Bibliometric Analysis of the Importance of Statistics and Stochastic Models in Flood Prevention for Colombia

Nubia Edith Céspedes Prieto, Camila Alexandra Vargas Buitrago

Escuela de Ingenieros Militares ESING, Ejercito Nacional de Colombia.
Email: necespedesp@unal.edu.co

Abstract

This documentary research and from a bibliometric analysis seeks to analyze the trends related to risk prevention in flood zones for Colombia in a national and international context based on the information reported by Scopus between 2000 and 2023. 292 documents were analyzed, including articles, books and book chapters using VOSviewer software. The units of analysis used were the KeyWords Plus (KW+) keywords, from which bibliometric maps were created applying a methodology based on co-word analysis, grouping techniques and visualization techniques, as well as the analytics provided by the database. The results obtained show four clusters: Stochastic systems, climate change, disaster risks, environmental risks, showing that Colombia requires studies based on statistical data for the regions that are complemented with stochastic models with which policies can be developed. prevention in risk areas for the prevention of disasters generated by natural phenomena such as floods.

Keywords: Stochastics, flood prevention, vulnerable regions.

1. Introduction

The detailed recording of precipitation data in space and time is essential to establish flood risks in a region, but if these are complemented with information related to climate, environmental, social, cultural conditions and are complemented with data related to policies governmental and economic, it is achieved not only to establish vulnerable areas, but also to generate prevention and care mechanisms through planning and management that lead to efficient and effective care. Because the collection of data allows us to know the areas with the greatest vulnerability, facilitates the determination of risk factors and provides the possibility of generating adaptation, prevention and response measures in emergencies. The data facilitates coordination between the different entities in the region, the authorities, communities, organizations and the media.

Studies have shown that the potential of data is achieved through a robust, integrated and updated information system with which it is possible to collect, process and disseminate information from an area or region in a timely and accessible manner, for this it is required of: collecting, processing, analyzing and disseminating data through a culture of data obtained with quality, transparency and integration of data at various levels to turn it into a strategic ally that leads to

the prevention of floods in Colombia to protect the well-being of populations and the environment.

One of the models with the greatest impact corresponds to those developed with stochastic studies because with this model it is possible to analyze the phenomena in a random manner and complement it with factors and phenomena related to the behavior of the climate, its relationship with other natural phenomena such as earthquakes, cyclones. and social, economic, cultural and environmental variables that feed the model and contribute in a way that facilitates the projection of the occurrence of floods in the study area that is significant to disaster prevention because they facilitate the estimation of the probability, frequency and magnitude of events. extremes that lead to disasters or human and material losses, helping to formulate contingency plans, mitigation strategies and early warning systems that reduce the risk and vulnerability of populations exposed to various types of disasters, such as: earthquakes, floods , droughts, forest fires, volcanic eruptions, landslides, tsunamis, hurricanes, among others.

The main finding is the absence of studies of this type in the Andean zone and the lack of applications of stochastic studies in Colombia for the identification of flood risks in the hydrographic basins and the non-use of stochastic models to prevent floods or manage plans. care in vulnerable municipalities

2. Materials and methods

This research is of a documentary nature that is obtained with a bibliometric analysis that is obtained from the quantitative analysis of the publications of international authors related to disaster prevention based on data or statistical models such as stochastics, with this method a descriptive and correlational analysis of the scientific production generated between 2020 and 2023 with which the bibliometric maps are constructed through the following stages: a) collection of information, b) selection of the units of analysis, c) calculation of the frequency of co-occurrence and similarity index between the information units and d) positioning and visualization of the corresponding analysis units on two-dimensional maps (Börner, 2003). To achieve the results, we work with an exhaustive search of articles related to the use of data and stochastic models for the prevention of environmental disasters in the scopus database. The search strategy used corresponded to the use of markers such as. Stochastic AND flood AND risk AND assessment in the titles, summaries or keywords in the period of time delimited between 2020 to 2023, only original articles, books and book chapters were selected (versus clinical notes, reviews, scientific letters or editorials) , because articles or chapters are a type of publication that accurately reflects the proven results of research. The study showed 292 publications between the years 2000 and 2023 (Table 1), of which 280 are articles (article), 10 book chapters (book chapter) and 2 books (book). For quantitative data processing, VOSviewer software (version 1.6.10) was used to analyze Co-authorship, Co-occurrence.

Table 1. Documents Returned from the Scopus Search

Tipo de documento	N. docs	% Tot docs
Article	280	96.5%
Book Chapter	10	3.4%
Book	2	0.1%

Total	292	100%

Own elaboration

Of which a total of 83 KW+ were obtained with a frequency ≥ 10, of these 10 key words with the highest frequency of occurrence were taken as reference (Table 2). After applying the clustering algorithm, with a resolution parameter with a value of 10, they were grouped into four themes, which reflected the degree of similarity of the KW+.

Table 2. Clpuster delimitados en la Búsqueda en Scopus

Clúster	KW+	Frecuencia
Clúster 1	Stochastic Systems	161
	Floods	160
	Flood Control	116
	Stochastic Models	112
	Stochasticity	90
	Uncertainty Analysis	80
	Risk Perception	49
	Risk Analysis	45
	Decision Making	41
	Reservoirs (Water)	39
Clúster 2	Numerical Model	35
	Reservoir	29
	Computer Simulation	24
	Stochastic Simulations	15
Clúster 3	Climate Change	50
	Rain	43
	Hydrological Modeling	32
	Climate Models	20
Clúster 4	Flood Forecasting	18
	Flood Risk Assessments	11
	Risk Assessment	234
	Flooding	71
	Flood	58
	Risk Management	41
	Probability	33
	Flood Risk	31
	Disasters	16
	Disaster Management	11
	Disaster Prevention	10
	Environmental Risk	10
	Rivers	28

Own elaboration

Clúster temático	Referencias
Stochastic Simulations	(Guo et al., 2015), (Avila-Aceves et al., 2023; Guo et al., 2015; Hughes & Guttorp, 1994; Katz & Parlange, 1998; Shinohara & Inatsu, 2023; Wilks, 1998, 1999, Shabestanipour, G., 2023, Rincón, D., 2023),
Use of statistics	(Vittorio Casella & Marica Franzini.. 2015, , Indhu Dasari & Vamsi Krishna Vema. 2023, Hiva Viseh & David N. Bristow., 2023, Lucie P. Et.al., 2022 Tal-Maon,et.al., 2024, Shaikh, M., et.al 2024, Pathan, A., 2024, Patel, A., 2024,))
Climate change	(Benevolenza & DeRigne, 2019; Cea & Costabile, 2022; Ishii & Mori, 2020; Jongman et al., 2012; Modarresi & Maleki, 2023; Wilks, 1999),

Own elaboration

As a result of the labeled bibliometric map, 4 thematic clusters were obtained that defined the main axes of the research, around the importance of data and stochastic studies for disaster prevention. Cluster 1, focuses on stochastic studies, acts as the axis of the map, which groups together topics related to cluster 2, relating the use of statistics compared to cluster 3, focused on the relationship with climate change, and finally cluster 4 refers to to the risks.

The thematic groupings were visualized through a labeled bibliometric map Figure 1. Labeled Bibliometric Map, in which the 17 KW+ were represented

Regarding the bibliometric density map, the size of the KW+ labels and the color of the different areas indicated the importance of the thematic nuclei (Figure 2). They were located, due to their significant density and a correlation

Figure 1. Labeled Bibliometric Map, in which the 17 KW+ analyzed were represented

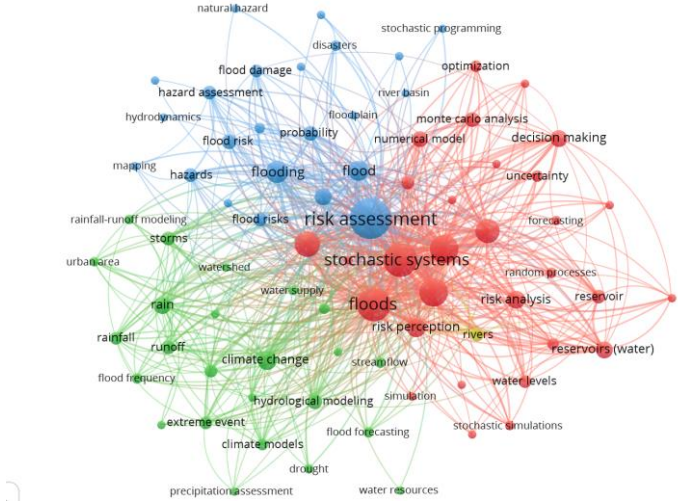
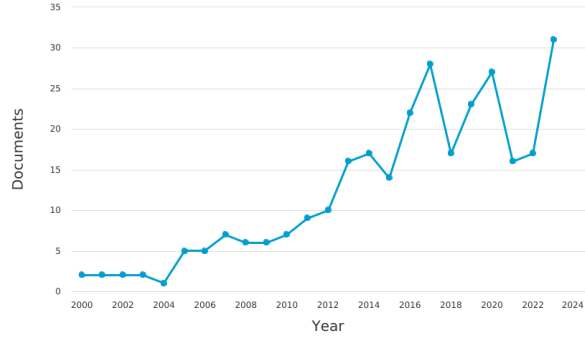
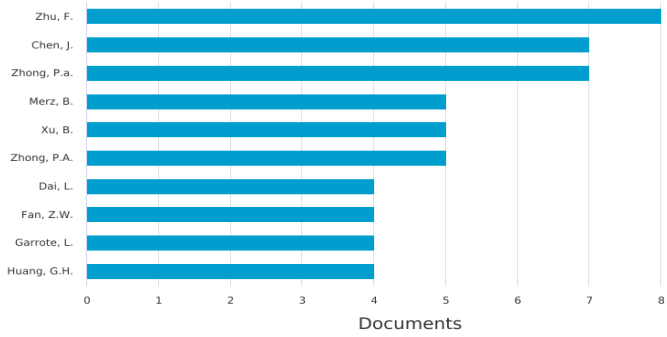


Figure 2. Number of Publications and Year of Publication



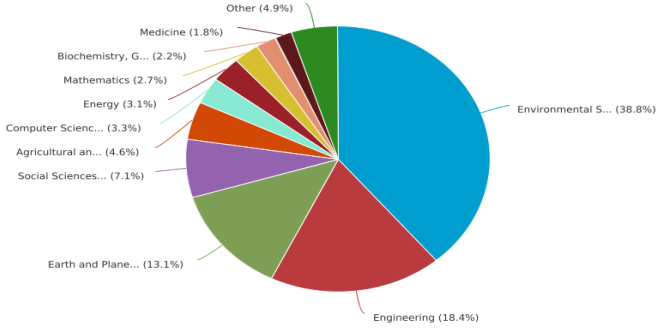
Own elaboration

Figure 3. Frequency of publications by author



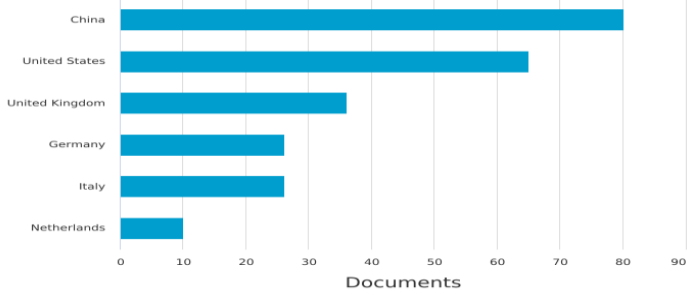
Own elaboration

Figure 4. Areas of knowledge that use these tools the most



Own elaboration

Figure 5. List of countries and publications related to the use of statistical data for disaster prevention with higher numbers of citations.



Own elaboration

Table 3. Most frequently referenced authors.

Autores	Título	Año de la publicación	Nombre de la fuente	N.º de citas
Apel, H., Thieken, A.H., Merz, B., Blöschl, G.	Flood risk assessment and associated uncertainty	2004	Natural Hazards and Earth System Science,	390
Apel, H., Thieken, A.H., Merz, B., Blöschl, G.	A probabilistic modelling system for assessing flood risks	2006	Natural Hazards,	209
Alem, D., Clark, A., Moreno, A.	Stochastic network models for logistics planning in disaster relief	2016	European Journal of Operational Research,	185
Wheater, H.S., Chandler, R.E., Onof, C.J., ...Lourmas, G., Segond, M.-L.	Spatial-temporal rainfall modelling for flood risk estimation	2005	Stochastic Environmental Research and Risk Assessment,	129
Chitsaz, N., Banihabib, M.E.	Comparison of Different Multi Criteria Decision-Making Models in Prioritizing Flood Management Alternatives	2015	Water Resources Management	126
Franks, S.W., Kuczera, G.	Flood frequency analysis: Evidence and implications of secular climate variability, New South Wales	2002	Water Resources Research	125
Salas, J.D., Fu, C., Cancelliere, A., ...Pineda, A., Vincent, E.	Characterizing the severity and risk of drought in the Poudre River, Colorado	2005	Journal of Water Resources Planning and Management,	116
Ward, P.J., Blauhut, V., Bloemendaal, N., ...Veldkamp, I.E.T., Winsemius, C.H.	Review article: Natural hazard risk assessments at the global scale	2020	Natural Hazards and Earth System Sciences,	105
Nadal, N.C., Zapata, R.E., Pagán, I., López, R., Agudelo, J.	Building damage due to riverine and coastal floods	2010	Journal of Water Resources Planning and Management,	99
Jamali, B., Bach, P.M., Deletic, A.	Rainwater harvesting for urban flood management – An integrated modelling framework	2020	Water Research,	96
Sharif, M., Burn, D.H.	Simulating climate change scenarios using an improved K-nearest neighbor model	2006	Journal of Hydrology,	95

3. Discussion of results

When reviewing the results obtained on the trends in the use of data to establish flood risks, a growth trend of more than 10 points is evident for the last 12 years and the countries that report the most research on this topic are China and United States, with a range of 80 -90 publications and the absence of production on this topic for South American countries and Colombia is evident, reflecting a potential for scientific production that is required for the study of this disaster prevention opportunity for countries on the continent. South America and especially Colombia.(Benevolenza & DeRigne, 2019) The studies show a relational trend with the influence of climate change and the potential risk of flooding for the rural and urban sector, substantially increasing interest in generating predictions for Europe.(Cea & Costabile, 2022)

Flood risk assessment is achieved with the identification, analysis and assessment of possible effects generated by extreme hydrological events that directly or indirectly affect a vulnerable population, its assets and the environment, which is why it is required to identify and manage optimal processes of collecting, processing, storing and analyzing information in the sector of interest that lead to the characterization of the threats, vulnerabilities and exposures of the different elements at risk of a specific sector or region. This is done in order to establish risk zones based on data management that provides a scientific and technical basis for making decisions related to the prevention, mitigation and modification of scenarios susceptible to flooding and the design and implementation of emergency and recovery plans for affected territories. (Jongman et al., 2012)

Hydrological modeling is a fundamental tool for the analysis and prevention of risks and floods associated with extreme weather phenomena. Hydrological modeling consists of the mathematical representation of the processes that occur in the hydrological cycle, such as precipitation, runoff, infiltration, evapotranspiration and water storage in soil and aquifers. (Barbero et al., 2022) Hydrological modeling allows estimating the quantity and spatial and temporal distribution of water in a hydrographic basin, as well as its response to different climate change, land use or water resource management scenarios. (Avila-Aceves et al., 2023) Hydrological modeling requires input data of good quality and sufficient quantity, which come from various sources such as weather stations, satellites, radars or remote sensors. (Alsumayt et al., 2023) The management of this data involves its collection, validation, storage, processing and analysis, as well as its integration and harmonization with hydrological models. (Shinohara & Inatsu, 2023) Data management is a key aspect for the success of hydrological modeling, since the reliability and precision of the results obtained depends on it. (Jongman et al., 2012).

It is of interest to find that several studies in countries such as England, the United Kingdom and the United States, among others, value the use of the Poisson distribution as the statistical and probability tool that facilitates modeling to establish the probability of flooding occurring at any given time. from the record in a space and time interval that allows modeling from the exponential distribution, as a continuous distribution that leads to determining the critical threshold that would cause the river to overflow, facilitating the study of hydrological phenomena in the region. The achievement of this information is useful tools for the design of planning measures, which contemplate the administration to mitigate the effect of this type of natural disasters. (Modarresi & Maleki, 2023)

Preventing flood risks depends not only on factors such as the space and time of rainfall occurrence, but also involves the intensity, duration of precipitation, the hydrological and hydraulic characteristics of the basin, land use, infrastructure and vulnerability. of the population and both natural and artificial resources and that are directly affected by environmental or climatic conditions and that in the specific case, climate change is a variable that has a direct impact, which is why research leads to the need for the use of stochastic models. , which incorporate the temporal and spatial variability of the occurrence rate as well as the correlation between events, which leads to estimating the risk of flooding and its uncertainties, making it easier for local administrations to determine the risk, plan, organize and act more efficiently at the moment. to attend to the emergency. Jerome G.F. (2020).

Studies related to statistics applied to hydrological behavior and the determination of flood risks have focused on regions that, due to their longitude and latitude, have the four seasons, but not on areas close to the equatorial region where the behavior of rain and drought varies. less regularly than what happens with the seasons, so applying this type of statistical studies and stochastic model in Colombia is of interest and that stochastic models would be more viable to project flood risks. The modeling achieved by studying the water dynamics of the basin where the observed and simulated hourly rainfall is recorded leads to results that show that the stochastic model for the study of rainfall at a point scale including point extremes and spatial correlations between sites and leads to establishing flood frequency curves on a daily scale with an average relative bias of 0.36% to 16.9% at 10-year return levels confirming their potential for use to infer long-term flood risk for basins of size medium for the UK.

The relationship between climate change and floods is directly proportional and there are several studies aimed at determining their incidence and especially the projection of disaster risks as a consequence of climate change and the vulnerability that rural and urban sectors may have at a global level, for example. Therefore, the help of stochastic models to generate possible scenarios from historical, statistical data or numerical simulations leads to the evaluation of risks for vulnerable areas (Cea & Costabile, 2022). However, there are several studies that have shown that the models Stochastics present limitations and challenges such as the quality and availability of the data, as well as the complexity and uncertainties of the physical, chemical, environmental, social or cultural processes for the validation and projection of results. Related to the effects of climate change. (Apel H, 2004). Effects that transcend global effects that involve several regions and that require being studied to generate prediction and preparation mechanisms interregionally, expanding the sources and variables of information and the use of mega data that facilitate the projection and use of stochastic models effectively and efficiently (Jongman et al., 2014).

4. Conclusions:

Within the review and bibliometric analysis, it is established that studies related to the identification of flood risks using stochastic statistical models are very few and with a tendency to increase in the last five years. That the studies focus on northern areas, but few or none in Andean areas and specifically for Colombia, these types of studies are not reported, becoming an opportunity to apply this type of studies in vulnerable areas of Colombia and strengthen the study for the prevention or response to floods in Andean areas.

Studies that involve statistical analysis and that are worked from stochastic models are a viable alternative to work and project management regarding the prevention and attention of flood disasters.

For Colombia, it is necessary to collect and systematize periodic information on the behavior of rainfall in vulnerable areas with which stochastic models can be worked and must be complemented with information on the context, social, cultural, economic, agricultural, among others in order to feed the databases and use them to process the information and generate risk maps and include them for risk management and attention plans in Colombian territory.

Risk studies around hydrological basins must be complemented with studies of social, cultural, economic and political behavior, in order to better assess the risks and clearly project disaster management and attention in the identified areas of the basin.

A potential use for the projection and management of flood risk management for Colombia is identified in stochastic models.

European and North American studies demonstrate the opportunity for success in the prevention and projection of administrative management, to address risks due to environmental disasters where there are several cases of success in the face of floods.

Climate change is a phenomenon that affects the world population and requires attention and especially prevention and monitoring studies, in order to prevent environmental disasters.

For stochastic models to be effective, meticulous monitoring and trajectory of the type is required, so it is advisable to articulate the information generated by Colombia through IDEAM, Agustín Codazzi, DANE, MINAMBIENTE and the local information they collect. the governorates and mayors' offices, in order to consolidate a database of several years and interacting elements, which lead to the strengthening of flood risk prevention and management in Colombia.

Thanks and recognition

Thanks to the support and financing of the School of Military Engineers of the National Army of Colombia.

WORKS CITED

Alsumayt, A., El-Haggar, N., Amouri, L., Alfawaer, Z. M., & Aljameel, S. S. (2023). Smart Flood Detection with AI and Blockchain Integration in Saudi Arabia Using Drones. Sensors, 23(11). https://doi.org/10.3390/s23115148

Avila-Aceves, E., Plata-Rocha, W., Monjardin-Armenta, S. A., & Rangel-Peraza, J. G. (2023). Correction to: Geospatial modelling of floods: a literature review (Stochastic Environmental Research and Risk Assessment, (2023), 37, 11, (4109-4128), 10.1007/s00477-023-02505-1). Stochastic Environmental Research and Risk Assessment, 37(11), 4539. https://doi.org/10.1007/s00477-023-02538-6.

Apel, H., Thieken, A. H., Merz, B. y Blöschl, G.: Evaluación del riesgo de inundaciones e incertidumbre asociada, Nat. Peligros Sistema Tierra Sci., 4, 295–308, https://doi.org/10.5194/nhess-4-295-2004, 2004.

Barbero, G., Costabile, P., Costanzo, C., Ferraro, D., & Petaccia, G. (2022). 2D hydrodynamic approach supporting evaluations of hydrological response in small watersheds: Implications for lag time estimation. Journal of Hydrology, 610. https://doi.org/10.1016/j.jhydrol.2022.127870

Benevolenza, M. A., & DeRigne, L. (2019). The impact of climate change and natural disasters on vulnerable populations: A systematic review of literature. Journal of Human Behavior in the Social Environment, 29(2), 266–281. https://doi.org/10.1080/10911359.2018.1527739

Cea, L., & Costabile, P. (2022). Flood Risk in Urban Areas: Modelling, Management and Adaptation to Climate Change: A Review. Hydrology, 9(3). https://doi.org/10.3390/hydrology9030050

Jongman, B., Ward, P. J., & Aerts, J. C. J. H. (2012). Global exposure to river and coastal flooding: Long term trends and changes. Global Environmental Change, 22(4), 823–835. https://doi.org/10.1016/j.gloenvcha.2012.07.004

Modarresi, S. A., & Maleki, M. R. (2023). Integrating pre and post-disaster activities for designing an equitable humanitarian relief supply chain. Computers and Industrial Engineering, 181. https://doi.org/10.1016/j.cie.2023.109342

Shinohara, M., & Inatsu, M. (2023). Stochastic Precipitation Model Using Large Ensemble Data. Journal of Disaster Research, 18(8), 868–876. https://doi.org/10.20965/jdr.2023.p0868

Jiang, X., Zhang, L., Liang, Z., Fu, X., Wang, J., Xu, J., Zhang, Y., & Zhong, Q. (2023). Study of early flood warning based on postprocessed predicted precipitation and Xinanjiang model. Weather and Climate Extremes, 42. https://doi.org/10.1016/j.wace.2023.100611

Mulonga, J., & Olago, D. (2023). Perceptions of climate variability and change in coastal: The case of mangrove-dependent communities in the Lower Tana Delta. Environmental Challenges, 13. https://doi.org/10.1016/j.envc.2023.100799

Tal-maon, M., Broitman, D., Portman, M. E., & Housh, M. (2024). Combining a hydrological model with ecological planning for optimal placement of water-sensitive solutions. Journal of Hydrology, 628. https://doi.org/10.1016/j.jhydrol.2023.130457

Shaikh, M. P., Yadav, S. M., & Manekar, V. L. (2024). Flood Hazards Mapping by Linking CF, AHP, and Fuzzy Logic Techniques in Urban Areas. Natural Hazards Review, 25(1). https://doi.org/10.1061/NHREFO.NHENG-1716

Pathan, A. I., Agnihotri, P. G., Patel, D., Gandhi, P. J., Prieto, C., Mohseni, U., & Patidar, N. (2024). Use of Big Data for Flood Assessment Through HEC-RAS Model: A Study of Purna River of Navsari. In Lecture Notes in Civil Engineering: Vol. 340 LNCE. https://doi.org/10.1007/978-981-99-1890-4_7

Patel, A., Keriwala, N., & Yadav, S. M. (2024). 2D Flood Simulation and Mapping Using Hydraulic Model and GIS Technology. In Lecture Notes in Civil Engineering: Vol. 340 LNCE. https://doi.org/10.1007/978-981-99-1890-4_6

Shabestanipour, G., Brodeur, Z., Farmer, W. H., Steinschneider, S., Vogel, R. M., & Lamontagne, J. R. (2023). Stochastic Watershed Model Ensembles for Long-Range Planning: Verification and Validation. Water Resources Research, 59(2). https://doi.org/10.1029/2022WR032201

Rincón, D., Velandia, J. F., Tsanis, I., & Khan, U. T. (2022). Stochastic Flood Risk Assessment under Climate Change Scenarios for Toronto, Canada Using CAPRA. Water (Switzerland), 14(2). https://doi.org/10.3390/w14020227

Indhu Dasari & Vamsi Krishna Vema. (2023) Evaluación de la incertidumbre estructural de los modelos hidrológicos y su impacto en el mapeo de inundaciones. Revista de Ciencias Hidrológicas 68:16, páginas 2404-2421.

Hiva Viseh & David N. Bristow. (2023) Riesgo de inundación residencial en Metro Vancouver debido al cambio climático utilizando cuadros de probabilidad. Revista Internacional de Gestión de Cuencas Hidrográficas 0:0, páginas 1-18.

Lucie Pheulpin, Nathalie Bertrand & Vito Bacchi. (2022) Cuantificación de la incertidumbre y análisis de sensibilidad global con parámetros de entrada dependientes: aplicación a un modelo hidráulico 2D básico. LHB 108:1.

Vittorio Casella & Marica Franzini. (2015) Estandarización de cifras y procedimientos de evaluación de la precisión vertical del MDT. Geomática, peligros y riesgos naturales 6:5-7, páginas 515-533.

Masoume Behrouz & Saeed Alimohammadi. (2023) Gestión de riesgos y evaluación de la coordinación de componentes en el diseño de un sistema compuesto de defensa contra inundaciones fluviales. Revista Internacional de Gestión de Cuencas Hidrográficas 0:0, páginas 1-17.

Omar M. Nofal & John W. van de Lindt. (2022) Comprensión del riesgo de inundaciones en el contexto del modelado de resiliencia comunitaria para el entorno construido: necesidades y tendencias de investigación. Infraestructura sostenible y resiliente 7:3, páginas 171-187.

Jim Hall y Dimitri Solomatine. (2008) Un marco para el análisis de incertidumbre en las decisiones de gestión del riesgo de inundaciones. Revista Internacional sobre Gestión de Cuencas Hidrográficas 6:2, páginas 85-98.

Jerome G. F. (2020); Flood Disaster Hazards; Causes, Impacts and Management: A State-of-the-Art Review Natural Hazards - Impacts, Adjustments and Resilience, páginas 1-18.