Características geográficas de los hallazgos de fosas clandestinas obervadas por la prensa en México: Una aproximación desde la estadística y la minería de textos.

Jorge Ruiz Reyes

22/04/2021

Los hallazgos de fosas clandestinas en México han sido recurrentes y extendidos durante quince años. Esto ha llevado a que existan diversos esfuerzos por documentar los hechos y aproximarse a la magnitud éstos, sobre todo para conocer el total de fosas y cuerpos de personas que han sido exhumadas en diferentes regiones. A pesar de esto, existe información sobre los hallazgos de fosas que todavía no ha sido analizada para identificar otros patrones durante este periodo. El presente artículo mostrará cómo se pueden utilizar herramientas de la estadística y la minería de textos para analizar las descripciones geográficas redactadas por notas de prensa escrita nacional y local, que han documentado hallazgos de fosas clandestinas entre 2007 y 2019. El objetivo es extraer el texto de estas notas e identificar cuáles son los sitios donde se han observado las fosas, así como de identificar diferencias cualitativas entre los diferentes estados de la República que han tenido reportes. Con esto se propone una nueva ruta para analizar información que aporte a estudios de contexto y procesos de búsqueda de personas en el país. **Palabras clave**: Fosas clandestinas, Minería de Textos, Prensa, México. Hidden grave findings have been recurrent and extended in Mexico for fifteen years. This has led to various efforts to document the events and approximate the phenomenon’s magnitude. Despite this, different types of data have not yet been analyzed to identify other hidden graves’ patterns during this period. This article will show how statistics and text mining tools can help us analyze graves’ geographical descriptions written by national and local press reports, which have documented hidden graves between 2007 and 2019. The objective is to extract the text of these notes and identify the sites where the graves have been documented, and recognize qualitative differences between the various states of the republic that have had reports. A new route is proposed to analyze information that contributes to context studies and missing persons’ search processes. **Keywords**: Hidden graves, Text Mining, Press, Mexico.

Glorario

**Fosas clandestinas**: Sitio donde una o más personas fueron enterradas de forma anónima y/o ilegal.

**Minería de textos**: Técnica proveniente de la ciencia de datos que conjunta la estadística, el aprendizaje de máquina (machine learning) y la lingüistica computacional para procesar, limpiar e identificar patrones de datos no estructurdos en forma de texto.

# Introducción

El fenómeno de la desaparición de personas y hallazgos de fosas clandestinas es uno de los problemas públicos más apremiantes en México. De acuerdo con cifras oficiales, actualmente existen más de 82 mil personas desaparecidas en el país [**1**]. Por su parte, los hallazgos de fosas han sido recurrentes y extendidos en la mayoría de las entidades del país, durante un periodo de quince años [**2**] Asimismo, entre 2018 y 2020, la Comisión Nacional de Búsqueda (CNB) ha registrado un total de 1 mil 394 fosas clandestinas en el país [**3**].

A pesar de la escala de los hallazgos de fosas, existen tres retos alrededor del fenómeno que deben ser atendidos para lograr implementar políticas públicas adecuadas en materia de búsqueda, exhumación e identificación de personas: 1) conocer la magnitud del problema; 2) superar los problemas de intercambio de información e identificación de cuerpos de personas que permancen sin identificar y; 3) conocer los patrones cuantitativos y cualitativos de los hallazgos de fosas que nos permitan generar nuevas búsquedas.

Este artículo presenta una aproximación que nos permite seguir generando respuestas respecto al tercer punto. Para hacerlo se realizó un ejercicio de análisis de texto cuantitativo sobre notas de prensa que han registrado hallazgos de fosas clandestinas en el periodo 2009-2019. El objetivo es identificar las caracterísiticas geográficas de los sitios descritos por las y los periodistas que han documentado los hallazgos de fosas clandestinas, y reconocer posibles diferencias y similitudes entre los municipios y regiones geográficas descritas. Con esto se pretende mostrar posibles nuevas rutas que sigan aportando al estudio de la desaparición y a la pregunta sobre dónde buscar a las personas desaparecidas.

# Análisis

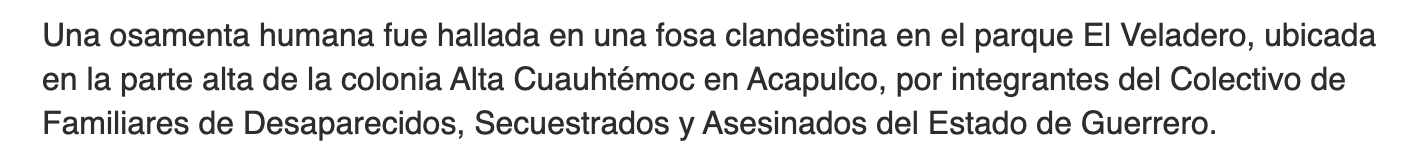
Los hallazgos de fosas clandestinas han sido registrados y observados en México por fuentes oficiales, prensa escrita nacional y local, así como por colectivos de familiares en búsqueda. Esta información cuenta con características específicas y se ve condicionada por el contexto de violencia y la capacidad política, económica o geográfica que tienen los actores para documentarla [**4**]. Sabemos que existen municipios y zonas que solo son observadas por la prensa, mientras que otras zonas sólo son documentadas por las autoridades o por colectivos de familiares. No solo esto, sino que la información que se produce de las fuentes es cualitativamente distinta, ya que los objetivos de su producción tiene fines distintos [**5**]. Esto significa que tenemos información dispersa y fragmentada, que responde a las capacidades que tenemos de documentar los hechos.

Dicha fragmentación o sesgos en la información son un reto, sin embargo son comunes en contextos de violencia extendida como el mexicano. Sobre todo porque, en términos estadísticos, no son muestras representativas de los hechos de la violencia [**6**]. Es decir, no podemos hacer inferencias sobre la población, ya que los procesos de generación de estos datos impide que todas las personas que han desaparecido, o todos los hallagos de fosas sean registrados. A pesar de esto, las fuentes contienen información que es valiosa porque nos permiten conocer los hechos y el contexto bajo el que se desarrolla esta violencia.

En México existen estudios que han documentado los hallazgos con miras a aproximarse a la magnitud de los hechos. Es decir, a responder cuántas fosas clandestinas se han localizado y cuántos cuerpos de personas se han exhumado en estados y municipios. [**2**,**7**] Sin embargo, estas fuentes contienen datos que no han sido aprovechados y que podrían ayudarnos a responder preguntas como ¿cuáles son las características geográficas de los sitios donde se localizan fosas clandestinas?, ¿existen patrones específicos de los lugares con base en los municipios y estados del país donde se localizan?

Las preguntas anteriores pueden responderse a través desde diversas aproximaciones. Por ejemplo, analizando fichas de documentación de los sitios de hallazgo generadas por colectivos de familiares, equipos forenses independientes o autoridades. El problema es que esta información puede ser limitada a ciertas zonas, encontrarse en diferentes formatos o inclusive puede ser negada por parte de las autoridades. A pesar de esto, fuentes como la prensa escrita también pueden darnos este tipo de descripiciones, como se puede observar en la siguiente imagen.

## Imagen 1



Fuente: El Sol de Acapulco, 2020

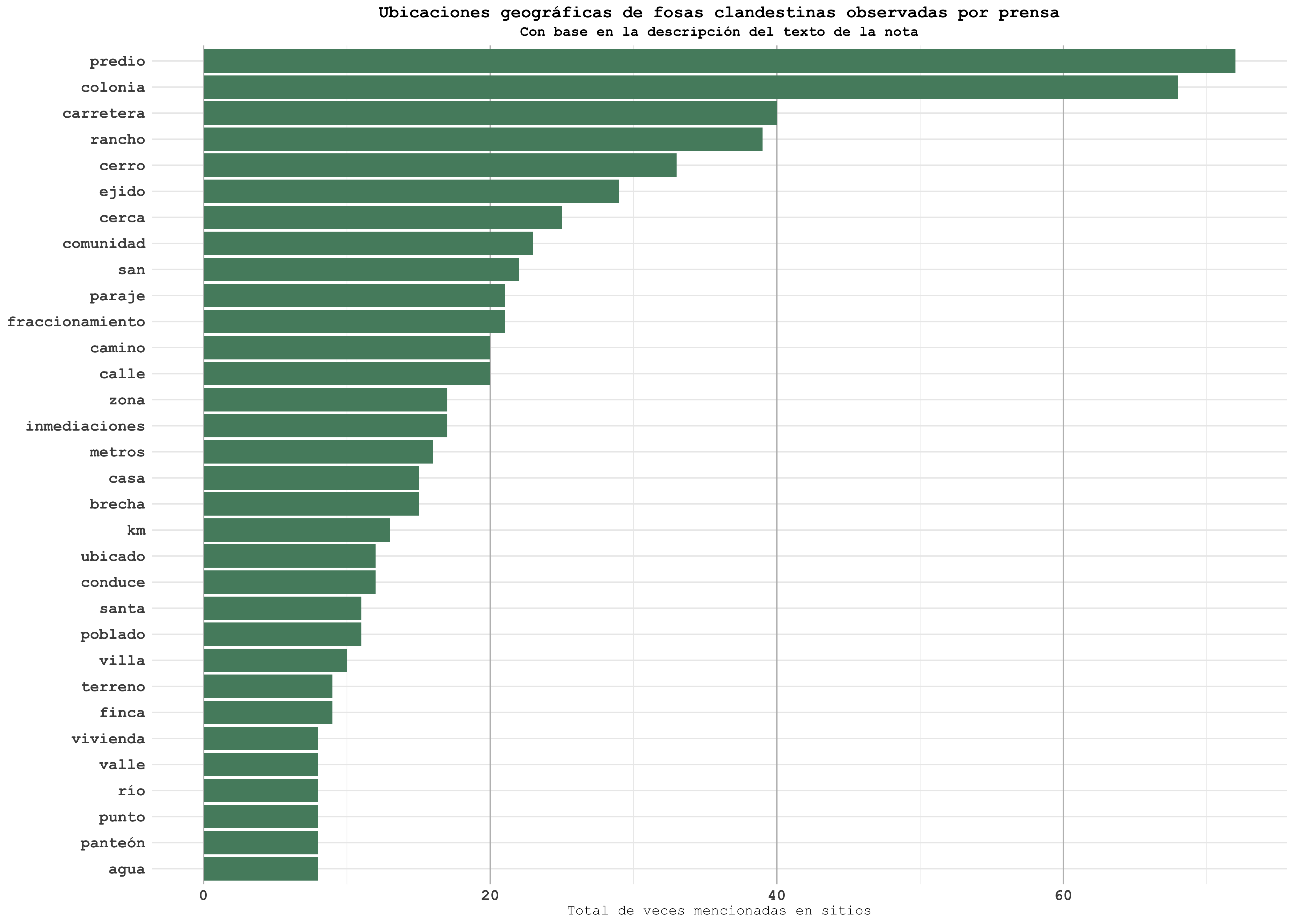
La **Imagen 1** recupera la descripción de una fosa clandestina localizada en Acapulco, Guerrero, en noviembre de 2020 [**8**]. Como se puede ver, este fragmento no solo nos da información de cuántas fosas y osamentas se localizaron, sino que también nos da detalles geográficos, como el parque “El Veladero” y la colonia “Alta Cuauhtémoc”. Este tipos de datos son comunes en las descripciones de las fosas que se han hecho por la prensa desde el año 2007 en el país.

Si tomamos un número mayor de notas y extraemos el texto de éstas, se pueden generar análisis cuantitativos de texto que nos permitan conocer las características de estos sitios en las diferentes regiones del país. Esta aproximación metodológica es conocida como *minería de textos*, la cual integra técnicas provenientes de la estadística, el aprendizaje de máquina (machine learning) y la lingüistica computacional para descrubir patrones de datos no estructurdos en forma de texto [**9**].

Utilizando la base de datos de fosas clandestinas generada por el Programa de Derechos Humanos de la Universidad Iberoamericana Ciudad de México (PDH Ibero), se tomaron las descripciones geográficas de 639 sitios que reportaron un total de 1491 fosas en 320 municipios del país entre 2007 y 2019 [**2**]. Cada uno de los sitios tuvo por lo menos un reporte prensa, sin embargo, existen hallazgos donde podían existir más de diez reportes, dada la atención mediática de los hechos. Lo anterior permitió tener una cantidad suficiente de texto con la que se exploró de manera preliminar, la posibilidad de extraer información cualitativa sobre las caracerísticas de las fosas que son observadas por la prensa.[[1]](#footnote-24)

Si tomamos en cuenta los 320 municipios donde se han localizado fosas reportadas por la prensa, se puede observar que las zonas geográficas se componen por las siguientes características:

## Figura 1



Fuente: Generada con información del PDH Ibero

La **figura 1** muestra las palabras (ubicaciones) mencionadas con mayor frecuencia por las notas de prensa que reportaron hallazgos de fosas entre los años 2007-2019. Los predios (72 menciones), colonias (68 menciones), carretera (40 menciones), rancho (39 menciones) y cerro (33 menciones) son las cinco palabras que se retoman con mayor frecuencia dentro de las descripciones de las y los periodistas. Sin embargo, la figura anterior también nos permite reconocer que los hallazgos de fosas se dan tanto en zonas rurales, como en zonas urbanas de los municipios de México.

Lo anterior indica las dificultades y las diferentes estrategias que deben ser planeadas para los procesos de exhumación y recuperación de restos de personas. Sobre todo porque no existe un solo sitio característico para iniciar una exhumación, ya que ésta puede ocurrir en una brecha o cuerpo de agua en una zona rural, así como un fraccionamiento en una ciudad con una densidad poblacional alta.

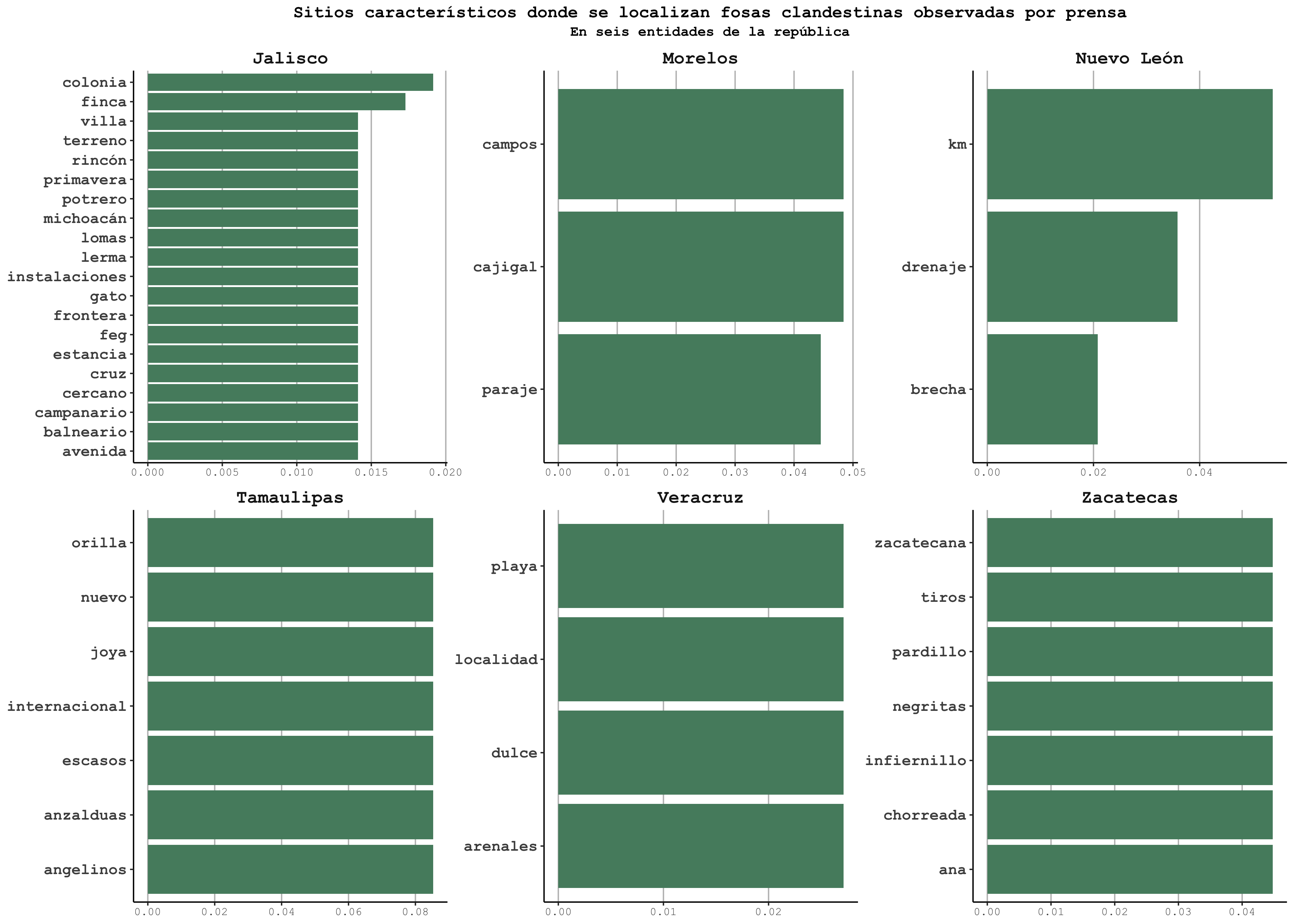
Consecuentemente, con la minería de textos podemos encontrar otros patrones que nos permitan conocer si hay sitios característicos o “distintivos” donde se localizan fosas clandestinas en estados del país, en comparación con otros.

Calculando una medida estadística conocida como *Frecuencia de término - frecuecia inversa de término*(TF-IDF por sus siglas en inglés), podemos conocer la relevancia que tiene una palabra para un subconjunto de textos, con base una colección de textos mayor [**10**].[[2]](#footnote-28)

En nuestro caso, agruparemos el texto de las notas con base a los estados del país, donde cada palabra que indique una descripción geográfica en un estado es un subcojunto de las descripciones geográficas de todos los estados. Posteriormente, analizamos si una palabra se menciona con frecuencia en un determinado estado o si no se menciona. De esta forma, mientras más frecuente sea una palabra en un estado, pero no en otro, podemos establecer que es un sitio distintivo donde se localizan fosas en comparación con otros.

Los resultados pueden observarse a continuación:

**Figura 2**



Fuente: Generada con información del PDH Ibero

En la **Figura 2** podemos ver cuáles son los sitios característicos donde se han observado fosas clandestinas en seis entidades de México, entre los años 2007-2019. Es importante aclarar que una palabra de alguna entidad es distitiva respecto a las otras, ya que tiene mayores menciones en ese estado. Sin embargo, esto no significa que palabras como colonias, localidades o kilómetros no se mencionen en el resto.

Si nos enfocamos en Jalisco, podemos ver que las palabras “distintivas” respecto a lo reportado por las notas reflejan sitios urbanos como avenidas, fincas, colonias y sus nombres (Colonia Primavera, Colonia Campanario) e instalaciones universitarias como el hallazgo de fosas en la Federación de Estudiantes de Guadalajara de la Universidad de Guadalajara (FEG) en 2011 [**11**]. Estos sitios son distintivos para este estado ya que una parte importante de los hallazgos de fosas en este periodo han sido dentro de la Zona Metropolitana de Guadalajara [**12**]. Sin embargo, también destaca Michoacán, haciendo alusión a las fosas de La Barca, localizadas entre la frontera de Jalisco y Michoacán en 2013-14 [**13**].

Por su parte, en Morelos podemos ver que el sitio distintivo donde se reportan fosas son los campos conocidos como “Cajigal”, localizadas en 2014 [**14**]. Asimismo, en Nuevo León se identifca que los hallazgos reportados en la prensa son distintivos por estar dentro de tramos carreteros (km) y cercanos a brechas o drenajes, siendo también consistente con los observado, por ejemplo, con las fosas de Salinas Victoria y de García [**15**].

Asimismo, en Tamaulipas el Ejido “La Joya” es distintivo, ya que en este lugar se localizaron las fosas de San Fernando en 2011, donde más de 150 cuerpos de personas fueron exhumados [**16**]. También se mencionan “Internacional” y “Anzalduas”, ya que fosas clandestinas también han sido localizadas cerca de este puente fronterizo con Estados Unidos [**17**]. En Veracruz, los sitios distintivos son la “Playa Agua Dulce”, en el municipio Lerdo de Tejada, donde elementos de la marina localizaron fosas en 2012 [**18**]. Por último, en Zacatecas se mencionan nombres de colonias y predios como “La Zacatecana” o “Las Negritas”, sin embargo, también se mencionan “tiros”, ya que algunas fosas en el estado se ha localizado cercanas a tiros de mina [**19**].

# Discusión y conclusión

Los resultados de la sección anterior muestran que la *minería de textos* puede ser una herramienta útil para obtener información cualitativa de los hallazgos de fosas clandestinas en el país. Es decir, podemos aproximarnos a identificar otras características y patrones más allá de las cifras, que pueden ser útiles para estudios de contexto y planeación en materia de políticas públicas y programas contemplados en la *Ley General en Materia de Desaparición Forzada y Desaparición Cometida por Particulares* [**20**]. Esto se da gracias al esfuerzo de documentación por parte de la prensa en estos años.

Entre los resultados destaca que la prensa ha podido documentar hallazgos en zonas rurales y urbanas del país, desde 2009 hasta 2019. Asimismo, utilizando una métrica como el TF-IDF, pudimos clasificar los lugares característicos o distintivos donde se observan fosas clandestinas en seis entidades del país por la prensa. Estos resultados resaltan, por ejemplo, que en Jalisco se localizan fosas clandestinas con más frecuencia dentro de la Zona Metropolitana de Guadalajara, mientras que en Nuevo León los sitios son distintivos por estar en tramos carreteros y brechas, mostrando que los patrones cambian entre regiones.

A pesar de esto, debemos tomar en cuenta dos consideraciones. La primera es que estos resultados no indican que todas las fosas clandestinas se van a localizar u observar en los sitios que se describen por las notas de prensa. Por el contrario, sabemos que no todos los municipios que han tenido hallazgos de fosas en el país han tenido reportes de fosas [**5**]. Algunos sitios pueden ser reportados ya que son de fácil acceso o la integridad de las y los reporteros se ve menos comprometida. Asimismo, es posible que la prensa reporte hallazgos donde hay una mayor cantidad de fosas o cuerpos de personas exhumados, dado que la magnitud del hecho los hace “mediáticos”, dejando hallazgos con menor número de fosas o cuerpos sin reportar [**21**].

La segunda es que no podemos establecer una causalidad respecto a los resultados. Es decir, no podemos afirmar que una fosa clandestina se localizará en un lugar dado que éste tiene *N* características geográficas. Esta es una aproximación preliminar donde todavía se tiene que incluir una mayor cantidad de información, sobre todo de otras fuentes oficiales y no oficiales. Con mayor información de calidad y abierta se podrían generar modelos de análisis de texto que nos permitan complementar otros estudios generados desde organizaciones nacionales e internacionales, colectivos de familiares, autoridades y academia.

De esta forma, el objetivo de este texto fue el de mostrar cómo es que podemos aprovechar datos no estructurados como es el texto, para conocer más sobre el fenómeno de la desaparición de personas y hallazgos de fosas clandestinas en el país. La información que se produce entorno a esta problemática puede estar fragmentada y tener sesgos importantes de identificar, sin embargo, también presenta una oportunidad para fortalecer estudios de contexto y planeación de políticas públicas para atenderla.

# Referencias

* 1. CNB. Versión Pública RNPDNO. [Internet]. febrero 2021 [citado 9 febrero 2021]. Disponible en: <https://versionpublicarnpdno.segob.gob.mx/Dashboard/Index>
  2. Gonzáles Núñez D, Ruiz Reyes J, Chávez Vargas LG, Guevara Bermúdez JA. Violencia y terror. Hallazgos sobre fosas clandestinas en México 2007-2017. . 1.ª ed. Mexico City: Universidad Iberoamericana ; 2019.
  3. Encinas, A. Búsqueda e Identificación de Personas Desaparecidas, Informe 2020 [Internet]. febrero 2021 [citado 9 febrero 2021]. Disponible en: <http://www.alejandroencinas.mx/home/busqueda-e-identificacion-de-personas-desaparecidas-informe-2020/>
  4. Manrique-Vallier D, Price ME, Gohdes A. Multiple Systems Estimation Techniques for Estimating Casualties in Armed Conflicts. . En: Seybolt TB, Aronson JD, Fischhoff B, eds. by. Counting Civilian Casualties. An Introduction to Recording and Estimating Nonmilitary Deaths in Conflict. . 1.ª ed. Nueva York : Oxford University Press ; 2013. p. 165.
  5. Meltis Véjar M, Ruiz Reyes J, Ball P. New Results For The Identification Of Municipalities With Clandestine Graves In Mexico [Internet]. 23 de noviembre 2017 [citado 09 febrero 2021]. Disponible en: <https://hrdag.org/2017/11/23/new-clandestine-graves-mexico/>
  6. Patrick B. The Bigness of Big Data: samples, models, and the facts we might find when looking at data. En: Alston P, ed. by. The Transformation of Human Rights Fact-Finding. 1.ª ed. Nueva York : Oxford University Press; 2016. p. 427.
  7. Guillén A, Torres M, Turati M. El país de las 2 mil fosas [Internet]. 2018 [citado 9 febrero 2021]. Disponible en: <https://quintoelab.org/project/el-pais-de-las-2-mil-fosas>
  8. Moreno C. Localizan osamenta humana en fosa clandestina en el Parque Veladero [Internet]. 2020 [citado 9 febrero 2021]. Disponible en: <https://www.elsoldeacapulco.com.mx/policiaca/localizan-osamenta-humana-en-fosa-clandestina-en-el-parque-veladero-acapulco-estado-noticia-guerrero-violencia-6027154.html>
  9. Han J, Kamber M, Pei J. Data Mining Trends and Research Frontiers. En: Han J, Kamber M, Pei J, eds. by. Data Mining: Concepts and Techniques [Internet]. 3.ª ed. Elsevier ; 2012 [citado 9 febrero 2021]. pp. 585–631. Disponible en: <https://www.sciencedirect.com/science/article/pii/B9780123814791000137>
  10. Silge J, Robinson D. Text mining with R. A tidy approach. [Internet]. 1.ª ed. O’Reilly Media; 2017 [citado 9 febrero 2020]. Cap.3. Disponible en: <https://www.tidytextmining.com/index.html>
  11. RUVALCABA O. Encuentran dos cadáveres en edificio de la FEG. EL INFORMADOR [Internet]. 14 de diciembre 2011 [citado 9 febrero 2021];. Disponible en: <https://www.informador.mx/Jalisco/Encuentran-dos-cadaveres-en-edificio-de-la-FEG-20111214-0020.html>
  12. Souza D, Franco D. Los sitios de exterminio que proliferan en Jalisco. Pie de Página [Internet]. 2020 [citado 9 febrero 2021];. Disponible en: <https://piedepagina.mx/los-sitios-de-exterminio-que-proliferan-en-mexico/>
  13. Chouza P. La Barca: el pueblo tranquilo al que le han surgido 67 cadáveres. El País [Internet]. 1 de dciembre 2013 [citado 9 febrero 2021];. Disponible en: <https://elpais.com/internacional/2013/12/02/actualidad/1385951092_787703.html>
  14. Redacción. Encuentran cuatro fosas clandestinas en Morelos. El Financiero [Internet]. 6 de febrero 2014 [citado 9 febrero 2021];. Disponible en: <https://www.elfinanciero.com.mx/sociedad/hallazgo-de-cuatro-fosas-clandestinas-en-morelos-pgje>
  15. Chio Peña Y. La búsqueda de familiares de desaparecidos en Nuevo León. A dónde van los desaparecidos [Internet]. 18 de abril 2019 [citado 9 febrero 2021];. Disponible en: <https://adondevanlosdesaparecidos.org/2019/04/18/la-busqueda-de-familiares-de-desaparecidos-en-nuevo-leon/>
  16. Redacción. Hallan 59 muertos en Tamaulipas. Animal Político [Internet]. 07 de abril 2011 [citado 9 febrero 2021];. Disponible en: <https://www.animalpolitico.com/2011/04/hallan-59-muertos-en-tamaulipas/>
  17. Cedillo J. Hallan restos humanos en fosa clandestina en Reynosa, Tamaulipas. BAJAPRESS [Internet].22 de octubre 2019 [citado 9 febrero 2021];. Disponible en: <https://www.bajapress.com/noticias/mexico/tamaulipas/hallan-restos-humanos-en-fosa-clandestina-en-reynosa-tamaulipas>
  18. Redacción. En Lerdo de Tejada, hallan cuatro fosas clandestinas con 11 cadáveres. Presencia [Internet]. 24 de junio 2012 [citado 9 febrero 2021];. Disponible en: <https://www.presencia.mx/nota.aspx?id=50378&s=7>
  19. Redacción. Hallan cuatro fosas clandestinas en La Zacatecana. NTR Peridismo crítico [Internet]. 4 de septiembre 2017 [citado 9 febrero 2021];. Disponible en: <http://ntrzacatecas.com/2017/09/04/hallan-cuatro-fosas-clandestinas-en-la-zacatecana/>
  20. Cámara de Diputados. LEY GENERAL EN MATERIA DE DESAPARICIÓN FORZADA DE PERSONAS, DESAPARICIÓN COMETIDA POR PARTICULARES Y DEL SISTEMA NACIONAL DE BÚSQUEDA DE PERSONAS [Internet]. 2017 [citado 9 febrero 2021]. Disponible en: <http://www.diputados.gob.mx/LeyesBiblio/pdf/LGMDFP_171117.pdf>
  21. Sloboda J, Dardagan H, Spagat M, Hsiao-Rei Hicks M. Iraq Body Count. A Case Study In The Uses of Incident-Based Conflict Casualty Data . En: Seybolt TB, Aronson JD, Fischhoff B, eds. by. Counting Civilian Casualties. An Introduction to Recording and Estimating Nonmilitary Deaths in Conflict. . 1.ª ed. Nueva York : Oxford University Press ; 2020. p. 68.

1. El análisis se realizó utilizando el lenguaje de programación R. Los datos y el código pueden consultarse en este repositorio: <https://github.com/JorgeRuRe/fosas-GIZ-CICR> [↑](#footnote-ref-24)
2. Específicamente, se multiplican dos métricas: 1) la frecuencia de término donde se identifica la frecuencia de la palabra en el subconjunto y 2) la frecuencia inversa de documento donde se divide el número total de documentos en la colección, por el número de subconjuntos que contienen el término; ambos términos en escala logarítmica. La fórmula puede verse como:

   donde:

   y donde: [↑](#footnote-ref-28)