



UNIVERSIDAD AUTÓNOMA DEL ESTADO DE HIDALGO
INSTITUTO DE CIENCIAS BÁSICAS E INGENIERÍAS
LICENCIATURA EN INGENIERÍA EN GEOLOGÍA AMBIENTAL

TESIS

**MAPEO DEL CONOCIMIENTO CIENTÍFICO
ACUMULADO EN GEOCIENCIAS A TRAVÉS DE
UN ANÁLISIS BIBLIOMÉTRICO: UNA
HERRAMIENTA POCO EXPLORADA.**

Para obtener el título de
Licenciada en Ingeniería en Geología Ambiental

PRESENTA

Akatzin Joselyn Melo Moreno

Director externo

Dr. David Yáñez Dávila

Instituto de Energías Renovables (IER, UNAM)

Director interno

Dr. Enrique Cruz Chávez (AACTyM, ICBI, UAEH)

Comité tutorial

Dr. Fred Yoan Pérez Corona

Dr. Gustavo Santos Raga

Mineral de la Reforma, Hidalgo, febrero, 2025

ÍNDICE GENERAL

ÍNDICE DE FIGURAS	IV
ÍNDICE DE TABLAS.....	VI
ÍNDICE DE ECUACIONES.....	VI
<i>DEDICATORIA</i>	VII
<i>AGRADECIMIENTOS</i>	VIII
RESUMEN.....	IX
ABSTRACT	X
CAPÍTULO 1: GENERALIDADES.....	1
1.1 Introducción.....	1
1.2 Antecedentes	4
1.3 Justificación	8
1.4 Objetivos	9
1.4.1 Objetivo general	9
1.4.2 Objetivos particulares.....	9
CAPÍTULO II MARCO TEÓRICO	10
2.1 Análisis bibliométrico	10
2.1.1 Base de datos bibliométricas	10
2.1.2 Método de revisión por pares	12
2.1.3 Literatura gris.....	12
2.2 Métricas bibliométricas	13
2.2.1 Métricas usadas en el análisis de desempeño científico.....	13
2.2.2 Métricas usadas en el análisis de la estructura CoInSo.....	16
2.2.3 Leyes bibliométricas	17
2.2.4 Software bibliométrico.....	21
2.2.5 Metodologías bibliométricas reportadas.....	29
2.2.6 Herramientas de Inteligencia Artificial para el BA.....	33
CAPÍTULO III METODOLOGÍA	36
CAPÍTULO IV RESULTADOS.....	39
4.1 Etapa 1: P-TeX	39
4.1.1 Elegibilidad de datos (subetapa i).	39

4.1.2 Recopilación de datos (subetapa ii).....	39
4.1.3 Semántica de los metadatos (subetapa iii).	40
4.2 Etapa 2: PA-A y CoInSo-A	41
4.2.1 Análisis y mapeo de datos (Subetapa iv)	42
CAPÍTULO V DISCUSIÓN.....	57
5.1 Producción científica mundial.....	57
5.2 Productividad de los documentos científicos por país.	57
5.3 Productividad de documentos científicos por institución.....	58
5.4 Productividad de los documentos científicos en revistas indexadas.	58
5.5 Productividad de los documentos científicos por autor.	59
5.6 Mapa bibliométrico de co-ocurrencias.....	60
CAPÍTULO VI CONCLUSIONES	61
5.1 Perspectivas (Subetapa v).....	61
REFERENCIAS	63

ÍNDICE DE FIGURAS

- Figura 1.** Conocimiento científico mapeado: a) análisis bibliométricos reportados en Web of Science para todas las áreas de conocimiento b) BA reportados en WoS para geociencias; c) análisis del comportamiento en la aplicación de los BA que describe la correlación entre el número de documentos publicados por año de todas las áreas de investigación; d) análisis de la productividad científica de los BA aplicada a geociencias.3
- Figura 2.** La metodología bibliométrica propuesta (BAM por su acrónimo en inglés: *Bibliometric Analysis Metodology*) busca reducir al mínimo el sesgo y los errores en el mapeo de estudios científicos acumulativos en diversos campos del conocimiento. Modificada de (Yáñez-Dávila *et al.*, 2023)38
- Figura 3.** Gráfico de la producción científica del BA Aplicada a documentos de geociencias, donde se observa un crecimiento exponencial en la publicación descrito por la correlación entre el número de documentos por año; También se puede observar las diferentes fases de Price, donde la fase 1 se define como «zona de precursores» en la investigación y la fase 2 se define como «zona de crecimiento exponencial» iniciada sustancialmente en 2020.....44
- Figura 4.** Distribución mundial de documentos publicados en revistas especializadas, revisiones y actas de congresos sobre análisis bibliométricos en las geociencias (1993-2024). El grosor de la línea roja indica el porcentaje de colaboración. Los nombres de los países se abrevian utilizando la convención internacional sugerida por la norma internacional ISO-3166. La escala logarítmica del número de citas va de 1 a 10000 y se solapa con el número de artículos publicados de manera multinacional o de un solo país.....46
- Figura 5.** Diagrama esquemático de redes bibliométricas que muestra seis grupos principales de colaboración más productivos entre instituciones identificada a partir de este estudio.48
- Figura 6.** Zonas de las revistas que cubren el tema de las geociencias análisis bibliométrico (1993-2024) según la ley de Branford, mostrando las zonas núcleo, intermedia y periférica.....49
- Figura 7.** Diagrama de Sankey donde muestra las preferencias de los diez primeros grupos de autores (Tabla 8) para publicar sus trabajos de investigación de BA en el tema de geociencias (1993-2024). El tamaño de las líneas grises representa la cantidad de información que fluye entre colaboración de autores y revistas. Así, las líneas más anchas indican un flujo mayor, mientras que las líneas más estrechas indican un flujo menor de información.52
- Figura 8.** Mapa bibliométrico esquemático de co-ocurrencia que muestra las palabras clave más representativas encontradas en DP y agrupadas en siete grandes grupos de palabras clave: procesos relacionados con la teledetección e inteligencia artificial (red de color rojo), temas referentes al cambio climático y sustentabilidad (color verde), temas con

contaminación atmosférica y salud (en color amarillo), temas relacionados con recursos hídricos (rosa, azul y café), y finalmente temas respecto a la remediación del suelo (morado).	54
Figura 9. Mapa temático basado en el análisis de las redes de KeyWords Plus® del mapeo bibliométrico en geociencias (1993-2024).	56
Figura 10. Crecimiento de las principales revistas a través de los años en BA aplicados a geociencias.....	59

ÍNDICE DE TABLAS

Tabla 1.	Recopilación de bases de datos bibliométricas indexadas y no indexadas de uso común dentro de la comunidad científica.....	11
Tabla 2.	Métricas de desempeño (PA-A), estructura conceptual, intelectual y social (CoInSo-A) y leyes bibliométricas utilizadas para el mapeo científico de información.....	20
Tabla 3.	Softwares bibliométricos utilizados para el mapeo y visualización de información.....	27
Tabla 4.	Metodologías bibliométricas reportadas en orden cronológico y sus diferencias en comparación con este estudio.	30
Tabla 5.	Herramientas de Inteligencia Artificial (IA) para el procesamiento de información científica	34
Tabla 6.	Cadenas booleanas utilizadas en esta investigación	40
Tabla 7.	Resumen de los indicadores bibliométricos encontrados en las bases de datos Scopus y Web of Science sobre el análisis bibliométrico de las geociencias (Periodo 1993 - 2024).....	41
Tabla 8.	Principales instituciones mundiales y país residente por productividad de artículos publicados sobre temas de geociencias (Periodo 1993 - 2024)	47
Tabla 9.	Principales grupos mundiales de investigadores según la productividad a lo largo del tiempo, el número de artículos y citas publicados sobre el tema de investigación de la BA aplicada a las geociencias.....	50

ÍNDICE DE ECUACIONES

Ecuación 1.	Coeficiente de colaboración.....	14
Ecuación 2.	Ley de Leimhkuler.....	19
Ecuación 3.	Expresión original de la Ley de Lotka.....	19
Ecuación 4.	Expresión actual de la Ley de Lotka.....	20
Ecuación 5.	Ley de Price	42
Ecuación 6.	Porcentaje de crecimiento anual	42
Ecuación 7.	Tiempo de duplicación	43

DEDICATORIA

A quien me cuida desde arriba, Anita, gracias por todo, por enseñarme a echarle ganas, a ponerle dedicación todos los días a todo lo que hago, te amo por siempre y te llevo en mi corazón.

A mi compañera y mejor amiga de vida, Alinne, por acompañarme en cada paso que doy, por aconsejarme, regañarme cuando es necesario, por los momentos divertidos, siempre estar orgullosa de mí y celebrar nuestros logros juntas, te amo hermana.

A mí mejor amiga Daniela, por hacer más divertida y amena la estancia en la universidad, por todos esos momentos que hemos pasado juntas y estos años de amistad, te amo amiga.

A mi mejor amigo, compañero, cómplice y amor, Cristian, por estar conmigo en las buenas y en las malas, por quererme, por siempre tener una palabra para mí, acompañarme, escucharme y ayudarme en lo que necesito, te amo.

A mis tíos y mi abuela, que me quieren y me apoyan en todo lo que necesito, los quiero.

A los amigos que hice en el camino, por compartir todo conmigo, por los mejores momentos, Lib, Juan, los amo.

A mis gatitas, que, a pesar de no estar cerca ahora, siempre están en mi corazón las amo y las extraño.

Y, por último, pero no menos importante, a mí, por ponerle dedicación a todo todos los días y no rendirme.

AGRADECIMIENTOS

A la Universidad Autónoma del Estado de Hidalgo (UAEH) por abrirme las puertas para formarme académicamente.

Al Instituto de Ciencias Básicas e Ingenierías (ICBI) que a través del Área Académica de Ciencias de la Tierra y Materiales me dieron las herramientas y aprendizajes necesarios para mi titulación.

A mi director de tesis Dr. David Yáñez Dávila, en conjunto con el Instituto de Energías Renovables (IER, UNAM) por darme la oportunidad de trabajar con él y confiar en mí para este trabajo.

Al Dr. Enrique Cruz Chávez, por tomarme en cuenta para este proyecto, aconsejarme y corregirme.

Al proyecto de ciencia de frontera de CONAHCyT (CF-2023-G-490) y al proyecto DGAPA-PAPIIT (IN 108322) por el soporte técnico - científico para desarrollar la presente tesis.

A mi sinodales, el Dr. Gustavo Santos Raga y al Dr. Fred Yoan Pérez Corona, por sus comentarios acertados que contribuyeron de manera sustanciosa a mejorar el presente trabajo de investigación.

RESUMEN

La presente tesis examina el análisis bibliométrico como una herramienta fundamental para la revisión crítica de la literatura en las geociencias (geografía, ciencias atmosféricas, geología, oceanografía y ciencias planetarias), destacando los diferentes indicadores bibliométricos, las diferentes metodologías empleadas para realizar este tipo de estudios, así como bases de citas indexadas más utilizadas, los tipos de softwares para el mapeo y visualización de datos más usados junto con las herramientas de inteligencia artificial actualmente utilizadas para el desarrollo de la investigación científica. La metodología integral empleada consta de dos principales fases: 1) preprocesamiento de texto donde se realiza la elegibilidad de datos, la recopilación de datos, y la semántica de los metadatos; 2) el análisis de desempeño y la estructura conceptual, intelectual y social. Esta última etapa permitió evaluar los análisis bibliométricos aplicados a geociencias a través de métodos estadísticos descriptivos y multivariantes, la producción científica de países, instituciones, autores, además de determinar áreas de investigación destacadas y redes de colaboración a partir de metadatos obtenidos de las dos principales bases de datos indexadas (WoS y Scopus), para el periodo de tiempo de 1993-2024. El actual análisis sugiere temas básicos de investigación ampliamente desarrollados como servicios ecosistémicos, uso y contaminación del suelo, cambio climático, así como temas motores abordados en la actualidad como en contaminación atmosférica, biorremediación, teledetección, inteligencia artificial, sustentabilidad, agricultura de precisión, manejo de recursos hídricos, entre otros. Estas áreas se identifican como los principales focos en los análisis bibliométricos, lo que indica un mayor interés en la investigación de estos temas, por lo que evidenciaron así un sesgo en la distribución de estudios, con áreas y regiones poco investigadas, en consecuencia destacando la importancia de los análisis sistemáticos para sintetizar y organizar información científica, identificar áreas emergentes, avances científicos y medir el impacto de investigaciones e investigadores promoviendo la colaboración global y facilitando el acceso al conocimiento geocientífico para la sociedad.

Palabras clave: *análisis de la literatura, metodología bibliométrica, métricas bibliométricas, Web of Science, Scopus, clustering, network analysis.*

ABSTRACT

The present thesis examines bibliometric analysis as a fundamental tool for the critical review of literature in the Earth Sciences (geography, atmospheric sciences, geology, oceanography and planetary sciences). The text provides a comprehensive overview of the various bibliometric indicators and methodologies employed in this field of study. It also examines the most used indexed citation databases, the predominant software for data mapping and visualisation, and the current artificial intelligence tools utilised in the development of scientific research. The comprehensive methodology under consideration comprises two primary phases: Firstly, there is text pre-processing, where data eligibility, data collection, and metadata semantics are performed. Secondly, there is performance analysis and conceptual, intellectual and social structure. The final stage of the methodological process enabled the evaluation of the bibliometric analyses applied to Earth Sciences through the utilisation of descriptive and multivariate statistical methods. This evaluation contained the scientific production of countries, institutions, and authors, as well as the identification of hotspot research areas and collaboration networks. These determinations were made based on metadata obtained from the two primary indexed databases, Web of Science and Scopus, for the period spanning from 1993 to 2024. The current analysis suggests that there is a broad development of fundamental research topics, including ecosystem services, land use and contamination, climate change, as well as current driving themes such as air pollution, bioremediation, remote sensing, artificial intelligence, sustainability, precision agriculture, water resource management, among others. These domains are identified as the primary hotspots in the bibliometric analyses, which demonstrate a bias in the distribution of studies, with under-researched domains and regions. Consequently, this highlights the importance of systematic analyses to synthesise and structure the scientific information, identify emerging themes, scientific advances, and measure the impact of research and researchers. This, in turn, promotes global collaboration. It also makes geoscientific knowledge more accessible to society.

KeyWords: *literature review, systematic review, bibliometric methodology, bibliometric metrics, Web of Science, Scopus.*

CAPÍTULO 1: GENERALIDADES

1.1 Introducción

El Análisis Bibliométrico (BA por su acrónimo en inglés; *Bibliometric Analysis*) es una herramienta útil para la revisión crítica de la literatura a partir del análisis de metadatos provenientes de bases de datos de citas indexadas (p. ej. Scopus, Web of Sciences, PubMed, etc.). Además, este análisis permite obtener información del conocimiento acumulado en diversas áreas de investigación científicas y generar resultados y/o conclusiones antes no dilucidadas (Yáñez-Dávila et al., 2023). El uso del BA se ha incrementado en los últimos años dentro de la investigación científica, principalmente en temas de *ciencias de la información y bibliotecarias*, *ciencias medioambientales* e *informática y aplicaciones interdisciplinarias* (figura 1-a) debido en gran medida al confinamiento ocasionado por COVID-19 (coronavirus), que afectó diversas actividades humanas, principalmente la movilidad urbana, la demanda eléctrica, la calidad del aire y las actividades científico-académicas (Corpus Mendoza et al., 2021; Donthu et al., 2021).

Referente a los BA reportados para todas las áreas de conocimiento en Web of ScienceTM (WoS), se observan dos periodos de desarrollo (figura 1-c): i) el periodo entre 1972-2020 se aprecia un crecimiento lineal de los BA, referido como la zona de estudios “precursores”; ii) el periodo 2021-2024 se observa un comportamiento de tipo exponencial de los BA, referido como zona de “crecimiento exponencial”, este comportamiento nos indica una progresiva adopción de este tipo de estudio en el análisis de la literatura para generar nueva información y áreas de investigación con un alto potencial a investigar.

Referente a los BA desarrollados a las geociencias, este representa solo el 2.4% de todos los estudios de BA registrados en la base de datos de WoS hasta el 12 diciembre 2024. Del total de 748 documentos publicados, el 45% se relaciona con temas en geografía, el 28% con temas de geología, el 17% con temas en ciencias atmosféricas, el 6% con temas de oceanografía y finalmente el 4% con temas relacionados a ciencias planetarias (figura 1-a y 1-d). En 2020 durante el confinamiento ocasionado por el COVID-19, las investigaciones presenciales en áreas de geociencias se vieron afectadas, por consecuencia hubo la necesidad de adaptarse, llevando a un incremento de actividades de investigación secundaria, a través de la aplicación de diversas metodologías referentes a revisiones sistemáticas de la literatura, metaanálisis y

análisis bibliométricos (Alaniz Álvarez et al., 2024). De igual manera, a raíz de la pandemia, las herramientas digitales y el análisis de datos se han convertido en elementos esenciales para el desarrollo de conocimiento científico y el aprendizaje geociencias (Alaniz Álvarez et al., 2024).

El BA, permite analizar la investigación científica a través de la minería de textos de los metadatos obtenidos de bases de datos de citas indexadas, de dos diferentes formas: i) el análisis de desempeño que se realiza mediante métodos estadísticos descriptivos (p. ej. la producción científica de países, instituciones, revistas y autores); ii) el análisis de la estructura conceptual, intelectual y social mediante métodos estadísticos multivariados (p. ej. correlación entre citas, red de colaboración, y áreas de investigación destacadas) relacionados con estudios sobre cualquier tema de carácter científico y su marco temporal (Bezak et al., 2021; Qin Yong et al., 2022; Martín-Martín et al., 2021; Donthu et al., 2021).

Una de las principales ventajas del BA consiste en su capacidad para identificar los aspectos claves de diferentes áreas de investigación (p. ej. geología, geografía, ciencias atmosféricas, oceanografía y ciencias planetarias, entre otras) y responder preguntas fundamentales como:

- 1) ¿Cuál ha sido la evolución temporal y el estado actual del desarrollo de la investigación?
- 2) ¿Cuáles son las principales instituciones y líderes de investigación a nivel mundial en este campo?
- 3) ¿Qué revistas de revisión por pares han publicado los trabajos más relevantes?
- 4) ¿Cuáles son las áreas de investigación más significativas en este tema y cuáles son los artículos más citados?
- 5) ¿Qué áreas de investigación emergentes podrían convertirse en puntos focales para futuros estudios y abordar cuestiones técnicas críticas en la investigación científica y/o industria?, entre otras.

En esta tesis se pretende responder a las anteriores preguntas con respecto a los BA reportados en geociencias dentro del marco temporal de 1993-2024. Por consiguiente, dentro del primer capítulo se analizan los antecedentes de los BA reportados en la literatura. Consecutivamente en el capítulo 2 se presenta un extenso marco teórico referente a: 1) diversas metodologías bibliométricas usadas; 2) diferentes bases de datos de citas disponibles para la obtención de los metadatos bibliométricos; 3) recopilación de métricas bibliométricas usadas en el análisis de la literatura científica; 4) análisis de software bibliométrico disponible para el mapeo y visualización de los metadatos; y finalmente 5) herramienta de inteligencias artificial para el

desarrollo del BA. Por consiguiente, en el capítulo 3 se desarrolló una metodología integral para el análisis de los metadatos bibliométricos. Con esta metodología aplicada, en el capítulo 4 se muestran los resultados para el mapeo de conocimiento científico acumulado en geociencias. Finalmente, en el capítulo 5 y 6 se concluye con las discusiones y perspectivas de esta investigación.

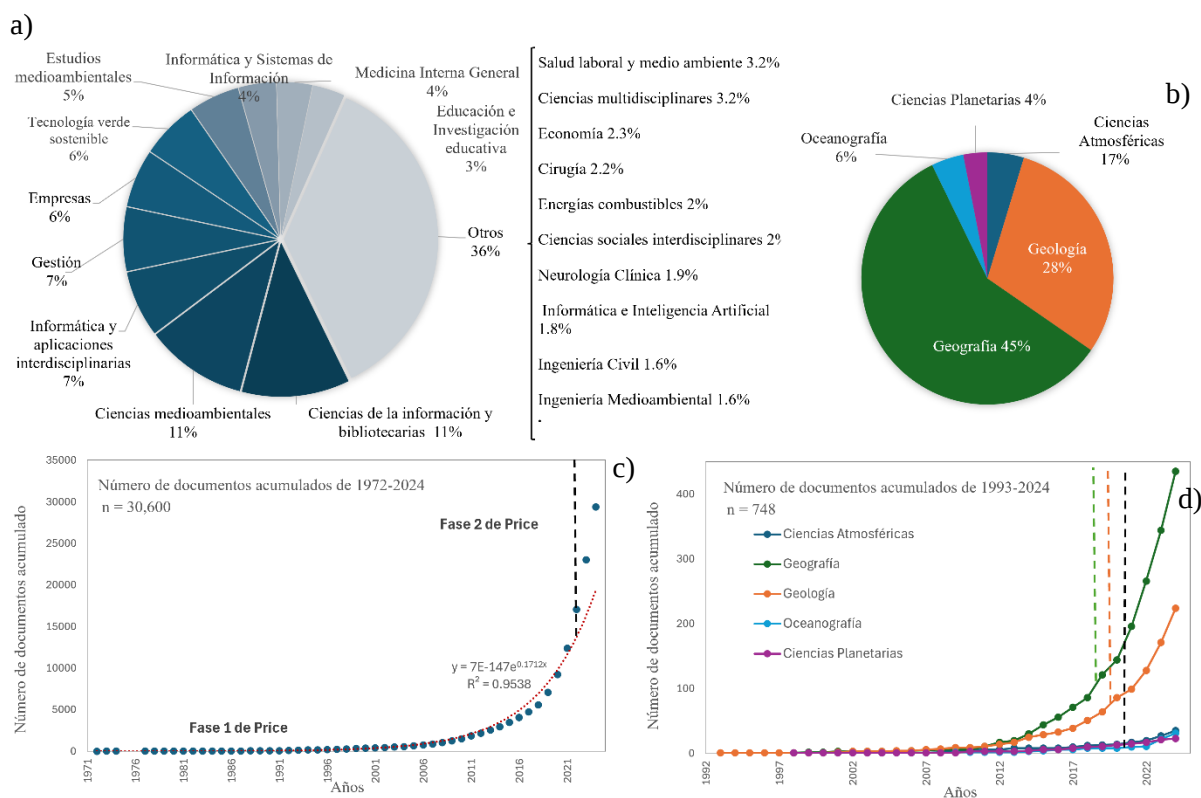


Figura 1. Conocimiento científico mapeado: a) análisis bibliométricos reportados en Web of Science para todas las áreas de conocimiento b) BA reportados en WoS para geociencias; c) análisis del comportamiento en la aplicación de los BA que describe la correlación entre el número de documentos publicados por año de todas las áreas de investigación; d) análisis de la productividad científica de los BA aplicada a geociencias.

1.2 Antecedentes

El análisis bibliométrico actualmente es un campo multidisciplinario que ha crecido de manera significativa en los últimos años, y se ha consolidado como una herramienta clave para evaluar la producción científica y académica. El BA inicia a través de la creación de métricas bibliométricas, como el *índice de colaboración*, el *coeficiente de colaboración*, el *índice h*, entre otros (Wallin, 2005). Así como la aplicación de métodos estadísticos descriptivos y multivariantes que han sido de ayuda al momento de medir el desempeño y la estructura conceptual, intelectual y social de las publicaciones científicas, autores y revistas. Por lo tanto, en esta sección se describe de manera general algunos antecedentes más relevantes en la creación de las metodologías de BA más recientes. Bajo este mismo argumento Garfield (1955) propuso por primera vez la creación de un índice de citas para la ciencia basándose en Shepard's Citations, y sugirió un sistema que permitiera consultar las métricas de un artículo. Garfield (1955) creó así el Instituto de Información Científica (ISI), así como las bases de datos que componen actualmente la Web of Science (WoS), publicada por Thomson Scientific (Obtenido de Oliveira Lucas et al., 2014).

Small et al., (1977) describió un algoritmo que lleva a cabo una clasificación de títulos utilizando técnicas estadísticas multivariantes (single-link clustering technique), que consiste en generar agrupamientos o clústeres bibliográficos. Este algoritmo se utilizó en el análisis de 890 revistas que se agruparon en 168 clústeres. Los resultados indican que este método de agrupación de revistas tiene una utilidad práctica significativa para la clasificación de estas.

Hirsch (2005) propuso un índice de cuantificación de la literatura denominándolo “*índice h*” útil para caracterizar, estimar la importancia, la trascendencia y la producción científica de un investigador o investigadora. Hirsch utilizó la plataforma Web of Science para encontrar los artículos citados y propuso una ecuación matemática de proporcionalidad para el número total de citas: $N_{c,tot} = ah^2$ (donde a = es la constante de proporcionalidad que se define entre 3 y 5, h = índice de Hirsch). Finalmente se concluyó que este índice h puede proporcionar un criterio útil con el que comparar de forma imparcial, a diferentes individuos que compiten por el mismo recurso.

Debido a lo anteriormente mencionado surge la necesidad e importancia del análisis bibliométrico el cual tiene la capacidad de poder cuantificar el impacto y la visibilidad de los

trabajos científico - académicos, facilitando la toma de decisiones para la evaluación del rendimiento de la investigación, la selección de publicaciones relevantes y la medición de la productividad científica de un autor(a), una institución o país en particular, permitiendo identificar tendencias para continuar con el desarrollo de la investigación científica en áreas de ciencia básica y/o de frontera. De manera paralela al crecimiento de esta ciencia bibliométrica, se ha visto la necesidad de desarrollar nuevos softwares especializados para transformar y facilitar la forma de realizar estos análisis. Estos programas permiten gestionar y procesar grandes volúmenes considerables de metadatos bibliográficos de manera más fácil, eficiente y precisa. Estas herramientas no solo facilitan la visualización de redes de citas y coautoría, sino que también ofrecen funciones avanzadas para el análisis de temas a través de técnicas utilizadas en la minería de textos y el análisis de redes (Moral Muñoz et al., 2020), es este contexto: Chiu et al., (2007) realizaron un análisis bibliométrico acerca de las publicaciones relacionadas con el tsunami ocurrido en Indonesia el 26 de diciembre del 2004, utilizando la base de datos Science Citation Index (SCI) para analizar los parámetros del tipo de documento, idioma, resultado y patrones de publicación, autoría, distribución de categoría temática, palabras clave, país de publicación, artículo citado con mayor frecuencia y la distribución del documento. Obteniendo como resultados que Estados Unidos y Japón produjeron el 53% de la producción total y el inglés fue el idioma dominante, con el 95% de los artículos.

Grawin Sebastián, (2012) diseñó dos de los primeros softwares, el primero de ellos denominado Bibliotools con la finalidad de poder crear mapas de la ciencia basados en datos bibliográficos, utilizó diferentes códigos Python que permiten realizar minería, procesamiento, análisis, visualización y reporte de datos; el segundo software nombrado BiblioMaps que tiene la finalidad de ayudar a sus usuarios a explorar la naturaleza del corpus estudiado, así como varios mapas temáticos y de co-ocurrencia, el cual contiene un conjunto de códigos html/css/js que generan una interfaz de visualización basada en la web.

Liu et al., (2012) realizaron un estudio bibliométrico sobre terremotos en todo el mundo durante 1900-2010 donde se evaluó el rendimiento de la investigación sobre terremotos a partir de un análisis bibliométrico de 84,051 documentos publicados en revistas y otros medios de difusión, utilizando las bases de datos bibliográficas Scientific Citation Index (SCI) y Social Science Citation Index (SSCI). Se resumieron los indicadores de publicación significativos en la investigación sobre terremotos, así mismo se evaluaron los resultados de la investigación

nacional e institucional y se presentó el desarrollo de la investigación sobre terremotos desde una perspectiva complementaria. Los resultados sugirieron un sólido desarrollo de la investigación sobre terremotos, en términos de aumento de la producción científica y de la colaboración en la investigación.

Wu et al., (2015) realizaron un análisis bibliométrico para estudiar las tendencias globales de investigación sobre deslizamientos de tierra durante 1991-2014 a partir de las bases de datos Science Citation Index-Expanded y Social Sciences Citation Index (SSCI), donde se analizaron 10,567 artículos relacionados. El análisis reveló categorías científicas, títulos de las fuentes, distribución geográfica global de los autores, autores productivos, colaboraciones internacionales, instituciones y la evolución de las frecuencias de palabras clave. Obtuvieron como resultados que La Academia China de Ciencias, el Servicio Geológico de los Estados Unidos y el Consejo Nacional de Investigación de Italia fueron las tres principales instituciones contribuyentes. Además, se encontró que la evaluación de la susceptibilidad a los deslizamientos de tierra, la estabilidad de los deslizamientos inducidos por lluvias y terremotos, fueron los temas que atrajeron la mayor atención.

Gariano et al., (2016) revisaron artículos publicados que abarcan el impacto pasado, presente y futuro esperado del cambio climático en los deslizamientos de tierra, la revisión se basó en algunos trabajos previos, y en particular en los trabajos de (Crozier, 2010). Examinaron 103 artículos de manera manual y determinaron las ventajas y los límites de los enfoques para evaluar los efectos de las variaciones climáticas en los deslizamientos de tierra, la revisión indicó que los resultados de la modelación de los estudios climáticos de deslizamientos de tierra dependen más de los escenarios de emisiones, los Modelos de Circulación Global y los métodos para reducir la escala de las variables climáticas, además, encontraron un sesgo en su distribución geográfica, con grandes partes del mundo no investigadas.

Emmer, (2018) estudió la investigación sobre amenazas naturales en todo el mundo durante 1900-2017 a partir del análisis de la base de datos Web of Science™. Se analizó un conjunto de 588,424 artículos de investigación, que abarcan diferentes tipos de amenazas naturales; así mismo se analizó la investigación sobre los diferentes tipos de amenazas naturales en el tiempo, el enfoque geográfico de la investigación mediante el análisis de contenido de títulos y resúmenes de los cuales las inundaciones, los terremotos, las tormentas y las sequías son los

tipos de amenazas naturales más frecuentemente investigados. Concluyendo que más de la mitad de todos los artículos de investigación se han publicado desde 2010 hasta 2017.

Bezak et al., (2021) realizaron un análisis bibliométrico acerca de la erosión de los suelos; se evaluaron 3,030 aplicaciones de modelos publicadas en 1,697 artículos durante el período de 1994-2018. Se utilizó el software VOSviewer para el análisis de citas, análisis de colaboración, coautoría y acoplamiento bibliográfico entre países. Se utilizó la modelización generalizada de árboles de regresión potenciada (BRT) para evaluar las variables más relevantes relacionadas con la modelización de la erosión del suelo, también, se analizaron y visualizaron redes bibliométricas. Como resultado, se creó la base de datos 'Global Applications of Soil Erosion Modelling Tracker (GASEMT) y se obtuvo que la selección del modelo de erosión del suelo tiene el mayor impacto en el número de citas de publicaciones, seguido de la escala de modelización y el CiteScore de la publicación.

Yáñez-Dávila et al., (2023) realizaron un análisis bibliométrico sobre los avances de investigación pasados y presentes realizados en experimentos de interacción agua-roca aplicados a la geotérmica desde 1963 hasta 2022, utilizaron las bases de datos Web of Science y Scopus para la recopilación de artículos publicados. Dentro de los resultados más relevante se identificaron las revistas mayormente citadas, así como los liderazgos de investigación (autores, grupos de redes, instituciones y países) y tendencias de investigación futuras. Finalmente, concluyeron en que estos análisis beneficiarán a la comunidad geotérmica mundial al proporcionar una visión más profunda de los experimentos de laboratorio de interacción agua-roca realizados hasta la fecha.

En conclusión, la producción de información científica y académica crece de manera exponencial como se pudo observar en la figura1-c del capítulo anterior, este crecimiento exponencial limita estar actualizado en el tema de investigación por la gran cantidad de información científica que se genera. Por lo tanto, los análisis bibliométricos son importantes para sintetizar y evaluar de manera objetiva y precisa en el impacto científico; además de medir la relevancia de investigaciones, áreas emergentes e investigadores influyentes. Continuar con estos BA es crucial para facilitar el acceso al conocimiento científico y académico alrededor del mundo, promover la colaboración global y fomentar el progreso científico.

1.3 Justificación

En un contexto de desarrollo constante en las formas de publicación científica, el análisis bibliométrico es clave en el mapeo del conocimiento científico acumulado. En el caso de las geociencias el BA es un enfoque poco explorado, lo que lleva a un vacío de información y/o datos que dificultan la visualización clara de las áreas de mayor impacto y los avances relevantes a través de los años en este campo. Por esta razón es que el BA puede ser una herramienta de investigación de gran relevancia para la comunidad geocientífica, debido a que permite verificar, sistematizar y organizar la información de los avances geocientíficos de manera eficiente. Esta investigación busca aplicar herramientas de análisis bibliométricas que organice, clasifique y verifique la información existente en análisis bibliométricos aplicados a geociencias, proporcionando un recurso accesible y confiable para investigadores, académicos y estudiantes interesados en esta área en específico.

La relevancia de este estudio radica en la necesidad urgente de ordenar y cuantificar grandes volúmenes de información geocientífica disponible, lo cual puede impulsar investigaciones más eficaces y fundamentadas, a través de la minería de textos de estudios realizados a metadatos obtenidos en bases de datos citas indexadas, este análisis permite obtener el desempeño (PA-A por su acrónimo en inglés: *performance appraisal analysis*), la estructura conceptual, intelectual y social (CoInSo-A por sus acrónimos en inglés: *conceptual, intellectual and social analysis*), así como el mapeo de información reportada. Los resultados obtenidos mediante esta tesis mejorarán el acceso y la comprensión de los conocimientos geocientíficos acumulados, lo que puede contribuir significativamente a la toma de decisiones informadas en investigaciones futuras y a la creación de nuevas líneas de estudio. Además, se espera que los resultados de esta investigación tengan un impacto positivo en la sociedad al obtener una mejor organización del conocimiento en todos los sectores involucrados, tales como la gestión ambiental, la geología aplicada y la ingeniería. En última instancia, este trabajo pretende ser una herramienta útil para científicos, investigadores, académicos y estudiantes interesados en las geociencias, con el objetivo de optimizar y focalizar los esfuerzos en áreas de alta relevancia científica.

1.4 Objetivos

1.4.1 Objetivo general

Aplicar las herramientas de análisis bibliométrico para mapear el conocimiento científico acumulado dentro de las geociencias, con la finalidad de conocer el desarrollo actual de la investigación geocientífica, su impacto y relevancia en la ciencia.

1.4.2 Objetivos particulares

- (1) Realizar un extenso marco teórico referente a todos los detalles técnicos para realizar un análisis bibliométrico.
- (2) Aplicar una metodología integral en el análisis de estudios bibliométricos aplicados en las geociencias.
- (3) Realizar un análisis de la producción científica, redes de colaboración entre países, instituciones y autores a través del tiempo en las geociencias.
- (4) Evaluar la evolución de los temas y áreas de investigación dentro del área de estudio, detectando los últimos avances y desafíos de la investigación científica, así como cambios en las prioridades científicas y el surgimiento de nuevas líneas de estudio.
- (5) Reconocer la importancia del análisis bibliométrico como una herramienta estratégica en la investigación en geociencias, destacando su potencial para la toma de decisiones geocientífica y la planificación de futuras investigaciones.

CAPÍTULO II MARCO TEÓRICO

2.1 Análisis bibliométrico

El BA es usado como herramienta para resumir y sintetizar grandes volúmenes de datos publicados en la literatura científica a través del uso de la minería de textos. Este análisis permite obtener una mejor comprensión de los patrones y tendencias de la investigación actual (Yáñez-Dávila et al., 2023; Donthu et al., 2021; Martín-Martín et al., 2021). Se recomienda el BA como un primer enfoque para el acercamiento en el análisis crítico de la literatura científica para el desarrollo de cualquier tema de investigación a través del uso de bases de datos de citas indexadas como Web of Science, Scopus, Google Scholar, entre otras. A través del BA se obtiene el PA-A y el CoInSo-A de la investigación reportada en Documentos Publicados (DP) principalmente revisados por pares, con la finalidad de identificar temas de investigación básica, temas declinados, temas emergentes, temas de relevancia en la actualidad, liderazgos dentro de investigación (autores, grupos de redes, instituciones y países) así como brechas y desafíos en el conocimiento científico de futuras investigaciones que deben realizarse para la solución de los problemas presentes.

La técnica de minería de textos en los BA permite la identificación y extracción de patrones y relaciones interesantes del contenido textual de documentos transformándola en un lenguaje de fácil comprensión para los usuarios, este propósito lo logra mediante el uso de técnicas como la identificación de términos o conceptos, el agrupamiento de documentos similares (clustering), la categorización automática de textos y la visualización de colecciones. Todo proceso bibliométrico es considerado minería de textos (Romero Pérez et al., 2022).

2.1.1 Bases de datos bibliométricas

En los BA es de gran importancia buscar una fuente de datos segura y concisa que se ajuste al área de investigación a desarrollar, estas bases son fundamentales para investigadores, instituciones académicas y organismos financiadores (Moral Muñoz et al., 2020). Es importante destacar que existen diferentes bases de datos de citas donde cada una de ellas proporciona información diferente que puede o no permitir realizar fácilmente el análisis bibliométrico (ver Tabla 1).

Estas bases de datos de citas son plataformas digitales que recopilan y organizan información sobre publicaciones científicas, autores, citas y otros aspectos clave de la producción científico-académica. Entre las bases de datos de citas indexadas más utilizadas se encuentran Web of Science y Scopus, que ofrecen datos que permiten realizar estudios de citas, análisis de tendencias, identificación de redes de colaboración, entre otros (Martín Martín et al., 2021). Por otra parte, cabe destacar el uso de la base de información de Google Scholar por ser también una de las más utilizadas, sin embargo, esta es una base de información es definida como literatura gris por contener información no revisada por pares.

Tabla 1. Recopilación de bases de datos bibliométricas indexadas y no indexadas de uso común dentro de la comunidad científica (Lingappa Sangam et al., 2012; Moral-Muñoz et al., 2020).

Bases de datos de citas para análisis bibliométricos					
Nombre de la base de datos	Tipo de suscripción	Registro de Publicaciones	Revistas revisadas por pares	Formatos descargables	URL
Web of Science	Suscripción	90 millones	Registro completo	Endnote, Plain text, RIS format, BibTex, Excel, Tab delimited, Printable HTML	https://clarivate.com/
Scopus	Suscripción	69 millones	Registro completo	Endnote, Mendeley, ExLibris, SciVal, RIS format, CSV excel, BibTex, Plain Text	https://www.scopus.com/home.uri
Google Scholar	Sin suscripción	389 millones	Registro parcial	BibTex, Endnote, RefMan, RefWords	https://scholar.google.es/schhp?hl=es
Microsoft Academic	Sin suscripción	230 millones	Registro parcial	BibTex	https://www.microsoft.com/en-us/research/project/academic/
Dimensions	Sin y con suscripción	102 millones	Registro adicional	BibTex, RIS format	https://www.dimensions.com/
Crossref	Sin y con suscripción	120 millones	Registro adicional	BibText, Metadata as JSON	https://www.crossref.org/

OpenCitations' (COCI)	Sin suscripción	72 millones	Registro adicional	CSV	https://opencitations.net/
-----------------------	-----------------	-------------	--------------------	-----	---

2.1.2 Método de revisión por pares

La revisión por pares es un proceso científico complejo y riguroso, de gran relevancia para la comunidad científica, utilizado para validar trabajos escritos por expertos en diferentes temas de investigación. Este método es una herramienta importante para la evaluación crítica de manuscritos enviados a revistas de carácter científico, donde son revisados por especialistas externos al equipo editorial, con el objetivo de evaluar su calidad, viabilidad y rigor científico. Su propósito es asegurar que las investigaciones que están próximas a ser publicadas cumplan con altos estándares de calidad y credibilidad (Ladrón De Guevara Cervera et al., 2008).

La revisión por pares ayuda a los editores a seleccionar los mejores manuscritos para sus publicaciones, además de proporcionar apoyo a autores y editores en la mejora de los artículos. Durante la revisión por pares, expertos independientes evalúan los textos, proponiendo correcciones que consideran necesarias. Estas observaciones permiten a los autores mejorar sus trabajos, al recibir críticas constructivas y señalamientos sobre posibles errores o áreas de mejora en sus investigaciones. (Ladrón De Guevara Cervera et al., 2008).

Se sugiere que el BA utilice bases de datos indexadas que contienen DP revisados por pares, con la finalidad de que la información obtenida mediante este análisis este sustentada por información previamente respaldada por expertos en los diferentes temas de investigación.

2.1.3 Literatura gris

La literatura gris, también conocida como “no convencional”, “semi-publicada” e “informal” es un tipo de documento que presenta resultados de investigaciones originales, pero que no se distribuye a través de los canales formales de publicación científica, como las revistas indexadas con revisión por pares. Aunque aporta conocimiento valioso, su difusión es limitada. Ejemplos de literatura gris incluyen:

- Tesis y tesinas de pregrado y posgrado.
- Memorias de congresos (en extenso y, en resumen).
- Preprints (documentos preliminares antes de la revisión formal).

- Artículos en revistas no indexadas o sin un proceso de arbitraje riguroso.

Estos documentos suelen carecer de visibilidad y no reciben el mismo nivel de reconocimiento o validación que la literatura publicada en revistas de carácter científico (Mendoza Núñez, 2022). Por otro lado, las bases de información de literatura gris más consultadas actualmente es Google Scholar y Microsoft Academic, por lo que se sugiere ser cuidadoso en la información obtenida de estas bases para realizar el BA.

2.2 Métricas bibliométricas

Las métricas bibliométricas son herramientas cuantitativas utilizadas en el BA para evaluar la producción y el impacto de las investigaciones científicas (ver Tabla 2). En consecuencia, estas métricas o indicadores bibliométricos se basan en el análisis de publicaciones científicas, citas y otros indicadores, con el objetivo de medir el alcance, la relevancia y la calidad de las contribuciones científicas en diferentes disciplinas. Las métricas bibliométricas (métricas desempeño, métricas de la estructura conceptual, intelectual, social y leyes bibliométricas) surgen de la necesidad de evaluar el rendimiento de la actividad científica y su impacto en la sociedad con el fin primordial de adecuar convenientemente la asignación de los recursos destinados a investigación y desarrollo. En síntesis, estas métricas han evolucionado como un componente clave para medir el desempeño en la ciencia y en el aspecto académico (González De Dios et al., 1997).

2.2.1 Métricas usadas en el análisis de desempeño científico

Índice de colaboración (IC)

Indicador bibliométrico que mide el grado de colaboración entre autores en la producción de publicaciones científicas. Se calcula como la proporción de artículos científicos en coautoría en relación con el total de artículos publicados por un investigador/investigadora o grupo. Es decir, refleja qué porcentaje de las publicaciones de un investigador o investigadora se realizaron en colaboración con otros autores.

Este índice puede revelar tendencias como el aumento de la colaboración interdisciplinaria y global, además de proporcionar una idea de cómo las redes de investigación están estructuradas en un área científica determinada (Universidad Distrital Francisco José de Caldas, 2023).

Coefficiente de colaboración (CC)

El CC es formulado por Ajiferuke et al., (1988) quien incorpora las propuestas de Lawani, (1980) y Subramanyam, (1983) sobre el Índice de Colaboración (IC) como el número medio de autores por artículo y el Grado de Colaboración (GC) definido como la tasa del número de artículos de investigación colaborados en relación con el número total de artículos de investigación publicados en la disciplina durante un periodo determinado, respectivamente. El coeficiente de colaboración está basado en la idea de la atribución fraccional de productividad en la producción de documentos es estimada por la siguiente ecuación:

$$CC = 1 \frac{\sum_{j=1}^K \left(\frac{1}{j}\right) f_j}{N} \quad \text{Ecuación 1}$$

Donde:

f_j = representa el número de artículos de investigación con j-autores múltiples publicados durante un determinado periodo.

N = número total de artículos de investigación publicados en la disciplina objeto de estudio durante un determinado periodo.

k = Representa el mayor número de autores por artículo en la disciplina objeto de estudio.

Estas tres variables son utilizadas para evaluar la producción científica de colaboración entre los autores en estudios bibliométricos (Alvarado Urbizagastegui R., 2011).

Índice h

El índice h también conocido como índice de Hirsch es un indicador bibliométrico con el cual se mide simultáneamente la calidad y la cantidad de la producción científica principalmente de un investigador o investigadora. Este indicador está basado en el número de citas que han recibido sus publicaciones, proporcionando así una evaluación equilibrada entre productividad e impacto (Universidad de Las Palmas de Gran Canaria, 2024).

El índice h se calcula ordenando las publicaciones de un autor(a) de mayor a menor con base en el número de citas recibidas. El índice h es el número máximo tal que el autor(a) tiene al menos h publicaciones con al menos h citas cada una. Por ejemplo, un índice $h = 7$ significa que el investigador/investigadora tiene 7 publicaciones que han sido citadas al menos 7 veces cada una. Este indicador es útil para detectar a los investigadores más destacados en un campo, debido

a que refleja tanto su productividad a lo largo del tiempo como el impacto de sus publicaciones en la comunidad científica (Universidad de Las Palmas de Gran Canaria, 2024).

El índice h se utiliza como herramienta principal para evaluar calidad y la cantidad de la producción científica en las principales bases de citas indexadas tales como WoS y Scopus. El índice h también es usado para evaluar calidad y la cantidad de la producción científica de revistas científicas o países, utilizadas principalmente por portal web de SCImago Journal & Country Rank (SJR).

Índice g

El índice g es un indicador bibliométrico que, al igual que el índice h , mide la productividad científica de un investigador o investigadora y el impacto de sus publicaciones. A diferencia del índice h , el índice g ofrece una mayor diferenciación entre investigadores que tienen un índice h similar, ya que este índice tiene en cuenta tanto la cantidad de publicaciones como el número total de citas recibidas por estas publicaciones (Universidad de Las Palmas de Gran Canaria, 2023).

El cálculo del índice g se realiza teniendo en cuenta las siguientes consideraciones:

1. Las publicaciones del autor(a) se ordenan en función del número de citas recibidas, de mayor a menor.
2. Se crea una columna con el número acumulado de citas.
3. Se genera otra columna con el número de la posición al cuadrado.
4. Por lo tanto, el índice g corresponde a la posición donde el número acumulado de citas es igual o mayor que el cuadrado de la posición.

Por ejemplo, si un investigador tiene un índice $g = 15$, significa que sus 15 publicaciones más citadas acumulan un número de citas superior a 15 al cuadrado, es decir, más de 225 citas en total.

El índice g es más sensible a los artículos altamente citados en comparación con el índice h , ya que da mayor peso a las publicaciones que reciben un número significativo de citas. Actualmente, este indicador es calculado principalmente por la aplicación Publish or Perish, que utiliza los datos de Google Scholar para determinar las citas (Universidad de Las Palmas de Gran Canaria, 2024).

Índice i

Este índice mejora el índice h previamente propuesto, ya que considera el número de autores de cada artículo y el porcentaje de auto citas del investigador/investigadora. Para calcular el índice i , primero se establece un índice pre- i , de manera similar al índice h , un investigador tiene un “índice pre- i ” si “ i ” de sus artículos tiene un número de “ i ” citas, tras dividir el número de citas de cada artículo por el número de coautores. Una vez determinado cuántos artículos cumplen con esta condición, se multiplica este número por el logaritmo ($\log_{10}$) de la proporción de alocitas (las alocitas son el total de citas recibidas menos las citas de artículos en los que el autor(a) principal es el mismo investigador). Además, se introduce una versión estandarizada, donde el índice i puede ajustarse según la “edad científica” de cada investigador (que corresponde a los años transcurridos desde la primera publicación), denominada índice i_s (Molina Montenegro et al., 2010).

2.2.2 Métricas usadas en el análisis de la estructura CoInSo

Clustering multivariable

Es una herramienta de aprendizaje automático no supervisado que tiene como objetivo principal crear clústeres temáticos o sociales. Dependiendo del tipo de análisis a utilizar (análisis factorial exploratorio, la agrupación jerárquica, el algoritmo de Island, el método de Lovaina, el escalamiento multidimensional y el algoritmo de centros simples) este puede ser útil para comprender cómo se manifiesta y desarrolla un campo de investigación. Si se utiliza, por ejemplo, los conglomerados temáticos creados mediante el análisis de co-citación y el acoplamiento bibliográfico, se puede saber el desarrollo a lo largo del tiempo en el campo de investigación (Donthu et al., 2021).

Red de co-ocurrencia

La co-ocurrencia de palabras se refiere a la presencia simultánea de dos términos dentro de un cuerpo del artículo. Esta técnica se utiliza para analizar el contenido de un texto científico mediante el estudio de la aparición conjunta de pares de palabras o términos, lo que facilita la identificación de relaciones entre conceptos en un área específica. A medida que aumenta la frecuencia con la que dos palabras aparecen juntas, se refuerza su conexión conceptual (Boeris, 2012).

Análisis factorial

Es una técnica estadística de interdependencia en la que no se distinguen variables dependientes o independientes, como ocurre en los modelos de regresión. En lugar de ello, todas las variables se analizan simultáneamente. Es sumamente versátil y su propósito principal es identificar una estructura subyacente entre las variables del análisis, basándose en las correlaciones entre ellas. En otras palabras, busca agrupar variables muy correlacionadas, conocidos como "factores". Además, esta técnica permite reducir la complejidad de un gran número de variables en un conjunto más pequeño, facilitando una explicación más precisa del fenómeno en estudio (Méndez Martínez et al., 2012).

2.2.3 Leyes bibliométricas

Ley de Bradford

La Ley de Bradford establece una relación entre la productividad de las revistas científicas y el número de artículos que publican sobre un tema específico. Según esta ley, si las revistas se ordenan de forma decreciente por el número de artículos publicados en un campo, pueden dividirse en un "núcleo de Bradford", que incluye las revistas más productivas sobre ese tema, y varias zonas o grupos sucesivos. Cada grupo contiene aproximadamente el mismo número de artículos que el núcleo, pero el número de revistas aumenta en una progresión geométrica, con una relación de $1: n: n^2$ (Ardanuy, 2012).

El cálculo para estas zonas se realiza normalmente de manera gráfica. Se utiliza una escala semilogarítmica para representar el número acumulado de títulos (revistas) por orden decreciente frente al número acumulado de artículos. El punto en que la curva se convierte en una línea recta indica el tamaño del núcleo de Bradford (Ardanuy, 2012).

En la práctica, muchas veces no es necesario realizar un cálculo preciso de las zonas. En lugar de eso, se determina el núcleo dividiendo las revistas por orden decreciente según el número de artículos, accesos o citas y calculando el núcleo a partir del 50% de los artículos acumulados. Esto simplifica el análisis de la dispersión de las publicaciones científicas (Ardanuy, 2012).

Ley de Price

Derek J. de Solla Price determinó en 1956 que el crecimiento de la información científica era exponencial y se producía a un ritmo tan rápido que en periodos de 10 a 15 años la información

global sería el doble de la existente después de este periodo. Sin embargo, cada disciplina sufre su evolución propia, pasando por diversas etapas:

- ✓ Crecimiento lineal (investigadores precursores): primeras publicaciones en un campo de investigación.
- ✓ Crecimiento exponencial: el campo de investigación se convierte en un frente científico.
- ✓ Crecimiento asintótico: el crecimiento del campo de investigación se ralentiza, las publicaciones en esta etapa tienen como finalidad primordial la revisión y el archivo de conocimiento.
- ✓ Zona de decrecimiento (envejecimiento de la investigación): en esta etapa el número de publicaciones referente al campo de investigación comienza a declinar, puesto que aparentemente ya no se genera nueva información dentro del campo.

También es importante recalcar que la literatura científica pierde actualidad cada vez más rápidamente, sin embargo, el envejecimiento de la bibliografía no es uniforme para todas las disciplinas científicas. Así, en las ciencias experimentales y en la tecnología el rápido crecimiento de la bibliografía conlleva también un rápido envejecimiento de esta, mientras que, en las humanidades es mucho más lento. Incluso dentro de las ciencias experimentales, las diferencias pueden ser significativas. Por ejemplo, la obsolescencia en bioquímica es mucho más rápida que en botánica (Ardanuy, 2012).

Ley de Pareto

La ley de Pareto también conocida como la ley 80/20. Estima que el 80% de las publicaciones, por ejemplo, número de artículos y citas lo producen el 20% de las fuentes, por ejemplo, revista y autor; (Çiftçi et al., 2016). Esta ley resulta especialmente útil en la organización de documentos para optimizar la información bibliográfica, al asignar de manera más eficiente los recursos de información y aumentar el valor de los documentos clave (Arsenova, 2013).

Esta ley puede aplicarse en el análisis bibliométrico de diversas maneras, como son:

- ✓ Gestión general de documentos.
- ✓ Construcción de recursos de artículos y documentos.
- ✓ Recolección de información prioritaria.

Ley de Leimkuhler

La ley de Leimkuhler fue desarrollada a partir de la formulación verbal de la Ley de Bradford (ecuación 2), a la cual se le denominó “distribución de Bradford”; (Chiroque, 2009). Esta ley bibliométrica suele utilizarse en el estudio de las necesidades de los lectores, la productividad y el precio de las revistas, ya que aporta información sobre la selección de documentos y ayuda eficazmente al análisis bibliométrico en la extracción de un libro y las actividades de recopilación (Arsenova, 2013).

$$R(r) = a \log(1 + br) \quad \text{Ecuación 2}$$

Donde:

$R(r)$ = es el número acumulado de ítems producidos por las r fuentes de mayor productividad.

a y b = son parámetros que varían de acuerdo con la temática estudiada.

$\log$ = es el logaritmo natural.

Ley de Lotka

Esta ley describe una relación cuantitativa entre los autores y las contribuciones producidas en un campo dado a lo largo de un periodo de tiempo. Esta distribución de probabilidades discreta resulta desigual puesto que la mayor parte de los artículos proceden de una pequeña porción de autores altamente productivos. Esta ley fue propuesta por Lotka en 1926 como una ley de cuadrado inverso actualmente se expresa de manera más general en forma de potencia inversa generalizada (Ardanuy, 2012).

Originalmente se expresa como:

$$A(x) = \frac{A_0}{x^2} \quad \text{Ecuación 3}$$

Donde:

$A(x)$ representa el número de autores con x trabajos publicados.

A_0 es el número de autores con 1 trabajo publicado.

La expresión actual:

$$A(x) = \frac{c}{x^n}$$

Ecuación 4

donde C y n deben ser estimados a partir de los datos observados.

Ley de Zipf

La Ley de Zipf clasifica las palabras por su frecuencia, desde las más comunes hasta las menos frecuentes, utilizando métodos estadísticos. Según esta ley, el producto de la frecuencia (f) y el rango (r) de las palabras tiende a ser aproximadamente constante. Aunque esta ley es compleja en su aplicación bibliométrica, demuestra la relación entre el valor del ranking de repetición y frecuencia de las palabras. Su investigación sugiere que los rangos comunes (r') son más efectivos que los valores extremos esperados, tanto los altos como los bajos (Çiftçi et al., 2016).

Tabla 2. Métricas de desempeño (PA-A), estructura conceptual, intelectual y social (CoInSo-A) y leyes bibliométricas utilizadas para el mapeo científico de información (Lingappa Sangam et al., 2012; Moral-Muñoz et al., 2020).

Métricas bibliométricas		
Análisis de desempeño		Análisis de la CoInSo
Métricas de cantidad	Métricas de calidad	Métricas análisis multivariable
Panorama general		Clustering
Publicación total	Cita total	Clúster por acoplamiento
Producción científica anual	Promedio de citas por año	Acoplamiento bibliográfico
Fuentes		Estructura conceptual
Publicaciones de los países	Países más citados	Red de co-ocurrencia
Publicaciones de afiliaciones	Afiliaciones más relevantes	Mapa temático
Publicaciones de Revistas	Revista más relevante	Evolución temática
Autores		Análisis factorial
Número de autores colaboradores	Índice de colaboración	Estructura intelectual
Publicación en coautoría	Coficiente de colaboración	Análisis de co-citación
Publicaciones de único autor	Producción del autor(a) a lo largo del tiempo	Análisis e palabras conjuntas
	Índice h	Análisis de coautoría
	Índice g	Historiografía

Número de años en activo para la publicación
Productividad por años activos de publicación

Número de publicaciones citadas
Proporción de publicaciones citadas
Citación por publicación citada

Red de colaboración
Mapa mundial de la colaboración

Leyes bibliométricas

Ley de Bradford,
Ley de Price,
Ley de Pareto,
Ley de Leimkuler,
Ley de Lotka,
Ley de Zipf.

2.2.4 Software bibliométrico

Un software bibliométrico es una herramienta informática diseñada para analizar la producción científica y académica a través de datos cuantitativos extraídos de publicaciones, como artículos, libros o patentes. Este tipo de software permite realizar estudios bibliométricos (ver Tabla 3). Estos programas permiten procesar grandes volúmenes de información proveniente de bases de datos de citas indexadas. Entre los softwares más destacados se encuentran VOSviewer, Bibliometrix y CiteSpace, que ofrecen fuertes capacidades de análisis y representación gráfica de los datos, lo que permite a los académicos, científicos e instituciones académicas obtener una comprensión más profunda de las tendencias científicas e identificar líderes en áreas de conocimiento específicas (Moral Muñoz et al., 2020).

Biblioshiny

Aplicación que proporciona una interfaz web para Bibliometrix, apoya en el uso fácil de las principales características de bibliometrix como son (Massimo Aria et al., 2016):

- ✓ Importación de datos y conversión a la recopilación de marcos de datos.
- ✓ Recopilación de datos mediante la recopilación de APIs de Dimensions, PubMed y Scopus.
- ✓ Filtrado de datos.

Análisis y gráficos para cuatro métricas de nivel diferentes:

- ✓ Fuentes.
- ✓ Autores.
- ✓ Documentos.
- ✓ Agrupación por acoplamiento.

Análisis de tres estructuras del Conocimiento:

- ✓ Estructura conceptual.
- ✓ Estructura intelectual.
- ✓ Estructura social.

Blibiometrix

Herramienta de código abierto diseñada para el análisis cuantitativo en cienciometría y bibliometría. Diseñado en lenguaje de programación R que permite realizar estudios completos sobre grandes volúmenes de datos bibliográficos (Massimo Aria et al., 2016).

Entre sus principales funciones, se destacan las siguientes:

- ✓ Importación de datos bibliográficos: puede extraer información de bases de datos científicas como Scopus, Web of Science, PubMed, Dimensions y Cochrane.
- ✓ Análisis bibliométrico: incluye métodos para estudiar patrones de citación, co-citación, acoplamiento bibliográfico y análisis de co-palabras.
- ✓ Construcción de matrices de datos: permite crear matrices de datos para estudiar la colaboración científica, analizar redes de citación y examinar el uso de palabras clave en publicaciones.
- ✓ Extensibilidad: al estar desarrollado en R, es fácil de personalizar y expandir, lo que permite crear nuevas funciones y automatizar análisis específicos.

Sciento-Text

Es un paquete de R utilizado para realizar análisis bibliométricos y cienciométricos a partir de datos obtenidos de bases de citas indexadas basados en indicadores. Funciona con datos de WoS y Scopus. Es capaz de obtener diferentes análisis, como matriz de coautoría, principales autores/países/instituciones con sus índices g y h y sus citas respectivas, autores altamente citados, colaboración internacional y frecuencia de términos (Moral Muñoz et al., 2020).

SciTool

Es un conjunto de herramientas modulares diseñado específicamente para realizar estudios cienciométricos. Admite análisis temporales, geoespaciales, temáticos y de redes, así como la visualización de conjuntos de datos a nivel micro (individual), meso (local) y macro (global).

Fue desarrollado por el Cyberinfrastructure for Network Science Center de la Universidad de Indiana (EE. UU.) y es de libre acceso (Cobo et al., 2011).

Bibliotools/BiblioMaps

Son un conjunto de scripts para realizar mapeos cuantitativos basado en datos obtenidos de bases de datos indexadas, ejecutadas bajo Python, por su parte BiblioMaps es una interfaz perteneciente a BiblioTools que está diseñado para no requerir Python y sea más manejable, Además puede trabajar con datos de WoS y Scopus. Este conjunto de scripts realiza las diferentes tareas como son la adquisición, preprocesado, análisis, visualización e informe de datos. BiblioMaps sólo ofrece la red como opción de visualización, para obtener los mapas es necesario el uso de un software externo. Aunque el usuario puede realizar el análisis controlando cada etapa del proceso, esta herramienta ofrece el script «all_in_one», que permite realizar todos los análisis a la vez (Moral-Muñoz et al., 2020).

Metaknowledge

Paquete de Python para realizar análisis bibliométricos y cuantitativos basados en el análisis de desempeño. Trabaja principalmente con datos de WoS, PubMed, Scopus y ProQuest Dissertation and Theses. Realiza análisis longitudinal, espectroscopia de años de publicación estándar y multirreferencia, análisis computacional de textos y análisis de redes (Moral Muñoz et al., 2020).

SciMAT

Herramienta de software de código abierto diseñada para asistir todos los pasos del flujo de trabajo, incorporando todos los elementos necesarios (métodos, algoritmos y medidas) para obtener diferentes análisis y visualizaciones. Los mapas científicos obtenidos cuentan con medidas bibliométricas basadas en citas, como la suma, el máximo, el mínimo y el promedio de citas. Utiliza diferentes índices bibliométricos (h , g , i) y permite añadir archivos en formatos WoS y RIS. Como resultado, elabora un informe en formato HTML y LaTeX; las imágenes y diagramas los exporta en formatos PNG y SVG y las redes de clústeres y mapas de evolución se exportan en formato Pajek (Moral Muñoz et al., 2020).

ScientoPy

Herramienta de software utilizada para el análisis bibliométrico de grandes conjuntos de datos científicos. Está diseñada para analizar y visualizar las tendencias de investigación a lo largo del tiempo, permitiendo la identificación de temas emergentes, autores, instituciones y fuentes relevantes en diferentes áreas del conocimiento.

Esta herramienta ayuda a integrar problemas relacionados con las fuentes de Scopus, Clarivate y WoS, extrayendo y mostrando el índice h para el tema en análisis. Además, ofrece variedad de opciones para realizar análisis temporales de autores, instituciones, términos clave y temas emergentes, utilizando cuatro tipos de visualización diferentes. Gracias a su capacidad de análisis bibliométrico, esta herramienta es de gran importancia para realizar estudios futuros en distintos campos emergentes (Ruiz Rosero et al., 2019).

CITAN (Citation Analysis)

Paquete de herramientas que permite realizar análisis de rendimiento a partir de los datos de Scopus en una base de datos SQL, permite modificar información, como fusionar documentos o autores, eliminar duplicados o un documento o autor(a) específico a través de códigos. En general, puede calcular el índice h , el índice g , el índice i_p y otros indicadores generales de impacto. Además, proporciona gráficos descriptivos con los indicadores calculados (Moral Muñoz et al., 2020).

CitNetExplorer

Herramienta eficaz para visualizar y analizar redes de citas, basada en la historiografía. En las redes obtenidas por CitNetExplorer, cada nodo representa una publicación y cada arista representa una relación de citas entre dos publicaciones. Puede utilizar información de WoS y Scopus para construir redes de citas y una vez formada la red de citas, puede exportarse al formato de archivo Pajek. Está enfocado en dos temas principales: proporcionar datos de publicaciones, relacionar citas y añadir datos directos de WoS. Esta herramienta ofrece cuatro opciones diferentes: extraer componentes conectados, agrupar publicaciones, identificar publicaciones centrales y encontrar el camino más corto o largo desde una publicación a otras (Moral Muñoz et al., 2020).

Network Workbench Tool (NWB)

Conjunto de herramientas de análisis, modelado y visualización de redes, esta herramienta proporciona algoritmos específicos para tratar los datos de las publicaciones para construir, analizar redes y generar mapas bibliométricos, utiliza diferentes formatos de datos WoS, Scopus, Bibtex y EndNote Export Format. La herramienta NWB permite preprocesar datos, construir distintos tipos de redes, realizar un análisis gráfico de las redes construidas y visualizarlas (Cobo et al., 2011).

CiteSpace

Es una herramienta informática desarrollada para detectar, analizar y visualizar patrones y tendencias en la literatura científica. Su principal objetivo es facilitar el análisis de las tendencias emergentes en un área específica del conocimiento. CiteSpace puede leer diferentes formatos de fuentes bibliográficas, como WoS, PubMed, arXiv y SAO/NASA Astrophysics Data System (ADS). Además, CiteSpace es capaz de leer datos de subvenciones como NSF Awards y datos de patentes de Derwent Innovations Index (Cobo et al., 2011).

VOSviewer

Es una herramienta informática diseñada específicamente para construir y visualizar mapas bibliométricos, prestando especial atención a la representación gráfica de dichos mapas. Está diseñada principalmente para representar mapas de gran tamaño, ya que utiliza funciones como:

- ✓ Zoom.
- ✓ Algoritmos especiales de etiquetado.
- ✓ Metáforas de densidad.

Esta herramienta de software es de libre disposición para la comunidad de investigadores bibliométricos. Aunque VOSViewer puede utilizarse para construir y visualizar mapas bibliométricos de cualquier tipo de datos de co-ocurrencia, esta herramienta de software no permite extraer y construir ninguna matriz de co-ocurrencia a partir de los datos bibliométricos. Para esto, es necesario un procesador externo. Además, esta herramienta no dispone de módulos de preprocesamiento para preparar los datos para su posterior análisis (Cobo et al., 2011).

Bibexcel

Herramienta diseñada para crear archivos de datos que puedan importarse a Excel o a cualquier programa que tome registros de datos tabulados para su posterior manipulación, puede leer información recuperada de diversas fuentes bibliográficas, (WoS, Scopus y ProCite). Con los parámetros que ofrece esta herramienta se pueden obtener diferentes redes bibliométricas, entre las principales redes se encuentra la co-citación, acoplamiento bibliográfico, coautor y co-palabra. Además, es posible crear diferentes matrices de co-ocurrencia tomando cualquier campo del documento, o combinándolos; permite someter los datos a distintas tareas de preprocesamiento, duplicación de documentos y transformación de textos. El tipo de visualización seleccionado dependerá de la naturaleza de la unidad de análisis (Moral Muñoz et al., 2020).

IN-SPIRE

Herramienta de análisis de documentos que permite encontrar relaciones, tendencias y temas declinados en los datos para obtener nuevos conocimientos y representaciones; utiliza patrones estadísticos de palabras para caracterizar los documentos en función de su contexto, puede leer documentos sin formato (texto ASCII) o documentos con formato como HTML, .xml, .xlsx y .csv, realizando así cálculos para encontrar similitudes entre documentos y otros meta-campos como el título de documentos y la fecha, obteniendo como resultado un algoritmo de clustering llamado «Fast Divisive Clustering». Para realizar estos cálculos utiliza el modelo de espacio vectorial (Salton G. et al., 1983), por lo que cada documento se representa como un vector de términos. Así, si se seleccionan palabras clave como campo, la medida de similitud mostrará si dos documentos tienen palabras clave similares. Al final del proceso de agrupación, se generarán varios temas (conjuntos de documentos). Cada tema tiene como nombre los términos que aparecen con más frecuencia de sus documentos (Cobo et al., 2011).

CoPalRed

Software diseñado para realizar análisis de co-palabras a partir de palabras clave de documentos científicos, con esta información recolectada crea nuevo conocimiento; puede leer archivos en formato .csv y cuenta con un módulo de preprocesamiento que ayuda a normalizar las palabras clave de forma sencilla. Una vez unificadas las palabras clave, CoPalRed construye una matriz de co-ocurrencias y la normaliza utilizando el índice de equivalencia (Callon et al., 1991). Con

esta información realiza tres tipos de análisis: análisis estructural, análisis estratégico y análisis dinámico. Finalmente visualiza los resultados mediante diagramas estratégicos, temas y redes temáticas (Cobo et al., 2011).

Leydesdorff's Software

Conjunto de programas que funcionan con líneas de comandos y permiten realizar un mapeo científico de información con distintas funciones de análisis como co-palabras, co-autores, acoplamiento de autores y revistas y co-citación para realizar análisis bibliométricos; utiliza las plataformas de WoS, Scopus, y Google Scholar como bases de datos dando como resultado visualizaciones en formato Pajek, UNICET, Network Workbench y Sci2. El conjunto de programas no permite preprocesar los datos; así, por ejemplo, para realizar un análisis longitudinal se necesita un software externo para dividir los datos en diferentes periodos de tiempo (Cobo et al., 2011).

KeyWords Plus

Los datos de KeyWords Plus son palabras o frases que aparecen con frecuencia en los títulos de las referencias de un artículo, pero que no aparecen necesariamente en el título del artículo. Basado en un algoritmo especial que es exclusivo de las bases de datos de Clarivate, KeyWords Plus mejora el poder de la búsqueda de referencias citadas mediante la búsqueda en todas las disciplinas de todos los artículos que tienen referencias citadas en común. Debido a que casi todos los artículos de Web of Science contienen información bibliográfica en inglés, incluido el título, habrá pocos o ningún término de KeyWords Plus que no esté en inglés (Clarivate, 2022).

Tabla 3. Softwares bibliométricos utilizados para el mapeo y visualización de información (Lingappa Sangam et al., 2012; Moral-Muñoz et al., 2020).

Software bibliométrico						
<i>Software o librerías</i>	<i>Desarrollador</i>	<i>País</i>	<i>Sistema Operativo</i>	<i>Lenguaje de programación</i>	<i>Año</i>	<i>URL</i>
Biblioshiny	Universidad de Nápoles Federico II	Italia		R	2017	https://www.bibliometrix.org/home/index.php/layout/biblioshiny
Bibliometrix	Universidad de Nápoles Federico II	Italia		R	2017	https://www.bibliometrix.org/home/

Sciento-Text	Universidad del Cauca	Colombia		Python	2016	https://github.com/Scientotext/sciento-text
SciTool	Ciberinfraestructura para el Centro de Ciencias de Redes	EUA	Windows, OSX, Linux		2016	https://scitools.com/
Bibliotools/ BiblioMaps	Universidad de Lyon	Francia		Python	2012	https://www.sebastian-grauwin.com/bibliomaps/
Metaknowledge	Universidad de Waterloo	Canadá		Python	2017	https://metaknowledge.readthedocs.io/en/latest/
SciMAT	Universidad de Granada	España	Windows, OSX, Linux		2016	https://sci2s.ugr.es/scimat/
scientoPy	Universidad de Asia Meridional	India		R	2016	https://www.scientopy.com/ http://www.ibspan.waw.pl/gagolews/CITAN/
CITAN	Universidad de Deakin	Australia		R	2015	(página web inhabilitada)
CitNetExplorer	Universidad de Leiden	Países Bajos	Windows, OSX, Linux		2014	https://www.citnetexplorer.nl/
Network Workbench Tool	Universidad de Indiana	EUA	Windows, OSX, Linux		2011	https://netlab.gmu.edu/networkbench/
CiteSpace	Drexel University	EUA	Windows		2010	

					https://citespace.podia.com/
VOSviewer	Universidad de Leiden	Países Bajos	Windows, OSX, Linux	2010	https://www.vosviewer.com/
Bibexcel	Universidad de Umea	Suecia	Windows	2010	
IN-SPIRE	Laboratorio Nacional del Noroeste del Pacífico	EUA	Windows	2010	https://inspirehep.net/
CoPalRed	Universidad de Granada	España		2005	http://ec3.ugr.es/copalred/ (página web inhabilitada)
Leydesdorff's Software	Universidad de Ámsterdam	Países Bajos		N/A	https://www.leydesdorff.net/software/index.htm

2.2.5 Metodologías bibliométricas reportadas

A lo largo del tiempo, diversos autores han propuesto diferentes metodologías y enfoques para llevar a cabo BA, cada uno con sus propias características y aplicaciones (ver Tabla 4). El objetivo de estos autores en conjunto es desarrollar una referencia que sea única y significativa para los académicos, científicos y organizaciones interesados en utilizar métodos y/o análisis bibliométricos para mapear diversos campos de investigación (Zupic et al., 2014). El objetivo de esta sección es contrastar las diversas metodologías bibliométrías reportadas, con el fin de saber cuándo y cómo utilizarlas en los BA, o en todo caso complementarlas y aplicarlas como métodos para revisar y evaluar la literatura científica (Donthu et al., 2021).

Tabla 4. Metodologías bibliométricas reportadas en orden cronológico y sus diferencias en comparación con este estudio (Zupic et al., 2014) (Aria et al., 2017) (Carrión Mero et al., 2020) (Donthu et al., 2021) (Goh et al., 2021) (Qin Y. et al., 2022).

Resumen de las metodologías bibliométricas					
Autor	Preprocesamiento de Texto (P-Text)			Análisis del desempeño y de la estructura conceptual, intelectual y social (CoInSo-A)	
	Elegibilidad de datos	Recopilación de datos	Semántica de los Metadatos	Análisis y visualización de datos	Perspectivas
Zupic and Cater (2015)		✓		✓	✓
Aria and Cuccurullo (2017)		✓		✓	✓
Carrión-Mero et al. (2020)	✓	✓	✓	✓	
Donthu et al. (2021)	✓	✓		✓	✓
Goh and See (2021)	✓	✓	✓	✓	
Qin et al. (2022)		✓		✓	✓
Este estudio	✓	✓	✓	✓	✓

Etapas 1: Preprocesamiento de Texto (P-Text)

La primera etapa del BA consiste en aplicar técnicas de minería de textos con la finalidad de transformar textos no estructurados en textos estructurados y normalizados que sean aptos para el posterior análisis, este tipo de procesamiento permite un análisis eficiente de información reduciendo el sesgo en la información (Baraibar Diez et al., 2020).

Subetapa (i): Elegibilidad de datos.

Parte fundamental del BA consiste en definir los criterios de búsqueda que describan de manera exhaustiva del tema de investigación objeto de estudio. Estos criterios de búsqueda o KeyWords se debe definir previamente con base a la experiencia del investigador(es) y/o aplicando la metodología PICO la cual considera evidencia científica en los trabajos de investigación más relevantes, a través de un análisis y minería de texto donde se agrupan KeyWords bajo los siguientes parámetros:

- i. **Población o problema (tema de investigación objeto).**

- ii. **Intervenciones, causas o factores** (aplicaciones científico – tecnológicas con los cuales se han resuelto problemas de investigación previos).
- iii. **Comparación y/o estudio** (técnicas o método de investigación previamente aplicados).
- iv. **Outcomes** (por sus siglas en inglés) o resultados (conclusiones de investigaciones previamente realizadas).

Estos criterios de búsqueda harán más fácil identificar y compilar publicaciones científicas representativas a partir de bases de datos de citas indexadas seleccionadas para un período de tiempo definido.

Subetapa (ii): Recopilación de datos

La recopilación de datos bibliométricos sobre el tema de investigación debe llevarse a cabo considerando un periodo de interés y seleccionando las bases de datos de citas más adecuadas. Es preferible optar por aquellas que ofrezcan registros de citas de literatura científica revisada por pares. Entre las bases de datos más fuertes disponibles se encuentran Web of Science®, Scopus® que pueden utilizarse para esta subetapa (Martín Martín et al., 2018).

Por ejemplo, una suscripción completa a Web of Science® (WoS) proporciona acceso a registros de citas científicas publicados en aproximadamente 21,894 revistas en revisión por pares y en algunas actas de congresos de prestigio, con datos que se remontan a 1900 hasta la actualidad (Vera Baceta et al., 2019).

Scopus®, es también una base de datos de citas indexadas accesible de igual modo mediante suscripción; aunque abarca un periodo más corto (1966 a la actualidad) ofrece registros de citas de hasta más de 25,000 revistas revisadas por pares, así como actas de congresos y libros (Martín Martín et al., 2021).

Otra base de datos opcional para usar en los BA es Google Scholar es un motor de búsqueda web abierto que proporciona bibliografía académica actualizada (Harzing et al., 2016). Este motor busca registros de citas en una variedad de fuentes académicas, como revistas con revisión por pares, libros académicos, ponencias en congresos y tesis. Sin embargo, algunas publicaciones referenciadas en Google Scholar no han sido sometidas a revisión por pares y pueden incluir obras con un escrutinio menos riguroso (literatura gris) en comparación con las ofrecidas por bases de datos como WoS y Scopus. Además, Google Scholar no permite una

búsqueda avanzada basada en palabras clave específicas o metadatos, lo que limita la precisión de sus resultados.

Una vez seleccionadas las bases de datos, es fundamental definir las cadenas booleanas adecuadas (incluyendo palabras clave y sus variantes sintácticas) para facilitar la búsqueda de publicaciones científicas sobre el tema de estudio y completar la recopilación de datos bibliométricos. Esto ofrece información valiosa sobre el comportamiento del área de investigación mediante diversos indicadores bibliométricos, como autores, universidades, revistas, países y citas recibidas, entre otros.

Subetapa (iii): Semántica de los metadatos

En esta subetapa se realiza un proceso de normalización de datos, esto con el fin de evitar errores ortográficos, como son; datos de autor, títulos de revistas y afiliación en las bases de datos de citas indexadas (Carrión Mero et al., 2020). Además, se puede utilizar una hoja de cálculo o un programa informático (p.ej. R o Python) para la depuración la base de datos primaria descargada de las bases de citas indexadas como Web of Science®, Scopus® o Google Scholar utilizando una cadena de palabras clave como criterios de inclusión y exclusión esto con la finalidad de refinar la información bibliométrica, así obteniendo la base de datos de trabajo final.

Etapla 2: Análisis del desempeño y de la estructura (CoInSo).

Subetapa (iv): Análisis y mapeo de datos.

Los registros bibliométricos de la base de trabajo de datos final obtenidos de la subetapa tres serán objeto de análisis mediante cualquier paquete informático como los reportados en la Tabla 3. A través del análisis y mapeo de los datos se obtiene el desempeño y la estructura CoInSo sobre el tema de investigación objeto de estudio.

En concreto, a través del análisis de desempeño se obtiene en esta etapa información sobre producción científica, autores destacados, instituciones, artículos más citados, contribuciones por países y otros datos bibliométricos obtenidos a partir de las métricas de cantidad y calidad de la Tabla 2. Al mismo tiempo, mediante el análisis de la estructura CoInSo y el uso de las métricas de análisis multivariable de la Tabla 2, se examina el dominio de la investigación científica, la composición disciplinaria, los temas de investigación dominantes y el patrón de estas interrelaciones.

Subetapa (v): Perspectivas.

Una vez que se haya completado el análisis y mapeo de los datos, se recomienda abordar las siguientes preguntas enfocadas en la investigación, que guiarán las conclusiones y las perspectivas en futuras investigaciones:

- i. ¿Dónde se llevó a cabo la investigación?
- ii. ¿Cuál fue el impacto en las áreas de investigación relacionadas?
- iii. ¿Qué áreas de investigación pueden tener mayor impacto en un futuro cercano?

Esta información es útil para identificar los principales atributos del artículo desde la perspectiva del autor describiendo el enfoque del estudio. (Goh et al., 2021).

2.2.6 Herramientas de Inteligencia Artificial para el BA

La Inteligencia Artificial (IA) ha emergido como una herramienta revolucionaria en el ámbito de la investigación, transformando la manera en que se abordan problemas con grandes volúmenes de datos como es el caso de los análisis bibliométricos. En los últimos años, el uso de la IA ha crecido exponencialmente, incluso en algunas ocasiones, suele ser tomado como herramienta principal para la investigación (González Ciriaco et al., 2023). En el contexto académico, la IA permite automatizar tareas que antes requerían una intervención manual significativa, como la búsqueda de información, el análisis de datos y la predicción de resultados (ver Tabla 5). Además, la IA facilita la personalización de la investigación, mejorando la precisión de los modelos y proporcionando soluciones más adaptadas a problemas específicos. A medida que la tecnología avanza, su integración en la investigación científica promete no solo aumentar la eficiencia, sino también abrir nuevas fronteras en el conocimiento, impulsando la innovación diferentes ámbitos de investigación. (Lim et al., 2023).

Respecto a los BA estas herramientas resultan ser útiles para la búsqueda de información, la interpretación y organización de resultados.

Tabla 5. Herramientas de Inteligencia Artificial (IA) para el procesamiento de información científica (Lim et al., 2023)

Herramientas de Inteligencia Artificial (IA)			
Funciones Principales	Herramienta	Función	URL
Mejorar la escritura y gramática	Grammarly	Mejora la gramática, corrige la ortografía y la puntuación.	https://www.grammarly.com
	Quillbot	Mejora la redacción, la gramática, ayuda con las citas y parafrasea el texto.	https://quillbot.com/
	Jenni	Ayuda y mejora del lenguaje escrito	https://jenni.ai/
	Hemingway y Editor	Mejora la calidad de la escritura	https://hemingwayapp.com/
	ProWritingAid		https://prowritingaid.com/
	Rytr.me	Ayuda en la redacción	https://rytr.me/
Ayudar a redactar y generar ideas	ChatGPT	Modelo de lenguaje que utiliza técnicas de aprendizaje supervisado y de refuerzo.	https://chat.openai.com/
	Andi	Similar a ChatGPT, actúa como chatbot y motor de búsqueda genera respuestas de forma convencional.	https://andisearch.com/
	Perplexity AI	Utiliza LLM para respuestas complejas.	https://everask.ai/
	Komo	Proporciona referencias y citas	https://komo.ai/
	BARD	Búsqueda sin anuncios y muestra referencias de donde obtuvo la información	https://bard.google.com/
	BERT	LLM de Google, que funciona de forma similar a ChatGPT	https://github.com/google-research/bert
	Claude	Familia de modelos de lenguaje de Google presentada en 2018 puede responder a las preguntas de los usuarios	https://console.anthropic.com/login
	Osteosynthesis	Un LLM basado en transformadores que permite generar respuestas de tipo humano a las entradas del usuario.	https://www.osteosynthesis.org/pubmed-search
	AI-Writer	Genera artículos de investigación basados en palabras clave introducidas por el usuario	https://ai-writer.com/
	DeepSeek	Genera artículos a partir de un simple titular	https://www.deepseek.com/
Optimización de motores de búsqueda	Consensus	Desarrolla modelos de lenguaje de código abierto, chatbot similar a ChatGPT.	https://www.consensus.app/
	Elicit	Busca bibliografía y proporciona una lista de artículos pertinentes y sus resúmenes	https://elicit.org/
	SEO.ai	Muestra al usuario las revisiones sistemáticas pertinentes y sus resúmenes Optimización de motores de búsqueda	https://seo.ai/

Generar/inter pretar imágenes	Dall-E 2	Genera imágenes a partir de los datos introducidos por el usuario	https://openai.com/dall-e-2/
	DermEngine	Genera diferenciales basados en la imagen de la lesión cutánea cargada	https://www.dermengine.com/
	V7	Capaz de procesar imágenes, vídeos, archivos de imágenes y documentos a partir de los datos cargados por el usuario.	https://www.v7labs.com
Organización	Lateral	Organiza automáticamente los documentos para utilizarlos en futuras investigaciones	https://www.lateral.io/
	Waldo	Crea tablas y permite descargarlas, exportar documentos de google y publica investigaciones	https://www.waldo.fyi/
Análisis de voz	Fireflies.ai	Analizar, grabar y transcribir entradas de voz	https://fireflies.ai/

CAPÍTULO III METODOLOGÍA

La metodología integral propuesta para llevar a cabo el BA fue expuesta por primera vez por Yáñez-Dávila et al., (2023), sin embargo, se modificó para realizar el mapeo del conocimiento científico acumulado en geociencias, con la finalidad de tener una metodología didáctica y comprensible para el área de geociencias. Esta metodología bibliométrica integral se divide en dos etapas principales:

1. Preprocesamiento de Texto (P-TeX).
2. Análisis de desempeño (PA-A) y estructura (CoInSo-A).

que, a su vez, se divide en cinco subetapas específicas:

P-TeX

- i) Elegibilidad de datos.
- ii) Recopilación de datos.
- iii) Semántica de los metadatos.

PA-A y CoInSo-A

- iv) Análisis y mapeo de datos.
- v) Perspectivas.

La etapa de P-TeX se diseñó para la detección de patrones de productividad científica de los datos obtenidos de las bases de citas indexadas a través de las tres subetapas que componen el P-TeX (Yáñez-Dávila et al., 2023):

- (i) *Elegibilidad de datos*: esta subetapa involucra definir el criterio general de búsqueda basado en el tema de investigación objeto de estudio.
- (ii) *Recopilación de datos*: esta subetapa involucra la recopilación de datos a partir las bases de citas indexadas, utilizando palabras clave definida en la subetapa anterior (i) dentro períodos de tiempo específicos.
- (iii) *Semántica de los metadatos*: esta subetapa involucra el ajuste y refinamiento de los metadatos obtenidos en la subetapa anterior (ii).

Por otro lado, la etapa PA-A y CoInSo-A busca confirmar los principales patrones de productividad científica de los datos obtenidos de las bases de citas indexadas mediante las siguientes subetapas (Yáñez-Dávila et al., 2023):

(iv) *Análisis y mapeo de datos*: esta subetapa involucra el análisis de desempeño y de la estructura CoInSo de la base de datos obtenida de la subetapa anterior (iii) para determinar la evolución temporal de los trabajos científicos sobre el tema de estudio, así como el mapeo y la visualización de datos.

(v) *Perspectivas*: esta subetapa involucra la interpretación de la información procesada de la subetapa anterior (iv) para generar nuevo conocimiento que involucra la perspectiva de la investigación objeto de estudio (geociencias), las tendencias de la investigación futura.

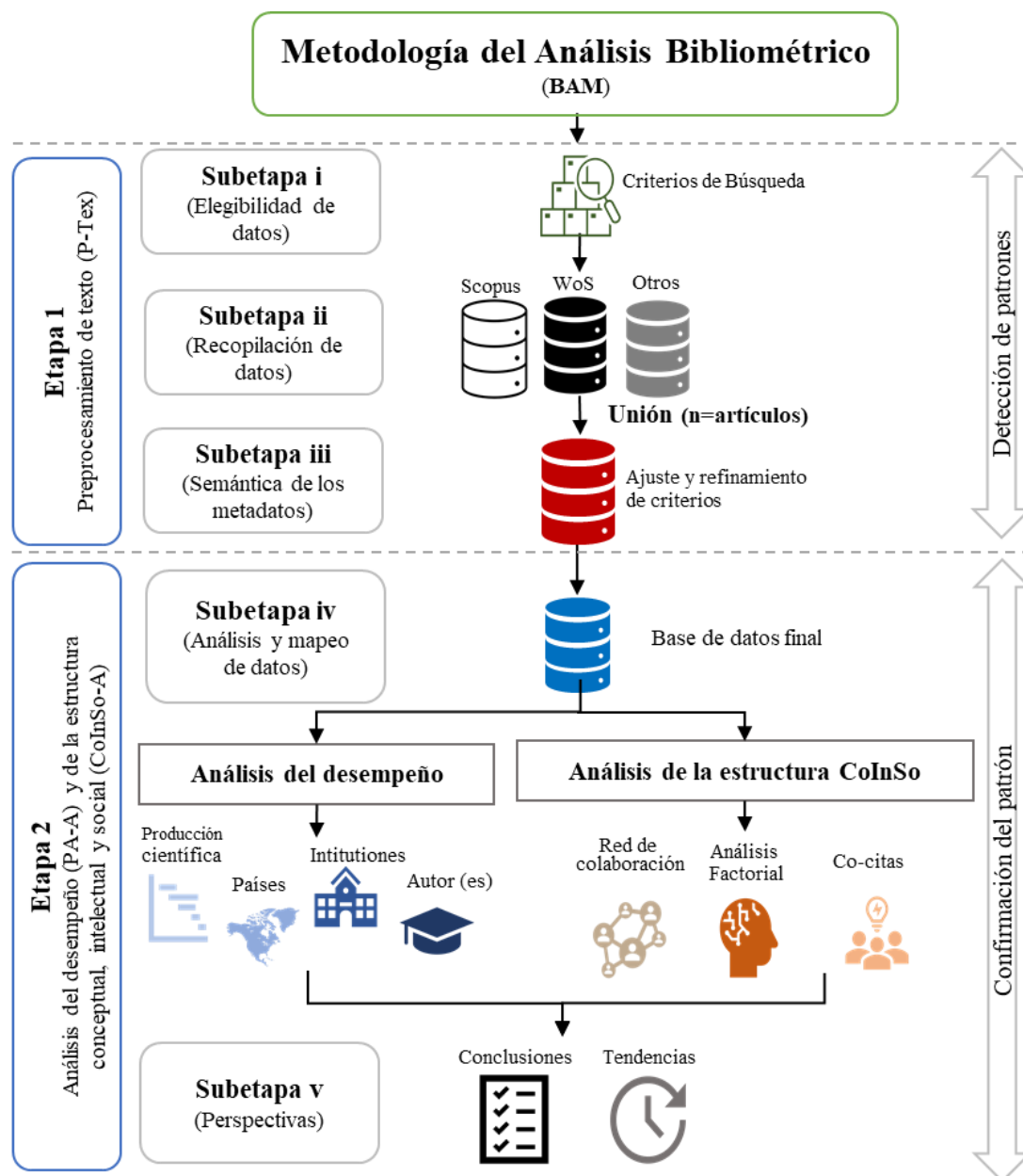


Figura 2. La metodología bibliométrica propuesta (BAM por su acrónimo en inglés: Bibliometric Analysis Methodology) busca reducir al mínimo el sesgo y los errores en el mapeo de estudios científicos acumulativos en diversos campos del conocimiento. Modificada de (Yáñez-Dávila et al., 2023)

CAPÍTULO IV RESULTADOS

4.1 Etapa 1: P-Text

4.1.1 Elegibilidad de datos (subetapa i).

Las palabras clave adecuadas para el presente BA se llevaron a cabo mediante la definición de los siguientes criterios de inclusión/exclusión:

- ✓ Idioma: inglés.
- ✓ Tipo de documento: artículo, revisiones y actas de conferencias.
- ✓ Periodo de tiempo: enero de 1993 a diciembre de 2024.
- ✓ Condición específica: Ramas de las geociencias.

4.1.2 Recopilación de datos (subetapa ii).

Teniendo en cuenta las ventajas que ofrecen las bases de datos de citas indexadas disponibles, se seleccionaron las plataformas de WoS y Scopus como herramientas de búsqueda principal. Se definieron cadenas booleanas adecuadas para la búsqueda de artículos de acuerdo con las diferentes sintaxis de estas bases de datos. Este proceso de búsqueda se llevó a cabo de forma independiente el 12 de diciembre de 2024 por dos técnicos diferentes para evitar sesgos basados en las diferentes ramas de geociencias (ver Tabla 6), con la siguiente cadena booleana sintetizada (para detalles de las cadenas booleanas usadas para WoS o Scopus consultar el material suplementario S-1 y S-2, respectivamente):

{TITLE-ABS-KEY (Análisis bibliométrico) OR TITLE-ABS-KEY (ramas de las ciencias de la tierra: seis ramas de ciencias atmosféricas; treinta y dos ramas de geología; diez ramas de geografía; nueve ramas de oceanografía; y cinco ramas de ciencias planetarias)}.

Las bases de datos citas primarias de documentos publicados (DP) fueron Scopus y WoS que resultaron 1051 y 869 DP, respectivamente. Las dos bases de datos primarias del presente BA se fusionaron dando como resultado un total de 1492 DP utilizando un método eficiente sugerido por Echchakoui, 2020.

El proceso de fusión se realizó utilizando programación R a través de los siguientes pasos computacionales:

1. Convertir los registros de citas obtenidos en las bases de datos primarias en archivos bibliográficos (*.bib).
2. Convertir los archivos Scopus.bib y WoS.bib en archivos «BibTex».
3. Formatear los archivos WoS y Scopus BibTex para obtener los mismos campos de etiquetas.
4. Fusionar los dos archivos WoS y Scopus BibTex en R-studio para rechazar duplicados de registros de publicaciones o citas.

Tras aplicar este procedimiento de fusión, se obtuvo una muestra estadística secundaria (1492 de DP).

Es importante destacar que las cadenas booleanas ofrecen información mayoritaria acerca de Ciencias de la Tierra, sin embargo, los resultados proyectan mayor crecimiento y prioridad en información de geociencias.

Tabla 6. Cadenas booleanas utilizadas en esta investigación.

Análisis bibliométricos					
Ciencias de la tierra					
Ciencias atmosféricas	Geología		Geografía	Oceanografía	Ciencias Planetarias
Ciencias atmosféricas	Geología Económica	Gemología	Palinología	Glaciología	Geología planetaria
Climatología	Geología e Ingeniería	Paleontología	Geografía	Oceanografía	Astronomía
Paleoclimatología	Geología ambiental	Petrología	Geografía física	Oceanografía biológica	Ciencias Planetarias
Química atmosférica	Geología cuaternaria	Geodesia física	Geografía humana	Oceanografía física	Selenografía
Física atmosférica	Geología petrolera	Sedimentología	Medio ambiente y química	Oceanografía química	Sistemas espaciales
Meteorología	Geología histórica	Sismología	Edafología	Paleoceanografía	
	Hidrogeología	Paleosismología	Navegación satelital	Geología marina	
	Geología estructural	Estratigrafía	Sensores remotos	Paleotempestología	
	Geoquímica	Vulcanología	Fotogrametría	Hidrología	
	Geocronología	Paleontología teórica	Topografía		
	Geodesia	Cartografía			
	Geomagnética	Geoestadística			
	Geomicrobiología	Pedología			
	Geofísica	Geomorfología			
	Micropaleontología	Limnología			
	Mineralogía				

4.1.3 Semántica de los metadatos (subetapa iii).

Se aplicó un subproceso de ajuste y refinamiento a la base de datos secundaria, en la que se realizó la normalización de datos, búsqueda de errores ortográficos y criterios de inclusión y exclusión. Se ejecutó un nuevo subproceso de filtrado utilizando una hoja de cálculo Excel y las siguientes cadenas de texto:

{include (keyword, «bibliometric analysis, » «bibliometric mapping») AND limit-to (doctype, «article» «conference proceedings, » «reviews») AND limit-to (language «English») AND limit-to («earth of science») AND exclude («health studies»)}.

Finalmente se seleccionó el resultado de este subproceso, una base de datos terciaria que contenía 971 DP (ver material suplementario S-3). Esta base de datos final fue la base de datos de trabajo para este estudio de BA (archivo M.bib o M.xlsx).

4.2 Etapa 2: PA-A y CoInSo-A

Como parte de la etapa PA-A y CoInSo-A, los resultados de la subetapa iv incluyeron los siguientes análisis:

- 1) Análisis de desempeño a través del BA aplicado a ramas de geociencias.
- 2) Análisis de la estructura CoInSo mediante el uso de clustering y análisis multicriterio en los principales indicadores bibliométricos de geociencias (ver Tabla 7).

Tabla 7. Resumen de los indicadores bibliométricos encontrados en las bases de datos Scopus y Web of Science sobre el análisis bibliométrico de las geociencias (Periodo 1993 - 2024).

Información principal sobre los datos	
Periodo de tiempo	1993-2024
Revistas	414
Documentos	971
Media de citas por artículo	20.16
Contenido del documento	
Palabras clave del autor	2676
Palabras clave Plus	5673
Autores	
Autores	2811
Autores del documento de autoría única	49
Autores de documentos multi-autor	2762
Documentos por año	31.32
Colaboración de autores	
Documentos por autor	0.345
Autores por documento	2.89
Coautores por documento	4.38

4.2.1 Análisis y mapeo de datos (Subetapa iv)

Análisis bibliométrico para evaluar la producción científica mundial.

El análisis de desempeño basado en la producción científica generada del BA en geociencias de los 971 documentos reportados de 1993-2024 se muestran en la figura 3, donde se observa un gráfico estadístico de frecuencias acumuladas de los DP por año con base a lo reportado en la literatura científica mundial, donde la línea discontinua de intersección (2020) representa el punto de inflexión más significativo de la curva exponencial y por lo tanto un cambio de tendencia lineal a exponencial en el comportamiento de los DP (Sinclair, 1974; Andreo-Martínez et al., 2020; Carrión-Mero et al., 2020).

La curva exponencial describe el crecimiento de los DP aplicado al tema de investigación de geociencias con el tiempo, que matemáticamente puede expresarse a través de la siguiente función exponencial (Price, 1963):

$$F(t) = ae^{bt} \quad \text{Ecuación 5}$$

Donde:

- ✓ $F(t)$ representa el tamaño de la muestra en un momento dado t .
- ✓ a es el tamaño inicial de la muestra.
- ✓ b es la tasa de crecimiento continúa relacionada con el porcentaje en que aumenta el tamaño de la muestra por año.

La tasa de crecimiento anual porcentual se calcula utilizando la Ecuación (5):

$$R = 100(e^b - 1) \quad \text{Ecuación 6}$$

Donde:

R representa la tasa de crecimiento anual de cualquier campo científico con respecto al año anterior.

Considerando las Ecs. (4) y (5), la tasa de crecimiento anual de la producción científica para los DP aplicado a temas de investigación en geociencias se calculó en ~26.7% (1993-2024), lo que equivale a una publicación media de ~31 artículos por año. Otro indicador bibliométrico útil

relacionado con este crecimiento exponencial viene dado por el tiempo de duplicación (periodo en el que se ha duplicado el tamaño de la muestra de publicaciones de literatura científica), que se determinó en ~3 años utilizando la Ec. (6) recomendada por (Tague et al., 1981):

$$D = \frac{\ln 2}{b} \quad \text{Ecuación 7}$$

La variabilidad del número de DP con respecto al tiempo se representa utilizando fases de la ley de Price (Price, 1963) . La fase 1 de Price se define como la zona de «precursores», donde la información se mantiene estable o continua, es decir, no crece de manera significativa en un periodo de tiempo, mientras que la fase 2 se denomina zona de «crecimiento exponencial», aquí la información obtenida por año crece de manera exponencial.

En la Fig.3 se observa una línea divisora en la información, que representa el límite temporal entre dos fases (esta línea puede cambiar a través del tiempo), para este BA se estimó a través de los datos de DP acumulados. La «zona de precursores» de la fase 1 de Price se describió para el límite temporal entre 1993-2020, en el que se publicaron 288 documentos científicos. Estos DP representan ~30 % con respecto al número total de artículos (n=971).

Por otro lado, la «zona de crecimiento exponencial» de la fase 2 de Price se identificó para un periodo de tiempo comprendido entre 2021 y 2024, en el que la producción científica exhibe un aumento exponencial de 683 DP, lo que representa ~70 % con respecto al número total de publicaciones.

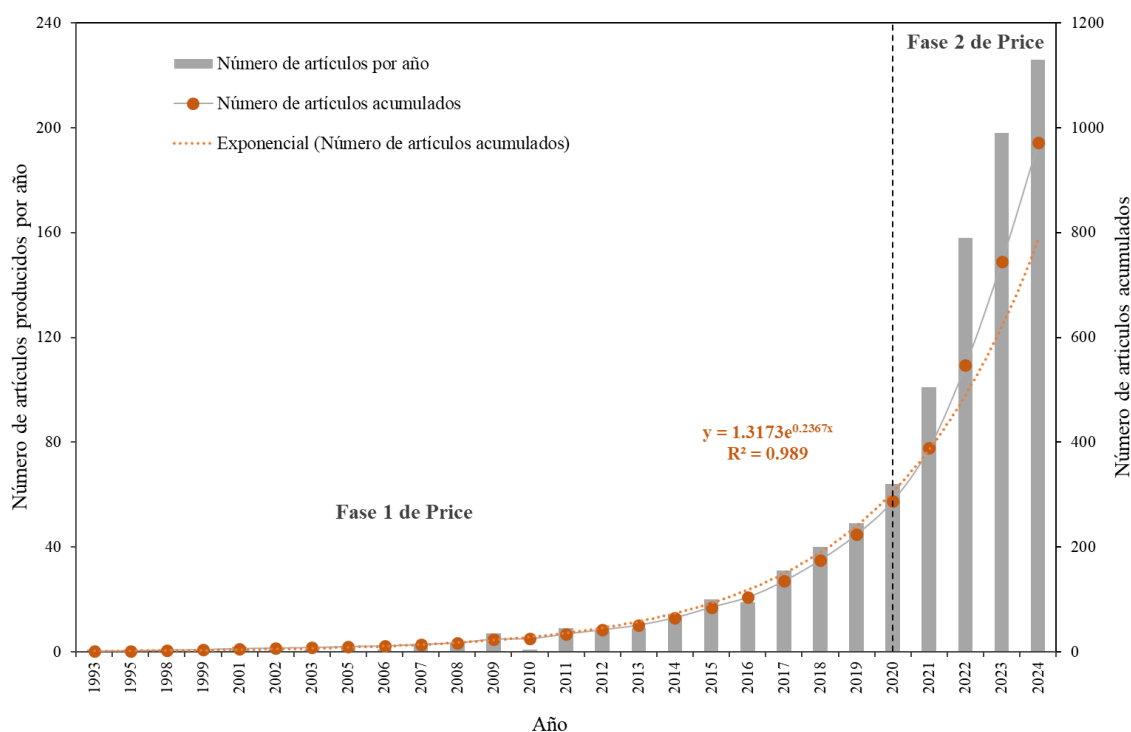


Figura 3. Gráfico de la producción científica del BA Aplicada a documentos de geociencias, donde se observa un crecimiento exponencial en la publicación descrito por la correlación entre el número de documentos por año; También se puede observar las diferentes fases de Price, donde la fase 1 se define como «zona de precursores» en la investigación y la fase 2 se define como «zona de crecimiento exponencial» iniciada sustancialmente en 2020.

Mapeo bibliométrico para evaluar la productividad de los documentos científicos por país.

A través de la información obtenida acerca de la contribución de DP por país, en la figura 4 se muestra la distribución mundial de los artículos del BA aplicados a las geociencias en revistas revisadas por pares, revisiones y actas de congresos. También se obtuvo el porcentaje de colaboración entre investigadores e instituciones de distintos países, donde el mayor vínculo de colaboración se logró principalmente para los pares:

- 3) China-Estados Unidos con el 4.2%
- 4) China-Canadá con el 1.7%
- 5) China-Reino Unido con 1.4%
- 6) Sudáfrica-Reino Unido con 1.4%

Se observaron vínculos de colaboración en otros pares de países menores o iguales al 1%.

Actualmente los trabajos de BA aplicados a geociencias han sido estudiados en 65 países. Los diez países que encabezan la publicación de artículos científicos sobre el tema de investigación y sus registros de citas de impacto se atribuyen a:

1. China, con un total de 398 DP, que representan una contribución de ~44%, y un total de 7,819 citas.
2. India con 52 DP, con una contribución de ~6%, y 460 registros de citas.
3. Brasil con 51 DP, una contribución de ~6%, y 834 citas.
4. España con 34 DP, una contribución de ~4%, y 1204 citas.
5. Estados Unidos con 33 DP, una contribución de ~4%, y 1358 citas.
6. Australia con 26 DP, una contribución de ~3%, y 887 citas.
7. Sudáfrica con 24 DP, una contribución de ~3%, y 472 citas.
8. Italia con 23 DP, una contribución de ~3%, y 597 citas.
9. Chile con 16 DP, una contribución de ~2% y 98 citas.
10. Malasia con 16 DP, una contribución de ~2 %, y 213 citas.

Otros países como Canadá, Ecuador, Polonia, Alemania, **México**, Reino Unido, Colombia y Turquía tienen una producción menor, con una contribución integrada de ~11%. Entre estos países, el Reino Unido y Ecuador destacan por un alto impacto de la investigación dado por 425 y 470 citas, respectivamente. Hasta el momento **México** representa solo el 0.3% de aportación de información a nivel global.

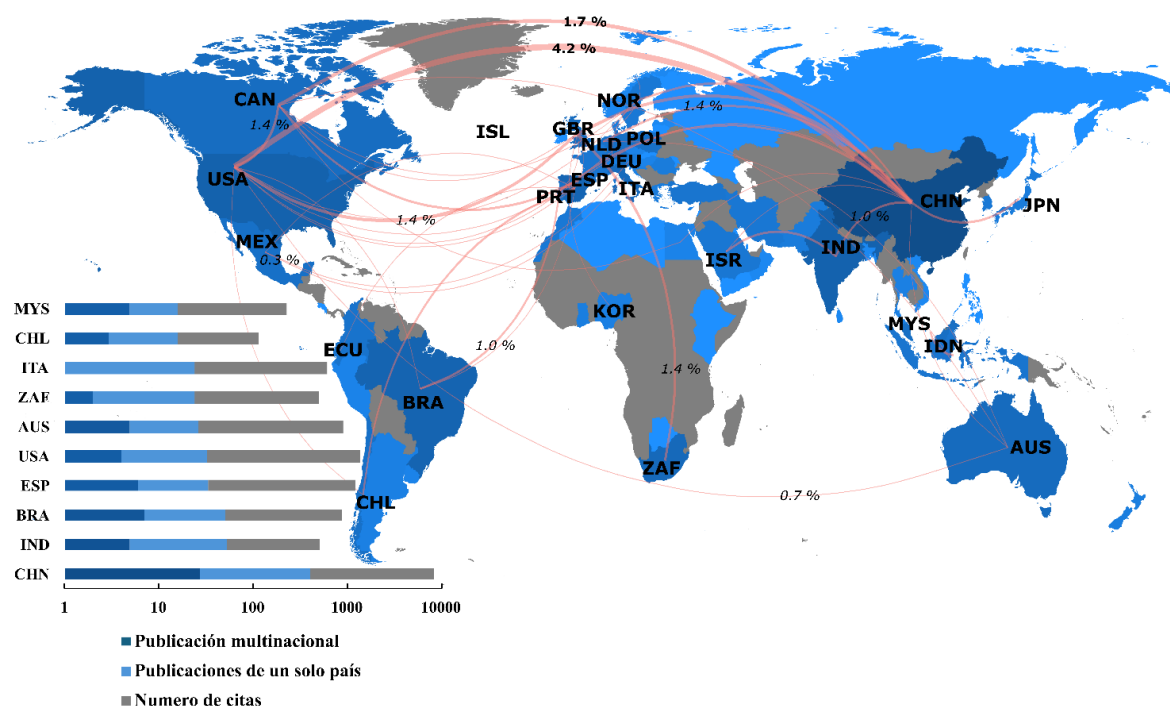


Figura 4. Distribución mundial de documentos publicados en revistas especializadas, revisiones y actas de congresos sobre análisis bibliométricos en las geociencias (1993-2024). El grosor de la línea roja indica el porcentaje de colaboración. Los nombres de los países se abrevian utilizando la convención internacional sugerida por la norma internacional ISO-3166. La escala logarítmica del número de citas va de 1 a 10000 y se solapa con el número de artículos publicados de manera multinacional o de un solo país.

Análisis bibliométrico para la evaluación de la productividad de documentos por institución.

Como resultado del BA para el periodo 1993-2024, se identificaron un total de 1,537 instituciones como muestra representativa, con un total de 4,136 artículos publicados en conjunto. Los resultados bibliométricos indican que las cinco principales instituciones que lideran la productividad de DP son:

1. Universidad de Wuhan (China) lidera con la publicación de 65 artículos.
2. Universidad de Pekin (China) con la publicación de 51 artículos.
3. Universidad Tianjin (China) con 40 artículos.
4. Escuela Superior Politécnica del Litoral (Ecuador) con 37 artículos.
5. Universidad Tecnológica de Malasya (Malasia) con 30 artículos.

También pueden observarse las doce instituciones que lideran la productividad de DP en geociencias en la siguiente tabla:

Tabla 8. Principales instituciones mundiales y país residente por productividad de documentos publicados sobre temas de geociencias (Periodo 1993 - 2024).

Institución	País	Documentos	DT-W (%)
Universidad de Wuhan	China	65	1.6
Universidad de Pekín	China	51	1.3
Universidad de Tianjin	China	40	1.0
Escuela Superior Politécnica del Litoral	Ecuador	37	0.9
Universidad Tecnológica de Malaysia	Malasia	30	0.8
Universidad de Almería	España	30	0.8
Universidad Normal de Pekin	China	29	0.7
Universidad de Guangzhou	China	28	0.7
Universidad de Geociencias de China	China	28	0.7
Academia China de Investigación en Ciencias Medio-Ambientales	China	26	0.7
Universidad de Hohai	China	26	0.7
La Universidad de Nápoles Parthenope	Italia	24	0.6
DT-W (%) Documentos ponderados con respecto al número total de publicaciones por instituciones			

Por otra parte, en la figura 5 se muestra un diagrama esquemático de la red con las agrupaciones de colaboración más productivas entre instituciones de investigación. Se identificaron claramente seis agrupaciones significativas de colaboración. Las universidades de Pekín (China) y Guangzhou (China) en colaboración con 3 universidades más fueron reconocidas como el clúster de colaboraciones más relevante. Los otros cinco clústeres se encontraron para las siguientes alianzas institucionales:

2. Universidad de Almería (España) y la Escuela Superior Politécnica del Litoral (Ecuador).
3. Universidad de Peking y Universidad de Asia.
4. Universidad de Lanzhou (China) y el Instituto del Noroeste de Eco-Ambiente y Recursos.
5. Universidad de Hohai y la Universidad Agrícola de Nanjing (China).
6. Universidad de Ciencia y Tecnología de la Información de Nanjing y la Universidad de Hainan (China).

Tras analizar estos resultados, se observó que la fuerza y el número de agrupaciones de colaboración depende de factores como la distancia geográfica, el idioma entre países y autores, el uso compartido o el acceso a infraestructuras, entre otros. Algunos de estos aspectos que intervienen en la colaboración fueron analizados por Melin y Persson (1996).



Figura 5. Diagrama esquemático de redes bibliométricas que muestra seis grupos principales de colaboración más productivos entre instituciones identificada a partir de este estudio.

Análisis bibliométrico para evaluar la productividad de los DP en revistas de revisión por pares.

Dentro de los 971 DP resultantes se encontró que 417 revistas de revisión por pares fueron utilizadas como principales fuentes de publicación. Estos DP recibieron 18,801 citas entre 1993 - 2024.

Con base en la ley de Bradford (para más detalle revisar la sección 2.2.3), en la figura 6 se representa una lista parcial de revistas de mayor a menor publicaciones que fueron de ayuda para lograr agrupar y clasificar esta información en tres zonas principales:

1. Zona Núcleo: 14 revistas agrupadas en el diagrama de productividad de Bradford utilizadas para la publicación de 327 artículos, con una contribución de ~33,6 % del número total de DP.
2. Zona Intermedia: 107 revistas agrupadas para la publicación de 325 artículos con una contribución de ~33,4 %.
3. Zona periférica: 293 revistas agrupadas utilizadas para la publicación de 319 documentos con una contribución de ~32.8%.

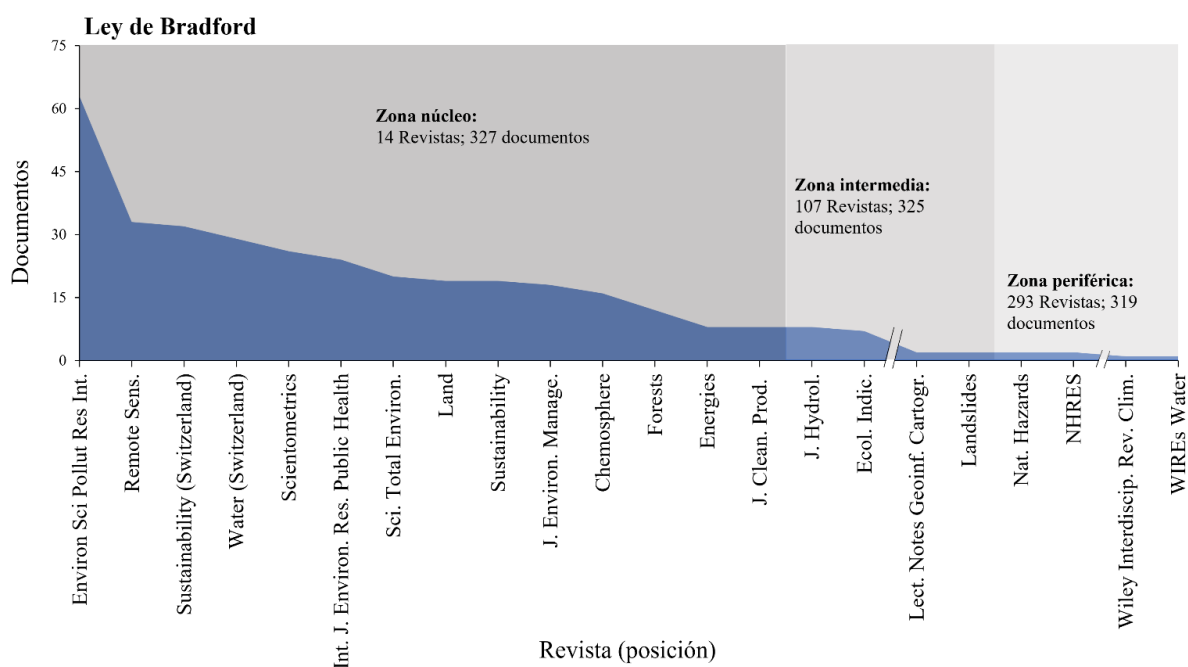


Figura 6. Zonas de las revistas que cubren el tema de las geociencias análisis bibliométrico (1993-2024) según la ley de Bradford, mostrando las zonas núcleo, intermedia y periférica.

Análisis bibliométrico para evaluar la productividad de DP por autor.

Después de analizar la información total resultante de los DP (n=971), se encontró que la mayoría de estos artículos fueron publicados predominantemente por grupos de colaboración de autores (de 2 a 18 coautores por artículo). La Tabla 9 presenta los diez principales grupos de colaboración de autores que lideran con 224 DP el tema de investigación de BA aplicado a las geociencias durante (2008 - 2024).

Dentro de estos grupos de autores lideran tres con alta productividad en DP:

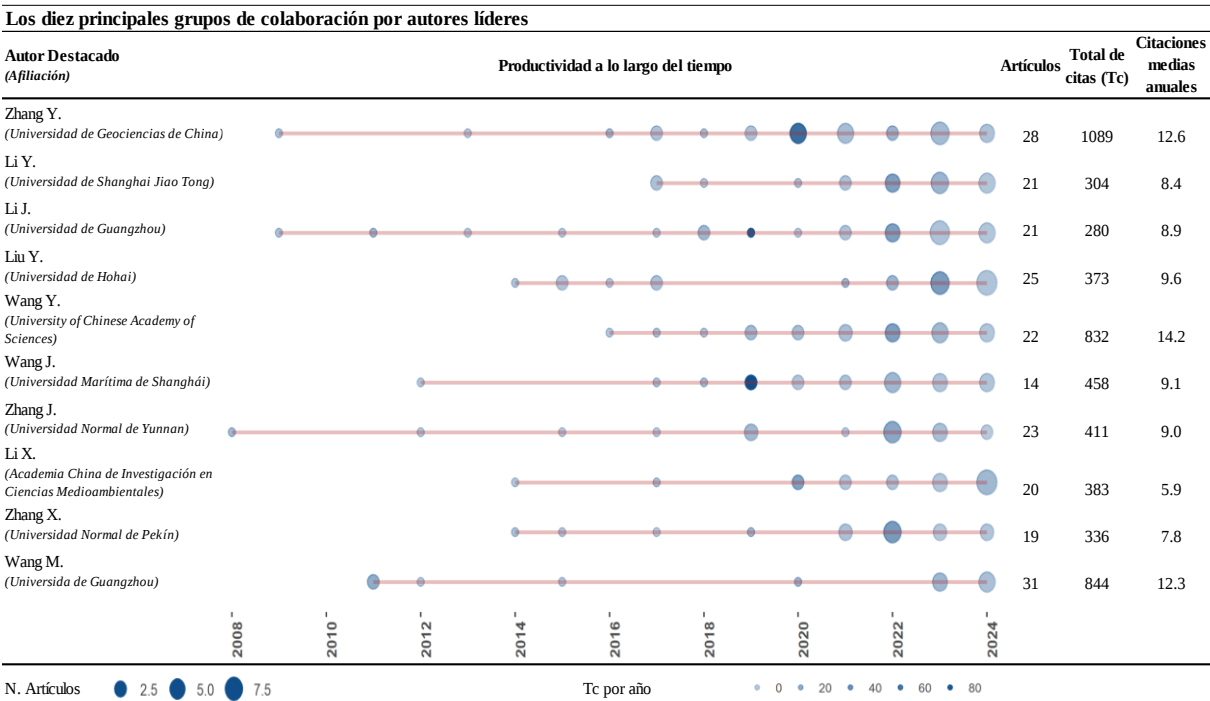
1. «M. Wang» con 31 DP.
2. «Y. Zhang» con 28 DP.
3. «Y. Liu» con 25 DP.

Respecto al número de citas hasta 2024, lideran los grupos de autores:

1. «Y. Zhang» con 1,089 registros de citas y un promedio de 12.6 citas medias anuales.
2. «M. Wang» con 844 registros de citas y 13.3 citas medias anuales.
3. «Y. Wang» con 832 registros de citas 14.2 citas medias anuales.

En la Tabla 9 también se enumeran otros grupos importantes de colaboración de autores y su productividad a lo largo del tiempo en BA aplicada a las geociencias.

Tabla 9 Principales grupos mundiales de investigadores según la productividad a lo largo del tiempo, el número de artículos y citas publicados sobre el tema de investigación de la BA aplicada a las geociencias.



Con esta información en la figura 7 se presenta un diagrama de Sankey esquemático con las preferencias de estos diez grupos de autores principales para publicar en determinadas revistas indexadas.

Desde el punto de vista del impacto científico, los grupos de autores «Y. Zhang *et al.*» e «Y. Li *et al.*» (con un índice *h* de 10 y 9, respectivamente) prefirieron publicar en revistas de alto factor de impacto, como:

1. Environmental Science and Pollution Research con una producción del 23.5% del total de DP.
2. Remote Sensing con una producción del 10.6% del total de DP.
3. Sustainability con una producción del 9.2% del total de DP.

Según la ley de Bradford estas revistas se sitúan en la zona núcleo. Los demás autores han seleccionado revistas situadas en la zona intermedia y periférica de la ley de Bradford, para las que el índice *h* oscila entre 5 y 8.

En la misma figura, se observa también la colaboración de autores entre países, donde los artículos publicados por los grupos de autores lideran los países:

1. «Y. Zang *et al.*» escritos principalmente por coautores de China y EE. UU.
2. «Y. Li *et al.*» también escritos por coautores de China y EE. UU.
3. «Y. Liu *et al.*» escritos por coautores de China, EE. UU. y Canadá.

En la figura 7 se encuentran también países con menores colaboraciones entre autores y revistas para publicaciones de BA en geociencias.

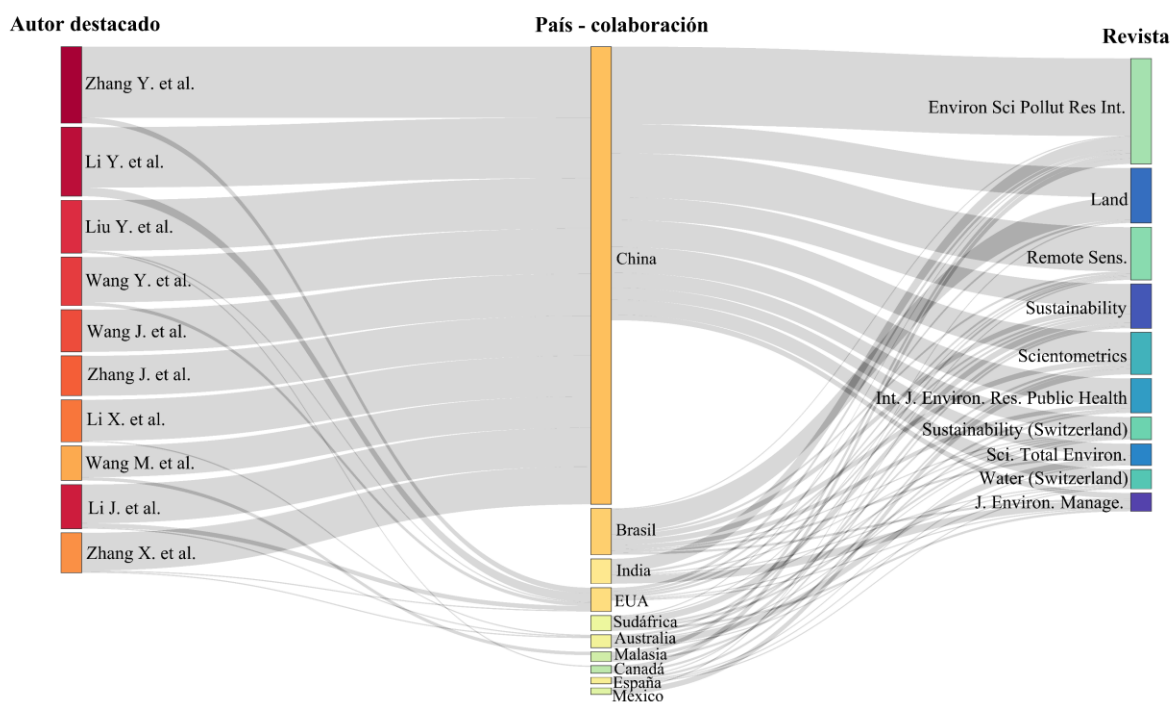


Figura 7. Diagrama de Sankey donde muestra las preferencias de los diez primeros grupos de autores (Tabla 8) para publicar sus trabajos de investigación de BA en el tema de geociencias (1993-2024). El tamaño de las líneas grises representa la cantidad de información que fluye entre colaboración de autores y revistas. Así, las líneas más anchas indican un flujo mayor, mientras que las líneas más estrechas indican un flujo menor de información.

Análisis de redes mediante mapa bibliométrico de co-ocurrencias.

Para analizar el dominio del conocimiento científico, se creó un mapa bibliométrico de co-ocurrencias utilizando KeyWords Plus® (figura 8). En esta red se identificaron las KeyWords Plus® que se utilizaron para analizar la base de datos de trabajo para el periodo 1993-2024.

El tamaño de estos clústeres permitió identificar los principales nodos que centralizaban las conexiones entre otras KeyWords Plus® relacionadas. La conexión de enlaces en el mapa muestra las relaciones de KeyWords Plus® entre nodos, mientras que el grosor de estas conexiones representa la fuerza de estas relaciones (Van Eck et al., 2014).

A partir de este análisis se generó una red que abarca las principales cuarenta y nueve KeyWords Plus® interconectadas. Se identificaron siete grandes grupos de KeyWords Plus®:

1. Teledetección e inteligencia artificial: el clúster más grande identificado es representado principalmente por procesos de desarrollo software que están fuertemente interconectados y referenciados con las KeyWords^{Plus®} relacionadas con aprendizaje automático, aprendizaje profundo, Sistemas de Información Geográfica (SIG), Inteligencia Artificial (IA), análisis de redes y agricultura de precisión.
2. Cambio climático y sustentabilidad: es el segundo clúster en tamaño mostrado por el mapa de co-ocurrencias que muestra una fuerte conexión con trabajos de sustentabilidad, desarrollo sustentable, agricultura, recursos hídricos, huella hídrica entre otros.
3. Contaminación atmosférica y salud: el tercer clúster muestra una fuerte interconexión de KeyWords^{Plus®} relacionadas con salud, covid-19, calidad del aire y modelación atmosférica.
4. Tratamiento de aguas residuales: es el cuarto clúster mostrado por el mapa de coocurrencias, este representa una conexión entre temas de contaminación, medio ambiente, tratamiento de agua, tratamiento aguas residuales y procesos de adsorción.
5. Remediación del suelo: el quinto clúster de acuerdo con las KeyWords^{Plus®} muestra una fuerte conexión con trabajos de investigación referentes a el estudio de microplásticos, biorremediación y arsénico.
6. Calidad del agua: el sexto clúster muestra conexiones con temas de agua dulce y calidad del agua.
7. Agua subterránea: el séptimo clúster y siendo el más pequeño, tiene una similitud con temas de inundaciones.

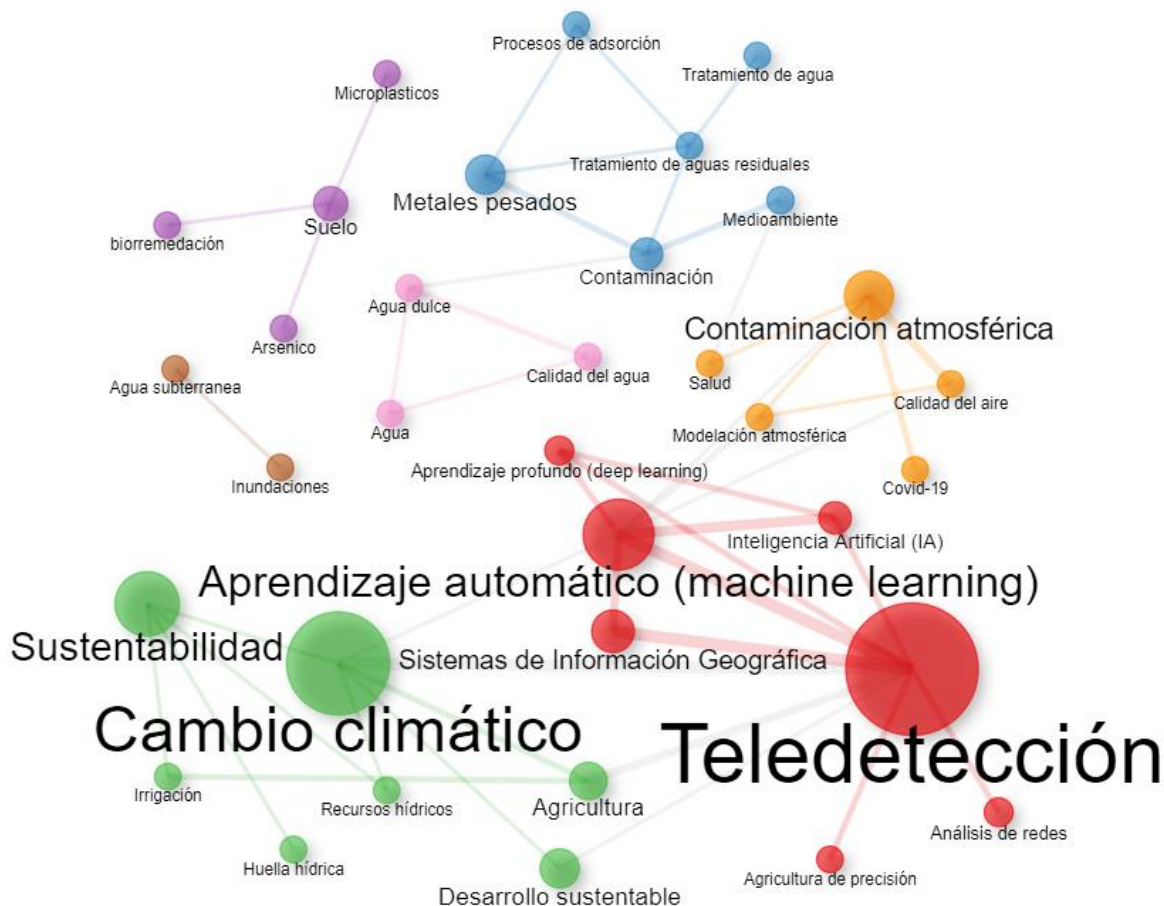


Figura 8. Mapa bibliométrico esquemático de co-ocurrencia que muestra las palabras clave más representativas encontradas en DP y agrupadas en siete grandes grupos de palabras clave: procesos relacionados con la teledetección e inteligencia artificial (red de color rojo), temas referentes al cambio climático y sustentabilidad (color verde), temas con contaminación atmosférica y salud (en color amarillo), temas relacionados con recursos hídricos (rosa, azul y café), y finalmente temas respecto a la remediación del suelo (morado).

Análisis de la evolución temática basado en el análisis de redes.

Dentro del análisis de la estructura conceptual se realizó un mapa de evolución temática para el análisis de los temas de investigación dominantes utilizando la variable categórica KeyWords Plus® incluida en la base de datos final o de trabajo de 971 DP (figura 8).

Las fronteras de investigación de BA aplicados a geociencias se realizan utilizando una matriz bidimensional que involucra dos tipos de medidas:

- Grado de relevancia o centralidad: indicado por el <eje x> en función de la importancia del tema.
- Grado de desarrollo o densidad: indicado por el <eje y> en función de la medida del desarrollo de los temas.

Además, los círculos dentro del mapa son aquellas que el KeyWords Plus® etiqueta con mayores ocurrencias, y su tamaño son proporcionales a las ocurrencias de la palabra clave. (Cobo et al., 2011; Aria et al., 2020).

Tras trazar el resultado de las dos variables (figura 9), se identificaron nueve conglomerados ubicados del más grande al más pequeño con el número de clúster.

Se encontraron cinco conglomerados ubicados en el cuadrante superior derecho que indica **temas motores** (temas muy importantes para el campo de investigación que, además están bien desarrollados; Kaiser et al., 2023):

- ✓ Clúster 1: círculo de mayor tamaño con temas centrales de metales pesados, suelo, contaminación, microplásticos, biorremediación y tratamiento de aguas residuales.
- ✓ Clúster 2: círculo de menor tamaño con los temas de teledetección, aprendizaje automático, SIG, desarrollo sustentable, IA y aprendizaje profundo.
- ✓ Clúster 3: círculo de menor tamaño que abarca los temas de sustentabilidad, agricultura y agua subterránea.
- ✓ Clúster 7: círculo de los más pequeños de la sección con temas de agricultura de precisión, calidad del agua, recursos hídricos.
- ✓ Clúster 5: círculo más pequeño de la sección, indican temas menores relacionados con contaminación atmosférica, calidad del aire y covid-19.

Un conglomerado ubicado en el área de transición entre **temas motores y básicos** (temas menos desarrollados, pero igualmente importantes; Kaiser et al., 2023):

- ✓ Clúster 4: abarca temas relacionados con análisis de redes, manejo de residuos y aguas residuales.

Tres conglomerados ubicados en el cuadrante inferior derecho que indica **temas básicos**:

- ✓ Clúster 6: temas de uso del suelo y cambio climático.
- ✓ Clúster 8: temas relacionados con servicios ecosistémicos, conservación.
- ✓ Clúster 9: tema de contaminación del suelo.

Siendo estos últimos, los círculos más pequeños de la sección en el diagrama en general.

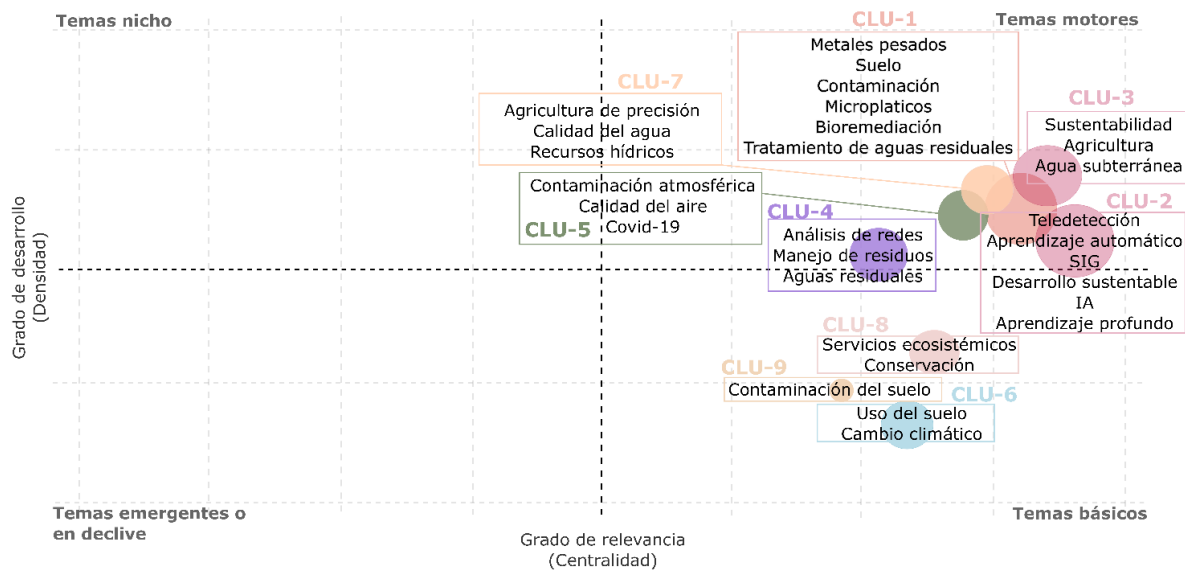


Figura 9. Mapa temático basado en el análisis de las redes de KeyWords Plus® del mapeo bibliométrico en geociencias (1993-2024).

CAPÍTULO V DISCUSIÓN

5.1 Producción científica mundial.

Con base en la Ley de Price podemos dilucidar la existencia de dos fases: la primera de ellas nombrada como “precursores en la investigación” que abarca el periodo de 1993 a 2020, el cual refleja un crecimiento lineal en la investigación en cuestión, la segunda fase denominada “crecimiento exponencial” con un periodo de 2021 a 2024 indica un claro cambio hacia un ritmo de producción mucho más rápido (Price, 1963). Esta transición puede ser resultado de varios factores como son el impacto de la pandemia y el confinamiento ya que estos redujeron la investigación de campo y de laboratorio por lo que la comunidad científica generó nueva información a partir de la información existente (Corpus-Mendoza et al., 2021). Esta situación inusual forzó una mayor digitalización y rapidez en ciertos campos de investigación (González Ciriaco et al., 2023). Esto se logró analizando la información existente teórica-experimental de manera sistemática a través de un análisis bibliométrico, análisis sistemático de la literatura, o metaanálisis, entre otros, además del aumento en la colaboración entre países, el uso de tecnologías como las herramientas de inteligencia artificial, además del rápido crecimiento e innovación de tecnologías de investigación.

5.2 Productividad de los documentos científicos por país.

La posición de China en la investigación mundial en BA aplicado a geociencias, ha crecido de manera exponencial en los últimos años, siendo un pionero en la generación de información científica a nivel mundial con una aportación del 44% del total de la información, este liderazgo se refleja en sus vínculos de colaboración principalmente. La predominancia de este país en la producción de información sobre BA puede estar relacionada con el volumen de inversión en investigación científica, la creación de infraestructuras académicas robustas, una fuerte orientación hacia la innovación tecnológica, así como su capacidad para atraer y exportar talento académico (Li et al., 2020). Además, es importante recalcar la estrecha colaboración entre EUA y China (líderes de la investigación en geociencias) en varias áreas de investigación a pesar de la guerra comercial entre ellos (González García, 2020). Estos resultados nos pueden indicar también los pioneros en investigación mundial de BA en otras líneas de investigación.

Una futura línea de investigación podría explorar la relación entre las políticas nacionales de investigación y el impacto de las publicaciones científicas, particularmente en países con baja

producción. Asimismo, sería útil investigar los factores sociales, económicos y políticos que influyen en la colaboración internacional, con especial énfasis en los vínculos entre potencias científicas y países en desarrollo. (Wagner et al., 2005).

5.3 Productividad de documentos científicos por institución.

El fuerte liderazgo de las instituciones chinas se ve reflejado con las tres primeras posiciones institucionales: Universidad de Wuhan, Universidad de Pekín, y Universidad Tianjin, que juntas representan una parte significativa de la producción de documentos en este campo (156 DP, que representa el 16%). Nuevamente esto subraya el creciente poder de China en la investigación científica global en áreas clave de geociencias. Este fenómeno reflejar tanto el desarrollo y expansión de las capacidades investigadoras de estas universidades, así como su estructura robusta para realizar investigación tanto en términos de financiación e infraestructura científica. El predominio de instituciones chinas en el top ten de la lista de las más productivas puede indicar también un fuerte enfoque de investigación en áreas prioritarias como el desarrollo tecnológico y la innovación científica (Zhou et al., 2006).

La presencia de la Escuela Superior Politécnica del Litoral (Ecuador) entre las cinco principales instituciones, resalta el impacto creciente de las universidades de América Latina en la producción científica a nivel global, lo cual podría sugerir un cambio en la dinámica de publicación, con más investigadores de esta región participando activamente en proyectos de alto impacto (Vessuri et al., 2013). Respecto a la colaboración entre universidades se logra identificar que la colaboración conformada por la Universidad de Pekín y la Universidad de Guangzhou, junto con otras tres universidades, refleja una tendencia común en la colaboración intra-nacional, mientras que las alianzas transnacionales, como las que involucran a universidades de España y Ecuador, sugieren que las barreras geográficas pueden ser superadas mediante estrategias colaborativas eficaces y el uso de plataformas digitales (Wagner et al., 2005).

5.4 Productividad de los documentos científicos en revistas indexadas.

Con respecto a las revistas indexadas para la publicación de los 971 documentos de investigación, encontramos que el número de DP entre 1993-2024 creció exponencialmente a partir del 2020 (figura 10), este hallazgo que confirma la importancia de las revistas académicas, la actualización del conocimiento y el avance de la ciencia en geociencias. Este patrón de

distribución no solo valida la ley de Bradford, sino que también muestra el impacto de cada revista con respecto a los DP, donde se observa que las revistas Environment Science Pollution Research International, Remote Sensing, and Sustainability (Switzerland), muestran una pendiente positiva a partir del 2016, 2018 y 2019 respectivamente, indicando un incremento en el número de DP acumulados con respecto al tiempo en BA en geociencias, lo que manifiesta la diversidad y la complejidad del panorama de edición en el campo de investigación.

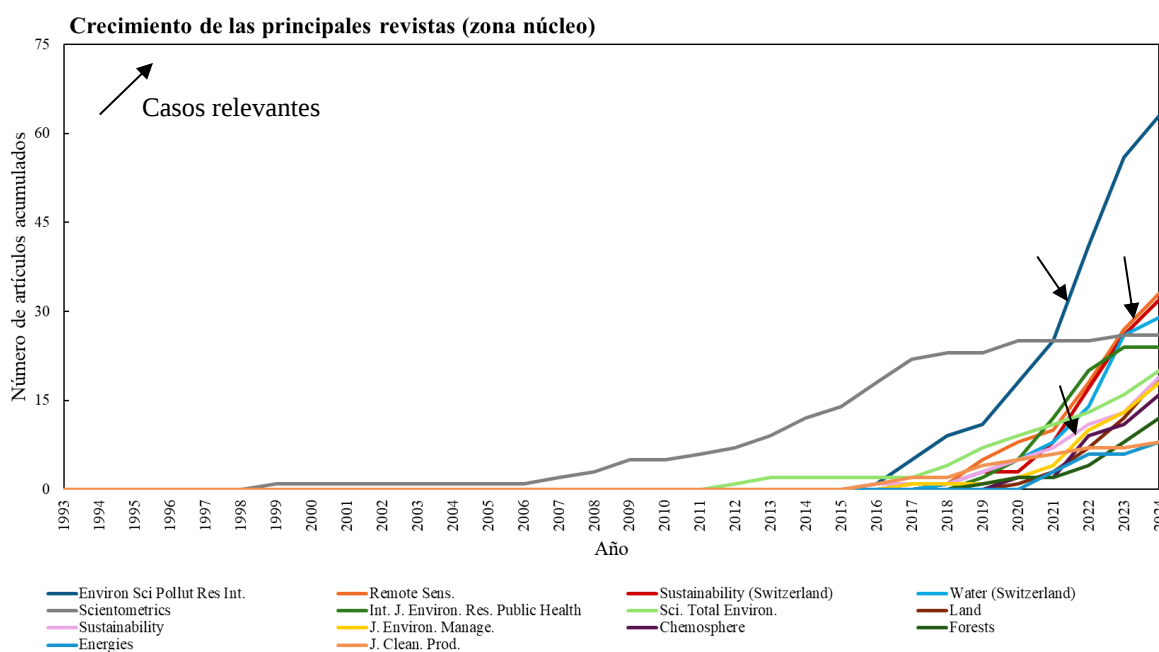


Figura 10. Crecimiento de las principales revistas a través de los años en BA aplicados a geociencias.

5.5 Productividad de los documentos científicos por autor.

Por lo que se refiere a la productividad por autor líder en la investigación en temas de BA en geociencias, China encabeza el top ten de los principales grupos de investigadores clave en el campo, cuyos trabajos han sido citados y seguidos por otros científicos. Se denota que estos autores no solo tienen una alta productividad, sino que también lideran en el número de citas, lo que indica que sus publicaciones han tenido un alto impacto en la comunidad geocientífica.

(Tabla 9). Este fenómeno puede deberse a su capacidad para abordar temas relevantes y actuales dentro de la investigación en los temas abordados en la presente investigación.

El hecho de que estos grupos de investigación prefieran publicar en revistas como *Environmental Science and Pollution Research*, *Remote Sensing* y *Sustainability* resalta la importancia de publicar en revistas de alto prestigio para asegurar una mayor visibilidad y un impacto más amplio. Estas revistas tienen una alta circulación y están indexadas en bases de datos reconocidas como WoS o Scopus, lo que facilita una mayor diseminación de los resultados de investigación. En consecuencia, la investigación globalizada apoyada por colaboraciones interdisciplinarias y multinacionales entre investigadores y países juega un importante papel en el éxito de la investigación sobre BA aplicado a las geociencias, el cual sigue siendo un motor esencial para el avance de las ciencias aplicadas y la sostenibilidad ambiental (Restrepo Botero, 2020).

5.6 Mapa bibliométrico de co-ocurrencias.

El mapa bibliométrico de co-ocurrencias de KeyWords Plus® revela una clara tendencia hacia la integración de tecnologías avanzadas como la inteligencia artificial y el aprendizaje automático en investigaciones centradas en el cambio climático, la contaminación y la gestión de recursos naturales. Permitiendo mapear las interconexiones entre diversas áreas temáticas dentro de la investigación científica, brindando una visión clara de las tendencias dominantes y emergentes en el campo de las geociencias. Este mapa de co-ocurrencias y los clústeres identificados brindan una visión valiosa sobre las áreas clave en la investigación científica aplicada, pero también presenta algunas limitaciones. El uso de KeyWords Plus® puede no capturar toda la riqueza temática de los trabajos publicados, ya que no todos los términos clave reflejan de manera precisa la complejidad de los temas tratados en los estudios. (Valderrama Zurián et al., 2021) Además, aunque este análisis bibliométrico proporciona una visión general de los temas más populares, pero no se detiene en análisis de los métodos ni en el impacto real de las investigaciones en el campo.

CAPÍTULO VI CONCLUSIONES

5.1 Perspectivas (*Subetapa v*)

Conforme a los objetivos principales propuesto en el presente trabajo de tesis, los resultados alcanzados permitieron establecer las siguientes conclusiones:

- (1) El exhaustivo marco teórico desarrollado en la presente tesis ha proporcionado una gama de herramientas y detalles técnicos que permitirá realizar un BA acerca de cualquier tema de investigación en geociencias.
- (2) La metodología integral desarrollada en la presente tesis (BAM) permitió evaluar el progreso de la investigación bibliométrica mundial aplicado a geociencias utilizando las dos bases más grandes de metadatos datos que existen en la actualidad (WoS y Scopus), para el período de tiempo de 1993-2024.
- (3) La evolución temporal de la productividad y el estado actual de desarrollo de los BA aplicados geociencias ha permitido establecer que este tipo de investigaciones se encuentran en la fase inicial de la «zona de crecimiento exponencial», indicando que el uso de este tipo de estudios para abordar diferentes temas de investigación aplicados a geociencias es de actual interés para los investigadores.
- (4) En este contexto la colaboración entre países, instituciones y autores, en investigaciones en geociencias, China, India y Brasil lideran la publicación de DP de BA. En cuanto a los grupos de investigación que lideran estos estudios bibliométricos, el grupo de Wang M. *et al.*, y Zhang Y. *et al.*, han sido los grupos de investigación más activos en este campo del conocimiento. En cuanto a uso de revistas indexadas preferentemente utilizadas para reportar este tipo de estudios bibliométricos con aplicación en geociencias, Environmental Science and Pollution Research lideró la producción DP, seguida de Remote Sensing, Sustainability, y otras once revistas de alto impacto, que, con base en la ley de Bradford, se sitúan en la zona núcleo.
- (5) La evolución de los temas y áreas de investigación dentro de las geociencias permitió establecer siete grupos principales a través de las KeyWords Plus®: (i) Teledetección e IA, (ii) cambio climático y sustentabilidad, (iii) contaminación atmosférica y salud, (iv) tratamiento de aguas residuales, (v) remediación del suelo, (vi) calidad del agua, y (vii)

agua subterránea e inundaciones), donde se identificaron como primer enfoque los trabajos sobre teledetección e inteligencia artificial como el principal hotspot de investigación de BA en geociencias. Otros temas de investigación relevantes a partir del mapa de evolución temática e identificados como temas motores en geociencias son agricultura de precisión, tratamiento de aguas residuales, aprendizaje automático y profundo, y sistemas de información geográfica como los principales hotspots bibliométricos siendo temas fuertes desarrollados. Por lo tanto, los temas no incluidos en el mapa de evolución temática son susceptibles para ser utilizados en el futuro diseño de nuevos análisis bibliométricos para abordar algunos de los últimos avances y retos de la investigación aplicada a las geociencias. Así mismo se sugieren que las soluciones a los problemas ambientales del futuro dependerán de una colaboración entre diversas disciplinas, utilizando tanto tecnologías de vanguardia como enfoques sostenibles para enfrentar los retos globales.

- (6) El crecimiento de los BA enfocado como una herramienta estratégica en la investigación de geociencias es de gran importancia ya que refleja la aceleración y estado actual de la ciencia y la investigación en esta área. El incremento fue impulsado inicialmente por el confinamiento debido al covid-19, sin embargo, en la actualidad puede estar siendo impulsado por factores externos como el aumento de fondos para ciertas áreas de investigación, y la expansión del conocimiento. Además, la producción científica observada en los BA se distribuye de manera diferente entre un número reducido de revistas de alto impacto y una mayor cantidad de revistas que, aunque menos citadas, son vitales para la diversidad y la expansión del conocimiento. Esto también pone de manifiesto la relevancia de considerar no solo el número de publicaciones, sino también la calidad y el impacto que estas tienen en la comunidad científica global destacando su potencial para la toma de decisiones geocientífica y la planificación de futuras investigaciones.

Finalmente, se espera que el presente BA beneficie a todos los investigadores, técnicos y estudiosos interesados en el tema de investigación en geociencias para la planificación o diseño de una amplia variedad de futuras aplicaciones de análisis bibliométrico que aporten ideas útiles a aquellos investigadores interesados en resolver algunos de los actuales problemas científicos y tecnológicos presentes que conlleven a nuevas fronteras del conocimiento.

REFERENCIAS

- Ajiferuke, I., Burell, Q., & Tague, J. (1988). Collaborative coefficient: A single measure of the degree of collaboration in research. *Scientometrics*, 14(5), 421–433. doi: 10.1007/BF02017100
- Alaniz Álvarez, S. A., Camprubí, A., & Villacorta, S. (2024). Navigating the Future of Geoscience Education in a Post-COVID World. *Episodes*, 47(4), 697–698. doi: 10.18814/epiiugs/2024/02404s00
- Alvarado Urbizagastegui R. (2011). La colaboración de los autores en la literatura producida sobre la Ley de Lotka. *Ciência Da Informação*, v. 40 n. 2 (2011). doi: <https://doi.org/10.18225/ci.inf.v40i2.1315>
- Andreo Martínez, P., Ortiz Martínez, V. M., García Martínez, N., de los Ríos, A. P., Hernández Fernández, F. J., & Quesada Medina, J. (2020). Production of biodiesel under supercritical conditions: State of the art and bibliometric analysis. *Applied Energy*, 264, 114753. doi: <https://doi.org/10.1016/j.apenergy.2020.114753>
- Ardanuy, J. (2012). Breve introducción a la bibliometría. *Facultat de Biblioteconomia. Universitat de Barcelona*, 1.
- Aria, M., & Cuccurullo, C. (2017). bibliometrix: An R-tool for comprehensive science mapping analysis. *Journal of Informetrics*, 11(4), 959–975. doi: <https://doi.org/10.1016/j.joi.2017.08.007>
- Aria, M., Misuraca, M., & Spano, M. (2020). Mapping the evolution of social research and data science on 30 years of social indicators research. *Social Indicators Research*, 149(3), 803–831. doi: 10.1007/s11205-020-02281-3
- Arsenova, I. (2013). New Application of Bibliometrics. *Procedia - Social and Behavioral Sciences*, 73, 678–682. doi: 10.1016/j.sbspro.2013.02.105
- Baraibar Diez, E., Luna, M., Odriozola, M. D., & Llorente, I. (2020). Mapping social impact: A bibliometric analysis. *Sustainability (Switzerland)*, 12(22), 1–21. doi: 10.3390/su12229389
- Bezák, N., Mikoš, M., Borrelli, P., Alewell, C., Alvarez, P., Anache, J. A. A., Baartman, J., Ballabio, C., Biddoccu, M., Cerdà, A., Chalise, D., Chen, S., Chen, W., De Girolamo, A. M., Gessesse, G. D., Deumlich, D., Diodato, N., Efthimiou, N., Erpul, G., ... Panagos, P. (2021). Soil erosion modelling: A bibliometric analysis. *Environmental Research*, 197. doi: 10.1016/j.envres.2021.111087
- Boeris, C. E. (2012). Aplicación de técnicas de análisis de redes sociales y de co-ocurrencia de palabras en la determinación de frentes de investigación. *Asociación Argentina de Astronomía*, 55.

- Callon, M., Courtial, J. P., & Laville, F. (1991). Co-word analysis as a tool for describing the network of interactions between basic and technological research: The case of polymer chemistry. *Scientometrics*, 22(1), 155–205. doi: 10.1007/BF02019280
- Carrión Mero, P., Montalván Burbano, N., Paz Salas, N., & Morante Carballo, F. (2020). Volcanic geomorphology: A review of worldwide research. *Geosciences (Switzerland)*, 10(9), 1–17. doi: 10.3390/geosciences10090347
- Chiroque, R. (2009). *Análisis bibliométrico en la revista de gastroenterología peruana durante los años 1995 al 2005*. Universidad Nacional Mayor de San Marcos.
- Çiftçi, Ş. K., Danişman, Ş., Yalçın, M., Tosuntaş, Ş. B., Ay, Y., Sölpük, N., & Karadağ, E. (2016). Map of scientific publication in the field of educational sciences and teacher education in Turkey: A bibliometric study. *Kuram ve Uygulamada Eğitim Bilimleri*, 16(4), 1097–1123. doi: 10.12738/estp.2016.4.0009
- Clarivate. (2022, June 9). *KeyWords Plus generation, creation, and changes*. Clarivate. Recuperado de: https://support.clarivate.com/ScientificandAcademicResearch/s/article/KeyWords-Plus-generation-creation-and-changes?language=en_US el 23 de enero del 2024.
- Cobo, M. J., López Herrera, A. G., Herrera Viedma, E., & Herrera, F. (2011). Science mapping software tools: Review, analysis, and cooperative study among tools. *Journal of the American Society for Information Science and Technology*, 62(7), 1382–1402. doi: 10.1002/asi.21525
- Corpus Mendoza, A. N., Ruiz Segoviano, H. S., Rodríguez Contreras, S. F., Yáñez Dávila, D., & Hernández Granados, A. (2021). Decrease of mobility, electricity demand, and NO₂ emissions on COVID-19 times and their feedback on prevention measures. *Science of the Total Environment*, 760. doi: 10.1016/j.scitotenv.2020.143382
- Crozier, M. J. (2010). Deciphering the effect of climate change on landslide activity: A review. *Geomorphology*, 124(3), 260–267. doi: <https://doi.org/10.1016/j.geomorph.2010.04.009>
- Donthu, N., Kumar, S., Mukherjee, D., Pandey, N., & Lim, W. M. (2021). How to conduct a bibliometric analysis: An overview and guidelines. *Journal of Business Research*, 133, 285–296. doi: 10.1016/j.jbusres.2021.04.070
- Echchakoui, S. (2020). Why and how to merge Scopus and Web of Science during bibliometric analysis: the case of sales force literature from 1912 to 2019. *Journal of Marketing Analytics*, 8(3), 165–184. doi: 10.1057/s41270-020-00081-9
- Emmer, A. (2018). Geographies and scientometrics of research on natural hazards. *Geosciences (Switzerland)*, 8(10). doi: 10.3390/geosciences8100382
- Gariano, S. L., & Guzzetti, F. (2016). Landslides in a changing climate. *Earth-Science Reviews*, 162, 227–252. doi: 10.1016/j.earscirev.2016.08.011

- Goh, K. H., & See, K. F. (2021). Twenty years of water utility benchmarking: A bibliometric analysis of emerging interest in water research and collaboration. *Journal of Cleaner Production*, 284, 124711. doi: <https://doi.org/10.1016/j.jclepro.2020.124711>
- González Ciriaco, L. A., & Medina Marín, A. J. (2023). Avances y desafíos éticos en la integración de la IA en la producción científica. *Journal of Scientific Metrics and Evaluation*, 1(1), 48–67. doi: 10.69821/josme.v1ii.2
- González De Dios, J., Moya, M., & Mateos Hernández, M. A. (1997). Indicadores bibliométricos: Características y limitaciones en el análisis de la actividad científica. *Anales Españoles de Pediatría*, 47(3), 235–244.
- González García, J. (2020). Causas, evolución y perspectivas de la guerra comercial para China. *Análisis Económico*, 35, 91-116,.
- Grawin S. (2012). *BiblioTools/BiblioMaps*. Recuperado de: <https://www.sebastian-grauwin.com/bibliomaps/> el 13 de noviembre de 2024.
- Harzing, A. W., & Alakangas, S. (2016). Google Scholar, Scopus and the Web of Science: a longitudinal and cross-disciplinary comparison. *Scientometrics*, 106(2), 787–804. doi: 10.1007/s11192-015-1798-9
- Hirsch, J. E. (2005). An index to quantify an individual's scientific research output. *Proceedings of the National Academy of Sciences of the United States of America*, 102(46), 16569–16572. doi: 10.1073/pnas.0507655102
- Kaiser, M., & Kuckertz, A. (2023). Bibliometrically mapping the research field of entrepreneurial communication: where we stand and where we need to go. *Management Review Quarterly*. doi: 10.1007/s11301-023-00355-3
- Ladrón De Guevara Cervera, M., Hincapié, J., Jackman, J., Herrera, O., Vinicio, C., & Uribe, C. (2008). Revisión por pares: ¿Qué es y para qué sirve? Peer Review: what it's and what it's for? *Salud Uniorde, Barranquilla (Col.)*, 24(2), 258–272.
- Lawani Stephen M. (1980). *Quality, collaboration and citations in cancer research: a bibliometric study*. The Florida State University ProQuest Dissertations & Theses.
- Li, Y., Ji, Q., & Zhang, D. (2020). Technological catching up and innovation policies in China: What is behind this largely successful story? *Technological Forecasting and Social Change*, 153, 119918. doi: <https://doi.org/10.1016/j.techfore.2020.119918>
- Lim, B., Seth, I., & Rozen, W. M. (2023). The role of Artificial Intelligence Tools on advancing scientific research. *Springer*, 48(15), 3036–3038. doi: 10.1007/s00266-023-03526-5
- Lingappa Sangam, S., & Shivayogi Mogali, S. (2012, December). Mapping and visualization softwares tools: a review. *Content Management in Networked Environment*.

- Liu, X., Zhan, F. B., Hong, S., Niu, B., & Liu, Y. (2012). A bibliometric study of earthquake research: 1900–2010. *Scientometrics*, 92(3), 747–765. doi: 10.1007/s11192-011-0599-z
- Martín Martín, A., Orduna Malea, E., Thelwall, M., & López Cózar, E. D. (2018). Google Scholar, Web of Science, and Scopus: A systematic comparison of citations in 252 subject categories. *Journal of Informetrics*, 12(4), 1160–1177. doi: <https://doi.org/10.1016/j.joi.2018.09.002>
- Martín Martín, A., Thelwall, M., Orduna Malea, E., & López Cózar, E. D. (2021). Google Scholar, Microsoft Academic, Scopus, Dimensions, Web of Science, and OpenCitations' COCI: a multidisciplinary comparison of coverage via citations. *Scientometrics*, 126(1), 871–906. doi: 10.1007/s11192-020-03690-4
- Massimo A, & Corrado C. (2016a, October 17). *Bibliometrix*. University of Naples Federico II. Recuperado de: <https://www.bibliometrix.org/home/index.php/layout/bibliometrix> el 23 de octubre del 2024.
- Massimo A, & Corrado C. (2016b, October 17). *Biblioshiny*. University of Naples Federico II. Recuperado de: <https://www.bibliometrix.org/home/index.php/layout/biblioshiny> el 23 de octubre del 2024.
- Melin, G., & Persson, O. (1996). Studying research collaboration using co-authorships. *Scientometrics*, 36(3), 363–377. doi: 10.1007/BF02129600
- Méndez Martínez, C., & Rondón Sepúlveda, M. A. (2012). Introducción al análisis factorial exploratorio. *Revista Colombiana de Psiquiatría*, 41(1), 197–207.
- Mendoza Núñez, V. M. (2022). ¿Cómo se define la literatura gris en el ámbito científico? Su importancia en las revisiones sistemáticas. *Casos y Revisiones de Salud*, 4(2), 5–10. doi: 10.22201/fesz.26831422e.2022.4.2.1
- Molina Montenegro, M. A., & Gianoli, E. (2010). El índice-I, un nuevo estimador del impacto de la productividad científica: Los ecólogos de Chile como caso de estudio. *Revista Chilena de Historia Natural*, 83, 219–227.
- Moral Muñoz, J. A., Herrera Viedma, E., Santisteban Espejo, A., & Cobo, M. J. (2020). Software tools for conducting bibliometric analysis in science: An up-to-date review. *Profesional de La Información*, 29(1). doi: 10.3145/epi.2020.ene.03
- Oliveira Lucas, E., & García Zorita, J. C. (2014). Produção Científica sobre Capital Social: estudo por acoplamento bibliográfico. *Revista da Faculdade de Biblioteconomia e Comunicação da UFRGS*, 20(3), 27–42.
- Price, D. J. D. S. (1963). Little Science, Big Science. *Columbia University Press*. doi: doi:10.7312/pric91844

- Qin Y., Xu Z., Wang X., & Škare M. (2022). Green energy adoption and its determinants: A bibliometric analysis. *Renewable and Sustainable Energy Reviews*, 153. doi: <https://doi.org/10.1016/j.rser.2021.111780>
- Restrepo Botero, J. C. (2020). La revisión por pares (peer review) en las revistas científicas: un proceso que requiere intervención. *Tempus Psicológico*, 3(1), 133–155. doi: <https://doi.org/10.30554/tempuspsi.3.1.3410.2020>
- Romero Pérez, I., & Muñoz Reyes, Y. (2022). Aplicación de la minería de texto a la identificación de enfoques temáticos en un grupo de investigación. *Investigación e Innovación En Ingenierías*, 10(2), 178–195. doi: 10.17081/invinno.10.2.5907
- Ruiz Rosero, J., Ramírez González, G., & Viveros Delgado, J. (2019). Software survey: ScientoPy, a scientometric tool for topics trend analysis in scientific publications. *Scientometrics*, 121(2), 1165–1188. doi: 10.1007/s11192-019-03213-w
- Salton G., & McGill M. J. (1983). *Introduction to modern information retrieval* (Vol. 3). McGraw-Hill Book Company.
- Sinclair, A. J. (1974). Selection of threshold values in geochemical data using probability graphs. *Journal of Geochemical Exploration*, 3(2), 129–149. doi: [https://doi.org/10.1016/0375-6742\(74\)90030-2](https://doi.org/10.1016/0375-6742(74)90030-2)
- Small, H. G., & Koenig, M. E. D. (1977). Journal clustering using a bibliographic coupling method. *Information Processing & Management*, 13(5), 277–288. doi: [https://doi.org/10.1016/0306-4573\(77\)90017-6](https://doi.org/10.1016/0306-4573(77)90017-6)
- Subramanyam, K. (1983). Bibliometric studies of research collaboration: A review. *Journal of Information Science*, 6(1), 33–38. doi: 10.1177/016555158300600105
- Tague, J., Beheshti, J., & Rees-Potter, L. (1981). The law of exponential growth: Evidence, implications and forecasts. *Library Trends*.
- Universidad de Las Palmas de Gran Canaria. (2023, February 23). *Indicadores e índices de la producción científica*. Biblioteca Universitaria. Recuperado de: https://biblioteca.ulpgc.es/factor_impacto el 23 de octubre del 2024.
- Universidad Distrital Francisco José de Caldas. (2023, October 17). *Métricas para medir la calidad y el impacto de las publicaciones científicas*. Biblioteca Universitaria. Recuperado de: <https://udistrital.libguides.com/c.php?g=1321303&p=9720956> el 23 de octubre del 2024.
- Valderrama Zurián, J. C., García Zorita, C., Marugán Lázaro, S., & Sanz Casado, E. (2021). Comparison of MeSH terms and KeyWords Plus terms for more accurate classification in medical research fields. A case study in cannabis research. *Information Processing & Management*, 58(5), 102658. doi: <https://doi.org/10.1016/j.ipm.2021.102658>

- Van Eck, N. J., & Waltman, L. (2014). Visualizing Bibliometric Networks. In *Measuring Scholarly Impact: Methods and Practice* (pp. 285–320). Cham: Springer International Publishing. doi: 10.1007/978-3-319-10377-8_13
- Vera Baceta, M. Á., Thelwall, M., & Kousha, K. (2019). Web of Science and Scopus language coverage. *Scientometrics*, 121(3), 1803–1813. doi: 10.1007/s11192-019-03264-z
- Vessuri, H., Guédon, J.-C., & Cetto, A. M. (2013). Excellence or quality? Impact of the current competition regime on science and scientific publishing in Latin America and its implications for development. *Current Sociology*, 62(5), 647–665. doi: 10.1177/0011392113512839
- Wagner, C. S., & Leydesdorff, L. (2005). Network structure, self-organization, and the growth of international collaboration in science. *Research Policy*, 34(10), 1608–1618. doi: <https://doi.org/10.1016/j.respol.2005.08.002>
- Wallin, J. A. (2005). Bibliometric methods: Pitfalls and possibilities. *Basic and Clinical Pharmacology and Toxicology*, 97(5), 261–275. doi: 10.1111/j.1742-7843.2005.pto_139.x
- Wen Ta, C., & Yuh Shan, H. (2007). Bibliometric analysis of tsunami research. *Scientometrics*, 73(1), 3–17. doi: 10.1007/s11192-005-1523-1
- Wu, X., Chen, X., Zhan, F. B., & Hong, S. (2015). Global research trends in landslides during 1991–2014: a bibliometric analysis. *Landslides*, 12(6), 1215–1226. doi: 10.1007/s10346-015-0624-z
- Yáñez-Dávila, D., Santoyo, E., & Santos-Raga, G. (2023). Worldwide research progress and trends on geothermal water–rock interaction experiments: a comprehensive bibliometric analysis. *Earth Science Informatics*, 1–24. doi: 10.1007/s12145-022-00926-0
- Zhou, P., & Leydesdorff, L. (2006). The emergence of China as a leading nation in science. *Research Policy*, 35(1), 83–104. doi: <https://doi.org/10.1016/j.respol.2005.08.006>
- Zupic, I., & Čater, T. (2014). Bibliometric Methods in Management and Organization. *Organizational Research Methods*, 18(3), 429–472. doi: 10.1177/1094428114562629