ARTIFICIAL INTELLIGENCE AND COPULA-PROBABILISTIC MODELS FOR EARLY FLOOD WARNING AND COMMUNITY RISK MANAGEMENT: CASE STUDIES IN FIJI ISLANDS

A Thesis submitted by

Ravinesh Chand

BSc GCED (Mathematics and Physics), PGCTT

For the award of

Master of Research

2024

ABSTRACT

Flooding is one of the most prevalent natural hazards, impacting numerous regions worldwide. The repercussions of such disasters are especially severe in developing nations, particularly in small island countries like Fiji. The absence of advanced flood risk monitoring resources and relevant data in these developing nations presents significant challenges to implementing effective early flood warning systems. To address this issue, this research develops innovative flood monitoring, assessment, and forecasting tools using artificial intelligence (AI) and copula-statistical methods to enhance and contribute to developing effective early flood warning systems, thereby assisting in better flood preparation and management strategies to mitigate the severe impacts of flooding. The first objective is to develop a novel hourly flood monitoring index ($SWRI_{24\text{-}hr\text{-}S}$) to identify flood events and compute their associated characteristics, including flood volume (V), duration (D) and peak (Q). The feasibility of this index as an hourly flood risk monitoring tool is demonstrated for various flood-prone sites in Fiji. The 3-dimensional (3D) vine copula model is employed to model the joint distribution between D, V, and Q to extract their joint exceedance probability for probabilistic flood risk assessment across these study sites. In the second objective, a hybrid deep learning model (C-GRU) is designed by fusing the Convolutional Neural Network (CNN) with the Gated Recurrent Unit (GRU) model to forecast the proposed $SWRI_{24\text{-}hr\text{-}S}$ over a short-term (i.e., 1-hourly forecast horizon) to assess the future flood risk for five flood-prone study sites in Fiji. The objective model is trained using the statistically significant lagged $SWRI_{24\text{-}hr\text{-}S}$ and real-time hourly rainfall data. The state-of-the-art models, i.e., CNN, GRU, Long Short-Term Memory (LSTM), and Random Forest Regression (RFR), were also developed for benchmarking. The hyperparameters of the objective and benchmarking models were optimised using the efficient Bayesian Optimization (BO) technique. Overall, the outcomes of this research are expected to assist Fiji and other flood-prone regions worldwide in enhancing their existing early flood warning systems by integrating the flood monitoring, assessment, and forecasting tools developed in this study. This integration will enhance the decision-support framework for flood preparedness and response efforts, thus mitigating the severe impacts of flooding and improving community risk management.

CERTIFICATION OF THESIS

I Ravinesh Chand declare that the Master of Research Thesis entitled *Artificial Intelligence and Copula-Probabilistic Models for Early Flood Warning and Community Risk Management: Case Studies in Fiji Islands* is not more than 100,000 words in length including quotes and exclusive of tables, figures, appendices, bibliography, references, and footnotes.

This Thesis is the work of Ravinesh Chand except where otherwise acknowledged, with the majority of the contribution to the papers presented as a Thesis by Publication undertaken by the student. The work is original and has not previously been submitted for any other award, except where acknowledged.

Date: 19/06/2024

Endorsed by:

Professor Ravinesh C. Deo
Principal Supervisor

Dr. Thong Nguyen-Huy
Associate Supervisor

Dr. Mumtaz Ali
Associate Supervisor

Dr. Sujan Ghimire
External Supervisor

Student and supervisors' signatures of endorsement are held at the University.

STATEMENT OF CONTRIBUTION

This Master of Research thesis by publication has produced two high quality (Quartile 1) journal papers during the candidature. The details of joint authorship and agreed share of these contributions are as follows:

Paper 1 (Chapter 4):

Chand, R., Nguyen-Huy, T., Deo, R. C., Ghimire, S., Ali, M., & Ghahramani, A. (2024). Copula-Probabilistic Flood Risk Analysis with an Hourly Flood Monitoring Index, *Water*, 16, 1560. https://doi.org/10.3390/w16111560.

Ravinesh Chand contributed 70% to this paper. Collectively Thong Nguyen-Huy, Ravinesh C. Deo, Sujan Ghimire, Mumtaz Ali, and Afshin Ghahramani contributed the remainder.

Paper 2 (Chapter 5):

Chand, R., Deo, R. C., Ghimire, S., Nguyen-Huy, T., & Ali, M. (2024). Hybrid Convolutional Neural Network Fused with Gated Recurrent Unit-based Hourly Flood Index Forecasting Framework: Case Studies for Fiji Islands. *Stochastic Environmental Research and Risk Assessment* (Under Review). Submission ID 9b9ebaac-d593-41a9-a7ed-cb326f0bc35e.

Ravinesh Chand contributed 70% to this paper. Collectively Ravinesh C. Deo, Sujan Ghimire, Thong Nguyen-Huy, and Mumtaz Ali contributed the remainder.

ACKNOWLEDGEMENTS

First and foremost, I express my heartfelt gratitude to the Almighty God for His continuous blessings and protection throughout my journey. Additionally, I am deeply indebted to the invaluable support I received from the following individuals and organisations, without whom this thesis would not have been completed.

I extend my deepest and most heartfelt thanks to my principal supervisor, Professor Ravinesh C. Deo, for his unwavering support, guidance, and feedback during my candidature, which proved instrumental in successfully completing my research. I am also grateful for his assistance and mentorship during my scholarship application process. I further thank him for sharing his expertise using Overleaf to write papers. Furthermore, I sincerely appreciate all my associate supervisors, Dr Thong Nguyen-Huy, Dr Mumtaz Ali, and Dr Sujan Ghimire, for their invaluable support, guidance, and insightful feedback throughout my research. I am incredibly grateful to Dr Thong Nguyen-Huy for his expertise and tremendous support, particularly in completing objective one of this research. Additionally, I acknowledge Dr Afshin Ghahramani for his expertise, guidance, and feedback while writing the research proposal and Paper 1.

Moreover, I am heartily thankful to the Australian Government's Department of Foreign Affairs and Trade (DFAT) for funding my studies through the Australia Awards scholarship scheme for Fiji, which enabled me to pursue this study. I would like to express my profound gratitude to the Fiji Meteorological Services for providing the data needed for my research. I am also grateful to the University of Southern Queensland (UniSQ), particularly the Graduate Research School (GRS), for providing additional support throughout my studies. I also take this opportunity to thank my research colleagues from UniSQ's Artificial Intelligence Applications Laboratory (managed by Professor Ravinesh C. Deo) for their moral support and motivation. I am especially grateful to Dr Salvin S. Prasad and Dr Lionel P. Joseph for their guidance and support during the development stage of the forecasting models (Objective 2).

Finally, I express my deepest gratitude to my parents, spouse, and siblings for their support and encouragement during my research. Their love, support, and belief in me kept me going during the challenging times of this journey. I also acknowledge all my friends, relatives, and peers for their motivation throughout my candidacy.

DEDICATION

This Master of Research thesis is lovingly dedicated to my beloved parents, Mr Lalesh Chand and Mrs Anjila Wati, and my dear wife, Mrs Ashna Gounder.

TABLE OF CONTENTS

LIST OF FIGURES

ABBREVIATIONS

API	Antecedent Precipitation Index
AdaGrad	Adaptive Gradient Algorithm
Adam	Adaptive Moment Estimator
AIC	Akaike Information Criterion
AI	Artificial Intelligence
AWRI	Available Water Resources Index
BIC	Bayesian Information Criterion
BO	Bayesian Optimization
BiLSTM	Bidirectional LSTM
CRED	Centre for Research on the Epidemiology of Disasters
R^2	Coefficient of Determination
CNN	Convolutional Neural Network
CCF	Cross-Correlation Function
DL	Deep Learning
DM	Diebold-Mariano
ECDF	Empirical Cumulative Distribution Function
xAI	Explainable Artificial Intelligence
FMS	Fiji Meteorological Services
NDMO	Fiji's National Disaster Management Office
FJD	Fijian Dollar
NDMO	Fiji's National Disaster Management Office
FFGS	Flash Flood Guidance System
D	Flood Duration
Q	Flood Peak
V	Flood Volume
FE	Forecast Error
FC	Fully Connected
GRU	Gated Recurrent Unit
GPI	Global Performance Indicator
GPU	Graphics Processing Unit
GDP	Gross Domestic Product
IQR	Interquartile Range

IKNN	Iterative K-nearest Neighbors
JCDF	Joint Cumulative Distribution Function
JPDF	Joint Density Distribution Function
KGE	Kling-Gupta Efficiency
E_{LM}	Legate-McCabe Efficiency Index
LIME	Local Interpretable Model-Agnostic Explanations
logLik	Log-Likelihood
LSTM	Long Short-Term Memory
ML	Machine Learning
MAE	Mean Absolute Error
mBICv	Modified Vine Copula Bayesian Information Criteria
MEMD	Multivariate Empirical Mode Decomposition
MI	Mutual Information
E_{NS}	Nash-Sutcliffe Efficiency Index
PSIDS	Pacific Small Island Developing State
PACF	Partial Autocorrelation Function
r	Pearson's Correlation Coefficient
PDF	Probability Density Function
RFR	Random Forest Regression
RNN	Recurrent Neural Network
RMSE	Root Mean Square Error
RMSProp	Root Mean Square Propagation
SPPs	Satellite-based precipitation products
SHAP	SHapley Additive exPlanations
SPCZ	South Pacific Convergence Zone
SAPI	Standardised Antecedent Precipitation Index
SPI	Standardised Precipitation Index
SWAP	Standardised WAP
SVR	Support Vector Regression
sMAPE	Symmetric Mean Absolute Percentage Error
TPU	Tensor Processing Unit
TPE	Tree-structured Parzen Estimator
USD	United States Dollar

WAP	Weighted Average of Precipitation
W	Weighting Factor
E_{WI}	Willmott's Index of Agreement

CHAPTER 1: INTRODUCTION

1.1. Background

Floods are among the most catastrophic natural disasters, capable of posing significant threats to populations through fatalities, injuries, displacement, and extensive damage to public and private infrastructure, cultural sites, the environment, and economic activities (Maranzoni et al., 2023). It results from various factors, including hurricanes, storm surges, excessive rainfall, and rapid snowmelt (Kumar et al., 2023). According to the Centre for Research on the Epidemiology of Disasters (CRED), floods constituted 44% of all disaster events between 2000 and 2019 (the highest figure compared to the other disaster types), impacting 1.65 billion people and resulting in 104,614 fatalities and economic losses totalling 651 billion USD worldwide (CRED, 2020; Maranzoni et al., 2023). The effects of climate change are expected to rapidly increase the intensity and frequency of flood events, leading to more severe damages in the near future (Nguyen-Huy et al., 2021). Therefore, it has become increasingly crucial to develop reliable and cost-effective tools for accurately monitoring, forecasting, and assessing the risk of extreme flood events to mitigate their severe impacts. In recent years, advanced mathematical and statistical methods and artificial intelligence (AI) algorithms have been used to develop practical tools for monitoring, assessing, and forecasting flood risks.

In developing nations with limited flood monitoring resources and data, a mathematically derived flood index based solely on rainfall offers a crucial approach to assessing potential flood risks. The 24-hourly water resources index ($WRI_{24\text{-}hr\text{-}S}$), proposed by Deo et al. (2018) in their pilot study, is one of the mathematical indices developed using only hourly rainfall data. This index was applied to monitor flood risk during sustained extreme rainfall periods in two study locations: Australia and South Korea. However, unlike the daily flood indices, such as the Flood Index (I_F) (Deo et al., 2015) and the Standardised Antecedent Precipitation Index (*SAPI*) (Nguyen-Huy et al., 2022), the primary limitation of the $WRI_{24\text{-}hr\text{-}S}$ is that it is unnormalised. Hence, it does not allow for the objective assessment of flood risk across geographically diverse study sites. Once normalised, the (normalised) index will serve as a uniform metric, ensuring consistent comparison of flood magnitudes and facilitating effective assessment of flood risks across geographically diverse regions. Consequently, it will

enable the identification of flood events and the derivation of their associated characteristics, such as flood volume, peak, and duration.

Floods are multivariate stochastic phenomena characterised by mutually correlated variables, including flood peak, volume, and duration (Aminuddin Jafry et al., 2024; Daneshkhah et al., 2016; Ganguli & Reddy, 2013; Klaho et al., 2022; Latif & Mustafa, 2021, 2020; Nguyen-Huy et al., 2022; Tosunoglu et al., 2020). Potential risk and damage from floods are likely dependent on these interrelated flood characteristics, which should, therefore, be prioritised in risk assessment (Latif & Mustafa, 2020; Nguyen-Huy et al., 2022). Hence, it is essential to analyse the joint distribution properties of these characteristics in multivariate analysis to estimate the actual probabilities of flood occurrence (Chebana & Ouarda, 2009; Nguyen-Huy et al., 2022). Previous studies have used vine copulas to model the joint distribution of flood event characteristics to derive important information for flood risk management, including the exceedance probability and joint and conditional return periods (Daneshkhah et al., 2016; Gräler et al., 2013; Latif & Mustafa, 2020; Nguyen-Huy et al., 2022; Shafaei et al., 2017; Tosunoglu et al., 2020). For instance, Nguyen-Huy et al. (2022), in their study, demonstrated the feasibility of the 3-dimensional (3D) vine copula in modelling the joint distribution of flood characteristics (i.e., flood duration, volume and peak) derived from the *SAPI* to extract their joint exceedance probability for probabilistic flood risk assessment. Such studies are critical for enhancing the evaluation of exceedance probability used in early flood warning systems and for accurately assessing flood risks, thereby facilitating the implementation of targeted risk management strategies. This is particularly crucial in developing nations with limited data and resources for flood monitoring and assessment.

Moreover, it is imperative to be able to forecast future flood situations in addition to analysing past flood events. Flood forecasting is a crucial element of an early warning system, demanding relevant data, advanced technologies, and specialised expertise for its development. This presents a challenge in developing countries where these resources are often lacking. Hence, these countries must prioritise implementing a cost-effective solution that leverages readily available technology for the early flood warning system.

Furthermore, it is essential to note that mathematically derived flood indices, such as I_F, cannot predict the flooded state in advance unless a forecasting model for these indices is developed (Moishin et al., 2021a). The index-based flood forecasting

approach has recently gained significant popularity despite being previously underexplored. Artificial Intelligence (AI)-based models, including Machine Learning (ML), Deep Learning (DL), and hybrid models, have been developed and have demonstrated promising forecasting performance for the daily flood index (I_F) to assess future flood risk for Australia (Prasad et al., 2021), Bangladesh (Ahmed et al., 2023), and Fiji (Moishin et al., 2021a). However, implementing an early flood warning system that utilises more efficient and cost-effective AI techniques, along with copula-statistical methods based on mathematically derived flood monitoring tools at shorter timescales (e.g., hourly), is crucial. This approach will enable relevant authorities and communities to take timely actions to mitigate potential risks associated with flooding, potentially saving lives and minimising property damage and economic losses.

1.2. Statement of the problem

Flooding is one of the most common natural disasters affecting numerous regions around the globe (CRED, 2020). Fiji, a Pacific Small Island Developing State (PSIDS) on which this study is focused, is also among the regions highly exposed to devastating natural hazards, where flooding is an annual event that causes loss of life, significant damage to housing, public infrastructure, agriculture, and economic disruption and losses (Government of Fiji, 2017; Lucas, 2020; McNamara, 2013). Apart from coastal flooding in low-lying areas, Fiji is significantly impacted by fluvial floods, which occur when rivers overflow due to prolonged or heavy rainfall, and pluvial floods (or flash floods) caused by intense precipitation overwhelming drainage systems, especially in flat and urban areas (Government of Fiji, 2017).

Fiji has experienced some of the most severe floods in its history (McGree et al., 2010; Yeo & Blong, 2010). Between 1970 and 2016, Fiji experienced 44 major flood events that affected approximately 563,310 people and resulted in 103 fatalities (Government of Fiji, 2017). The March 2012 flash flooding was one of the most catastrophic events that hit Fiji, particularly the western division, where the country's core industries, such as sugar, gold mining, and tourism, are located (McNamara, 2013). A survey conducted by McNamara (2013) that assessed the aftermath of the March 2012 flooding in Fiji, reported four fatalities, the temporary displacement of 15,000 individuals in evacuation centres, and initial damages exceeding 71 million FJD to key economic sectors.

The estimated average annual flood losses in Fiji exceed 400 million FJD, equivalent to 4.2% of the country's Gross Domestic Product (GDP) (Government of Fiji, 2017). Such losses are significant for a small island nation with a population of less than a million and a GDP of less than 5 billion USD (The World Bank Group, 2022). Given the assumption of significant increases in rainfall due to climate change, annual losses of flood-related assets in Fiji could surpass 5% of the GDP by 2050 without adaptation measures (Government of Fiji, 2017). This poses a considerable challenge to the country's long-term sustainable development and an economic threat. Consequently, effective early warning systems are much needed in Fiji to mitigate the severe impacts of flooding, a recommendation also highlighted by McNamara (2013) following the devastating effects of the March 2012 flooding in Fiji. Over the years, Fiji has significantly invested in flood monitoring and forecasting systems and recently adopted the Flash Flood Guidance System to enhance its flood preparedness and response efforts.

Overall, the primary goal of this Master of Research thesis is to develop innovative flood monitoring, assessment, and forecasting tools using both AI and copula-statistical methods to enhance the existing and future early flood warning systems in Fiji. This research initiative aims to assist relevant stakeholders and the public in being better prepared, thereby potentially mitigating the severe impacts of such natural hazards.

1.3. Research objectives

The primary aim of this Master of Research thesis is to (a) develop a novel hourly flood monitoring with a probabilistic flood risk assessment and (b) develop an AI-based flood forecasting tool and test its practical utility in various flood-prone areas of Fiji. It is anticipated that these tools will assist relevant organisations in understanding the attributes of past flood events and forecasting future flood occurrences, thus enhancing the country's early flood warning systems, potentially leading to improved flood preparedness, mitigation, and response strategies.

The two specific objectives have been designed to meet the primary aim of this thesis, which are:

1. **To develop a novel hourly flood index** ($SWRI_{24\text{-}hr\text{-}S}$) (reported in the journal paper 1, Chapter 4). This index is then used to identify flood events and compute their characteristics, i.e., flood duration (D), volume (V), and

peak (Q) across various study sites in Fiji over the study period. The 3D vine copula is employed to model the joint distribution of flood characteristics to extract their joint exceedance probability for probabilistic flood risk assessment. The $SWRI_{24\text{-}hr\text{-}S}$ is formulated by normalising the existing 24-hourly water resources index ($WRI_{24\text{-}hr\text{-}S}$) in the literature, enabling objective flood risk assessment across geographically diverse regions. A comprehensive analysis is conducted to test and validate its practical utility in assessing flood risk on an hourly scale. The joint exceedance probability between the flood characteristics is derived by first quantifying the probability that the flood characteristics, i.e., D, V, and Q, exceed specific thresholds simultaneously. The thresholds were selected at the 50th-quantile (median), 75th-quantile (moderate), and 95th-quantile (extreme).

2. **To develop a novel hybrid deep learning (DL) algorithm, C-GRU, for flood risk forecasting** (reported in the journal paper 2, Chapter 5). The hybrid C-GRU model integrates Convolutional Neural Networks (CNN) with the Gated Recurrent Unit (GRU) model to forecast the $SWRI_{24\text{-}hr\text{-}S}$ over a short-term, i.e., 1-hourly forecast horizon to assess future flood risk for five flood-prone study sites in Fiji. The hybrid C-GRU model uses statistically significant lagged values of $SWRI_{24\text{-}hr\text{-}S}$ and real-time hourly rainfall as inputs. The hyperparameters of the proposed hybrid model (and benchmark models) are optimised using the Bayesian Optimization (BO) technique. The forecast performance of the hybrid C-GRU model is evaluated against the state-of-the-art benchmark models: CNN, GRU, Long Short-Term Memory (LSTM) and Random Forest Regression (RFR).

The journal paper resulting from completing objective one of this research has been published in *Water*, 2024, Vol. 11, page 1560, https://doi.org/10.3390/w16111560. This journal is ranked *Q*1 in Water Science and Technology. The journal paper resulting from completing objective two of this research has been submitted for publication in *Stochastic Environmental Research and Risk Assessment* and is under review. This journal is ranked as *Q*1 in Water Science and Technology.

1.4. Thesis Layout

The schematic representation of the objectives formulated in this thesis is illustrated in Figure 1. It clearly outlines the processes from data acquisition to the application of innovative AI and copula-statistical methods for early flood warning and community risk management. This Master of Research thesis is organised into six chapters, given as follows:

Chapter 1 This chapter provides the background of the research, presents the problem statement to highlight the study's importance and outlines the study's objectives.

Chapter 2 This chapter presents a comprehensive literature review to summarise the existing research and identify gaps in the domain knowledge.

Chapter 3 This chapter describes the study sites, dataset, and general methodology adopted in this research. While it provides a general overview, the specific study sites, datasets, and methodologies for each objective are detailed in their respective chapters.

Chapter 4 This chapter is presented as a published article in the *Water* Journal (https://doi.org/10.3390/w16111560). It addresses the first objective of the research, introducing a novel hourly flood index ($SWRI_{24\text{-}hr\text{-}S}$). This index is applied to seven study sites in Fiji to identify flood events and compute their characteristics over the study period. Subsequently, the 3D vine copula model is employed to model the joint distribution of flood characteristics to extract their joint exceedance probability for probabilistic flood risk assessment across these study sites.

Chapter 5 This chapter is presented as a journal article submitted for publication in the *Stochastic Environmental Research and Risk Assessment* Journal and is under review. It addresses the second objective of the research. This chapter introduces a hybrid C-GRU model designed to forecast the $SWRI_{24\text{-}hr\text{-}S}$ over a short-term, i.e., 1-hourly forecast horizon, to assess future flood risk for five flood-prone study sites in Fiji. The Bayesian Optimization (BO) with the Tree-structured Parzen Estimator (TPE) algorithm is employed to optimise the hyperparameters of the proposed hybrid C-GRU and benchmark models. The performance of the proposed hybrid C-GRU model is compared to other benchmark models using

various performance evaluation metrics, visual diagnostic plots, and a statistical test.

Chapter 6 This chapter offers a comprehensive summary of the study, discusses its limitations, and provides recommendations for future research.

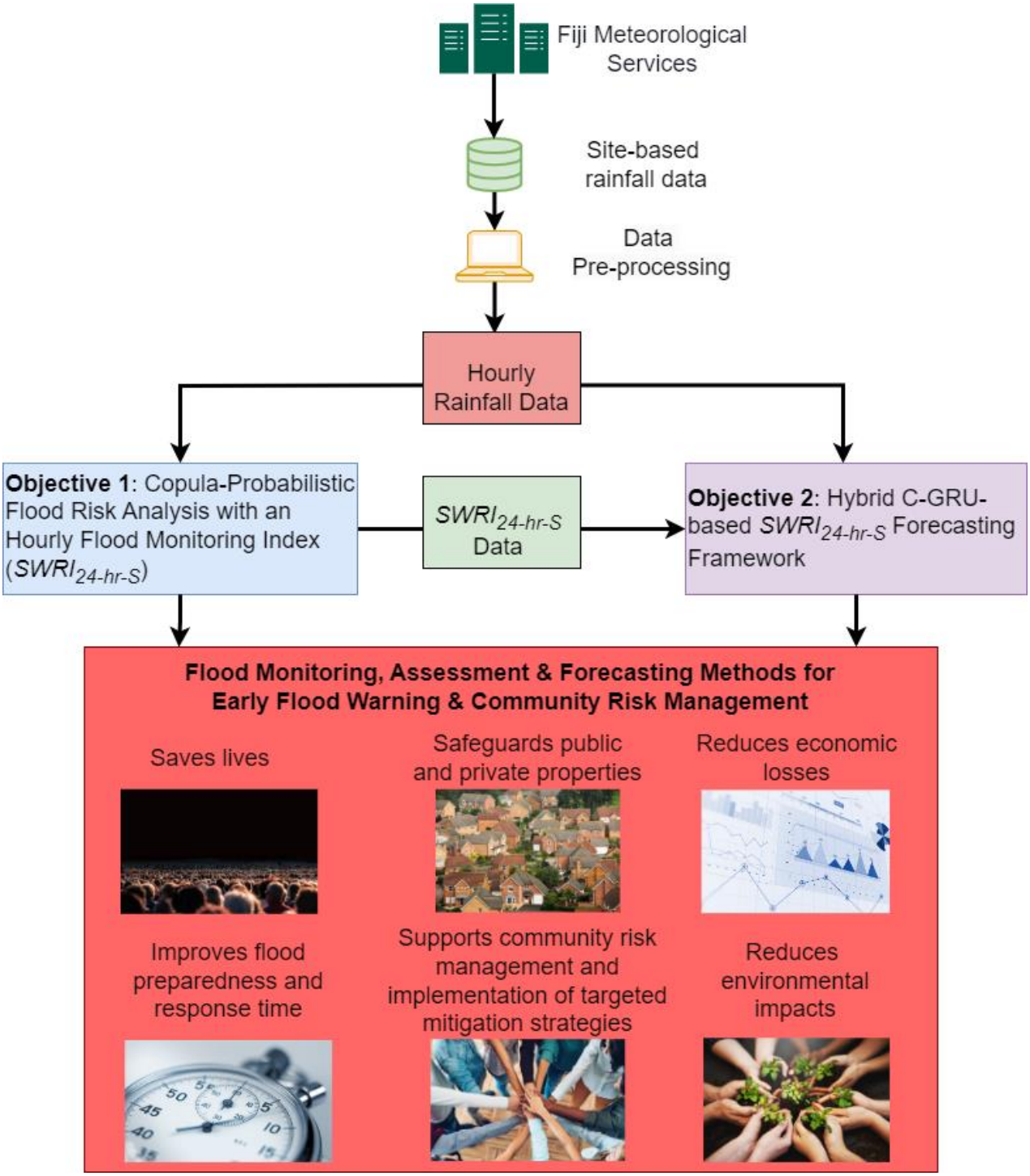

Figure 1 Schematic view of this Master of Research thesis.

CHAPTER 2: LITERATURE REVIEW

This chapter presents the literature and relevant backgrounds of the study. It further outlines the knowledge gaps in the existing literature, consequently highlighting the innovative tools that can be adopted for designing an early flood warning system to mitigate the severe impacts of flood risk.

2.1. Flood monitoring indices

Floods are complex events influenced by several factors. These factors include the cumulative impact of various weather or climate events, such as temperature and precipitation, which may not be extreme individually, but their cumulative effect can cause severe consequences, along with the vulnerability and exposure of the affected area (Leonard et al., 2014; Seneviratne et al., 2012). Other factors, including topography, soil saturation, catchment size and shape, drainage networks, dam or levee construction, sea-level rise, and climate change, also contribute to flood occurrence. Therefore, an index that combines all these factors and the cumulative impacts of other weather or climate characteristics is crucial for a comprehensive flood risk assessment. However, in many developing countries, the lack of access to relevant data and flood monitoring and risk assessment facilities poses a significant challenge. Therefore, employing a mathematically derived flood index solely based on rainfall data offers an alternative strategy for evaluating an imminent flood risk situation.

Over the years, numerous mathematical tools have been devised for flood monitoring solely based on rainfall data. These include the Standardised Precipitation Index (*SPI*) (Seiler et al., 2002), the Available Water Resources Index (*AWRI*) (Byun & Lee, 2002), the Weighted Average of Precipitation (*WAP*) (Lu, 2009), the Standardised WAP (*SWAP*) (Lu et al., 2014), the Flood Index (I_F) (Deo et al., 2015), and the Standardised Antecedent Precipitation Index (*SAPI*) (Nguyen-Huy et al., 2022). However, when conducting flood risk assessment with indices (or indicators), emphasis should be placed on monitoring the changes in the remaining water volume due to rainfall over time and the capability of flood indices to reflect changing hydrological conditions accurately (Nguyen-Huy et al., 2022). Given that the start and end dates of flood events and monitoring can be unpredictable, ranging from spontaneous to short-term or long-term occurrences, flood monitoring based on flood indices should consider antecedent rainfall within each respective period (Nguyen-Huy

et al., 2022). Consequently, flood indices such as *SWAP*, I_F, and *SAPI* are robust since they are designed to accommodate changes in antecedent rainfall by utilising appropriate time-dependent reduction functions that consider the depletion of water resources through various hydrological processes. The flood index, $I_{F,}$ has been applied in several locations and has demonstrated superior performance in daily flood monitoring, including Australia (Deo et al., 2018), Iran (Nosrati et al., 2010), Bangladesh (Deo et al., 2019), and Fiji (Moishin et al., 2021b). However, the primary limitation of I_F and other indices, such as *SPI*, is their dependence on daily or monthly/annual accumulated rainfall data, which operate on significantly longer timescales than necessary for a real-time flood monitoring system. Consequently, these indices may not effectively capture flood risk during sudden bursts of high-intensity rainfall, leading to rapid responses (i.e., flash floods).

To enable real-time flood monitoring, Deo et al. (2018) proposed a 24-hourly water resources index ($WRI_{24\text{-}hr\text{-}S}$) in their pilot study to monitor flood risk during sustained extreme rainfall periods and demonstrated its application in two study locations: Brisbane, Australia, and Dobong Observatory, South Korea. The $WRI_{24\text{-}hr\text{-}S}$ in that study was developed using hourly rainfall data to monitor flood risk by considering rainfall accumulation over the preceding 24 hours, whereby the rainfall contribution from preceding hours is subjected to a time-dependent reduction function, which accounts for the depletion of water resources through various hydrological processes, including evaporation, percolation, seepage, runoff, and drainage (Deo et al., 2018). However, unlike daily flood indices, I_F, and *SAPI*, which are normalised metrics, the main limitation of the $WRI_{24\text{-}hr\text{-}S}$ is that it is unnormalised. Hence, it cannot be used to compare and objectively assess flood risk across geographically diverse study sites. This highlights a significant research gap addressed in the current study.

2.2. Copula-based flood characteristics modelling

Flooding is a multivariate probabilistic event that exhibits a complex interdependence among its three intercorrelated characteristics: flood peak, volume, and duration (Latif & Mustafa, 2020). Hence, the univariate flood frequency analysis approach for probabilistic flood risk assessment based on a single characteristic has been recognised as unreliable because such approaches give limited information and may underestimate or overestimate flood-related risk (Latif & Mustafa, 2020; Nguyen-Huy et al., 2022). For instance, a flood risk assessment based solely on duration will

fail to account for the risks associated with high peak and volume (Nguyen-Huy et al., 2022). Therefore, since flood event characteristics are generally interrelated, multivariate models that accurately represent the dependence structures of these characteristics are crucial for a comprehensive flood risk assessment (Nguyen-Huy et al., 2022; Tosunoglu et al., 2020).

Copula models, first introduced by Sklar (1959), are considered the most efficient statistical tool for modelling multivariate distributions compared to conventional multivariate models (Latif & Mustafa, 2020). Copulas are functions that combine several univariate marginal cumulative distribution functions into a joint cumulative distribution (Daneshkhah et al., 2016; Gräler et al., 2013). By doing so, copulas describe the dependence structure between random variables and enable the calculation of joint probabilities independent of the marginal behaviour of the involved variables (Ganguli & Reddy, 2013; Gräler et al., 2013; Latif & Mustafa, 2020; Nguyen-Huy et al., 2022; Tosunoglu et al., 2020). Among multivariate copula models, the vine copula (or pair-copula) provides the most flexible approach for modelling the dependence structure between multiple variables based on the mixing of (conditional) bivariate copulas (Aas et al., 2009; Daneshkhah et al., 2016; Gräler et al., 2013; Nguyen-Huy et al., 2022; Tosunoglu et al., 2020).

The underlying theory for vine copula construction is detailed in Bedford and Cooke (2001, 2002), originating from the work presented by Joe (1997). The fundamental concept of vine copula-based methodology is to decompose the multivariate joint density function into a sequence of local building blocks comprised of best-fitted bivariate copula functions and their associated conditional and unconditional distribution functions rather than relying on a single fixed copula across all intercorrelated random variables (Aas et al., 2009; Daneshkhah et al., 2016; Latif & Mustafa, 2020; Latif & Simonovic, 2022a). Several studies have employed the vine copula approach to model the joint distribution of flood characteristics, such as flood volume, peak, and duration, to evaluate exceedance probability and joint and conditional return periods (Daneshkhah et al., 2016; Gräler et al., 2013; Latif & Mustafa, 2020; Nguyen-Huy et al., 2022; Shafaei et al., 2017; Tosunoglu et al., 2020).

The present study draws inspiration from the research conducted by Nguyen-Huy et al. (2022). In that study, the mathematically derived flood index (*SAPI*), which incorporates a time-dependent reduction function, was first calculated using only satellite-derived daily rainfall data. This index was then employed to identify all flood

events and compute their associated characteristics (i.e., flood duration, volume, and peak) in Myanmar. The 3D vine copula was used to model the joint distribution of these flood event characteristics to extract the joint exceedance probabilities in different combined scenarios (where flood characteristics simultaneously exceed various thresholds) for probabilistic flood risk assessment. The methodologies presented in that study are crucial for enhancing the evaluation of exceedance probability utilised in early flood warning systems, ensuring accurate flood risk assessment, and implementing targeted risk management strategies. However, such approaches remain unexplored for many flood-prone regions around the globe.

2.3. Data-driven flood index forecasting

Flood forecasting is crucial for an early flood warning system, yet it remains one of the most complex and challenging tasks in hydrology, primarily due to the complex dynamic processes involved (Nguyen-Huy et al., 2021; Prasad et al., 2021). Flood forecasting is more practical when conducted over a short-term (e.g., hourly), as it allows for better flood risk estimation, preparation time and implementation of appropriate mitigation strategies (Alexander et al., 2018; Hapuarachchi et al., 2011; Kant et al., 2013; Tiwari & Chatterjee, 2010). Hydrodynamic models are commonly employed for simulating detailed flood dynamics, as they can be directly integrated with hydrological and river models to enhance flood risk assessment, real-time flood forecasting, and scenario analysis (Teng et al., 2017). However, these models have high data requirements, demand intensive computational resources, and require specialised expertise, making them challenging to apply in operational flood forecasting (Kabir et al., 2020; Nevo et al., 2022; Teng et al., 2017).

An index-based flood forecasting system using AI techniques to assess future flood risk has recently been proposed as a cost-effective tool, particularly useful for developing countries lacking hydro-meteorological datasets and with underdeveloped flood risk monitoring and forecasting facilities and resources. As previously mentioned, the feasibility of the daily flood index, I_F in daily flood monitoring has already been demonstrated in Australia (Deo et al., 2018), Iran (Nosrati et al., 2010), Bangladesh (Deo et al., 2019), and Fiji (Moishin et al., 2021b). However, as a flood monitoring tool, I_F cannot predict the flooded state in advance until a forecasting model is developed and tested (Moishin et al., 2021a). Consequently, a few recent studies have demonstrated the feasibility of the hybrid ML/DL-based I_F forecasting system in

assessing future flood risk. For instance, in a study by Prasad et al. (2021), a hybrid ML by combining the multivariate empirical mode decomposition (MEMD) technique with the M5 tree (MEMD-M5 tree) model to forecast I_F for the flood-prone Lockyer Valley region of Queensland, Australia was demonstrated. The data utilised in that study included the Effective Drought Index, Daily Precipitation, Available Water Resources Index, and Prediction Return to Normal. The performance of the MEMD-M5 tree model was compared with the hybrid MEMD-RFR and the standalone M5 tree and RFR models. The results of that study demonstrated that the MEMD-M5 tree model outperformed the benchmark models in accurately emulating future I_F values.

Moreover, the present study draws inspiration from the study by Moishin et al. (2021a) that developed a hybrid DL algorithm, ConvLSTM, by integrating two DL algorithms, i.e., Convolutional Neural Network (CNN) with Long Short-Term Memory (LSTM) models to forecast I_F across multiple forecast horizons for nine flood-prone sites in Fiji. The performance of the hybrid DL model was evaluated against benchmark models: CNN-LSTM, LSTM, and Support Vector Regression (SVR). All models developed in that study were trained using statistically significant lagged I_F and real-time daily rainfall data. The results of that study demonstrated the feasibility of the ConvLSTM model in forecasting I_F to assess the possibility of flood situations at 1, 3, 7, and 14 days ahead forecast horizon in Fiji's case studies. In a recent study, Ahmed et al. (2023) also proposed a hybrid DL (i.e., CNN-BiLSTM) model that combined CNN with bidirectional long short-term memory (BiLSTM) to forecast I_F a week ahead for thirty-four selected stations in Bangladesh. The data utilised in that study included fifteen synoptic-scale climatic indices. The results of that study demonstrated the superior forecasting performance of the hybrid CNN-BiLSTM model compared to the benchmark models, SVR and BiLSTM. These studies demonstrate that hybrid ML/DL can be effectively utilised to accurately model and forecast I_F in assessing future flood risk on a daily scale. However, designing a robust tool for assessing future flood risks on a shorter timescale (e.g. hourly) is essential for real-time flood monitoring and forecasting.

Moreover, while the choice between ML and DL models highly depends on the problem, DL models offer distinct advantages when working with large datasets. Their capacity to process a vast number of features enables the design of highly effective data-driven models, making DL models a superior choice over traditional ML models (Sarker, 2021). Another distinct advantage of DL models over ML models is their ability

to extract high-level features directly from the data, significantly reducing the time and effort needed to construct a feature extractor for each problem (Sarker, 2021). One of the DL models specialised for modelling sequential or time-series data is Recurrent Neural Networks (RNNs), which utilise a hidden state to retain information from previous inputs. However, due to problems with vanishing and exploding gradients, standard RNNs encounter short-term memory issues that hinder their ability to effectively learn long-term dependencies in the data, limiting their accuracy in long-term forecasting (Kim et al., 2018; Sarker, 2021).

The LSTM, initially introduced by Hochreiter and Schmidhuber (1997) and later enhanced by Graves (2013), and the Gated Recurrent Unit (GRU), introduced by Cho et al. (2014), are two variants of RNN capable of effectively capturing long-term dependencies in sequential data. While the LSTM and GRU share similar architectures, the GRU has fewer trainable parameters than the LSTM, as it employs a simplified gating mechanism with only two gates (reset and update gates), whereas the LSTM incorporates three gates (input, output, and forget gates). The GRU's structure allows it to adaptively capture dependencies from large data sequences without discarding information from earlier parts of the sequence (Sarker, 2021). Therefore, compared to LSTM, the GRU achieves comparable performance while significantly reducing computation time, owing to its streamlined structure and fewer parameters, resulting in faster training times (Kisvari et al., 2021; Li, 2023; Sarker, 2021; Sharma et al., 2022; Wang et al., 2020; Zhang et al., 2022). In addition to temporal dependence models, one-dimensional CNN (Conv1D) can autonomously extract pertinent features from the input sequential data through their convolutional kernels with minimal human intervention (Ghimire et al., 2022). Therefore, integrating CNN with a temporal model such as GRU combines CNN's feature extraction capabilities with GRU's efficiency in temporal modelling, resulting in a robust approach to modelling sequential data.

The integration of CNN with the GRU (i.e., hybrid C-GRU) model has proven effective in various applications, including water level prediction (Pan et al., 2020), river flooding forecasting and anomaly detection (Miau & Hung, 2020), short-term residential load forecasting (Sajjad et al., 2020), soil moisture prediction (Yu et al., 2021), short-term canyon wind speed prediction (Ji et al., 2022), PM_{10} forecasting (Sharma et al., 2022) and evapotranspiration forecasting (Ahmed et al., 2021).

Therefore, the hybrid C-GRU model should also be explored for flood index forecasting.

2.4. Research gaps

The literature review outlined the key findings and identified a few notable research gaps in mathematically derived flood indices for flood risk monitoring, assessment, and forecasting. The $WRI_{24\text{-}hr\text{-}S}$ in the existing literature can be used to monitor flood risk during sustained extreme rainfall periods. This index is formulated to consider rainfall accumulation over the preceding 24 hours, whereby the rainfall contribution from the preceding hours is subjected to a time-dependent reduction function, accounting for water resource depletion through various hydrological processes. However, the primary limitation of this index is that it cannot objectively be used to monitor and compare flood risk across geographically diverse study sites as it is unnormalised. To overcome this issue, the hourly flood index ($SWRI_{24\text{-}hr\text{-}S}$) derived from normalising the $WRI_{24\text{-}hr\text{-}S}$ is proposed in the current study as a universal index to identify flood events across geographically diverse study regions and derive its associated characteristics (i.e., flood volume, peak and duration).

The literature also asserts that because flood event characteristics are generally interrelated, multivariate models that accurately capture these dependencies are essential for a comprehensive flood risk assessment. Consequently, an advanced statistical tool, i.e., copulas, can be employed for modelling multivariate distributions between random variables. The vine copula models among multivariate copula models have shown to be the most flexible approach for modelling the joint distribution between flood event characteristics to evaluate exceedance probability and joint and conditional return periods. Thus, the vine copulas can be utilised to model the joint distribution between the flood event characteristics derived from the proposed $SWRI_{24\text{-}hr\text{-}S}$ to extract their joint exceedance probability for probabilistic flood risk assessment. Evaluating exceedance probabilities is essential for early flood warning systems to enable improved flood risk assessment and the development of targeted mitigation strategies. This approach is crucial for flood-prone regions, including Fiji, where no research of such nature has been conducted before.

Lastly, the literature also shows the feasibility of index-based flood forecasting systems using hybrid DL/ML models to assess future flood risk, specifically demonstrated for the daily flood index (I_F). Consequently, a hybrid model such as C-

GRU, which integrates two powerful DL models—CNN for feature extraction and GRU for temporal modelling—should be designed and tested to model and forecast the proposed $SWRI_{24\text{-}hr\text{-}S}$ for assessing future flood risk. The current research, therefore, addresses and fills the aforementioned gap in the literature by developing novel and innovative flood monitoring, assessment, and forecasting tools using AI and copula-statistical methods to enhance flood warning systems. These tools will assist decision-makers in assessing flood risk more accurately, enabling them to better prepare for flood events and develop efficient plans to mitigate their severe impacts on communities.

CHAPTER 3: STUDY AREA, DATA, AND METHODOLOGY

This chapter offers a general overview of the case study area. Detailed information on the study locations and the data utilised to accomplish Objectives 1 and 2 are provided in Chapters 4 and 5 of the thesis. Similarly, while the methodologies for each objective are elaborated in their respective chapters, this chapter only presents a general summary of the methods employed to achieve the two research objectives.

3.1. Case study area

The case study area for this research was Fiji, which is centrally located over the International Date Line (180°) within the South-West Pacific region (Sharma et al., 2021), between latitudes 15°S and 22°S and longitudes 177°W and 174°E (see Figure 2(a)). Fiji has a tropical marine climate, with warm temperatures year-round. The nation has two distinct seasons: a warm, wet period from November to April and a cooler, drier season from May to October. This seasonal variation is primarily due to the South Pacific Convergence Zone (SPCZ), the main rainfall-producing system in the region, which typically lies over Fiji during the wet season.

Fiji has a total land area of about 18,333 km^2 spread over 332 islands, of which approximately 111 are inhabited (Kuleshov et al., 2014). Viti Levu and Vanua Levu are two large mountainous islands covering about 87% of the total land area (Feresi et al., 2000), with a maximum elevation of up to 1300 meters above sea level. Fiji has four divisions: Central, Eastern, Northern, and Western, with a population of 929,766 people (The World Bank Group, 2022).

Most of the population and infrastructure in Fiji are located on large floodplains susceptible to long-duration flooding or in small catchments prone to flash flooding (Government of Fiji, 2017). Fiji is severely affected by recurrent floods, primarily high-frequency, low-intensity events; although not all of them get recorded, their frequent occurrence leads to substantial cumulative losses over time (Government of Fiji, 2017). Therefore, developing and applying cost-effective and innovative tools to mitigate its severe impacts have become increasingly crucial. This research focused only on study sites in Viti Levu due to the lack of required rainfall data in other locations. Despite this limitation, the study is crucial for Viti Levu, which has the largest population and hosts the country's core economic activities, including sugar production, gold mining, and tourism. Figure 2(b) shows the map of Fiji with

corresponding study sites for Objective 1 of this research. Two of these study sites, Navua and Nausori, are in the Central Division of Fiji, and the rest are in the Western Division. For Objective 2, this study focused on five flood-prone sites (Rakiraki, Tavua, Lautoka, Nadi, and Sigatoka) in the Western Division of Fiji, shown in Figure 2(c).

3.2. Data

The rainfall data for the Rakiraki, Tavua, Ba, Lautoka, Nadi, Sigatoka, Navua, and Nasinu sites from 1 January 2014 to 31 December 2018 (5 years) were successfully acquired from Fiji Meteorological Services (FMS). For the Tavua, Rakiraki, Nasinu, Sigatoka, and Lautoka sites, the rainfall data were provided in 5-minute intervals, whereas for the Ba, Nadi, and Navua sites, the data were provided in 10-minute intervals. The rainfall data for each site was aggregated to calculate the hourly rainfall required for this study. If at least 66.67% of the data points for a given hour were available, they were summed to determine the total rainfall for that hour. Otherwise, the rainfall for that hour was recorded as missing. The Ba site, which had approximately 24% missing data, was excluded from further experiments. The remaining sites had less than 5% missing data. Thus, the Iterative K-nearest Neighbors (IKNN) technique (Oriani et al., 2020) was employed to fill in all the missing data for these study sites.

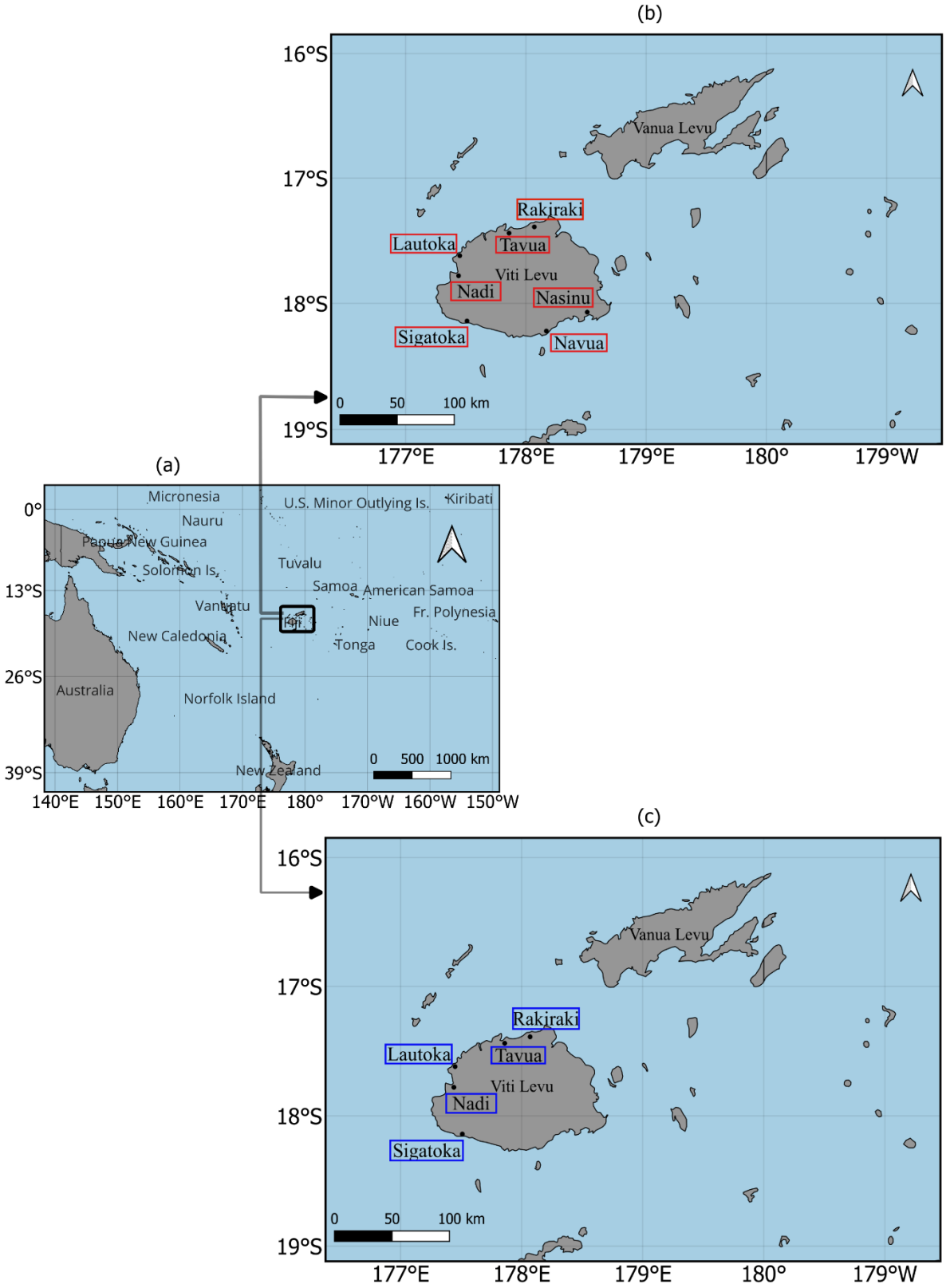

Figure 2 The geographical map shows Fiji's location in the South Pacific region (a) and a detailed inset map highlighting various study sites within Fiji for (b) Objective 1 and (c) Objective 2.

3.3. General methodology

The main steps undertaken to accomplish Objective 1 of this research are shown in Figure 3. The code to obtain the $SWRI_{24\text{-}hr\text{-}S}$ was developed using Python. Once the missing values were filled for each of the seven selected sites (Rakiraki, Tavua, Lautoka, Nadi, Sigatoka, Navua, and Nasinu), the rainfall data from 1 Jan 2014 to 31 December 2018 was used to calculate the $WRI_{24\text{-}hr\text{-}S}$ first. The $SWRI_{24\text{-}hr\text{-}S}$, a normalised form of the $WRI_{24\text{-}hr\text{-}S}$, was then calculated. The $SWRI_{24\text{-}hr\text{-}S}$ was then used to identify all the floods that occurred at each study site from 2014 to 2018 using the criteria $SWRI_{24\text{-}hr\text{-}S} > 0$. For each of these flood events, their onset time, duration (*D*), volume (*V*), and peak (*Q*), along with total precipitation, total $WRI_{24\text{-}hr\text{-}S}$, and maximum $WRI_{24\text{-}hr\text{-}S}$, were calculated and analysed. The annual and monthly climate summaries published by the FMS were then utilised to validate some of the major flood events identified by the $SWRI_{24\text{-}hr\text{-}S}$ at each study site, evaluating its practical utility in identifying flood events at an hourly scale. A comprehensive analysis was also carried out to thoroughly examine the total flood frequency by site, month, and year, among other aspects.

Moreover, for the probabilistic flood risk assessment across all study sites, this study focused on tri-variate cases to model the joint distribution of *D*, *V*, and *Q* (Nguyen-Huy et al., 2022). Before modelling, the flood characteristics dataset for each study site, which contained the duration (*D*), volume (*V*), and peak (*Q*) of each flood event, was further analysed by calculating its descriptive statistics, such as the five-number summary, mean, standard deviation, skewness, and kurtosis. These statistics were utilised to understand the shape and distribution of the dataset. A comprehensive correlation analysis was conducted to understand the relationship between each characteristic pair for modelling their joint distribution (Daneshkhah et al., 2016; Latif & Mustafa, 2020; Nguyen-Huy et al., 2022). Both parametric measures, i.e., Pearson's correlation coefficient (*r*), and nonparametric rank-based correlation measures, i.e., Spearman's correlation coefficient (ρ) and Kendall's tau (τ), were used to examine the relationship between each pair of flood characteristics. Mutual Information (MI) was also employed to assess the degree of dependence between each pair of flood characteristics.

After thoroughly analysing the attributes and dependency between each pair of flood characteristics across all study sites, the joint distribution between each flood characteristic was modelled using the 3D vine copula for each study site. These models were implemented using the R programming language, utilising the 'rvinecopulib' library package (Nagler & Vatter, 2023). The best-fitted 3D vine copula selected at each study site was then utilised to derive the joint exceedance probability between *D*, *V*, and *Q* under different combination scenarios. This was achieved by first quantifying that the *D*, *V*, and *Q* exceed specific thresholds simultaneously. The thresholds were selected at the 50th-quantile (median), 75th-quantile (moderate), and 95th-quantile (extreme). Finally, the joint exceedance probability between *D*, *V*, and *Q*, derived under various combination scenarios, was thoroughly analysed to understand the spatial pattern of flood risk across the study sites.

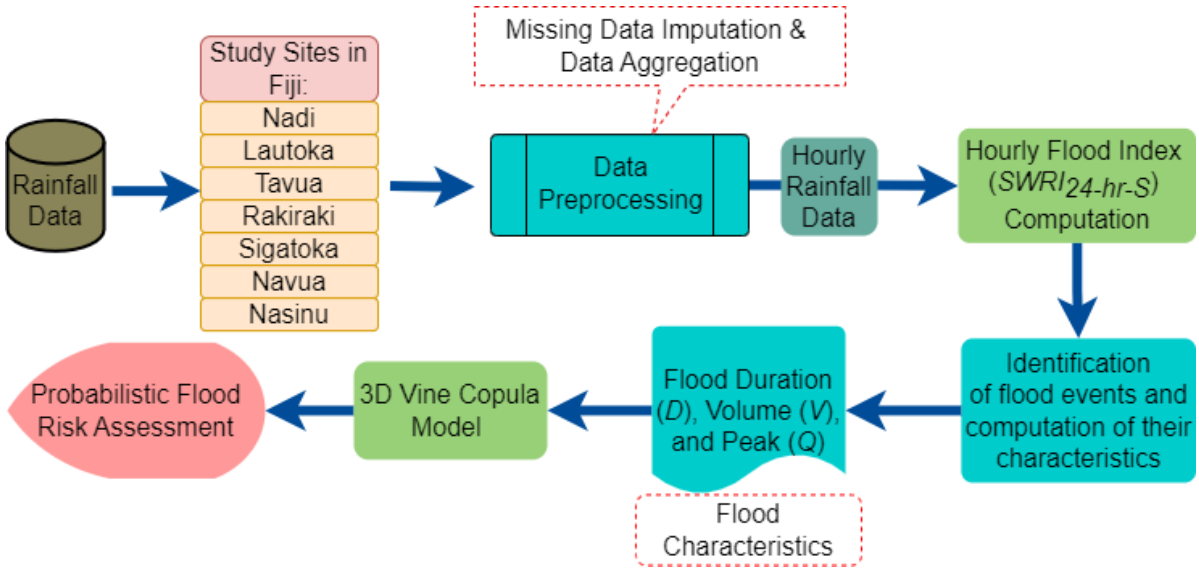

Figure 3 Schematic view of the steps untaken to develop an hourly flood monitoring tool and copula-probabilistic flood risk assessment system.

Furthermore, Figure 4 illustrates a brief schematic of the main steps undertaken to accomplish Objective 2 of this research. The hybrid C-GRU model was designed to forecast the $SWRI_{24\text{-}hr\text{-}S}$ over a short-term, i.e., 1-hourly forecast horizon for five flood-prone sites (Rakiraki, Tavua, Lautoka, Nadi, and Sigatoka) in Fiji. For this objective, only two datasets were utilised for each study site: the raw hourly rainfall data and the $SWRI_{24\text{-}hr\text{-}S}$ datasets, already available from Objective 1 for these study sites. Before commencing the model development, the Augmented Dickey-Fuller test (Cheung &

Lai, 1995; Dickey & Fuller, 1979) was first employed to check the stationarity of the individual time series data, i.e., hourly rainfall and the $SWRI_{24\text{-}hr\text{-}S}$ datasets for each study site. This step was crucial because statistical properties change over time for nonstationary data, making it difficult for ML/DL models to capture consistent patterns. This, in turn, impedes the models' ability to generalise to new data. After this, for each study site, the partial autocorrelation function (PACF) and cross-correlation function (CCF) were utilised to determine the most statistically significant lags for the $SWRI_{24\text{-}hr\text{-}S}$ and hourly rainfall time series, respectively, to forecast $SWRI_{24\text{-}hr\text{-}S}$ at time t. Hence, the features used as model inputs included the lagged $SWRI_{24\text{-}hr\text{-}S}$ and hourly rainfall. Subsequently, the predictor (or input features) and target variables for each site were concatenated to form the final dataset for developing the forecasting models.

The input features were then normalised in the range of 0 to 1 using the min-max scaling technique. This step ensures that each input feature shares the same order of magnitude, facilitating faster and more efficient model training (Prasad et al., 2024). The mathematical formula for this normalisation technique is given as follows (Joseph et al., 2024a):

$$X_n = \frac{X_{act} - X_{\min}}{X_{\max} - X_{\min}} \tag{1}$$

where X_n, X_{act}, $X_{\min}$, and $X_{\max}$ represent the normalised, actual, minimum, and maximum values of the input feature.

The dataset for model development at each study site was partitioned into training, validation, and testing subsets. Given the absence of consensus on data-splitting ratios, the first 80% of the data was allocated for training, with 20% of the training set reserved for validation and the remaining 20% for testing the model. These ratios were adopted from a related study (Moishin et al., 2021a). The validation dataset in this study was used for two purposes. First, it was used for model hyperparameter tuning. Second, for all DL models developed in this study, the validation data was employed to monitor the model's performance during training using the early stopping technique. Once the training, validation, and testing data were obtained for each study site, the proposed hybrid C-GRU model was developed. The hybrid C-GRU model combines CNN's feature extraction capabilities with GRU's efficiency in temporal modelling, resulting in a robust approach to sequential modelling. In addition, three DL

models—CNN, GRU, and LSTM—and one ML model, RFR, were developed for benchmarking.

All the models developed were highly parametric, particularly the hybrid and all DL models. Thus, effective hyperparameter tuning was necessary to optimise their performance and enhance their generalisation capabilities (Andonie, 2019; Eggensperger et al., 2013; Nguyen et al., 2020). This study employed an advanced BO technique utilising the Tree-structured Parzen Estimator (TPE) algorithm to fine-tune the hyperparameters of all developed models. All forecasting models were implemented using Python programming. This implementation was done via the Google Colaboratory (Google Colab) platform, which provides a freely available Jupyter Notebook interface supported by both a Tensor Processing Unit (TPU) and a Graphics Processing Unit (GPU). The DL models were developed using the Keras (Ketkar, 2017) and TensorFlow (Abadi et al., 2016) libraries, while the RFR model was implemented using the Scikit-learn library (Pedregosa et al., 2011). The 'Hyperopt' library (Komer et al., 2014) was utilised to implement BO with the TPE algorithm for efficient hyperparameter selection.

A diverse range of performance evaluation metrics was employed to evaluate the forecasting efficacy of the proposed hybrid C-GRU model against benchmark models across all study sites. Two categories of statistical metrics were utilised: Category A (ideal value = 1) and Category B (ideal value = 0). Within Category A, a total of five statistical metrics were employed, including Pearson's Correlation Coefficient (r), Nash-Sutcliffe Efficiency Index (E_{NS}), Willmott's Index of Agreement (E_{WI}), Legate-McCabe Efficiency Index (E_{LM}), and Kling-Gupta Efficiency (KGE). In Category B, a total of three error metrics were included: Mean Absolute Error (MAE), Root Mean Square Error ($RMSE$), and Symmetric Mean Absolute Percentage Error ($sMAPE$; %). The Python package 'HydroErr' (Roberts et al., 2018) was used to implement these performance evaluation metrics. A Global Performance Indicator (GPI) was additionally utilised to combine the outcomes of all metrics, enabling convenient ranking and comparison of numerous models. Furthermore, the Diebold–Mariano (DM) statistical test (Diebold & Mariano, 2002) was also employed to determine whether the proposed hybrid C-GRU model's performance was statistically significantly superior to the benchmark models. In addition to the diverse statistical indicators, the performance of the proposed hybrid C-GRU models was compared with

benchmark models using various visual diagnostic plots, including line, scatter, and empirical cumulative distribution function (ECDF) plots across all study sites.

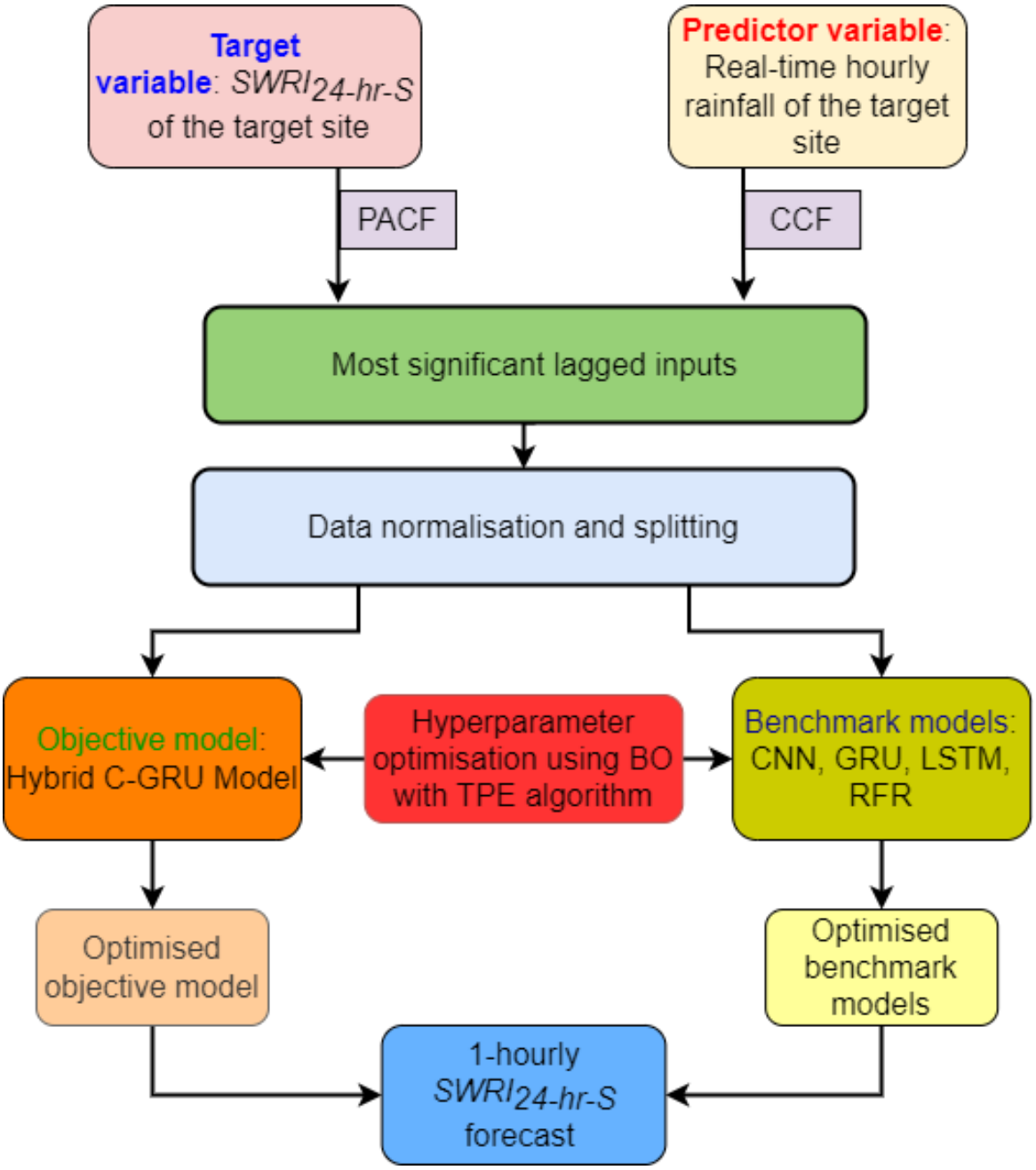

Figure 4 Schematic view of the proposed hybrid C-GRU-based $SWRI_{24\text{-}hr\text{-}S}$ forecasting system.

CHAPTER 4: PAPER 1 – COPULA-PROBABILISTIC FLOOD RISK ANALYSIS WITH AN HOURLY FLOOD MONITORING INDEX

4.1. Introduction

This chapter is an exact copy of the published article in the *Water* Journal, Vol. 11, page 1560 (2024) (Scopus Impact Factor 5.8).

The chapter introduces a novel hourly flood index ($SWRI_{24\text{-}hr\text{-}S}$) to monitor flood risk across geographically diverse study sites. The proposed $SWRI_{24\text{-}hr\text{-}S}$ is computed by first calculating the $WRI_{24\text{-}hr\text{-}S}$ using the site-based hourly rainfall data. The $WRI_{24\text{-}hr\text{-}S}$ takes into account the rainfall accumulation over the preceding 24 hours, whereby the rainfall contribution from preceding hours is subjected to a time-dependent reduction function, which accounts for the depletion of water resources through various hydrological processes (Deo et al., 2018). The $SWRI_{24\text{-}hr\text{-}S}$ is then derived from normalising the $WRI_{24\text{-}hr\text{-}S}$. The practical utility of the $SWRI_{24\text{-}hr\text{-}S}$ in identifying flood events on an hourly scale and computing their characteristics (i.e., flood volume (V), duration (D) and peak (Q)) is demonstrated for the seven flood-prone sites in Fiji. The results indicate that Fiji mainly experiences high rainfall during the wet/cyclone season from November to April, including the months of May and October. As a result, the frequency of flood events is higher, with greater flood volume during these months than others. Therefore, it is imperative that relevant authorities, such as Fiji's NDMO, implement comprehensive and well-structured flood preparedness and risk mitigation measures for the wet/cyclone season (November to April), extending to include the months of May and October.

The flood event characteristics and water-intensive properties for the five severe flood events at each of the seven study sites are also presented. The most severe flood event at each study site is validated using the climate summaries published by FMS. Relevant decision-makers are expected to use these findings to understand the characteristics of past flood occurrences at these sites, thus facilitating future decision-making and planning to mitigate the impacts of flooding for these study sites.

The flood event characteristics data, which included *D*, *V*, and *Q* of each flood event identified at each study site, are statistically analysed. Subsequently, the 3D vine copula is employed for each study site to model the joint distribution between flood characteristics to extract their joint exceedance probability for probabilistic flood risk assessment. The results indicate a moderate but notable variation in the spatial patterns of joint exceedance probabilities for flood event characteristics across different combination scenarios.

4.2. Published paper

water

MDPI

Article

Copula-Probabilistic Flood Risk Analysis with an Hourly Flood Monitoring Index

Ravinesh Chand [1], Thong Nguyen-Huy [2,3,*], Ravinesh C. Deo [1], Sujan Ghimire [1], Mumtaz Ali [4,5] and Afshin Ghahramani [6,†]

1 School of Mathematics, Physics and Computing, University of Southern Queensland, Springfield, QLD 4300, Australia
2 Centre for Applied Climate Sciences, University of Southern Queensland, Toowoomba, QLD 4350, Australia
3 Faculty of Information Technology, Thanh Do University, Kim Chung, Hoai Duc, Ha Noi 100000, Vietnam
4 UniSQ College, University of Southern Queensland, Toowoomba, QLD 4350, Australia; mumtaz.ali@unisq.edu.au
5 New Era and Development in Civil Engineering Research Group, Scientific Research Center, Al-Ayen University, Thi-Qar, Nasiriyah 64001, Iraq
6 University of Southern Queensland, Toowoomba, QLD 4350, Australia
* Correspondence: thong.nguyen-huy@unisq.edu.au or nhthong@thanhdouni.edu.vn
† Current address: Department of Environment and Science and Innovation, Queensland Government, Toowoomba, QLD 4350, Australia.

Abstract: Floods are a common natural disaster whose severity in terms of duration, water resource volume, peak, and accumulated rainfall-based damage is likely to differ significantly for different geographical regions. In this paper, we first propose a novel hourly flood index ($SWRI_{24-hr-S}$) derived from normalising the existing 24-hourly water resources index ($WRI_{24-hr-S}$) in the literature to monitor flood risk on an hourly scale. The proposed $SWRI_{24-hr-S}$ is adopted to identify a flood situation and derive its characteristics, such as the duration (D), volume (V), and peak (Q). The comprehensive result analysis establishes the practical utility of $SWRI_{24-hr-S}$ in identifying flood situations at seven study sites in Fiji between 2014 and 2018 and deriving their characteristics (i.e., D, V, and Q). Secondly, this study develops a vine copula-probabilistic risk analysis system that models the joint distribution of flood characteristics (i.e., D, V, and Q) to extract their joint exceedance probability for the seven study sites in Fiji, enabling probabilistic flood risk assessment. The vine copula approach, particularly suited to Fiji's study sites, introduces a novel probabilistic framework for flood risk assessment. The results show moderate differences in the spatial patterns of joint exceedance probability of flood characteristics in different combination scenarios generated by the proposed vine copula approach. In the worst-case scenario, the probability of any flood event occurring where the flood volume, peak, and duration are likely to exceed the 95th-quantile value (representing an extreme flood event) is found to be less than 5% for all study sites. The proposed hourly flood index and the vine copula approach can be feasible and cost-effective tools for flood risk monitoring and assessment. The methodologies proposed in this study can be applied to other data-scarce regions where only rainfall data are available, offering crucial information for flood risk monitoring and assessment and for the development of effective mitigation strategies.

Keywords: flood characteristics; flood monitoring; hourly flood index; joint distribution; risk mitigation; vine copulas

Citation: Chand, R.; Nguyen-Huy, T.; Deo, R.C.; Ghimire, S.; Ali, M.; Ghahramani, A. Copula-Probabilistic Flood Risk Analysis with an Hourly Flood Monitoring Index. *Water* **2024**, *16*, 1560. https://doi.org/10.3390/w16111560

Academic Editor: Francesco De Paola

Received: 15 April 2024
Revised: 18 May 2024
Accepted: 22 May 2024
Published: 29 May 2024

1. Introduction

Flooding is a catastrophic natural disaster progressively increasing in frequency and severity, primarily attributed to climate change–induced phenomena such as increased rainfall intensity. Generally, there are three prevalent flood types: fluvial or river floods, pluvial or flash floods, and coastal floods [1,2]. A flash flood is a sudden and severe local

inundation often resulting from high-intensity rainfall (e.g., tropical cyclones, slow-moving tropical depressions, or thunderstorms) within a short period (usually less than six hours) and/or may also be caused by sudden discharge of impounded water (e.g., dam or levee failures, ice jam release, or a glacier lake outburst) [3,4]. Flash floods can affect a range of locations, including river plains, valleys, and areas with steep terrain, elevated surface runoff rates, constrained stream channels, and persistent heavy convective rainfall [3]. They often necessitate prompt action to mitigate their severe impact, typically relying on expeditious decision-making and emergency response. Flash floods make up about 85% of all floods, resulting in more than 5000 deaths annually and causing severe social, economic, and environmental impacts [5]. The repercussions of flood disasters are more devastating in developing countries such as Fiji [6], where this study is focused. Therefore, developing a real-time flood risk monitoring tool remains an ongoing research motivation to enable an assessment of flood occurrences for early warning systems in Fiji.

Fiji experiences regular flooding events arising from orographic rainfall due to the topography of its larger islands, including Viti Levu and Vanua Levu, which have a maximum elevation of 1300 m above sea level, along with the impact of prevailing southeast trade winds [7]. For Fiji, about 90% of its population resides in coastal areas susceptible to floods [8]. Between 1970 and 2016, Fiji faced 44 major flood events, impacting approximately 563,310 people and resulting in 103 fatalities [9]. The most catastrophic floods occurred in 2004, 2009, 2012 (including the January and March flood events), and 2014. The 2009 and 2012 events, considered among the worst in the nation's history, resulted in over 200 million FJD in damages and losses, causing 15 fatalities and directly affecting more than 160,000 people [9]. For Fiji, the estimated average annual flood losses exceed 400 million FJD, equivalent to 4.2% of Fiji's Gross Domestic Product (GDP) [9]. These are substantial losses for a small island nation with a population of less than a million and a GDP of less than 5 billion USD [10]. Under the assumption that climate change conditions will significantly increase rainfall, the annual flood-related asset losses could exceed 5% of Fiji's GDP by 2050 without adaptation measures [9]. As a result, it is imperative to develop reliable methods for accurately monitoring flood risk on near real-time (e.g., hourly) timescales to mitigate the severe impacts of flooding.

Intense and/or prolonged precipitation is among the primary causes of floods. However, to better understand flood risk, multiple factors must be considered. These include the combined impact of various weather or climate events such as temperature and precipitation (while these events may not individually reach extreme levels, their cumulative effect can result in severe impacts) as well as the vulnerability and exposure of the affected area [11,12]. Climate change also induces changes in various flood-related factors, including precipitation, soil moisture content, sea level, and glacial lake conditions, potentially changing flood characteristics [12]. Other factors, such as land use and cover, catchment size and shape, drainage networks, and dam or levee construction, can also influence flood dynamics. Hence, an index that integrates all these factors and the accumulative impacts of other weather or climate characteristics is essential for a comprehensive flood risk assessment.

In many developing nations where flood monitoring resources, hydro-meteorological datasets, and risk monitoring facilities are underdeveloped, applying a mathematically derived flood index utilising only rainfall data provides a key strategy for assessing an impending flood risk situation. Some of the key mathematical indices used previously in flood risk monitoring include the Standardised Precipitation Index (SPI), the Available Water Resources Index ($AWRI$), the Weighted Average of Precipitation (WAP), the Standardised WAP ($SWAP$), the Flood Index (I_F), and the Standardised Antecedent Precipitation Index ($SAPI$) [13–18]. The flood indices, such as $AWRI$, $SWAP$, I_F, and $SAPI$, are robust as they are designed to account for changes in antecedent or immediate past rainfall by employing a suitable time-dependent reduction function that accounts for the depletion of water resources through various hydrological processes. For example, the daily flood index, I_F, applied in Australia [17,19,20], Iran [21], Bangladesh [22,23], and Fiji [6], has shown good

performance in monitoring flood events on a daily scale. Despite its benefits, one primary weakness of I_F and other indices, such as SPI, lies in their utilisation of daily, monthly, or annual accumulated rainfall data, which represent much longer timescales than what is required in a flash flood monitoring system. Consequently, these indices fail to adequately represent the flood risk caused by bursts of high-intensity rainfall and rapid responses leading to flash flood events.

The present study draws relevance from a pilot study conducted by Deo et al. [24] that proposed a 24-h water resources index ($WRI_{24-hr-S}$) based on a concept similar to the $AWRI$, which was applied in two study locations, Australia and South Korea, to monitor the flash flood risk in sustained extreme rainfall periods. The $WRI_{24-hr-S}$ monitors flood risk by considering the contribution of accumulated rainfall in the past 24 h, whereby the rainfall contribution from the preceding hours is subjected to the time-dependent reduction function that accounts for the depletion of water resources through various hydrological processes such as evaporation, percolation, seepage, runoff, and drainage. However, unlike $SAPI$ and I_F, which are normalised values derived from the Antecedent Precipitation Index (API) and the $AWRI$, respectively, the identification of flood events and the computation of their characteristics are not achievable with the current form of $WRI_{24-hr-S}$. This is primarily because this index is unnormalised and does not enable objective assessment of flood risk across geographically diverse study sites.

To enhance the understanding of flood risks, it is essential to calculate the flood characteristics, including the volume (V), peak (Q), and duration (D) that concurrently result in major collateral damage. As the flood characteristics such as the D, V, and Q are mostly interrelated, we envisage that these characteristics should be jointly considered in a multivariate analysis model to estimate the actual probability of a flood occurrence [18,25]. Importantly, any model representing the joint distribution of D, V, Q, and other crucial flood characteristics, such as the onset and withdrawal of a flood event, can provide significant insights into the relative severity of any flood event. After the initial study of Sklar [26], copula-based models became attractive in modelling interrelated variables, albeit using a multivariate approach. As such, copulas can jointly model the distribution of flood characteristics such as the D, V, and Q, regardless of their marginal distributions or whether their dependence structure is linear or non-linear [18]. Vine copulas have recently shown superior capabilities in modelling flood characteristics compared to traditional Archimedean and elliptical copulas [18,27–29]. Therefore, this research follows a recent study by Nguyen-Huy et al. [18], which used vine copulas to model the joint distribution of extreme flood characteristics derived using the $SAPI$ in Myanmar. Their study has provided interesting insights into the probabilistic flood risk analysis. To the best of our knowledge, no prior research has applied vine copulas to analyse the probabilistic flood risk in Fiji despite floods being a catastrophic phenomenon on this small island nation.

The scientific contributions of this paper, with significant implications for flood risk monitoring and assessment, are threefold. Firstly, the paper advances the concept of the 24 h water resources index pioneered by Deo et al. [24] and formulates a novel hourly flood index ($SWRI_{24-hr-S}$) (a normalised metric) by normalising the 24-hourly water resources index in such a way that enables the objective assessment of flood risk across geographically diverse study sites. Secondly, the present study adopts the $SWRI_{24-hr-S}$, which is computed using real-time hourly rainfall data from 2014 to 2018 obtained from the Fiji Meteorological Services (FMS), to evaluate its practical utility in identifying flood situations and computing their associated flood characteristics (i.e., D, V, and Q) for seven different study sites in Fiji. Thirdly, the present study develops the vine copula approach for the first time to model the joint distribution of D, V, and Q derived from the $SWRI_{24-hr-S}$ for specific cases of Fiji's flood events to extract their joint exceedance probabilities for probabilistic flood risk assessment. Fiji is a Pacific Small Island Developing State (PSIDS) that frequently experiences recurrent flooding. The lack of advanced infrastructure and necessary data in Fiji makes continuous flood risk monitoring and assessment challenging. The main aim of the present study is, therefore, to develop a novel hourly flood index,

$SWRI_{24-hr-S}$, using only hourly rainfall data (which are readily available for the present study sites) and to conduct a probabilistic flood risk assessment by modelling the joint distribution of extreme flood characteristics derived from the $SWRI_{24-hr-S}$ for Fiji's case studies. Hence, the methodologies presented in this study aim to enhance and contribute to existing monitoring and early warning systems for flash floods in Fiji. Moreover, these approaches may also be applied in other flood-prone regions globally, particularly those facing similar data scarcity challenges. By adapting these methodologies, vulnerable communities can benefit from improved flood preparedness and mitigation strategies.

The rest of the paper is structured as follows. Section 2 provides information on the study area, the dataset used, and data pre-processing steps. It also encompasses the mathematical methodology for computing the hourly flood index and the flood characteristics. Additionally, it provides details on Vine copula models and equations used for computing the joint exceedance probability of flood characteristics. Section 3 provides the results and discussion. Section 4 highlights the key findings, the study's limitations, and insights for future research.

2. Materials and Methods

2.1. Study Area

The proposed copula-probabilistic flood risk analysis system based on the hourly flood index was applied to geographically diverse sites in Fiji. Fiji is located in the South Pacific Ocean at a latitude of 15° S to 22° S and a longitude of 177° W to 174° E (Figure 1), with a tropical maritime climate characterised by warm temperatures throughout the year [30,31]. The nation comprises an archipelago of 332 islands, 111 of which are permanently inhabited, with a total land area of about 18,333 km^2 [30]. Viti Levu (10,400 km^2) and Vanua Levu (5540 km^2) are two large mountainous islands covering about 87% of the total land area [31].

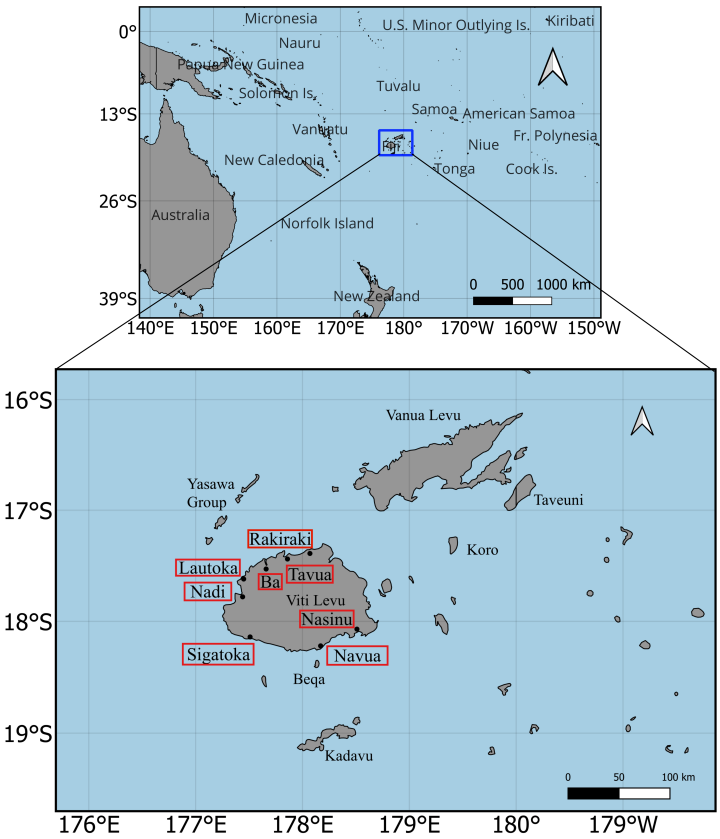

Figure 1. The geographical map shows Fiji's location in the South Pacific region and a detailed inset map highlighting various study sites within Fiji.

Fiji has a distinct dry season (May–October) and wet season (November–April). This seasonal variation is mainly attributed to the South Pacific Convergence Zone (SPCZ), the primary rainfall-producing system for the region, which lies typically over Fiji during the wet season [31,32]. The rivers and stream basins in Fiji are predominantly small in size and flow from steep mountainous terrain, resulting in rapid shifts in water levels during periods of intense rainfall, which can lead to flash floods within a few hours [33]. This study included sites only in Viti Levu due to the lack of rainfall data in other locations. These sites are areas in Fiji prone to recurrent and severe flooding events. Figure 1 shows a map of the study area and the corresponding study sites.

2.2. Dataset

The rainfall data for the Lautoka, Sigatoka, Nasinu, Rakiraki, Navua, Nadi, Ba, and Tavua sites from 1 January 2014 to 31 December 2018 (5 years) were successfully acquired from Fiji Meteorological Services. The rainfall data were provided in 5 min intervals for Tavua, Rakiraki, Nasinu, Sigatoka, and Lautoka and in 10 min intervals for the Ba, Nadi, and Navua sites. During the data pre-processing phase, the rainfall data for each site were aggregated to obtain the hourly rainfall needed for this study. If at least 66.67% of the data points (i.e., at least 4 out of 6 data points for a 10 min interval or at least 8 out of 12 data points for a 5 min interval) were available for a particular hour, they were summed to determine the total rainfall for that hour. Otherwise, the rainfall value for that hour was marked as missing. This approach was adopted to maximise the recovery of data values.

Table 1 summarises the hourly rainfall datasets and geographic settings of the study sites. The Ba site, which had a high percentage of missing values, was excluded from this study. The remaining sites had less than 5% missing values; therefore, any gap-filling method could fill in the missing values [34]. Based on the study by Oriani et al. [35], the Iterative K-nearest Neighbour (IKNN) technique was used to fill in all the missing data. The data from 1 January 2014 to 31 December 2018 were used for all the computations. However, $WRI_{24-hr-S}$ followed by $SWRI_{24-hr-S}$ were calculated from 2 January 2014 as antecedent rainfall of 24 h (the hourly rainfall data for 1 January 2014), which was necessary to allow the calculations of these metrics.

Table 1. Overview of the hourly rainfall datasets for the 8 sites in Fiji. (Note: The hourly rainfall spans from 1 January 2014 to 31 December 2018, with 43,824 expected observations.)

Study Site	Location	Missing Data (%)	Average Hourly Rainfall (mm)	Maximum Hourly Rainfall (mm)
Ba	17.53° S, 177.66° E	23.76	0.24	56.00
Lautoka	17.62° S, 177.45° E	0.83	0.19	83.50
Nadi	17.78° S, 177.44° E	1.17	0.27	260.00
Nasinu	18.07° S, 178.51° E	1.18	0.33	72.00
Navua	18.22° S, 178.17° E	1.57	0.36	62.50
Rakiraki	17.39° S, 178.07° E	3.76	0.23	68.50
Sigatoka	18.14° S, 177.51° E	1.99	0.21	59.00
Tavua	17.44° S, 177.86° E	3.45	0.16	57.50

2.3. Development of the Hourly Flood Index and Vine Copula Model

2.3.1. Hourly Flood Index and Flood Characteristics

This research formulates a novel hourly flood index ($SWRI_{24-hr-S}$), which is a normalised version of the 24-hourly water resources index ($WRI_{24-hr-S}$) proposed by Deo et al. [24]. Applying this index to Fiji is significantly advantageous because it is mathematically derived using only hourly rainfall data, which are readily available for the present study sites. The proposed $SWRI_{24-hr-S}$ is implemented using the Python programming language.

The following steps are taken to obtain the $SWRI_{24-hr-S}$. The first step is calculating the $WRI_{24-hr-S}$. The $WRI_{24-hr-S}$ for the current (i^{th}) hour is given by the following equation [24]:

$$WRI^{i}_{24-hr-S} = P_1 + \frac{[P_2(W-1)]}{W} + \frac{[P_3(W-1-1/2)]}{W} + ... + \frac{[P_{24}(W-1-1/2-\cdots-1/23)]}{W} \quad (1)$$

where P_1 is the total rainfall recorded an hour before, P_2 is the total rainfall recorded 2 h before, and so on; W is the time-reduction weighting factor ($W \equiv 1 + 1/2 + 1/3 + ... + 1/24 \approx 3.8$) that incorporates the contributions of accumulated rainfall in the latest 24 h [24]. This weighting factor ensures that the decay of accumulated rainfall or its potential impact on a flood event depends on several hydrological effects such as evapotranspiration, percolation, seepage, runoff, drainage, etc., in accordance with earlier works [15,24]. The substitution of W = 3.8 into Equation (1) yields the following:

$$WRI^{i}_{24-hr-S} \approx P_1 + 0.74P_2 + 0.61P_3 + ... + 0.02P_{24} \quad (2)$$

Notably, the $WRI_{24-hr-S}$ for a current (ith) hour is expected to accumulate 100% of rainfall received an hour before, ≈74% of that received two hours before, ≈61% of that received three hours before, and eventually ≈2% of that received 24 h before.

After calculating $WRI_{24-hr-S}$ for any study period, the mathematical form of $SWRI_{24-hr-S}$ for a current (ith) hour is calculated as a normalised version of $WRI_{24-hr-S}$:

$$SWRI^{i}_{24-hr-S} = \frac{WRI^{i}_{24-hr-S} - \overline{WRI^{max}_{24-hr-S}}}{\sigma(WRI^{max}_{24-hr-S})} \quad (3)$$

where $\overline{WRI^{max}_{24-hr-S}}$ is the mean monthly maximum values of $WRI_{24-hr-S}$ for the respective study period and $\sigma(WRI^{max}_{24-hr-S})$ is the standard deviation of the monthly maximum values of $WRI_{24-hr-S}$ for the respective study period.

For the purpose of this paper, we follow the notion that if the magnitude of $SWRI_{24-hr-S}$ for the current (ith) hour is greater than zero (or that the water resources are higher than normal), it is regarded as a flood situation. In this paper, we defined flood characteristics using the running-sum methodology of Yevjevich [36], which has also been used in several other studies [6,17,18]. With reference to Figure 2, the flood duration, D, is estimated as the number of hours between the start of a flood, t_{onset}, and the end of a flood, t_{end}, derived from the $SWRI_{24-hr-S}$ time series. In accordance with the onset of a flood, the volume of the flood, V, is calculated as the sum of all values of $SWRI_{24-hr-S}$ between t_{onset} and t_{end} of a flood situation, and the peak of the flood, Q, is determined as the maximum $SWRI_{24-hr-S}$ during any flood situation.

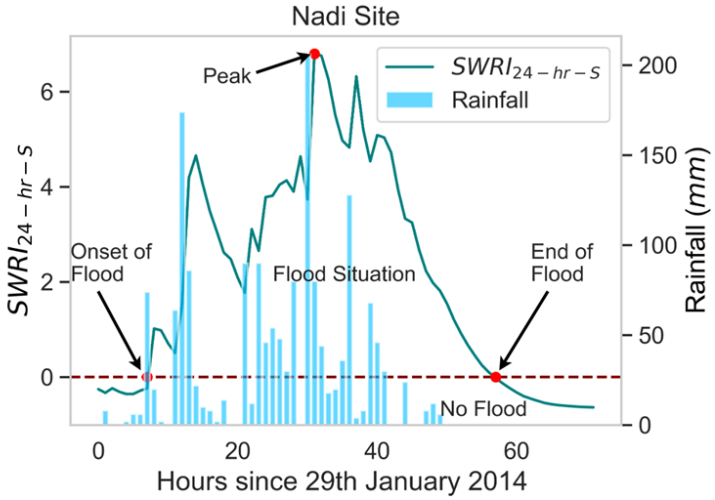

Figure 2. The $SWRI_{24-hr-S}$ developed to identify a flood event in January 2014 at the Nadi study site, demonstrating its capability to determine the duration, volume, and peak of any flood event. The flood volume, representing the accumulated water resources, is the cumulative $SWRI_{24-hr-S}$ under the curve closed by the onset and end of a flood situation and the zero line.

Equations (4)–(6) show the mathematical equations used to calculate the flood characteristics before developing the copula-probabilistic flood risk analysis model.

$$V = \sum_{t=t_{onset}}^{t=t_{end}} SWRI_{24-hr-S} \tag{4}$$

where $SWRI_{24\text{-}hr\text{-}S} > 0$

$$D = t_{end} - t_{onset}(hours) \tag{5}$$

$$Q = max(SWRI_{24-hr-S})_{t_{onset}-t_{end}} \tag{6}$$

where t_{onset} and t_{end} are the onset and end timestamps of a flood situation, respectively.

To demonstrate the practical use of $SWRI_{24-hr-S}$ for hourly risk flood monitoring, Figure 2 illustrates the $SWRI_{24-hr-S}$ applied to identify flood events in January 2014 at the Nadi site located in the western division of Fiji. As illustrated in Figure 2, the onset timestamp of the flood situation, i.e., the exact hour when the magnitude of $SWRI_{24-hr-S}$ starts to rise above zero, was on 29 January 2014 at 8 a.m. To verify this particular flood situation, we now refer to the report from the FMS [37], which showed indeed that an active trough that caused widespread rain across Fiji was noticeable from 29 to 30 January 2014 and resulted in flooding, particularly in the western division of Fiji, where this study site is located. Thus, this verification confirms that the proposed $SWRI_{24-hr-S}$ has correctly identified this flood event, demonstrating its practicality in identifying a flood situation at an hourly scale.

To further verify the potency of $SWRI_{24-hr-S}$ for hourly flood risk monitoring, in Figure 3, $WRI_{24-hr-S}$ is plotted alongside the hourly rainfall data for the same study site during the same period. Compared to $SWRI_{24-hr-S}$ or the raw hourly rainfall data, $SWRI_{24-hr-S}$ simplifies the process of identifying a flood situation. This is because a simple criterion, $SWRI_{24-hr-S} > 0$, provides a good indicator of flood risk, which is impossible when using $WRI_{24-hr-S}$ and the raw hourly rainfall values.

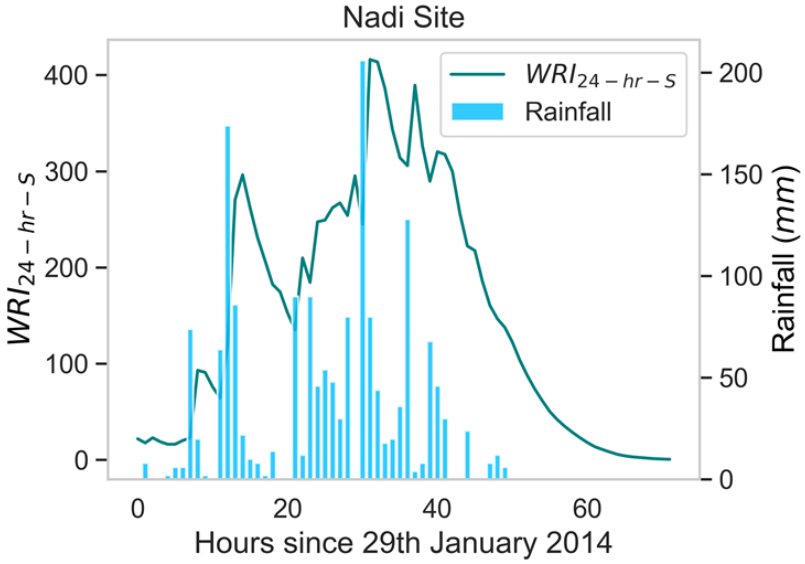

Figure 3. The $WRI_{24-hr-S}$ and rainfall since 29 January 2014 (72 h) for the Nadi site.

2.3.2. Joint Exceedance Probability between Flood Characteristics

For a comprehensive flood risk assessment, this study follows the original approach of Nguyen-Huy et al. [18] to develop vine copula-based joint exceedance probability models. This task entails developing a multivariate analysis system of flood characteristics that considers the joint exceedance probability of a flood duration D, volume V, and peak Q for the present study sites. This study specifically aimed to estimate the probability that

the duration, volume, and peak were concurrently greater than or equal to some threshold scenarios, as presented below:

$$P(D \geq d \wedge V \geq v \wedge Q \geq q) \tag{7}$$

Equation (7) requires the modelling of a joint distribution function of three variables, $F(x_d, x_v, x_q)$. Thus, in this study, we have developed a copula-based model, described in the following section, to estimate the joint exceedance probability of the flood characteristics, i.e., D, V, and Q to perform a probabilistic flood risk analysis.

2.3.3. Copula Analytical Approach

A copula $C(.) : [0,1]^n \rightarrow [0,1]$ is a function that links univariate marginal distribution functions $P(X_i \leq x_i) = F_i(x_i)$ of random variables $X_1, \ldots, X_n$ to form a joint cumulative distribution function (JCDF), $P(X_1 \leq x_1, \ldots, X_n \leq x_n) = F(x_1, \ldots, x_n)$, i.e., [26]:

$$F(x_1, \ldots, x_n) = C[F_1(x_1), \ldots, F_n(x_n)] \tag{8}$$

with the corresponding joint density distribution function (JPDF) in terms of marginal and copula probability density functions [26]:

$$f(x_1, \ldots, x_n) = \left[\prod_{i=1}^{n} f_i(x_i)\right] c[F_i(x_i), \ldots, F_n(x_n)] \tag{9}$$

where $f_i(x_i)$ and $c(.)$ are the corresponding marginal and copula PDFs, respectively. When the marginal distributions are continuous, a unique copula exists. Equations (8) and (9) demonstrate an advanced capability of copulas, allowing a JCDF of random variables to be constructed through two separate processes: (i) modelling a copula function that captures the dependence structure among correlated variables and (ii) modelling the univariate marginal distributions. This aspect of copulas presents a more flexible approach for choosing suitable univariate distribution functions to fit the observed data in practical applications. From Equation (8), the joint distribution of duration, volume, and peak given in Equation (7) can be written using copulas as follows:

$$\begin{aligned} &P(D \geq d \wedge V \geq v \wedge Q \geq q) \\ &= 1 - F_D(d) - F_V(v) - F_Q(q) + F_{DV}(d,v) + F_{VQ}(v,q) + F_{DQ}(d,q) - F_{DVQ}(d,v,q) \\ &= 1 - F_D(d) - F_V(v) - F_Q(q) + C_{DV}[F_D(d), F_V(v)] + C_{VQ}\left[F_V(v), F_Q(q)\right] + \\ &\qquad C_{DQ}\left[F_D(d), F_Q(q)\right] - C_{DVQ}\left[F_D(d), F_V(v), F_Q(q)\right] \end{aligned} \tag{10}$$

Different copula families, such as empirical, Archimedean, extreme value, elliptical, vine, and entropy copulas, can be used to model the copula function given in Equation (10) [38,39]. Vine copulas, among other copulas, can be used to achieve the utmost flexibility in constructing the JCDF and JPDF, given in Equations (8) and (9), respectively. Vine copulas have been applied in recent studies across various fields, such as weather and climate risk in agriculture [40,41], hydrology and water resources [27,42–48], and finance and insurance [49–51]. The following section provides more details on vine copulas.

2.3.4. Vine Copulas

The vine copula was first introduced by Joe [52], whose concept was to decompose the JPDF into a cascade of iteratively conditioned bivariate copulas, also called pair copulas. While this decomposition is not unique, all possible decompositions can be organised into a graphical model called a regular vine (R-vine) [53].

Within the R-vine framework, two main types of vine copula decomposition exist: the canonical (C-vine) and drawable (D-vine) distributions. These modes determine the parametric construction of an R-vine. The D-vine copula offers higher flexibility than

the C-vine copula, especially when considering all mutual inter-correlations between targeted random variables one after another [47]. However, the C- and D-vine frameworks are the same when considering a 3-dimensional (3D) (or tri-variate) joint distribution framework [18,47].

In this study, we have focused on tri-variate cases to model the joint distribution of flood duration, volume, and peak for a detailed probabilistic risk analysis of any flood event. Figure 4 shows a graphical representation of D- and C-vine copulas in the form of trees, edges, and nodes.

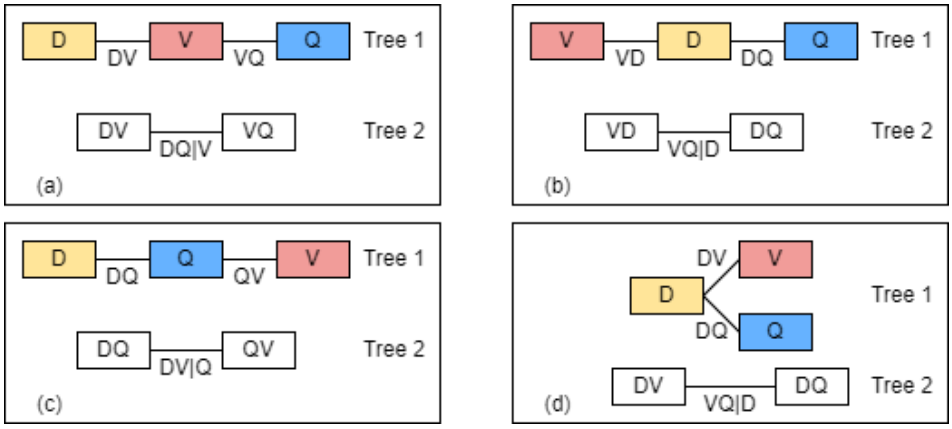

Figure 4. Schematic diagram of different ways of constructing the 3D D-vine copula structure in this study. (**a**) Case 1: the flood volume (V) as a conditioning variable, (**b**) Case 2: the flood duration (D) as a conditioning variable, (**c**) Case 3: the flood peak (Q) as a conditioning variable, and (**d**) an example of the C-vine copula structure. Source: Adapted from Nguyen-Huy et al. [18].

In the tri-variate case where D, V, and Q are modelled simultaneously, the C-vine copula is the D-vine copula with a specified centre variable, as previously mentioned. For instance, when the flood duration (D) variable serves as the centre variable, the D-vine copula depicted in Figure 4b is identical to the C-vine copula shown in Figure 4d. The edges are linked to bivariate copulas, such as the edge DV associated with the bivariate copula C_{DV}, which captures the dependence structure between D and V.

To fit the univariate marginal distribution functions, we employed the univariate local-polynomial likelihood kernel density estimation method capable of handling discrete (duration) and continuous (volume and peak) data [54]. Additionally, the following bivariate copula families were utilised to develop the 3D vine copula models in this study: independence, parametric (elliptical, Archimedean, and their rotated versions), and non-parametric (transformation kernel) families [54–56] (Table A1).

To estimate the parameters of bivariate copulas, we employed maximum likelihood for parametric models and local-likelihood approximations for non-parametric models. Moreover, the modified vine copula Bayesian information criteria (mBICv) [57] was utilised to select the bivariate copulas, and Kendall's tau (τ) was adopted to select tree sequences [18]. The mBICv can address the issues of the Bayesian information criterion (BIC), which assumes that the number of possible parameters grows sufficiently slowly with the sample size n and that all models are equally likely. Additionally, the mBICv was explicitly tailored to vine copula models [57]. The vine copula models were developed using the R programming language utilising the 'rvinecopulib' library package [54].

3. Results and Discussion

We now provide a detailed appraisal of the hourly flood index $SWRI_{24-hr-S}$ for detecting hourly flood possibility in terms of the onset and the end timestamps, duration, peak, volume, and total accumulated rainfall during any flood situation for seven study sites in Fiji over the study period (2014–2018). We also provide joint distribution model

results using the newly proposed vine copulas to provide a probabilistic risk analysis framework for flash floods.

3.1. Application of the Hourly Flood Index for Flood Event Analysis

The $WRI_{24-hr-S}$, followed by the $SWRI_{24-hr-S}$ for the study period (2014–2018), were successfully computed for each of the seven study sites. The practicality of $SWRI_{24-hr-S}$ for determining a flood situation has already been demonstrated in Figure 2. Similarly, the flood events between 3 and 6 April 2016 were quantified. This was first done for the Tavua site as it was one of the severely flooding-impacted areas in the western division of Fiji [58].

Our analysis identified four flood events using the criterion $SWRI_{24-hr-S} > 0$ to indicate a flood situation (Figure 5) for the Tavua site. Of four flood events, the two significant events were predominantly caused by heavy rain in the past 24 h. The first major flood event started on 3 April at 5 p.m. and ended on 4 April at 4 p.m. with a total duration of 23 h, a volume of 20.37, and a peak of 2.36. Approximately 154 mm of rainfall was recorded during this event. The second major flood event started on the 6th of April at 3 a.m. and lasted for 14 h. This flood event recorded a total volume of 8.77, with a peak of 1.03, while approximately 69.50 mm of rainfall was recorded for this event. The combined volume of all four flood events for the Tavua site between 3 April 2016 and 6 April 2016 (96 h) was 30.59.

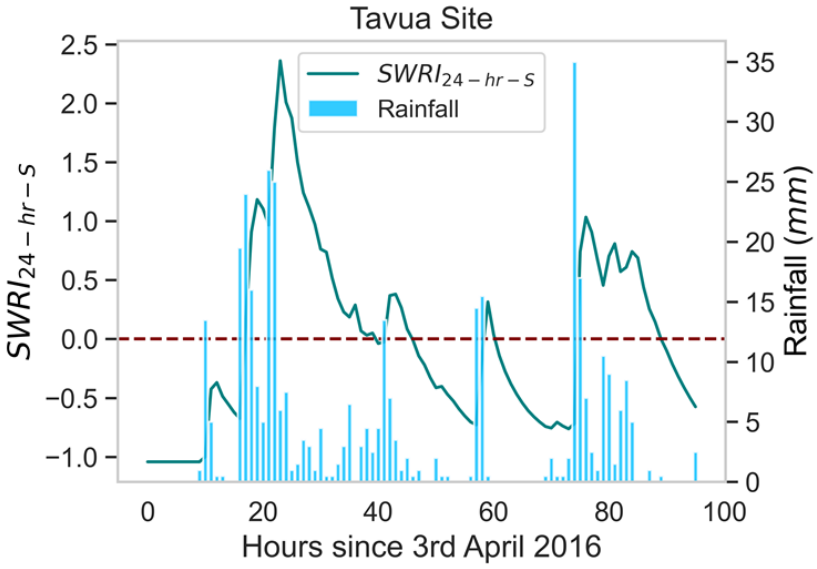

Figure 5. The $SWRI_{24-hr-S}$ applied to identify the flood events in April 2016 at the Tavua site (96 h).

The flood events for the same period were also determined for the other six sites. The flood characteristics, i.e., D, V, and Q of the floods, varied among these sites, as shown in Figure 6. The results showed that areas in the western division of Fiji were severely affected by flooding, as was reported by FMS [58] [Tavua ($V \approx 30.59$), Lautoka ($V \approx 25.63$), Sigatoka ($V \approx 19.45$), Nadi ($V \approx 9.12$), and Rakiraki ($V \approx 8.79$)] compared to the areas in the central division [Navua ($V \approx 0.12$) (minor flood event), and Nasinu (no floods)]. These results demonstrate the utility of $SWRI_{24-hr-S}$ in identifying flood situations and determining their characteristics. Consequently, the proposed $SWRI_{24-hr-S}$ can be considered a feasible and cost-effective tool to monitor the flash flood risk in Fiji. The variation in flood characteristics among our study sites demonstrates the importance of flood risk assessments for each site separately, despite the proximity of these sites, as also highlighted in previous work [6]. Figure 7 depicts the occurrence of floods, encompassing minor events with minimal volume and potentially insignificant impacts at each of the seven study sites over 2014–2018. Over this five-year study period, a slight fluctuation in flood frequency was observed across the study sites, as illustrated in Figure 7. Notably, the Tavua study site

exhibited the highest frequency, while the Nasinu study site recorded the lowest frequency of flood events.

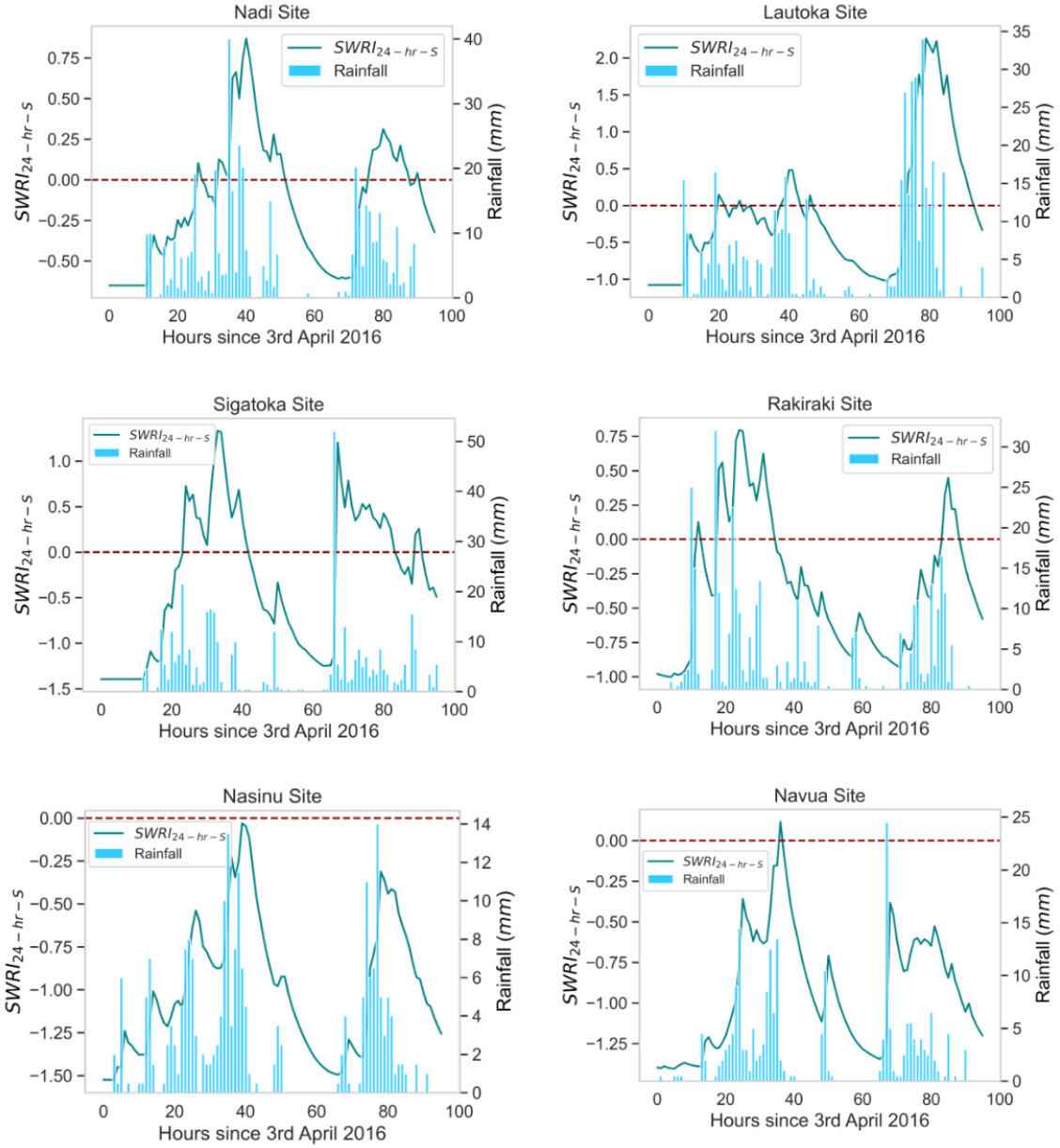

Figure 6. The $SWRI_{24-hr-S}$ applied to identify the flood events in April 2016 at the other study sites (96 h).

The occurrence of frequent severe weather events, such as tropical cyclones and depressions, results in significant flood events in Fiji, and this usually occurs during the wet season (November–April) and occasionally in the dry season (May–October), especially in La Niña years [33]. This is evident in Figure 8, which indicates the wet season, including May and October, experiencing high rainfall, leading to a higher frequency of flood events and greater flood volume (Figure 9) compared to the other months. This emphasises the need for Fiji's National Disaster Management Office (NDMO) and other relevant stakeholders to implement comprehensive flood mitigation and resilience measures. Public education on flood safety and preparedness for the wet season is also crucial, particularly for those residing in flood-prone areas.

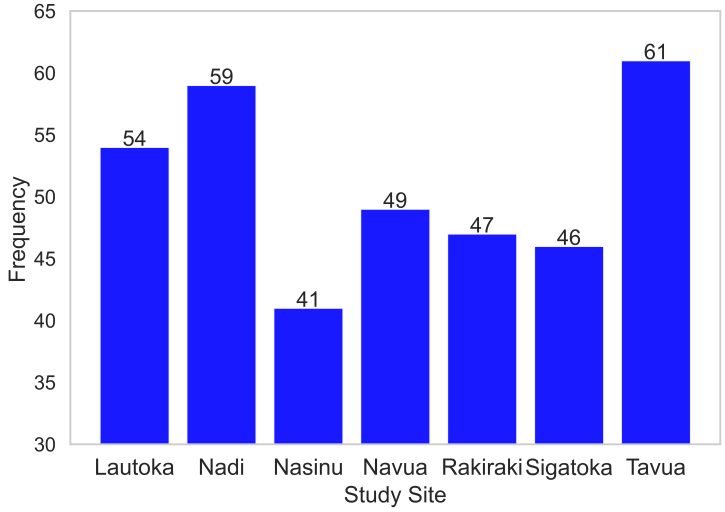

Figure 7. Geographic analysis of flood frequency between 2014 and 2018.

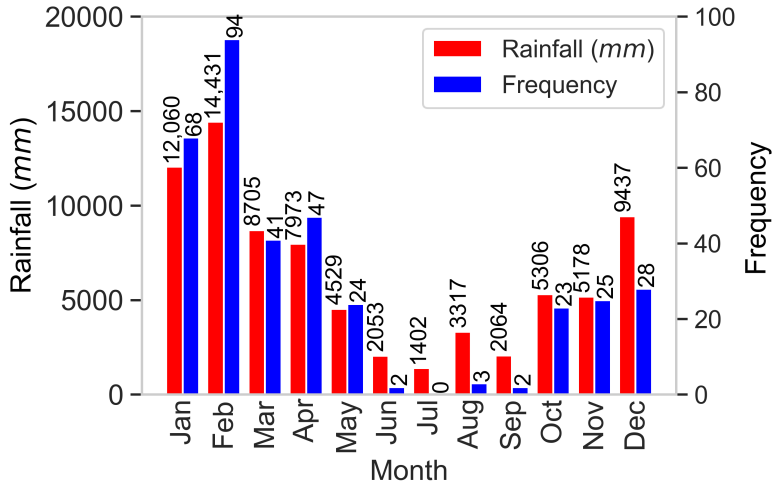

Figure 8. Temporal (monthly) analysis of flood frequency and total monthly rainfall aggregated from 2014 to 2018.

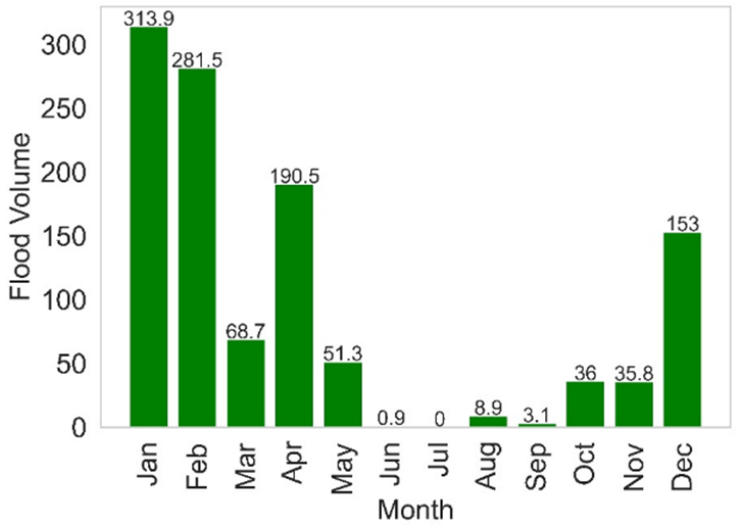

Figure 9. Temporal (monthly) analysis of the combined volume of flood events across 7 study sites from 2014 to 2018.

The annual rainfall and the occurrence of flood events across seven sites from 2014 to 2018 are illustrated in Figure 10. This figure shows that the year 2015 had the lowest rainfall among all years examined. According to FMS [59], the rainfall trends in 2015 exhibited a typical El Niño pattern, characterised by below-average rainfall at most of the study stations. Consequently, there were fewer flood events (Figure 10) and a lower flood volume (Figure 11) in 2015 compared to the other years in the present study.

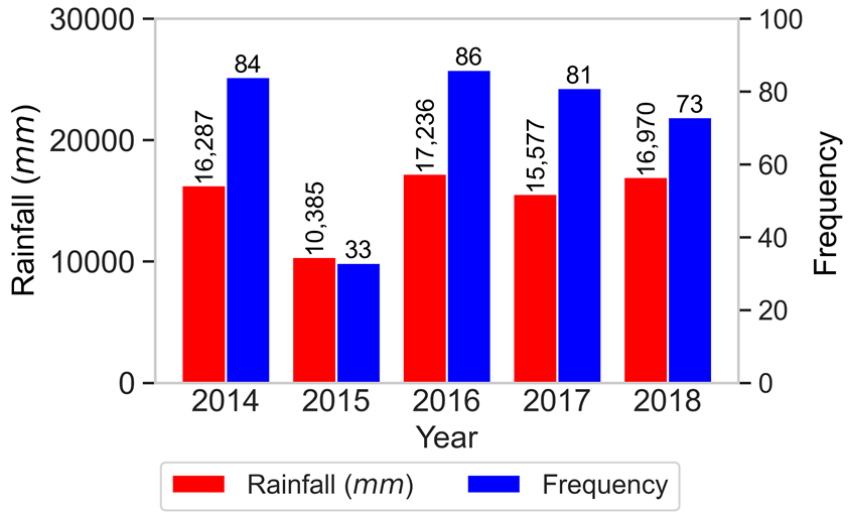

Figure 10. Frequency of floods and total rainfall across 7 study sites aggregated from 2014 to 2018.

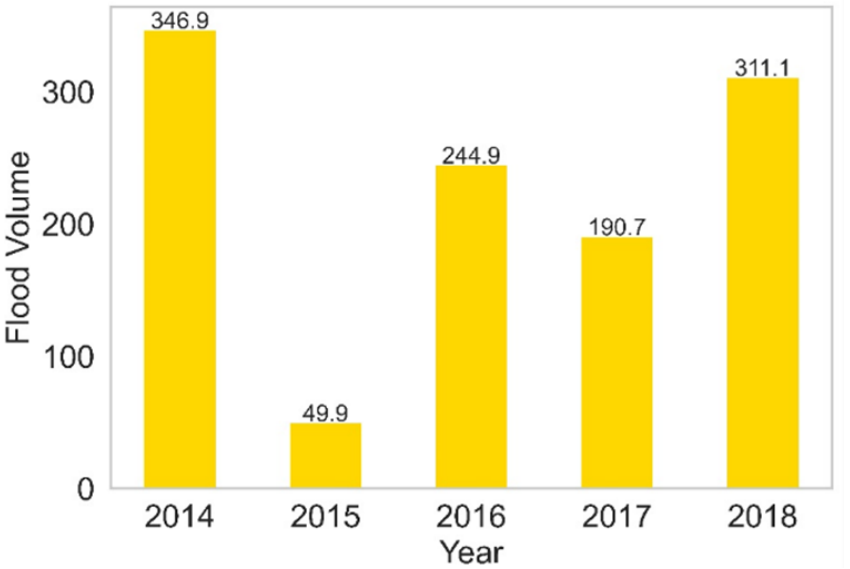

Figure 11. Yearly combined volume of flood events across 7 study sites from 2014 to 2018.

Table 2 lists the five severe floods at each of the seven sites during the study period. The flood severity was determined by ranking the flood events at each site based on their volume, with 1 indicating the most severe and 5 indicating the least severe. For each of the seven sites, the table displays the onset time, duration, volume, peak, total $WRI_{24-hr-S}$, total rain, and maximum $WRI_{24-hr-S}$ for each flood event. Statistics such as these may aid relevant organisations in understanding past flood events at these sites, which will facilitate future flood mitigation decisions to minimise the severity of floods at these locations.

A brief analysis of the most severe flood event at each study site was performed and validated using the annual climate summaries published by the Fiji Meteorological Services to ensure that the $SWRI_{24-hr-S}$ accurately identified them. The analysis of floods in Nadi (Table 2a) showed that the most severe flood started on 29 January 2014 at 8 a.m. and

recorded a volume of 157.28. This flood lasted 49 h and reached a peak of 6.80, making it the most severe flood event with the most prolonged duration among all the study sites during the 5-year study period. During this flood event, about 1590 mm of rainfall was recorded. This flood event was described in Figure 2 and validated using Fiji's annual climate summary for 2014 [37].

Table 2b shows that the most severe flood event in Lautoka recorded a total volume of 25.05. This flood event started on 14 January 2018 at 2 p.m. and lasted 19 h, reaching a peak of 3.29. During this flood event, about 175 mm of rainfall was recorded. This flood event was verified using Fiji's climate summaries 2018, which stated that heavy rainfall occurred from 13 to 15 January 2018 due to an active trough of low pressure, resulting in widespread flooding [60,61].

Table 2. Analysis of the 5 most severe flood events for different study sites in Fiji from 2014 to 2018. (**a**) Nadi, (**b**) Lautoka, (**c**) Nasinu, (**d**) Navua, (**e**) Rakiraki, (**f**) Sigatoka, and (**g**) Tavua.

Study Site		Onset Time (t_{onset})	Volume (V)	Duration (D) (hrs)	Peak (Q)	Total $WRI_{24-hr-S}$	Total Rain (mm)	Maximum $WRI_{24-hr-S}$
a	Nadi							
	1	29 January 2014 at 8 a.m.	157.28	49	6.80	10,557.06	1590	416.05
	2	8 February 2017 at 4 a.m.	11.88	18	1.31	1316.32	195.40	109.56
	3	4 April 2016 at 8 a.m.	6.86	20	0.87	1108.52	161.40	84.87
	4	7 January 2014 at 6 p.m.	6.31	10	1.73	715.09	84	133.10
	5	15 January 2014 at 6 p.m.	5.77	13	1.18	794.02	84	102.03
b	Lautoka							
	1	14 January 2018 at 2 p.m.	25.05	19	3.29	1315.15	175	126.10
	2	6 April 2016 at 2 a.m.	23.95	19	2.27	1283.34	180	96.59
	3	1 April 2014 at 7 a.m.	18.39	13	3.49	935.98	109	131.92
	4	6 February 2017 at 3 p.m.	11.48	20	1.53	954.55	131.50	75.40
	5	8 February 2017 at 9 a.m.	10.85	13	1.49	718.16	98	74.18
c	Nasinu							
	1	27 February 2014 at 9 a.m.	24.90	18	2.83	1388.83	173	115.53
	2	21 February 2015 at 6 p.m.	23.15	16	2.48	1261.37	163.50	106.16
	3	6 December 2014 at 4 a.m.	19.15	16	2.92	1155.37	140	117.79
	4	11 November 2018 at 11 p.m.	7.46	8	1.91	521.64	37	91.16
	5	28 May 2018 at 8 a.m.	7.20	7	2.10	474.23	21	96.15
d	Navua							
	1	15 December 2016 at 6 a.m.	56.98	28	4.29	2732.99	392.50	161.35
	2	17 March 2017 at 4 a.m.	16.37	14	2.10	1023.68	114.50	99.48
	3	16 January 2014 at 1 a.m.	9.82	14	1.16	838.28	123.50	72.87
	4	2 May 2016 at 6 a.m.	8.15	13	1.38	751.01	110	79.03
	5	27 February 2014 at 3 p.m.	7.98	10	1.61	626.25	83.50	85.60
e	Rakiraki							
	1	19 December 2016 at 5 a.m.	33.99	21	4.28	1783.89	265.50	170.39
	2	14 January 2018 at 2 p.m.	19.89	17	2	1199.35	156.50	97.02
	3	20 February 2016 at 8 p.m.	16.89	17	2.68	1103.01	112.50	119.05
	4	17 December 2016 at 4 p.m.	11.28	19	1.25	989.33	145.50	73.09
	5	5 March 2017 at 9 a.m.	10.06	15	1.52	818.03	114.50	81.73
f	Sigatoka							
	1	30 January 2014 at 10 a.m.	23.32	16	2.84	999.93	121	92.83
	2	3 February 2018 at 3 p.m.	15	12	2.25	695.39	93	79.78
	3	1 May 2018 at 6 p.m.	11.10	14	1.67	671.12	65	67.25
	4	4 April 2016 at 12 a.m.	10.91	18	1.33	789.16	103	59.79
	5	1 April 2018 at 6 a.m.	9.74	11	1.55	549.66	73	64.62
g	Tavua							
	1	8 February 2017 at 10 a.m.	45.88	23	4.04	1612.79	238	117.34
	2	3 April 2016 at 5 p.m.	20.37	23	2.36	1023.74	154	78.55
	3	17 May 2014 at 10 a.m.	17.16	17	1.81	805.30	103.50	65.93
	4	6 February 2017 at 11 a.m.	14.77	18	1.69	774.08	89.50	63.11
	5	6 March 2017 at 1 p.m.	14.23	17	1.75	737.52	98.50	64.47

According to Table 2c, the most severe flood in Nasinu started on 27 February 2014 at 9 a.m. and lasted 18 h. This flood had a volume of 24.90 and reached a peak of 2.83.

Approximately 173 mm of rainfall was recorded during this flood event. As reported by FMS [37], the tropical depressions TD14F and TD15F caused heavy rainfall in Fiji's central and eastern divisions between 25 and 27 February 2014. As a result, parts of Fiji, particularly the major river systems in the central division (where this study site is located), experienced flooding during this period.

The most severe flood, both in Navua (Table 2d) and Rakiraki (Table 2e), occurred in December 2016. For Navua, this event started on 15 December 2016 at 6 a.m. and lasted for 28 h, during which it recorded a volume of 56.98 and reached a peak of 4.29. For Rakiraki, it started on the 19th of December at 5 a.m. and lasted for 21 h, during which it recorded a volume of 33.99 and reached a peak of 4.28. Approximately 392.50 mm and 265.50 mm of rainfall were recorded during this flood event for Navua and Rakiraki, respectively. The most severe flood event that occurred at the Navua and Rakiraki sites was validated using Fiji's climate summaries 2016, which stated that the trough of low pressure and active rain bands associated with the tropical depression TD04F resulted in heavy rainfall that caused severe flooding in some parts of the country's main island of Viti Levu (where these study sites are located) from 12 to 20 December 2016 [58].

Based on Table 2f, the most severe flood event in Sigatoka started on 30 January 2014 at 10 a.m. and lasted for 16 h. This flood had a volume of 23.32 and reached a peak of 2.84. Approximately 121 mm of rainfall was recorded during this flood event. As mentioned earlier, an active trough that caused widespread rain across Fiji from the 29 to the 30 January 2014 resulted in flooding, particularly in the western division of Fiji, where this site is located [37].

Lastly, the analysis of floods in Tavua (Table 2g) showed that the most severe flood started on 8 February 2017 at 10 a.m. and recorded a volume of 45.88. This flood lasted 23 h and reached a peak of 4.04. During this flood event, about 238 mm of rainfall was recorded. As per FMS [62], the tropical depression TD09F affected the country between 6 and 8 February 2017 and led to flooding in parts of the western division of Fiji, where this study site is located.

3.2. Application of the Vine Copula Model for Probabilistic Flood Risk Analysis

The frequency of flood events at each study site is demonstrated in Figure 7. Similarly, the flood characteristics, i.e., D, V, and Q, for each flood event were calculated for all study sites. Table 3 shows the five-number summary, including the mean, standard deviation, skewness, and kurtosis for each flood characteristic at each study site.

Moreover, as shown in Table 3, the minimum flood duration was 1 h at all study sites. The maximum flood duration, volume, and peak were recorded at the Nadi site (this flood event is described in Figure 2). The skewness and kurtosis information of each flood characteristic, which describes the shape and distribution of a dataset, were more than +1 and +3, respectively, for all study sites, indicating that their distribution is highly right-skewed. This means the flood characteristics dataset for all study sites contains extreme flood duration, volume, and peak values.

The results in Table 3 also show that flood characteristics exhibit high variability across all study sites in terms of their median and inter-quartile range (IQR). The median flood duration for all study sites was 3 h, while the median volume and peak ranged from approximately 0.52 to 0.89 and 0.24 to 0.42, respectively. The IQR for flood duration, volume, and peak varied from 4 to 7 h, 1.78 to 2.94, and 0.52 to 0.91, respectively. The high spatiotemporal variation in flood characteristics highlights the importance of modelling these characteristics simultaneously, and employing a robust model like the copulas used in this study is crucial for accurately capturing the dependence among these flood characteristics.

Table 3. Descriptive statistics of flood characteristics at each study site.

Flood Characteristic	Site	Min.	Lower Quartile (*Q*1)	Median (*Q*2)	Upper Quartile (*Q*3)	Max.	Mean	Standard Dev.	Skewness	Kurtosis
Duration (*D*) (hours)	Lautoka	1	1	3	6	20	4.796	4.791	1.771	5.511
	Nadi	1	1	3	7	49	5.424	7.135	4.286	24.981
	Rakiraki	1	1	3	8	21	5.681	5.801	1.248	3.243
	Tavua	1	2	3	7	23	5.525	5.749	1.590	4.492
	Sigatoka	1	1	3	5	18	4.413	4.578	1.746	4.848
	Navua	1	1	3	5	28	4.367	5.003	2.753	11.676
	Nasinu	1	2	3	6	18	4.415	4.153	1.962	6.236
Volume (*V*)	Lautoka	0.003	0.204	0.633	1.983	25	2.746	5.473	3.022	11.168
	Nadi	0.002	0.132	0.517	2.271	157.276	4.156	20.404	7.538	55.625
	Rakiraki	0.005	0.097	0.529	3.038	33.993	3.272	6.329	3.257	13.998
	Tavua	0.016	0.185	0.617	2.272	45.879	3.190	7.123	4.230	22.984
	Sigatoka	0.021	0.220	0.632	2.834	23.316	2.826	4.777	2.523	9.293
	Navua	0.005	0.167	0.557	2.058	56.983	3.012	8.495	5.656	34.803
	Nasinu	0.033	0.159	0.886	2.854	24.903	3.028	5.853	2.951	10.098
Peak (*Q*)	Lautoka	0.003	0.179	0.348	0.701	3.490	0.597	0.725	2.519	9.356
	Nadi	0.002	0.108	0.244	0.674	6.803	0.527	0.940	5.369	35.042
	Rakiraki	0.005	0.097	0.323	0.816	4.282	0.606	0.809	2.663	10.954
	Tavua	0.016	0.124	0.338	0.799	4.040	0.573	0.688	2.719	12.339
	Sigatoka	0.021	0.207	0.393	1.121	2.841	0.718	0.714	1.369	3.945
	Navua	0.005	0.167	0.401	0.723	4.289	0.601	0.747	2.924	13.398
	Nasinu	0.033	0.139	0.419	0.916	2.915	0.720	0.763	1.566	4.596

Conducting a comprehensive correlation analysis among flood characteristics and understanding the relationship between each characteristic pair is another crucial step in modelling their joint distribution [18]. In this study, both parametric measures—Pearson's correlation coefficient (r)—and nonparametric rank-based correlation measures—Spearman's correlation coefficient (ρ) and Kendall's tau (τ)—were employed to examine the relationship between each pair of flood characteristics. Additionally, Mutual Information (MI) was utilised to examine the degree of dependence between each pair of flood characteristics. The results obtained are presented in Table 4. The correlation coefficients, i.e., r, ρ, and τ, obtained between each pair of flood characteristics, were statistically significant at the 1% level of significance. Overall, a strong positive dependency is observed between each pair of flood characteristics across all study sites, as shown in Table 4. However, there would be cases where the peak value is extremely high while the duration is low or the peak is moderate while the duration is long, and the volumes would also be different. For example, as shown in Table 4, at the Tavua site, the linear correlation measured by Pearson's correlation coefficient between D and Q ($r_{D\&Q} \approx 0.859$) is higher than that between D and V ($r_{D\&Q} \approx 0.838$). However, the rank correlation measured by Spearman's and Kendall's tau correlation coefficients between D and Q ($\rho_{D\&Q} \approx 0.902, \tau_{D\&Q} \approx 0.698$) are lower than those between D and V ($\tau_{D\&V} \approx 0.946, \tau_{D\&V} \approx 0.836$). This implies complex and non-linear relationships among flood characteristics. Subsequently, the risks of different flood events are different. Therefore, this requires a robust model like copulas used in this study to capture the full dependence among flood characteristics.

Additionally, it must be noted that across all study sites, the strongest dependencies exist between $D - V$ and $V - Q$ compared to $D - Q$. This is particularly evident based on Kendall's tau (τ) coefficient utilised in this study to select the most optimal structure of the vine copula model at each study site. This implies that the flood volume, V, can be positioned between the other two flood characteristics (i.e., D and Q), as illustrated in Figure 4a, to model the joint distribution of D, V, and Q using the 3D D-vine copula.

Table 4. The statistical correlation in terms of the Pearson's correlation coefficient (r), Spearman's rank correlation coefficient (ρ), Kendall's tau (τ), and Mutual Information (MI) computed between the pairs of flood characteristics, i.e., Duration (D, in hours), Volume (V), and Peak (Q) for each study site.

Site	*D&V*				*D&Q*				*V&Q*			
	r	ρ	τ	MI	r	ρ	τ	MI	r	ρ	τ	MI
Lautoka	0.895	0.941	0.831	1.051	0.860	0.877	0.738	0.584	0.931	0.971	0.883	1.044
Nadi	0.863	0.950	0.849	0.966	0.929	0.891	0.740	0.619	0.924	0.978	0.890	0.771
Rakiraki	0.855	0.934	0.828	0.823	0.842	0.902	0.760	0.647	0.949	0.987	0.926	1.123
Sigatoka	0.895	0.929	0.820	0.851	0.810	0.869	0.721	0.616	0.887	0.979	0.888	1.129
Tavua	0.838	0.946	0.836	0.983	0.859	0.902	0.698	0.651	0.942	0.960	0.859	1.103
Navua	0.891	0.937	0.838	1.066	0.937	0.892	0.752	0.616	0.899	0.982	0.903	0.933
Nasinu	0.942	0.945	0.831	0.903	0.895	0.844	0.687	0.501	0.896	0.964	0.844	0.904

Table 5 shows the results obtained when the 3D D-vine copula is fitted to the flood characteristics data at each study site. As depicted in Table 5, the results confirm that the D-vine structure illustrated in Figure 4a with flood volume (V) as the conditioning variable is the most appropriate to model the joint distribution of flood characteristics (i.e., D, V, and Q) across all study sites. The table also shows the best-fitted bivariate copula function and its associated parameters at each tree level for each study site. For instance, at the Sigatoka site, in the first tree (Tree 1), the Frank (C_{DV}) and Survival Gumbel (C_{VQ}) copulas are selected between $D - V$ and $V - Q$, respectively. In the second tree (Tree 2), the Frank copula is chosen as the most parsimonious for identifying the bivariate copula ($C_{DQ|V}$).

Table 5. Overall summary of fitted 3D D-vine copula framework at each study site. Note: τ = Kendall's tau, logLik = Log-Likelihood, AIC = Akaike Information Criterion, and BIC = Bayesian Information Criterion.

Site	D-Vine Structure (Conditioning Variable)	Tree Level	Flood Characteristic Pairs	Best-Fitted Copula	Copula Dependence Parameter (s)	τ	logLik	AIC	BIC
Lautoka	D-V-Q (V is placed in the centre)	Tree 1	D-V	Gaussian	$\rho = 0.93$	0.77	102.58	−199.16	−193.19
			V-Q	Gaussian	$\rho = 0.94$	0.78			
		Tree 2	DQ | V	Gumbel	$\theta = 1.1$	0.08			
Nadi		Tree 1	D-V	Frank	$\theta = 15$	0.77	142.45	−278.89	−272.66
			V-Q	BB7	$\theta = 2.1$; $\delta = 11.6$	0.81			
		Tree 2	DQ | V	Independence	NA	0			
Nasinu		Tree 1	D-V	Gaussian	$\rho = 0.93$	0.75	72.98	−141.96	−138.53
			V-Q	Gaussian	$\rho = 0.92$	0.75			
		Tree 2	DQ | V	Independence	NA	0			
Navua		Tree 1	D-V	Frank	$\theta = 16$	0.78	111.65	−219.29	−215.51
			V-Q	Survival Gumbel (Rotated Gumbel 180 degrees)	$\theta = 6.4$	0.84			
		Tree 2	DQ | V	Independence	NA	0			
Rakiraki		Tree 1	D-V	Gaussian	$\rho = 0.93$	0.75	118.13	−230.25	−224.70
			V-Q	BB7	$\theta = 3$; $\delta = 15$	0.83			
		Tree 2	DQ | V	Independence	NA	0			
Sigatoka		Tree 1	D-V	Frank	$\theta = 20$	0.82	123.88	−241.76	−236.27
			V-Q	Survival Gumbel (Rotated Gumbel 180 degrees)	$\theta = 7$	0.86			
		Tree 2	DQ | V	Frank	$\theta = -3.6$	−0.36			
Tavua		Tree 1	D-V	Survival BB7 (Rotated BB7 180 degrees)	$\theta = 5.2$; $\delta = 3.6$	0.77	148.44	−288.87	−280.43
			V-Q	BB7	$\theta = 4.3$; $\delta = 13.4$	0.82			
		Tree 2	DQ | V	Independence	NA	0			

To derive the joint exceedance probability of the flood event characteristics (i.e., D, V, and Q) for different combination scenarios using the best-fitted 3D D-vine copula selected at each study site, we first quantify the probability that the flood duration, volume, and peak exceed specific thresholds simultaneously (Equation (7)). The thresholds were selected at the 50th quantile (median), 75th quantile (moderate), and 95th quantile (extreme). The quantile values of each flood characteristic were computed and subsequently averaged for all study sites, as presented in Table 6. As seen in Table 6, for example, the averaged 50th-quantile value of duration is $q_D(0.5) = 3$ h. Similarly, the averaged 75th-quantile value of duration is $q_D(0.75) = 6$ h, and the averaged 95th-quantile value of duration is $q_D(0.95) = 15$ h. As for the spatial pattern, Table 6 demonstrates a moderate variation in the quantile values of each flood characteristic across all study sites.

Table 6. The duration (D) (hours), volume (V), and peak (Q) at the 50th quantile (median), 75th quantile (moderate), and 95th quantile (extreme) for each study site.

Flood Characteristic	Study Site	50th Quantile	75th Quantile	95th Quantile
D (hours)	Lautoka	3	6	15
	Nadi	3	7	13
	Rakiraki	3	8	17
	Tavua	3	7	17
	Sigatoka	3	5	16
	Navua	3	5	14
	Nasinu	3	6	16
	Average	3	6	15
V	Lautoka	0.633	1.983	13.898
	Nadi	0.517	2.271	6.365
	Rakiraki	0.529	3.038	15.205
	Tavua	0.617	2.272	14.767
	Sigatoka	0.632	2.834	11.054
	Navua	0.557	2.058	9.149
	Nasinu	0.866	2.854	19.153
	Average	0.622	2.473	12.799
Q	Lautoka	0.348	0.701	1.840
	Nadi	0.244	0.674	1.365
	Rakiraki	0.323	0.816	1.976
	Tavua	0.338	0.799	1.750
	Sigatoka	0.393	1.121	2.305
	Navua	0.401	0.723	1.694
	Nasinu	0.419	0.916	2.477
	Average	0.352	0.821	1.915

By applying the vine copula probabilistic model, we show the joint exceedance probabilities of the duration, volume, and peak in different combination scenarios for each study site in Figures 12–14. From a flood risk analysis perspective, the present results clearly demonstrate a moderate yet notable difference in spatial patterns of the joint exceedance probability of flood event characteristics in different combination scenarios. As shown in Figure 12a, the probabilities of a flood event occurring where both the volume and peak exceed the 50th-quantile (median) values (i.e., $V \geq q$ (0.50) = 0.622 and $Q \geq q$ (0.50) = 0.352) and the duration (D) exceeds the median (i.e., $D \geq q$ (0.50) = 3 h), moderate (i.e., $D \geq q$ (0.75) = 6 h), and extreme (i.e., $D \geq q$ (0.95) = 15 h) values are approximately 50–59%, 23–39%, and 4–10% across all study sites, respectively.

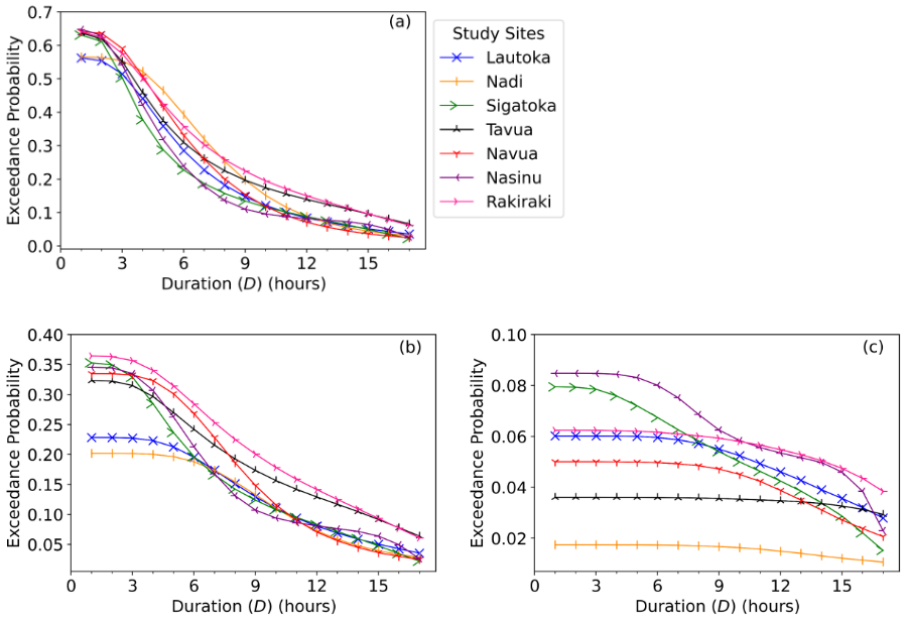

Figure 12. Flood risk assessment presented in terms of the joint exceedance probability of a flood event's duration (D) being greater than or equal to the 50th quantile (median) (i.e., $D \geq q$ (0.5) = 3 h), 75th quantile (moderate) (i.e., $D \geq q$ (0.75) = 6 h), and 95th quantile (extreme) (i.e., $D \geq q$ (0.95) = 15 h) combined with the following: (**a**) both the volume and peak being greater than or equal to the 50th quantile (i.e., $V \geq q$ (0.5) = 6.222 & $Q \geq q$ (0.5) = 0.352)), (**b**) the volume being greater than or equal to the 50th quantile and the peak being greater than or equal to the 75th quantile (i.e., $V \geq q$ (0.5) = 6.222 & $Q \geq q$ (0.75) = 0.821), and (**c**) the volume being greater than or equal to the 50th quantile and the peak being greater than or equal to the 95th quantile (i.e., $V \geq$ q (0.5) = 6.222 & $Q \geq q$ (0.95) = 1.915).

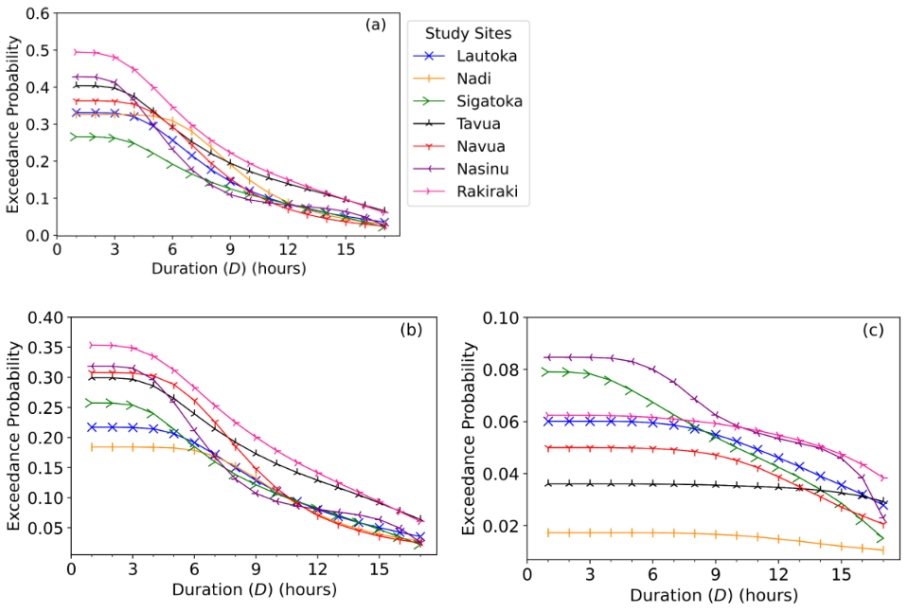

Figure 13. Flood risk assessment presented in terms of the joint exceedance probability of the flood duration (D) being greater than or equal to the 50th quantile (median) (i.e., $D \geq q$ (0.5) = 3 h), 75th quantile (moderate) (i.e., $D \geq q$ (0.75) = 6 h), and 95th quantile (extreme) (i.e., $D \geq q$ (0.95) = 15 h) combined with the following: (**a**) the volume being greater than or equal to the 75th quantile and the peak being greater than or equal to the 50th quantile (i.e., $V \geq q$ (0.75) = 2.473 & $Q \geq q$ (0.5) = 0.352), (**b**) both the volume and peak being greater than or equal to the 75th quantile (i.e., $V \geq q$ (0.75) = 2.473 & $Q \geq q$ (0.75) = 0.821), and (**c**) the volume being greater than or equal to the 75th quantile and the peak being greater than or equal to the 95th quantile (i.e., $V \geq q$ (0.75) = 2.473 & $Q \geq q$ (0.95) = 1.915).

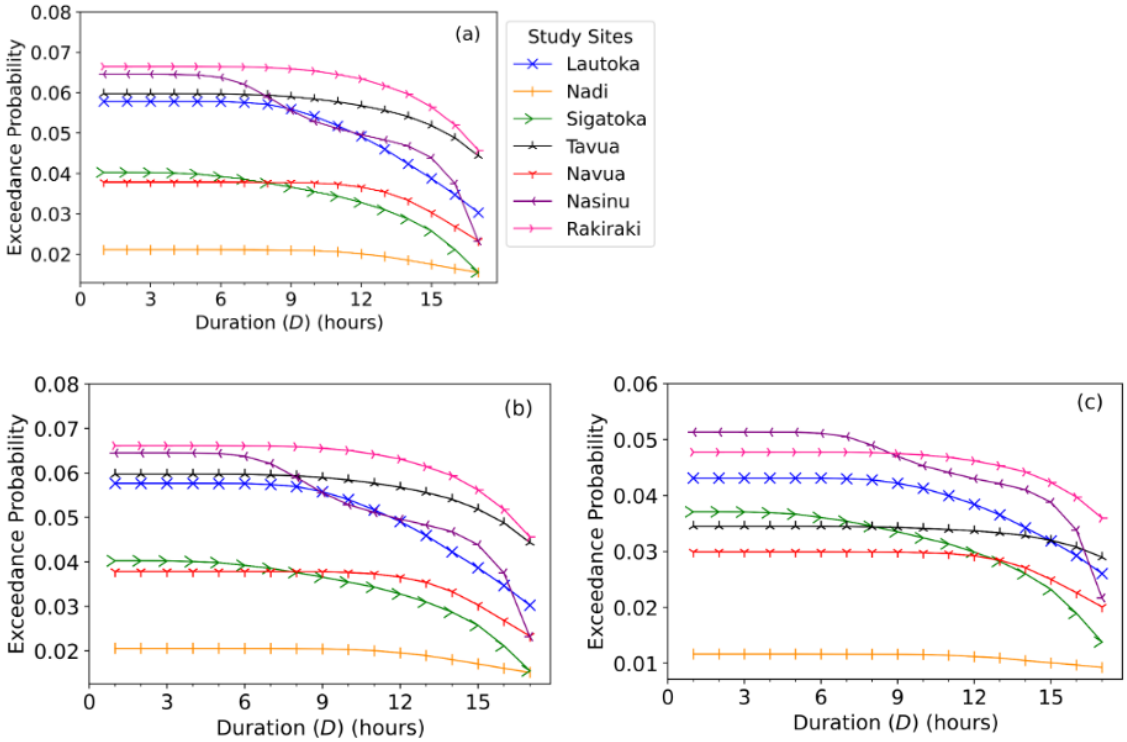

Figure 14. Flood risk assessment presented in terms of the joint exceedance probability of the flood duration (D) being greater than or equal to the 50th quantile (median) (i.e., $D \geq q$ (0.5) = 3 h), 75th quantile (moderate) (i.e., $D \geq q$ (0.75) = 6 h), and 95th quantile (extreme) (i.e., $D \geq q$ (0.95) = 15 h) combined with the following: (**a**) the volume being greater than or equal to the 95th quantile and the peak being greater than or equal to the 50th quantile (i.e., $V \geq q$ (0.95) = 12.799 & $Q \geq q$ (0.5) = 0.352), (**b**) the volume being greater than or equal to the 95th quantile and the peak being greater than or equal to the 75th quantile (i.e., $V \geq q$ (0.95) = 12.799 & $Q \geq q$ (0.75) = 0.821), and (**c**) both the volume and peak being greater than or equal to the 95th quantile (i.e., $V \geq q$ (0.95) = 12.799 & $Q \geq q$ (0.95) = 1.915).

A similar probabilistic flood risk analysis conducted with both the volume and peak exceeding the 75th-quantile (moderate) values (i.e., $V \geq q$ (0.75) = 2.473 and $Q \geq q$ (0.75) = 0.821) and the duration (D) exceeding the median (i.e., $D \geq q$ (0.50) = 3 h), moderate (i.e., $D \geq q$ (0.75) = 6 h), and extreme (i.e., $D \geq q$ (0.95) = 15 h) values showed that the probability of flood occurrence was approximately 18–35%, 18–28%, and 4–9% across all study sites, respectively (Figure 13b).

In general, the probability of a flood event with a volume exceeding the 50th-quantile (median) or 75th-quantile (moderate) values and both the peak and duration exceeding the 95th-quantile (extreme) value was less than 5% across all study sites. In the case when both the flood volume and duration exceeded the 95th-quantile (extreme) value, the probability of a flood event with the peak exceeding the 50th-quantile (median) or 75th-quantile (moderate) values was less than 6% across all study sites.

In the worst-case scenario, when the flood risk could be more severe, we found that the probability of a flood event occurring where the volume, peak, and duration exceeded the extreme values (i.e., $V \geq q$ (0.95) = 12.799, $Q \geq q$ (0.95) = 1.915, and $D \geq q$ (0.95) = 15 h) was less than 5% at all study sites (Figure 14c). These findings imply a moderate probability of a flood event characterised by median (i.e., 50th-quantile) duration, volume, and peak values across all study sites. The results also suggest that the likelihood of a flood event characterised by extreme duration, volume, and peak is exceptionally low across all study sites.

4. Conclusions, Limitations of the Study and Future Research Directions

This study has made novel contributions to flood risk monitoring and assessment by developing a mathematically convenient hourly flood index, $SWRI_{24-hr-S}$, and testing its practical use in identifying flood situations over 2014-2018 at seven different study sites in Fiji while jointly modelling flood characteristics such as flood duration, volume, and peak using a vine copula model for probabilistic flood risk assessment.

The results have unambiguously established the practical use of the newly proposed $SWRI_{24-hr-S}$ as a potent indicator to identify the flood situation at an hourly scale while computing the associated flood characteristics that were impossible with a 24-hourly water resources index used in the literature. The results also showed that Fiji mainly experienced high rainfall during the wet/cyclone season (November to April), including May and October. Consequently, the number of flood events was higher in these months than in the other months. This highlights the critical importance of implementing comprehensive and well-structured flood preparedness and risk mitigation strategies tailored explicitly for these months characterised by increased rainfall and flood events, thus ensuring the safety and security of communities and their properties. This study also presented the flood characteristics and water-intensive properties of five severe flood events for each study site. Relevant organisations, such as Fiji's NDMO, are expected to use these findings to understand the attributes of past flood events at these study sites. This, in turn, can support future decision-making on flood mitigation, ultimately reducing the severe impacts of such events.

The results also demonstrated a strong positive dependency between each pair of flood characteristics across all study sites. The D-vine structure with flood volume (V) as the conditioning variable was identified as the most appropriate for modelling the joint distribution of flood characteristics (i.e., D, V, and Q) across all study sites. Therefore, it was utilised to model the joint distributions among the extreme flood characteristics to extract their joint exceedance probability, providing crucial information for probabilistic flood risk assessment at each study site. The findings revealed moderate variations in the spatial patterns of joint exceedance probability of extreme flood event characteristics across different combination scenarios, underscoring exceptionally low probabilities of floods with extreme duration, volume, and peak at all study sites.

Despite the merits of the present study, a primary limitation of this research was the unavailability of rainfall data required for many of the flood-prone sites across Fiji. Consequently, this research was confined to selected sites within the western and central divisions of the country. As a result, this study could not perform a comparative analysis across all four divisions (i.e., the western, central, northern, and eastern divisions), which could have provided valuable insights into extreme flood risk areas in the nation. It is important to note that the Ba study site in the western division of Fiji, a frequent flooding zone, had to be excluded due to a high percentage of missing rainfall data. Therefore, in future research, our methodology can be improved with $SWRI_{24-hr-S}$ derived from satellite-based rainfall products covering a wider area, including major towns and cities, following the recent approach for Myanmar [18].

Another limitation of the present study was using a prior/fixed time-dependent reduction function with the weighting factor, W = 3.8, derived in an earlier study [24] to determine the contributions of accumulated rainfall in the latest 24 h. As discussed, the proposed $SWRI_{24-hr-S}$ is a normalised version of an existing $WRI_{24-hr-S}$ that used a suitable time-dependent reduction function to account for the depletion of water resources through various hydrological processes. It must be noted that the results of this study are sensitive to the value of W. A change in W will alter the contributions of accumulated rainfall in the latest 24 h in Equation (2). Consequently, both $WRI_{24-hr-S}$ and the proposed $SWRI_{24-hr-S}$ will change accordingly.

In the future, further studies can test the correctness of this weighting factor (W) more comprehensively for study sites where the topography may vary considerably. This could require a major correlation of this weighting factor against rainfall-runoff and other

physical models to capture more accurately the actual value of the decay of accumulated rainfall and its impacts on a flood event [15,24]. Several tests with hydrological parameters, including evapotranspiration rates, percolation, seepage, surface runoff, and drainage conditions, etc., may be required to ascertain the time-dependent reduction function for $SWRI_{24-hr-S}$. In regions with different decay rates of rainfall-accumulated water volume, it is crucial to incorporate them when formulating $SWRI_{24-hr-S}$. The proposed $SWRI_{24-hr-S}$ must also be verified for its broader adoption as an index-based risk monitoring tool. Therefore, its feasibility is expected to be demonstrated in other flood-prone regions globally in future studies, contingent on the availability of well-documented flood records for validation and hourly rainfall data. While doing so, a different technique to normalise the existing $WRI_{24-hr-S}$ may be selected, depending on how the normalised $WRI_{24-hr-S}$ index represents flood risk situations in those climatic conditions.

This study has undertaken a purely mathematical-based approach to monitoring flood risk, so in future studies, it is anticipated that the proposed $SWRI_{24-hr-S}$, in conjunction with additional data such as the catchment hydrology, drainage information, and river flows, will be utilised to develop hourly hydrographs for various sites. This approach will further cement the accuracy of flood characteristic estimation and the monitoring of flash flood events. There is also the potential to develop an innovative $SWRI_{24-hr-S}$-based forecasting system with sufficient lead time, presenting a novel approach for early flash flood warnings in Fiji and other regions.

A key advantage of $SWRI_{24-hr-S}$, as an hourly flood risk monitoring tool, is its simplistic mathematical formula that is easy to compute, analyse, and interpret for non-expert audiences compared to physical or hydrological models, including rainfall-runoff models for flood risk monitoring that involves complex development. However, in the future, especially in varied hydrological and topographic settings, it is crucial to comprehensively compare the proposed $SWRI_{24-hr-S}$ with other established flood monitoring systems, including the Flash Flood Guidance System (FFGS).

Despite these limitations, using $SWRI_{24-hr-S}$ has demonstrated acceptable accuracy in detecting flood situations on an hourly scale. Therefore, our proposed methodology can be considered a feasible and cost-effective tool for hourly flood risk monitoring in Fiji and perhaps other similar geographical locations. Applying the proposed probabilistic flood risk analysis using vine copulas can enhance the nation's overall flood risk assessment and mitigation strategies.

Author Contributions: Conceptualization, R.C., T.N.-H. and R.C.D.; methodology, R.C., T.N.-H. and R.C.D.; software, R.C. and T.N.-H.; validation, R.C., T.N.-H. and R.C.D.; formal analysis, R.C.; investigation, R.C.; resources, T.N.-H. and R.C.D.; data curation, R.C.; writing—original draft preparation, R.C.; writing—review and editing, R.C., T.N.-H., R.C.D., S.G., A.G. and M.A.; visualisation, R.C.; supervision, T.N.-H. and R.C.D; project administration, T.N.-H. and R.C.D.; funding acquisition, T.N.-H. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the Asia-Pacific Network for Global Change Research grant number CRRP2023-07MY-Nguyen Huy.

Data Availability Statement: The data presented in this study are available on request from the first author, excluding some data protected by copyright.

Acknowledgments: The first author is an Australia Awards Scholar supported by the Australian Government Department of Foreign Affairs and Trade (DFAT). We thank DFAT for funding this study via the Australia Awards Scholarship Scheme 2022. The authors are also grateful to Fiji Meteorological Services for providing the rainfall data needed for this study. Disclaimer: The views and opinions expressed in this paper belong to the authors and do not represent the views of the Australian Government or the Fiji Meteorological Services.

Conflicts of Interest: The authors declare no conflicts of interest.

Abbreviations

The following abbreviations are used in this manuscript:

AIC	Akaike Information Criterion
API	Antecedent Precipitation Index
AWRI	Available Water Resources Index
BIC	Bayesian Information Criterion
D	Flood Duration
FFGS	Flash Flood Guidance System
FJD	Fijian Dollar
FMS	Fiji Meteorological Services
GDP	Gross Domestic Product
I_F	Daily Flood Index
IQR	Interquartile Range
JCDF	Joint Cumulative Distribution Function
JPDF	Joint Density Distribution Function
logLik	Log-Likelihood
mBICv	Modified Vine Copula Bayesian Information Criteria
NDMO	Fiji's National Disaster Management Office
PDF	Probability Density Function
PSIDS	Pacific Small Island Developing State
Q	Flood Peak
r	Pearson's Correlation Coefficient
SAPI	Standardised Antecedent Precipitation Index
SPCZ	South Pacific Convergence Zone
SPI	Standardised Precipitation Index
SWAP	Standardised Weighted Average of Precipitation
$SWRI_{24-hr-S}$	Hourly Flood Index
USD	United States Dollar
V	Flood Volume
WAP	Weighted Average of Precipitation
$WRI_{24-hr-S}$	24-Hourly Water Resources Index
ρ	Spearman's Rank Correlation Coefficient
τ	Kendall's Tau Correlation Coefficient

Appendix A

Table A1. The bivariate copula families utilised to develop the 3D D-vine copula models in this study.

Copula Type	Bivariate Copula Family	Name
Parametric	Elliptical	Gaussian
		Student-t
	Archimedean	Frank
		Gumble
		Rotated Gumbel 90 degrees
		Rotated Gumbel 180 degrees (Survival Gumbel)
		Rotated Gumbel 270 degrees
		Clayton
		Rotated Clayton 90 degrees
		Rotated Clayton 180 degrees (Survival Clayton)
		Rotated Clayton 270 degrees
		Joe
		Rotated Joe 90 degrees
		Rotated Joe 180 degrees (Survival Joe)
		Rotated Joe 270 degrees

Table A1. *Cont.*

Copula Type	Bivariate Copula Family	Name
		Clayton-Gumbel (BB1) Rotated BB1 90 degrees Rotated BB1 180 degrees (Survival BB1) Rotated BB1 270 degrees Joe-Gumbel (BB6) Rotated BB6 90 degrees Rotated BB6 180 degrees (Survival BB6) Rotated BB6 270 degrees Joe- Clayton (BB7) Rotated BB7 90 degrees Rotated BB7 180 degrees (Survival BB7) Rotated BB7 270 degrees Joe-Frank (BB8) Rotated BB8 90 degrees Rotated BB8 180 degrees (Survival BB8) Rotated BB8 270 degrees
Non-parametric	-	Transformation kernel
-	-	Independence

References

1. Taherizadeh, M.; Khushemehr, J.H.; Niknam, A.; Nguyen-Huy, T.; Mezősi, G. Revealing the effect of an industrial flash flood on vegetation area: A case study of Khusheh Mehr in Maragheh-Bonab Plain, Iran. *Remote Sens. Appl. Soc. Environ.* **2023**, *32*, 101016. [CrossRef]
2. Taherizadeh, M.; Niknam, A.; Nguyen-Huy, T.; Mezősi, G.; Sarli, R. Flash flood-risk areas zoning using integration of decision-making trial and evaluation laboratory, GIS-based analytic network process and satellite-derived information. *Nat. Hazards* **2023**, *118*, 2309–2335. [CrossRef]
3. Associated Programme on Flood Management. Management of Flash Floods. Integrated Flood Management Tools Series No. 16, World Meteorological Organization. 2012. Available online: https://www.floodmanagement.info/publications/tools/APFM_Tool_16.pdf (accessed on 11 January 2024).
4. Shah, V.; Kirsch, K.R.; Cervantes, D.; Zane, D.F.; Haywood, T.; Horney, J.A. Flash flood swift water rescues, Texas, 2005–2014. *Clim. Risk Manag.* **2017**, *17*, 11–20. [CrossRef]
5. Dordevic, M.; Mutic, P.; Kim, H. Flash Flood Guidance System: Response to One of the Deadliest Hazards. WMO Bulletin 1, World Meteorological Organization. 2020. Available online: https://www.imgw.pl/sites/default/files/2020-04/wmo_klimat_woda_en-min.pdf (accessed on 11 January 2024).
6. Moishin, M.; Deo, R.C.; Prasad, R.; Raj, N.; Abdulla, S. Development of Flood Monitoring Index for daily flood risk evaluation: Case studies in Fiji. *Stoch. Environ. Res. Risk Assess.* **2021**, *35*, 1387–1402. [CrossRef]
7. Sharma, K.K.; Verdon-Kidd, D.C.; Magee, A.D. A decision tree approach to identify predictors of extreme rainfall events–a case study for the Fiji Islands. *Weather Clim. Extrem.* **2021**, *34*, 100405. [CrossRef]
8. Davis, J.; Henion, D.; Murro, M.J. *The Impacts of Coastal Flooding on Physical Infrastructure: Case Studies in Fiji, Kiribati, and Papua New Guinea*; Report; Center for Excellence in Disaster Management & Humanitarian Assistance: Ford Island, HI, USA , 2022.
9. Government of Fiji. Climate Vulnerability Assessment: Making Fiji Climate Resilient. Report, Government of Fiji. 2017. Available online: https://reliefweb.int/report/fiji/climate-vulnerability-assessment-making-fiji-climate-resilient (accessed on 11 January 2024).
10. The World Bank Group. Data. 2024. Available online: https://data.worldbank.org/country/fiji?view=chart (accessed on 11 January 2024).
11. Leonard, M.; Westra, S.; Phatak, A.; Lambert, M.; van den Hurk, B.; McInnes, K.; Risbey, J.; Schuster, S.; Jakob, D.; Stafford-Smith, M. A compound event framework for understanding extreme impacts. *Wiley Interdiscip. Rev. Clim. Change* **2014**, *5*, 113–128. [CrossRef]
12. Seneviratne, S.I.; Nicholls, N.; Easterling, D.; Goodess, C.M.; Kanae, S.; Kossin, J.; Luo, Y.; Marengo, J.; McInnes, K.; Rahimi, M.R.; et al. Changes in Climate Extremes and their Impacts on the Natural Physical Environment. In *Managing the Risks of Extreme Events and Disasters to Advance Climate Change Adaptation: Special Report of the Intergovernmental Panel on Climate Change*; Field, C.B., Barros, V., Stocker, T.F., Dahe, Q., Dokken, D.J., Ebi, K.L., Mastrandrea, M.D., Mach, K.J., Plattner, G.K., Allen, S.K., et al., Eds.; Cambridge University Press: Cambridge, UK; New York, NY, USA, 2012; pp. 109–230. [CrossRef]

13. Seiler, R.; Hayes, M.; Bressan, L. Using the standardized precipitation index for flood risk monitoring. *Int. J. Climatol. J. R. Meteorol. Soc.* **2002**, *22*, 1365–1376. [CrossRef]
14. Byun, H.R.; Lee, D.K. Defining three rainy seasons and the hydrological summer monsoon in Korea using available water resources index. *J. Meteorol. Soc. Jpn.* **2002**, *80*, 33–44. [CrossRef]
15. Lu, E. Determining the start, duration, and strength of flood and drought with daily precipitation: Rationale. *Geophys. Res. Lett.* **2009**, *36*, L12707. [CrossRef]
16. Lu, E.; Cai, W.; Jiang, Z.; Zhang, Q.; Zhang, C.; Higgins, R.W.; Halpert, M.S. The day-to-day monitoring of the 2011 severe drought in China. *Clim. Dyn.* **2014**, *43*, 1–9. [CrossRef]
17. Deo, R.C.; Byun, H.R.; Adamowski, J.F.; Kim, D.W. A real-time flood monitoring index based on daily effective precipitation and its application to Brisbane and Lockyer Valley flood events. *Water Resour. Manag.* **2015**, *29*, 4075–4093. [CrossRef]
18. Nguyen-Huy, T.; Kath, J.; Nagler, T.; Khaung, Y.; Aung, T.S.S.; Mushtaq, S.; Marcussen, T.; Stone, R. A satellite-based Standardized Antecedent Precipitation Index (SAPI) for mapping extreme rainfall risk in Myanmar. *Remote Sens. Appl. Soc. Environ.* **2022**, *26*, 100733. [CrossRef]
19. Nguyen-Huy, T.; Deo, R.C.; Yaseen, Z.M.; Prasad, R.; Mushtaq, S. Bayesian Markov chain Monte Carlo-based copulas: Factoring the role of large-scale climate indices in monthly flood prediction. In *Intelligent Data Analytics for Decision-Support Systems in Hazard Mitigation: Theory and Practice of Hazard Mitigation*; Deo, R.C., Kisi, O., Samui, P., Yaseen, Z.M., Eds.; Springer Nature: Singapore, 2021; pp. 29–47. [CrossRef]
20. Prasad, R.; Charan, D.; Joseph, L.; Nguyen-Huy, T.; Deo, R.C.; Singh, S. Daily flood forecasts with intelligent data analytic models: Multivariate empirical mode decomposition-based modeling methods. In *Intelligent Data Analytics for Decision-Support Systems in Hazard Mitigation: Theory and Practice of Hazard Mitigation*; Deo, R.C., Kisi, O., Samui, P., Yaseen, Z.M., Eds.; Springer Nature: Singapore, 2021; pp. 359–381. [CrossRef]
21. Nosrati, K.; Saravi, M.M.; Shahbazi, A. Investigation of flood event possibility over Iran using Flood Index. In *Survival and Sustainability: Environmental Concerns in the 21st Century*; Gökçekus, H., Türker, U., LaMoreaux, J.W., Eds.; Environmental Earth Sciences; Springer: Berlin/Heidelberg, Germany, 2011; pp. 1355–1361.
22. Deo, R.C.; Adamowski, J.F.; Begum, K.; Salcedo-Sanz, S.; Kim, D.W.; Dayal, K.S.; Byun, H.R. Quantifying flood events in Bangladesh with a daily-step flood monitoring index based on the concept of daily effective precipitation. *Theor. Appl. Climatol.* **2019**, *137*, 1201–1215. [CrossRef]
23. Ahmed, A.M.; Farheen, S.; Nguyen-Huy, T.; Raj, N.; Jui, S.J.J.; Farzana, S. Real-time prediction of the week-ahead flood index using hybrid deep learning algorithms with synoptic climate mode indices. *Res. Sq.* **2023**. [CrossRef]
24. Deo, R.C.; Byun, H.R.; Kim, G.B.; Adamowski, J.F. A real-time hourly water index for flood risk monitoring: Pilot studies in Brisbane, Australia, and Dobong Observatory, South Korea. *Environ. Monit. Assess.* **2018**, *190*, 1–27. [CrossRef]
25. Chebana, F.; Ouarda, T.B. Index flood–based multivariate regional frequency analysis. *Water Resour. Res.* **2009**, *45*, W10435. [CrossRef]
26. Sklar, M. Fonctions de répartition à n dimensions et leurs marges. *Ann. l'ISUP* **1959**, *8*, 229–231.
27. Daneshkhah, A.; Remesan, R.; Chatrabgoun, O.; Holman, I.P. Probabilistic modeling of flood characterizations with parametric and minimum information pair-copula model. *J. Hydrol.* **2016**, *540*, 469–487. [CrossRef]
28. Shafaei, M.; Fakheri-Fard, A.; Dinpashoh, Y.; Mirabbasi, R.; De Michele, C. Modeling flood event characteristics using D-vine structures. *Theor. Appl. Climatol.* **2017**, *130*, 713–724. [CrossRef]
29. Tosunoglu, F.; Gürbüz, F.; İspirli, M.N. Multivariate modeling of flood characteristics using Vine copulas. *Environ. Earth Sci.* **2020**, *79*, 1–21. [CrossRef]
30. Kuleshov, Y.; McGree, S.; Jones, D.; Charles, A.; Cottrill, A.; Prakash, B.; Atalifo, T.; Nihmei, S.; Seuseu, F.L.S.K. Extreme weather and climate events and their impacts on island countries in the Western Pacific: Cyclones, floods and droughts. *Atmos. Clim. Sci.* **2014**, *4*, 803. [CrossRef]
31. Feresi, J.; Kenny, G.J.; de Wet, N.; Limalevu, L.; Bhusan, J.; Ratukalou, I. *Climate Change Vulnerability and Adaptation Assessment for Fiji*; Technical Report; The International Global Change Institute (IGCI), University of Waikato: Hamilton, New Zealand, 2000.
32. Kumar, R.; Stephens, M.; Weir, T. Rainfall trends in Fiji. *Int. J. Climatol.* **2014**, *34*, 1501–1510. [CrossRef]
33. McGree, S.; Yeo, S.W.; Devi, S. *Flooding in the Fiji Islands between 1840 and 2009*; Risk Frontiers Technical Report, Risk Frontiers; Macquarie University: Macquarie Park, Australia, 2010.
34. Lo Presti, R.; Barca, E.; Passarella, G. A methodology for treating missing data applied to daily rainfall data in the Candelaro River Basin (Italy). *Environ. Monit. Assess.* **2010**, *160*, 1–22. [CrossRef]
35. Oriani, F.; Stisen, S.; Demirel, M.C.; Mariethoz, G. Missing data imputation for multisite rainfall networks: A comparison between geostatistical interpolation and pattern-based estimation on different terrain types. *J. Hydrometeorol.* **2020**, *21*, 2325–2341. [CrossRef]
36. Yevjevich, V.M. Objective Approach to Definitions and Investigations of Continental Hydrologic Droughts. Ph.D. Thesis, Colorado State University, Fort Collins, CO, USA, 1967.
37. FMS. Fiji Annual Climate Summary 2014. Report, Fiji Meteorological Service. 2015. Available online: https://www.met.gov.fj/aifs_prods/ACS-2014.pdf (accessed on 11 January 2024).
38. Nguyen-Huy, T.; Deo, R.C.; Mushtaq, S.; Kath, J.; Khan, S. Copula-based agricultural conditional value-at-risk modelling for geographical diversifications in wheat farming portfolio management. *Weather Clim. Extrem.* **2018**, *21*, 76–89. [CrossRef]

39. Nguyen-Huy, T.; Deo, R.C.; Khan, S.; Devi, A.; Adeyinka, A.A.; Apan, A.A.; Yaseen, Z.M. Student performance predictions for advanced engineering mathematics course with new multivariate copula models. *IEEE Access* **2022**, *10*, 45112–45136. [CrossRef]
40. Nguyen-Huy, T.; Deo, R.C.; Mushtaq, S.; Kath, J.; Khan, S. Copula statistical models for analyzing stochastic dependencies of systemic drought risk and potential adaptation strategies. *Stoch. Environ. Res. Risk Assess.* **2019**, *33*, 779–799. [CrossRef]
41. Nguyen-Huy, T.; Kath, J.; Mushtaq, S.; Cobon, D.; Stone, G.; Stone, R. Integrating El Niño-Southern Oscillation information and spatial diversification to minimize risk and maximize profit for Australian grazing enterprises. *Agron. Sustain. Dev.* **2020**, *40*, 4. [CrossRef]
42. Ni, L.; Wang, D.; Wu, J.; Wang, Y.; Tao, Y.; Zhang, J.; Liu, J.; Xie, F. Vine copula selection using mutual information for hydrological dependence modeling. *Environ. Res.* **2020**, *186*, 109604. [CrossRef]
43. Nguyen-Huy, T.; Deo, R.C.; Mushtaq, S.; Khan, S. Probabilistic seasonal rainfall forecasts using semiparametric d-vine copula-based quantile regression. In *Handbook of Probabilistic Models*; Samui, P., Chakraborty, S., Bui, D.T., Deo, R.C., Eds.; Elsevier: Oxford, UK, 2020; pp. 203–227. [CrossRef]
44. Cheng, Y.; Du, J.; Ji, H. Multivariate joint probability function of earthquake ground motion prediction equations based on vine copula approach. *Math. Probl. Eng.* **2020**, *2020*, 1697352. [CrossRef]
45. Lü, T.J.; Tang, X.S.; Li, D.Q.; Qi, X.H. Modeling multivariate distribution of multiple soil parameters using vine copula model. *Comput. Geotech.* **2020**, *118*, 103340. [CrossRef]
46. Latif, S.; Mustafa, F. Parametric vine copula construction for flood analysis for Kelantan river basin in Malaysia. *Civ. Eng. J.* **2020**, *6*, 1470–1491. [CrossRef]
47. Latif, S.; Simonovic, S.P. Parametric Vine copula framework in the trivariate probability analysis of compound flooding events. *Water* **2022**, *14*, 2214. [CrossRef]
48. Latif, S.; Simonovic, S.P. Trivariate Joint Distribution Modelling of Compound Events Using the Nonparametric D-Vine Copula Developed Based on a Bernstein and Beta Kernel Copula Density Framework. *Hydrology* **2022**, *9*, 221. [CrossRef]
49. Czado, C.; Bax, K.; Sahin, Ö.; Nagler, T.; Min, A.; Paterlini, S. Vine copula based dependence modeling in sustainable finance. *J. Financ. Data Sci.* **2022**, *8*, 309–330. [CrossRef]
50. Jeong, H.; Dey, D. Application of a vine copula for multi-line insurance reserving. *Risks* **2020**, *8*, 111. [CrossRef]
51. Allen, D.E.; McAleer, M.; Singh, A.K. Risk measurement and risk modelling using applications of vine copulas. *Sustainability* **2017**, *9*, 1762. [CrossRef]
52. Joe, H. *Multivariate Models and Multivariate Dependence Concepts*; Chapman & Hall/CRC: Boca Raton, FL, USA, 1997.
53. Nagler, T.; Krüger, D.; Min, A. Stationary vine copula models for multivariate time series. *J. Econom.* **2022**, *227*, 305–324. [CrossRef]
54. Nagler, T.; Vatter, T. Rvinecopulib: High Performance Algorithms for Vine Copula Modeling. R Package Version 3, CRAN. 2023. Available online: https://cran.r-project.org/web/packages/rvinecopulib/rvinecopulib.pdf (accessed on 11 January 2024).
55. Nagler, T.; Czado, C. Evading the curse of dimensionality in nonparametric density estimation with simplified vine copulas. *J. Multivar. Anal.* **2016**, *151*, 69–89. [CrossRef]
56. Nagler, T.; Schellhase, C.; Czado, C. Nonparametric estimation of simplified vine copula models: Comparison of methods. *Depend. Model.* **2017**, *5*, 99–120. [CrossRef]
57. Nagler, T.; Bumann, C.; Czado, C. Model selection in sparse high-dimensional vine copula models with an application to portfolio risk. *J. Multivar. Anal.* **2019**, *172*, 180–192. [CrossRef]
58. FMS. Fiji Annual Climate Summary 2016. Report, Fiji Meteorological Service. 2017. Available online: https://www.met.gov.fj/Summary2.pdf (accessed on 11 January 2024).
59. FMS. Fiji Annual Climate Summary 2015. Report, Fiji Meteorological Service. 2016. Available online: https://www.met.gov.fj/aifs_prods/ACS-2015.pdf (accessed on 11 January 2024).
60. FMS. Fiji's Climate in 2018. Report, Fiji Meteorological Service. 2019. Available online: https://www.met.gov.fj/aifs_prods/Climate_Products/December (accessed on 11 January 2024).
61. FMS. Fiji Climate Summary January 2018. Report, Fiji Meteorological Service. 2018. Available online: https://www.met.gov.fj/aifs_prods/FSCJAN18.pdf (accessed on 11 January 2024).
62. FMS. Fiji Annual Climate Summary 2017. Report, Fiji Meteorological Service. 2018. Available online: https://www.met.gov.fj/aifs_prods/Climate_Products/2017annualSum2020.01.28 (accessed on 11 January 2024).

4.3. Links and implications

Developing innovative and cost-effective flood monitoring and assessment tools is crucial for flood-prone regions lacking advanced risk monitoring resources and relevant datasets. In this chapter (Paper 1), a novel hourly flood monitoring tool, i.e., $SWRI_{24\text{-}hr\text{-}S}$, is proposed, which is a normalised metric derived from normalising the existing $WRI_{24\text{-}hr\text{-}S}$ (Deo et al., 2018) in the literature. The $SWRI_{24\text{-}hr\text{-}S}$ can serve as a universal index for objectively assessing flood risk across geographically diverse study sites. The feasibility of the proposed $SWRI_{24\text{-}hr\text{-}S}$ to identify flood events on an hourly scale and derive their associated characteristics, i.e., flood volume (V), duration (D), and peak (Q), is demonstrated for seven flood-prone sites in Fiji.

Furthermore, the vine copula is employed to model the joint distribution of D, V, and Q at each study site to extract their joint exceedance probability under various combination scenarios for probabilistic flood risk assessment. The joint exceedance probability between D, V, and Q, derived under various combination scenarios, provides insights into the spatial pattern of flood risk across the study sites. Consequently, integrating the $SWRI_{24\text{-}hr\text{-}S}$ as an hourly flood monitoring tool with the copula-statistical method to more accurately assess flood risk across the study sites is beneficial for enhancing flood warning systems. This, in turn, will aid relevant authorities in accurately assessing flood risk, thereby improving preparedness for flood events and devising effective plans and strategies to mitigate their severe impacts on communities.

Although the approaches proposed in this chapter (Paper 1) are practical and can be adopted by other flood-prone regions around the globe to enhance flood risk monitoring and assessment, the main limitation could be the required rainfall data, especially for data-scarce regions. It is imperative to note that the proposed $SWRI_{24\text{-}hr\text{-}S}$, in its formulation, only requires one externally sourced data, i.e., hourly rainfall. For Fiji, most flood-prone sites could not be included in the research due to the unavailability of the required rainfall data. In addition, the Ba site, a high flood-risk area, had to be excluded due to the high rate of missing data. Therefore, further studies are suggested to use satellite-based rainfall products to address this issue while covering wider locations, following the recent approach for Myanmar for better flood risk monitoring and assessment.

Another limitation of this study is using a predetermined time-reduction weighting factor ($W \approx 3.8$) established in prior research (Deo et al., 2018). As mentioned

earlier, the proposed $SWRI_{24\text{-}hr\text{-}S}$ is a normalised metric derived from normalising the existing $WRI_{24\text{-}hr\text{-}S}$ that employs a suitable time-dependent reduction function incorporating a weighting factor to account for the depletion of water resources through various hydrological processes. However, future research can assess the appropriateness of this weighting factor ($W \approx 3.8$) in areas characterised by significant variations in topography and climatic conditions, aiming for a more accurate representation of the decay of accumulated rainfall.

Moreover, this research focused on only three flood event characteristics (i.e., D, V, and Q) derived from the proposed $SWRI_{24\text{-}hr\text{-}S}$ to develop the vine copula model for probabilistic flood risk assessment. However, future studies can include additional flood event characteristics, such as peak time, annual maximum 24-hour rainfall and highest storm surges and river discharge (Latif & Simonovic, 2022a, 2022b; Shafaei et al., 2017). While the proposed $SWRI_{24\text{-}hr\text{-}S}$ is a practical tool for assessing flood risk on an hourly scale, as demonstrated in Fiji's case studies, it cannot predict the flooded state in advance unless a forecasting model is developed and tested. Therefore, in its second objective, this research focused on developing a hybrid DL model to forecast the $SWRI_{24\text{-}hr\text{-}S}$ to assess future flood risk. The next chapter will elaborate on the research findings of this second objective.

CHAPTER 5: PAPER 2 – HYBRID CONVOLUTIONAL NEURAL NETWORK FUSED WITH GATED RECURRENT UNIT-BASED HOURLY FLOOD INDEX FORECASTING FRAMEWORK: CASE STUDIES FOR FIJI ISLANDS

5.1. Introduction

This chapter is an exact copy of the article submitted for publication in the *Stochastic Environmental Research and Risk Assessment* Journal and is under review (Scopus Impact Factor 7.1).

This chapter designs a hybrid C-GRU model to forecast the $SWRI_{24\text{-}hr\text{-}S}$ over a short-term (i.e., 1-hourly forecast horizon) to assess future flood risk for five flood-prone study sites in Fiji. The study utilises only two datasets, i.e., the $SWRI_{24\text{-}hr\text{-}S}$ and hourly rainfall for each study site for model development. The PACF and CCF are utilised to determine the most statistically significant lags for the $SWRI_{24\text{-}hr\text{-}S}$ and hourly rainfall time series, respectively. Three DL models (i.e., CNN, GRU, and LSTM) and one ML model (i.e. RFR) are developed for performance comparison. An advanced BO with the TPE algorithm is employed to fine-tune the hyperparameters of all developed models. For all study sites, the forecasting performance of the proposed hybrid C-GRU model is meticulously evaluated against the four benchmark models using various performance evaluation metrics and visual diagnostics plots. The Diebold–Mariano (DM) statistical test is additionally used to ascertain whether the proposed hybrid C-GRU model's performance is statistically significantly better than that of the benchmark models.

The comprehensive performance evaluation shows the proposed hybrid C-GRU model's superior forecasting capabilities. The integration of CNN for feature extraction with GRU for temporal modelling in the hybrid C-GRU model effectively enhances the accuracy of hourly $SWRI_{24\text{-}hr\text{-}S}$ forecasts across five study sites in Fiji. Hence, the effectiveness of this newly proposed hybrid C-GRU-based $SWRI_{24\text{-}hr\text{-}S}$ forecasting framework underscores its potential use in decision support systems for early flood warning and risk evaluation in Fiji. It can also be adopted by other flood-prone regions to enhance their disaster preparedness and risk mitigation strategies.

5.2. Published paper

Hybrid Convolutional Neural Network Fused with Gated Recurrent Unit-based Hourly Flood Index Forecasting Framework: Case Studies for Fiji Islands

Ravinesh Chand[1], Ravinesh C Deo[1*], Sujan Ghimire[1], Thong Nguyen-Huy[2], Mumtaz Ali[3]

[1]School of Mathematics, Physics and Computing, University of Southern Queensland, Springfield, 4300, QLD, Australia.

[2]Centre for Applied Climate Sciences, University of Southern Queensland, Toowoomba, 4350, QLD, Australia.

[3]UniSQ College, University of Southern Queensland, Springfield, 4300, QLD, Australia.

*Corresponding author(s). E-mail(s): ravinesh.deo@unisq.edu.au;
Contributing authors: u1151334@umail.unisq.edu.au;cravinesh13@gmail.com;
sujan.ghimire@unisq.edu.au; thong.nguyen-huy@unisq.edu.au; mumtaz.ali@unisq.edu.au;

Abstract

Developing flood forecasting techniques at short timescales can improve an early warning system, mitigate severe flood risk and facilitate effective emergency response strategies in vulnerable regions. In this study, we introduce an hourly flood index ($\boldsymbol{SWRI_{24-hr-S}}$) derived by normalising the water resources index ($\boldsymbol{WRI_{24-hr-S}}$) to monitor flood risk. We develop a hybrid deep learning algorithm, C-GRU, by integrating Convolutional Neural Networks (CNN) with Gated Recurrent Unit (GRU) model and evaluate its effectiveness in forecasting hourly $\boldsymbol{SWRI_{24-hr-S}}$ in five flood-prone sites in Fiji. The model incorporates statistically significant lagged $\boldsymbol{SWRI_{24-hr-S}}$ with real-time hourly rainfall, and comparative analysis is performed against benchmark models: CNN, GRU, Long Short-Term Memory (LSTM) and Random Forest Regression (RFR). The results demonstrate that the proposed hybrid C-GRU model outperforms all the other models in accurately forecasting $\boldsymbol{SWRI_{24-hr-S}}$ over a 1-hourly forecast horizon. Across all of the study sites, the proposed model consistently generates the highest $\boldsymbol{r}$ ($\mathbf{0.996-0.999}$) and the lowest $\boldsymbol{RMSE}$ ($\mathbf{0.007-0.014}$) and $\boldsymbol{MAE}$ ($\mathbf{0.003-0.004}$) in the testing phase. The proposed hybrid C-GRU model also achieves the highest Global Performance Index ($\boldsymbol{GPI}$) values and the largest percentage of forecast errors ($\boldsymbol{FE}$) ($\approx$ 98.9-99.9%) within smaller error brackets (i.e., $|\boldsymbol{FE}| < \mathbf{0.05}$) across all study sites. Using the methodologies developed, we show the practical application of the proposed framework as a decision support system for early flood warning, demonstrating its potential to enhance real-time monitoring and early warning systems with broader application to flood-prone regions.

Keywords: Floods, deep learning, hourly flood forecasting, flood index, flood risk mitigation

1 Introduction

Flooding, a global crisis, is a devastating natural disaster affecting numerous regions worldwide. Flood occurs when excessive water overflows onto land which is usually dry, often due to heavy rainfall, snow melt, storm surges, dam release, or water overflow from natural watercourses (e.g., rivers). Floods can cause widespread devastation, including significant economic loss, loss of life, and substantial damage to public and personal property, agriculture, and the environment (Nguyen-Huy et al., 2021). According to the Centre for Research on the Epidemiology of Disasters (CRED), flooding was the predominant natural disaster, representing 43% of all incidents, impacting roughly 2.5 billion individuals and resulting in 160,000 fatalities from 1994 to 2013 (CRED, 2015). The estimated economic loss from flooding during this period (1994-2013) amounted to 636 billion USD (CRED, 2015).

The consequences of a flood disaster are particularly devastating in developing countries like Fiji (Moishin et al., 2021b), where this study is focused. Such countries do not have the advanced infrastructure for monitoring flood events, and most, if not all, evaluations of flood risk are carried out using accumulated rainfall over days or weeks. However, the exact measurement of a flash flood event due to a sudden downpour is somewhat unrealistic due to the estimated probability of flooding and the lack of an objective method for risk evaluation. Hence, developing a real-time flood monitoring and forecasting system that uses rainfall data with a time-dependent reduction function can offer a new promise for risk management in developing nations. This remains a crucial area of research driven by an urgent need to enhance early warning systems to mitigate the devastating consequences of flooding in Fiji and other small Pacific Island nations.

Fiji is a Pacific Small Island Developing State (PSIDS), with most of its population and infrastructure situated on large floodplains susceptible to long-duration flooding or in small catchments prone to flash flooding (Government of Fiji, 2017). The estimated average annual flood losses exceed 400 million FJD, equivalent to 4.2 percent of Fiji's Gross Domestic Product (GDP) (Government of Fiji, 2017). Between 1970 and 2016, Fiji experienced 44 major flood events, which impacted approximately 563,310 people and resulted in 103 fatalities (Government of Fiji, 2017). With the anticipated substantial increases in rainfall intensity due to climate change, flood-related asset losses in Fiji are projected to escalate, potentially exceeding 5 percent of the GDP by 2050 (Government of Fiji, 2017). This projection underscores the long-term economic threat posed by flooding in Fiji. Therefore, developing reliable methods for accurate flood forecasting and risk assessment is crucial to mitigate the severe impacts of flooding in Fiji.

Flood forecasting is essential in flood warning systems and remains among the most critical tasks in hydrology (Prasad et al., 2021). Flood forecasting is more beneficial when done in near real-time (e.g., on an hourly scale) with sufficient lead time as it allows for better estimation of flood risk and implementation of appropriate flood mitigation plans, evacuation, and rehabilitation measures (Alexander et al., 2018; Hapuarachchi et al., 2011; Kant et al., 2013; Tiwari and Chatterjee, 2010). Hydrodynamic models are the most widely used tool for simulating detailed flood dynamics (Teng et al., 2017). They can be directly integrated with hydrological and river models to facilitate flood risk assessment, real-time flood forecasting, and scenario analysis (Teng et al., 2017). However, studies have shown that these models are challenging to apply in operational flood forecasting, have high data requirements, and are highly computationally expensive (Kabir et al., 2020; Nevo et al., 2022; Teng et al., 2017).

In many developing nations with limited flood monitoring resources, hydrometeorological datasets, and risk monitoring facilities, a mathematically derived flood index based solely on rainfall data offers a valuable means to evaluate an impending flood risk situation. For instance, the flood index (I_F) is one of the most robust flood monitoring indices widely applied in various places globally, including Australia (Deo et al., 2015), Iran (Nosrati et al., 2011), Bangladesh (Deo et al., 2019; Ahmed et al., 2023), Myanmar (Nguyen-Huy et al., 2022), and Fiji (Moishin et al., 2021b), to monitor flood events on a daily scale. Despite its benefits, one primary limitation of I_F is its reliance on daily rainfall data, which spans a much longer timeframe than necessary for a near real-time flood risk monitoring system. Hence, flood indices based on shorter-term rainfall data (e.g., hourly) can be more practical for real-time assessment of flood situations.

In their pilot study, Deo et al. (2018) proposed the 24-hourly water resources index ($WRI_{24-hr-S}$) as a near real-time flood risk monitoring tool. This index was applied at two study locations, Australia and South Korea, demonstrating its potential for continuous flash flood risk monitoring during sustained extreme rainfall. The $WRI_{24-hr-S}$ monitors flood risk by considering the contribution of accumulated rainfall in the past 24 hours, whereby the rainfall contribution from the preceding hours is subjected to the time-dependent reduction function that accounts for the depletion of water resources through various hydrological processes such as evaporation, percolation, seepage, runoff, and drainage (Deo et al., 2018). However, unlike I_F, which is a normalised index, the $WRI_{24-hr-S}$ in its current form cannot effectively be used to identify a flood risk as it is unnormalised.

In this study, we propose a novel hourly flood index ($SWRI_{24-hr-S}$) by normalising the $WRI_{24-hr-S}$, originally proposed by Deo et al. (2018), and test and validate its practical utility to enable an objective assessment of flood risk at an hourly scale. However, it is essential to emphasise that the $SWRI_{24-hr-S}$ cannot be used to predict the flooded state in advance unless a forecasting model for this index is developed and thoroughly tested, which is the primary goal of this paper for Fiji's case studies. Accurate and reliable forecasting of the $SWRI_{24-hr-S}$ is crucial for assessing flood risk on an hourly scale, thereby enhancing decision-support systems for early flood warnings and enabling more effective flood risk management and mitigation strategies in Fiji.

It is imperative to note that although index-based flood forecasting has developed rapidly in recent years, no research has developed a forecast model for $SWRI_{24-hr-S}$ and explored its implications for hourly flood forecasting in a region like Fiji. Notably, Artificial Intelligence (AI)-based, i.e., Machine Learning (ML), Deep Learning (DL) and hybrid models have been developed to forecast I_F for daily flood forecasts. For instance, Prasad et al. (2021) in their study proposed the hybrid ML by combining the multivariate empirical mode decomposition (MEMD) technique with the M5 tree model to forecast daily I_F for the flood-prone Lockyer Valley region of Queensland, Australia.

The present study is also inspired by an earlier study of Moishin et al. (2021a) that has developed a hybrid DL algorithm, ConvLSTM, by integrating the predictive capabilities of Convolutional Neural Network (CNN) and Long Short-Term Memory (LSTM) models to forecast I_F across multiple forecast horizons. The hybrid DL model's performance was compared to benchmark models, including CNN-LSTM, LSTM, and Support Vector Regression (SVR). All the models developed were trained using statistically significant lagged values of I_F and real-time daily precipitation data. The study's results demonstrated the feasibility of the ConvLSTM-based I_F forecasting model in determining the possibility of flood situations in Fiji on a daily scale. Similarly, Ahmed et al. (2021) also proposed a hybrid DL model that integrated a CNN with a bi-directional long-short term memory (BiLSTM) to forecast I_F a week ahead for thirty-four selected stations in Bangladesh. The results of this study also demonstrated the superior forecasting performance of the hybrid DL, CNN-BiLSTM model, compared to the benchmark models, including SVR and BiLSTM.

While ML models are generally more interpretable and computationally efficient, requiring relatively less training time than DL models, the latter can automatically learn and extract crucial features from raw data without explicit feature engineering (Sarker, 2021). This autonomous feature extraction capability enables DL models to capture complex patterns and dependencies in the data, often resulting in superior performance when provided with sufficient training data (Sarker, 2021). Recurrent Neural Networks (RNNs) are a type of DL model known for their capability to capture sequential dependencies facilitated by their internal memory. However, RNNs often suffer from short-term memory caused by vanishing and exploding gradient problems, which impede their capacity to learn long-term dependencies in the data.

To address this issue, more advanced RNN variants have been developed, such as LSTM, initially introduced by Hochreiter and Schmidhuber (1997) in 1997 and further improved by Graves (2013) in 2013, and the Gated Recurrent Unit (GRU), introduced by Cho et al. (2014) in 2014. The LSTM and GRU models share similar architectures, except that the GRU model features a simplified gating mechanism compared to the LSTM. Specifically, while the LSTM incorporates three gates (i.e., input, forget, and output), the GRU utilises only two (i.e., reset and update). Consequently, compared to LSTMs, GRUs have simpler architecture and fewer trainable parameters, often leading to faster training times while effectively capturing long-term dependencies in sequential data (Kisvari et al., 2021; Li, 2023; Sharma et al., 2022; Wang et al., 2020; Zhang et al., 2022).

Considering the advantages of the GRU over an LSTM model, the present study proposes a hybrid DL algorithm that integrates CNN with GRU algorithms (C-GRU, hereafter) to forecast $SWRI_{24-hr-S}$ over a 1-hourly forecast. This integration is robust, leveraging the CNN algorithm's

capability to extract crucial temporal features from sequential data through convolutional operations (Ghimire et al., 2019; Joseph et al., 2024), and the GRU algorithm's proficiency in learning long-term dependencies and effectively modelling sequential data. The proposed hybrid C-GRU model has been successfully applied in various other studies, demonstrating its efficacy in areas such as water level prediction (Pan et al., 2020), river flooding forecasting and anomaly detection (Miau and Hung, 2020), short-term residential load forecasting (Sajjad et al., 2020), soil moisture prediction (Yu et al., 2021), short-term wind power forecasting (Zhao et al., 2023), wind speed prediction (Ji et al., 2022), PM_{10} forecasting (Sharma et al., 2022), evapotranspiration forecasting (Ahmed et al., 2021), classification of dust sources (Gholami and Mohammadifar, 2022), fault diagnosis for chiller system (Wang et al., 2020), prediction of heart disease (Almulihi et al., 2022), Dysarthria speech detection (Shih et al., 2022), and human activity recognition (Dua et al., 2023). However, as mentioned earlier, no prior study has developed a hybrid C-GRU model and tested its capability to forecast $SWRI_{24-hr-S}$ for hourly flood predictions in any region, including Fiji.

The novelty and scientific contributions of this study are as follows:

(a) To design and evaluate the performance of the hybrid C-GRU model that can forecast $SWRI_{24-hr-S}$ representing flood risk over a short-term, i.e., 1-hourly forecast horizon.

(b) To train the proposed hybrid C-GRU model on statistically significant lagged $SWRI_{24-hr-S}$ with real-time hourly rainfall data following a similar methodology to Moishin et al. (2021a) in such a way that only the rainfall data are required to build the model and provide a realistic assessment of flood risks.

(c) To enhance the performance of the proposed hybrid C-GRU model by adopting the Bayesian Optimization (BO) algorithm for efficient hyperparameter selection.

(d) To fully ascertain the proposed hybrid C-GRU model's performance against competing benchmark models: CNN, GRU, LSTM, and Random Forest Regression (RFR) using a diverse range of performance evaluation metrics and visual analysis of forecasted and observed (or actual) $SWRI_{24-hr-S}$ values.

This study, therefore, presents a practical framework using the hybrid C-GRU model, which is tested at diverse sites in the Fiji Islands, with an opportunity for its potential use in decision support systems for early flood warning and risk evaluation. The outcomes of this study are expected to contribute significantly towards disaster risk reduction and mitigation strategies by enhancing or strengthening Fiji's real-time monitoring and early warning systems, thus improving disaster preparedness, mitigation, and response efforts. Overall, the results show that the proposed hybrid C-GRU model outperforms all benchmarked models to accurately forecast $SWRI_{24-hr-S}$ over a 1-hourly forecast horizon. Therefore, the methodologies proposed could also be explored in other flood-prone regions around the globe.

2 Theoretical Overview

This section provides an overview of the proposed hybrid C-GRU (objective model) network developed to forecast $SWRI_{24-hr-S}$. This section also provides theoretical details of CNN and GRU algorithms used for model benchmarking. The theoretical details of other benchmark models, i.e., LSTM (Chung et al., 2014; Ghimire et al., 2022, 2019; Hochreiter and Schmidhuber, 1997; Nguyen et al., 2020; Wang et al., 2020) and RFR (Breiman, 2001; Chen et al., 2017; Ghimire et al., 2023; Liaw et al., 2002), are comprehensively elucidated elsewhere since they are well-known methodologies. We also provide an overview of the Bayesian Optimization algorithm used for hyperparameter tuning.

2.1 Convolutional Neural Network (CNN)

The one-dimensional convolutional neural network (Conv1D) was adopted to develop a hybrid C-GRU model for $SWRI_{24-hr-S}$ forecasting. CNN is a popular feedforward neural network originally

introduced by LeCun et al. (2015). It has two main features: weight sharing and local connections, which reduce the number of trainable parameters, thus reducing computations (Ghimire et al., 2022). Another advantage of CNN is its ability to automatically learn crucial spatial features without manual intervention (Ghimire et al., 2022). While CNN models are frequently employed for image recognition tasks, using the Conv1D model for prediction tasks involving time series data has emerged more recently (Ghimire et al., 2022). A notable advantage of the Conv1D model lies in its simple and compact design, which facilitates efficient and cost-effective implementation due to its operation with one-dimensional convolutions (Ghimire et al., 2022). A standard CNN typically comprises two interconnected layers: a feature extraction layer and a fully connected (FC) layer. The feature extraction layer, positioned after the input layer in the architecture, consists of multiple layers. Within the feature extraction layer, there are two types of layers: the convolutional layer and the pooling layer. Each convolutional layer employs several convolutional kernels to extract hidden features and generate a feature map. This feature map then undergoes a nonlinear activation function $f(.)$ to produce the output c_i of the i^{th} input as follows (Joseph et al., 2024):

$$c_i = f\left(w_i * x_i + b_i\right) \tag{1}$$

where $*$ represents the convolution operation, and w_i, x_i, and b_i are the weight matrix, input, and bias vector, respectively.

The output of the convolutional layer is then reduced by the pooling layer, also known as subsampling (Ghimire et al., 2022). The pooling layer in CNNs serves two main functions. Firstly, it reduces the spatial dimensions of feature maps, thereby reducing computational complexity and expediting training time (Zhao and Zhang, 2024). Secondly, it extracts crucial features while reducing redundant information, thus enhancing the model's generalisation capability and interpretability (Zhao and Zhang, 2024). The pooling layer partitions each feature map into fixed-size regions and performs operations such as selecting the maximum value (max pooling) or computing the average value (average pooling) within each region (Ghimire et al., 2022; Zhao and Zhang, 2024). In this study, we have used max pooling layers for pooling operations. The resulting reduced feature map is then flattened into a one-dimensional (1-D) array (using the flattening layer) and passed into one or more FC layers. The output of the last FC layer is usually fed into a softmax layer (for classification tasks) or a regression layer (for regression tasks) to produce the final output.

2.2 Gated Recurrent Unit (GRU)

GRU, introduced by Cho et al. (2014), is one of the variants of the RNN algorithm known for capturing long-term dependencies and effectively modelling sequential data. Unlike traditional RNNs, which suffer from the vanishing and exploding gradient problem during backpropagation through time, GRUs mitigate this problem through their simplified gating mechanisms that control the flow of information through the network (Chung et al., 2014; Li, 2023; Sharma et al., 2022; Zhang et al., 2022). The GRU unit features two gates regulating information flow: the reset and update gates. The update gate determines the extent to which the previous hidden layer state h_{t-1} is retained in the current hidden layer state h_t. The update gate first receives information from h_{t-1} and current input vector x_t and subsequently processes this information using an activation function σ. The update gate z_t can be expressed as (Zhang et al., 2022):

$$z_t = \sigma\left(W_{zx}x_t + W_{zh}h_{t-1} + b_z\right) \tag{2}$$

where W_{zx} and W_{zh} represent the learnable weight matrix for the update gate; b_z denotes the bias for the update gate.

The reset gate determines how much information from the previous time step is written into the candidate memory state $\widetilde{h}_t$. Like the update gate, the reset gate processes h_{t-1} and x_t using an activation function σ. The reset gate r_t can be expressed as (Zhang et al., 2022):

$$r_t = \sigma\left(W_{rx}x_t + W_{rh}h_{t-1} + b_r\right) \tag{3}$$

where W_{rx} and W_{rh} represent the learnable weight matrix for the reset gate; b_r denotes the bias in the reset gate.

Next, the reset gate r_t is combined with h_{t-1} and x_t to create a candidate memory state $\widetilde{h}_t$. The expression for $\widetilde{h}_t$ is as follows (Zhang et al., 2022):

$$\widetilde{h}_t = \tanh\left(\widetilde{W}_{hx}x_t + \widetilde{W}_{hh}\left[r_t * h_{t-1}\right] + b_{\widetilde{h}}\right) \tag{4}$$

where $\widetilde{W}_{hx}$ and $\widetilde{W}_{hh}$ represent the learnable weight matrix for the candidate memory state; $b_{\widetilde{h}}$ denotes the bias in the candidate memory state.

The current hidden layer state h_t is then derived by combining the previous hidden layer state h_{t-1} with the candidate memory state $\widetilde{h}_t$ The expression for h_t is as follows (Zhang et al., 2022):

$$h_t = (1 - z_t) * h_{t-1} + z_t * \widetilde{h}_t \tag{5}$$

2.3 Hybrid C-GRU Model

This study designs the proposed hybrid C-GRU model by combining CNN's robust feature extraction capability with GRU's powerful nonlinear time sequential predictive ability, enhancing its overall predictive performance for $SWRI_{24-hr-S}$ forecasting. Fig. 1 illustrates the topological structure of the proposed hybrid C-GRU model. As shown in Fig. 1, the hybrid C-GRU model comprises two Conv1D layers, two max-pooling layers, one flattening layer, and two GRU layers. The GRU layers have replaced the FC layer of CNN. In this configuration, crucial features extracted from the CNN are flattened into 1-D arrays and then fed into the GRU layers to incorporate these features for the low-latency forecasting of the $SWRI_{24-hr-S}$.

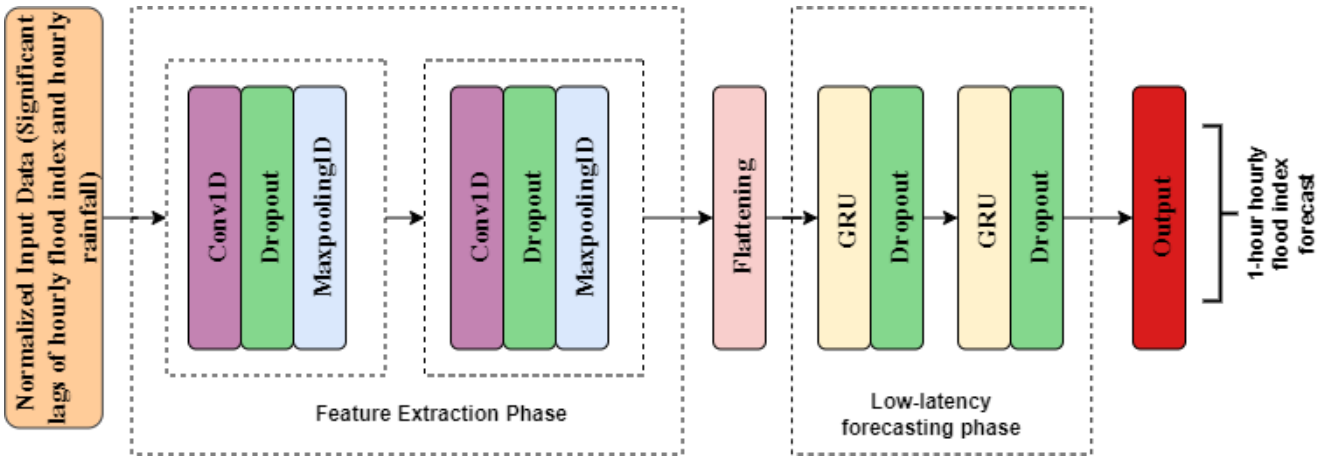

Fig. 1: The topological structure of the proposed hybrid C-GRU model for near real-time hourly forecasting of $SWRI_{24-hr-S}$ for flood risk evaluation.

2.4 Hyperparameter Tuning Using Bayesian Optimization (BO) Algorithm

Tuning hyperparameters to enhance prediction accuracy is challenging and time-consuming (Ghimire et al., 2022). The performance of numerous ML and DL algorithms depends heavily on the values assigned to these hyperparameters, making it essential to employ an effective method of configuring them (Eggensperger et al., 2013; Nguyen et al., 2020). Bayesian optimisation (BO) methods have been shown to outperform traditional optimisation approaches like grid search and random search (Bergstra and Bengio, 2012; Nguyen et al., 2020; Eggensperger et al., 2013).

The primary advantage of BO, which sets it apart from traditional optimisation methods, is its ability to achieve optimal hyperparameter configuration with fewer function evaluations (Nguyen

et al., 2020). This efficiency emanates from its unique capability to model the distribution of hyperparameter configurations and their corresponding fitness scores from previous iterations (Nguyen et al., 2020). By learning from this distribution, BO strategically selects the most promising configurations to evaluate next, thus accelerating the search for optimal hyperparameters. Consequently, BO thoroughly evaluates the most promising candidates for hyperparameter choices by probabilistically guiding and reducing the number of samples drawn from the hyperparameter search space (Nguyen et al., 2020). This process prioritises evaluating promising candidates, effectively guiding the optimisation process towards regions of the hyperparameter space expected to yield better performance. Conversely, in the grid and random search, each evaluation within their iterations is independent of prior iterations, leading to unavoidable assessments of hyperparameter search space regions with unsatisfactory performance, ultimately resulting in high computational costs (Prasad et al., 2023).

In this study, we utilise a variant of BO, known as Tree-structured Parzen Estimator (TPE) (Bergstra et al., 2011; Komer et al., 2014), to automatically optimise the hyperparameters of both the proposed hybrid C-GRU and benchmarking models. TPE has been introduced recently to overcome the limitations of conventional BO methods when dealing with categorical and conditional parameters, aiming to enhance the hyperparameter selection process (Bergstra et al., 2011; Komer et al., 2014; Nguyen et al., 2020). In the TPE algorithm, every sample from the empirical data defines a Gaussian distribution characterised by a mean equivalent to the hyperparameter value and a specified standard deviation (Nguyen et al., 2020). To initiate the optimisation iterations, the TPE algorithm employs a random search to initialise the distributions by sampling the response surface $\left\{\theta^{(i)}, y^{(i)}, i = 1, \ldots, N_{init}\right\}$ where θ represents the hyperparameter set, y is the corresponding value on the response surface (i.e., the validation loss or the fitness value), and N_{init} is the number of start-up iterations (Nguyen et al., 2020). Next, the hyperparameter space is partitioned into two groups, namely *good* and *bad* samples, determined by their fitness values and a predefined threshold value y^*, as follows (Nguyen et al., 2020):

$$p\left(\theta \,|y\right) = \begin{cases} \Pr_{good}\left(\theta\right) \text{if} y < y^* \\ \Pr_{bad}\left(\theta\right) \text{if} y \geq y^* \end{cases} \tag{6}$$

where $\Pr_{good}$ and $\Pr_{bad}$ are the probabilities that the hyperparameter set θ is in the good and bad groups, respectively.

This approach ensures that the selection of optimal hyperparameters depends not solely on the best observation but on a collection of the best observations and their distributions. Following this, the algorithm computes an expected improvement (EI) as follows (Nguyen et al., 2020):

$$EI\left(\theta\right) = \frac{\Pr_{good}\left(\theta\right)}{\Pr_{bad}\left(\theta\right)} \tag{7}$$

Finally, the hyperparameter configuration θ^* that maximises the EI at each iteration is selected. For more information, readers can refer to Feurer and Hutter (2019). This study implemented the TPE algorithm using the 'Hyperopt' package, an open-source Python library for hyperparameter optimisation using BO, developed by Komer et al. (2014).

3 Materials and Method

3.1 Study Area and Dataset

The proposed hybrid C-GRU model for near real-time $SWRI_{24-hr-S}$ forecasting is applied to geographically diverse sites in Fiji. Fiji is an archipelago of 332 islands with two main islands (Viti Levu and Vanua Levu) in the South Pacific Ocean. It is part of the continent of Oceania. The nation experiences two distinct seasons: a warm, wet period from November to April and a cooler, drier season from May to October. This seasonal variation is mainly attributed to the South Pacific Convergence Zone (SPCZ), the primary rainfall-producing system for the region, which typically lies over Fiji during the wet season (Feresi et al., 2000; Kumar et al., 2014). River

flooding is common during almost every wet season in Fiji and occasionally during the dry season, particularly during La Niña events (McGree et al., 2010). The current study focuses on five sites within the western division of Fiji, which are prone to recurrent and severe flooding events. Fig. 2 shows the map of the study area and the corresponding study sites.

The rainfall data for Lautoka, Sigatoka, Rakiraki, Nadi, and Tavua sites from 1st January 2014 to 31st December 2018 (5 years) were successfully acquired from the Fiji Meteorological Services. During the data pre-processing phase, rainfall data for each site was aggregated to derive the hourly rainfall necessary for this study. If at least 66.67% of the data points (i.e., at least 4 out of 6 data points for a 10-minute interval and at least 8 out of 12 data points for a 5-minute interval) were available within a particular hour, they were summed to calculate the total rainfall for that hour; otherwise, the rainfall value for that hour was recorded as missing. This approach aimed to maximise data recovery. After data aggregation, all study sites were found to have less than 5% missing values. Following the methodology outlined in Oriani et al. (2020), the Iterative K nearest Neighbour (IKNN) technique was employed to fill in all missing data. Table 1 summarises the hourly rainfall datasets and geographic settings of the 5 study sites. As depicted in Table 1, the average hourly rainfall is spatially different. The maximum hourly rainfall of 260 mm was recorded for the Nadi site over the study period. The skewness and kurtosis of the hourly rainfall data, which describe the shape and distribution of the dataset, were found to be greater than +1 and +3, respectively, for all study sites. This indicates that their distribution is highly right-skewed. Such skewness is primarily attributed to the frequent presence of zero values within these datasets. Consequently, extreme hourly rainfall values significantly impact the distribution, resulting in a highly right-skewed distribution. To visualise, Fig. 3 illustrates the hourly rainfall trend over a 5-year period using data from the Lautoka site.

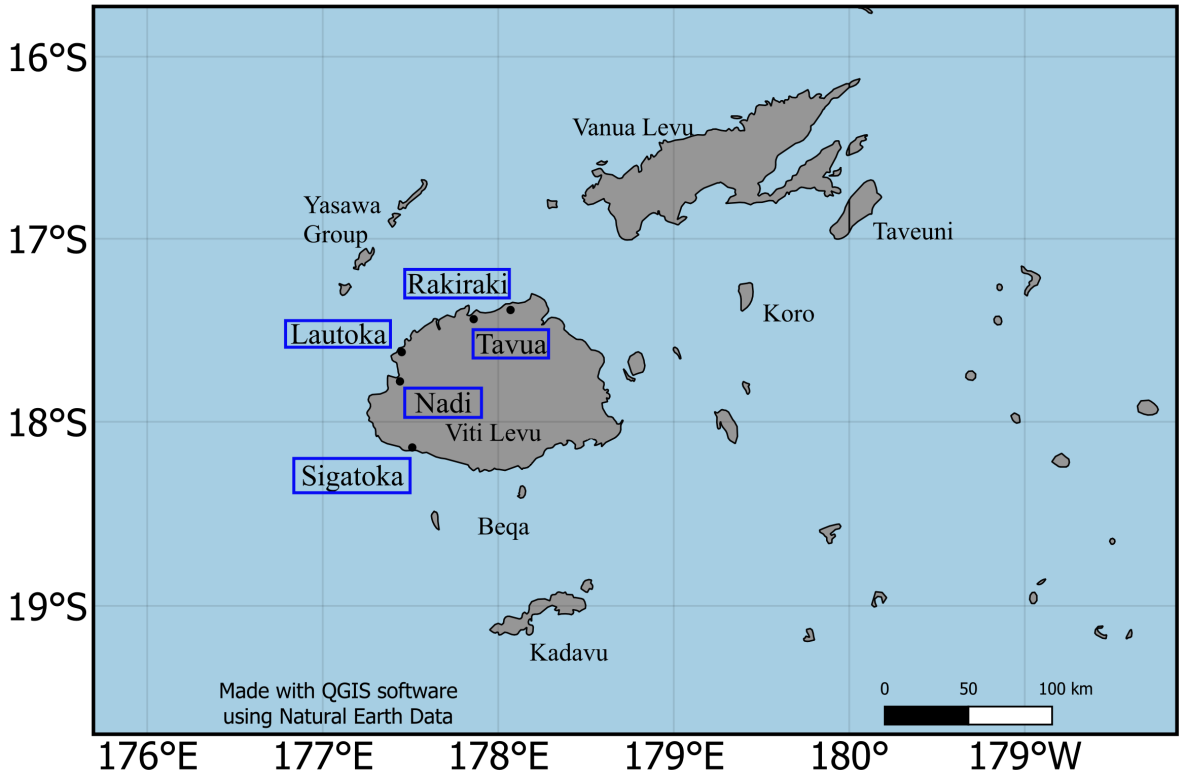

Fig. 2: The map of Fiji shows the various study sites where the hybrid C-GRU flood forecasting framework was developed

3.2 Proposed Hourly Flood Index

In this study, we propose the hourly flood index ($SWRI_{24-hr-S}$), a normalised version of the $WRI_{24-hr-S}$ and develop the proposed hybrid C-GRU model to forecast this index to assess flood risk on an hourly scale. The $SWRI_{24-hr-S}$ is a practical tool for real-time flood risk monitoring

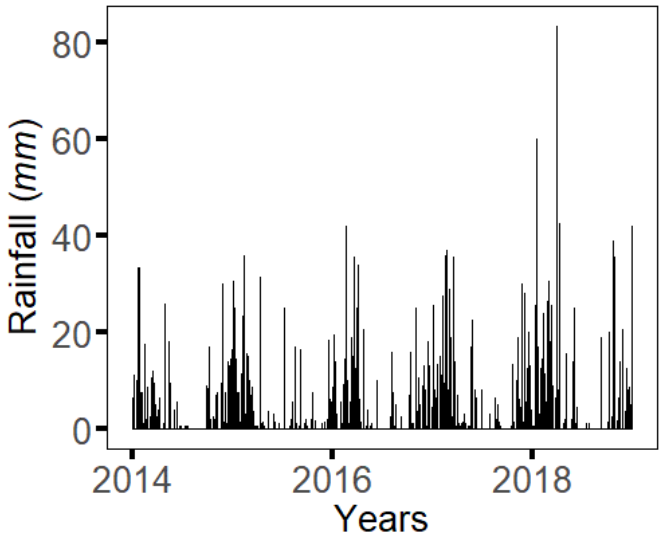

Fig. 3: Hourly rainfall for the Latuoka study site from 2014 to 2018

Table 1: Geographic settings and descriptive statistics of hourly rainfall dataset for the five study sites. Note that the hourly rainfall spans from January 1, 2014, to December 31, 2018, with 43,824 observations

Site	Location (Latitude, Longitude)	Ave. hourly rainfall (mm)	Max. hourly rainfall (mm)	Skewness	Kurtosis
Lautoka	17.62 °S, 177.45 °E	0.19	83.50	16.56	421.51
Nadi	17.78 °S, 177.44 °E	0.27	260.00	36.35	2090.02
Rakiraki	17.39 °S, 178.07 °E	0.23	68.50	17.00	428.86
Sigatoka	18.14 °S, 177.51 °E	0.21	59.00	15.24	316.25
Tavua	17.44 °S, 177.86 °E	0.15	57.50	16.40	381.76

as it is mathematically derived using only the hourly rainfall data, which are readily available for the present study sites. The proposed hourly flood index is implemented using the Python programming language.

The following steps are taken to obtain the $SWRI_{24-hr-S}$. The first step is calculating the $WRI_{24-hr-S}$. The $WRI_{24-hr-S}$ for the current (i_{th}) hour proposed by Deo et al. (2018) is given by the following equation:

$$WRI^{(i)}_{24-hr-S} = P_1 + \frac{[P_2\,(W-1)]}{W} + \frac{[P_3\,(W-1-1/2\,)]}{W} + \ldots + \frac{[P_{24}\,(W-1-1/2-\ldots-1/23\;\;)]}{W} \tag{8}$$

where P_1 is the total rainfall recorded an hour before, P_1 is the total rainfall recorded 2 hours before, and so on; W is the time-reduction weighting factor ($W = 3.8$) verified by Deo et al. (2018) that incorporates the contributions of accumulated rainfall in the latest 24 hours. This weighting factor ensures that the decay of accumulated rainfall or its potential impact on a flood event depends on several hydrological phenomena, including evapotranspiration, percolation, seepage, run-off, drainage, etc., as established in prior studies (Deo et al., 2018; Lu, 2009). The substitution of $W = 3.8$ into Eq. 8 yields:

$$WRI^{(i)}_{24-hr-S} \approx P_1 + 0.74P_2 + 0.61P_3 + \ldots + 0.02P_{24} \tag{9}$$

It is noted that $WRI_{24-hr-S}$ for a current (i^{th}) hour is expected to accumulate $\approx$ 100% of rainfall received an hour before, $\approx$74% of that received two hours before, $\approx$ 61% of that received three hours before, and eventually, $\approx$2% of that received 24 hours before. Following the calculation

of $WRI_{24-hr-S}$ for any study period, the mathematical form of the novel $SWRI_{24-hr-S}$ for a current (i^{th}) hour is calculated as a normalised version of $WRI_{24-hr-S}$:

$$SWRI^{i}_{24-hr-S} = \frac{WRI^{i}_{24-hr-S} - \overline{WRI^{max}_{24-hr-S}}}{\sigma(WRI^{max}_{24-hr-S})} \tag{10}$$

where $\overline{WRI^{max}_{24-hr-S}}$ is the mean monthly maximum values of $WRI_{24-hr-S}$ for the respective study period and $\sigma(WRI^{max}_{24-hr-S})$ is the standard deviation of the monthly maximum values of $WRI_{24-hr-S}$ for the respective study period.

In this paper, we follow the notion that if the magnitude of $SWRI_{24-hr-S}$ for the current (i^{th}) hour is greater than zero (or that the water resources are higher than normal), it is considered as a flood situation. For all the study sites, the $WRI_{24-hr-S}$ followed by $SWRI_{24-hr-S}$ were computed starting from January 2, 2014, as antecedent rainfall data for 24 hours (the hourly rainfall data for January 1, 2014) was required to enable the computation of these metrics. Fig. 4 demonstrates how the $SWRI_{24-hr-S}$ can be practically applied to identify flood events at the Nadi site in April 2016. As shown in Fig. 4, the flood situation at the Nadi site began at 8 a.m. on April 4, 2016, precisely when the magnitude of $SWRI_{24-hr-S}$ first exceeded zero and lasted for 20 hours. To confirm the occurrence of this flood event, we refer to the Fiji Meteorological Services (FMS) annual report 2016 (Fiji Meteorological Service, 2016), which showed indeed that a tropical depression TD14F caused heavy rainfall in some parts of the county between the 3rd and 5th of April 2016. This led to severe flooding, particularly in some major towns in western Viti Levu, including Nadi. Therefore, this verification confirms the practicality of the proposed $SWRI_{24-hr-S}$ in identifying a flood situation on an hourly scale.

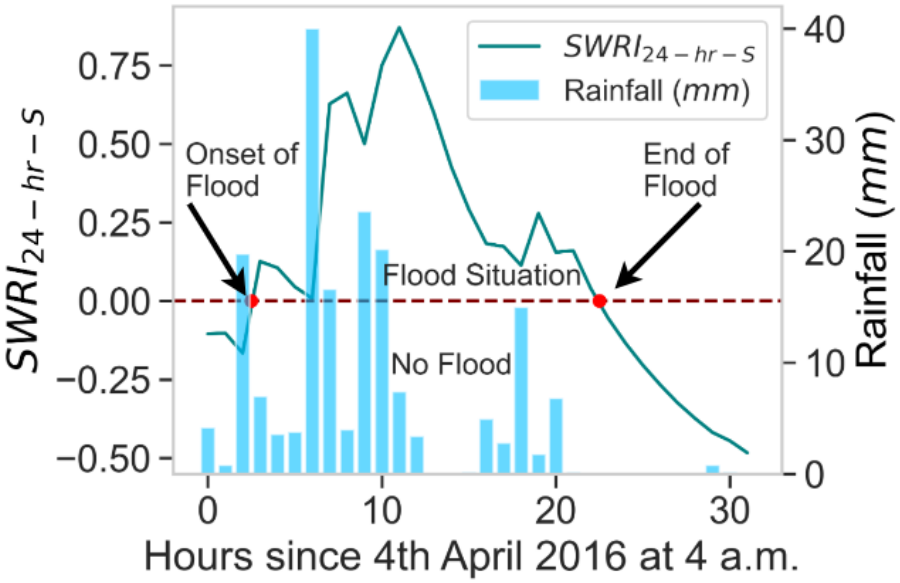

Fig. 4: The $SWRI_{24-hr-S}$ is applied to identify a flood event in April 2016 at the Nadi study site, demonstrating its ability to monitor the flood risk hourly

3.3 Design of the proposed hybrid C-GRU model

This study used Python programming to implement the objective model, i.e., the hybrid C-GRU and all other benchmarking models. The implementation was done via a Google Colaboratory (Google Colab) platform offering a freely available Jupyter Notebook interface supported by a Tensor Processing Unit (TPU) and a Graphical Processing Unit (GPU). Google Colab provides an advanced virtual environment for executing ML and DL algorithms. The DL models were developed using the Keras (Ketkar, 2017) and Tensorflow (Abadi et al., 2016) libraries, and the RFR model was developed using the Sklearn library (Pedregosa et al., 2011). The 'Hyperopt'

library (Komer et al., 2014) was used for BO. The R programming language was also used to plot the correlograms to determine the relevant model input features.

The primary scope of this study is to design a hybrid C-GRU model illustrated in Fig. 1 that can forecast the $SWRI_{24-hr-S}$ for each study site over a 1-hourly forecast horizon. To accomplish this, we have utilised the real-time hourly rainfall and the hourly flood index ($SWRI_{24-hr-S}$) datasets for each study site, following a similar methodology outlined in the related study by Moishin et al. (2021a). The step-by-step procedure to develop the forecasting models at each study site is as follows:

Step 1: Using the Augmented Dickey-Fuller test, we first checked whether the $SWRI_{24-hr-S}$ and hourly rainfall time series data were stationary for each study site (Cheung and Lai, 1995). The null hypothesis (H_0) of this test must be rejected for the data to be stationary. The result showed that both datasets were stationary for all the study sites.

Step 2: The partial auto-correlation function (PACF) and cross-correlation function (CCF) were statistically assessed using the correlogram plots to determine the significant time-lagged inputs for the forecasting models at each study site. The correlogram plot of CCF and PACF was used to determine the most statistically significant lags of hourly rainfall and $SWRI_{24-hr-S}$ time series, respectively, to forecast $SWRI_{24-hr-S}$ at time t. A 95% confidence band was employed in the correlogram plots to assess the significance of variable lags, whereby any lags within this boundary were deemed insignificant. Both the plots were generated by considering only 20 lags (i.e., past 20 hours) of each input variable.

Fig. 5 shows the PACF and CCF plots for the Lautoka site. Upon analysing the PACF plot, only the first three lags of the $SWRI_{24-hr-S}$ time series were the most significant for the Lautoka site. Similarly, upon examining the CCF plot for the Lautoka site, only the most significant lag with the highest cross-correlation coefficient (r_{cross}) of the hourly rainfall time series was selected. This process was repeated for the other study sites to ascertain significant model input features, and the results obtained are furnished in Table 2. Hence, the features used as model inputs included the antecedent $SWRI_{24-hr-S}$ and hourly rainfall, and the target variable consisted of $SWRI_{24-hr-S}$ at time t for each study site. Subsequently, we concatenated the predictor and target variables for each site to form the final dataset to develop the forecasting models.

Step 3: The input data for the model were then normalised between 0 and 1 using the min-max scaling technique provided in the Sklearn library in Python (Pedregosa et al., 2011). This was done to ensure that each input variable has the same order of magnitude, leading to faster and more efficient training of the forecasting model (Prasad et al., 2024). Next, the dataset for model development for each study site was partitioned into training, validation, and testing subsets. Given the lack of consensus on data splitting ratios for training, validation and testing, this study, following the approach of a related study (Moishin et al., 2021a), allocated the first 80% of the dataset for training the model, with 20% of the training data used for validation and utilising the remaining 20% for testing the model. This train-test split was consistently applied across all study sites, as outlined in Table 2.

The validation dataset served two purposes in this study. Firstly, it was used for model hyper-parameter tuning. Secondly, for all DL models developed in this study, the validation data was used to monitor the model's performance during training using the early stopping technique, which is discussed later in this section. The benchmarking DL models, i.e., the CNN model was constructed using three Conv1D layers; the LSTM model was constructed using three LSTM layers; the GRU model was constructed using three GRU layers. During training, the hyperparameter optimisation for all the forecasting models was executed using BO with the TPE algorithm, with a maximum of 40 evaluations. The parameter search space used for BO optimisation for each forecasting model is furnished in Table 3. While some hyperparameters are specific to the model, four common hyperparameters used in any DL model, which are also utilised in this study, are as follows:

- Activation Function: All layers within a network, except for the output layer, typically use the same activation function, the Rectified Linear Unit (ReLU). The primary advantage of

using ReLU compared to other activation functions, such as sigmoid and *tanh*, is that ReLU introduces nonlinearities into the model by setting its negative values to zero. This helps overcome the issue of vanishing gradients, enabling a model to learn faster and achieve better performance (Ghimire et al., 2022).

- Dropout (Ghimire et al., 2019): Dropout is employed as a regularisation technique to mitigate overfitting and enhance training performance. It accomplishes this by randomly selecting a fraction of neurons during each iteration and preventing them from undergoing training. This fraction of neurons, termed a dropout rate, is a real hyperparameter ranging from 0 to 1. In this study, this parameter was set to a fixed value of 0.1 for all DL models.

- Epochs and early stopping technique (Ghimire et al., 2023, 2024a, 2022, 2019): The number of epochs was set to 1000 during model training. However, we implemented early stopping as a regularisation method to monitor the model's performance during training. A stopping criterion was defined based on the performance metric (i.e., mean squared error (MSE)) on the validation data, such that the training process was terminated when the validation loss stopped decreasing for a certain number of epochs specified by the "patience" term, or when the validation loss started to increase, indicating model's potential overfitting to the training data. The training was terminated once the early stopping criterion was met, and the model with the lowest validation loss was saved. This method helped prevent overfitting and saved computational time. In this study, the "patience" hyperparameter for early stopping criteria was set to 20.

- Backpropagation optimisation algorithm: Optimisation algorithms are used in backpropagation to update a network's weights during training. In this study, the Adaptive Moment Estimator (Adam) optimiser with a learning rate of 0.001 was used as the backpropagation learning algorithm. Adam combines the advantages of the Adaptive Gradient Algorithm (AdaGrad) and the Root Mean Square Propagation (RMSProp) (Kingma and Ba, 2014; Zou et al., 2019). This approach calculates adaptive learning rates for each parameter based on estimates of the first and second moments of the gradients (Kingma and Ba, 2014). Adam optimiser is computationally efficient, has low memory requirements, and is well-suited for large datasets (Ghimire et al., 2022).

Step 4: Finally, the various performance evaluation metrics (discussed in the next section) were used to compare the performance of the proposed hybrid C-GRU model with the benchmark models, i.e., CNN, LSTM, GRU, and RFR models.

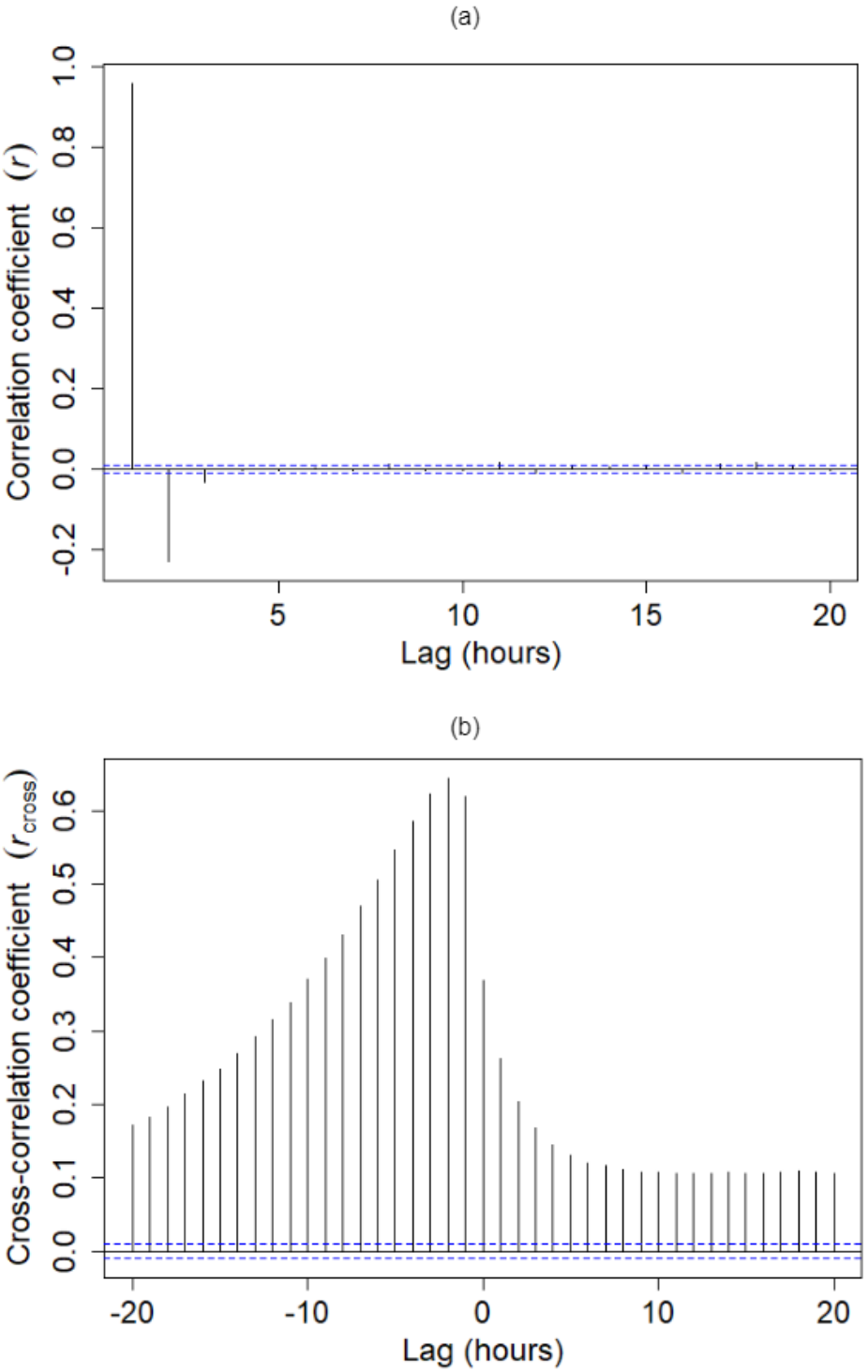

Fig. 5: Significant model input feature selection using (a) the correlogram plot of the partial auto-correlation function (PACF) of the $SWRI_{24-hr-S}$ series showing the three most significant lags for the Lautoka site and (b) the correlogram plot of the cross-correlation function (CCF) for the $SWRI_{24-hr-S}$ versus the real-time hourly rainfall for the Lautoka site

Table 2: Model input data at each study site, with site-based data partition into training, validation, and testing sets for model development

Study Site	Model Input		Data Points	Training Data (≈80% of all data points)	Validation Data (≈20% of training data)	Test Data (≈20% of all data points)
	Significant lags of the $SWRI_{24-hr-S}$ series	Significant lags of the real-time hourly rainfall series				
Lautoka	Lag 1, Lag 2, and Lag 3 ($SWRI_{24-hr-S}$ at $t-1$, $t-2$, and $t-3$)	Lag 1 (hourly rainfall at $t-1$)	43797	35038	7008	8759
Nadi	Lag 1, Lag 2, Lag 3, and Lag 4 ($SWRI_{24-hr-S}$ at $t-1$, $t-2$, $t-3$, and $t-4$)		43796	35037	7007	8759
Rakiraki	Lag 1, Lag 2, and Lag 3 ($SWRI_{24-hr-S}$ at $t-1$, $t-2$, and $t-3$)		43797	35038	7008	8759
Sigatoka	Lag 1, Lag 2, and Lag 3 ($SWRI_{24-hr-S}$ at $t-1$, $t-2$, and $t-3$)		43797	35038	7008	8759
Tavua	Lag 1 and Lag 2 ($SWRI_{24-hr-S}$ at $t-1$ and $t-2$)		43798	35038	7008	8760

Table 3: Parameter search space assigned to the Bayesian optimisation (BO) algorithm for developing the proposed hybrid C-GRU and benchmark models

Forecasting models	Model hyperparameters	Hyperparameter search space
Hybrid C-GRU (Objective Model)	CNN filter 1	{50, 60, 100, 200, 250, 300}
	CNN filter 2	{50, 60, 100, 200, 250, 300}
	GRU cell unit 1	{30, 40, 50, 60, 80, 100, 150, 200}
	GRU cell unit 2	{30, 40, 50, 60, 80, 100, 150, 200}
	Batch	{64, 128, 256, 512}
GRU	GRU cell unit 1	{30, 40, 50, 60, 80, 100, 150, 200}
	GRU cell unit 2	{30, 40, 50, 60, 80, 100, 150, 200}
	GRU cell unit 3	{30, 40, 50, 60, 80, 100, 150, 200}
	Batch	{64, 128, 256, 512}
LSTM	LSTM cell unit 1	{30, 40, 50, 60, 80, 100, 150, 200}
	LSTM cell unit 2	{30, 40, 50, 60, 80, 100, 150, 200}
	LSTM cell unit 3	{30, 40, 50, 60, 80, 100, 150, 200}
	Batch	{64, 128, 256, 512}
CNN	CNN filter 1	{50, 60, 100, 200, 250, 300}
	CNN filter 2	{50, 60, 100, 200, 250, 300}
	CNN filter 3	{10, 20, 30, 40, 50, 80}
	Batch	{64, 128, 256, 512}
RFR	Number of trees in the forest (n_estimators)	{50, 100, 150, 200, 250, 300, 350, 400}
	Minimum number of samples required to be at the leaf node (min_samples_leaf)	Uniform {0, 0.5}
	Minimum samples to split an internal node (min_samples_split)	Uniform {0, 1}
	Maximum features to consider for the best split (max_features)	{1, 'sqrt', 'log2', 'None'}

3.4 Model Performance Evaluation Criteria

The performance of the hybrid C-GRU model against the benchmark models was evaluated using two sets of statistical metrics: Category A (ideal value = 1) and Category B (ideal value = 0). Different metrics were utilised within each category to address limitations and take advantage of various statistical metrics available (Joseph et al., 2024).

In this study, within category A, five statistical metrics were employed, namely the Pearson's Correlation Coefficient (r), Nash-Sutcliffe Efficiency Index (E_{NS}), Willmott's Index of Agreement (E_{WI}), Legate-McCabe Efficiency Index (E_{LM}), and Kling-Gupta Efficiency (KGE). In category B, three error metrics were utilised: Mean Absolute Error (MAE), Root Mean Square Error ($RMSE$), and Symmetric Mean Absolute Percentage Error ($sMAPE$) (%). The statistical metrics in Category A assessed the variance between forecasted and observed $SWRI_{24-hr-S}$ values, while the error metrics in Category B were utilised to examine model bias (Joseph et al., 2023). As bias and variance contribute to reducible error, the models were compared based on their ability to minimise bias and variance (Joseph et al., 2023).

The Python package 'HydroErr' (Roberts et al., 2018) was used to implement these performance evaluation metrics. The mathematical expression of these metrics is as follows:

$$r = \frac{\sum_{i=1}^{n} \left(O_i - \overline{O}\right)\left(S_i - \overline{S}\right)}{\sqrt{\sum_{i=1}^{n} \left(O_i - \overline{O}\right)^2} \sqrt{\sum_{i=1}^{n} \left(S_i - \overline{S}\right)^2}}, (-1 \leq r \leq 1) \tag{11}$$

$$E_{NS} = 1 - \frac{\sum_{i=1}^{n} \left(S_i - O_i\right)^2}{\sum_{i=1}^{n} \left(O_i - \overline{O}\right)^2}, (-\infty < E_{NS} \leq 1) \tag{12}$$

$$E_{WI} = 1 - \frac{\sum_{i=1}^{n} (S_i - O_i)^2}{\sum_{i=1}^{n} \left(|S_i - \overline{O}| + |O_i - \overline{O}|\right)^2}, (0 \leq E_{WI} \leq 1) \tag{13}$$

$$E_{LM} = 1 - \frac{\sum_{i=1}^{n} |S_i - O_i|}{\sum_{i=1}^{n} |O_i - \overline{O}|}, (-\infty < E_{LM} \leq 1) \tag{14}$$

$$KGE = \sqrt{(r-1)^2 + \left(\frac{CV_S}{CV_O} - 1\right)^2 + \left(\frac{\overline{S}}{\overline{O}} - 1\right)^2}, (-\infty < KGE \leq 1) \tag{15}$$

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^{n} (S_i - O_i)^2}, (0 \leq RMSE < +\infty) \tag{16}$$

$$MAE = \frac{1}{n} \sum_{i=1}^{n} |S_i - O_i|, (0 \leq MAE < +\infty) \tag{17}$$

$$sMAPE = \frac{1}{n} \sum_{i=1}^{n} \frac{|S_i - O_i|}{|S_i| + |O_i|} \times 100, (0 \leq sMAPE < +\infty) \tag{18}$$

where S is the forecasted $SWRI_{24-hr-S}$, O is the observed (or actual) $SWRI_{24-hr-S}$, $\overline{S}$ is the mean of forecasted $SWRI_{24-hr-S}$, $\overline{O}$ is the mean of the observed $SWRI_{24-hr-S}$, n is the number of values, CV_S is the coefficient of variation of forecasted $SWRI_{24-hr-S}$ and CV_O is the coefficient of variation of observed $SWRI_{24-hr-S}$.

Despite using various performance evaluation metrics to compare the proposed hybrid C-GRU model's performance with the benchmark models, ranking the forecasting models solely based on such metrics is challenging as each metric has distinct advantages and limitations (Ghimire et al., 2024b). To overcome this challenge, this study used a robust global performance indicator (GPI) (Behar et al., 2015; Ghimire et al., 2022, 2023, 2022, 2024b, 2023; Joseph et al., 2023, 2024) to rank and establish overall model performance. The GPI combines the outcomes of all eight metrics used in this study for a comprehensive model performance evaluation, with a higher GPI indicating greater model accuracy. For the i^{th} model, the GPI is calculated as (Joseph et al., 2024):

$$GPI = \sum_{j=1}^{N} \alpha_j \left(\bar{y}_j - y_{ij}\right), (-\infty < GPI < +\infty) \tag{19}$$

where N is the total number of performance evaluation metrics used (i.e., 8 in this study), $\alpha_j = -1$ for Category A metrics and $\alpha_j = +1$ for Category B metrics, y_{ij} is the scaled value of metric j for model i, and $\bar{y}_j$ is the median value of scaled values of metric j.

4 Results and Discussion

This section provides an account of the empirical results of the modelling experiments carried out and the assessments of the hybrid C-GRU model's performance in forecasting the $SWRI_{24-hr-S}$ over a 1-hourly forecast horizon for each study site compared to benchmark models, including CNN, LSTM, GRU, and RFR. The proposed hybrid C-GRU model's robustness against the benchmark models to forecast the $SWRI_{24-hr-S}$ was comprehensively assessed using various performance evaluation metrics given in section 3.4 and visual plots using the testing datasets for five study sites. In this section, we also propose the practical application of the proposed hybrid C-GRU model in the decision support system for early flood warnings.

4.1 Forecasting performance of the hybrid C-GRU vs. benchmark models

The initial evaluation of the hybrid C-GRU model's forecasting capability was based on the scatterplot between the forecasted $SWRI_{24-hr-S}$ ($SWRI^{for}_{24-hr-S}$) and observed $SWRI_{24-hr-S}$ ($SWRI^{obs}_{24-hr-S}$) in the testing phase for all five study sites (Fig. 6).

The scatter plots also include the coefficient of determination R^2, the 1:1 line and the equation of the goodness-of-fit line; $SWRI^{for}_{24-hr-S} = m \times SWRI^{obs}_{24-hr-S} + C$, where m is the gradient, and C is the y-intercept of the goodness-of-fit line. The 1:1 line represents an exact match between observed and forecasted $SWRI_{24-hr-S}$ values. The closer the data points align with this line, the better the model fits the data, indicating minimal discrepancies between observed and forecasted $SWRI_{24-hr-S}$ values. Consequently, a perfectly fitting regression model should have $m = 1$, $C = 0$, and $R^2 = 1$. While there was generally a good agreement between observed and forecasted $SWRI_{24-hr-S}$ values across all tested models, it was observed that the hybrid C-GRU model yielded notably more accurate forecasted $SWRI_{24-hr-S}$ values compared to all other models for all the study sites.

As depicted in Fig. 6, across all the study sites, the data points closely align with the 1:1 line in the scatterplot, and the values of m and R^2 were close to unity for the hybrid C-GRU model. This indicates a high level of agreement between the forecasted and observed $SWRI_{24-hr-S}$ values compared to other benchmark models. The values of m and R^2 in pairs ($m|R^2$) for the hybrid C-GRU model were 1.01|0.996 for the Lautoka site, 1.02|0.993 for the Nadi site, 0.99|0.998 for the Rakiraki site, 0.97|0.998 for the Sigatoka site, and 1.01|0.997 for the Tavua site. Alternatively, the y-intercept (Ideal value=0) for the hybrid C-GRU model for all the study sites was found to be close to naught, i.e., 0.01 for the Lautoka, Nadi, Rakiraki and Tavua sites and 0.04 for the Sigatoka site. Among the models tested, RFR consistently performed poorly across all study sites, as shown in Fig. 6.

More specifically, Fig. 6 revealed that RFR, compared to the other models, was underfitting for $SWRI_{24-hr-S} > 0$ across all the study sites. Hence, it is not a very suitable forecasting model in this study, as accurately forecasting $SWRI_{24-hr-S} > 0$ is crucial since it indicates a flood situation. To further visualise the similarity between forecasted and observed $SWRI_{24-hr-S}$ values, we employed a line plot, as depicted in Fig. 7, for the Rakiraki site. This plot compares the $SWRI_{24-hr-S}$ generated by the hybrid C-CGRU model with those of benchmark models, i.e., GRU, LSTM, CNN, and RFR. The plot illustrates that $SWRI_{24-hr-S}$ values forecasted by the hybrid C-CRU model exhibit greater similarity to the observed $SWRI_{24-hr-S}$ values than the other models. These primary results already demonstrate the superior forecasting capability of the proposed hybrid C-GRU model compared to other benchmarking models.

The hybrid C-GRU model's superior forecasting capability against the benchmarking models was further assessed using various performance evaluation metrics for each site study in the testing phase (Table 4, Figs. 8 and 9). The metric r is a non-dimensional and absolute metric that assesses the strength and direction of the linear relationship between the observed and forecasted $SWRI_{24-hr-S}$ values (Joseph et al., 2023). The $RMSE$ and MAE measures are derived from the aggregating residuals between observed and forecasted $SWRI_{24-hr-S}$ values (Joseph et al., 2023). The value of r closer to 1, along with lower values of $RMSE$ and MAE (ideally around 0), indicates the optimal model.

Table 4 shows that the proposed hybrid C-GRU model obtained the highest r $(0.996 - 0.999)$ and the lowest $RMSE$ $(0.007 - 0.014)$ and MAE $(0.003 - 0.004)$ for all study sites. The r metric for three sites (Nadi, Sigatoka, and Tavua) was the same for the proposed hybrid C-GRU model, and one or two other benchmarking models (i.e., GRU model for Nadi, GRU and CNN models for Sigatoka, GRU and LSTM models for Tavua). Also, the $RMSE$ of the proposed hybrid C-GRU and LSTM models for the Tavua site were the same. On the contrary, the MAE at these sites was lower for the proposed hybrid C-GRU model than these benchmarking models. It should be noted that all DL models developed in this study performed exceptionally well based on the r metric. Nevertheless, a primary drawback of r is its susceptibility to outliers. Also, while r is a scale and offset invariant of the data, it can sometimes yield higher values for models that

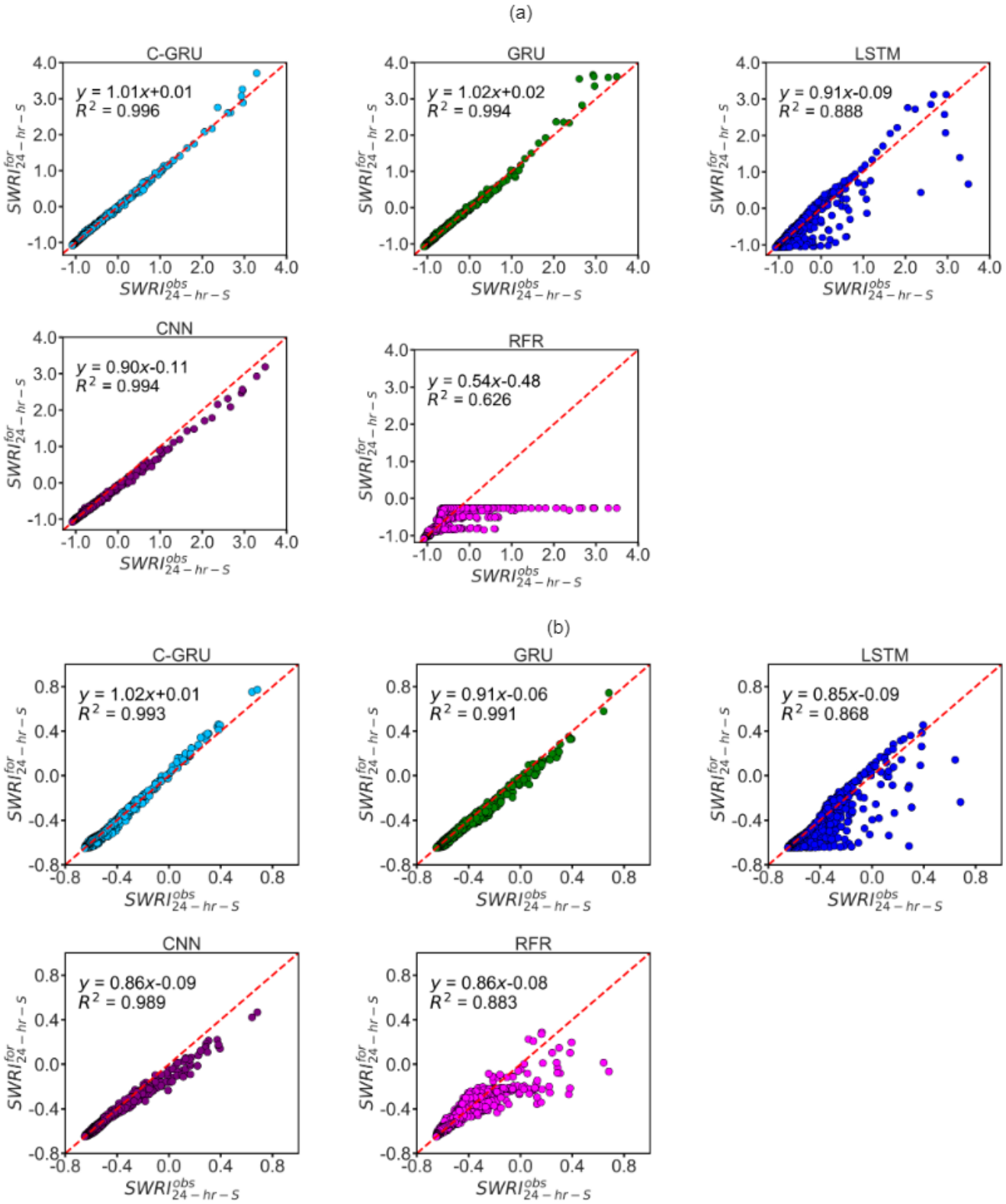

Fig. 6: Scatterplot of the forecasted $SWRI_{24-hr-S}$ ($SWRI^{for}_{24-hr-S}$) vs. observed $SWRI_{24-hr-S}$ ($SWRI^{obs}_{24-hr-S}$) generated from the proposed hybrid C-GRU model, compared with the four other benchmarking models (i.e., GRU, LSTM, CNN, and RFR) for the five study sites: (a) Lautoka, (b) Nadi, (c) Rakiraki, (d) Sigatoka, and (e) Tavua, in the testing phase. The scatterplots also show the 1:1 line (red dashed line), the equation of the goodness-of-fit line, and the coefficient of determination (R^2) displayed in each panel

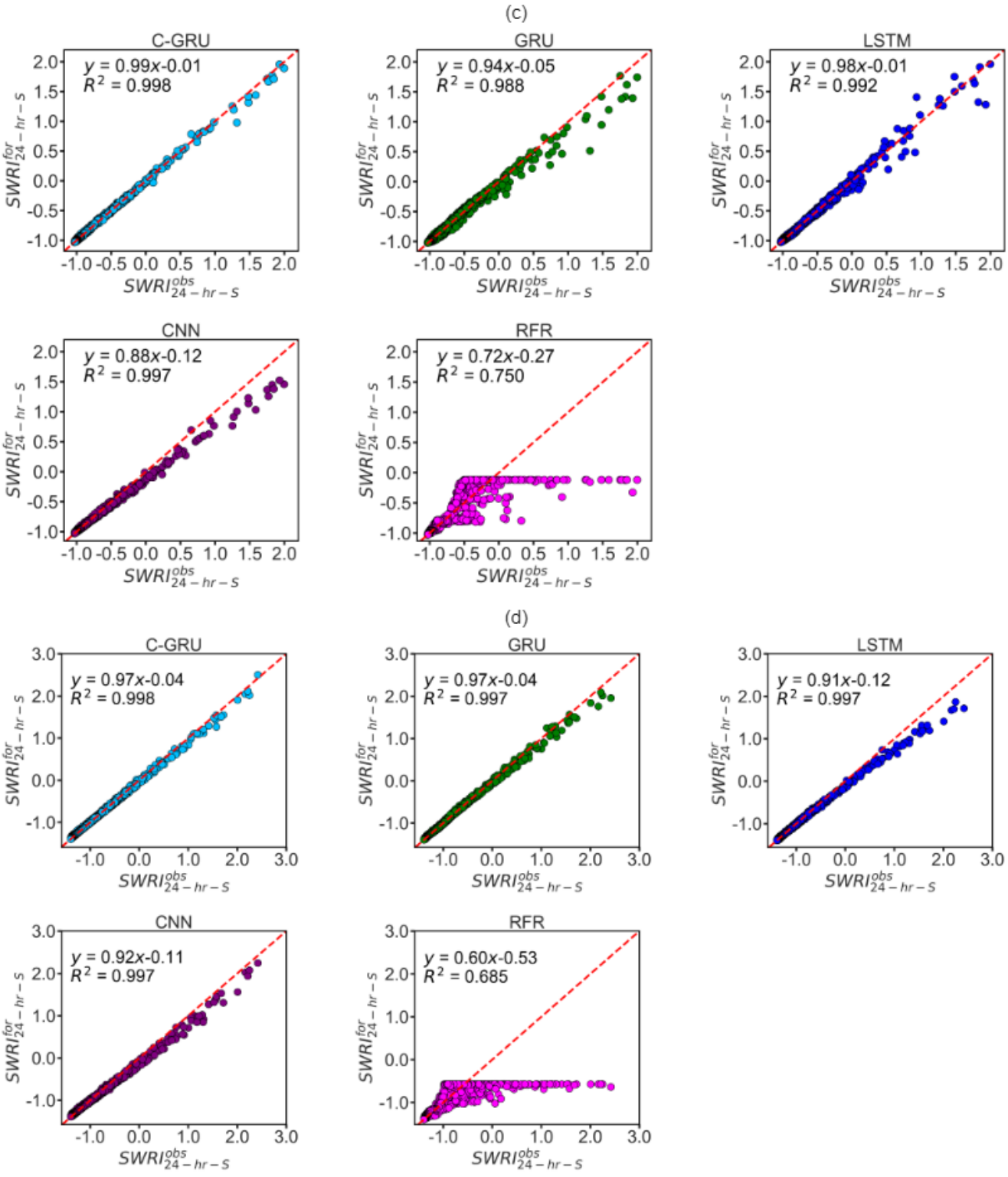

Fig. 6: (continued)

perform only moderately well (Joseph et al., 2023). Additionally, squaring residuals in $RMSE$ can introduce a bias towards higher $SWRI_{24-hr-S}$ values (Joseph et al., 2023). Therefore, these two metrics (i.e., r and $RMSE$) can sometimes be unreliable. The absolute computation of residuals in MAE mitigates biases, making it more reliable than r and $RMSE$. However, while MAE is often considered an alternative to $RMSE$, both are absolute error indicators unsuitable for comparing models across geographically diverse sites (Joseph et al., 2024, 2023). This is simply because sites with higher $SWRI_{24-hr-S}$ values will essentially yield larger absolute error values than sites with lower $SWRI_{24-hr-S}$ values, regardless of model performance.

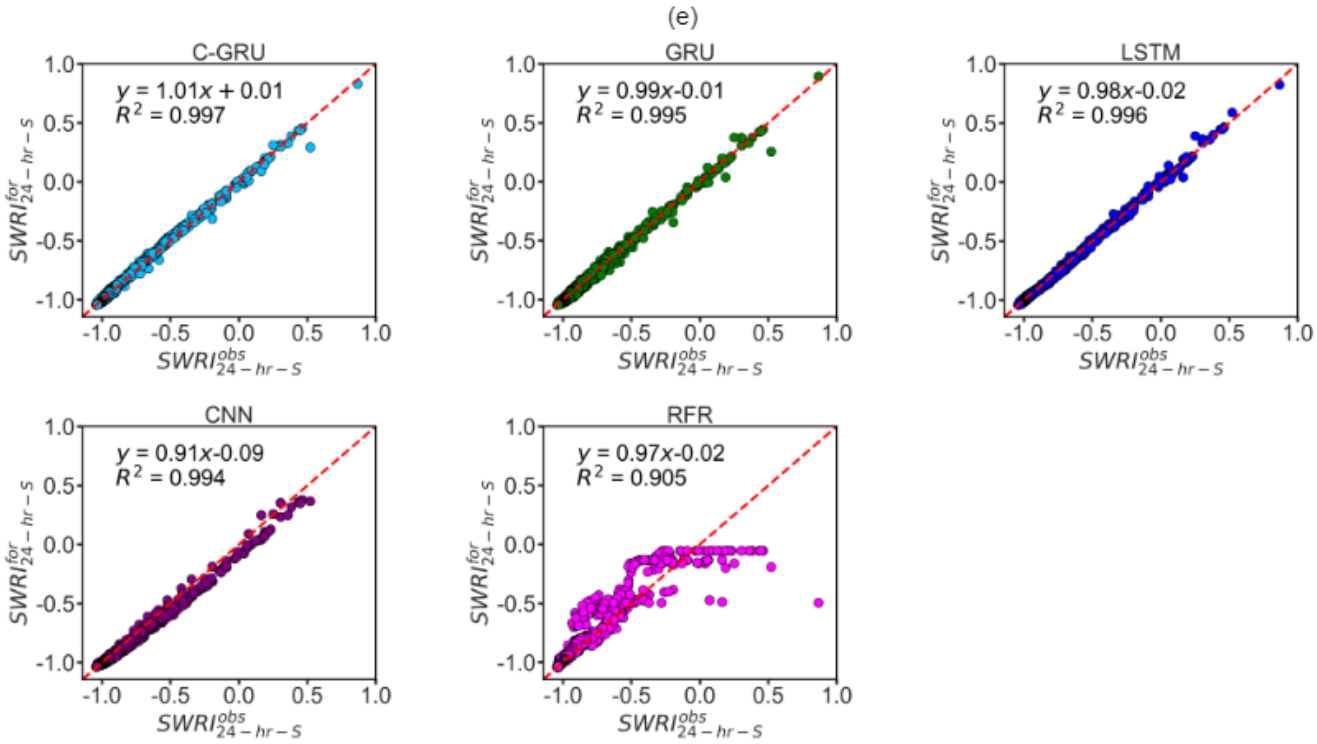

Fig. 6: (continued)

Table 4: Evaluation of the proposed objective model (i.e., C-GRU) vs. all other comparative models in the testing phase for all study sites using the Pearson's Correlation Coefficient (r), Root Mean Square Error ($RMSE$), Mean Absolute Error (MAE), Nash-Sutcliffe Efficiency Index (E_{NS}), Willmott's Index of Agreement (E_{WI}), and Legate-McCabe Efficiency Index (E_{LM})

Site	Forecasting models	r	$RMSE$	MAE	E_{NS}	E_{WI}	E_{LM}
	C-GRU	0.998	0.014	0.003	0.996	0.999	0.963
	GRU	0.997	0.020	0.006	0.993	0.996	0.940
Lautoka	LSTM	0.942	0.077	0.015	0.887	0.970	0.835
	CNN	0.997	0.025	0.007	0.988	0.997	0.929
	RFR	0.791	0.413	0.024	0.612	0.847	0.745
	C-GRU	0.996	0.008	0.003	0.992	0.998	0.933
	GRU	0.996	0.011	0.006	0.985	0.996	0.868
Nadi	LSTM	0.932	0.032	0.008	0.888	0.963	0.803
	CNN	0.995	0.015	0.005	0.971	0.992	0.879
	RFR	0.940	0.031	0.008	0.882	0.967	0.812
	C-GRU	0.999	0.008	0.004	0.998	0.999	0.956
	GRU	0.994	0.023	0.009	0.985	0.996	0.888
Rakiraki	LSTM	0.996	0.017	0.006	0.992	0.998	0.928
	CNN	0.998	0.025	0.007	0.982	0.995	0.914
	RFR	0.866	0.093	0.018	0.749	0.920	0.782
	C-GRU	0.999	0.012	0.004	0.998	0.999	0.965
	GRU	0.999	0.015	0.005	0.996	0.999	0.957
Sigatoka	CNN	0.999	0.025	0.009	0.991	0.998	0.926
	LSTM	0.998	0.026	0.006	0.990	0.997	0.955
	RFR	0.828	0.149	0.031	0.673	0.879	0.754
	C-GRU	0.998	0.007	0.003	0.997	0.999	0.946
	GRU	0.998	0.009	0.005	0.994	0.999	0.906
Tavua	LSTM	0.998	0.007	0.004	0.997	0.999	0.926
	CNN	0.997	0.014	0.007	0.987	0.996	0.869
	RFR	0.951	0.040	0.009	0.899	0.975	0.840

To overcome the limitations of the absolute measures, in this study, we have used the relative error measure, i.e., Symmetric Mean Absolute Percentage Error ($sMAPE$) (%) (Fig. 8), to assess

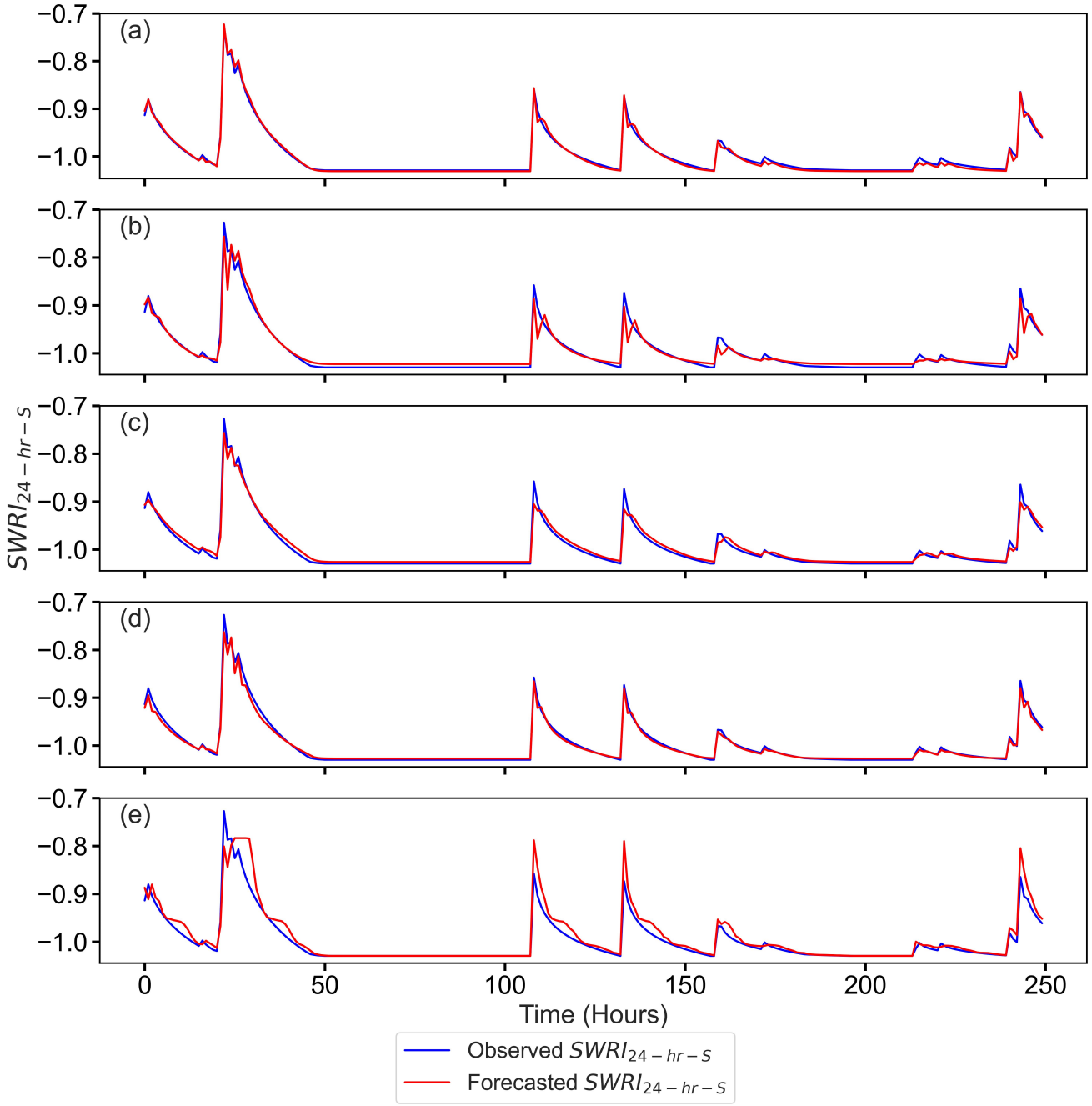

Fig. 7: Observed vs. forecasted $SWRI_{24-hr-S}$ generated by the (a) proposed hybrid C-GRU model compared to four benchmarking models: (b) GRU, (c) LSTM, (d) CNN and (e) RFR for the Rakiraki Site in the testing phase (for closer examination, the plot displays results for only 250 data points from the test set, equivalent to 250 hours)

model bias across different study sites. The $sMAPE$ is a symmetrical measure that avoids the issue of division by zero (Ghimire et al., 2024b). Conversely, the conventional Mean Absolute Percentage Error ($MAPE$) metric tends to be overinflated when the observed value is close to zero (note that we do have observed and forecasted $SWRI_{24-hr-S}$ values around zero for all the study sites), whereas $sMAPE$ does not encounter this problem (Ghimire et al., 2024b). The model yielding the lowest $sMAPE$ is deemed superior.

As illustrated in Fig. 8, the proposed hybrid C-GRU model exhibited superior performance by consistently achieving lower $sMAPE$ values than all benchmark models. A closer examination of Fig. 8 revealed that the proposed hybrid C-GRU model, compared with the standalone models, i.e., GRU and CNN, demonstrated a significant percentage decrease in $sMAPE$ across all the study sites. The proposed model achieved significant $sMAPE$ reductions ranging from -56.5% to

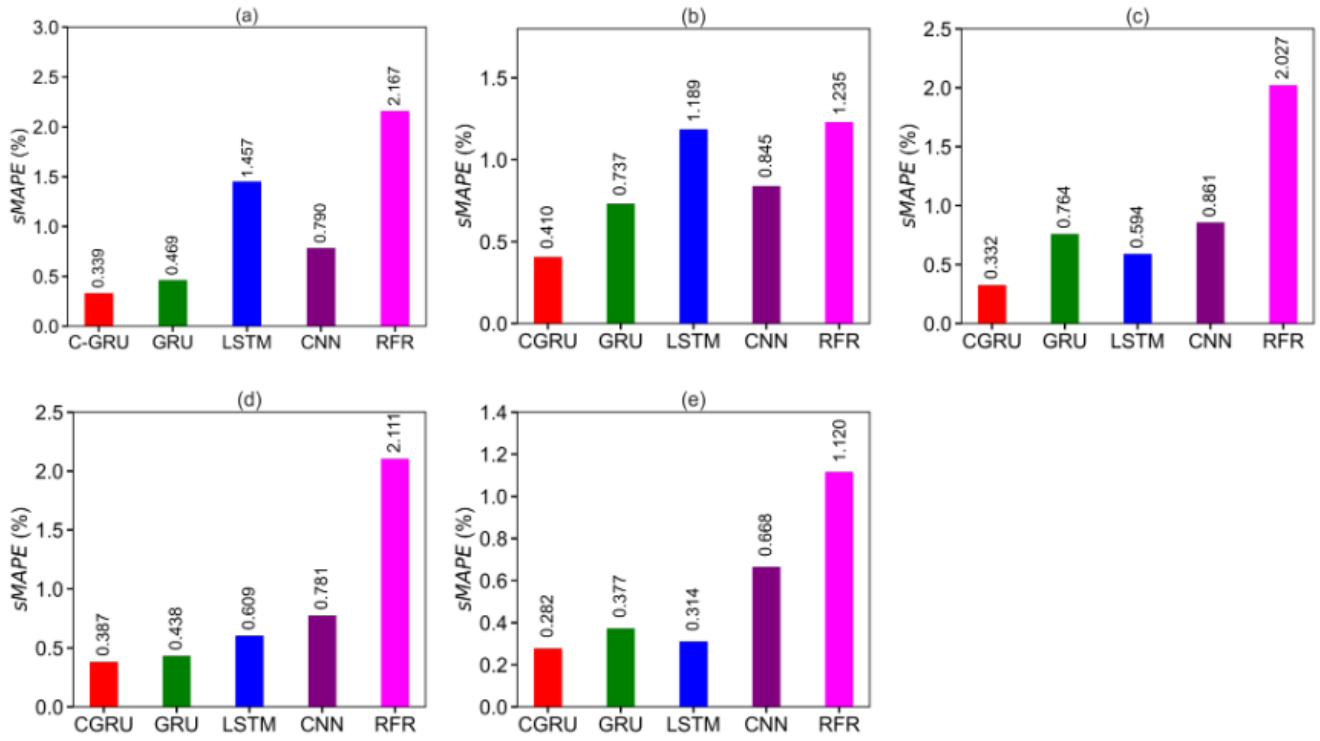

Fig. 8: Evaluation of the proposed objective model (i.e., C-GRU) vs. all other comparative models using the Symmetric Mean Absolute Percentage Error ($sMAPE$) (%) for the five study sites: (a) Lautoka, (b) Nadi, (c) Rakiraki, (d) Sigatoka, and (e) Tavua, in the testing phase

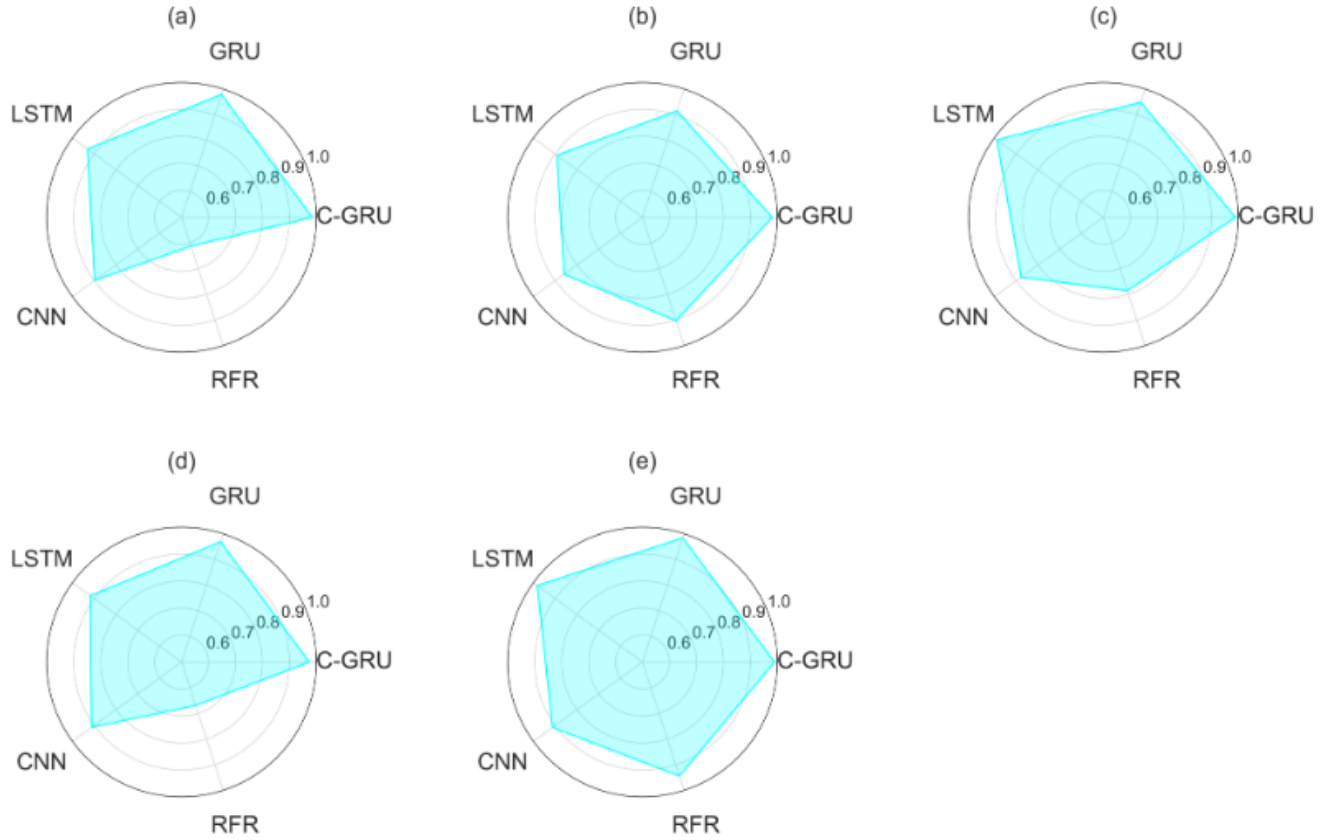

Fig. 9: Evaluation of the proposed objective model (i.e., C-GRU) vs. all other comparative models using the Kling-Gupta Efficiency (KGE) for the five study sites: (a) Lautoka, (b) Nadi, (c) Rakiraki, (d) Sigatoka, and (e) Tavua, in the testing phase

577 -11.6% compared to GRU and from -61.4% to -50.4% compared to CNN across all study sites.

When compared with the poorly performing benchmark model RFR, the proposed hybrid C-GRU model also demonstrated a significant percentage decrease in $sMAPE$: Lautoka (-84.4%), Nadi (-66.8%), Rakiraki (-83.6%), Sigatoka (-81.7%), and Tavua (-74.8%). This clearly establishes the proposed hybrid C-GRU model's superior performance compared to all benchmark models developed in this study.

Furthermore, Table 4 also provides additional metrics, including E_{NS}, E_{WI}, and E_{LM}, to analyse the forecasting performance of the proposed hybrid C-GRU model and the benchmarking models. The E_{NS} (Nash and Sutcliffe, 1970) is a dimensionless metric, which is a scaled version of MSE. It assesses the goodness-of-fit between the observed and forecasted data by comparing the residual variance and observed data variance (Joseph et al., 2023; Prasad et al., 2019). The ideal value of $E_{NS} = 1$, indicating a perfect agreement between the forecasted and observed $SWRI_{24-hr-S}$ values. The E_{NS} for the proposed hybrid C-GRU was closer to unity for all the study site sites (Lautoka: 0.996; Nadi: 0.992; Rakiraki and Sigatoka: 0.998; Tavua: 0.997), showcasing its superior forecasting performance compared to all benchmarking models. However, like $RMSE$ and MSE, which are biased towards larger values, the E_{NS} tends to overestimate larger $SWRI_{24-hr-S}$ values while neglecting lower $SWRI_{24-hr-S}$ values (Joseph et al., 2023).

The E_{WI} (Willmott, 1984, 1981) addresses this issue by examining the ratio of MSE instead of differences. This approach proves advantageous in detecting additive and proportional disparities between the forecasted and observed means and variances (Joseph et al., 2023, 2024; Prasad et al., 2019). The E_{WI} ranges from 0 to 1, where values closer to unity indicate a higher agreement between the forecasted and observed $SWRI_{24-hr-S}$ values. Across all five study sites, the average E_{WI} of the proposed hybrid C-GRU model demonstrated improvements of 0.16%, 1.34%, 0.34%, and 8.85% over the benchmark GRU, LSTM, CNN, and RFR models, respectively. While the E_{WI} is an improvement over r and E_{NS}, it remains sensitive to peak residuals due to the squaring of residuals in the numerator (Joseph et al., 2024; Krause et al., 2005).

Therefore, the E_{WI} can assign higher values to even poor-performing models (Joseph et al., 2024; Krause et al., 2005). The E_{LM} (Legates and McCabe Jr, 1999) (ideal value = +1) addresses these issues by substituting the squaring of the residual term in the numerator with the absolute value (Joseph et al., 2023, 2024; Prasad et al., 2019). Consequently, E_{LM} is not inflated and is unaffected by extreme $SWRI_{24-hr-S}$ values. Hence, E_{LM} can used as a reliable model assessment metric that is also easy to interpret (Joseph et al., 2023; Prasad et al., 2019). Table 4 illustrates that the proposed hybrid C-GRU model attained the highest E_{LM} values, which were also close to unity across all study sites (Lautoka: 0.963; Nadi: 0.933; Rakiraki: 0.956; Sigatoka: 0.965; Tavua: 0.946) compared to all benchmarking models.

The efficiency of the proposed hybrid C-GRU model was also verified using the Kling-Gupta efficiency (KGE) (Gupta et al., 2009; Kling et al., 2012) metric, which also overcomes the limitations of E_{NS}. The KGE assigns an equal weighting to the three components (i.e., correlation, bias, and variability measures) of the observed and forecasted $SWRI_{24-hr-S}$ values to ensure that the bias and variability ratios are not cross-correlated (Ghimire et al., 2019; Joseph et al., 2023). The KGE ranges from $-\infty$ to 1, where values closer to unity indicate a perfect fit. As illustrated in Fig. 9, the proposed hybrid C-GRU model demonstrates superior performance, exemplified by its high KGE value close to unity, outperforming all benchmark models.

Our proposed hybrid C-GRU model is further appraised using the GPI, as depicted in Fig. 10. As previously discussed, while various performance evaluation metrics were employed to compare all the forecasting models developed, solely relying on these metrics to identify the best-performing model can be challenging. Therefore, the GPI, which incorporates all the performance evaluation metrics in this study, was employed as a more robust metric.

Fig. 10 illustrates that the proposed hybrid C-GRU model achieved the highest GPI across all study sites compared to the benchmarking models. Therefore, after conducting a comprehensive performance evaluation of all forecasting models developed in this study, using various performance evaluation metrics, including the GPI metric, it is evident that the proposed hybrid

C-GRU model outperforms all benchmark models. Hence, it can be considered the optimal model for accurately forecasting the $SWRI_{24-hr-S}$ over a 1-hourly forecast horizon.

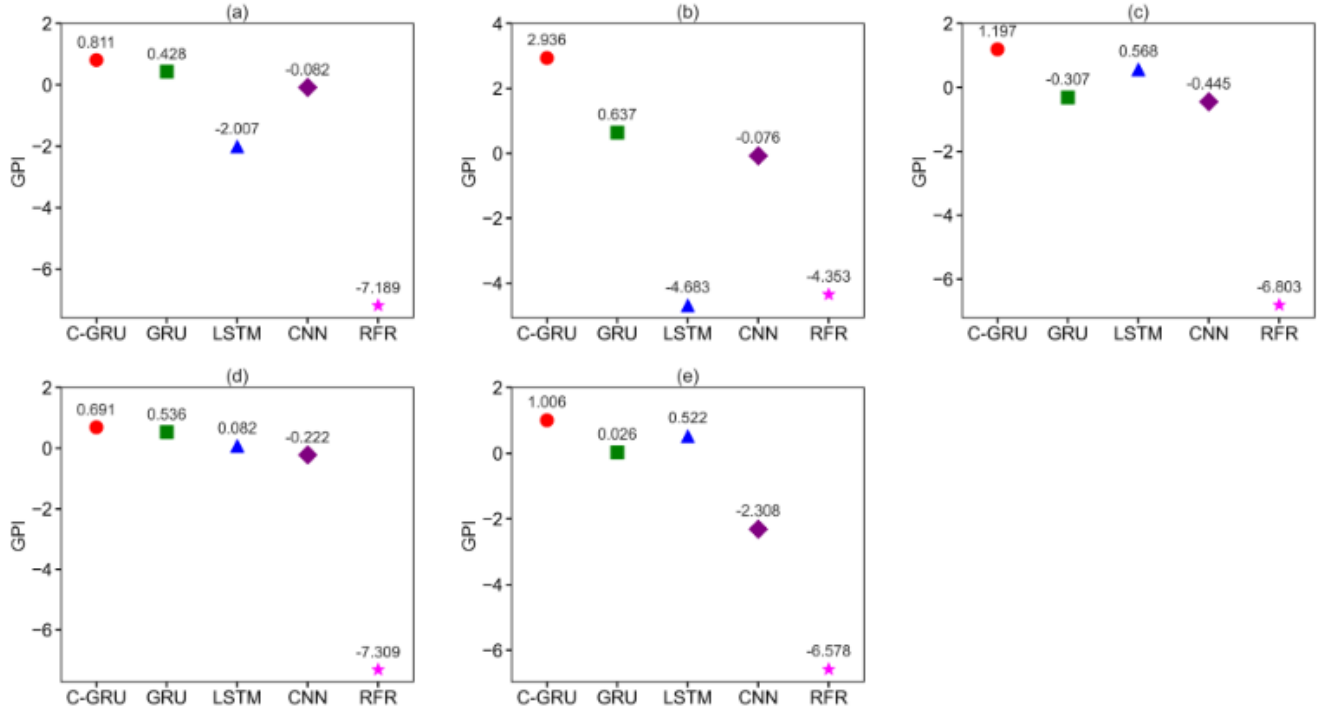

Fig. 10: Evaluation of the proposed objective model (i.e., C-GRU) vs. all other comparative models using the Global Performance Indicator (GPI) for the five study sites: (a) Lautoka, (b) Nadi, (c) Rakiraki, (d) Sigatoka, and (e) Tavua, in the testing phase

In addition to the model evaluation conducted thus far, we further assess our objective model using empirical cumulative distribution functions ($ECDF$) of absolute forecast error ($|FE|$), as depicted in Fig. 11. Ideally, the $|FE|$ value should be close to zero for the best-performing model, with its distribution closely clustered around zero. The $ECDF$ plot of $|FE|$ across all the study sites revealed that while the profiles for the benchmarking models exhibited similarities, the proposed hybrid C-GRU model demonstrated a distinctly narrower distribution across all the study sites. This indicates that forecast errors for the proposed C-GRU model consistently registered minimal spreads in forecasting errors, which were closer to 0, indicating superior forecasting accuracy compared to the other models evaluated.

A detailed analysis of Fig. 11 revealed that for $|FE| < 0.05$, the proposed hybrid C-GRU model registered significantly high percentages across all the study sites: $\approx 99.5\%$ for the Lautoka and Nadi sites, $\approx 99.8\%$ for the Rakiraki site, $\approx 98.9\%$ for the Sigatoka site, and $\approx 99.9\%$ for the Tavua site. In contrast, the standalone GRU model registered a slightly lower percentage, with $\approx 99\%$ for the Lautoka site, $\approx 98.9\%$ for the Nadi site, $\approx 98.4\%$ for the Rakiraki site, $\approx 98.6\%$ for the Sigatoka site, and $\approx 99.6\%$ for the Tavua site. The RFR model, however, recorded the lowest percentage for $|FE| < 0.05$, with $\approx 92.5\%$ for the Lautoka site, $\approx 96.4\%$ for the Nadi site, $\approx 92.8\%$ for the Rakiraki site, $\approx 89.8\%$ for the Sigatoka site, and $\approx 95.5\%$ for the Tavua site. These findings provide additional evidence supporting the efficiency of the proposed hybrid C-GRU model over the benchmark models.

Lastly, the Diebold–Mariano (DM) (Diebold and Mariano, 2002) statistical test was used to determine whether the proposed hybrid C-GRU model's performance is statistically significantly better than that of the benchmark models. The null hypothesis (H_0) and the alternative hypothesis (H_A) of the DM test were set as follows (Prasad et al., 2022, 2024): H_0: There is no significant observed difference between the performances of the two predictive models, and H_A: the observed difference is significant. The DM test was performed at a 5% level of significance such that we reject H_0 if the DM test statistic is > 1.96 or < -1.96. The outcomes of the DM test for all study sites are presented in Table 5, where the DM test statistic is consistently less than -1.96

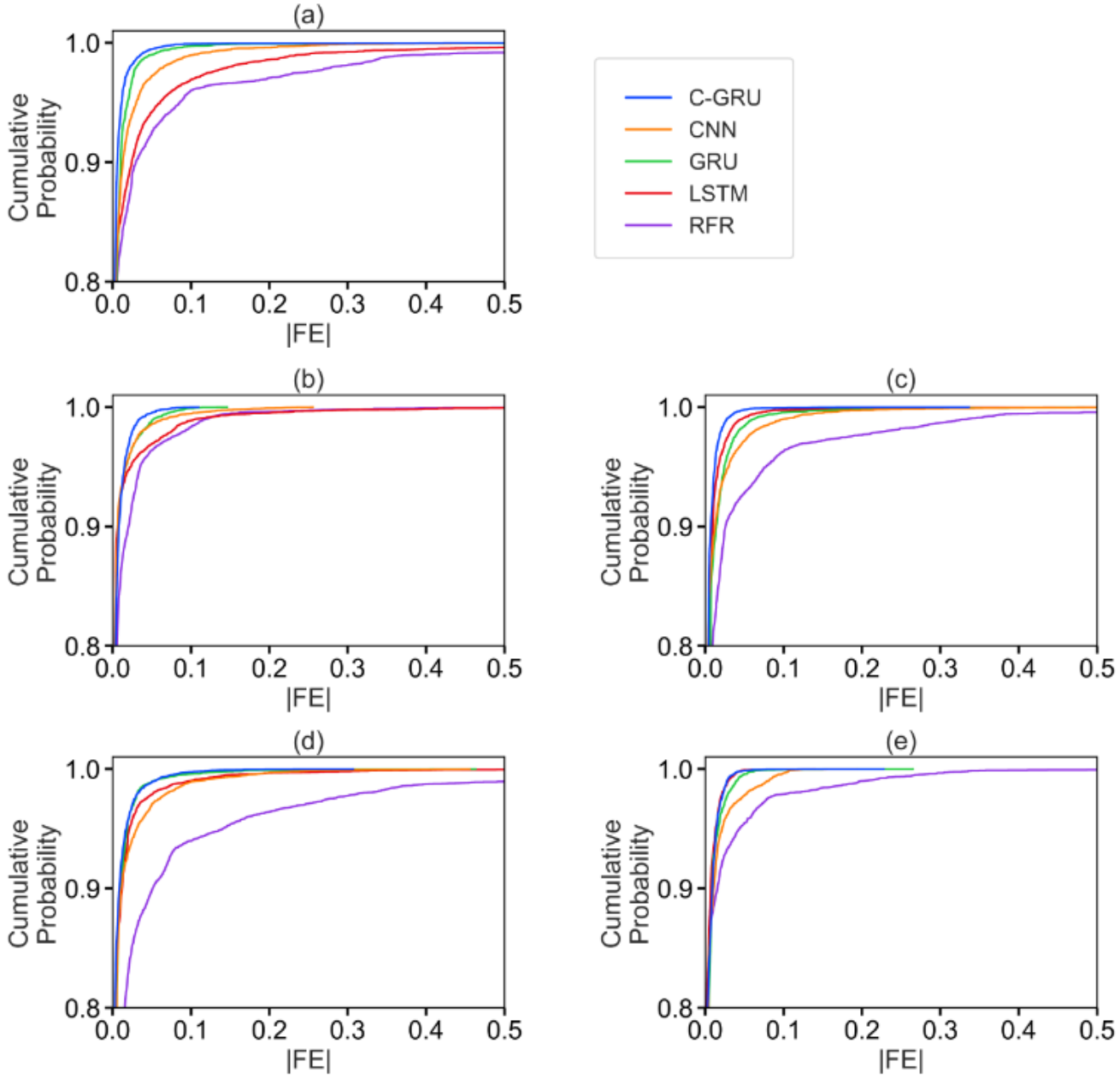

Fig. 11: Empirical cumulative distribution function ($ECDF$) of the absolute forecasting error ($|FE|$) generated by the proposed C-GRU vs. CNN, GRU, LSTM and RFR models for the five study sites: (a) Lautoka, (b) Nadi, (c) Rakiraki, (d) Sigatoka, and (e) Tavua, in the testing phase

across all the study sites. Consequently, we reject H_0, concluding that the proposed hybrid C-GRU model has demonstrated higher forecasting accuracy than all the benchmark models across all the study sites. Hence, we assert that the our proposed hybrid C-GRU model is optimal for forecasting $SWRI_{24-hr-S}$ over a 1-hourly forecast horizon.

Overall, the results demonstrate the robustness and efficiency of the proposed hybrid C-GRU model in forecasting $SWRI_{24-hr-S}$ over a 1-hourly forecast horizon compared to its counterparts. Various performance evaluation metrics, diagnostic plots, and statistical tests were employed to comprehensively assess the proposed model's performance and compare it with all benchmark models. The findings reveal that our proposed hybrid C-GRU model demonstrated superior performance, achieving high R^2 values and showing excellent agreement between forecasted and observed $SWRI_{24-hr-S}$ values, as evidenced by the scatterplot across all the study sites. The proposed model also registered high r and very low MAE and $RMSE$ values. The results also demonstrated that the proposed model achieved significant reductions in $sMAPE$ compared to

Table 5: Evaluation of the proposed hybrid CGRU model with benchmark models using Diebold–Mariano (DM) test at a 5% significance level in the testing phase across all the study sites

Site	DM test	C-GRU vs. GRU	C-GRU vs. LSTM	C-GRU vs. CNN	C-GRU vs. RFR
Lautoka	DM test statistic	-11.329	-16.675	-14.386	-14.326
	p-value	0.000	0.000	0.000	0.000
	H_o	Reject	Reject	Reject	Reject
Nadi	DM test statistic	-34.999	-18.430	-18.240	-18.630
	p-value	0.000	0.000	0.000	0.000
	H_o	Reject	Reject	Reject	Reject
Rakiraki	DM test statistic	-33.220	-15.046	-16.551	-15.686
	p-value	0.000	0.000	0.000	0.000
	H_o	Reject	Reject	Reject	Reject
Sigatoka	DM test statistic	-9.683	-6.459	-31.061	-17.838
	p-value	0.000	0.000	0.000	0.000
	H_o	Reject	Reject	Reject	Reject
Tavua	DM test statistic	-37.031	-18.718	-37.070	-14.890
	p-value	0.000	0.000	0.000	0.000
	H_o	Reject	Reject	Reject	Reject

all benchmarking models across all study sites. The proposed hybrid C-GRU model's superiority was highlighted by its higher values on the most stringent and reliable metric, E_{LM}, across all the study sites. Moreover, the proposed model consistently attained the highest GPI values across all the study sites, showcasing its superior performance. Analysis of the empirical cumulative distribution functions ($ECDF$) of absolute forecast error ($|FE|$) revealed that the proposed hybrid C-GRU model exhibited significantly higher percentages ($\approx 98.9 - 99.9\%$) within smaller error brackets (i.e., $|FE| < 0.05$) across all study sites, highlighting its superior forecasting accuracy compared to other evaluated models. Additionally, the DM statistical test also confirmed the efficiency of the proposed hybrid C-GRU model over the benchmark models.

The superior performance of the proposed hybrid C-GRU model primarily stems from its integrated architecture, combining CNN and GRU layers. Within the C-GRU algorithm, CNN layers extract crucial features from the input data while minimising redundant information. Subsequently, the salient feature map produced by the CNN is fed to the GRU layers, effectively capturing both past and future long-term dependencies in the historical sequential data. The results also showed that the RFR model exhibited poor performance among all forecasting models due to its limited capability to capture complex patterns in the time series data. Specifically, the results indicated that the RFR model could not accurately forecast the value of $SWRI_{24-hr-S} > 0$ across all study sites, which is a crucial aspect in this study as $SWRI_{24-hr-S} > 0$ signifies a flood situation.

4.2 Practical application of the proposed framework

Considering the promising forecasting results of the proposed hybrid C-GRU model, we have further exemplified its potential for real-life application in the decision support system for early flood warnings, as illustrated in Fig. 12.

The proposed system is designed to operate through both offline and online systems. The online system utilises the optimal pre-trained hybrid C-GRU model to forecast $SWRI_{24-hr-S}$ over a 1-hourly forecast horizon using the new input data as it becomes available. An expert end-user, preferably a flood forecaster, interprets the results from the online system. After reviewing the forecasted $SWRI_{24-hr-S}$ value, if $SWRI_{24-hr-S} > 0$, indicating a potential flood situation, the public will be promptly informed about the risk of floods. Concurrently, the proposed hybrid C-GRU model is continuously trained and fine-tuned through the offline system using historical data from the database. The updated model from the offline system periodically replaces the pre-trained online model, ensuring accurate and reliable forecasts are consistently generated. This decision support system can be implemented in Fiji's major towns and cities. It aims to enhance

and strengthen Fiji's real-time monitoring and early warning systems for floods, thus improving disaster preparedness, mitigation, and response efforts.

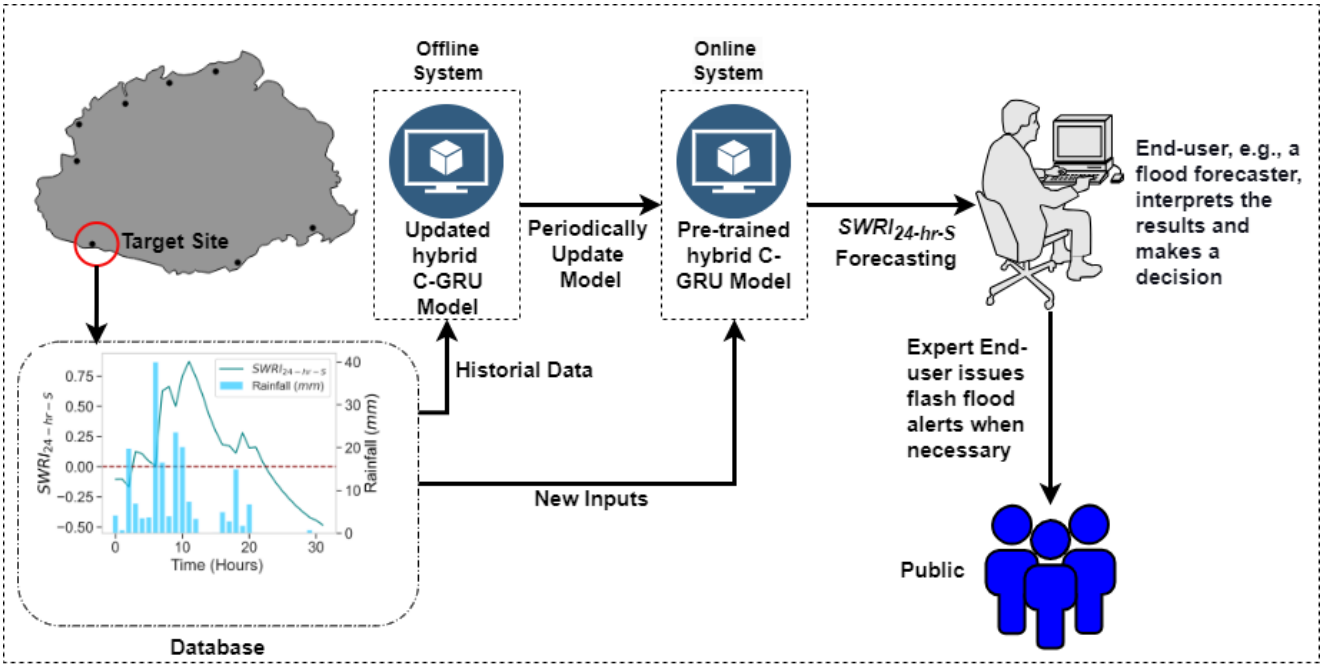

Fig. 12: Schematic representation of the proposed hybrid C-GRU model for practical application in the decision support system for early flood warnings

5 Conclusions, limitation and future research directions

This study supports ongoing efforts to enhance early flood warning systems by developing an accurate and reliable flood forecasting model for near real-time forecasting.

In this study, we proposed a hybrid C-GRU model, integrating two powerful deep learning algorithms, i.e., CNN integrated with GRU, to forecast the proposed hourly flood index ($SWRI_{24-hr-S}$) over a 1-hourly forecast horizon to assess flood risk at an hourly scale at five flood-prone sites in Fiji. The proposed model is trained using statistically significant lagged values of $SWRI_{24-hr-S}$ and real-time hourly rainfall data for each study site. Bayesian optimisation (BO) is utilised to efficiently optimise the hyperparameters of the proposed model. The performance of the proposed hybrid C-GRU model is compared with other benchmarking models, including CNN, GRU, LSTM and RFR. Various performance evaluation metrics and diagnostic plots confirm the excellent forecasting capability of the proposed hybrid C-GRU model compared to other counterpart models.

The results demonstrate that the proposed model outperforms all benchmarking models with substantial reductions in $sMAPE$ observed across all study sites. The proposed model also consistently achieved the highest GPI values across all sites. It also registered the largest percentage of forecast errors ($\approx 98.9 - 99.9\%$) within smaller error brackets (i.e., $|FE| < 0.05$) amongst all evaluated sites. Furthermore, the DM statistical test confirmed the efficiency of the proposed hybrid C-GRU model over the benchmark models. Moreover, the practical implementation of the proposed framework in a decision support system for early flood warnings is demonstrated, showcasing its potential to enhance Fiji's real-time monitoring and early warning systems for floods, thereby improving disaster preparedness, mitigation, and response efforts.

Nonetheless, it is crucial to recognise that the methodologies proposed in this study have certain limitations, and addressing these limitations is a potential future research direction. These are as follows:

1. This research was the first to use the proposed $SWRI_{24-hr-S}$ and develop the hybrid DL algorithm for hourly flood forecasting. However, a comprehensive study must validate the

proposed $SWRI_{24-hr-S}$ for its broader adoption as an index-based flood risk monitoring tool for Fiji and other flood-prone regions globally. This is contingent upon the availability of well-documented flood records for validation and hourly rainfall data.

2. Several factors, including land topography, land use changes, soil conditions, and catchment and drainage systems, are crucial for effective flood risk management and mitigation strategies. The proposed model was trained using only antecedent real-time hourly rainfall and $SWRI_{24-hr-S}$ data. Despite incorporating only two features, the forecasting performance was excellent. However, in future studies, enhancing the model's robustness is recommended by identifying and incorporating additional useful features.

3. In this study, a single-step forecasting strategy was utilised, which does not forecast $SWRI_{24-hr-S}$ at a longer forecast horizon than 1 hour. Forecasting $SWRI_{24-hr-S}$ with sufficient lead times is crucial. This is paramount for early warning systems, ensuring the effective implementation of flood mitigation strategies and better preparedness for flood risk. Consequently, given the excellent performance of the proposed hybrid C-GRU model at single-step forecasting, in future studies, a multiple-input multiple-output ($MIMO$) strategy, as demonstrated in the related study by Moishin et al. (2021a), should be tested to forecast $SWRI_{24-hr-S}$ at a longer forecast horizon.

4. The single-step outputs from the hybrid C-GRU model are depicted as point forecasts. Nonetheless, it is recommended that future studies delve into interval and probabilistic forecasting methodologies. These approaches offer the ability to estimate both the potential future range of $SWRI_{24-hr-S}$ values and the uncertainty associated with the forecasts.

5. The study does not test model hybridisation through data decomposition strategies. To further enhance the proposed model's performance, it is recommended that multivariate decomposition methods such as MEMD, as demonstrated in a related study by Prasad et al. (2021), and stationary wavelet transform (SWT), etc., be explored. In addition, other DL, such as the BiLSTM algorithm and hybrid algorithms, should also be tested for $SWRI_{24-hr-S}$ forecasting. The results of this study can serve as a benchmark for new models.

6. The proposed hybrid C-GRU is presented as a non-interpretable "black-box" model. Therefore, in future studies, it would be beneficial to utilise model-agnostic eXplainable Artificial Intelligence (xAI) methods like Local Interpretable Model-Agnostic Explanations (LIME) and SHapley Additive exPlanations (SHAP) to gain insight into the underlying mechanism of this proposed black-box model.

7. The proposed model's practical application in a decision support system for early flood warnings has been demonstrated. However, finding expertise to implement these advanced techniques within relevant organisations is a significant challenge. Therefore, it is recommended that future studies focus on developing more user-friendly tools, perhaps leveraging platforms like "Streamlit" for enhanced accessibility and usability.

Finally, despite these limitations, our proposed methodologies can be considered a viable and cost-effective tool for hourly flood forecasting in Fiji and can be applied to other flood-prone regions worldwide.

Acknowledgements. The first author, Ravinesh Chand is an Australia Awards Scholar supported by the Australian Government Department of Foreign Affairs and Trade (DFAT). He thanks DFAT for funding this study via the Australia Awards Scholarship scheme 2022. The views and opinions expressed in this paper belong to the authors and do not represent the views of the Australian Government or the Fiji Meteorological Services. The authors are also grateful to Fiji Meteorological Services for providing the rainfall data needed for this study.

Author Contribution. Conceptualization; Data curation; Formal analysis; Investigation; Methodology; Project administration; Resources; Software; Validation; Visualization; Writing - original draft; Writing - review & editing: **Ravinesh Chand**. Resources; Methodology; Project administration; Supervision; Writing - review & editing: **Ravinesh C Deo**. Resources; Methodology; Supervision; Writing - review & editing: **Sujan Ghimire**. Writing - review & editing: **Thong Nguyen-Huy**. Writing - review & editing: **Mumtaz Ali**.

Declarations

Conflict of interest. The authors declare no conflict of interests.

References

Abadi, M., Barham, P., Chen, J., Chen, Z., Davis, A., Dean, J., Devin, M., Ghemawat, S., Irving, G., Isard, M., *et al.*: Tensorflow: A system for large-scale machine learning. In: 12th USENIX Symposium on Operating Systems Design and Implementation (OSDI 2016), pp. 265–283 (2016)

Ahmed, A.M., Deo, R.C., Feng, Q., Ghahramani, A., Raj, N., Yin, Z., Yang, L.: Hybrid deep learning method for a week-ahead evapotranspiration forecasting. Stochastic Environmental Research and Risk Assessment, 1–19 (2021) https://doi.org/10.1007/s00477-021-02078-x

Ahmed, A.M., Farheen, S., Nguyen-Huy, T., Raj, N., Jui, S.J.J., Farzana, S.: Real-time prediction of the week-ahead flood index using hybrid deep learning algorithms with synoptic climate mode indices [preprint]. Research Square (2023) https://doi.org/10.21203/rs.3.rs-2654880/v1

Almulihi, A., Saleh, H., Hussien, A.M., Mostafa, S., El-Sappagh, S., Alnowaiser, K., Ali, A.A., Refaat Hassan, M.: Ensemble learning based on hybrid deep learning model for heart disease early prediction. Diagnostics **12**(12), 3215 (2022) https://doi.org/10.3390/diagnostics12123215

Alexander, A.A., Thampi, S.G., NR, C.: Development of hybrid wavelet-ann model for hourly flood stage forecasting. ISH Journal of Hydraulic Engineering **24**(2), 266–274 (2018) https://doi.org/10.1080/09715010.2017.1422192

Bergstra, J., Bengio, Y.: Random search for hyper-parameter optimization. Journal of Machine Learning Research **13**(2) (2012)

Bergstra, J., Bardenet, R., Bengio, Y., Kégl, B.: Algorithms for hyper-parameter optimization. Advances in neural information processing systems **24** (2011)

Behar, O., Khellaf, A., Mohammedi, K.: Comparison of solar radiation models and their validation under algerian climate–the case of direct irradiance. Energy Conversion and Management **98**, 236–251 (2015) https://doi.org/10.1016/j.enconman.2015.03.067

Breiman, L.: Random forests. Machine learning **45**, 5–32 (2001) https://doi.org/10.1023/A:1010933404324

Chung, J., Gulcehre, C., Cho, K., Bengio, Y.: Empirical evaluation of gated recurrent neural networks on sequence modeling. arXiv preprint arXiv:1412.3555 (2014)

Cheung, Y.-W., Lai, K.S.: Lag order and critical values of the augmented dickey–fuller test. Journal of Business & Economic Statistics **13**(3), 277–280 (1995) https://doi.org/10.1080/07350015.1995.10524601

CRED: The human cost of natural disasters 2015: A global perspective. Report, Centre of Research on Epidemiology of Disasters (CRED) (2015). https://www.preventionweb.net/files/42895_cerdthehumancostofdisastersglobalpe.pdf

Cho, K., Van Merriënboer, B., Gulcehre, C., Bahdanau, D., Bougares, F., Schwenk, H., Bengio, Y.: Learning phrase representations using rnn encoder-decoder for statistical machine translation. arXiv preprint arXiv:1406.1078 (2014)

Chen, W., Xie, X., Wang, J., Pradhan, B., Hong, H., Bui, D.T., Duan, Z., Ma, J.: A comparative study of logistic model tree, random forest, and classification and regression tree models for spatial prediction of landslide susceptibility. Catena **151**, 147–160 (2017) https://doi.org/10.1016/j.catena.2016.11.032

Deo, R.C., Adamowski, J.F., Begum, K., Salcedo-Sanz, S., Kim, D.-W., Dayal, K.S., Byun, H.-R.: Quantifying flood events in bangladesh with a daily-step flood monitoring index based on the concept of daily effective precipitation. Theoretical and applied climatology **137**, 1201–1215 (2019) https://doi.org/10.1007/s00704-018-2657-4

Deo, R.C., Byun, H.-R., Adamowski, J.F., Kim, D.-W.: A real-time flood monitoring index based on daily effective precipitation and its application to brisbane and lockyer valley flood events. Water Resources Management **29**, 4075–4093 (2015) https://doi.org/10.1007/s11269-015-1046-3

Deo, R.C., Byun, H.-R., Kim, G.-B., Adamowski, J.F.: A real-time hourly water index for flood risk monitoring: Pilot studies in brisbane, australia, and dobong observatory, south korea. Environmental monitoring and assessment **190**, 1–27 (2018) https://doi.org/10.1007/s10661-018-6806-0

Diebold, F.X., Mariano, R.S.: Comparing predictive accuracy. Journal of Business & economic statistics **20**(1), 134–144 (2002) https://doi.org/10.1198/073500102753410444

Dua, N., Singh, S.N., Semwal, V.B., Challa, S.K.: Inception inspired cnn-gru hybrid network for human activity recognition. Multimedia Tools and Applications **82**(4), 5369–5403 (2023) https://doi.org/10.1007/s11042-021-11885-x

Eggensperger, K., Feurer, M., Hutter, F., Bergstra, J., Snoek, J., Hoos, H., Leyton-Brown, K., *et al.*: Towards an empirical foundation for assessing bayesian optimization of hyperparameters. In: NIPS Workshop on Bayesian Optimization in Theory and Practice, vol. 10 (2013)

Feurer, M., Hutter, F.: Hyperparameter optimization. Automated machine learning: Methods, systems, challenges, 3–33 (2019)

Fiji Meteorological Service: Fiji annual climate summary 2016. Report, Fiji Meteorological Service (2016). https://www.met.gov.fj/Summary2.pdf

Feresi, J., Kenny, G.J., Wet, N., Limalevu, L., Bhusan, J., Ratukalou, I.: Climate change vulnerability and adaptation assessment for fiji. Report, The International Global Change Institute (IGCI), University of Waikato. (2000). https://researchcommons.waikato.ac.nz/items/6148f4a7-2b7d-4e9c-bc86-baa523e4e114

Ghimire, S., Bhandari, B., Casillas-Perez, D., Deo, R.C., Salcedo-Sanz, S.: Hybrid deep cnn-svr algorithm for solar radiation prediction problems in queensland, australia. Engineering Applications of Artificial Intelligence **112**, 104860 (2022) https://doi.org/10.1016/j.engappai.2022.104860

Ghimire, S., Deo, R.C., Casillas-Pérez, D., Salcedo-Sanz, S., Sharma, E., Ali, M.: Deep learning cnn-lstm-mlp hybrid fusion model for feature optimizations and daily solar radiation prediction. Measurement **202**, 111759 (2022) https://doi.org/10.1016/j.measurement.2022.111759

Ghimire, S., Deo, R.C., Casillas-Perez, D., Salcedo-Sanz, S.: Improved complete ensemble empirical mode decomposition with adaptive noise deep residual model for short-term multi-step solar radiation prediction. Renewable Energy **190**, 408–424 (2022) https://doi.org/10.1016/j.renene.2022.03.120

Ghimire, S., Deo, R.C., Casillas-Pérez, D., Salcedo-Sanz, S.: Efficient daily electricity demand prediction with hybrid deep-learning multi-algorithm approach. Energy Conversion and Management **297**, 117707 (2023) https://doi.org/10.1016/j.enconman.2023.117707

Ghimire, S., Deo, R.C., Casillas-Pérez, D., Salcedo-Sanz, S.: Electricity demand error corrections with attention bi-directional neural networks. Energy **291**, 129938 (2024) https://doi.org/10.1016/j.energy.2023.129938

Ghimire, S., Deo, R.C., Casillas-Pérez, D., Salcedo-Sanz, S.: Two-step deep learning framework with error compensation technique for short-term, half-hourly electricity price forecasting. Applied Energy **353**, 122059 (2024) https://doi.org/10.1016/j.apenergy.2023.122059

Ghimire, S., Deo, R.C., Raj, N., Mi, J.: Deep solar radiation forecasting with convolutional neural network and long short-term memory network algorithms. Applied Energy **253**, 113541 (2019) https://doi.org/10.1016/j.apenergy.2019.113541

Gupta, H.V., Kling, H., Yilmaz, K.K., Martinez, G.F.: Decomposition of the mean squared error and nse performance criteria: Implications for improving hydrological modelling. Journal of Hydrology **377**(1-2), 80–91 (2009) https://doi.org/10.1016/j.jhydrol.2009.08.003

Gholami, H., Mohammadifar, A.: Novel deep learning hybrid models (cnn-gru and dldl-rf) for the susceptibility classification of dust sources in the middle east: a global source. Scientific Reports **12**(1), 19342 (2022) https://doi.org/10.1038/s41598-022-24036-5

Ghimire, S., Nguyen-Huy, T., Deo, R.C., Casillas-Perez, D., Salcedo-Sanz, S.: Efficient daily solar radiation prediction with deep learning 4-phase convolutional neural network, dual stage stacked regression and support vector machine cnn-regst hybrid model. Sustainable Materials and Technologies **32**, 00429 (2022) https://doi.org/10.1016/j.susmat.2022.e00429

Ghimire, S., Nguyen-Huy, T., Prasad, R., Deo, R.C., Casillas-Pérez, D., Salcedo-Sanz, S., Bhandari, B.: Hybrid convolutional neural network-multilayer perceptron model for solar radiation prediction. Cognitive Computation **15**(2), 645–671 (2023) https://doi.org/10.1007/s12559-022-10070-y

Government of Fiji: Climate vulnerability assessment: Making fiji climate resilient. Report, Government of Fiji (2017). https://reliefweb.int/report/fiji/climate-vulnerability-assessment-making-fiji-climate-resilient

Graves, A.: Generating sequences with recurrent neural networks. arXiv preprint arXiv:1308.0850 (2013)

Hochreiter, S., Schmidhuber, J.: Long short-term memory. Neural computation **9**(8), 1735–1780 (1997) https://doi.org/10.1162/neco.1997.9.8.1735

Hapuarachchi, H., Wang, Q., Pagano, T.: A review of advances in flash flood forecasting. Hydrological processes **25**(18), 2771–2784 (2011) https://doi.org/10.1002/hyp.8040

Joseph, L.P., Deo, R.C., Casillas-Pérez, D., Prasad, R., Raj, N., Salcedo-Sanz, S.: Short-term wind speed forecasting using an optimized three-phase convolutional neural network fused with bidirectional long short-term memory network model. Applied Energy **359**, 122624 (2024) https://doi.org/10.1016/j.apenergy.2024.122624

Joseph, L.P., Deo, R.C., Prasad, R., Salcedo-Sanz, S., Raj, N., Soar, J.: Near real-time wind speed forecast model with bidirectional lstm networks. Renewable Energy **204**, 39–58 (2023) https://doi.org/10.1016/j.renene.2022.12.123

Ji, L., Fu, C., Ju, Z., Shi, Y., Wu, S., Tao, L.: Short-term canyon wind speed prediction based on cnn—gru transfer learning. Atmosphere **13**(5), 813 (2022) https://doi.org/10.3390/atmos13050813

Kingma, D.P., Ba, J.: Adam: A method for stochastic optimization. arXiv preprint arXiv:1412.6980 (2014)

Krause, P., Boyle, D., Bäse, F.: Comparison of different efficiency criteria for hydrological model assessment. Advances in geosciences **5**, 89–97 (2005) https://doi.org/10.5194/adgeo-5-89-2005

Komer, B., Bergstra, J., Eliasmith, C.: Hyperopt-sklearn: Automatic hyperparameter configuration for scikit-learn. In: Scipy, pp. 32–37 (2014)

Ketkar, N.: Introduction to Keras, pp. 97–111. Apress, Berkeley, CA (2017). https://doi.org/10.1007/978-1-4842-2766-4_7

Kling, H., Fuchs, M., Paulin, M.: Runoff conditions in the upper danube basin under an ensemble of climate change scenarios. Journal of Hydrology **424**, 264–277 (2012) https://doi.org/10.1016/j.jhydrol.2012.01.011

Kisvari, A., Lin, Z., Liu, X.: Wind power forecasting–a data-driven method along with gated recurrent neural network. Renewable Energy **163**, 1895–1909 (2021) https://doi.org/10.1016/j.renene.2020.10.119

Kabir, S., Patidar, S., Xia, X., Liang, Q., Neal, J., Pender, G.: A deep convolutional neural network model for rapid prediction of fluvial flood inundation. Journal of Hydrology **590**, 125481 (2020) https://doi.org/10.1016/j.jhydrol.2020.125481

Kant, A., Suman, P.K., Giri, B.K., Tiwari, M.K., Chatterjee, C., Nayak, P.C., Kumar, S.: Comparison of multi-objective evolutionary neural network, adaptive neuro-fuzzy inference system and bootstrap-based neural network for flood forecasting. Neural Computing and Applications **23**, 231–246 (2013) https://doi.org/10.1007/s00521-013-1344-8

Kumar, R., Stephens, M., Weir, T.: Rainfall trends in fiji. International Journal of Climatology **34**(5), 1501–1510 (2014) https://doi.org/10.1002/joc.3779

LeCun, Y., Bengio, Y., Hinton, G.: Deep learning. Nature **521**(7553), 436–444 (2015) https://doi.org/10.1038/nature14539

Li, X.: Cnn-gru model based on attention mechanism for large-scale energy storage optimization in smart grid. Frontiers in Energy Research **11**, 1228256 (2023) https://doi.org/10.3389/fenrg.2023.1228256

Legates, D.R., McCabe Jr, G.J.: Evaluating the use of "goodness-of-fit" measures in hydrologic and hydroclimatic model validation. Water resources research **35**(1), 233–241 (1999) https://doi.org/10.1029/1998WR900018

Lu, E.: Determining the start, duration, and strength of flood and drought with daily precipitation: Rationale. Geophysical research letters **36**(12) (2009) https://doi.org/10.1029/2009GL038817

Liaw, A., Wiener, M., *et al.*: Classification and regression by randomforest. R news **2**(3), 18–22 (2002)

947 Moishin, M., Deo, R.C., Prasad, R., Raj, N., Abdulla, S.: Designing deep-based learning flood
948 forecast model with convlstm hybrid algorithm. IEEE Access **9**, 50982–50993 (2021) https:
949 //doi.org/10.1109/ACCESS.2021.3065939

950 Moishin, M., Deo, R.C., Prasad, R., Raj, N., Abdulla, S.: Development of flood monitoring index
951 for daily flood risk evaluation: case studies in fiji. Stochastic Environmental Research and Risk
952 Assessment **35**, 1387–1402 (2021) https://doi.org/10.1007/s00477-020-01899-6

953 Miau, S., Hung, W.-H.: River flooding forecasting and anomaly detection based on deep learning.
954 IEEE Access **8**, 198384–198402 (2020) https://doi.org/10.1109/ACCESS.2020.3034875

955 McGree, S., Yeo, S.W., Devi, S.: Flooding in the fiji islands between 1840 and 2009. Risk Frontiers,
956 1–69 (2010)

957 Nguyen-Huy, T., Deo, R.C., Yaseen, Z.M., Prasad, R., Mushtaq, S.: Bayesian markov chain monte
958 carlo-based copulas: factoring the role of large-scale climate indices in monthly flood prediction.
959 Intelligent data analytics for decision-support systems in hazard mitigation: Theory and practice
960 of hazard mitigation, 29–47 (2021) https://doi.org/10.1007/978-981-15-5772-9_2

961 Nguyen-Huy, T., Kath, J., Nagler, T., Khaung, Y., Aung, T.S.S., Mushtaq, S., Marcussen, T.,
962 Stone, R.: A satellite-based standardized antecedent precipitation index (sapi) for mapping
963 extreme rainfall risk in myanmar. Remote Sensing Applications: Society and Environment **26**,
964 100733 (2022) https://doi.org/10.1016/j.rsase.2022.100733

965 Nguyen, H.-P., Liu, J., Zio, E.: A long-term prediction approach based on long short-term memory
966 neural networks with automatic parameter optimization by tree-structured parzen estimator
967 and applied to time-series data of npp steam generators. Applied Soft Computing **89**, 106116
968 (2020) https://doi.org/10.1016/j.asoc.2020.106116

969 Nevo, S., Morin, E., Gerzi Rosenthal, A., Metzger, A., Barshai, C., Weitzner, D., Voloshin, D.,
970 Kratzert, F., Elidan, G., Dror, G., *et al.*: Flood forecasting with machine learning models in
971 an operational framework. Hydrology and Earth System Sciences **26**(15), 4013–4032 (2022)
972 https://doi.org/10.5194/hess-26-4013-2022

973 Nash, J.E., Sutcliffe, J.V.: River flow forecasting through conceptual models part i—a discussion of
974 principles. Journal of Hydrology **10**(3), 282–290 (1970) https://doi.org/10.1016/0022-1694(70)
975 90255-6

976 Nosrati, K., Saravi, M.M., Shahbazi, A.: Investigation of flood event possibility over iran using
977 flood index. Survival and sustainability: environmental concerns in the 21st century, 1355–1361
978 (2011) https://doi.org/10.1007/978-3-540-95991-5_127

979 Oriani, F., Stisen, S., Demirel, M.C., Mariethoz, G.: Missing data imputation for multisite rainfall
980 networks: a comparison between geostatistical interpolation and pattern-based estimation on
981 different terrain types. Journal of Hydrometeorology **21**(10), 2325–2341 (2020) https://doi.org/
982 10.1175/JHM-D-19-0220.1

983 Prasad, R., Charan, D., Joseph, L., Nguyen-Huy, T., Deo, R.C., Singh, S.: Daily flood fore-
984 casts with intelligent data analytic models: multivariate empirical mode decomposition-based
985 modeling methods. Intelligent Data Analytics for Decision-Support Systems in Hazard Mit-
986 igation: Theory and Practice of Hazard Mitigation, 359–381 (2021) https://doi.org/10.1007/
987 978-981-15-5772-9_17

988 Prasad, S.S., Deo, R.C., Downs, N., Igoe, D., Parisi, A.V., Soar, J.: Cloud affected solar uv predic-
989 tion with three-phase wavelet hybrid convolutional long short-term memory network multi-step
990 forecast system. IEEE Access **10**, 24704–24720 (2022) https://doi.org/10.1109/ACCESS.2022.

3153475

Prasad, S.S., Deo, R.C., Downs, N.J., Casillas-Pérez, D., Salcedo-Sanz, S., Parisi, A.V.: Very short-term solar ultraviolet-a radiation forecasting system with cloud cover images and a bayesian optimized interpretable artificial intelligence model. Expert Systems with Applications **236**, 121273 (2024) https://doi.org/10.1016/j.eswa.2023.121273

Prasad, R., Deo, R.C., Li, Y., Maraseni, T.: Weekly soil moisture forecasting with multivariate sequential, ensemble empirical mode decomposition and boruta-random forest hybridizer algorithm approach. Catena **177**, 149–166 (2019) https://doi.org/10.1016/j.catena.2019.02.012

Prasad, S.S., Deo, R.C., Salcedo-Sanz, S., Downs, N.J., Casillas-Pérez, D., Parisi, A.V.: Enhanced joint hybrid deep neural network explainable artificial intelligence model for 1-hr ahead solar ultraviolet index prediction. Computer Methods and Programs in Biomedicine **241**, 107737 (2023) https://doi.org/10.1016/j.cmpb.2023.107737

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., *et al.*: Scikit-learn: Machine learning in python. Journal of Machine Learning Research **12**, 2825–2830 (2011)

Pan, M., Zhou, H., Cao, J., Liu, Y., Hao, J., Li, S., Chen, C.-H.: Water level prediction model based on gru and cnn. IEEE Access **8**, 60090–60100 (2020) https://doi.org/10.1109/ACCESS.2020.2982433

Roberts, W., Williams, G.P., Jackson, E., Nelson, E.J., Ames, D.P.: Hydrostats: A python package for characterizing errors between observed and predicted time series. Hydrology **5**(4), 66 (2018) https://doi.org/10.3390/hydrology5040066

Sarker, I.H.: Deep learning: a comprehensive overview on techniques, taxonomy, applications and research directions. SN Computer Science **2**(6), 420 (2021) https://doi.org/10.1007/s42979-021-00815-1

Sharma, E., Deo, R.C., Soar, J., Prasad, R., Parisi, A.V., Raj, N.: Novel hybrid deep learning model for satellite based pm10 forecasting in the most polluted australian hotspots. Atmospheric Environment **279**, 119111 (2022) https://doi.org/10.1016/j.atmosenv.2022.119111

Sajjad, M., Khan, Z.A., Ullah, A., Hussain, T., Ullah, W., Lee, M.Y., Baik, S.W.: A novel cnn-gru-based hybrid approach for short-term residential load forecasting. IEEE Access **8**, 143759–143768 (2020) https://doi.org/10.1109/ACCESS.2020.3009537

Shih, D.-H., Liao, C.-H., Wu, T.-W., Xu, X.-Y., Shih, M.-H.: Dysarthria speech detection using convolutional neural networks with gated recurrent unit. In: Healthcare, vol. 10, p. 1956 (2022). https://doi.org/10.3390/healthcare10101956 . MDPI

Tiwari, M.K., Chatterjee, C.: Development of an accurate and reliable hourly flood forecasting model using wavelet–bootstrap–ann (wbann) hybrid approach. Journal of Hydrology **394**(3-4), 458–470 (2010) https://doi.org/10.1016/j.jhydrol.2010.10.001

Teng, J., Jakeman, A.J., Vaze, J., Croke, B.F., Dutta, D., Kim, S.: Flood inundation modelling: A review of methods, recent advances and uncertainty analysis. Environmental modelling & software **90**, 201–216 (2017) https://doi.org/10.1016/j.envsoft.2017.01.006

Wang, Z., Dong, Y., Liu, W., Ma, Z.: A novel fault diagnosis approach for chillers based on 1-d convolutional neural network and gated recurrent unit. Sensors **20**(9), 2458 (2020) https://doi.org/10.3390/s20092458

Willmott, C.J.: On the validation of models. Physical geography **2**(2), 184–194 (1981) https://doi.org/10.1080/02723646.1981.10642213

Willmott, C.J.: On the evaluation of model performance in physical geography. Spatial statistics and models, 443–460 (1984) https://doi.org/10.1007/978-94-017-3048-8_23

Yu, J., Zhang, X., Xu, L., Dong, J., Zhangzhong, L.: A hybrid cnn-gru model for predicting soil moisture in maize root zone. Agricultural Water Management **245**, 106649 (2021) https://doi.org/10.1016/j.agwat.2020.106649

Zhang, C., Peng, T., Nazir, M.S.: A novel integrated photovoltaic power forecasting model based on variational mode decomposition and cnn-bigru considering meteorological variables. Electric Power Systems Research **213**, 108796 (2022) https://doi.org/10.1016/j.epsr.2022.108796

Zou, F., Shen, L., Jie, Z., Zhang, W., Liu, W.: A sufficient condition for convergences of adam and rmsprop. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 11127–11135 (2019)

Zhao, Z., Yun, S., Jia, L., Guo, J., Meng, Y., He, N., Li, X., Shi, J., Yang, L.: Hybrid vmd-cnn-gru-based model for short-term forecasting of wind power considering spatio-temporal features. Engineering Applications of Artificial Intelligence **121**, 105982 (2023) https://doi.org/10.1016/j.engappai.2023.105982

Zhao, L., Zhang, Z.: A improved pooling method for convolutional neural networks. Scientific Reports **14**(1), 1589 (2024) https://doi.org/10.1038/s41598-024-51258-6

5.3. Links and implications

Hourly flood forecasting is an essential component of early flood warning systems, providing timely information and supporting the implementation of effective mitigation strategies to protect lives, properties, and the environment. This study uses an AI-based model to design an innovative and cost-effective hourly flood forecasting framework for early warning systems. The outcomes of objective 1 of this research already demonstrated the feasibility of the novel $SWRI_{24\text{-}hr\text{-}S}$ in assessing flood risk on an hourly scale for Fiji's case studies. However, this index cannot predict the flooded state in advance unless a forecasting model is developed and tested. Consequently, the newly developed hybrid C-GRU model demonstrates outstanding performance in forecasting the $SWRI_{24\text{-}hr\text{-}S}$ over a short-term, i.e., 1-hourly forecast horizon. The diverse performance evaluation metrics employed for model comparison strongly affirm the superiority of the hybrid C-GRU model against all benchmark models tested across all five case study stations. Therefore, the newly proposed hybrid C-GRU-based $SWRI_{24\text{-}hr\text{-}S}$ forecasting framework can be integrated into the decision-support system for early flood warnings. This integration will strengthen Fiji's real-time flood monitoring and forecasting capabilities, improving flood risk preparedness, mitigation, and response efforts.

Despite the superior performance of the proposed hybrid C-GRU model in forecasting $SWRI_{24\text{-}hr\text{-}S}$ at a 1-hourly forecast horizon, it suggests several potential directions for future research aimed at broader applications. For instance, the proposed model is trained using only two features, i.e., the lagged $SWRI_{24\text{-}hr\text{-}S}$ and hourly rainfall. However, future studies can identify and include more useful features, such as synoptic-climate indices (Ahmed et al., 2023), river discharge and streamflow data, and hydrometeorological data to enhance its effectiveness. The proposed hybrid C-GRU model is highly parametric. Therefore, the BO with the TPE algorithm is employed to efficiently optimise model hyperparameters. However, future studies can explore using the BOHB technique, which combines Bayesian optimisation (BO) and Hyperband (HB) for more efficient hyperparameter optimisation (Falkner et al., 2018).

Moreover, although the hybrid C-GRU model yields highly accurate forecast results, it is a complex and non-interpretable "black-box" model. To address this issue, it is essential to integrate model-agnostic eXplainable Artificial Intelligence (*xAI*) tools, as they provide explanations for both local and global model outcomes (Joseph et al., 2024b; Prasad et al., 2023). Lastly, given the hybrid C-GRU model's superior

performance in forecasting $SWRI_{24\text{-}hr\text{-}S}$ at a 1-hourly forecast horizon, future studies should explore its effectiveness for longer forecast horizons to determine its robustness. Additionally, accurate forecasting of $SWRI_{24\text{-}hr\text{-}S}$ over a longer forecast horizon is crucial for assessing impending flood risk, thereby improving flood preparedness and implementing effective mitigation strategies.

CHAPTER 6: CONCLUSIONS AND FUTURE SCOPE

6.1. Synthesis and important findings

This research aims to provide innovative flood monitoring, assessment, and forecasting tools using AI and copula-statistical methods to enhance early warning systems and contribute to disaster risk reduction and mitigation strategies. Flooding is a catastrophic natural disaster that affects many nations worldwide. Its severe impact is particularly felt in developing countries, such as Fiji, on which this research focused. The methodologies proposed in this research rely on only one externally sourced data, specifically site-based, real-time hourly rainfall data (the other datasets were generated using the hourly rainfall data). For Fiji, rainfall data for various sites were obtained from FMS and pre-processed to generate the hourly rainfall data necessary for this study. However, this research could only focus on one of the two main islands in Fiji, Viti Levu, due to the unavailability of required data at other sites. The proposed methodologies can be applied to other data-scarce regions with similar rainfall data availability, aiding in effective flood risk monitoring, assessment, forecasting and mitigation strategies.

The following are the key findings of this research, achieved through its two main objectives. The outcomes from Objective 1 (Chapter 4) revealed that the proposed novel hourly flood index ($SWRI_{24\text{-}hr\text{-}S}$), a normalised metric, could be utilised as a practical tool to monitor flood risk and compute the flood event characteristics (i.e., D, V, and Q). This was impossible with the existing $WRI_{24\text{-}hr\text{-}S}$ (Deo et al., 2018) in the literature. The $SWRI_{24\text{-}hr\text{-}S}$ was employed to identify all the flood events for seven flood-prone sites and compute their characteristics between 2014 and 2018. The results were analysed, which showed that the wet season (November to April), including May and October, receives substantially higher rainfall than other months. Consequently, the frequency of floods and flood volume increased during these months. This has important implications for relevant authorities, such as Fiji's NDMO, in developing comprehensive flood preparedness and risk management strategies to mitigate the severe impacts of flood risk during these periods. The flood event characteristics and water-intensive properties, including the total rainfall recorded for five severe flood events at each of the seven study sites, were also provided. These

findings will help relevant organisations evaluate past flood events at these study sites and develop risk management strategies to mitigate their severe impacts.

Furthermore, the results also revealed that flood event characteristics (i.e., *D*, *V*, and *Q*) showed significant spatiotemporal variability and highly complex, non-linear relationships among them. Consequently, the 3D vine copula was employed across all study sites to capture the full dependence among flood characteristics by modelling the joint distribution and extracting their joint exceedance probability under different combination scenarios (i.e., flood event characteristics simultaneously exceed various thresholds) for probabilistic flood risk assessment. The results revealed that the 3D D-vine copula with flood volume (*V*) as the conditioning variable was the most parsimonious to model the joint distribution of flood characteristics across all study sites. The results evidently showed a moderate yet significant disparity in spatial patterns of the joint exceedance probability of flood event characteristics under different combination scenarios. The probability of a flood event characterised by median (i.e., 50th-quantile) duration, volume, and peak values across all study sites was found to be moderate, while the probability of an extreme flood event (where the flood volume, peak, and duration exceed the 95th-quantile value) was found to be exceptionally low across all study sites. This suggests that relevant authorities and communities in Fiji must implement risk management strategies not only for infrequent large floods but also for frequent smaller flood events with lower volume, peak and duration that may still cause considerable impacts (Government of Fiji, 2017). The hourly flood monitoring tool and evaluation of exceedance probability demonstrated in Objective 1 can be integrated into early flood warning systems to monitor and assess flood risks accurately, thus facilitating the implementation of targeted risk management strategies.

Moreover, although the outcomes of Objective 1 demonstrated the practical utility of $SWRI_{24\text{-}hr\text{-}S}$ to monitor flood risk, it is imperative to note that this index cannot predict the flood risk ahead of time unless a forecasting system is implemented and tested. Consequently, the outcomes of Objective 2 (Chapter 5) demonstrate the superiority of the proposed hybrid C-GRU model in forecasting the $SWRI_{24\text{-}hr\text{-}S}$ over a short-term, i.e., 1-hourly forecast horizon to assess future flood risk for the five flood-prone sites in the Western Division of Fiji. The hybrid C-GRU model was trained using the statistically significant lagged $SWRI_{24\text{-}hr\text{-}S}$ and hourly rainfall. Three DL models, i.e., CNN, GRU, and LSTM, and one ML model, i.e., RFR, were also developed for

comparison. To enhance the forecasting performance, the hyperparameters of all the models developed were optimised using an efficient BO with the TPE algorithm. The CNN and GRU layers of the hybrid C-GRU model proved highly effective, with the CNN layers excelling in feature extraction and the GRU layers adept at capturing long-term dependencies in sequential data. The outstanding performance of the hybrid C-GRU model in forecasting the $SWRI_{24\text{-}hr\text{-}S}$ over a short-term can be integrated within a decision support framework of early flood warning systems, thus improving Fiji's disaster preparedness and implementing appropriate flood mitigation strategies and response efforts.

The outcomes of this research offer innovative frameworks for an early flood warning system, particularly suited for data-scarce and flood-prone regions such as other developing Pacific Island countries experiencing recurrent flooding. Therefore, the major contributions of this research are succinctly summarised as follows:

- This research was the first to propose a novel hourly flood index ($SWRI_{24\text{-}hr\text{-}S}$) to identify flood events and compute their associated characteristics. Its practical utility in identifying flood events was tested and validated for various flood-prone sites in Fiji.
- The vine copula model was employed to model the joint distribution of flood characteristics and extract their joint exceedance for probabilistic flood risk assessment in Fiji's case studies. This approach is novel for Fiji, as no previous studies have demonstrated this methodology.
- A novel hybrid C-GRU-based $SWRI_{24\text{-}hr\text{-}S}$ flood forecasting system, a cost-effective and efficient tool, was developed and tested for Fiji's case studies.
- An efficient hyperparameter fine-tuning algorithm, i.e., the BO with the TPE algorithm, was integrated into its design architecture to improve the hybrid DL forecasting model's forecasting accuracy.
- The research also demonstrated the practical implementation of the proposed forecasting framework in the decision-support of early flood warning systems. This will enable relevant authorities and communities to better prepare for floods and implement appropriate mitigation measures to save lives, safeguard essential resources, and reduce economic losses.

6.2. Limitations and recommendations for future research

This research designed innovative tools for flood monitoring, assessment, and forecasting. These tools can be easily integrated into the decision-support systems of early warning, particularly in data-scarce regions. This integration will enhance and improve flood prediction and assessment strategies, thereby mitigating the devastating impacts of floods. However, the study elucidates some minor limitations that could be addressed in the future to further enhance its practical applications. The limitations and recommendations are summarised as follows:

- Due to the unavailability of the required rainfall data across various flood-prone sites, this study was limited to locations in the Western and Central Divisions of Fiji. Consequently, no sites from Fiji's Northern and Eastern Divisions were included. As a result, this study could not perform a comparative analysis across all four divisions to better identify high flood-risk areas in Fiji. Additionally, the Ba site, one of the main towns in the Western Division frequently affected by flooding, had to be excluded due to a high percentage of missing data. However, future studies can employ satellite-based rainfall products and apply the proposed methodologies to cover all major towns and cities in Fiji susceptible to flooding, following the recent approach used for Myanmar (Nguyen-Huy et al., 2022).
- This study employed a predetermined time-reduction weighting factor ($W \approx 3.8$) established in prior research (Deo et al., 2018). It must be noted that the proposed $SWRI_{24\text{-}hr\text{-}S}$ is the normalised metric derived from normalising the existing $WRI_{24\text{-}hr\text{-}S}$, which used a suitable time-dependent reduction function incorporating a weighting factor to account for the depletion of water resources through various hydrological processes. Consequently, the deviation of $WRI_{24\text{-}hr\text{-}S}$ followed by $SWRI_{24\text{-}hr\text{-}S}$ is contingent upon the value of W. However, future studies can test the appropriateness of this weighting factor (W) in locations where topography and climatic conditions exhibit substantial variability. Future studies could involve extensive correlational analyses of this weighting factor against rainfall-runoff relationships and other physical models to capture the decay of accumulated rainfall more accurately and its implications for flood events across diverse regions.

- Moreover, in future studies, the feasibility of the proposed $SWRI_{24\text{-}hr\text{-}S}$ as an hourly flood risk monitoring tool must be demonstrated to other flood-prone regions globally, contingent upon the availability of well-documented flood records for validation and hourly rainfall data. During this process, an alternative method for normalising the existing $WRI_{24\text{-}hr\text{-}S}$ could be chosen based on how effectively the normalised $WRI_{24\text{-}hr\text{-}S}$ index reflects flood risk under those specific climatic conditions.
- This study focused on only three flood event characteristics derived using the proposed $SWRI_{24\text{-}hr\text{-}S}$ to develop the vine copula model for probabilistic flood risk assessment. Nevertheless, future studies can consider incorporating additional flood event characteristics, where available, such as peak time, annual maximum 24-hour rainfall and highest storm surges and river discharge (Latif & Simonovic, 2022a, 2022b; Shafaei et al., 2017).
- This study developed a novel hybrid C-GRU-based $SWRI_{24\text{-}hr\text{-}S}$ flood forecasting system utilising only two features: lagged $SWRI_{24\text{-}hr\text{-}S}$ and hourly rainfall data. However, future studies can enhance the model's robustness by identifying and integrating additional pertinent features, such as synoptic-scale climate indices (Ahmed et al., 2023), river discharge and streamflow data, and hydrometeorological data to enhance the effectiveness of the proposed hybrid model for broader application.
- This study focused only on short-term (1-hour) $SWRI_{24\text{-}hr\text{-}S}$ forecasting. However, future studies should also test the model's performance for medium-term and long-term forecasting to scrutinise its robustness. Accurately forecasting $SWRI_{24\text{-}hr\text{-}S}$ at a longer forecast horizon is also critical for early warning systems, as it ensures that decision-makers receive crucial insights into evolving flood patterns and can take proactive measures to mitigate potential impacts and protect vulnerable communities.
- This study presents the prediction results as point forecasts, providing single-value estimates of $SWRI_{24\text{-}hr\text{-}S}$. While point forecasts are straightforward to interpret, they do not capture the inherent uncertainty in the predicted values. Future studies should investigate interval forecasting, which presents the expected range within which the predicted value will fall. Alternatively, probabilistic forecasting represents a viable approach, offering a nuanced

understanding of uncertainty by assigning probabilities to different predicted values.

- This study proposed a hybrid C-GRU model for accurately modelling $SWRI_{24\text{-}hr\text{-}S}$, which is a complex DL model with a large number of hyperparameters. Therefore, implementing these models for real-time applications may incur significant computational costs. Although an efficient BO with the TPE algorithm was employed for hyperparameter optimisation, future studies can also explore other techniques, such as the BOHB technique (Falkner et al., 2018). In addition, other DL and hybrid algorithms, such as CNN-BiLSTM, demonstrated in a related study (Ahmed et al., 2023), should also be explored for $SWRI_{24\text{-}hr\text{-}S}$ forecasting. The results of this study can establish a benchmark for new models.
- The proposed hybrid C-GRU model for $SWRI_{24\text{-}hr\text{-}S}$ forecasting is a non-interpretable "black-box" model. Thus, in future studies, it is recommended that the model-agnostic *xAI* methods like Local Interpretable Model-Agnostic Explanations (LIME) and SHapley Additive exPlanations (SHAP) are employed to understand the underlying mechanism of this proposed black-box model (Joseph et al., 2024b; Prasad et al., 2023).

To conclude, this Master of Research presented innovative, cost-effective tools that primarily rely on rainfall data alone. These tools are particularly useful for flood-prone areas lacking the financial and scientific resources to invest in advanced flood monitoring, assessment, and forecasting systems. The limitations and future research directions outlined above establish a baseline for further enhancement and improvement of the methods proposed in this study. Further improvement is crucial, as it significantly impacts the effectiveness of integrating these tools into decision-support systems for early flood warning. Finally, it is anticipated that more developing nations will adopt the methodologies presented in this study for flood risk management and mitigation.

REFERENCES

Aas, K., Czado, C., Frigessi, A., & Bakken, H. (2009). Pair-copula constructions of multiple dependence. *Insurance: Mathematics and economics, 44*(2), 182-198.

Abadi, M., Barham, P., Chen, J., Chen, Z., Davis, A., Dean, J., Devin, M., Ghemawat, S., Irving, G., & Isard, M. (2016). *TensorFlow: A system for Large-Scale Machine Learning.* Paper presented at the 12th USENIX symposium on operating systems design and implementation (OSDI 2016).

Ahmed, A. M., Deo, R. C., Feng, Q., Ghahramani, A., Raj, N., Yin, Z., & Yang, L. (2021). Hybrid deep learning method for a week-ahead evapotranspiration forecasting. *Stochastic Environmental Research and Risk Assessment*, 1-19.

Ahmed, A. M., Farheen, S., Nguyen-Huy, T., Raj, N., Jui, S. J. J., & Farzana, S. (2023). Real-time prediction of the week-ahead flood index using hybrid deep learning algorithms with synoptic climate mode indices [Preprint]. *Research Square*. doi:https://doi.org/10.21203/rs.3.rs-2654880/v1

Alexander, A. A., Thampi, S. G., & NR, C. (2018). Development of hybrid wavelet-ANN model for hourly flood stage forecasting. *ISH Journal of Hydraulic Engineering, 24*(2), 266-274.

Aminuddin Jafry, N., Suhaila, J., Yusof, F., Mohd Nor, S. R., & Alias, N. E. (2024). Enhancing flood risk assessment in the Johor River Basin through trivariate copula. *Journal of Water and Climate Change, 15*(4), 1820-1839.

Andonie, R. (2019). Hyperparameter optimization in learning systems. *Journal of Membrane Computing, 1*(4), 279-291.

Bedford, T., & Cooke, R. M. (2001). Probability density decomposition for conditionally dependent random variables modeled by vines. *Annals of Mathematics and Artificial intelligence, 32*, 245-268.

Bedford, T., & Cooke, R. M. (2002). Vines--a new graphical model for dependent random variables. *The Annals of Statistics, 30*(4), 1031-1068.

Byun, H.-R., & Lee, D.-K. (2002). Defining three rainy seasons and the hydrological summer monsoon in Korea using available water resources index. *Journal of the Meteorological Society of Japan. Ser. II, 80*(1), 33-44.

Chebana, F., & Ouarda, T. B. (2009). Index flood–based multivariate regional frequency analysis. *Water Resources Research, 45*(10).

Cheung, Y.-W., & Lai, K. S. (1995). Lag order and critical values of the augmented Dickey–Fuller test. *Journal of Business & Economic Statistics, 13*(3), 277-280.

Cho, K., Van Merriënboer, B., Gulcehre, C., Bahdanau, D., Bougares, F., Schwenk, H., & Bengio, Y. (2014). Learning phrase representations using RNN encoder-decoder for statistical machine translation. *arXiv preprint arXiv:1406.1078*.

CRED. (2020). *The human cost of disasters: an overview of the last 20 years (2000-2019)*. Retrieved from https://www.preventionweb.net/files/74124_humancostofdisasters20002019reportu.pdf

Daneshkhah, A., Remesan, R., Chatrabgoun, O., & Holman, I. P. (2016). Probabilistic modeling of flood characterizations with parametric and minimum information pair-copula model. *Journal of Hydrology, 540*, 469-487.

Deo, R. C., Adamowski, J. F., Begum, K., Salcedo-Sanz, S., Kim, D.-W., Dayal, K. S., & Byun, H.-R. (2019). Quantifying flood events in Bangladesh with a daily-step flood monitoring index based on the concept of daily effective precipitation. *Theoretical and Applied Climatology, 137*(1), 1201-1215.

Deo, R. C., Byun, H.-R., Adamowski, J. F., & Kim, D.-W. (2015). A real-time flood monitoring index based on daily effective precipitation and its application to Brisbane and Lockyer Valley flood events. *Water Resources Management, 29*(11), 4075-4093.

Deo, R. C., Byun, H.-R., Kim, G.-B., & Adamowski, J. F. (2018). A real-time hourly water index for flood risk monitoring: Pilot studies in Brisbane, Australia, and Dobong Observatory, South Korea. *Environmental monitoring and assessment, 190*(8), 1-27.

Dickey, D. A., & Fuller, W. A. (1979). Distribution of the estimators for autoregressive time series with a unit root. *Journal of the American statistical association, 74*(366a), 427-431.

Diebold, F. X., & Mariano, R. S. (2002). Comparing predictive accuracy. *Journal of Business & Economic Statistics, 20*(1), 134-144.

Eggensperger, K., Feurer, M., Hutter, F., Bergstra, J., Snoek, J., Hoos, H., & Leyton-Brown, K. (2013). *Towards an empirical foundation for assessing bayesian optimization of hyperparameters.* Paper presented at the NIPS workshop on Bayesian Optimization in Theory and Practice.

Falkner, S., Klein, A., & Hutter, F. (2018). *BOHB: Robust and efficient hyperparameter optimization at scale.* Paper presented at the International conference on machine learning.

Feresi, J., Kenny, G. J., de Wet, N., Limalevu, L., Bhusan, J., & Ratukalou, I. (2000). *Climate change vulnerability and adaptation assessment for Fiji*. Retrieved from https://researchcommons.waikato.ac.nz/items/6148f4a7-2b7d-4e9c-bc86-baa523e4e114

Ganguli, P., & Reddy, M. J. (2013). Probabilistic assessment of flood risks using trivariate copulas. *Theoretical and Applied Climatology, 111*, 341-360.

Ghimire, S., Bhandari, B., Casillas-Perez, D., Deo, R. C., & Salcedo-Sanz, S. (2022). Hybrid deep CNN-SVR algorithm for solar radiation prediction problems in Queensland, Australia. *Engineering Applications of Artificial Intelligence, 112*, 104860.

Government of Fiji. (2017). Climate vulnerability assessment: Making Fiji climate resilient. Retrieved from https://reliefweb.int/report/fiji/climate-vulnerability-assessment-making-fiji-climate-resilient

Gräler, B., van den BERG, M. J., Vandenberghe, S., Petroselli, A., Grimaldi, S., De Baets, B., & Verhoest, N. (2013). Multivariate return periods in hydrology: a critical and practical review focusing on synthetic design hydrograph estimation. *Hydrology and Earth System Sciences, 17*(4), 1281-1296.

Graves, A. (2013). Generating sequences with recurrent neural networks. *arXiv preprint arXiv:1308.0850*.

Hapuarachchi, H., Wang, Q., & Pagano, T. (2011). A review of advances in flash flood forecasting. *Hydrological processes, 25*(18), 2771-2784.

Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. *Neural computation, 9*(8), 1735-1780.

Ji, L., Fu, C., Ju, Z., Shi, Y., Wu, S., & Tao, L. (2022). Short-term canyon wind speed prediction based on CNN—GRU transfer learning. *Atmosphere, 13*(5), 813.

Joe, H. (1997). *Multivariate models and multivariate dependence concepts*: CRC press.

Joseph, L. P., Deo, R. C., Casillas-Perez, D., Prasad, R., Raj, N., & Salcedo-Sanz, S. (2024a). Multi-step-ahead wind speed forecast system: Hybrid multivariate decomposition and feature selection-based gated additive tree ensemble model. *IEEE Access*.

Joseph, L. P., Deo, R. C., Casillas-Pérez, D., Prasad, R., Raj, N., & Salcedo-Sanz, S. (2024b). Short-term wind speed forecasting using an optimized three-phase convolutional neural network fused with bidirectional long short-term memory network model. *Applied Energy, 359*, 122624.

Kabir, S., Patidar, S., Xia, X., Liang, Q., Neal, J., & Pender, G. (2020). A deep convolutional neural network model for rapid prediction of fluvial flood inundation. *Journal of Hydrology, 590*, 125481.

Kant, A., Suman, P. K., Giri, B. K., Tiwari, M. K., Chatterjee, C., Nayak, P. C., & Kumar, S. (2013). Comparison of multi-objective evolutionary neural network, adaptive neuro-fuzzy inference system and bootstrap-based neural network for flood forecasting. *Neural Computing and Applications, 23*(1), 231-246.

Ketkar, N. (2017). Introduction to Keras. In *Deep learning with python: a hands-on introduction* (pp. 97-111). Berkeley, CA: Apress.

Kim, K., Kim, D.-K., Noh, J., & Kim, M. (2018). Stable forecasting of environmental time series via long short term memory recurrent neural network. *IEEE Access, 6*, 75216-75228.

Kisvari, A., Lin, Z., & Liu, X. (2021). Wind power forecasting–A data-driven method along with gated recurrent neural network. *Renewable Energy, 163*, 1895-1909.

Klaho, M. H., Safavi, H. R., Golmohammadi, M. H., & Alkntar, M. (2022). Comparison between bivariate and trivariate flood frequency analysis using the Archimedean copula functions, a case study of the Karun River in Iran. *Natural hazards, 112*(2), 1589-1610.

Komer, B., Bergstra, J., & Eliasmith, C. (2014). *Hyperopt-Sklearn: Automatic Hyperparameter Configuration for Scikit-Learn.* Paper presented at the Scipy.

Kuleshov, Y., McGree, S., Jones, D., Charles, A., Cottrill, A., Prakash, B., Atalifo, T., Nihmei, S., & Seuseu, F. L. S. K. (2014). Extreme weather and climate events and their impacts on island countries in the Western Pacific: cyclones, floods and droughts. *Atmospheric and Climate Sciences, 4*(05), 803.

Kumar, V., Sharma, K., Caloiero, T., Mehta, D., & Singh, K. (2023). Comprehensive overview of flood modeling approaches: a review of recent advances. Hydrology 10 (7): 141. In.

Latif, S., & Mustafa, F. (2021). Bivariate joint distribution analysis of the flood characteristics under semiparametric copula distribution framework for the Kelantan River basin in Malaysia. *Journal of Ocean Engineering and Science, 6*(2), 128-145.

Latif, S., & Mustafa, F. (2020). Parametric vine copula construction for flood analysis for Kelantan river basin in Malaysia. *Civil Engineering Journal, 6*(8), 1470-1491.

Latif, S., & Simonovic, S. P. (2022a). Parametric Vine copula framework in the trivariate probability analysis of compound flooding events. *Water, 14*(14), 2214.

Latif, S., & Simonovic, S. P. (2022b). Trivariate Joint Distribution Modelling of Compound Events Using the Nonparametric D-Vine Copula Developed Based on a Bernstein and Beta Kernel Copula Density Framework. *Hydrology, 9*(12), 221.

Leonard, M., Westra, S., Phatak, A., Lambert, M., van den Hurk, B., McInnes, K., Risbey, J., Schuster, S., Jakob, D., & Stafford-Smith, M. (2014). A compound event framework for understanding extreme impacts. *Wiley Interdisciplinary Reviews: Climate Change, 5*(1), 113-128.

Li, X. (2023). CNN-GRU model based on attention mechanism for large-scale energy storage optimization in smart grid. *Frontiers in Energy Research, 11*, 1228256.

Lu, E. (2009). Determining the start, duration, and strength of flood and drought with daily precipitation: Rationale. *Geophysical research letters, 36*(12).

Lu, E., Cai, W., Jiang, Z., Zhang, Q., Zhang, C., Higgins, R. W., & Halpert, M. S. (2014). The day-to-day monitoring of the 2011 severe drought in China. *Climate Dynamics, 43*(1), 1-9.

Lucas, B. (2020). *Economic Impacts of Natural Hazards on Vulnerable Populations in FIJI*. Retrieved from https://climate-insurance.org/wp-content/uploads/2021/01/Fiji-Economic-Impacts-Report-27Nov2020.pdf

Maranzoni, A., D'Oria, M., & Rizzo, C. (2023). Quantitative flood hazard assessment methods: A review. *Journal of Flood Risk Management, 16*(1), e12855.

McGree, S., Yeo, S. W., & Devi, S. (2010). Flooding in the Fiji Islands between 1840 and 2009. *Risk Frontiers*, 1-69.

McNamara, K. E. (2013). A state of emergency: How local businesses experienced the 2012 flood in Fiji. *Australian Journal of Emergency Management, The, 28*(3), 17-23.

Miau, S., & Hung, W.-H. (2020). River flooding forecasting and anomaly detection based on deep learning. *IEEE Access, 8*, 198384-198402.

Moishin, M., Deo, R. C., Prasad, R., Raj, N., & Abdulla, S. (2021a). Designing deep-based learning flood forecast model with ConvLSTM hybrid algorithm. *IEEE Access, 9*, 50982-50993.

Moishin, M., Deo, R. C., Prasad, R., Raj, N., & Abdulla, S. (2021b). Development of Flood Monitoring Index for daily flood risk evaluation: case studies in Fiji. *Stochastic Environmental Research and Risk Assessment, 35*(7), 1387-1402.

Nagler, T., & Vatter, T. (2023). rvinecopulib: High performance algorithms for vine copula modeling. *R package version 3*.

Nevo, S., Morin, E., Gerzi Rosenthal, A., Metzger, A., Barshai, C., Weitzner, D., Voloshin, D., Kratzert, F., Elidan, G., & Dror, G. (2022). Flood forecasting with machine learning models in an operational framework. *Hydrology and Earth System Sciences, 26*(15), 4013-4032.

Nguyen-Huy, T., Deo, R. C., Yaseen, Z. M., Prasad, R., & Mushtaq, S. (2021). Bayesian Markov chain Monte Carlo-based copulas: factoring the role of large-scale climate indices in monthly flood prediction. In R. C. Deo, P. Samui, O. Kisi, & Z. M. Yaseen (Eds.), *Intelligent Data Analytics for Decision-Support Systems in Hazard Mitigation: Theory and Practice of Hazard Mitigation* (pp. 29-47). 152 Beach Road, #21-01/04 Gateway East, Singapore 189721, Singapore: Springer Nature Singapore Pte Ltd.

Nguyen-Huy, T., Kath, J., Nagler, T., Khaung, Y., Aung, T. S. S., Mushtaq, S., Marcussen, T., & Stone, R. (2022). A satellite-based Standardized Antecedent

Precipitation Index (SAPI) for mapping extreme rainfall risk in Myanmar. *Remote Sensing Applications: Society and Environment, 26*, 100733.

Nguyen, H.-P., Liu, J., & Zio, E. (2020). A long-term prediction approach based on long short-term memory neural networks with automatic parameter optimization by Tree-structured Parzen Estimator and applied to time-series data of NPP steam generators. *Applied Soft Computing, 89*, 106116.

Nosrati, K., Saravi, M. M., & Shahbazi, A. (2010). Investigation of flood event possibility over Iran using Flood Index. In H. Gökçekus, U. Türker, & J. W. LaMoreaux (Eds.), *Survival and Sustainability: Environmental Concerns in the 21st Century* (pp. 1355-1361). Berlin/Heidelberg, Germany: Springer.

Oriani, F., Stisen, S., Demirel, M. C., & Mariethoz, G. (2020). Missing data imputation for multisite rainfall networks: a comparison between geostatistical interpolation and pattern-based estimation on different terrain types. *Journal of Hydrometeorology, 21*(10), 2325-2341.

Pan, M., Zhou, H., Cao, J., Liu, Y., Hao, J., Li, S., & Chen, C.-H. (2020). Water level prediction model based on GRU and CNN. *IEEE Access, 8*, 60090-60100.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., & Dubourg, V. (2011). Scikit-learn: Machine learning in Python. *the Journal of machine Learning research, 12*, 2825-2830.

Prasad, R., Charan, D., Joseph, L., Nguyen-Huy, T., Deo, R. C., & Singh, S. (2021). Daily flood forecasts with intelligent data analytic models: multivariate empirical mode decomposition-based modeling methods. In R. C. Deo, P. Samui, O. Kisi, & Z. M. Yaseen (Eds.), *Intelligent Data Analytics for Decision-Support Systems in Hazard Mitigation: Theory and Practice of Hazard Mitigation* (pp. 359-381). 152 Beach Road, #21-01/04 Gateway East, Singapore 189721, Singapore: Springer Nature Singapore Pte Ltd.

Prasad, S. S., Deo, R. C., Downs, N. J., Casillas-Pérez, D., Salcedo-Sanz, S., & Parisi, A. V. (2024). Very short-term solar ultraviolet-A radiation forecasting system with cloud cover images and a Bayesian optimized interpretable artificial intelligence model. *Expert Systems with Applications, 236*, 121273.

Prasad, S. S., Deo, R. C., Salcedo-Sanz, S., Downs, N. J., Casillas-Pérez, D., & Parisi, A. V. (2023). Enhanced joint hybrid deep neural network explainable artificial intelligence model for 1-hr ahead solar ultraviolet index prediction. *Computer Methods and Programs in Biomedicine, 241*, 107737.

Roberts, W., Williams, G. P., Jackson, E., Nelson, E. J., & Ames, D. P. (2018). Hydrostats: A Python package for characterizing errors between observed and predicted time series. *Hydrology, 5*(4), 66.

Sajjad, M., Khan, Z. A., Ullah, A., Hussain, T., Ullah, W., Lee, M. Y., & Baik, S. W. (2020). A novel CNN-GRU-based hybrid approach for short-term residential load forecasting. *IEEE Access, 8*, 143759-143768.

Sarker, I. H. (2021). Deep learning: a comprehensive overview on techniques, taxonomy, applications and research directions. *SN Computer Science, 2*(6), 420.

Seiler, R., Hayes, M., & Bressan, L. (2002). Using the standardized precipitation index for flood risk monitoring. *International Journal of Climatology: A Journal of the Royal Meteorological Society, 22*(11), 1365-1376.

Seneviratne, S. I., Nicholls, N., Easterling, D., Goodess, C. M., Kanae, S., Kossin, J., Luo, Y., Marengo, J., McInnes, K., Rahimi, M. R., Reichstein, M., Sorteberg, A., Vera, C., & Zhang, X. (2012). Changes in Climate Extremes and their Impacts on the Natural Physical Environment. In C. B. Field, V. Barros, T. F. Stocker, Q. Dahe, D. J. Dokken, K. L. Ebi, M. D. Mastrandrea, K. J. Mach, G.-K. Plattner, S. K. Allen, M. Tignor, & P. M. Midgley (Eds.), *Managing the Risks of Extreme Events and Disasters to Advance Climate Change Adaptation: Special Report of the Intergovernmental Panel on Climate Change* (pp. 109-230). Cambridge, UK, and New York, NY, USA: Cambridge University Press.

Shafaei, M., Fakheri-Fard, A., Dinpashoh, Y., Mirabbasi, R., & De Michele, C. (2017). Modeling flood event characteristics using D-vine structures. *Theoretical and Applied Climatology, 130*(3), 713-724.

Sharma, E., Deo, R. C., Soar, J., Prasad, R., Parisi, A. V., & Raj, N. (2022). Novel hybrid deep learning model for satellite based PM10 forecasting in the most polluted Australian hotspots. *Atmospheric Environment, 279*, 119111.

Sharma, K. K., Verdon-Kidd, D. C., & Magee, A. D. (2021). A decision tree approach to identify predictors of extreme rainfall events–A case study for the Fiji Islands. *Weather and Climate Extremes, 34*, 100405.

Sklar, M. (1959). Fonctions de repartition an dimensions et leurs marges. *Publ. inst. statist. univ. Paris, 8*, 229-231.

Teng, J., Jakeman, A. J., Vaze, J., Croke, B. F., Dutta, D., & Kim, S. (2017). Flood inundation modelling: A review of methods, recent advances and uncertainty analysis. *Environmental modelling & software, 90*, 201-216.

The World Bank Group. (2022). Data. Retrieved from https://data.worldbank.org/country/fiji?view=chart

Tiwari, M. K., & Chatterjee, C. (2010). Development of an accurate and reliable hourly flood forecasting model using wavelet–bootstrap–ANN (WBANN) hybrid approach. *Journal of Hydrology, 394*(3-4), 458-470.

Tosunoglu, F., Gürbüz, F., & İspirli, M. N. (2020). Multivariate modeling of flood characteristics using Vine copulas. *Environmental Earth Sciences, 79*(19), 1-21.

Wang, Z., Dong, Y., Liu, W., & Ma, Z. (2020). A novel fault diagnosis approach for chillers based on 1-D convolutional neural network and gated recurrent unit. *Sensors, 20*(9), 2458.

Yeo, S. W., & Blong, R. J. (2010). Fiji's worst natural disaster: the 1931 hurricane and flood. *Disasters, 34*(3), 657-683.

Yu, J., Zhang, X., Xu, L., Dong, J., & Zhangzhong, L. (2021). A hybrid CNN-GRU model for predicting soil moisture in maize root zone. *Agricultural Water Management, 245*, 106649.

Zhang, C., Peng, T., & Nazir, M. S. (2022). A novel integrated photovoltaic power forecasting model based on variational mode decomposition and CNN-BiGRU considering meteorological variables. *Electric Power Systems Research, 213*, 108796.