

LIVING LAB

Dr. Yanbin Xu, Dr. Wenjun Zhu, Dr. Shinichi Kamiya

Insurance Risk and Finance Research Centre, Nanyang Technological University

23 May 2025

Inland Flood Risk Modelling for Southeast Asia: A Case Study of the Chao Phraya River – Now and Future



Acknowledgement

We acknowledge the Japan Agency for Marine-Earth Science and Technology, the Atmosphere and Ocean Research Institute of the University of Tokyo, and the National Institute for Environmental Studies for developing the MIROC6 climate model. We also thank the international climate modeling community contributing to the Coupled Model Intercomparison Project Phase 6, and the Earth System Grid Federation for providing access to the simulation data. We would also like to thank the GAIP Partners for their valuable input and feedback into the paper.

Foreword



Earlier this year, the Global Asia Insurance Partnership (GAIP) adopted its 2025–2029 strategy, marking a pivotal shift from being primarily a research-oriented organisation to one that drives impact by turning rigorous research into actionable solutions. Guided by its vision of building a risk-resilient and sustainable world, GAIP’s efforts are now organised around four strategic themes:

1. An Integrated Approach to Protection Gaps
2. Addressing Health & Retirement Gaps
3. Climate Change & Insurance
4. Navigating the Technological Landscape

Under this strategy, GAIP aims to convene its tripartite partners—across industry, regulatory, and academic sectors—and broader stakeholders in the risk ecosystem to transform insight into implementation, evidence into policy, and pilots into scale.

Nanyang Technological University (NTU), GAIP’s key academic partner, is central to this transition. The publication of this third Living Lab report marks a key milestone in our collaboration, reflecting our commitment to producing work that bridges scientific innovation and practical utility.

This report addresses the increasing unpredictability of natural hazards, such as inland flooding, in the context of climate change. Traditional flood risk models are under pressure due to the intensifying frequency and severity of events—particularly in regions with limited historical data. In response, this report introduces the Geo-Hierarchical Deep Learning (GHDL) framework, developed and tested by Dr. Yanbin Xu, GAIP Research Fellow at NTU.

Designed for data-scarce environments, the GHDL framework leverages deep learning techniques to provide an alternative approach to flood risk modelling. While the model is illustrated using Thailand's Chao Phraya River Basin, its applicability extends across Asia and beyond.

To support adoption and experimentation, GAIP will have a programming package available to its partners. We invite interested stakeholders to explore the model and collaborate with us in integrating it into your flood risk modelling capabilities.

We would like to thank Dr. Xu for his dedication, and all NTU colleagues who supported this work. We are also grateful to GAIP partners for their continued engagement and insights, and to GAIP Director Dickson Wong and Associate Director Yao Lei for their support in bringing this work to fruition.

As GAIP embarks on this new chapter, we look forward to working together with our partners to turn knowledge into action—and to deliver measurable impact for those most exposed to today's and tomorrow's risks.

Sincerely,



Min Hung Cheng
CEO, Global Asia Insurance
Partnership



Jun-koo Kang
Director, Insurance Risk and Finance
Centre (IRFRC)

Executive Summary

In the face of escalating climate change, the uncertainty surrounding flood risks has become a critical concern for the insurance industry across Asia. Traditional flood models often struggle to account for the increasing frequency and severity of flood events driven by changing climate patterns and will require extensive data. This report explores the Geo-Hierarchical Deep Learning (GHDL) framework as an innovative solution to address some of these challenges. By leveraging deep learning techniques and addressing critical data limitations, the GHDL framework offers a robust and adaptable approach to inland flood risk modelling, providing stakeholders with the tools needed to enhance flood risk management and insurance pricing strategies. Although the Chao Phraya River Basin is provided as a case study, this model can be used effectively in various inland flood situations relevant across Asia.

This report presents the development and application of the Geo-Hierarchical Deep Learning (GHDL) framework for inland flood hazards modelling (Section 3). While demonstrated using the Chao Phraya River Basin as a case study (Section 4), the GHDL framework is broadly applicable to inland flood risks across diverse regions (Section 2.4). By addressing critical gaps in data limitation and leveraging a deep learning approach (Section 2.1), this work provides a scalable and transferable solution for stakeholders aiming to enhance flood risk management and insurance pricing strategies (Section 2.2-2.3).

The GHDL framework offers a valuable alternative to existing commercial models. While commercial models like the RMS and Oasis LMF provide high-resolution flood mapping and modular frameworks, they often require extensive data inputs and are resource-intensive. In comparison, the GHDL model integrates upstream and downstream flood dependencies within a geo-hierarchical structure, allowing for effective flood hazard predictions in data-limited settings (Section 7). Both this report and its foundational reference paper (Xu et al, 2022) have validated the strong predictive performance of the GHDL model, outperforming simpler yet equally complex benchmarks in capturing spatial and temporal flood hazard dynamics.

In this report, we incorporate the GHDL model with future climate projections through Shared Socioeconomic Pathways (SSPs), enabling scenario-based risk assessments that are crucial for long-term planning in the insurance sector (Section 6).

To bridge the gap between research and industry application, a programming package designed to replicate the modelling process described herein will be provided to our GAIP partners. They can leverage this tool for in-house testing using their proprietary policy and claim datasets, offering a cost-effective way to explore flood risk modelling tailored to specific portfolios (Section 5). This adaptability empowers insurers to conduct rigorous testing, refine underwriting strategies, and explore innovative insurance products, such as parametric insurance, which the GHDL framework readily supports.

In summary, the GHDL framework represents an advancement in flood risk modelling by addressing critical challenges in standardisation and data constraints. We hope it can serve as a practical and accessible tool for the insurance industry, offering a roadmap for enhancing flood risk assessment and advancing climate adaptation efforts. This report invites them to explore the utility of the GHDL model and its supporting tools, with opportunities for further development and customisation to meet their unique needs.

Contents

1	<i>Introduction</i>	9
2	<i>Current Challenges of Flood Risk Modelling and the Generality of GHDL Model</i>	14
2.1	Challenges of Flood Risk Modelling	14
2.2	Advantage of GHDL in Developing Country Flood Risk Modelling	15
2.2.1	Strong in Addressing Data Scarcity	15
2.2.2	Cost-efficient and Physical Information Embedding.....	16
2.3	GHDL's Data Requirements and Assumptions	19
2.4	Flood Hazard Measures for Other River Basins in APAC	21
3	<i>Deep Learning Approaches to Flood Hazard Modelling</i>	22
3.1	Objective.....	22
3.2	Commonly Deep Learning Frameworks for Flood Hazard Factorisation	22
3.2.1	Convolutional Neural Networks.....	22
3.2.2	Wavelet Neural Networks.....	24
3.3	Geo-Hierarchical Deep Learning Framework.....	24
4	<i>Application of the framework on the Chao Phraya Basin</i>	27
4.1	Inland Flood Hazard Metrics	27
4.2	Data	29
4.2.1	Flood Hazard Data.....	29
4.2.2	Independent Data	35
4.3	Model Specifications and Experimental Design.....	36
4.4	Estimation of Flood Hazard.....	38
5	<i>Premium Pricing and Loss Estimation with Flood Hazard Factors</i>	42
5.1	Methodological Framework.....	42
6	<i>Projecting Future Flood Hazards for the Chao Phraya River Basin</i>	44
6.1	Climate Change Simulations.....	44
6.2	Shared socioeconomic pathways.....	44

6.3	Climate Simulation Dataset.....	45
6.4	Flood Risk Projections and Results.....	48
7	<i>Comparison with Commercial Flood Risk Models.....</i>	<i>51</i>
8	<i>Implications.....</i>	<i>54</i>
8.1	More Collaboration.....	54
8.2	Risk Reduction.....	55
	Strengthening Riverbank Infrastructure.....	55
	Enhancing Coastal Protection.....	56
9	<i>Conclusions.....</i>	<i>58</i>
A	<i>Appendix.....</i>	<i>60</i>
A.1	Discrete Wavelet Transform.....	60
10	<i>Reference.....</i>	<i>64</i>

1 Introduction

In recent years, the increasing frequency and severity of flood events due to climate change have underscored the need for more robust and adaptable flood risk modelling frameworks. Traditional models sometimes could fall short in predicting these evolving risks, necessitating innovative approaches that can better account for the complexities of flood dynamics in a changing climate. This report introduces the Geo-Hierarchical Deep Learning (GHDL) framework as a promising solution to these challenges, offering enhanced predictive capabilities and adaptability across diverse regions. Importantly, the GHDL framework is designed to complement existing models, particularly in areas with inland flood risk where the data required for traditional models is limited.

The importance of developing alternative flood risk models lies in the growing uncertainty and variability of flood events driven by climate change. As traditional models struggle to keep pace with these changes, there is a critical need for new methodologies that can provide more accurate and reliable predictions. The GHDL framework addresses this need by leveraging advanced deep learning techniques and addressing critical data limitations, offering a scalable and transferable solution for enhancing flood risk management and optimising insurance pricing strategies.

Our exploration of the GHDL framework began with its application to the Mississippi River Basin, where it demonstrated strong predictive performance and adaptability to various flood scenarios. The success of this initial application highlighted the potential of the GHDL model to be applied to other regions with similar challenges. This report now shifts focus to Asia, using the Chao Phraya River Basin in central Thailand as a case study to further validate the model's effectiveness.

The Chao Phraya River Basin, located in central Thailand, is one of the most critical regions for the nation's socio-economic and environmental well-being. Spanning an area of approximately 160,000 square kilometres, the basin supports over 13 million residents who rely on its waters for agriculture, industry, and domestic use (Budhathoki et al., 2024; Visessri and Ekkawatpanit, 2020). The basin encompasses key urban and rural areas, including the bustling capital city of

Bangkok, which alone is home to over 11 million people¹. The region is a hub of economic activity, contributing significantly to Thailand's GDP through agriculture, manufacturing, and services. Its fertile plains are vital for rice cultivation, making Thailand one of the world's top rice exporters, while the densely populated urban centres drive industrial and commercial growth.

Historically, the Chao Phraya River Basin has experienced significant flood events, which have had profound impacts on lives and livelihoods. The devastating floods of 2011, for instance, inundated large parts of Bangkok and its surrounding provinces [area affected 97,000 square kilometres (Gale and Saunders, 2013)], resulting in economic losses exceeding \$46 billion at the time (equivalent to \$55 billion today) and insured losses surpassing \$15 billion at the time (equivalent to \$18 billion today), while displacing millions of people (Bevere and Dhore, 2021). These floods highlighted the region's vulnerability to extreme weather events exacerbated by climate change and urbanisation. Other notable flood events occurred in 1995 and 2002 (area affected 444,000 and 372,000 square kilometres, respectively), each bringing widespread disruption and underscoring the critical need for effective flood management strategies. As a lifeline for Thailand, the Chao Phraya River Basin remains a focal point for research and investment in sustainable water resource management and disaster resilience.

¹ Office of the National Economic and Social Development Council (retrieved on December 1, 2024): https://www.nesdc.go.th/nesdb_en/more_news.php?cid=156&filename=index

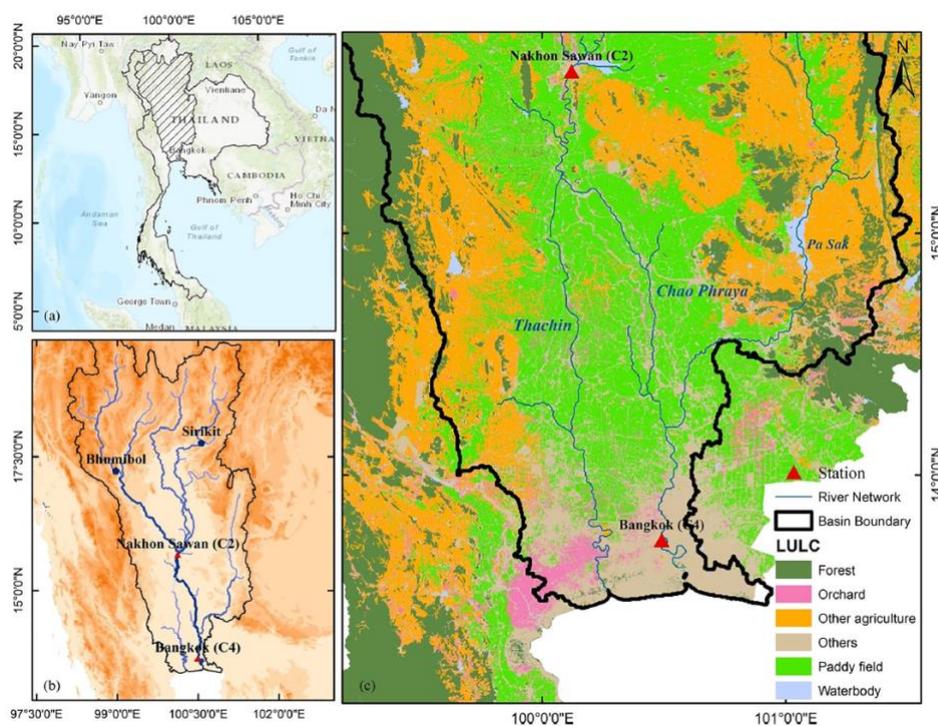


Figure 1: Location map of the study area: (a) Thailand, (b) Chao Phraya River Basin (CPRB), and (c) Lower Chao Phraya River Basin (LCPRB).

Source: Figure 1 of Budhathoki et al. (2024).

Well-devised risk management, including insurance, is essential for climate adaptation, protecting individuals, businesses, and communities from the financial impacts of flooding. Despite its importance, the flood insurance market suffers from low coverage. Rising flood losses already compel insurance companies to increase their capital base, potentially leading to prolonged periods of unprofitability. Uninsured risks remain a significant concern, as inadequate financial resources for relief and recovery adversely affect society, the economy, and the well-being of people (Jongman et al., 2014). The insurance gap has grown significantly, from \$117 billion in 2020 to \$161 billion in 2021. While nearly 29% of the global population is exposed to flood risks, only 7% of flood losses in emerging markets and 31% in developed economies were insured over the two decades (Swiss Re, 2022). This gap highlights the need for enhanced awareness and accessibility of flood insurance. What makes things worse is that while risk management generally reduces the impacts of floods, it struggles with unprecedented events, which are becoming more frequent owing to climate change (Kreibich et al., 2022).

Accurate models serve as an essential foundation for policymakers to design resilient infrastructures and implement proactive mitigation strategies for flood risks. Accurate flood risk prediction enhances the resilience of communities and economies, reducing the overall socioeconomic impact of flooding events (Surminski et al., 2015). For the insurance industry, precise flood risk assessments underpin the development of appropriate insurance products, inform pricing strategies, and ensure the financial stability of insurance providers (Kousky and Kunreuther, 2018). Moreover, a robust framework for flood risk modelling that is both calibratable and transferable across different markets is also critical for flood insurance development.

Despite their critical importance, the development of accurate flood risk models faces numerous challenges. Traditional statistical models often struggle to capture the complex interactions between hydrological and meteorological factors that drive flood events. Additionally, the limited spatial and temporal resolution of available data contributes to significant prediction inaccuracies (Bouwer, 2013). The dynamic and evolving nature of climate change further exacerbates these challenges, introducing additional variability that complicates reliable flood risk predictions.

Physical models, which aim to simulate flood mechanisms, are resource-intensive, costly, and laborious to develop. They also require frequent updates and recalibration, making them less adaptable to rapidly changing conditions. These shortcomings in flood risk modelling not only undermine program effectiveness but also hinder public awareness, reducing the perceived value of insurance and the willingness to pay for coverage (Thistlethwaite et al., 2020).

In the context of developing regions, these challenges are compounded by resource and data limitations, creating additional barriers to achieving accurate flood risk assessments. Such constraints underscore the need for innovative and efficient modelling approaches that can address these gaps while maintaining robustness and reliability.

In this report, we employed the Geo-Hierarchical Deep Learning (GHDL) framework (Xu et al., 2022), a model specifically developed to address data and resource limitations commonly faced in developing regions. This framework combines the robustness of statistical risk modelling with a structure that

preserves key physical characteristics, making it particularly suited for flood risk prediction and flood insurance pricing. The GHDL framework leverages high-resolution meteorological and hydraulic data within a uniquely coded geographical hierarchy. By integrating the geographical connectivity of regions directly into the deep learning structure, the model enhances its applicability and transferability across diverse physical and anthropogenic contexts. This approach not only improves prediction accuracy but also ensures the model remains adaptable to varying conditions (see Section 2).

With the flood hazard associations established, we utilise climate simulations for four widely discussed socio-economic pathways to project future climate change impacts.

2 *Current Challenges of Flood Risk Modelling and the Generality of GHDL Model*

2.1 **Challenges of Flood Risk Modelling**

The industry's capacity to model flood risks remains limited (Swiss Re, 2022). While probabilistic models exist, they are still at a relatively early stage of development. As summarised by Nkwunonwo et al. (2020), the advancement of flood risk modelling in developing countries faces significant challenges due to its nature as a secondary catastrophe—characterised by higher frequency and lower per-event damage compared to primary catastrophes like tropical cyclones and earthquakes. Referring to Nkwunonwo et al. (2020), we list two key non-technical gaps that pose particular challenges:

1. Data Availability and Quality

Developed countries such as the Netherlands, the UK, and the US typically have access to high-resolution hydrologic, hydraulic, and topographic data necessary for accurate flood modelling. In contrast, developing countries often face severe data shortages, driven by financial constraints or political factors.

Low-resolution global datasets are frequently used in developing countries but lack the precision needed for accurate flood inundation modelling, particularly in urbanised areas.

2. Cost Barriers for Moderate Exposure Regions

Advances in geospatial and remote sensing technologies offer promising solutions to data challenges. However, the cost of data acquisition, processing expertise, and required software remains prohibitive for many developing countries.

Building precise flood models for regions with moderate exposure may not be cost-effective. However, these regions often cover large areas and collectively account for a significant portion of insurers' exposure. For instance, the 2011 Thailand floods impacted 69 out of 79 provinces (Poapongsakorn and Meethom, 2013). While it might not be feasible to develop costly physical models for the non-Bangkok regions, these areas still contribute substantially to overall losses.

Beyond these two non-technical challenges, flood risk modelling for developing countries also faces additional barriers, such as

3. Challenges in Flood Frequency Analysis² and 4. Localisation of Global Flood Models.³

This report demonstrates how the GHDL model can partially address these challenges using the Chao Phraya River Basin as a case study. The structure of the GHDL model (described in Section 3.3) is designed to be generalisable, enabling the simultaneous modelling of inland flood risk dependencies across multiple regions.

In the remainder of this section, we discuss how the GHDL model addresses the challenges of flood risk modelling in developing countries, its data requirements and assumptions, and how its data-driven approach makes it applicable to other regions.

2.2 Advantage of GHDL in Developing Country Flood Risk Modelling

2.2.1 Strong in Addressing Data Scarcity

The GHDL flood hazard model demonstrates strong capabilities in addressing the data constraints commonly faced by developing countries for two key reasons:

1. Utilisation of Satellite Data

GHDL can directly incorporate raw satellite data, which offers a significant advantage in regions where conventional flood hazard data is sparse. Compared to traditional data sources, high-resolution satellite data is more readily available thanks to advancements in satellite technology.

Conventional event-based flood models heavily rely on accurately recorded historical flood events, making their effectiveness dependent on the availability and reliability of long-term observational data. This poses a significant challenge

² Reliable flood frequency analysis is often lacking in developing countries, limiting the ability to accurately estimate flood risks and inundation probabilities.

³ Global flood models often struggle to account for localised conditions and data constraints in developing countries, reducing their accuracy and applicability.

in many regions of Asia, where historical flood records are inconsistent and incomplete due to variations in data collection practices. By contrast, the use of satellite-derived data offers a more comprehensive and consistent approach to flood modelling.

2. High Time-series Resolution

Conventional probabilistic models typically focus on modelling the intensity and probability of flood events. However, these methods are constrained by the low frequency of flood events, with only a few occurring each year. In the situation that only several years of data are available, this event-based modelling is not sufficient to power a model with strong prediction power. In contrast, GHDL models flood hazards as a daily time series (see Section 4), with the capability to model intraday variations. By combining this approach with high-resolution satellite data, the issue of data scarcity is partially mitigated.

Floods, as a typical secondary catastrophe risk, occur with relatively higher frequency than primary catastrophe risks. As a result, just a few years of high time-granularity data can significantly enhance the capacity to model flood hazards. However, the absence of rare, high-impact events—such as the 2011 Thailand floods—from the data sample can pose challenges for accurate out-of-sample predictions. Despite this, the GHDL model's ability to learn associations between high-resolution meteorological variables (e.g., precipitation) and flood hazards provides a degree of robustness in addressing these challenges. Unlike conventional flood models that may struggle with limited historical observations of extreme events, the GHDL model leverages its ability to learn associations between high-resolution meteorological variables (e.g., precipitation patterns) and flood hazards. This enables it to extrapolate risk beyond observed events, capturing the likelihood and severity of rare but high-impact floods. By emphasising tail risk modelling, the GHDL framework enhances predictive robustness, offering insurers and policymakers a more reliable tool for assessing extreme flood scenarios even in data-limited regions.

2.2.2 Cost-efficient and Physical Information Embedding

While capturing some physical information, GHDL is generally more cost efficient than a conventional physical model. Unlike traditional physical models, which

require intensive resources to build and have limited transferability across regions, the GHDL model integrates multiple major and moderate exposure regions (Section 3.3) to leverage deep learning techniques. By capturing the relationships between flood hazards and climate variables, GHDL establishes dependencies across regions effectively.

The GHDL model demonstrates strong transferability across regions, as it relies on limited region-specific information (Section 3.3). This transferability has been validated through its successful performance in flood hazard prediction for the Mississippi River (Xu et al., 2022) and the Chao Phraya River, as presented in this report.

While the GHDL model requires maintenance to account for changes in flood mechanisms due to infrastructure developments or climate change, this process is straightforward. Maintenance involves feeding newly observed data into the model, without requiring significant computational resources, ensuring its continued accuracy and performance.

We summarise the challenges of flood risk pricing and how GHDL addresses these challenges in the following table:

Table 1: Challenges of flood risk modelling and adaptations made by GHDL

Challenge	Impact on Traditional Flood Models	How the GHDL Model Addresses This Challenge
A. Limited Event Data for Hazard Modelling		
A.1 Dependence on Historical Flood Records	Traditional models rely on well-documented past flood events, making them ineffective in data-scarce regions.	GHDL models flood hazards using daily satellite-derived meteorological data and percentile-based flood risk estimation, enabling robust risk assessment even in regions with limited historical records.
A.2 Absence of Rare, High-Impact Events	Lack of extreme flood events in the data can lead to inaccurate out-of-sample predictions.	GHDL focuses on modelling the tail of the flood risk distribution, improving estimates of low-probability, high-severity events.
A.3 Limited Applicability in Regions with Evolving Climate Patterns	Climate change affects flood frequency and intensity, but event-based models struggle to adapt due to their reliance on past events.	GHDL dynamically learns from meteorological variables, making it more adaptable to changing climate conditions and future flood risks.
B. Cost Barriers for Moderate Exposure Regions		
	Traditional flood models using physical methods can be costly. As a result, only high-risk regions tend to be modelled, while moderate exposure regions are often neglected.	GHDL incorporates geo-connectivity between moderate and high-exposure regions, allowing for flood risk dependencies to be modelled while also improving the interpretability of hazard assessments.
C. Localisation of Global Flood Models		
	Conventional models may struggle with localised flood dynamics due to insufficient resolution in historical data.	GHDL leverages high-resolution meteorological inputs, capturing fine-scale flood patterns more effectively.
D Limited Loss Data for Damage Modelling		
	Lack of reliable damage data makes it difficult to validate and refine flood impact assessments.	GHDL does not directly address damage modelling limitations.

2.3 GHDL's Data Requirements and Assumptions

Empirical analyses have demonstrated that the GHDL model outperforms models of similar complexity (Section 4.4) when the following assumptions are satisfied:

Assumptions for Optimal Performance:

1. **Inland Flood Risk Modelling:**

The GHDL model has been tested and validated in the context of inland flood risk modelling. Both this report and the foundational paper (Xu et al., 2020) focus on the relationship between climate variables and inland flood hazards. Its applicability to coastal floods has not yet been evaluated.

2. **Multiple Regions with Interdependent Flood Hazards:**

A key factor in GHDL's superior performance is its ability to detect and model dependencies among regions. This feature is most effective when several regions with interdependent flood hazards are analysed together.

Data Requirements:

As outline in the previous section, one of the key advantages of the GHDL framework is its ability to operate effectively where conventional flood hazard data is sparse. Instead of relying on traditional data sources, the GHDL model utilises high-resolution satellite data, making it particularly suitable for regions where detailed data may be scarce or unavailable. Below, we outline the specific data requirements for implementing the GHDL model:

1. **Flood Hazard Measure:**

A fairly accurate measure of flood hazard is required (see Section 4.1 for an in-depth discussion).

2. **Climate Variables:**

Climate data with predictive power for flood risks, such as precipitation, temperature, and air pressure, are essential for the model.

3. **Insurance Exposure and Loss Data (Optional):**

If the objective includes building a loss model, insurance exposure and loss data are also required.

We use the context of this report to further illustrate the data requirement and assumptions of GHDL. As a country with high flood risk, Thailand has a long history of gauging hydrological variables at various stations. These variables can serve as measures of flood hazard; however, we find it challenging to map the data acquired from public sources into a generally complete time series. For the Chao Phraya River Basin, we identified two credible public sources of flood hazard data:

- **Global Runoff Data Centre (GRDC):**⁴ GRDC is an international data repository that provides reliable hydrological data for researchers and policymakers. It hosts a global network of gauging station data, including daily water discharge records for the Chao Phraya River Basin from 1960 to 2000. However, the GRDC dataset lacks recent updates, limiting its applicability to contemporary challenges.
- **National Hydroinformatics Data Center (NHDC):**⁵ NHDC is Thailand's official data centre for hydrological and water management information. It provides daily water level data from January 2019 to April 2020 and hourly water flow data from April 2020 onward. This dataset is well-maintained and offers real-time insights into water dynamics within Thailand's river basins.

For this report, we selected the NHDC dataset, utilising daily water level data from January 1, 2019, to October 31, 2024, as our measure of flood hazard. Compared to the GRDC dataset, NHDC provides up-to-date data, making it more relevant to the challenges currently faced by insurers. While neither dataset covers the catastrophic 2011 Thailand flood, the NHDC dataset includes two relatively significant flood events in 2021 and 2022.

Another challenge in preparing this report is the lack of granular insurance loss or economic loss data. In the primary reference for this study, Xu et al. (2022) tested their proposed Geo-Hierarchical Deep Learning (GHDL) method using policy-level claims and exposure data from the National Flood Insurance Program in the

⁴ Global Runoff Data Centre (GRDC), Date of access December 1, 2024: <https://portal.grdc.bafg.de/applications/public.html?publicuser=PublicUser>

⁵ National Hydroinformatics Data Center (NHDC), Date of access December 1, 2024: <https://www.thaiwater.net/>

United States. Unfortunately, such granular claims and exposure data are unavailable in Thailand publicly. Detailed claims and exposure data at the required locations were unavailable, and publicly accessible flood loss data at a sufficiently granular level could not be identified. However, this limitation does not apply to insurers who can use their own loss data to estimate loss information using this model. By leveraging their proprietary data, insurers can effectively apply the GHDL framework to enhance flood risk assessment and develop more accurate pricing and underwriting strategies.

2.4 Flood Hazard Measures for Other River Basins in APAC

As demonstrated by Xu et al. (2022) and further elaborated in this report (Section 4), the GHDL model is highly flexible and unselective regarding the types of data it can utilise. For reference, we provide several sources of flood hazard data for other major river basins:

1. Mississippi River Basin (U.S.), United States Geological Survey Water Data for the Nation: <https://waterdata.usgs.gov/nwis/sw>
2. China Yangtze River (China), 长江水文网: http://www.cjh.com.cn/swyb_sssq.html
3. Mekong River (Vietnam, Laos, Cambodia, Thailand), Mekong River Commission: <https://www.mrcmekong.org/data-and-information-systems-and-services/>
4. Ganges Basin (India, Bangladesh), Hydrological Dataset for Ganges Basin (Sufian, 2024).

3 Deep Learning Approaches to Flood Hazard Modelling

3.1 Objective

Flood hazard modelling is essential for accurate flood insurance pricing. The primary objective of such modelling is to provide the best possible estimation of flood hazards for future periods based on all available information up to the present. We formulate this problem as follows.

Let Q_{it} represent the flood hazard at location $i = 1, \dots, I$ at time $t = 1, \dots, T$. The goal is to forecast Q_{it} with precision, leveraging information from previous time periods up to $t - 1$. Specifically, the forecast is expressed as:

$$\hat{Q}_{it} = \mathbb{E}[Q_{it} \mid \mathbf{Q}_{t-1}, \mathbf{W}_{t-1}] = f(\mathbf{Q}_{t-1}, \mathbf{W}_{t-1}),$$

where $\mathbf{Q}_{t-1} = \{Q_{it} : i = 1, \dots, I, t = t - 1, \dots, t - l\}$ and $\mathbf{W}_{t-1} = \{W_{it} : i = 1, \dots, I, t = t - 1, \dots, t - l\}$ represent the flood hazard and weather information sets for all valid locations from $t - l$ to $t - 1$, respectively.

The function $f(\cdot)$ denotes the mean estimation produced by the constructed deep learning model, which is designed to capture the underlying relationships between the historical flood hazard and weather information to predict future flood hazards effectively.

3.2 Commonly Deep Learning Frameworks for Flood Hazard Factorisation

3.2.1 Convolutional Neural Networks

Convolutional neural networks (CNNs) are a specialised class of deep learning models widely utilised in image recognition and computer vision tasks. These models are particularly suitable for processing image-based precipitation data. Unlike aggregated precipitation data, which may obscure localised variations, image-based precipitation data provides high spatial resolution, capturing detailed patterns across geographical areas. Such granularity is crucial as local landscape variations can cause heavy rainfall at specific locations to result in

disproportionately higher flood hazards. These localised patterns, often lost in aggregated datasets, can be effectively captured and analysed using image-based precipitation data.

The architecture and capabilities of CNNs have been profoundly shaped by several key advancements in the field. The seminal work by LeCun et al. (1998), introducing the LeNet-5 model, laid the foundation for the application of CNNs in document recognition. Building on this groundwork, Krizhevsky et al. (2012) introduced AlexNet, a deeper CNN architecture that achieved a breakthrough in image classification on the ImageNet dataset, underscoring the potential of deep learning in computer vision. Subsequent refinements in CNN design have further enhanced their performance, as demonstrated by models such as VGGNet (Simonyan and Zisserman, 2014), GoogLeNet (Szegedy et al., 2015), U-Net (Ronneberger et al., 2015), ResNet (He et al., 2016), and YOLO (Redmon et al., 2016), among others.

CNNs offer several notable advantages:

- **Local connections:** Instead of connecting each neuron to all neurons in the previous layer, CNNs establish connections with only a subset of neurons, reducing the number of parameters and expediting convergence.
- **Weight sharing:** Groups of connections share the same weights, further decreasing the number of parameters and enhancing computational efficiency.
- **Down-sampling:** Pooling layers leverage local correlation principles to down-sample feature maps, reducing data size while preserving salient information (Li et al., 2022).

A typical CNN system comprises four main components. The **convolution** process is fundamental for feature extraction, producing *feature maps* as output. To address potential information loss at the edges during convolution with a specific kernel size, **padding** is applied by adding zero values around the input, effectively adjusting its dimensions. The **stride** parameter determines the density of convolution operations, with larger strides resulting in less dense feature coverage. Following convolution, the feature maps may contain redundant information, increasing the risk of overfitting. To mitigate this, **pooling** layers

condense the feature maps, reducing redundancy and preserving essential features. The complete workflow of a 2-dimensional CNN system is illustrated in Figure 2.

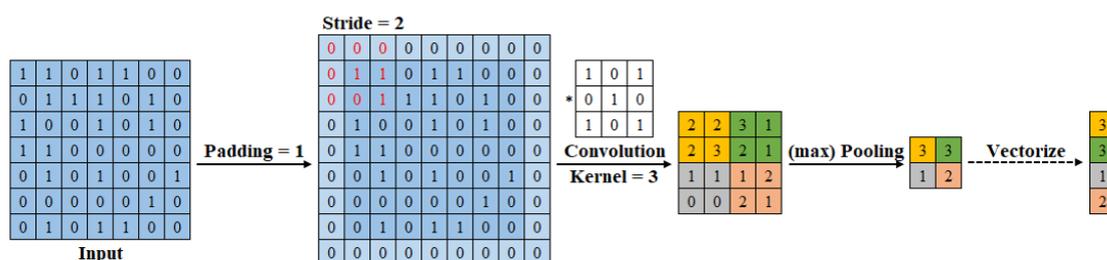


Figure 2: Illustration of padding, stride, convolution, and pooling procedure

This figure demonstrates the procedure of a 2-dimensional CNN system. Here, we take the padding = 1, stride = 2, convolutional kernel size = 3, and max pooling = 2. In the last step, the convolution output is vectorised and ready to be utilised in the following layers.

3.2.2 Wavelet Neural Networks

Wavelet Neural Networks (WNNs) present significant advantages for flood hazard management, making them a valuable tool within our proposed GHDL framework. This section elaborates on the benefits of WNNs and outlines the rationale for their application in analysing the fluvial component of flood hazards. The foundational concept of the WNN approach, the discrete wavelet transform (DWT), is introduced in Section 9.1.

WNNs combine the strengths of DWT and neural networks (NNs) by feeding wavelet-decomposed coefficients into a fully connected neural network architecture. Recognized as a state-of-the-art methodology in flood risk management (Mosavi et al., 2018), WNNs have demonstrated superior model accuracy compared to traditional Fourier transform approaches (Shafaei and Kisi, 2016). These capabilities make WNNs particularly well-suited for capturing complex, non-linear patterns in fluvial flood dynamics.

3.3 Geo-Hierarchical Deep Learning Framework

Leveraging its strengths in handling geo-connected dependencies and hazard correlations in extreme events, this study employs the Geo-Hierarchical Deep Learning (GHDL) framework, specifically designed for flood insurance pricing, to

forecast flood hazards (Xu et al., 2022). This framework integrates hydraulic and meteorological data, hierarchically organised by geographical locations, to enhance predictive accuracy.

Inland flood hazards within a specified region are generally categorised into two types: fluvial and pluvial floods. Fluvial floods are predominantly driven by upstream flood hazard dynamics, whereas pluvial floods are caused by intense localised rainfall. The concurrent occurrence of both types can significantly exacerbate their combined impacts (Chen et al., 2010).

The GHDL framework employs a dual-system methodology to capture the relationships between inland flood hazards and the climatic conditions contributing to both fluvial and pluvial floods simultaneously. Figure 3 illustrates the information integration process in the proposed GHDL model. Precipitation data relevant to pluvial flood analysis are processed through a CNN framework, while flood hazard data critical for fluvial flood analysis are processed using a WNN system. This bifurcated approach allows for a detailed understanding of the distinct mechanisms driving each type of flood. The locally processed data are then merged with integrated data from upstream cities, creating a comprehensive dataset for local flood hazard forecasting. This integrated dataset is subsequently forwarded to downstream cities for further analysis.

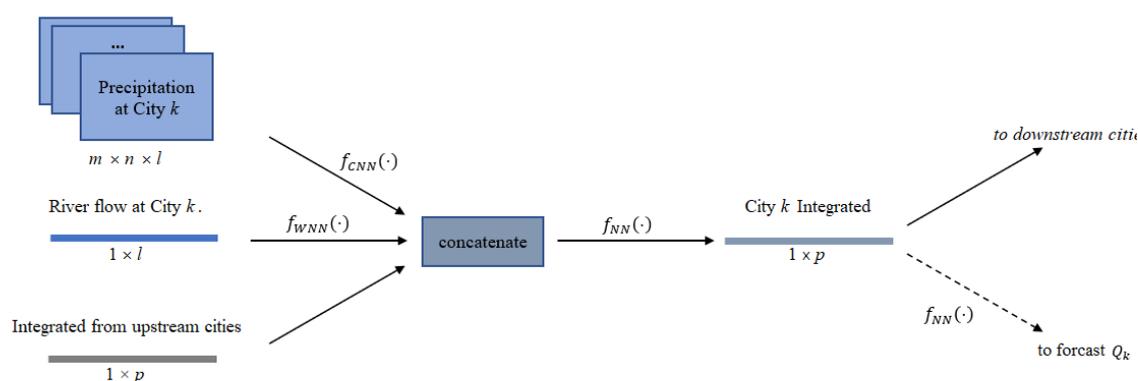


Figure 3: Information integration process of GHDL structure

Source: Figure 2 of Xu et al. (2022). This figure illustrates the information integration process of GHDL structure. For a given city k , the meteorological data (e.g., precipitation) are processed by a CNN system, while the hydraulic data (river flow) are processed by a WNN system. The processed meteorological and hydraulic data are then combined with the integrated information obtained

from the upstream cities. The resulting data are fed into a NN system to generate the integrated output for City k .

The geographical hierarchy of cities (or locations) within a river basin is determined by their connectivity along the river channel. For locations i and j situated along the same river, where i is upstream of j , it is reasonable to assume that the flood hazard at location i may causally impact the flood hazard at location j , but the reverse influence is less significant or negligible. Under the GHDL framework, the flood hazard forecasted at city i is propagated downstream to location j to support flood hazard predictions for j . Concurrently, the observed ground truth at location j is utilised in a feedback mechanism, in the process of deep learning model fitting, to refine the training of the flood prediction submodel for city i .

This framework offers three key advantages:

- **Enhanced interpretability:** By incorporating connectivity into the model, it becomes easier to understand how upstream information influences flood hazard predictions downstream.
- **Streamlined optimisation:** Encoding geographical connectivity narrows the search space for optimisation algorithms, resulting in faster convergence and increased model efficiency.
- **Reduced risk of overfitting:** The integration of connectivity acts as a constraint, simplifying the model by reducing the number of parameters, thereby mitigating the risk of overfitting in complex models.

4 Application of the framework on the Chao Phraya Basin

This section provides a detailed technical overview of the flood hazard estimation process for the Chao Phraya Basin. It also outlines the procedure for setting up the GHDL model for a specific region.

Procedure of setting up GHDL	
1	Select the flood hazard metric (Section 4.1).
2	Collect the data: flood hazard, climate variables, insurance data (Section 4.2).
3	Code the GHDL model with physical connectivity (Section 4.3)
4	Training and select best model for the area of focus (Section 4.3)
5	Estimate the flood hazard for future (Section 4.4)
6	incorporate the flood hazard into the pricing model

In this report, we examined the flood hazard across 11 cities situated within the Chao Phraya River Basin, grouped into five clusters as illustrated in Figure 4. These clusters encompass the majority of the population and economic exposures within the region, ensuring comprehensive coverage of the basin's critical areas.

4.1 Inland Flood Hazard Metrics

Inland flood hazards can be monitored using various measures derived from hydrological and meteorological methods. Here, we present two widely discussed flood hazard measures as examples: water level and water discharge (Quinn et al., 2019):

- **Water level (m):** This metric represents the height of water in a river, serving as a direct indicator of flood risk. When the water level at a specific location exceeds a certain threshold, a flood hazard can be identified. For instance, in Thailand, the National Hydroinformatics Data Center (NHDC)

categorises a location as having a “high” water level if it surpasses the 70th percentile of historical observations at the station. Additionally, the NHDC assigns an “overflow” category when the water level exceeds the riverbank height.

- **Water discharge** (m^3/s): Also known as water flow, this metric measures the volume of water flowing through a specific location over a given period. It is a valid flood risk indicator because of the generally non-decreasing relationship between water level and discharge at any given location. Water discharge is commonly used as a flood hazard measure in countries like the United States (Quinn et al., 2019), where gauging stations often have longer historical records for water flow than for water levels.

Insurers may be more familiar with measures that quantify the intensity of specific flood events, such as:

- **Peak discharge** (m^3/s): The maximum water discharge recorded during a flood event.
- **Mean annual flood** (m^3/s): The average peak flood discharge over a series of years.

These two metrics are essentially derived from water discharge and provide valuable insights into flood risk. They are also easier to use for insurance pricing and claims settlement. However, their application in deep learning models is limited due to the scarcity of observations, which makes it challenging to establish relationships between flood hazards and climate variables. In contrast, water level and water discharge data are often available at finer temporal resolutions, such as daily, hourly, or even minute-level observations.

Given the constraints of data availability, this report uses water level as the primary flood hazard measure for the Chao Phraya River Basin.

4.2 Data

4.2.1 Flood Hazard Data

As previously established, we obtained daily water level data from January 1, 2019, to October 31, 2024, from the NHDC. The NHDC provides historical water level data in the format of images, as demonstrated in Figure 4. The water levels at each gauging station are categorised by NHDC into five levels:

- **Overflow:** Water level above the riverbank (red).
- **High:** Water level above the 70% quantile (blue).
- **Normal:** Water level between the 30% and 70% percentiles (green).
- **Low:** Water level between the 10% and 30% percentiles (yellow).
- **Critically Low:** Water level below the 10% percentile (pink).

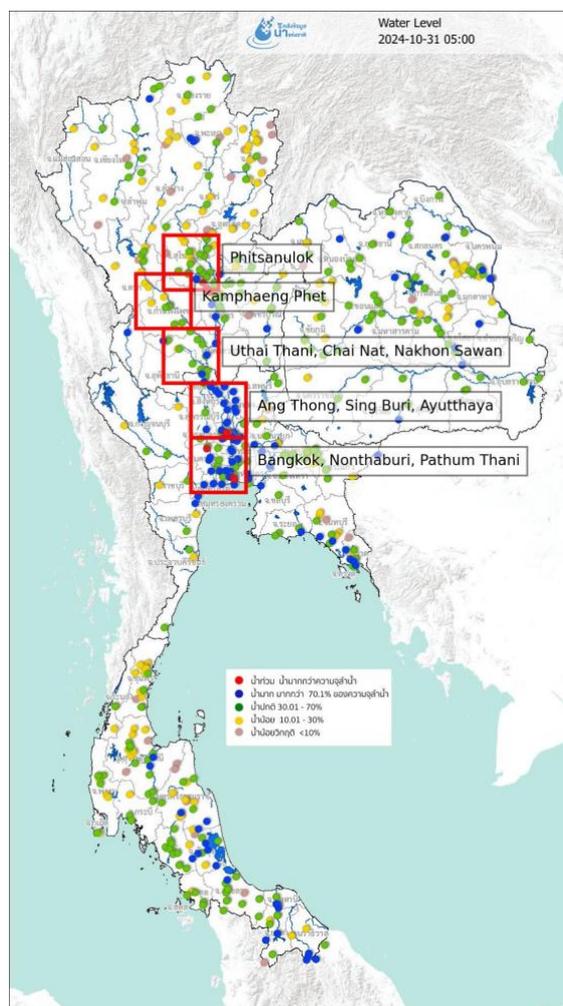


Figure 4: Thailand Water Flow Data as of 2024 from the National Hydroinformatics Data Center (NHDC).

Flood risk was determined using image processing techniques. For a particular cluster and date, pixel colours from the images were rounded, and the number of pixels of each colour was counted within predefined rectangles. To focus on relevant information, we excluded colour counts such as water bodies, land, and other complementary information. Additionally, the “Critically Low” category was excluded since gauging stations typically report this level when they are out of function. This left four relevant categories: Overflow (red), High (blue), Normal (green), and Low (yellow). We calculated the ratio of each category for each date and cluster.

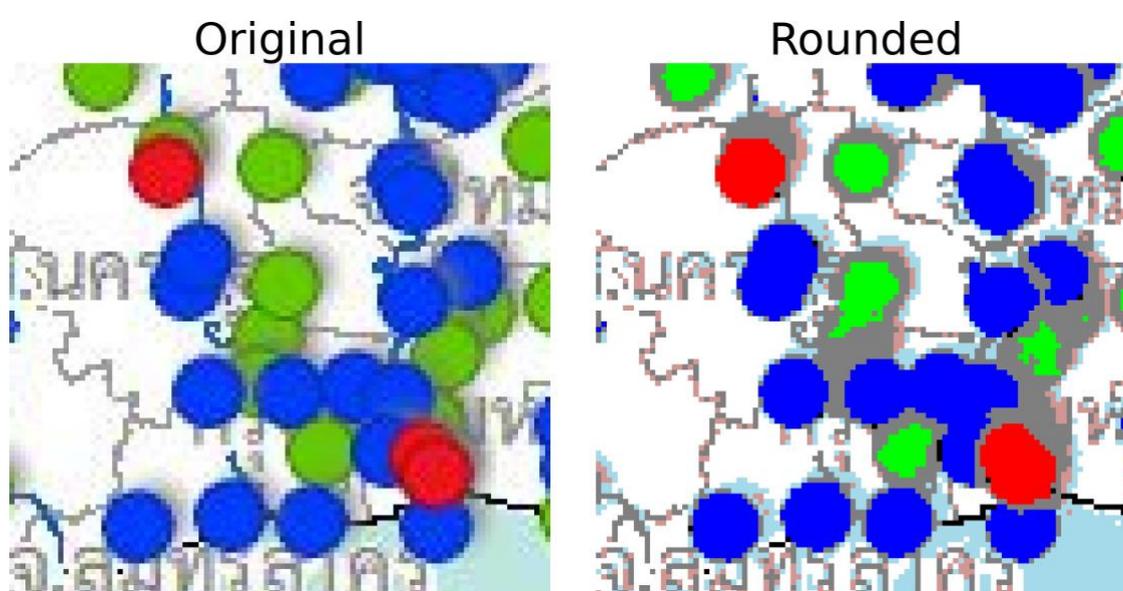


Figure 5: Image processing of water level data.

Figure 5 illustrates the distribution of category ratios across the sample period. The flood hazard distribution varies significantly across regions:

- **Bangkok, Nonthaburi, Pathum Thani:**
 - a. This region exhibits a significant proportion of Overflow hazard events, indicating heightened flood risk likely due to low elevation, urban density, and proximity to the river mouth.
 - b. While extreme hazards are frequent, the Normal hazard category dominates overall coverage.

- **Ang Thong, Sing Buri, Ayutthaya:**
 - a. The region shows a balanced mix of Normal, Low, and High hazard categories. Overflow events are less frequent compared to Bangkok.
 - b. This distribution reflects the interplay of upstream river dynamics and localised rainfall.
- **Uthai Thani, Chai Nat, Nakhon Sawan:**
 - a. Higher frequencies of Normal and Low hazard events are observed, with minimal Overflow occurrences.
 - b. This distribution reflects the upstream region's role as a source of river flow and reduced localised flood impacts.
- **Kamphaeng Phet:**
 - a. The region experiences localised Overflow hotspots, potentially due to specific terrain features or hydrological conditions.
 - b. High and Normal hazard categories dominate, indicating diverse flood conditions.
- **Phitsanulok:**
 - a. Flood hazards are dominated by Normal and Low categories, with limited Overflow events. This is likely due to the region's upstream location and higher elevation.

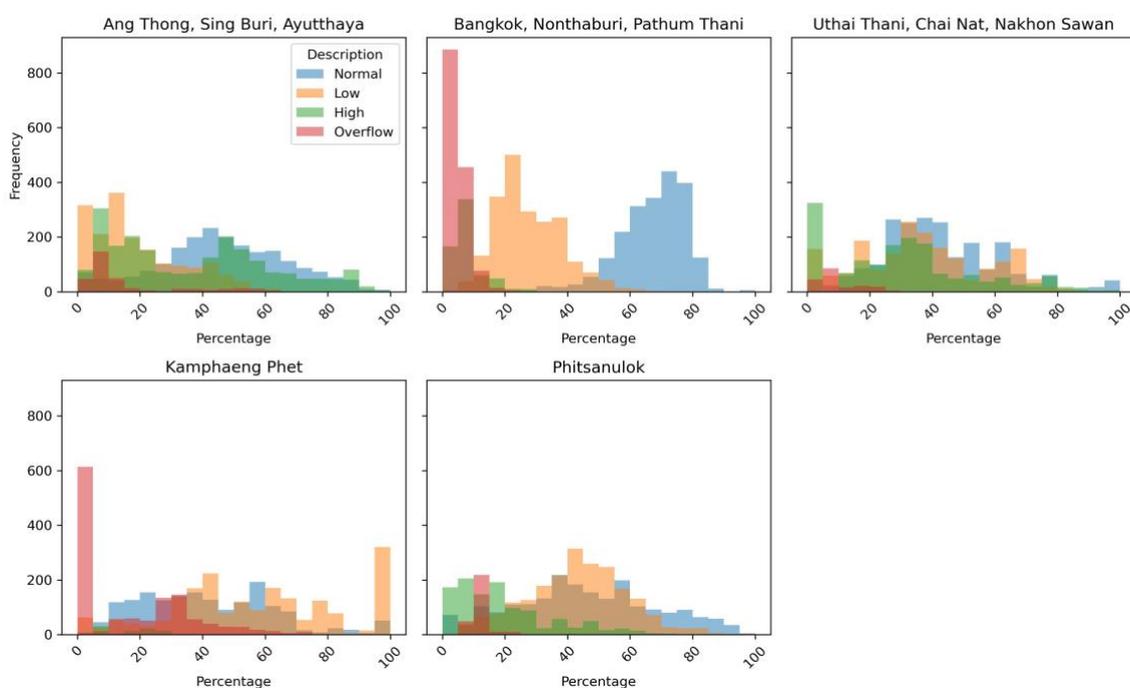


Figure 6: Flood hazard distribution by category for each location. For each day, the percentages of Normal, Low, High, and Overflow categories sum to 100%. Each subplot presents a histogram of the daily distribution of these four categories.

To facilitate analysis, we aggregated the flood hazard ratios into a single variable representing the flood hazard at each location and time, using the following equation:

$$Q_{it} = 5 \cdot Overflow_{it} + 2 \cdot High_{it} + 0 \cdot Normal_{it} + 0 \cdot Low_{it}.$$

For the 23 missing dates in the sample, we imputed values using the historical mean of the corresponding day of the year. Figure 7 shows the histogram of the aggregated flood hazard for each location, and Figure 8 illustrates the time series of aggregated flood hazard values across the sample period. For clarity, the time series was smoothed using a moving average from $t - 1$ to $t + 1$. Missing dates and significant flood events in 2021 and 2022 are also highlighted.

The aggregated flood hazard measure aligns well with the categorical flood hazard data. Based on the weight selected, an aggregated flood hazard value exceeding 2 indicates a high flood hazard at a given location and time. Hereafter, we refer to the aggregated flood hazard as “flood hazard” for simplicity and use it for subsequent analysis and discussions.

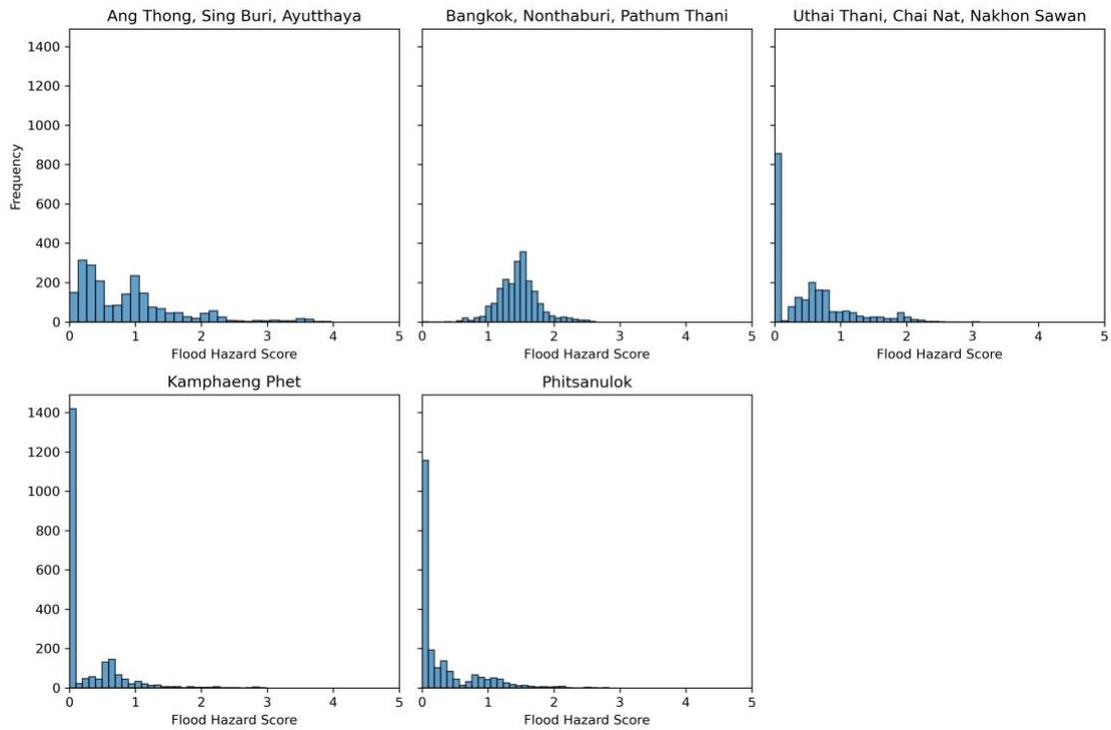


Figure 7: Histogram of flood hazard.

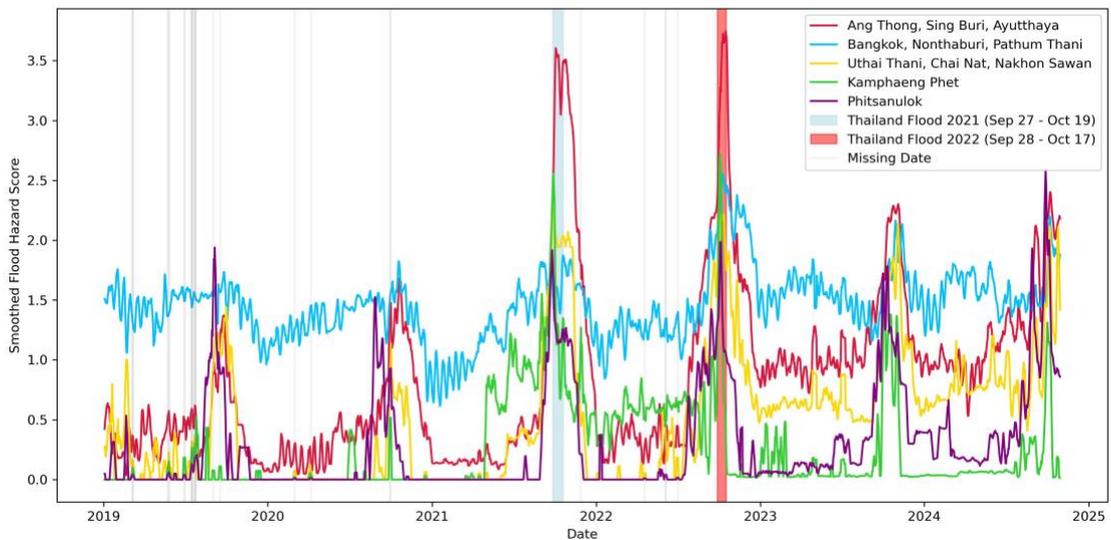


Figure 8: Time series of flood hazards for each location.

Flood hazard also exhibits significant seasonality at both short and long time horizons. While such seasonality will naturally be carried by independent variables used for forecasting, we removed yearly seasonality for convenience by subtracting the 30-day rolling historical average centred on each observation’s day of the year. The demeaned flood hazard values are depicted in Figure 9.

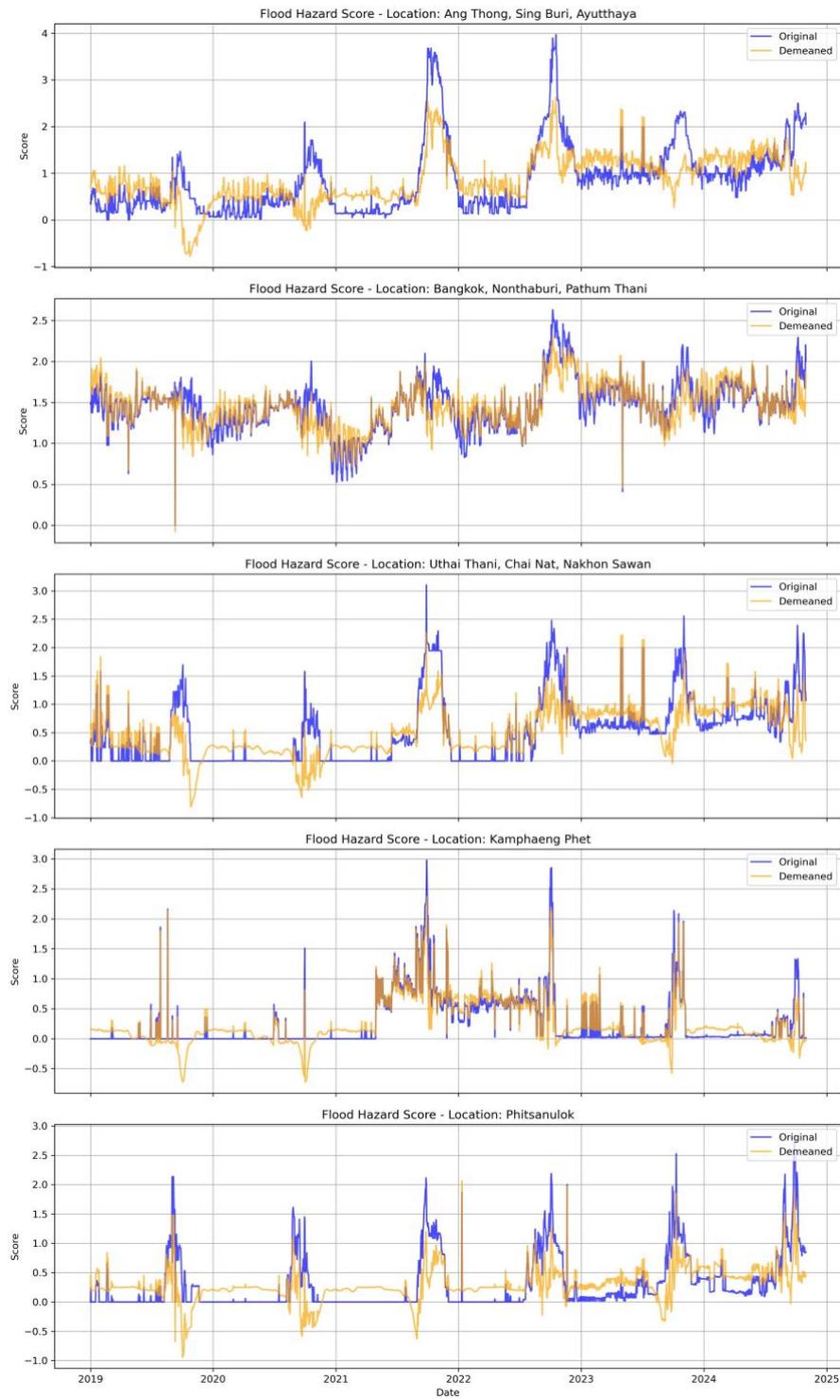


Figure 9: Demeaned vs original flood hazard time series by locations.

4.2.2 Independent Data

This report utilizes precipitation image data as the independent variable to predict flood hazard risk. The fluvial component of the GHDL framework is omitted in this analysis (Figure 3). We focus solely on the pluvial segment of GHDL because there are no credible long-term projections for fluvial data (such as water level or discharge). This is due to fluvial dynamics being influenced not only by climate change but also by infrastructure developments and other human interventions. Moreover, pluvial information plays a more significant role in flood modelling and has a broader impact than fluvial data. As a result, climate scientists often exclude fluvial components from climate simulations. By omitting the fluvial segment, we simplify the model without compromising its practical application—particularly for one-year forward projections in insurance pricing. Our results demonstrate robust predictive accuracy (Section 4.4). Precipitation data for the period from December 1, 2022, to October 31, 2024, was obtained from the ERA5 global reanalysis hourly dataset on single levels (Hersbach et al., 2020). The selected geographical area spans from 4°N to 21.5°N and 96°E to 107.25°E, encompassing the entire region of Thailand. The ERA5 hourly dataset provides total precipitation data at a spatial granularity of 0.25° by 0.25°, resulting in images with dimensions of 70-by-45 grids for each hourly observation.

To ensure compatibility with the flood hazard data processing pipeline, the precipitation data underwent the following preprocessing steps:

1. **Temporal Aggregation:** Hourly precipitation data were aggregated to daily frequency by summing precipitation values for each grid over 24-hour periods.
2. **Seasonality Adjustment:** The yearly seasonality of each grid was removed by subtracting the 30-day rolling historical average, centred on each observation's day of the year, grid by grid.

Figure 10 illustrates a comparison between the original and demeaned precipitation images for September 9, 2021. This preprocessing ensures the data is properly aligned with the flood hazard analysis framework, enhancing the accuracy of the subsequent modelling.

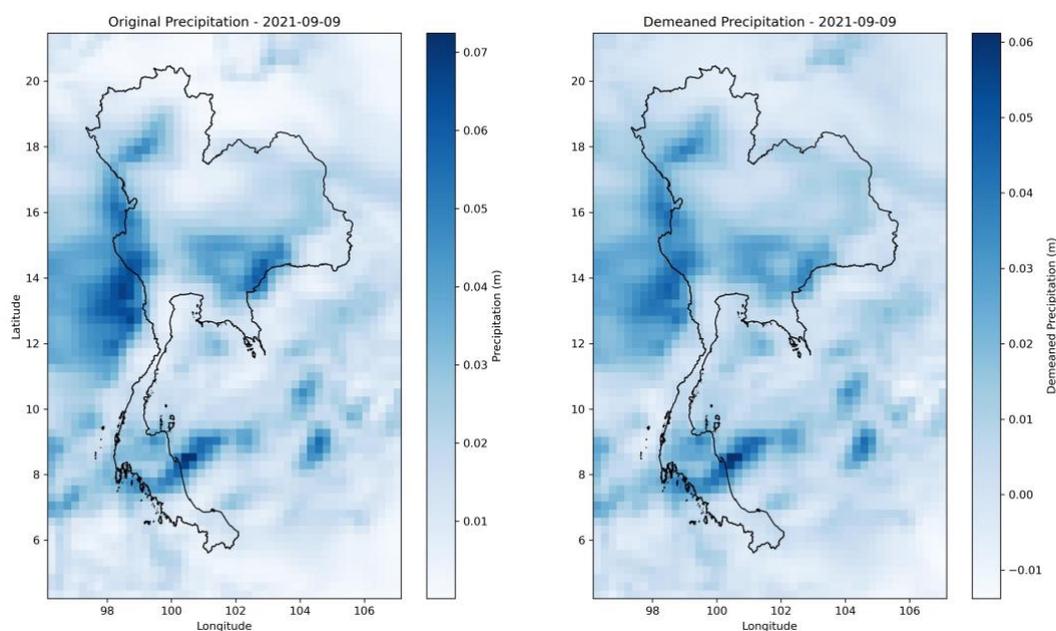


Figure 10: Demeaned vs original precipitation September 9, 2021.

4.3 Model Specifications and Experimental Design

Following the underwriting and renewal practices of property and casualty (P&C) insurance, which typically operate on an annual frequency, we extend the prediction gap to 365 days and use the previous 30 days of precipitation data as input. For example, to predict the flood hazard on date t , precipitation images from $t - 395$ to $t - 366$ are utilised.

The Geo-Hierarchical Deep Learning (GHDL) connectivity structure for the area of interest is illustrated in Figure 11. Phitsanulok and Kamphaeng Phet are two leaf regions with no upstream areas. The remaining three regions are descendant regions that can utilise upstream information to refine their predictions.

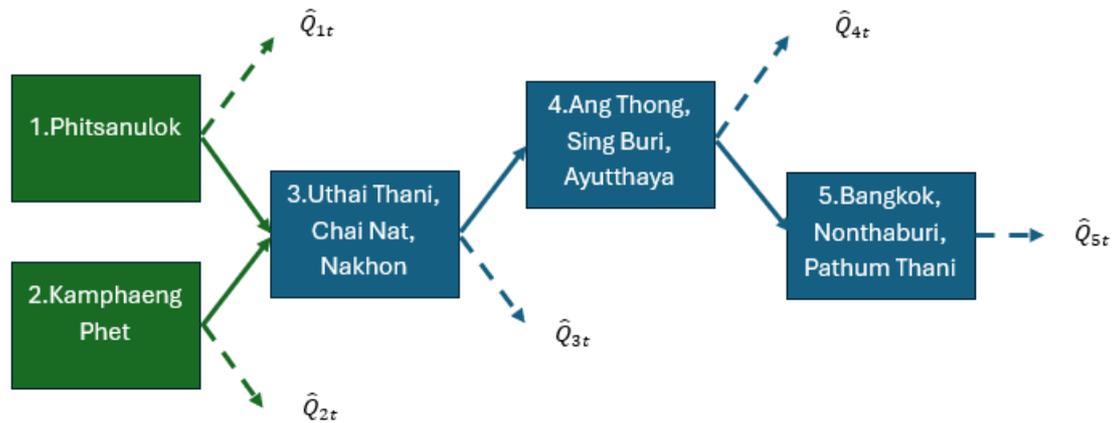


Figure 11: GHDL model structure for Chao Phraya River basin.

For each submodel, we employ a three-dimensional CNN with two hidden layers of architecture 32–64, using cubic kernels of size 3, with stride and padding set to 1. Three-dimensional max-pooling layers with kernel size 2 and stride 2 are applied to reduce the height and width of feature maps after each CNN layer. Following the second CNN layer, the flattened output is transformed into a vector of length 128 using a fully connected neural network (NN) layer. This vector is utilised both for predicting the flood hazard at the current location and for feeding forward to refine the predictions of downstream regions. Each deep learning layer employs the Rectified Linear Unit (ReLU) activation function to introduce non-linearity.

The objective function for the GHDL model combines the mean squared error (MSE) and an L2 norm penalty. The MSE evaluates the accuracy of the model, while the L2 penalty mitigates overfitting by constraining the magnitude of the model parameters.

To ensure a complete calendar year of data is included in the testing set, we randomly draw one year for each day of the calendar year from the sample period (2019–2024). For instance, for January 1, a random draw is made from January 1 of 2019 to 2024, and this sample is added to the testing set. This process is repeated for all other days of the year. The remaining data forms the training set. We use a three-fold cross-validation approach to validate the model. The random sampling process is repeated multiple times, and the results demonstrate consistent robustness.

Additionally, to enhance model convergence, we normalize the flood hazard data for each location.

As a benchmark, we select a local-context deep learning (LCDL) model with the same submodel architecture as the GHDL. LCDL does not incorporate the geo-connectivity structure but is given the same number of parameters to ensure complexity fairness. Additionally, LCDL utilizes exactly the same input information as GHDL, ensuring that any performance differences can be attributed to the inclusion of geo-connectivity rather than differences in model capacity or data availability.

4.4 Estimation of Flood Hazard

Figure 12 presents the loss curves, measured as MSE with an L2 regularization penalty, to evaluate the modelling performance of the GHDL and LCDL models across training, validation, and testing datasets over multiple epochs. The red and grey dashed lines indicate the loss value and epoch at which convergence occurs, respectively.

The GHDL model demonstrates faster convergence compared to the benchmark LCDL model and achieves superior performance across both training and validation metrics. This highlights the effectiveness of incorporating geo-hierarchical connectivity for flood hazard prediction. The GHDL model's ability to leverage upstream and downstream information provides a significant advantage in capturing the spatial and temporal dynamics of flood hazards.

In contrast, the LCDL model, while simpler, is unable to achieve comparable accuracy. The lack of hierarchical dependencies limits its capacity to fully capture the complex relationships inherent in flood hazard prediction tasks.

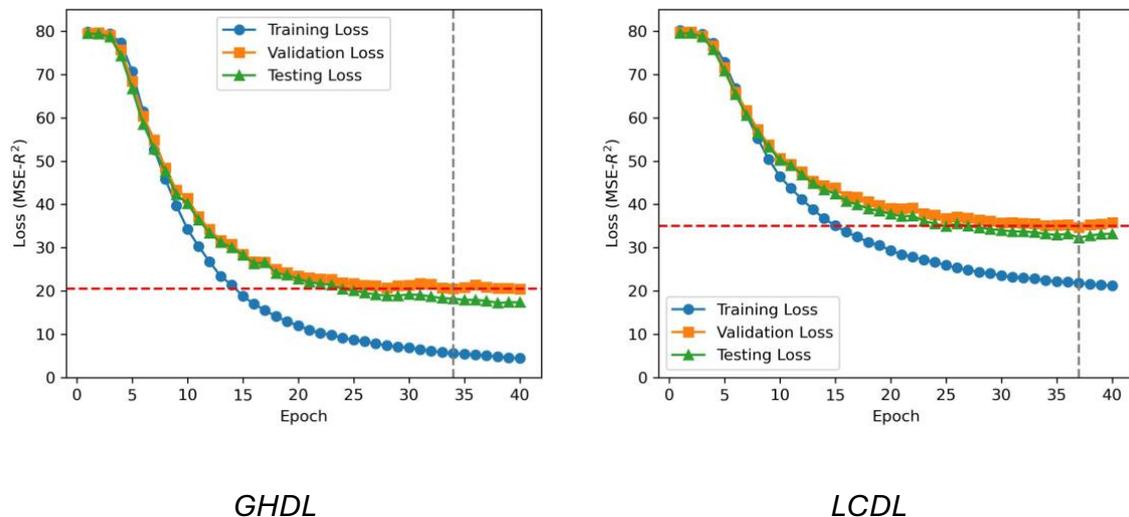


Figure 12: Training, validation, and testing loss by epoch. GHDL achieves convergence at a lower loss across all three phases than LCDL.

We further evaluate the forecasted flood hazards by examining model performance for each observation, with particular attention to extreme situations where flood hazards are high and trigger concerns. Figures 13 and 14 present the time series of observed versus predicted flood hazards for the testing set, along with their kernel density estimates (KDE) for GHDL and LCDL, respectively.

In both figures, high flood hazard dates (observed ≥ 2) are highlighted, with red circles indicating missed predictions (predicted ≤ 1.5) and green triangles indicating correctly captured predictions (predicted > 1.5). The lower-right corner of the left panels reports the MSE and the ratio of correctly captured high-risk events. The right panels of both figures show the KDE of predicted and observed flood hazards by location.

Consistent with the earlier loss analysis, the MSE for GHDL (ranging from 0.18 to 0.27 across locations) is consistently lower than that of LCDL (ranging from 0.29 to 0.56). In terms of high-risk event capture, the GHDL model significantly outperforms the LCDL model in descendant regions, while the performance is comparable in leaf regions. The KDE analysis further reveals that the GHDL model's predicted hazard distribution aligns more closely with the observed distribution than the LCDL model, particularly in the high-risk tail, indicating better predictive accuracy for extreme events.

The observations presented in this report are generally consistent with the findings of Xu et al. (2022).

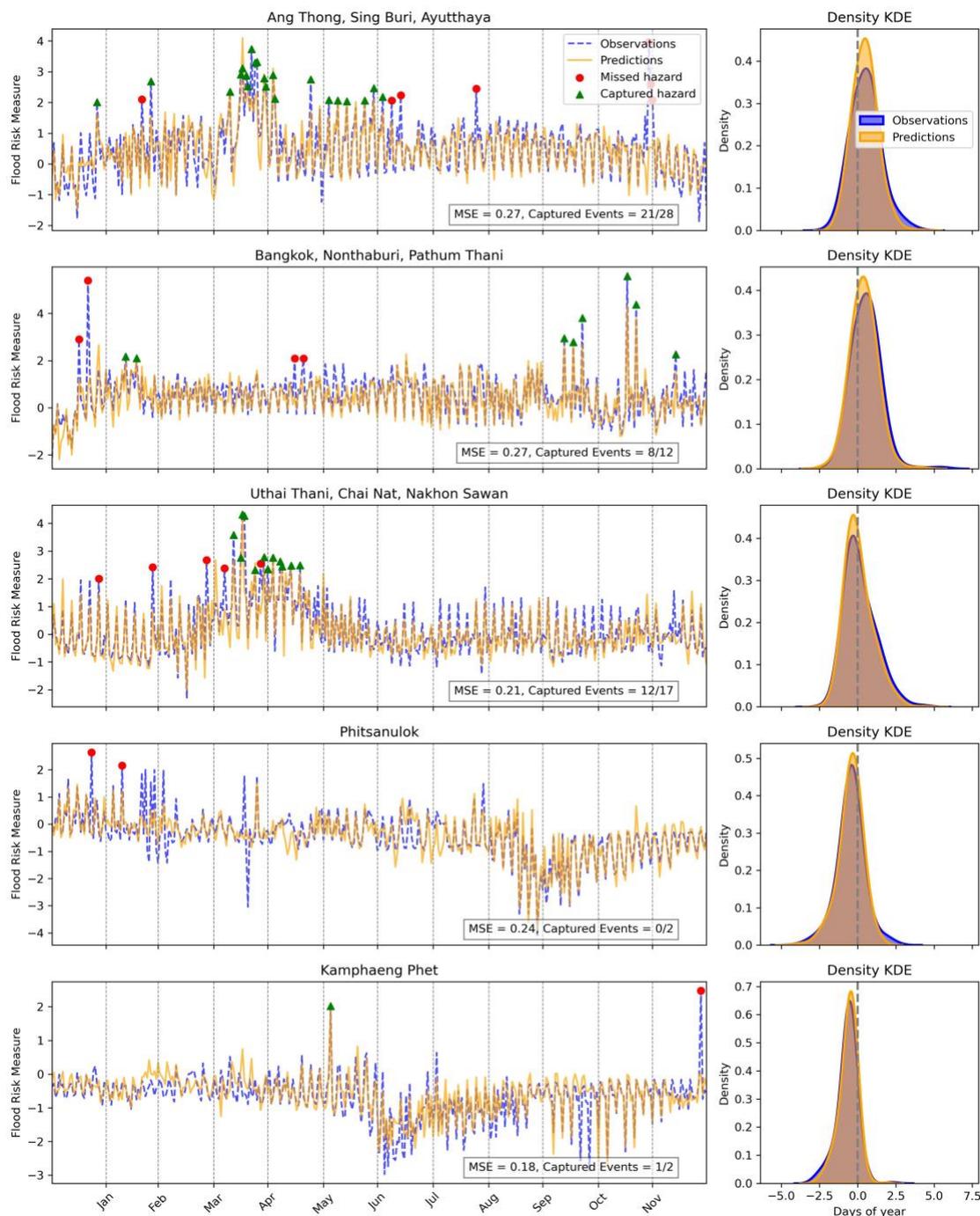


Figure 13: GHDL testing set prediction vs observation. Results for five locations are shown in separate panels from top to bottom. In each panel, the left subfigure presents the observed (blue) vs. predicted (orange) flood hazard time series. Flood hazard dates exceeding two standard deviations above the mean are highlighted, with correctly captured events marked by green triangles and missed events marked by red circles.

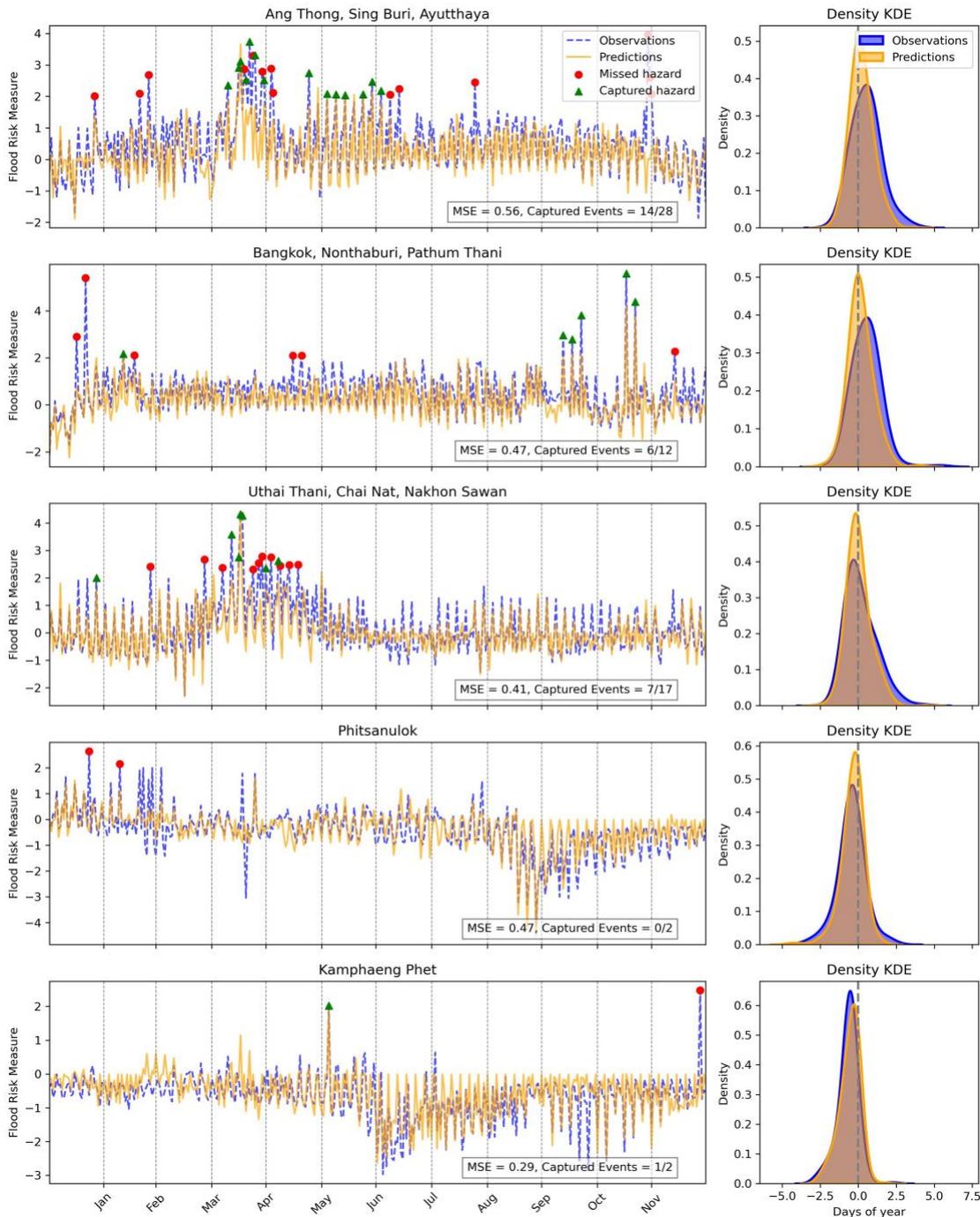


Figure 14: LCDL testing set prediction vs observation. Results for five locations are shown in separate panels from top to bottom. In each panel, the left subfigure presents the observed (blue) vs. predicted (orange) flood hazard time series. Flood hazard dates exceeding two standard deviations above the mean are highlighted, with correctly captured events marked by green triangles and missed events marked by red circles.

5 Premium Pricing and Loss Estimation with Flood Hazard Factors

As previously noted, we were unable to obtain credible insurance loss data with sufficient granularity to facilitate pricing analysis for this report. However, to demonstrate how the flood hazard factors predicted by the deep learning models can be applied in insurance pricing, we outline the methodological framework in this section.

5.1 Methodological Framework

Let Y_{kt} be a random variable representing the total flood loss associated with the k th insured building at location i during the time period $[t - 1, t]$ ($i = 1, \dots, I; k = 1, \dots, K, t = 1, \dots, T$)⁶. In practice, we estimate Y_{kt} by multiplying its corresponding damage ratio $U_{kt} \in [0,1]$, with the insured building value θ_k that is, $Y_{kt} = \theta_k \cdot U_{kt}$. Given a deductible d_k and a limit l_k the total claim of a flood insurance policy Y_{kt}^L is defined as below:

$$Y_{kt}^L = \begin{cases} 0 & \text{if } Y_{kt} < d_k, \\ Y_{kt} - d_k & \text{if } d_k \leq Y_{kt} < l_k, \\ u_k - d_k & \text{if } Y_{kt} \geq l_k. \end{cases}$$

To simplify the analysis, we take $d_k = 0, l_k = \theta_k$, that is $Y_{kt}^L = Y_{kt}$. Conceptually, for a specific insured building k , there should exist a monotonically increasing damage function $h_k: Q_{it} \rightarrow U_{kt}$ that maps the flood hazard $\{Q_{it} = q_{it}\}$ to the corresponding loss event $\{U_{kt} = u_{kt}\}$. Let $f_{Q_{it}}$ represent the probability density function of Q_{it} . The pure premium, p_{kt} , can be obtained by applying this damage function to the expected value of the flood hazard:

$$\begin{aligned} p_{kt} = \mathbb{E}[Y_{kt}^L] &= \int_0^\infty u_{kt} \theta_k \cdot f_{Q_{it}}(q_{it}) dq_{it} = \int_0^\infty h_k(q_{it}) \theta_k \cdot f_{Q_{it}}(q_{it}) dq_{it} \\ &= \theta_k \mathbb{E}[h_k(Q_{it})]. \end{aligned}$$

⁶ Note that Y_{kt} does not include the indexing of i because the k th insured building corresponds to only one gauging site in our setting. Thus, indexing flood loss by k is sufficient to represent it.

In this paper, we utilize a set of generalized linear models (GLM) to establish the relationship between flood hazard, building characteristics, and the loss ratio. We denote the claim frequencies (i.e., number of claims incurred) as N_{kt} , and the severity (i.e., loss ratio per claim) as U_{kt} , then we have

$$\mathbb{E}[Y_{kt}^L] = \theta_k \mathbb{E}[N_{kt}] \mathbb{E}[U_{kt}].$$

We create separate GLMs for N_{kt} and U_{kt} , utilizing the forecasted expectation of river flow, \hat{Q}_{it} , provided by the GHDL model, and the individual-level building characteristics, $\mathbf{x}_k = (x_{k1}, \dots, x_{kp})'$:

$$\begin{aligned} \nu_{kt} &= \mathbb{E}[N_{kt} | \hat{Q}_{it}, \mathbf{x}_k] = g_n^{-1}(\lambda_n \hat{Q}_{it} + \mathbf{x}'_k \beta_n), \\ \mu_{kt} &= \mathbb{E}[U_{kt} | \hat{Q}_{it}, \mathbf{x}_k] = g_u^{-1}(\lambda_u \hat{Q}_{it} + \mathbf{x}'_k \beta_u), \end{aligned}$$

where g_n and g_u are link functions with respect to frequency and severity GLM models, respectively; λ_n and λ_u are their regression coefficients which correspond to flood hazard factor; and β_n and β_u are $p \times 1$ regression coefficients associated with building characteristics. Here, (λ_n, β_n) and (λ_u, β_u) are estimated from historical observations. Finally, the net premium can be calculated as follow:

$$p_{kt} = \theta_k \cdot \nu_{kt} \cdot \mu_{kt}.$$

6 Projecting Future Flood Hazards for the Chao Phraya River Basin

Understanding and mitigating future flood hazards require robust projections that integrate climate simulation data with advanced modelling techniques. In this section, we employ climate projections from the latest generation of climate simulations, coupled with our deep learning framework, to assess potential flood hazards in the Chao Phraya River Basin under various climate scenarios.

By integrating these climate projections with the flood hazard relationships learned by the deep learning model, we aim to generate spatiotemporally resolved forecasts of future flood risks. This approach provides valuable insights into how climate change could alter flood dynamics in the Chao Phraya Basin and enables scenario-based risk assessments, which are critical for underwriting decisions, pricing strategies, and portfolio risk management in the insurance industry.

6.1 Climate Change Simulations

Climate simulations are computational models that replicate Earth's climate system. These simulations are built on physical laws, enabling predictions of climate variables like temperature, precipitation, and sea-level changes under different greenhouse gas emission scenarios (Geophysical Fluid Dynamics Laboratory, 2024).

While these models operate on a global scale, downscaling techniques allow their results to be tailored for regional analyses, making them invaluable for understanding localized impacts, such as those on river basins.

6.2 Shared socioeconomic pathways

Climate projections are derived from distinct Shared Socioeconomic Pathway (SSP) scenarios, which encapsulate a spectrum of plausible futures shaped by varying socioeconomic trends, emissions trajectories, and climate-driving

processes. In our projection analysis, we focus on four key pathways that are most extensively discussed in the climate research literature:

- **SSP1-2.6:** A strong mitigation scenario that represents a slightly less stringent alternative. Emissions reach net zero shortly after 2050, with global temperature rise constrained to below 2.0°C by 2100.
- **SSP2-4.5:** A “middle-of-the-road” scenario where societal and technological progress continues at current rates. Emissions begin to decline after 2050, but net zero is not achieved by 2100. This results in a global temperature rise of approximately 2.7°C by 2100.
- **SSP3-7.0:** A medium-to-high emissions scenario where global cooperation weakens, and regional competition dominates. Emissions roughly double by 2100, leading to a temperature rise of about 3.6°C above pre-industrial levels by the end of the century.
- **SSP5-8.5:** A high-emissions pathway characterized by fossil fuel-driven economic growth and minimal climate policy intervention. Emissions more than double by 2050 and continue to rise throughout the century, causing a global temperature increase of approximately 4.4°C by 2100.

These scenarios provide a framework for exploring the range of potential climate futures, helping policymakers and researchers assess the implications of socioeconomic decisions and climate action.

6.3 Climate Simulation Dataset

We obtain daily precipitation (*pr*) simulation data from January 1, 2015, to December 31, 2100, from the Coupled Model Intercomparison Project Phase 6 (CMIP6) dataset⁷ for four SSP scenarios. For each scenario, we select the default variant (*r1i1p1f1*) from MIROC6⁸ as the representative model. Since the CMIP6

⁷ CMIP6, accessed on December 1, 2024: <https://esgf-ui.ceda.ac.uk/cog/search/cmip6-ceda/>

⁸ MIROC6 (Model for Interdisciplinary Research on Climate version 6) is a state-of-the-art global climate model (GCM) developed by a collaboration of Japanese research institutions, including the University of Tokyo, National Institute for Environmental Studies (NIES), and the Japan Agency for Marine-Earth Science and Technology (JAMSTEC). It is a key contributor to the Coupled Model Intercomparison Project Phase 6 (CMIP6), which provides standardized climate projections for research and policy-making.

precipitation data covers the global scale, we crop it to focus on our region of interest and apply an interpolation method to make the precipitation images compatible with the input requirements of our deep learning model.

Given that the ERA5 dataset used for model training provides precipitation data in meters per day (m/day), while CMIP6 *pr* is reported in units of $\text{kg m}^{-2} \text{s}^{-1}$, we convert the CMIP6 *pr* values to total daily precipitation. The processed precipitation images are then used to infer future flood hazards for the Chao Phraya River Basin, enabling projections for the 21st century.

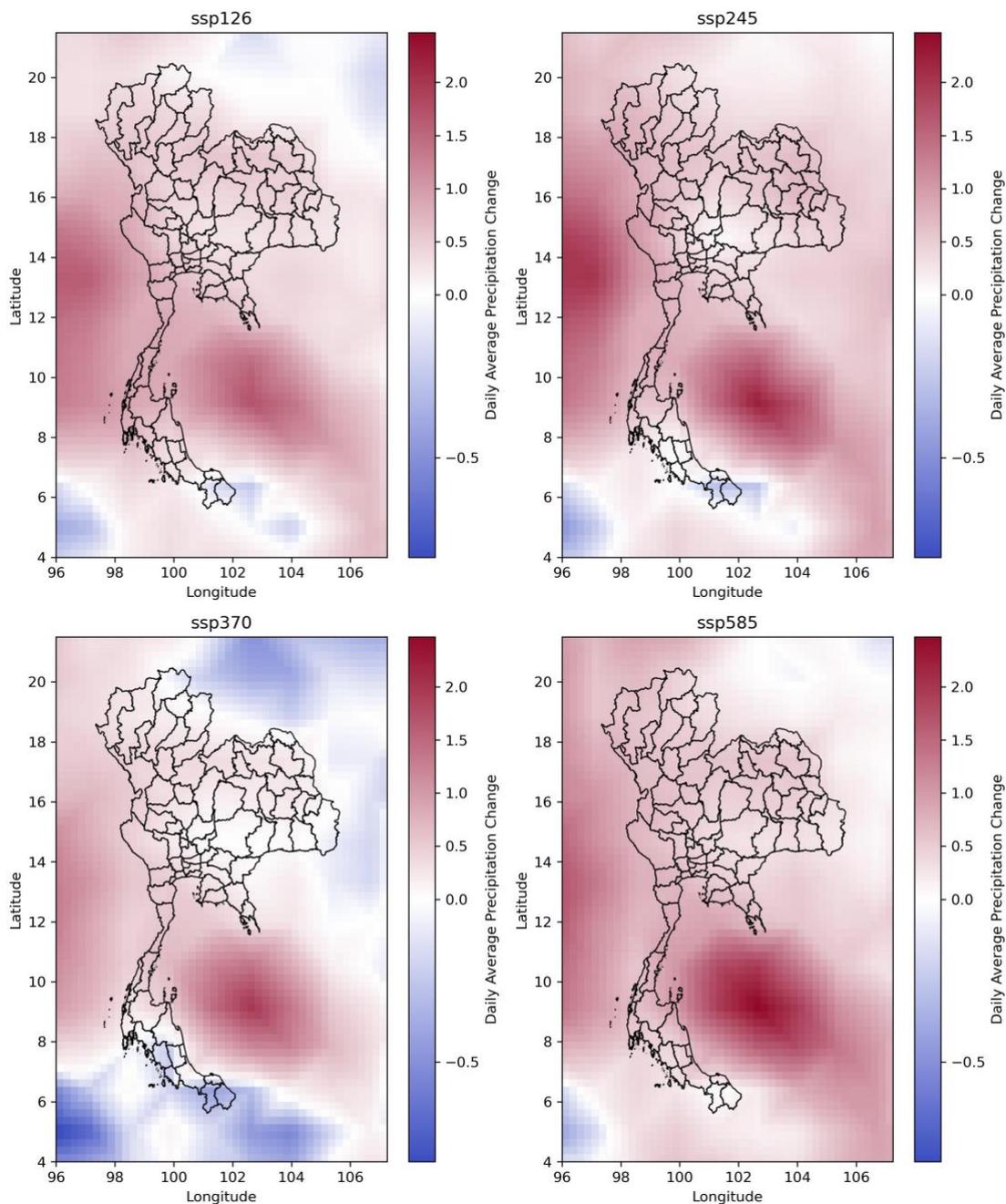


Figure 15: Projected changes in CMIP6 daily average precipitation between 2050--2100 and 2000--2014.

6.4 Flood Risk Projections and Results

Figure 16 presents the annual total precipitation time series for the Thailand region under each SSP scenario. The blue dots represent the total precipitation values, while the red line indicates the fitted linear trend over the years. Except for SSP126, the other three scenarios generally show an increasing trend in total precipitation.

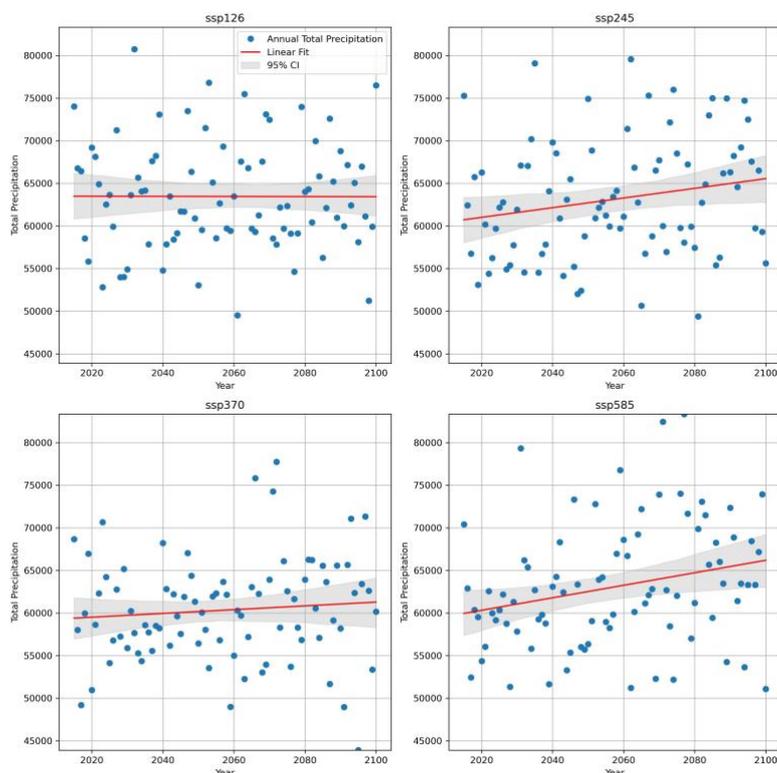


Figure 16: CMIP6 annual total precipitation by year.

The climate simulations reveal a significant increase in precipitation over the Gulf of Thailand and the Andaman Sea, with the SSP585 scenario showing the most pronounced increase, followed by SSP245, SSP370, and SSP126. In inland Thailand, precipitation also generally increases, albeit at varying magnitudes and locations. In the middle of the country, we see less significant increases even decreases in total precipitation.

Following the climate projections, we quantified the annual number of high flood hazard days (defined as flood risk hazard > 2) and high drought hazard days (defined as flood risk hazard < -2) for each location and SSP scenario. To assess relative changes compared to current climate conditions, we subtracted the

average annual high flood and drought hazard counts for the baseline period 2019–2024 from the projected values.

Figure 17 presents the changes in high flood and drought hazards based on CMIP6 climate projection inferences for the period 2019–2024. The box plots display the distributions of high flood hazard days (right) and high drought hazard days (left) for each location and SSP scenario. As the results represent changes relative to the baseline, distributions predominantly above the zero line (dashed black) indicate an increase in risk, while distributions below the zero line suggest a decrease in risk.

Regions near the Gulf of Thailand, particularly Bangkok, Nonthaburi, and Pathum Thani, exhibit a significant increase in flood risk and a decrease in drought risk, likely driven by the pronounced rise in precipitation (Figure 17). Similar patterns are observed in Ang Thong, Sing Buri, Ayutthaya, as well as Uthai Thani, Chai Nat, and Nakhon Sawan for SSP245 and SSP370, though the magnitude of changes is less pronounced compared to Bangkok.

Phitsanulok demonstrates relatively stable risks for both flood and drought, with some improvements under SSP126. In contrast, Kamphaeng Phet shows a consistent decrease in flood risk and an increase in drought risk across all SSP scenarios.

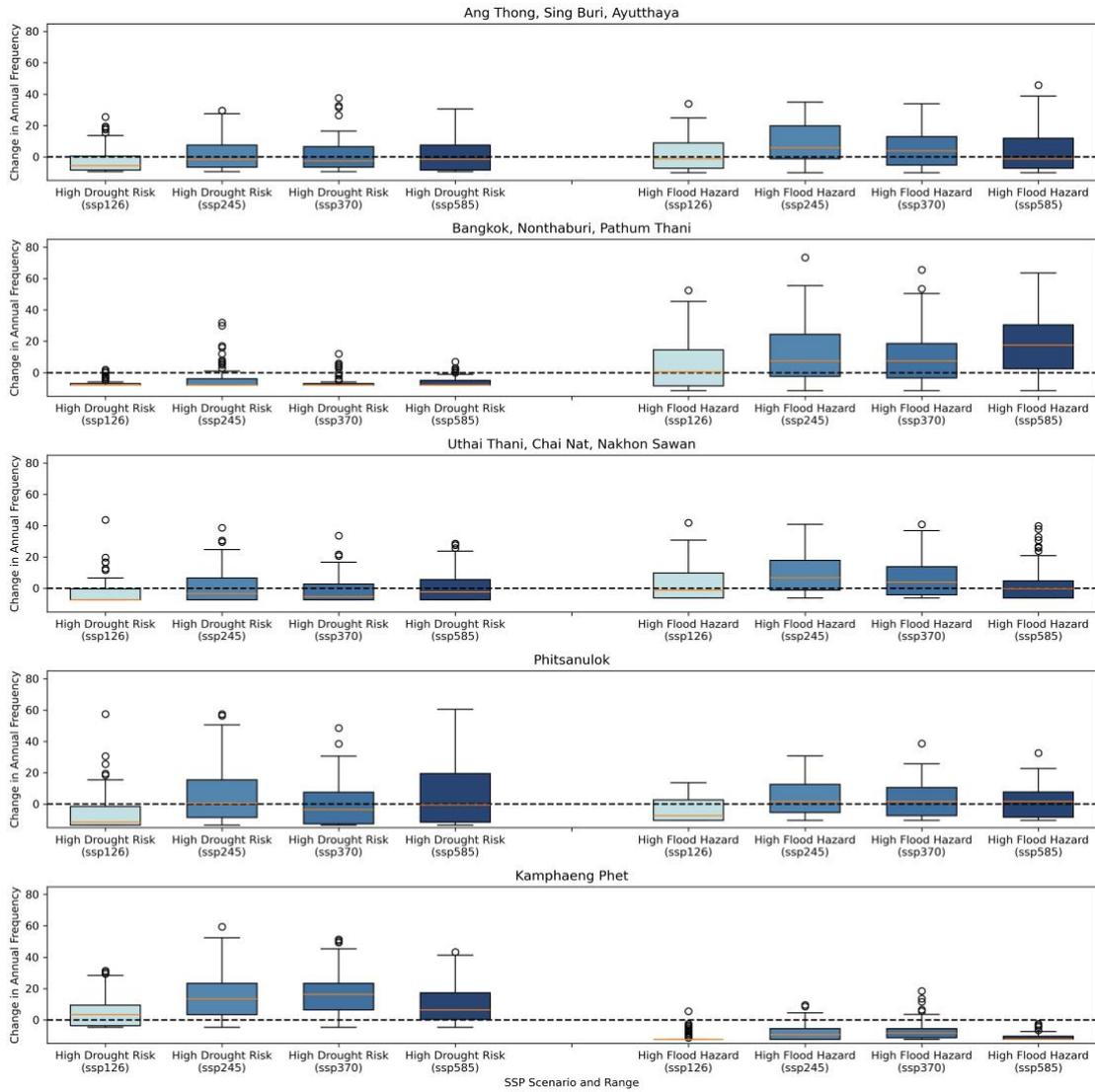


Figure 17: Changes in flood hazard based on CMIP6 climate projection inferences for the period 2019–2024.

7 *Comparison with Commercial Flood Risk Models*

In this section, we compare the GHDL model specifications for the Chao Phraya River Basin with two prominent commercial flood risk models in the market: the RMS Inland Flood Risk Model (Jankowfsky et al, 2016) and the SCOR Oasis LMF Flood Model (Rimkus et al, 2016). For the RMS model, since the methodology manual for Moody's RMS Southeast Asia Inland Flood Risk Model is not publicly available, we use the manual for the RMS US Inland Flood Risk Model as a reference.

Given the absence of insurance loss data, our discussion in this section will take a more qualitative approach. The GHDL model offers benefits as a complementary and alternative tool compared to commercial models developed with extensive resources and engineering-driven pricing methodologies. The GHDL model excels in integrating future climate projections and addressing data limitations, making it particularly valuable for regions with limited data availability.

By providing a climate-informed approach tailored for data-limited regions, the GHDL model enhances flood risk assessment and supports the development of innovative insurance products. We hope that the GHDL model serves as a reference for inland flood pricing for insurers, offering a foundation for further development and refinement. This model is designed to supplement existing commercial models, providing additional insights and improving the overall robustness of flood risk management strategies.

The following table summarises the comparison of GHDL-based index insurance model with the two commercial models from various perspectives.

Table 2: Comparison with Commercial Flood Risk Models

Feature	RMS US Inland Flood Model	SCOR Oasis LMF Flood Model	GHDL-based Index Insurance Model
Objective	Develop a probabilistic flood risk assessment model and support insurance pricing.	Develop a modular, scalable framework for flood risk prediction and insurance pricing.	Develop a flexible flood risk prediction model under the constraints of data limitation to support insurance pricing with actuarial meaning.
Key Flood Types	Fluvial, pluvial.	Fluvial, pluvial, groundwater flooding, coastal flooding.	Fluvial, pluvial.
Core Methodology	Combines rainfall-runoff models, routing models, and inundation modelling with probabilistic event generation.	Modular framework integrating stochastic models, physical models, and scenario-based assessments.	Geo-Hierarchical Deep Learning (GHDL) combining CNNs and WNNs within a spatially hierarchical structure.
Strengths	High-resolution mapping.	Modular, adaptable framework for different risk and hazard types.	Robust handling of spatial dependencies through hierarchical modelling.
	Extensive calibration and validation with long historical data series.	Integration of stochastic and scenario-based approaches.	Effective even with limited data (Section 2.1).
Data	PCA simulated Precipitation, Tropical cyclone rain.	Historical hydrological data.	Historical hydrological data, precipitation data.
Flood Hazard Metrics	Peak discharge, water level, and flood extent.	Water level and discharge.	Water level based aggregated flood hazard index derived from NHDC data.

Hydrologic Models	Semi-distributed rainfall-runoff modelling.	Statistical and physical models capturing river flow and flood extent.	Omitted in future projections.
Hydraulic Models	2D shallow water equations for high-resolution flood mapping (10-30m grids).	1D (river flow) and 2D (floodplain) models combined with stochastic scenarios.	Upstream/downstream interactions in GHDL.
Deep Learning/AI	None.	None.	GHDL integrates CNNs for local precipitation analysis and WNNs for fluvial flood dynamics.
Climate Projections	Not integrated.	Not integrated.	CMIP6 climate simulations under four SSP scenarios for future flood hazard projection.
Validation	Extensive validation against FEMA maps and gauge station data.	Limited validation due to scarce granular claims and exposure data.	Validated against observed flood hazards with performance compared to LCDL models.
Output Resolution	High (10-30m grids for inundation maps).	Variable depending on data resolution and model configuration.	Daily aggregated hazard indices at the regional level.

In summary, each model has unique strengths and limitations. The RMS Inland Flood Risk Model excels in high-resolution flood mapping and probabilistic accuracy but lacks integration with future climate scenarios. The SCOR Oasis LMF Flood Model offers flexibility and adaptability but is often constrained by data gaps and coarser resolution. In contrast, the GHDL Model provides a climate-informed approach tailored for data-limited regions, making it an effective tool for flood risk assessment and insurance applications in developing economies.

8 *Implications*

As anthropogenic climate change intensifies, flood risks that were once considered secondary have now emerged as primary threats, particularly in vulnerable regions across Asia where data relevant data is limited, like the Chao Phraya River Basin. These escalating risks necessitate proactive measures in flood risk management. This report presents several implications based on the results, with a particular focus on addressing challenges within the insurance industry. While the Chao Phraya River Basin serves as a case study, the model has been validated in broader contexts, including the Mississippi River Basin (Xu et al., 2022), making its findings relevant for other countries in Asia.

In this section, we conclude several implications based on the modelling result of this report when writing this report, such as including industry-academia collaboration, infrastructure investment, risk diversification, and coastal protection. These approaches can enhance predictive modelling, strengthen flood defences, and improve insurance resilience against climate-induced risks.

8.1 **More Collaboration**

Collaboration between industry and academia is essential for addressing the complexities of flood risk. Insurers can benefit from partnerships with academic institutions to develop advanced predictive models that integrate climate projections and flood hazard assessments.

A key aspect of this collaboration is data sharing—insurers can provide historical claims data, exposure information, and business insights, while researchers contribute expertise in climate modelling and risk assessment methodologies. Joint efforts can enable the creation of robust databases and improve access to historical loss and damage data, critical for refining insurance pricing and reserving strategies.

A programming package designed to replicate the modelling process described herein will also be provided to GAIP partners⁹. They can leverage this tool for in-house testing using their proprietary policy and claim datasets, offering a cost-effective way to explore flood risk modelling tailored to specific portfolios. By utilising this package, they can collaborate closely with researchers to fit the models to their specific needs and analyse the outputs. This collaborative approach not only enhances the practical application of the GHDL model but also ensures continuous improvement through shared insights and data.

While centralised research can save resources and expand our ability in handling the emerging risks, regular knowledge exchange through industry-academic workshops, joint publications, and technical reports can further bridge the gap between research and practical application, ensuring findings are translated into actionable risk management solutions.

By fostering collaboration, insurers and researchers can address data scarcity issues prevalent in developing regions, promoting a more informed and cohesive approach to managing future flood risks.

8.2 Risk Reduction

Strengthening Riverbank Infrastructure

The results of this study underscore the pivotal role of riverbank infrastructure in mitigating flood risks. Climate projections indicate that the Chao Phraya River Basin will experience an increase in both the frequency and severity of flood events as climate change intensifies. Overflow incidents are projected to become more frequent, posing significant challenges to flood risk management in the region. Strengthening riverbank systems is essential to maintaining the safety and sustainability of communities and economies in this vulnerable area.

For insurers, the resilience of riverbank infrastructure directly impacts the frequency and magnitude of flood-related claims. Enhancing riverbank defences, particularly in downstream areas like Bangkok where flood hazards are most

⁹ This package will be available on the website of Global-Asia Insurance Partnership (GAIP), accessible only to GAIP partners through their registered accounts.

acute, can substantially reduce financial exposure. Investments in flood control measures such as levees, embankments, and other protective structures can mitigate risks and improve loss ratios.

Enhancing Coastal Protection

Coastal regions, including areas near the mouth of the Chao Phraya River, face dual threats from riverine flooding and sea-level rise. The insurance industry has a vested interest in advocating for and investing in coastal protection measures, such as seawalls, mangrove restoration, and sustainable land-use practices. These initiatives can significantly reduce flood-related damages and, by extension, insurance payouts. Enhanced coastal protection also offers insurers opportunities to develop tailored products, such as coverage for coastal property owners who adopt adaptive measures, thereby aligning economic incentives with risk reduction goals.

To address these challenges, it is crucial for both public and private stakeholders to recognise the implications of flood risks and support efforts to enhance riverbank and coastal resilience. While insurers may not directly finance infrastructure projects, they can play a key role in advocating for and advising on risk reduction strategies, highlighting the impact of flood defences on insurance costs and long-term resilience. By engaging with policymakers, urban planners, and local communities, insurers can contribute valuable insights that help shape effective flood mitigation measures, ensuring a more sustainable approach to managing flood risks.

By supporting risk reduction initiatives, insurers can protect insured assets while also fostering long-term resilience. Beyond risk transfer, insurers can play a broader role in shaping climate resilience by advising on risk mitigation strategies, incentivising protective measures (e.g., through premium discounts), and engaging with policymakers to enhance flood preparedness. These efforts align with a holistic, integrated approach to closing protection gaps—a key theme in GAIP's work—which emphasises a three-pronged strategy: (1) risk reduction, (2) increasing insurance penetration, and (3) fiscal risk financing.

Furthermore, this discussion connects with GAIP's "Beyond Protection" framework, which highlights the diverse roles insurers can play in supporting climate action, including (1) insurer, (2) risk advisor, (3) educator, (4) influencer, (5)

investor, and (6) wealth manager. By embracing these roles, insurers can move beyond traditional coverage models and contribute to systemic risk reduction, ultimately strengthening resilience against climate-induced flooding.

9 *Conclusions*

In this report, we studied and presented a different approach to modelling flood risks and applied this approach to the Chao Praya River as a case study.

Leveraging GHDL as the deep learning framework, we proposed a methodology tailored to the resource and data constraints of developing regions. This approach enables insurers to create flood hazard forecasts with minimal resource demands and limited data availability, making it particularly suitable for areas with sparse observational records (Section 2.1). Using water level as a proxy for flood hazard, we established a robust relationship between precipitation and flood risk, with the model successfully capturing the majority of high flood hazards at the time of underwriting (Section 4.4).

Building on this foundation, we utilised climate simulations under four SSP scenarios to forecast future flood risks across various regions in Thailand. Our flood and drought hazard assessments highlight the disproportionate impacts of climate change across provinces. Regions closer to the Gulf of Thailand, such as Bangkok, Nonthaburi, and Pathum Thani, face heightened flood risks due to significant increases in precipitation, while other areas experience more complex patterns, with some regions showing stable or even improved risks under certain scenarios.

This study underscores the importance of integrating localised climate projections into hazard modelling and risk management strategies, particularly in regions vulnerable to the uneven impacts of climate change. The results of this case study provide critical insights for policymakers and insurers to design targeted adaptation measures and enhance resilience against future flood and drought risks in Southeast Asia (Section 8).

To make GHDL actionable for insurers, the next step is to test its applicability in real-world settings by integrating insurer-held exposure and claims data with the model's flood hazard predictions. We encourage GAIP partners to utilise the provided programming package for in-house testing, allowing them to explore flood risk modelling tailored to their specific portfolios. By collaborating closely with researchers, they can fit the models to their unique needs, analyse the outputs, and continuously improve the model through shared insights and

data. This collaborative approach will help validate the GHDL model's effectiveness in pricing, underwriting, and risk assessment, ultimately enhancing flood risk management strategies.

Beyond flood risk, GHDL's deep learning approach can serve as a foundation for developing similar frameworks to model other climate-related risks, such as typhoons, coastal flooding, or heatwave-related mortality. Insurers can explore how AI-driven models enhance catastrophe risk assessment, parametric insurance solutions, and loss reserving strategies. Engaging in joint research and proof-of-concept applications will help insurers shape the next generation of data-driven risk models tailored to their needs.

A Appendix

A.1 Discrete Wavelet Transform

The discrete wavelet transform (DWT) is a linear operation that decomposes a signal in $L^2(\mathbb{R})$ into a series of segments, facilitating a more efficient and accurate representation of signals. In the context of DWT, wavelets are defined as families of functions h_{ab} ,

$$h_{ab}(x) = |a|^{-1/2} h\left(\frac{t-b}{a}\right), \quad a, b \in \mathbb{R}, a \neq 0,$$

generated from a single base function h through the processes of dilation and translation (Daubechies, 1988). This method allows for the analysis of various signal components at different scales, making DWT a powerful tool in signal processing.

Select a function ϕ from $L^2(\mathbb{R})$ such that its family of translations, denoted as $\{\phi(x-k), k \in \mathbb{Z}\}$, forms an orthonormal set. The wavelet family associated with ϕ is defined as follows:

$$\phi_{jk} = 2^{j/2} \phi(2^j t - k), \quad j, k \in \mathbb{Z},$$

It is evident that ϕ_{jk} can be reformulated to match the format of function h_{ab} , by setting $a = 2^{-j}$ and $b = 2^j k$. Utilising ϕ_{jk} as base functions, we can construct a series of linear spaces $\{V_j \subset L^2(\mathbb{R}), j \in \mathbb{Z}\}$, where

$$V_j = \left\{ \sum_k A_k \phi_{jk}(x) : \sum_k |A_k|^2 < \infty \right\}, \quad j \in \mathbb{Z}.$$

Assuming that ϕ is selected such that the spaces are nested,

$$V_j \subset V_{j+1}, \quad j \in \mathbb{Z},$$

and that

$$\bigcup_{j \in \mathbb{Z}} V_j \text{ is dense in } L^2(\mathbb{R}),$$

we then define the complementary set of V_j relative to V_{j+1} , creating another sequence of linear spaces:

$$W_j = V_{j+1} - V_j.$$

There exist a function ψ , whose family of translations is orthonormal. The wavelet space of ψ , expressed as

$$\psi_{jk} = 2^{j/2} \psi(2^j t - k), \quad j, k \in \mathbb{Z},$$

linearly spans the sequence of spaces $\{W_j \subset L^2(\mathbb{R}), j \in \mathbb{Z}\}$, where

$$W_j = \left\{ \sum_k D_k \psi_{jk}(x) : \sum_k |D_k|^2 < \infty \right\}, \quad j \in \mathbb{Z}.$$

If a function ϕ is chosen in accordance with the nested and dense condition, any function $g \in L^2(\mathbb{R})$ can be uniquely represented as a convergent series within $L^2(\mathbb{R})$, with respect to ϕ and ψ (see Härdle et al., 1998, for details of proofs and the methodology for deriving ψ once ϕ is selecte.):

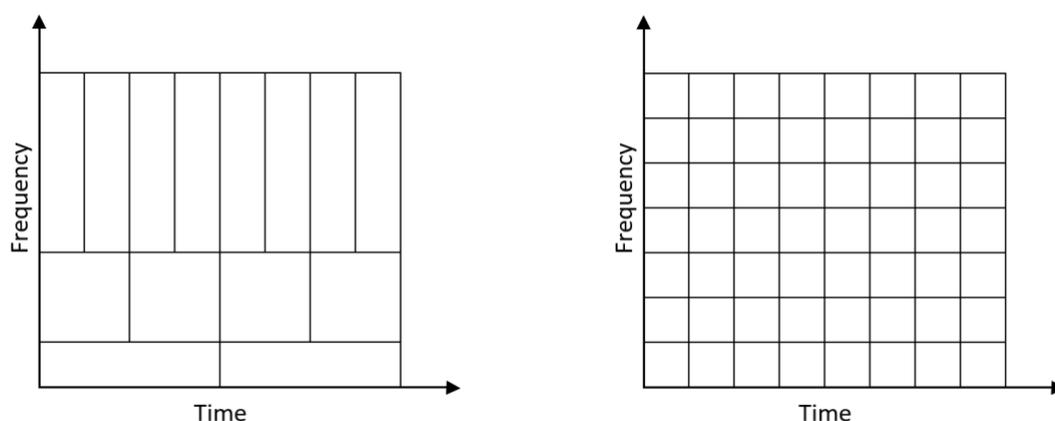
$$g(x) = \sum_k A_k \phi_{0k}(x) + \sum_{j=0}^{\infty} \sum_k D_{jk} \cdot \psi_{jk}(x).$$

This relation is known as the discrete wavelet transform (DWT) of g . The selected function ϕ is called scale wavelet, and the derivation of ψ is called mother wavelet.

The multiresolution expansion exhibits the property of localisation in both time and frequency domains. The summation over k corresponds to localisation in time (shifts of functions $\phi_{j0}(x)$ and $\psi_{jk}(x)$). Conversely, summation over j corresponds to localisation in frequency. As j increases, the associated frequency of $\psi_{jk}(x)$ becomes higher, allowing for the analysis of finer details in the frequency domain.

Given this nature of representation, the coefficients D_{jk} are referred to as “detail coefficients.” These coefficients capture the high-frequency information at each level of the decomposition, pinpointing the more nuanced aspects of the signal. Conversely, the coefficients A_k , known as “approximation coefficients,” represent the remaining signal information that is not captured by the detail coefficients. They essentially provide a smoothed or averaged version of the signal, reflecting its broader trends.

The time and frequency localisation properties of the DWT endow it with several desired characteristics, making it a powerful tool for analysing river flow data in flood risk management. First, river flow data is typically non-stationary, influenced by systemic factors such as climate change and seasonality, as well as occasional events like heavy rainfall. The ability of DWT to localise both time and frequency allows for a more precise analysis of transient features and anomalies in the signal (Daubechies, 1988). Second, DWT facilitates multi-resolution analysis of river flow data. This enables the decomposition of data into different scales, effectively capturing long-term trends, such as seasonal variations, and short-term events, such as flash floods, within a unified framework. Third, river flow data can often be noisy, affected by factors like sensor inaccuracies or environmental interference. DWTs are particularly useful in denoising this data, thereby enhancing the quality of information crucial for making informed predictions and decisions. Owing to these capabilities, DWTs have been extensively applied in flood risk management to transform and analyse various data types (Kumar et al., 2015, among others; Seo et al., 2015; see Shafaei and Kisi, 2016).



(a) *Discrete wavelet transform*

(b) *Fast Fourier transform*

Figure A1: Coverage of the time-frequency plane of the wavelet and Fourier based transforms

In practical applications, when dealing with a discrete signal $g[n]$ and a scaling function $\phi: \mathbb{R} \rightarrow \mathbb{R}$, the wavelet coefficients A_k and D_{jk} are computed by progressively projecting the signal onto the scaling and mother wavelet functions. This projection starts from the highest selected frequency and proceeds to the lowest selected frequency (see Figure A2 for an example):

$$A_k = \sum_n g[n] \phi_{0k}(n),$$

$$D_{jk} = \sum_n g[n] \psi_{jk}(n).$$

Notably, filtering a low-frequency signal using a higher-frequency filter does not result in the loss of information. Consequently, the original signal $g[n]$ can always be reconstructed accurately using the acquired wavelet coefficients.

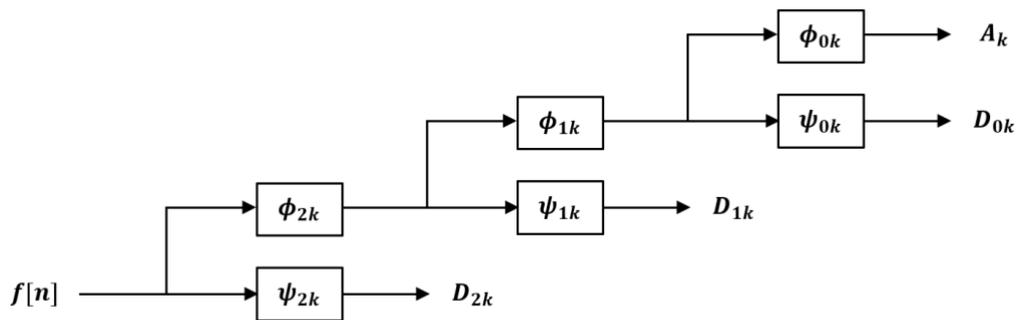


Figure A2: A 3-level discrete wavelet transform system

10 Reference

Bevere, L., and Dhore, K. (2021). *The World's Costliest Flood: The 2011 Thailand Flood, 10 Years On* (E1-21-2021). Swiss Re Institute.

https://www.nesdc.go.th/nesdb_en/more_news.php?cid=156&filename=index

Bouwer, L. M. (2013). Projections of future extreme weather losses under changes in climate and exposure. *Risk Analysis*, 33(5), 915–930.

Budhathoki, A., Tanaka, T., and Tachikawa, Y. (2024). Developing flood risk curves of agricultural economic damage under climate change in the Lower Chao Phraya River Basin, Thailand. *Journal of Flood Risk Management*, e13031.

Chen, A. S., Djordjević, S., Leandro, J., and Savić, D. A. (2010). An analysis of the combined consequences of pluvial and fluvial flooding. *Water Science and Technology*, 62(7), 1491–1498.

<https://doi.org/10.2166/wst.2010.486>

Daubechies, I. (1988). Orthonormal bases of compactly supported wavelets. *Communications on Pure and Applied Mathematics*, 41(7), 909–996. <https://doi.org/10.1002/cpa.3160410705>

Gale, E. L., and Saunders, M. A. (2013). The 2011 thailand flood: Climate causes and return periods. *Weather*, 68(9), 233–237.

Geophysical Fluid Dynamics Laboratory. (2024). *Climate modelling*.

<https://www.gfdl.noaa.gov/climate-modeling/>

Härdle, W., Kerkycharian, G., Picard, D., and Tsybakov, A. (1998). *Wavelets, approximation, and statistical applications* (Vol. 129). Springer.

He, K., Zhang, X., Ren, S., and Sun, J. (2016). Deep residual learning for image recognition.

Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 770–778.

Hersbach, H., Bell, B., Berrisford, P., Hirahara, S., Horányi, A., Muñoz-Sabater, J., Nicolas, J., Peubey, C., Radu, R., Schepers, D., et al. (2020). The ERA5 global reanalysis. *Quarterly Journal of the Royal Meteorological Society*, 146(730), 1999–2049.

Jankowsky, S., Hilberts, A., Mortgat, C., Li, S., Xu, N., Mei, Y., ... & Yang, Y. (2016). The RMS US inland flood model. In *E3S Web of Conferences* (Vol. 7, p. 04014). *EDP Sciences*.

Jongman, B., Hochrainer-Stigler, S., Feyen, L., Aerts, J. C., Mechler, R., Botzen, W. W., Bouwer, L. M., Pflug, G., Rojas, R., and Ward, P. J. (2014). Increasing stress on disaster-risk finance due to large floods. *Nature Climate Change*, 4(4), 264–268.

Kousky, C., and Kunreuther, H. (2018). Risk management roles of the public and private sector. *Risk Management and Insurance Review*, 21(1), 181–204.

- Kreibich, H., Van Loon, A. F., Schröter, K., Ward, P. J., Mazzoleni, M., Sairam, N., Abeshu, G. W., Agafonova, S., AghaKouchak, A., Aksoy, H., et al. (2022). The challenge of unprecedented floods and droughts in risk management. *Nature*, 608(7921), 80–86.
- Krizhevsky, A., Sutskever, I., and Hinton, G. E. (2012). Imagenet classification with deep convolutional neural networks. *Advances in Neural Information Processing Systems*, 1097–1105.
- Kumar, S., Tiwari, M. K., Chatterjee, C., and Mishra, A. (2015). Reservoir inflow forecasting using ensemble models based on neural networks, wavelet analysis and Bootstrap Method. *Water Resources Management*, 29(13), 4863–4883. <https://doi.org/10.1007/s11269-015-1095-7>
- LeCun, Y., Bottou, L., Bengio, Y., and Haffner, P. (1998). Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, 86(11), 2278–2324.
- Li, Z., Liu, F., Yang, W., Peng, S., and Zhou, J. (2022). A survey of Convolutional Neural Networks: Analysis, applications, and prospects. *IEEE Transactions on Neural Networks and Learning Systems*, 33(12), 6999–7019. <https://doi.org/10.1109/tnnls.2021.3084827>
- Mosavi, A., Ozturk, P., and Chau, K.-W. (2018). Flood prediction using machine learning models: Literature review. *Water*, 10(11), 1536. <https://doi.org/10.3390/w10111536>
- Nkwunonwo, U. C., Whitworth, M., & Baily, B. (2020). A review of the current status of flood modelling for urban flood risk management in the developing countries. *Scientific African*, 7, e00269.
- Poapongsakorn, N., & Meethom, P. (2013). Impact of the 2011 floods, and flood management in Thailand. *ERIA Discussion Paper Series*, 34, 2013.
- Quinn, N., Bates, P. D., Neal, J., Smith, A., Wing, O., Sampson, C., Smith, J., and Heffernan, J. (2019). The spatial dependence of flood hazard and risk in the United States. *Water Resources Research*, 55(3), 1890–1911. <https://doi.org/10.1029/2018wr024205>
- Redmon, J., Divvala, S., Girshick, R., and Farhadi, A. (2016). You only look once: Unified, real-time object detection. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 779–788.
- Rimkus, S., Patnaukuni, S., Bovy, H., Souch, C., Nunn, P. (2016). Guide to flood modeling. In *Oasis LMF*. https://oasislmf.org/application/files/4317/1624/0848/Hazards_-_Flood.pdf
- Ronneberger, O., Fischer, P., and Brox, T. (2015). U-net: Convolutional networks for biomedical image segmentation. *International Conference on Medical Image Computing and Computer-Assisted Intervention*, 234–241.
- Seo, Y., Kim, S., Kisi, O., and Singh, V. P. (2015). Daily water level forecasting using wavelet decomposition and artificial intelligence techniques. *Journal of Hydrology*, 520, 224–243. <https://doi.org/10.1016/j.jhydrol.2014.11.050>

- Shafaei, M., and Kisi, O. (2016). Predicting river daily flow using wavelet-artificial neural networks based on regression analyses in comparison with artificial neural networks and support vector machine models. *Neural Computing and Applications*, 28(S1), 15–28. <https://doi.org/10.1007/s00521-016-2293-9>
- Simonyan, K., and Zisserman, A. (2014). Very deep convolutional networks for large-scale image recognition. *arXiv Preprint arXiv:1409.1556*.
- Surminski, S., Aerts, J. C., Botzen, W. J., Hudson, P., Mysiak, J., and Pérez-Blanco, C. D. (2015). Reflections on the current debate on how to link flood insurance and disaster risk reduction in the European Union. *Natural Hazards*, 79, 1451–1479.
- Sufian, Abu (2024), *Hydrological Dataset for Ganges Basin*, Mendeley Data, V1, doi: 10.17632/sx4m6tdwb7.1
- Swiss Re. (2022). *Sigma report natural catastrophes in 2021: The floodgates are open*. <https://www.swissre.com/institute/research/sigma-research/sigma-2022-01.html>.
- Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., Erhan, D., Vanhoucke, V., and Rabinovich, A. (2015). Going deeper with convolutions. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 1–9.
- Thistlethwaite, J., Henstra, D., Brown, C., and Scott, D. (2020). Barriers to insurance as a flood risk management tool: Evidence from a survey of property owners. *International Journal of Disaster Risk Science*, 11(3), 263–273. <https://doi.org/10.1007/s13753-020-00272-z>
- Visessri, S., and Ekkawatpanit, C. (2020). Flood management in the context of climate and land-use changes and adaptation within the Chao Phraya River basin. *Journal of Disaster Research*, 15(5), 579–587.
- Xu, Y., Tan, K. S., and Zhu, W. (2022). From Meteorology to Market: A Geo-Hierarchical Deep Learning Approach for Flood Risk Pricing. *Nanyang Business School Research Paper No. 24-04*. <https://ssrn.com/abstract=4692475>