiento y jerarquía.
 2. T1.2 Validar los cluster generados mediante validación cruzada y con un conjunto de prueba
 3. T1.3 Documentar los resultados obtenidos
- T2 Validación con sismólogos de los agrupamientos encontrados
 1. T2.1 Confrontar los agrupamientos encontrados contra los lineamientos de las fallas obtenidas por sismólogos
 2. T2.2 Determinar posibles nuevos patrones espaciales encontrados con el clustering o confirmar los ya conocidos
 3. T2.3 Documentar los resultados obtenidos
- T3 Elaboración de un manual con la metodología propuesta
 1. T3.1 Consolidar la documentación obtenida en las tareas T0, T1 y T2 para escribir un manual que contenga la metodología propuesta para el análisis de datos espaciales sísmicos.

0.7. Otros

0.7.1. Impacto ambiental del proyecto

El proyecto Reconocimiento de patrones espaciales sísmicos en el sur occidente colombiano se encuentra en el área de tecnologías de la información y las comunicaciones, específicamente en la minería de datos, la teoría de la información y el reconocimiento de patrones. Estas tecnologías son consideradas tecnologías limpias por lo tanto no hace uso de material biológico que ponga en riesgo la salud ni genera procesos lesivos para el medio ambiente a corto, mediano o largo plazo haciendo nulo su impacto ambiental.

0.7.2. Pertinencia Social

Al aplicar el Reconocimiento de Patrones para correlacionar sismos del sur occidente colombiano se está generando nuevo conocimiento, pertinente para una sociedad que demanda la identificación de todas las posibles amenazas sísmicas existentes en la región, y útil para las entidades de gobierno que deben planear y ejecutar planes de ordenamiento del territorio, planes de emergencia y atención de desastres. El insumo básico para la toma de decisiones sobre planes de ordenamiento del territorio y prevención y atención de desastres de origen sísmológico es un catálogo regional de sismos depurado, corregido y agrupado donde se puedan identificar patrones espaciales.

0.8. Aspectos éticos

El proyecto Reconocimiento de patrones espaciales sísmicos en el sur occidente colombiano se encuentra en el área de tecnologías de la información y las comunicaciones, específicamente en la minería de datos, la teoría de la información y el reconocimiento de patrones. Este proyecto no necesita de la experimentación con seres humanos o animales por esta razón no se necesita del consentimiento informado.

0.9. Divulgación

Los resultados del proyecto serán divulgados a través de la presentación de resultados en una conferencia internacional sobre reconocimiento de patrones o sísmología y la publicación de un artículo en una revista categoría B según COLCIENCIAS.

0.10. Formación de recurso humano

El proyecto involucrará un estudiante del programa de Maestría en Ingeniería con énfasis en Ingeniería de Sistemas o Electrónica de la Pontificia Universidad Javeriana-Cali cuya tesis de maestría estará relacionada con el objetivo general del

proyecto presentado. El director del proyecto se compromete a usar parte de los recursos destinados por la Fundación para la formación del estudiante.

BIBLIOGRAFÍA

- [1] A.Ñivia. Memoria explicativa del mapa geológico del Departamento del Valle del Cauca. publicación digital. Technical report, Ingeominas, Bogotá, 2001.
- [2] G. Paris, N. Machette, R. L. Dart, and K. M Haller. Map and database of quaternary faults and folds in Colombia and its offshore regions. Technical report, USGS-00-0284, 2000.
- [3] Ingeominas. Estudio de microzonificación sísmica de Santiago de Cali. Technical report, Alcaldía de Santiago de Cali, 2006.
- [4] Ingeominas-DAGMA. Estudio de microzonificación sísmica de Santiago de Cali-actualización del catalogo de sismos del occidente colombiano-CATOCC. Technical report, Ingeominas-DAGMA, 2005.
- [5] P. Berkhin. Survey of clustering data mining techniques. Technical report, Accrue Software, 2009.
- [6] Y Song, H.D Meng, M.J O Grady, and G.M O Hare. The application of cluster analysis in geophysical data interpretation. *Computers and Geoscience*, 14:263271, 2010.
- [7] A.K Jain, M.N Murty, and P.J. Flynn. Data clustering: a review. *ACM Computing Surveys*, 31(3):264–323, 1999.
- [8] Committee on Applied and National Research Council Theoretical Statistics. *Massive Data Sets: Proceedings of a Workshop*. National Academic Press, 1997.
- [9] E.H Isaaks and R.M Srivastava. *Applied Geostatistics*. Oxford University Press, 1989.
- [10] K. Aki. Characterization of barriers on an earthquake fault. *J. Geophys. Res*, 84:6140–6148, 1979.
- [11] A. Zamani and N.A Heshami. Computer-based self-organized tectonic zoning: a tentative pattern recognition for Iran. *Computers and Geosciences*, 30:705–718, 2004.
- [12] N. Bethoux, G. Ouillon, and M.Ñicolas. The instrumental seismicity of the western alps: spatio temporal patterns analysed with wavelets transform. *Geophysics. J. Int*, 135:177–194, 1998.

- [13] A. Ansari, N. Assadollah, and H. Zafarani. Clustering analysis of the seismic catalog of iran. *Computers & geosciences*, 35(3):475–486, 2009.