Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 245 (2024) 255–262

Procedia Computer Science

www.elsevier.com/locate/procedia

9th International Conference on Computer Science and Computational Intelligence 2024 (ICCSCI 2024)

Nonparametric Predictive Inference Forest Fire Dashboard

Amirah Hazwani Roslin[a], Noryanti Muhammad[a]*

[1]*Centre for Mathematical Sciences, Universiti Malaysia Pahang Al-Sultan Abdullah, Lebuh Persiaran Tun Khalil Yaakob, Kuantan, 26300, Pahang, Malaysia*

Abstract

Forest fires have significantly increased over the years, leading to serious and expensive damages. These damages can be minimised, through information sharing before the forest fire occurrence using a dashboard. The existing forest fire dashboards' main function is for monitoring based on past historical data. It will be helpful for many communities to know about the forest fire possibility of happening early for proper damage control precautions. Although the current forest fire prediction models are highly accurate, they still struggle with some issues. Hence, the Nonparametric Predictive Inference Forest Fire (NPIFF) dashboard was introduced in this paper to offer a new perspective on forest fire prediction study. It also included a predictive function using the R programming language software as its main task. The NPIFF dashboard utilised Malaysia's and Indonesia's past wildfire locations datasets to generate an imprecise probability of the next forest fire event. The results of a novel method, nonparametric predictive inference (NPI) with copula were displayed in the NPIFF dashboard using the new R algorithm. Then, the NPIFF dashboard was published online via the shinyapps.io server after it was successfully developed and functional. The information from the NPIFF dashboard can be relied on since it provides certainty confidence and portrayed highly accurate results which are useful for several organisations for various reasons. NPIFF dashboard will consider several improvements in the future to offer more information for more efficient forest fire prediction.

Keywords: dashboard; forest fire; Malaysia and Indonesia; nonparametric predictive inference; R software

* Corresponding author. Tel.: +609-4316678.
E-mail address: noryanti@umpsa.edu.my

1877-0509
10.1016/j.procs.2024.10.250

1. Introduction

A forest fire or wildfire is one of the natural disasters that cause huge damage to multiple aspects at once such as in economy, environment and living beings, both humans and animals. It is a seasonal disaster that commonly occurs in dry countries with extreme heat and its occurrence trend has risen over the past 20 years. Particularly, this past decade had a concerning surging trend of wildfire due to the technological advancement in human consumption and recent El-Niño events. This would lead to more serious damages such as property wreckage, vegetation loss, severe injuries and deaths which require expensive repairing expenses and compensations [1, 2]. Hence, many related parties in the forest fire events wanted to minimise these damages as much as possible including the government agencies, business investors and local villagers in nearby wildfire hotspots. Previously, forest fire occurrence detection relied on human expertise, judgement and skills, which can be restricted, even with the help of technology. In today's data-driven world, many forest fire responders and policymakers started to turn to the more scientific method in analysing past data to predict future forest fire events. However, understanding the intricate analysis behind the prediction results can be complicated and incomprehensible to those who are involved with the forest fire phenomenon.

Subsequently, the public and forest firefighters need a platform that is straightforward for them to understand. The dashboard is one of the most organised and powerful visualisation tools for presenting data and result summaries to others on one screen. It is the top choice among professional and public users to be briefly informed on a certain situation in a short amount of time. Also, a dashboard is crucial for data comparison and to simultaneously monitor differences between multiple groups or conditions. There are some main qualities of dashboard development design which is to keep it simple, relevant, direct and concise to ensure its user-friendliness feature. Interactive dashboards enable users to interact and manipulate the data to produce the desired result as per their interests [3, 4]. Some dashboards are even equipped with downloading and report generation functions to help their users save their customised results for future reference. According to [3], there are three types of dashboards namely, strategical, analytical and operational. A strategical dashboard helps management in the decision-making process and strategy formulations. As opposed to the operational dashboard which carries the same concept but requires more attention as it deals with constantly changing circumstances. Analytical dashboards focus on interactions between layers of data. Regardless, these dashboards are informative to their users and appropriate to their needs.

Numerous past studies have ventured into mathematical and statistical methods application in forest fire prediction studies. Most of them portray excellent performance in analysing complex ignition factors to predict forest fire occurrences. Nonetheless, it is undeniable that there is still room for improvement in dealing with uncertainties in wildfire prediction. This can hinder the prediction result reliability despite they were highly accurate. The spatial correlation influence in wildfire location prediction was highlighted by [5] to enhance the prediction accuracy result. Besides, the existing forest fire dashboard that is publicly accessed mainly provides the historical events of forest fires regardless of their locations which tend to be specific and limited too. Only a few of them include the predictive function but as a side component. Those only apply basic forecasting methods that most likely ignore other influential ignition factors when forecasting the next forest fire event. The information gained from these dashboards may be insufficient for related parties to prepare themselves to face the wildfire prior to its occurrence. This situation makes it difficult to control forest fire damage.

Therefore, this paper offers a recommendation based on the current forest fire studies and dashboard circumstances. A dashboard solely with predictive function as its main feature is developed with a novel method in forest fire studies. This paper introduces a new algorithm that extends from the [6] and [7] for predicting the next wildfire hotspot. It is based on a nonparametric predictive inference (NPI) with a copula framework for bivariate data outlined by [8] but they only focused on the parametric copulae integration. As a brief explanation of the new method introduced in [6] and [7], NPI quantifies the uncertainties of the next forest fire occurrence by providing an imprecise probability $(\underline{P}, \overline{P})$, comprised of lower probability $(\underline{P})$ and upper probability $(\overline{P})$. Meanwhile, the copula integration considers the inter-correlation between the bivariate data, which were the latitude and longitude data. At this stage, the new method considers the spatial correlation as suggested by [5] to enhance the prediction accuracy result. Thus, the NPI with parametric copula result is anticipated to be more accurate than the existing predictive models. These results were visualised in an interactive dashboard, Nonparametric Predictive Inference Forest Fire (NPIFF) dashboard, which can be accessed online. The NPIFF dashboard algorithm was developed using the R programming software with proper R packages. For starters, the NPIFF dashboard presented the proposed method's results for Malaysia's and Indonesia's next wildfire location based on their past forest fire locations data in 2020.

2. Method

R is a free and open-source software that is flexible in data analytics and data-sharing tasks. At first, the NPI with parametric copula analysis was performed via R software before developing the NPIFF dashboard. The parametric copula parameters used in this study were the Normal, Clayton, Gumbel and Frank copulae. They were estimated using the maximum likelihood estimator method which later were integrated with the NPI. Afterward, the $(\underline{P}, \overline{P})$ was generated which indicated the probability of the next forest fire occurrence at all locations in Malaysia and Indonesia. The errors or differences $\bar{d}$ within the $\underline{P}$ and $\overline{P}$ gap quantifies the uncertainties of the next forest fire event which reflects its accuracy. Afterward, the interactive NPIFF dashboard was built using the R software as an extension of the existing NPI with parametric copula analysis background in the R software environment.

Several packages are required to develop the NPIFF dashboard. Primarily, the 'shiny' package can easily build the interactive web dashboard application with a quick 'reactive' interactivity between the input and output. The page layout and components within the package are easy to customize depending on the choices made in the dashboard algorithm structure. The output rendering is available for multiple types of graphical, tabular, numerical and string formats [9]. Alongside that, the 'shinythemes' package is always used with the 'shiny' package for aesthetic purposes. There are several templates themes in this package to use and allow customization as well [10]. The NPIFF dashboard utilizes the "journal" theme from the 'shinythemes' package. As for the output panel, the 'data.table' package specifies building the data frame based on the input by the users. This package is suitable for manipulating and viewing the data frame faster than other tabular types format [11]. Apart from that 'leaflet' package offers interactive map output for viewing and user control such as dragging, zooming and marking. It is compatible to deploy with the 'shiny' package [12]. Lastly, the 'magrittr' package allows the simplification of chaining commands with the forward-pipe operator [13].

The datasets employed in this study were extracted from the Moderate Resolution Imaging Spectroradiometer (MODIS) satellite. Malaysia (2020) and Indonesia (2020) datasets contain several attributes of environmental surroundings during the forest fire events throughout the year 2020. Since the main focus of the NPIFF is the location, only the longitude (x) and latitude (y) attributes were extracted as the bivariate data to consider the spatial correlation. Both attributes were recorded as degrees East (°E) and degrees North (°N), respectively. The notation for the raw data in this paper is (y, x), follows the standard coordinate system order. Whenever the (y, x) values reached the negative value, they were signified as degrees West (°W) for x and degrees South (°S) for y values. The coordinate input enables the NPIFF to be applied to different countries. Initially, only Malaysia and Indonesia were available in the current NPIFF. Then, the association between the bivariate data was demonstrated as a threshold (t) value, which is the sum of x and y. The unique t value was the primary key in determining the output result displayed on the NPIFF dashboard page.

The finished algorithm for the NPIFF dashboard was outlined in a single R Markdown (Rmd) file, which is the suitable file format for the sharing function. There were three code chunks in the Rmd file. At the top of the file, yet another markup language (YAML) code chunk defined the title, output type and runtime. Each of the elements in the YAML header was listed as "Forest Fire Prediction with NPI", "html_document" and "shiny" respectively. Secondly, the code chunk is for loading the data environment (RData) to be presented in the NPIFF dashboard. The RData file consisted of the (y, x) raw data, all copula parameter values and $(\underline{P}, \overline{P})$ and $\bar{d}$ for Malaysia and Indonesia from the NPI with parametric copula analysis results. The last code chunk was the shinyApp(ui, server) development coding in dictating the information setting, layout and user interface (UI) of the NPIFF dashboard. Note that all code chunks needed to define 'include=FALSE' to ensure the codes were hidden from the final display. Then, the Rmd file was knitted and published online through the shinyapps.io server. The related R files were uploaded to the shinyapps.io server and a link was generated for shared accessibility.

3. Result and Discussion

The NPIFF dashboard UI layout is divided into two parts; input and output sections that are located at the left and right sides concurrently. At this initial development stage, the NPIFF dashboard only has a single tab incorporated of the title ("Forest Fire Prediction with NPI"), year ("2020"), input section widgets and output section result (Fig. 1). The UI design for this dashboard was targeted to the non-academic communities who lack understanding in statistical knowledge. Among the communities are the forest firefighter departments, business investors, local villagers and others. The forest firefighter department can devise their action planning appropriately with the input location demographic and the $\underline{d}$ provide confidence to them to prepare before deploying to the fire scene. As for the business investors in a certain field can plan their financial and supply management strategy based on the resource availability that is related with the forest fire events such as in agricultural or furniture manufacturing sectors [14]. Another example is the local human settlements near the forest fire hotspot can try to limit the fire spread as far as possible from reaching their villages or farms. Apart from that, there are other implementations in more communities affected by wildfires either directly or indirectly that can benefit from the NPIFF dashboard information.

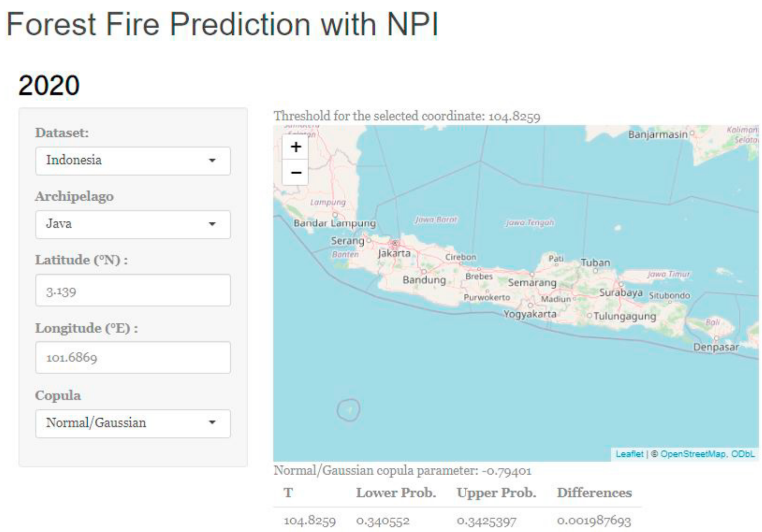

Fig. 1. NPIFF Dashboard User Interface

Firstly, the input section widgets are compiled within a sidebar panel on the right comprised of the selection list for dataset or country choices, two numeric entry boxes of y and x coordinates and another selection list for copula type choices. Moreover, an additional selection list of archipelago choices will become visible depending on a certain selected choice for the dataset selection list as portrayed in Fig. 1. Further metadata about these widgets inputs are listed in Table 1. The default selection of each widget is "Malaysia", 3.139, 101.6869 and "Normal/Gaussian" in that order. The users are required to select the country before entering the (y, x) values to know about the probability of the next forest fire event at the location of interest. They can either directly write the (y, x) values into the boxes or by clicking the up and down arrow at the end of the boxes which consider up to 4th decimal places interval. Next, the users need to choose a copula family type that delivers different probability prediction results. The optimal copula is the copula that contributes to the most accurate $(\underline{P}, \overline{P})$ prediction result after being integrated with the NPI. To show the NPIFF dashboard functionality, this paper provides the guidelines for using this application by generating the $(\underline{P}, \overline{P})$ for Pekan, Pahang, which is one of the areas with the highest total tree cover loss due to the forest fire in Malaysia. The coordinate of the desired location is (3.5221°N, 103.2754°E), an area where there were a few nearby villages. Note that, the optimal copula for a majority of Malaysia locations is the Gumbel copula. Fig. 2 shows the appropriate input based on the Pekan example.

Table 1. Input Widgets Metadata.

Widget Components	Actions	R command
Dataset:	Malaysia, Indonesia	selectInput("d", …)
Archipelago:	Borneo, Java, Maluku, Lesser Sunda Islands, Sulawesi, Sumatra, Western New Guine	conditionalPanel(condition = "input.d == 'Indonesia'", selectInput("g", …))
Latitude (°N):	• Allowed range: [-13, -8] • Step: 4 decimal places	numericInput("y", …)
Longitude (°E):	• Allowed range: [94, 141] • Step: 4 decimal places	numericInput("x", …)
Copula:	Normal/Gaussian, Clayton, Frank, Gumbel	selectInput("c", …)

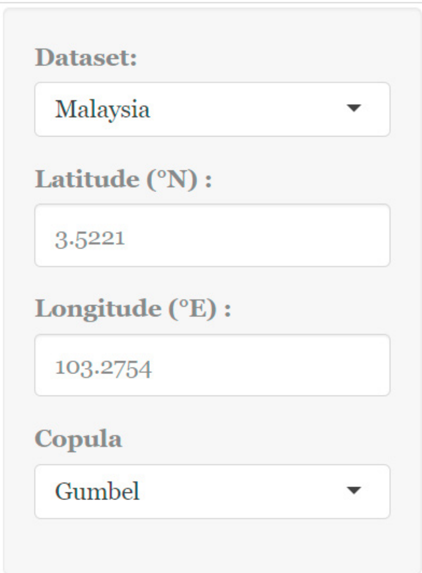

Fig. 2. Pekan Input Section in NPIFF Dashboard

Next, the output section is in the main panel layout where the string, graphical and tabular results are assembled. The components in the output section are the t value, interactive map, copula parameter value and a 4-column probabilities table. Table 2 explains the details of the output results. All output components depend on the input entry in Fig. 2 earlier. Different combinations will generate different sets of output. The only numerical results that are presented in string format are t and copula parameter values. These values are only for statistical reference in generating the output that is informative to the users. Hence, these values can be ignored by non-academic users. Other results are either translated into visualization or organized inside a table. By default, the interactive map feature had fix bound view based on the country or archipelago coordinate limit to get an overall land view on one screen. The user can adjust the view by zooming in or out accordingly and dragging the map view as they prefer. Additionally, the colour intensity within the circle indicated the probability of forest fire occurrence at the marked spot. The users need to be alert if the colour gets darker as it indicates a higher probability of forest fire happening at the selected location. Following this, the tabular results provide a detailed interpretation of the visual information. The $(\underline{P}, \overline{P})$ derived from the second and third columns reflect the colour intensity from the map while the $\overline{d}$ portrays the uncertainties for the users' confidence to rely on the prediction result. Fig. 3 was the result of the Pekan input earlier.

Table 2. Output Results.

Result Components	Actions	R command
Threshold	input$x + input$y	renderText(…)
Interactive Map	• Marker: (input$y, input$x) coordinate • Circle: Range of other locations covered by the same *t* value • Red intensity: Darker red = higher probability	renderLeaflet(…)
Copula Parameter	• Depend on input$c • Show copula type and parameter value	renderText(…)
Probabilities Table	Column: $t, \underline{P}, \overline{P}, \bar{d}$	renderTable(…)

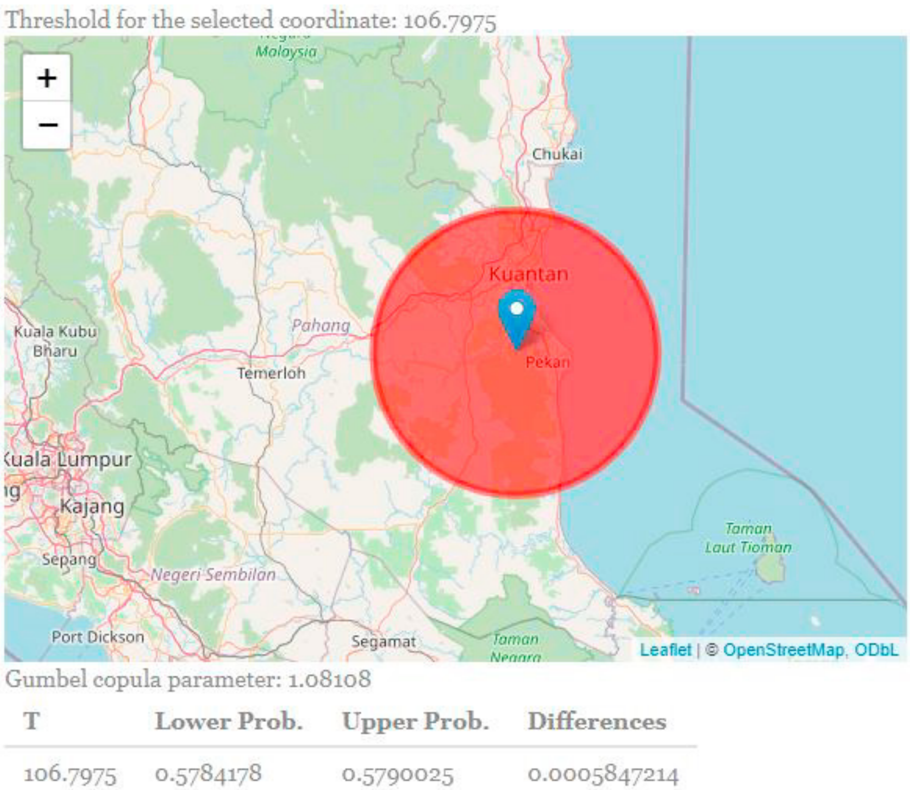

Fig. 3. Pekan Output Section in NPIFF Dashboard

The result interpretations from Fig. 3, the first output was the *t* value of 106.7975° which is the sum of Pekan's *y* and *x* coordinates (3.5521°N + 103.2754°E). This value would be the main reference in extracting data from the RData file to build a new data frame later. The blue marker on the map pinpoints the (3.5521°N, 103.2754°E) coordinate from the numeric inputs in Fig. 2, in the meantime, the circle represents nearby locations that have the same *t* value. The circle has a 111km radius from the blue marker which is equivalent to approximately 1° for both (*y*, *x*) axes. The $(\underline{P}, \overline{P})$ displayed at the bottom table applied to other locations within the circle range too. Moving to the next output was the copula parameter value of the Gumbel copula parameter that was integrated with the NPI framework in constructing the $(\underline{P}, \overline{P})$ for the next forest fire occurrence at Pekan. Lastly, the probabilities table was created based on the t value. Simultaneously, $(\underline{P}, \overline{P})$ for Pekan was (0.5784, 0.5790), which was interpreted as there is a 57.84% to 57.9% chance that the next forest fire will occur at Pekan. The colour intensity could be inferred as the medium level as it was neither too dark nor too faint. Although not an alarming probability, it is more likely the next forest fire will happen at Pekan based on the NPIFF dashboard result. Furthermore, the $\overline{d}$ value shows a very minimal error indicating a near-perfect and highly accurate prediction result to rely on. From this interpretation, the related parties will proceed accordingly depending on their objective and roles in the forest fire disaster.

4. Conclusion

In conclusion, this study provided a new dashboard development using a novel method in the forest fire study. The NPI with parametric copula is a valid alternative and it provides a new perspective in forest fire prediction studies by considering a nonparametric method. Also, the spatial correlation consideration with the copula integration enhances the prediction result to be more accurate in generating the probability of the next forest fire occurrence. This NPI with parametric copula analysis result can provide useful insight and information to those organizations or communities that are involved with the forest fire phenomenon. On top of that, the alarming rising trend of forest fire occurrence in this past decade motivates them to acquire information from calculated and reliable sources before proceeding with any sort of planning. The most effective way to share wildfire information is through dashboard sharing. However, the existing forest fire dashboards tend to focus on past trends instead of future events. Knowing the forest fire occurrence probabilities beforehand can help many organizations reduce forest fire damage and cost at the same time. Aside from that, the existing dashboards suffer from information overload within a single screen until it can become overwhelming and needs some time to understand their functionalities. Other current issues outlined by [4] that are present in the forest fire dashboards are colour overuse, unclear icon usage, lack of context and others. [3] recommended several considerations in improving the dashboard design and practicality such as minimalistic and user-friendly UI, colour consideration, interactivity on location function and data analytics prediction.

Hence, this paper introduced the NPIFF dashboard with predictive function as its main feature that offers most of [3] considerations in its algorithm to fill the gap identified by [4]. The NPIFF dashboard was developed using the R software along with the NPI with parametric copula analysis due to its flexibility in conveniently creating an interactive dashboard and deploying it online. A brief NPIFF dashboard development codes and methodology was shared in this paper. Also, the guideline for the NPIFF dashboard user to operate this application is included. Last but not least, the interpretation of the output result in the NPIFF dashboard can help the involved related parties in utilizing the information gained in the real-life implementation such as resource management planning, action strategy formulation, and other prevention arrangement. The NPIFF dashboard only has basic functionalities so far due to it being a newly developed dashboard. In future works, the NPIFF dashboard may incorporate other ignition attributes, different countries and more advanced features that can provide more information regarding the forest fires nature before their occurrence.

Acknowledgements

The authors thank the Universiti Malaysia Pahang Al-Sultan Abdullah (UMPSA) for providing financial support under the International Matching Grant UMP RDU222706 (UIC221519) and the UMPSA facilities. The authors also would like to thank the reviewers for the valuable comments for improvements to this paper. The NPIFF dashboard access and resources of the related R files are not shared in this paper but are available upon request.

References

[1] Gong A, Li J, and Chen Y. A Spatio-Temporal Brightness Temperature Prediction Method for Forest Fire Detection with MODIS Data: A Case Study in San Diego. Remote Sens. 2021;13(15).

[2] Adji NFS. Forest fires in Kalimantan: study case Central Kalimantan. Mangrove Watch. 2024;1(1):1-6.

[3] Chatzos D. Project for the development of a higher education management dashboard in R: Universitat Politècnica de Catalunya; 2018.

[4] Bach B, Freeman E, Abdul-Rahman A, Turkay C, Khan S, Fan Y, et al. Dashboard Design Patterns. IEEE Transactions on Visualization and Computer Graphics. 2023;29(1):342-52.

[5] Jain P, Coogan SC, Subramanian SG, Crowley M, Taylor S, and Flannigan MD. A review of machine learning applications in wildfire science and management. Environmental Reviews. 2020;28(4):478-505.

[6] Noryanti M, Amirah HR, Hanita D, Evizal AK, and Warih M. Predicting Forest Fire Spots Using Nonparametric Predictive Inference with Parametric Copula: Malaysia Case Study (to be published). AIP Conference Proceedings. [Conference Proceedings]. In press 2024.

[7] Amirah HR, Noryanti M, Evizal AK, Warih M, and Hanita D. Forecasting Locations of Forest Fires in Indonesia Through Nonparametric Predictive Inference with Parametric Copula: A Case Study (to be published). Seminar Kebangsaan Institut Statistik Malaysia Kali Ke-18. [Conference Proceedings]. In press 2024.

[8] Muhammad N. Predictive inference with copulas for bivariate data: Durham University; 2016.

[9] Chang W, Cheng J, Allaire J, Sievert C, Schloerke B, Xie Y, et al. Package ‘shiny’: Web Application Framework for R [Package]. 2024 [updated April 2, 2024. Available from: https://shiny.posit.co/, https://github.com/rstudio/shiny.

[10] Chang W. Themes for Shiny: Package ‘shinythemes’ 2021 [updated October 14, 2022. Available from: https://rstudio.github.io/shinythemes/.

[11] Barret T, Dowle M, Srinivasan A, Gorecki J, Chirico M, and Hocking T. Extension of `data.frame`: Package ‘data.table’ 2024 [updated March 31, 2024. Available from: https://r-datatable.com, https://Rdatatable.gitlab.io/data.table.

[12] Cheng J, Schloerke B, Karambelkar B, and Xie Y. Create Interactive Web Maps with the JavaScript 'Leaflet' Library: Package ‘leaflet’ [Package]. 2024 [updated March 26, 2024. Available from: https://rstudio.github.io/leaflet/.

[13] Bache SM, and Wickham H. A Forward-Pipe Operator for R: Package ‘magrittr’ [Package]. 2022 [updated October 13, 2022. Available from: https://magrittr.tidyverse.org.

[14] Tan L, Guo J, Mohanarajah S, and Zhou K. Can we detect trends in natural disaster management with artificial intelligence? A review of modeling practices. Natural Hazards. 2021;107:2389-417.