Doctoral Dissertation

Multi-Source Data Integration for Urban Disaster Estimation and Assessment Based on Remote Sensing and Social Sensing

YANG Yifan

Supervisor: GOKON Hideomi

Division of Advanced Science and Technology
Japan Advanced Institute of Science and Technology
(Knowledge Science)

March 2026

Abstract

Urban flood disasters pose increasing risks to human life, infrastructure, and socio-economic systems, particularly in densely populated cities where physical hazard processes and human behavioral responses interact in complex ways. From a theoretical perspective, this dissertation adopts a knowledge-oriented integration perspective that emphasizes the combined use of physical sensing data and human behavioral information for urban flood extent estimation. This perspective is grounded in Knowledge Science, emphasizing the transformation of diverse data sources into interpretable and actionable knowledge for disaster management.

To operationalize this concept, this dissertation proposes a comprehensive multi-source data integration framework that combines remote sensing, geographic information, and social sensing for urban flood estimation and assessment. The framework is applied to the 2019 Typhoon Hagibis flood in Nagano City, Japan, serving as a representative case of an urban riverine disaster.

The research is structured around three incremental experiments. Experiment I examines temporal and spatial anomalies in large-scale mobile phone GPS data, demonstrating that population mobility patterns reflect flood-related disruption and public response. Experiment II develops Random Forest–based flood estimation models that integrate Synthetic Aperture Radar (SAR), Digital Elevation Model (DEM), terrain indices (TWI, SPI), and GPS-derived population density, showing that the inclusion of behavioral information improves spatial coherence and reduces misclassification. Experiment III advances the framework to a fine-gridded (70 m × 70 m) urban model by incorporating Twitter-derived information, enabling near–real-time situational awareness through the integration of physical indicators and human-reported flood impacts.

Across all experiments, the results confirm that multi-source data fusion enhances the timeliness, spatial precision, and interpretability of urban flood estimation. Physical sensing data provides stable representations of inundation and terrain conditions, while social sensing data capture dynamic population displacement and disaster perception, which together reduce false positives and improve the spatial coherence of flood extent estimation across all experiments. Their integration establishes a data-driven foundation for resilient city assessment. This dissertation contributes a unified methodological

workflow linking sensing, data fusion, and application stages, demonstrating how theory-guided multi-source integration can support both real-time disaster response and long-term urban resilience planning.

Keywords:

Remote sensing; Social sensing; multi-source data integration; Urban flood estimation; Random Forest; SAR; GPS mobility data; Twitter data; Machine learning; Urban resilience; Disaster assessment

Acknowledgement

I would first like to express my deepest gratitude to Professor Gokon, to whom I owe the completion of this dissertation. I feel extremely fortunate to have been his student. His continuous guidance, support, and encouragement shaped every stage of my research. I am especially grateful for his respect for my ideas, which allowed me to pursue the topics I truly wished to explore. Even during difficult periods of my doctoral study, he never gave up on me and always found ways to motivate and guide me. This dissertation would not have been possible without his patience, insight, and unwavering support.

I would also like to extend my sincere appreciation to Professor Takeuchi. It has been an invaluable experience to study in his laboratory, where I learned not only technical knowledge but also how to think deeply and approach research with curiosity and perseverance. The warm and collaborative atmosphere of the laboratory, as well as the support from all its members, broadened my academic perspective and greatly enriched my doctoral journey. I am deeply thankful for this opportunity, which strengthened both my research abilities and my confidence in overcoming challenges.

I would also like to express my sincere gratitude to the members of the thesis examination committee for their careful review of this dissertation and for providing valuable comments that greatly improved the quality of this work.

My heartfelt thanks also go to my family and friends, who have supported me throughout this journey. Their kindness and encouragement sustained me through many difficult moments. I sincerely hope that I have lived up to their expectations and support.

I also acknowledge the organizations that provided essential data for this research, including Agoop Corporation, MLIT, JAXA, ESA, and related institutions. Their support enabled the multi-source analysis conducted in this dissertation.

Finally, I wish to express my sincere gratitude to the foundations and institutions that generously supported my research and academic activities: Shibuya Science Culture and Sports Foundation, The Telecommunications Advancement Foundation, JAIST Foundation Research Grant, Tateisi Science and Technology Foundation, GRSS Travel Support, and Rotary Yoneyama Memorial Foundation. Their financial support enabled field investigations, conference participation, and the completion of this dissertation.

To all who have supported me—directly or indirectly—thank you.

Contents

List of Figures

List of Tables

Chapter 1 Introduction

1.1 Background

In recent decades, the frequency and intensity of natural disasters have increased significantly due to climate change, rapid urbanization, and environmental degradation[1], [2], [3]. Among these, floods have emerged as one of the most destructive and recurrent hazards, threatening human life, infrastructure, and urban sustainability[4], [5], [6]. Globally, urban areas are becoming increasingly vulnerable because of their high population density, aging infrastructure, and exposure to extreme weather events[7], [8].

Japan, located in the East Asian monsoon zone and along major tectonic plate boundaries, faces particularly severe disaster risks from typhoons, floods, and earthquakes[9], [10]. The 2019 Typhoon Hagibis, for example, caused catastrophic flooding in Nagano City and other regions, highlighting the limitations of existing urban disaster management systems[11]. Although Japan possesses one of the world's most advanced disaster monitoring and early warning infrastructures, the unprecedented scale and dynamics of recent disasters underscore a critical challenge: how to integrate real-time, multi-dimensional data for more adaptive and knowledge-driven disaster estimation [12], [13], [14], [15].

With the advancement of geospatial technologies, data derived from remote sensing (e.g., Synthetic Aperture Radar, optical imagery) and social sensing (e.g., GPS mobility, social media) provide new opportunities for capturing both the physical processes and human dimensions of disasters[16], [17], [18]. Remote sensing enables large-scale, all-weather observation of surface changes, while social sensing reflects dynamic human behavior and perceptions during emergencies[18], [19]. However, integrating these heterogeneous data sources into a coherent analytical framework remains a significant challenge[20], [21]. Most existing disaster prediction and estimation models rely on single-source or static datasets, lacking the ability to incorporate real-time human behavior and contextual knowledge[22], [23].

Against this backdrop, this dissertation proposes a multi-source data integration framework for urban disaster estimation and assessment, combining remote sensing, social sensing, and GIS-based environmental data. By doing so, it aims to bridge the gap between physical hazard monitoring and social response analysis, contributing to the development of knowledge-based urban resilience.

1.2 Problem Statement

Traditional flood modeling approaches have long relied on hydrological parameters, static topographic information, or single-source remote sensing observations[24], [25]. Although these methods have contributed to understanding flood formation and extent,

they struggle to capture the rapidly changing characteristics of urban disasters[26], [27]. Hydrological models often depend on predefined assumptions that cannot fully represent real-time flood progression, and satellite-based observations are restricted by acquisition schedules and weather conditions[28], [29]. More importantly, these physical sensing methods provide limited insight into how urban residents behave during disasters[30]. In recent years, the growing availability of social sensing information such as mobile-phone GPS mobility patterns and Twitter messages has introduced new possibilities for observing human responses during extreme events[31], [32], [33]. However, integrating this behavioral information with physical sensing remains challenging because of the large differences in data structure, spatial resolution, and temporal frequency[34].

A significant research gap emerges from the limited degree of multi-source data integration in existing studies[35], [36]. Many flood estimation models rely exclusively on physical data such as SAR backscatter values, DEM-derived elevation and slope, or land use characteristics[37], [38]. Other studies focus primarily on human-centered indicators, including mobility anomalies extracted from GPS datasets or social media content reflecting public awareness[33], [34]. These separate approaches result in models that capture only one side of disaster dynamics. Without combining environmental processes with behavioral adjustments, estimation models cannot fully represent urban-scale flood impacts[39]. As a result, current flood assessments often overlook how physical flooding interacts with human activities, leading to incomplete situational awareness.

Another limitation concerns the insufficient spatiotemporal granularity of conventional models[24]. Many existing frameworks operate at aggregated levels such as municipal districts or large grid cells[40]. This level of resolution cannot adequately represent the distributed nature of flooding within dense urban environments, were minor differences in elevation, infrastructure, or land cover can significantly influence water movement[24], [40]. Similarly, the lack of fine-grained temporal data prevents these models from capturing the evolving nature of human responses during disasters[41], [42]. Improving flood estimation accuracy therefore requires integrating data from diverse sensing modalities into unified and finer spatial units.

Beyond technical issues of data sources and spatial scale, there is also a lack of knowledge-based interpretation within current flood modeling practices[24]. Many studies focus primarily on improving classification accuracy or algorithmic performance but pay limited attention to how different data sources contribute to knowledge creation about urban hazards[43]. From the perspective of Knowledge Science, disaster estimation is an integrated process that links environmental sensing, human behavior, decision-making, and learning[41], [44]. When assessment models overlook this conceptual dimension, their usefulness for long-term resilience planning becomes constrained[44]. There is a need for theoretical frameworks that translate multi-source data into actionable knowledge and support continuous learning at the city level.

These research gaps collectively highlight the necessity of developing a unified

analytical framework that integrates physical sensing and social sensing data while maintaining fine spatiotemporal resolution and contributing to knowledge formation. This dissertation addresses these challenges by combining remote sensing information, GPS mobility patterns, social media data, and machine learning methods in order to achieve a deeper and more comprehensive understanding of urban flood processes. Through this approach, the study aims to advance multi-source data integration techniques and support the development of resilient urban systems.

1.3 Research Questions and Hypotheses

Building upon the background and research gaps identified in Sections 1.1 and 1.2, this dissertation formulates several core research questions aimed at understanding how physical sensing and social sensing can be integrated to improve urban flood estimation. These questions guide the methodological design of Experiments I–III and provide a conceptual foundation for the development of a multi-source disaster assessment framework.

When a large-scale natural disaster occurs, the destruction of information and communication systems and transportation infrastructure gives rise to a period known as the “information blackout,” during which it becomes unclear where and what kinds of damage have occurred. Disaster response immediately after an event must be conducted rapidly and efficiently; however, this information blackout can cause significant delays in such responses. To address this issue, it is essential to promptly grasp the overall extent of damage across the vast affected area.

For this purpose, satellite remote sensing technology, which enables the acquisition of wide-area damage information from remote locations, is highly effective. For example, in the case of flood disasters, various remote sensing techniques—such as optical satellite imagery and synthetic aperture radar (SAR)—have been utilized for flood detection.

These technologies are extremely effective if satellite images of the target area can be acquired immediately after a disaster. However, wide-area damage assessment based on satellite remote sensing faces limitations in terms of observation frequency and imaging coverage, and it is not always possible to observe the entire affected area immediately after a disaster. Therefore, it is also necessary to develop technologies that enable disaster situation assessment from remote locations by utilizing spatial information processing techniques other than satellite remote sensing.

In addition, satellite remote sensing technology may include misclassifications in its analysis results due to limitations in the spatial resolution of satellite imagery and the complexity of the relationship between the electromagnetic waves observed by satellites and physical changes on the Earth’s surface. As a result, it is currently difficult to grasp the wide-area disaster situation with 100% accuracy using satellite imagery alone. To overcome this challenge, it is necessary to construct new damage extraction methods that combine satellite remote sensing with other types of spatial information.

Against this background, this study focuses not only on physical changes on the Earth’s

surface that can be inferred from satellite imagery, but also on human behavior in disaster-affected areas. The greater the severity of damage, the more significantly human movement patterns in the affected area are expected to change, and such changes in human mobility are likely to have some relationship with flood-inundated areas. Furthermore, this study also focuses on information disseminated through social networking services (SNS), in addition to physical human movement. It is hypothesized that some correlation exists between the severity of damage and the textual information posted on SNS. Based on this idea, this dissertation proposes that the accuracy of flood area estimation can be improved by constructing a new analytical framework that combines satellite image analysis with human mobility data estimated from smartphone GPS data and SNS data from affected areas. Applying changes in human society exposed to disasters—not only direct physical observation data—to physical flood detection represents a novel academic approach.

Accordingly, this study aims to propose, develop, and validate a new flood estimation method that integrates satellite imagery, human mobility data, and SNS data, using the 2019 Typhoon Hagibis flood in Nagano City, Japan, as a case study. In particular, satellite imagery, human mobility data, and SNS data differ in temporal resolution, spatial resolution, and data format, and exploring effective methods to integrate these multidimensional data sources constitutes a key challenge of this research.

Based on this conceptual foundation, the following research questions and hypotheses are formulated:

RQ1: How do population mobility patterns captured through GPS data change before, during, and after a major flood event, and to what extent do these patterns reflect the temporal and spatial dynamics of disaster impacts?

RQ2: Does integrating social sensing information with physical sensing data improve the accuracy and spatial coherence of urban flood estimation models?

RQ3: Can real-time communication signals extracted from georeferenced Twitter posts provide additional value for fine-resolution urban flood assessment when combined with GPS and remote sensing data?

RQ4: How can multi-source data integration contribute to knowledge creation in disaster estimation, as viewed through the lens of Knowledge Science, and what does this imply for the development of resilient urban systems?

Based on the above research questions, the following hypotheses are proposed:

H1: GPS-derived human mobility anomalies exhibit significant temporal and spatial deviations during flood events and can serve as reliable behavioral indicators for detecting disaster-affected areas.

H2: Integrating social sensing with physical sensing significantly improves machine-learning–based flood estimation compared with using physical data alone.

H3: The incorporation of Twitter-derived disaster-related information enhances fine-grid flood estimation by providing real-time situational cues that reduce false positives and refine spatial boundaries.

H4: Multi-source data fusion supports a knowledge-creation cycle (sensing → integration → application), enabling a more comprehensive understanding of disaster dynamics and contributing to data-driven urban resilience.

1.4 Objectives and Case Studies

The primary objective of this dissertation is to develop a comprehensive analytical framework for urban flood estimation and assessment through the integration of multi-source data derived from both physical and social sensing. Building upon the research gaps identified above, the study aims to create a unified approach capable of capturing the physical mechanisms of flooding along with the behavioral responses of urban residents. By combining remote sensing information, geospatial indicators, GPS-based population mobility, and social media communication, the research seeks to enhance the spatial resolution, real-time applicability, and interpretative value of flood estimation models.

The first research objective is to investigate how population mobility patterns derived from GPS data reflect flood-related behavioral anomalies. This objective focuses on understanding whether temporal and spatial fluctuations in mobility can serve as indicators of disaster impact and how such signals can be incorporated into urban-scale hazard assessment. The corresponding contribution lies in demonstrating that GPS data can be used as a behavioral sensor, allowing representation of human responses that are not observable through traditional physical data sources.

The second research objective is to develop and evaluate a multi-source flood estimation model that integrates remote sensing observations, DEM-derived terrain attributes, and GPS-based social sensing information. This part of the research examines how the combination of physical and behavioral indicators improves the identification of inundated areas. The contribution of this component lies in showing that social sensing information enhances the model's ability to estimate flood extent, leading to more accurate and spatially coherent inundation mapping.

The third research objective is to extend multi-source integration to a finer spatial scale by introducing social media information as an additional source of real-time disaster awareness. By constructing a high-resolution grid model that incorporates Twitter content, GPS mobility patterns, and remote sensing features, the study seeks to capture urban flood dynamics with greater precision. The corresponding contribution is the development of an integrated fine-grid model capable of representing subtle spatial variations and offering near-real-time insights into disaster conditions.

Another important objective is to examine how multi-source data integration contributes to knowledge creation within the context of Knowledge Science. By analyzing how physical information and human-centered data interact within estimation models, the research identifies new pathways for generating urban resilience knowledge. The theoretical contribution is the formulation of a knowledge-oriented perspective in which disaster estimation becomes a process of continuous sensing, interpretation, and

learning supported by heterogeneous information sources.

Finally, the dissertation contributes methodological advances through the design of multi-stage data processing workflows, the implementation of Random Forest models using diverse data combinations, and the application of evidence-based interpretations for understanding model behavior. These contributions establish a practical and scalable framework for urban disaster assessment that can be adapted to various cities and hazard types.

Although the machine learning models technically generate predictions in their computational process, the scientific task of this dissertation is flood extent estimation. Accordingly, the model outputs are referred to as estimation results throughout this dissertation.

Overall, the research objectives and contributions collectively demonstrate how integrating multiple types of sensing information can transform urban flood estimation from a static analytical task into a dynamic and knowledge-rich process that supports resilient city development.

1.5 Structure of the Dissertation

This dissertation is organized into eight chapters, each addressing a specific component of the multi-source data integration framework developed for urban flood estimation and assessment. The structure follows progressive logic, moving from foundational concepts to methodological development, experimental implementation, theoretical discussion, and final conclusions.

Chapter 1 introduces the background, motivation, research gaps, research objectives, and overall structure of the dissertation. It establishes the need for an integrated approach that combines physical sensing and social sensing to improve the understanding of urban disaster dynamics.

Chapter 2 presents a synthesis of theoretical foundations and related research. It reviews the development of remote sensing techniques, social sensing data, machine learning approaches, and urban resilience research. This chapter provides the academic context that supports the methodological framework proposed in later chapters.

Chapter 3 describes the overall methodological framework of the dissertation. It outlines the multi-source data integration strategy, the rationale for selecting specific data types, and the general modeling workflow that guides Experiments I to III. The chapter emphasizes the conceptual transition from single-source analysis to a unified multi-dimensional system for flood estimation.

Chapter 4 reports the findings of Experiment I, which evaluates the role of GPS mobility data in detecting flood-induced behavioral anomalies. It introduces the study area, data preprocessing procedures, analytical methods, and experimental results. The chapter demonstrates that changes in population mobility can serve as effective indicators of disaster occurrence and severity.

Chapter 5 presents Experiment II, which integrates synthetic aperture radar (SAR),

terrain features, and GPS-based social sensing into Random Forest models for flood-extent estimation. The chapter includes detailed descriptions of two modeling scenarios and provides comparative analyses of model performance. The results show that the inclusion of social sensing data enhances the spatial accuracy and robustness of flood estimation.

Chapter 6 describes Experiment III, which further extends the integration to a fine spatial resolution by incorporating georeferenced Twitter data together with GPS and remote sensing features. The chapter introduces a high-resolution grid-based model and evaluates the effectiveness of multi-source real-time data in capturing urban flood dynamics. The results highlight the value of combining physical indicators with human communication patterns for real-time disaster assessment.

Chapter 7 provides a comprehensive discussion of the findings across all experiments. It examines the implications of multi-source data integration, the mechanisms of knowledge creation in disaster estimation, and the theoretical contributions to Knowledge Science. It also identifies methodological challenges and conceptual insights derived from the research.

Chapter 8 concludes the dissertation by summarizing the key findings and contributions of the study. It also presents the limitations of the current work and outlines directions for future research, including the potential for real-time systems, cross-city generalization, and integration with advanced sensing technologies.

Overall, the organization of the dissertation reflects a coherent progression from conceptual foundations to applied modeling and theoretical reflection. Each chapter builds upon the previous one to form a comprehensive understanding of how multi-source sensing data can support urban disaster estimation and resilience assessment.

1.6 Significance of the Research

This dissertation holds theoretical, methodological, and practical significance for the field of urban disaster estimation and resilience assessment[45]. At the theoretical level, it advances the understanding of how heterogeneous data sources—ranging from satellite-based physical sensing to human-centered social sensing—can be integrated to form a more comprehensive representation of disaster processes. Traditional flood assessment models have typically emphasized hydrological or physical variables, often overlooking the behavioral dimension of urban residents during disaster events[24], [25]. By incorporating GPS mobility traces and Twitter communication patterns, this research expands disaster analysis from a purely environmental perspective to a socio-environmental framework aligned with contemporary principles of urban resilience and Knowledge Science.

Methodologically, the study introduces a multi-layered data fusion approach that bridges physical sensing, social sensing, and machine learning. The three experimental designs progressively demonstrate the feasibility and advantages of integrating spatial, temporal, and behavioral indicators for flood estimation. The Random Forest models

developed in this work illustrate how different data modalities contribute to estimation performance and how their combination yields more accurate and spatially coherent flood-extent results. Furthermore, the use of fine-grid spatial modeling and real-time social media information highlights a methodological step toward higher-resolution, near-real-time disaster assessment systems. This methodological framework can be generalized to other types of urban disasters and extended for cross-city applications.

Practically, the findings provide meaningful support for disaster management and urban planning. The ability to detect population mobility anomalies and identify flood-affected areas with improved spatial accuracy can assist local governments in making timely evacuation decisions, allocating emergency resources, and evaluating the impacts of flood hazards. The integration of social sensing data offers a cost-effective and scalable alternative to traditional monitoring systems, particularly in urban areas where mobile phone penetration and social media activity are high. In addition, the multi-source modeling framework contributes to the development of smart, resilient cities by enabling more adaptive, data-informed decision-making processes.

Overall, this research demonstrates the transformative potential of multi-source data integration for enhancing the accuracy, timeliness, and societal relevance of urban disaster estimation. By combining physical indicators with human behavioral signals, the study provides new insights into the dynamics of urban flooding and contributes to the advancement of resilient city strategies.

Chapter 2 Literature Review and Theoretical Framework

2.1 Introduction

Understanding and estimating urban flood disasters requires a comprehensive review of the scientific foundations on which current modeling approaches are built. Flood events are inherently multi-dimensional phenomena shaped by hydrometeorological processes, topographic and land-surface conditions, infrastructure configurations and, critically, human behavior[46], [47]. As modern cities become more complex and densely populated, traditional flood assessment methods relying solely on physical environmental information are no longer sufficient to describe the dynamic interactions between natural hazards and human activity[25]. To address these challenges, recent research has increasingly incorporated diverse sensing technologies, including satellite remote sensing, digital elevation models, mobile phone GPS traces and social media platforms, which capture both the physical and social dimensions of disaster events[33], [37], [48].

This chapter provides the theoretical and empirical foundation for the multi-source integration framework developed in this dissertation. It systematically reviews existing research on remote sensing for flood monitoring, social sensing data for disaster analysis, and machine learning models used in geospatial estimation tasks. In doing so, the chapter identifies the major strengths and limitations of each data source, along with the methodological challenges associated with combining them within a unified modeling architecture. Particular attention is given to Synthetic Aperture Radar (SAR) data, which offer all-weather, high-resolution flood detection capabilities; GPS-based mobility data, which reflect real-time population responses during disasters; and Twitter data, which reveal public attention, perception, and situational information.

The chapter further situates this dissertation within the broader context of knowledge science and urban resilience research. By reviewing literature on multi-source data fusion, urban sensing ecosystems, and knowledge creation processes, it highlights how integrating physical and social sensing advances both analytical performance and theoretical understanding of how cities respond to extreme events. Ultimately, the goal of this chapter is to articulate the existing research landscape, identify the gaps that motivate the present study, and establish the conceptual framework supporting the multi-source flood estimation models developed in later chapters.

2.2 Urban Floods and Disaster Estimation

2.2.1. Characteristics of Urban Flooding

Urban flooding represents one of the most destructive and complex forms of natural

hazards, arising from the interaction of extreme meteorological events with highly modified land surfaces and dense socio-economic systems[49]. Rapid urbanization has significantly altered the hydrological cycle: impervious surfaces such as roads, rooftops, and parking areas reduce infiltration capacity and accelerate surface runoff, resulting in sharp increases in peak discharge during heavy rainfall events[8], [50]. These changes amplify flood hazards even under rainfall conditions that would have posed minimal risk in less urbanized landscapes[51].

In riverine cities, such as Nagano City along the Chikuma River, the characteristics of flooding are further shaped by the spatial configuration of the river system[52]. The Chikuma River flows through a wide alluvial plain surrounded by mountainous terrain, creating a natural convergence zone for rainfall-induced runoff[53]. Floodwaters spread rapidly across low-lying residential and industrial zones when the river overtops or breaches its embankments, as occurred during Typhoon Hagibis in 2019[11], [53]. The flat topography, narrow drainage pathways, and high exposure of built infrastructure contribute to extensive inundation and prolonged water retention[11]. These features typify many riverine urban environments in Japan, where the combination of monsoon climate and mountainous catchments generates sudden and severe flood events[54].

Urban floods also have distinct socio-spatial impacts due to high population density and concentrated economic activities[55]. Roads, commercial centers, and public service facilities attract large daily mobility flows; during a flood, disruptions to these urban functions can cause cascading effects such as transportation paralysis, delayed evacuation, or uneven exposure across neighborhoods[56]. Moreover, human behavior becomes an integral component of the hazard process: changes in mobility patterns reflect public awareness, risk perception, and adaptive actions[55], [57]. Understanding these dynamics is essential for estimating not only the physical extent of flooding but also the social consequences that shape the overall disaster impact.

2.2.2. Traditional Flood Estimation Approaches

Flood prediction and estimation has traditionally relied on hydrological and hydrodynamic modeling frameworks such as HEC-RAS, SWMM, and MIKE-FLOOD[58], [59], [60]. The aim of these models is to reproduce the physical mechanisms that govern the transformation of rainfall into runoff, river discharge and floodplain inundation[58], [59], [60]. Hydrological models simulate how precipitation is converted to runoff through soil infiltration and surface flow processes[61]. Hydrodynamic models further calculate water movement by solving flow continuity and momentum equations, allowing for detailed representation of water depth, velocity, and inundation extent across a river basin[24], [28].

When supported by sufficient field observations, including river cross-sections, channel roughness, rainfall records, and soil characteristics, these models can provide robust and physically interpretable estimates of flood behavior[58], [59], [60], [61]. They are therefore widely used for long-term risk assessment, infrastructure design, and the

planning of flood-control measures.

Despite these strengths, traditional flood models face several constraints in urban disaster applications[24], [25]. First, they require extensive parameterization and calibration. Many of the input variables, such as roughness coefficients, infiltration rates and channel geometry, are time-consuming to obtain and can vary significantly across urban environments[62]. Second, the numerical solutions used in hydrodynamic simulations often involve high computational costs, making real-time estimation difficult when rapid decision-making is required[63], [64]. Third, these models primarily focus on physical processes and provide limited insight into the human dimensions of disasters, such as mobility changes, communication patterns, or spontaneous evacuations[61]. As a result, they may not fully reflect the actual evolution of hazard impacts within densely populated and highly dynamic urban areas.

These limitations have contributed to a growing interest in data-driven modeling techniques that incorporate earth observation data, machine learning methods, social sensing signals, and real-time information streams[65], [66]. The integration of these alternative data sources enables rapid analysis at finer spatial scales and offers a complementary pathway to enhance the timeliness, adaptability, and contextual awareness of modern flood estimation systems.

2.3 Remote Sensing and Flood Modeling

Remote sensing provides a critical foundation for flood monitoring, offering wide-area, repeatable, and objective measurements of the Earth's surface[67]. In recent years, the combination of Synthetic Aperture Radar (SAR), Digital Elevation Models (DEM), and terrain indices have significantly advanced the ability to detect inundation patterns in urban environments[37], [68]. These datasets serve as primary physical indicators in many flood estimation models, including the Random Forest framework developed later in this dissertation.

2.3.1. SAR for Flood Detection

Synthetic Aperture Radar (SAR) has become one of the most widely used data sources for flood monitoring because it can acquire imagery independent of cloud cover, lighting conditions, and time of day[37]. This capability is particularly important during flood events, as heavy rainfall and thick cloud layers often limit the usability of optical satellite imagery[68]. SAR systems transmit microwave signals towards the Earth's surface and then record the energy that is reflected to the sensor[69]. Water surfaces typically exhibit very low backscatter due to their smoothness, producing distinct dark pixels in SAR images that can be readily distinguished from land areas[69].

Sentinel-1, used extensively in this research, offers additional advantages including a short revisit period, free accessibility, and dual viewing geometry from ascending and descending orbits[70]. The difference between these orbits improves the detection of

flooded surfaces by providing complementary viewing angles[71]. In built-up areas, single-orbit observations may be affected by radar shadowing or layover from high-rise structures[72]. The combination of both orbits increases the likelihood of correctly identifying inundated regions, especially in complex urban terrain[71], [72].

Despite its strengths, SAR-based flood detection in cities remains challenging[70]. Buildings can cause geometric distortions such as shadow, foreshortening, and layover, which may mask flooded roads or low-lying structures[73]. Urban materials with high reflectivity can also produce strong backscatter signals that resemble water in certain imaging conditions[74]. Therefore, SAR alone may misclassify flooded and non-flooded surfaces, especially in densely built environments[73]. These challenges highlight the importance of integrating SAR with additional data sources such as DEM-derived terrain features and social sensing data to improve accuracy and reduce uncertainty.

2.3.2. DEM and Terrain Features

Digital Elevation Models (DEM) play a central role in flood assessment because elevation and slope strongly influence water accumulation and flow direction[38], [75]. Low-lying areas are naturally more prone to inundations, while steep slopes promote rapid runoff[75]. Accordingly, elevation and slope serve as fundamental explanatory variables in most machine-learning-based flood-extent estimation models, including those implemented in this dissertation.

Beyond basic topography, hydrological terrain indices such as the Topographic Wetness Index (TWI) and the Stream Power Index (SPI) provide valuable information about surface water accumulation potential and flow energy[76], [77]. TWI reflects the tendency of water to accumulate at a given location, incorporating both local slope and upstream contributing area[76]. SPI describes the erosive power of flowing water and can help identify pathways of concentrated flow and regions susceptible to flood-induced erosion[76]. When combined with DEM and SAR measurements, these indices help represent the physical processes that shape flood extent and contribute to more robust feature representation in machine learning frameworks[77].

In the experiments conducted in Chapters 5 and 6, DEM-derived indices were used to quantify terrain influence at both coarse and fine spatial resolutions. Their integration with SAR backscatter and social sensing variables strengthens the model's capacity to distinguish between flooded and non-flooded surfaces in heterogeneous urban environments.

2.3.3. Limitations of Remote Sensing-Based Methods

Although remote sensing provides indispensable physical information, it has several limitations that restrict its ability to independently characterize flood events[19], [67]. Satellite observations primarily capture the physical environment and cannot represent population mobility, evacuation patterns, or other social behaviors that influence the real

impacts of disasters[78], [79]. As demonstrated in previous studies and validated in this dissertation, physical inundation patterns do not always align with human responses, especially in highly urbanized settings[79].

Another limitation relates to the temporal resolution of satellite missions. Even though Sentinel-1 offers frequent revisits, the available observations may not coincide with the peak of a rapidly evolving flood event[80], [81]. This temporal mismatch limits the capacity of remote sensing to support real-time emergency response. Furthermore, SAR images may suffer from noise, geometric distortions, and speckles, particularly in-built environments where complex structures interact with radar waves[82]. These issues create ambiguity in distinguishing inundated areas from dry surfaces, especially without supplementary data[82].

For these reasons, remote sensing-based methods benefit greatly from integration with social sensing data, which provide near-real-time insights into human activity, communication patterns, and behavioral responses during disasters. This dissertation builds on this idea by combining SAR, DEM, terrain indices, GPS mobility data, and Twitter messages to develop a more comprehensive and human-centered flood estimation framework.

2.4 Social Sensing for Disaster Research

Social sensing refers to the use of human-generated digital traces—such as mobile phone GPS logs, social media posts, and communication data—to understand collective behaviors and environmental changes[83]. During disasters, these data streams provide real-time insights into how populations respond, move, perceive risk, and share information[41]. Social sensing thus complements physical sensing by adding a behavioral dimension that cannot be captured by satellites or traditional hydrological measurements alone[41], [84].

2.4.1. GPS Mobility Data

Mobile phone GPS data have emerged as a powerful indicator of population dynamics during disasters[85], [86]. This data is generated when mobile devices communicate with nearby cellular base stations or when GPS-enabled applications send location updates[86]. Because mobile phones are widely carried in daily life, GPS traces provide continuous and large-scale representations of human movement across urban spaces[87], [88].

During flooding events, mobility patterns change dramatically[89]. The number of GPS records in an area can decrease due to evacuation, movement restrictions, or infrastructure disruption, whereas sudden increases may indicate active evacuation flows or emergency activities[90]. Such anomalies can reveal the timing, scale, and severity of disaster impacts more rapidly than official reports or satellite images. Recent studies have used GPS data to detect behavioral changes during earthquakes, typhoons, floods, and epidemic lockdowns, demonstrating its value for monitoring both physical disruptions

and human responses[86], [89]. In this dissertation, GPS mobility data play a central role in capturing population activity before, during, and after the 2019 Nagano flood, forming a key input for Random Forest models and fine-grid urban analysis.

2.4.2. Social Media Data (Twitter)

Social media platforms such as Twitter provide another critical form of real-time social sensing[91], [92], [93]. Tweets include time-stamped textual content and often contain geographic information, enabling researchers to analyze public reactions, concerns, and observations during emergencies[91]. The spatial and temporal distribution of tweets helps identify where and when residents report anomalies, hazards, or disruptions[92].

Natural Language Processing (NLP) techniques enable the extraction of disaster-related information from large volumes of social media text[94]. Methods such as keyword extraction, sentiment analysis, and topic modeling can identify urgent messages, detect hazard mentions, and reveal evolving public perceptions[94]. Prior research has shown that Twitter data can support flood detection, earthquake monitoring, wildfire tracking, and situational awareness during extreme weather events[91], [92], [95]. In this dissertation, NLP-based keyword filtering is used to detect flood-related terms, assess tweet concentration across time, and locate areas where residents actively report flooding. These results provide a valuable real-time layer that enhances the fine-grid flood estimation model introduced in Chapter 6.

2.4.3. Limitations of Social Sensing

Despite its value, social sensing presents several inherent limitations. Social data is often noisy, containing irrelevant messages, duplicated content, or ambiguous statements that require extensive preprocessing[94], [96]. The spatial distribution of GPS traces and tweets is uneven; densely populated districts produce abundant data, whereas rural or sparsely inhabited regions may generate very few signals, leading to biased interpretations[97]. Data accessibility and privacy remain significant concerns, as mobile phone records and social media content require anonymization and often depend on cooperation from private companies or platform providers[98].

Additionally, social behavior does not always align precisely with physical events[91], [99]. For example, mobility reductions may occur due to heavy rain rather than flooding, and residents might tweet about hazards that are far from their actual location[100]. These discrepancies highlight the need to integrate social sensing with physical sensing to minimize misclassification and ensure reliable disaster assessment[99], [101]. By combining SAR, DEM, GPS, and Twitter data, this dissertation addresses these limitations and demonstrates how heterogeneous datasets can collectively enhance flood estimation accuracy.

2.5 Multi-Source Data Integration in Disaster Estimation

The increasing complexity of urban flood events has revealed the limitations of relying on any single source of information[40]. To fully describe both the physical dynamics of flooding and the societal consequences that follow, disaster researchers increasingly employ multi-source data integration frameworks[102], [103]. It is important to distinguish between *hazard* and *disaster* in this context. A hazard refers to physical phenomenon, such as rainfall, river overflow, or inundation depth, while a disaster captures the human and social impacts generated by that hazard, including population displacement, infrastructure disruption, or changes in mobility[104]. Remote sensing typically reflects hazard characteristics, whereas social sensing represents disaster impacts[105]. Multi-source integration aims to connect these two components, enabling a more complete understanding of flood behavior and societal response[102].

2.5.1. Concepts of Multi-Source Integration

Multi-source data integration refers to the systematic combination of heterogeneous datasets, such as remote sensing imagery, GIS-based terrain indicators and human behavioral data, to produce richer, more comprehensive insights into disaster processes[21], [35]. Physical sensing data (e.g., SAR backscatter, DEM-derived elevation and slope) provide objective measurements of the flood environment, while social sensing data (e.g., GPS mobility patterns, Twitter messages) offer real-time indicators of population reactions and situational awareness[106], [107].

This complementarity makes integrated approaches particularly suitable for urban flood research, where both the physical extent of inundation and human behavioral responses play essential roles. Typical frameworks combine remote sensing, GIS layers, and social sensing inputs to support flood detection, impact estimation, evacuation analysis, and resilience assessment[108], [109]. The approach adopted in this dissertation builds on this foundation by integrating SAR-derived flood indicators, terrain features, GPS activity signals, and Twitter-derived information across multiple spatial scales, resulting in a unified and comprehensive estimation framework.

2.5.2. Integration Methods

Several methodological strategies exist for integrating multi-source datasets, each suited to different data types and research goals.

Rule-based integration uses expert-defined conditions or thresholds[37]. For example, it can identify potential flood zones where SAR backscatter drops below a certain value, and the elevation is low[110]. These methods are intuitive but often lack flexibility when applied to heterogeneous urban environments.

Statistical integration methods, including Bayesian approaches and Dempster–Shafer Theory (DST), incorporate uncertainty into the integration process[111], [112]. Bayesian

models integrate prior knowledge with observational data, while DST supports combining multiple uncertain or partially conflicting pieces of evidence, such as estimations from different machine learning models or heterogeneous physical and social indicators[111], [112].

Machine learning–based integration has gained prominence due to its ability to handle high-dimensional, nonlinear relationships among diverse features[113]. Algorithms such as Random Forest, SVM, XGBoost, and deep learning models allow the simultaneous use of SAR backscatter, DEM-based terrain indicators, distance features, and human mobility or communication signals[114], [115]. Machine learning approaches require minimal assumptions about the underlying data distribution and can adapt effectively to complex urban environments[116].

2.5.3. Machine Learning for Flood Estimation

Within machine learning methods, Random Forest (RF) is particularly well suited for integrating multi-source geospatial datasets[117]. Its ensemble-based structure enables it to model nonlinear interactions between features and remain robust to noise, missing data, and multicollinearity[118]. Importantly, RF provides feature importance metrics that support interpretation, which is an important advantage in disaster and Knowledge Science contexts[119].

Previous studies have demonstrated the applicability of RF combined with SAR data for flood detection, and with DEM-derived terrain attributes for inundation susceptibility analysis[120]. However, most existing research integrates only physical indicators[121]. In contrast, this dissertation expands the application of RF to include human behavioral indicators such as GPS-based mobility patterns and Twitter-derived textual evidence, demonstrating that these variables substantially enhance the accuracy and spatial reliability of flood-extent estimation in urban environments[122], [123]. Moreover, the dissertation extends integration from coarse administrative or grid units to a fine-scale 70 m × 70 m estimation model, addressing the need for high-resolution urban flood assessment.

2.5.4. Research Gaps in Multi-Source Integration

Despite the advancements in this field, several gaps remain, many of which directly motivate and justify the contributions of this dissertation.

First, studies rarely combine SAR, DEM, GPS, and Twitter data within one integrated modeling framework. Most research relies on only one or two sensing domains, resulting in incomplete representations of flood processes.

Second, previous approaches lack scalability across spatial resolutions. While some studies utilize city-wide grids and others focus on fine-resolution modeling, there is little work demonstrating how integrated models can operate effectively from coarse to fine

scales. This dissertation addresses this by developing a multi-level modeling approach ranging from 500 m grids to 70 m × 70 m grids.

Third, while many studies focus on improving modeling accuracy, they give limited attention to how integrated data contributes to knowledge creation and interpretative understanding—a core dimension of Knowledge Science. The ways in which environmental signals and human behavior co-evolve during a disaster remain insufficiently examined.

Fourth, although social sensing data are increasingly used to capture disaster impacts, few studies explicitly analyze how human behavior interacts with physical flood dynamics. GPS and Twitter data are often used only descriptively, without integration into quantitative models that can explain how government alerts, rainfall intensity, or population mobility relate to observed flood outcomes.

By addressing these gaps, the present dissertation positions multi-source data integration not merely as a technical enhancement but as a conceptual framework for understanding the interrelationship between hazard processes, human behavior, and urban resilience.

2.6 Urban Resilience and Knowledge Science Framework

Understanding urban disaster estimation and assessment requires not only technical modeling but also a conceptual grounding in resilience theory and knowledge creation[124]. Flood events emerge from the interaction between physical hazards and human systems, and the ability of a city to prepare for, respond to, and learn from such events reflects its broader resilience[125]. At the same time, Knowledge Science offers a framework for explaining how diverse data sources can be transformed into actionable knowledge[126]. This section establishes the theoretical basis for connecting multi-source data integration with urban resilience and the knowledge creation process.

2.6.1. Urban Resilience Concepts

Urban resilience has multiple conceptual origins, broadly categorized into engineering resilience, ecological resilience, and social resilience[127]. Engineering resilience focuses on the speed of recovery and the capacity of a system to return to its pre-disaster state[128]. This perspective emphasizes robustness, structural strength, and the ability to absorb shocks[129]. By contrast, ecological resilience highlights the capacity of a system to adapt and reorganize while maintaining essential functions[130]. From this lens, a resilient urban system is not simply one that withstands disturbance, but one that adapts to evolving environmental conditions[124]. Social resilience extends the concept to communities and institutions, referencing the capacities of individuals, social networks, infrastructure systems, and governance structures to anticipate risk, respond collectively, and mobilize resources[125].

In the context of urban flooding, resilience involves the interplay between physical

environments and social systems[127]. Topography, drainage capacity, river morphology, and building density shape the spatial distribution of hazard exposure[131]. At the same time, mobility patterns, social communication, community responses, and institutional actions determine how a hazard evolves into a disaster[122]. Resilience therefore cannot be understood solely by analyzing physical inundation; it must also incorporate the behavioral and cognitive responses of urban residents[125]. Social sensing data, such as GPS mobility traces and Twitter messages, provide a direct means of observing these responses, thereby expanding the conventional physical-centric understanding of resilience[122], [132].

2.6.2. Disaster Knowledge Creation

Knowledge Science provides a conceptual lens to understand how raw data becomes meaningful knowledge that supports decision-making[133]. A disaster situation generates massive quantities of environmental signals and human behavioral responses, but without systematic processing, these signals remain fragmented[134]. The transformation from data to actionable knowledge follows a hierarchical process: data must first be processed into structured information, which can then be interpreted and contextualized to form knowledge[135]. This conceptual trajectory aligns with the SECI knowledge-creation model, in which knowledge evolves through socialization, externalization, combination, and internalization[136].

Social sensing plays a particularly key role in this transformation[137]. During a flood event, residents collectively generate information through their movement patterns, social media communication, and local reporting[134]. This collective sensing mechanism produces emergent knowledge about the disaster's progression, population reactions, and public awareness[138]. When integrated with remote sensing observations, these human-generated signals enrich the situational understanding of the event, enabling interpretations that go beyond purely physical hazard indicators[132].

This dissertation adopts the perspective that urban resilience is not only determined by physical resistance but also by a city's ability to sense, interpret, and learn from disaster-related information. Real-time behavioral data reflect the adaptive responses of residents, while remote sensing capture's objective environmental conditions. Knowledge creation thus occurs at the intersection of these two domains, where physical and social signals collectively form a coherent understanding of disaster dynamics.

2.6.3. Linking Multi-Source Data to Knowledge Science

The integration of multi-source data in this research serves not only as a methodological advancement but also as a knowledge science contribution. Each data modality represents a distinct form of knowledge about the urban environment[139]. Remote sensing data provide objective physical knowledge of the hazard: inundation extent, terrain characteristics, and environmental conditions[131]. Social sensing data

offers behavioral and cognitive knowledge, revealing how residents perceive risk, modify movement patterns, and share information during the disaster[140]. Machine learning models serve as a mechanism for integrating these forms of knowledge by capturing nonlinear relationships, synthesizing large volumes of heterogeneous data, and producing interpretable indicators such as feature importance or predicted flood probability[116].

Through this perspective, multi-source integration becomes a knowledge creation process, not merely an analytical technique. It enables the construction of a spatially explicit knowledge base in which environmental signals and human responses mutually inform predictive modeling. This integrated knowledge supports more resilient decision-making by providing disaster managers with a unified understanding of physical hazard dynamics and societal impacts[141]. Moreover, it establishes a conceptual framework in which data-driven estimations contribute to urban learning, thereby strengthening long-term resilience planning[142].

2.7 Summary

This chapter reviewed the theoretical and empirical foundations relevant to urban flood estimation and resilience assessment, drawing on previous studies covering remote sensing technologies, social sensing data, multi-source data integration, machine-learning approaches, and the Knowledge Science perspective. Existing literature has shown that although remote sensing and hydrological modeling have long served as standard tools for flood assessment, these approaches predominantly focus on the physical dimensions of hazards and often struggle to capture human behavioral responses or real-time urban dynamics. Similarly, prior research indicates that while social sensing data sources such as GPS mobility traces and Twitter messages provide valuable behavioral and cognitive information, they are frequently noisy, spatially uneven, and rarely integrated systematically with physical sensing frameworks.

From a methodological viewpoint, many existing flood estimation studies rely on single-source data, typically SAR imagery or DEM-derived terrain features, or combine only physical geospatial variables. Previous applications of social sensing data in disaster studies have primarily focused on situational awareness and qualitative analysis rather than quantitative flood-extent estimation. As a result, earlier models reported in the literature often operate at coarse spatial resolutions, have limited near–real-time capability, and lack mechanisms for translating heterogeneous inputs into a unified analytical framework. Although machine-learning techniques such as Random Forest have demonstrated strong performance in handling nonlinear geospatial features, their application to fully integrated, multi-source urban flood estimation remains limited in existing research.

The gaps identified across the literature highlight the necessity of this research. First, at the data level, no prior study has systematically integrated SAR imagery, DEM-derived terrain features, GPS mobility data, and Twitter-based social media information into a

coherent analytical framework for flood-extent estimation. Previous studies typically address only one or two of these data modalities, despite each capturing a distinct dimension of urban flood dynamics. Second, at the methodological level, the literature reveals a lack of multi-scale models capable of transitioning from coarse-grid behavioral anomaly detection to fine-grid urban flood estimation. The Random Forest–based framework proposed in this dissertation addresses this gap by enabling multi-resolution analysis and interpretable feature assessment. Third, at the theoretical level, existing disaster-related studies rarely link multi-source data integration with the knowledge creation processes emphasized in Knowledge Science. This dissertation extends prior work by explicitly framing flood estimation as a process in which data are transformed into information, integrated into knowledge, and applied to support resilient urban decision-making.

Overall, Chapter 2 synthesizes existing research findings and establishes the conceptual and methodological motivation for this dissertation. The limitations reported in prior remote sensing–based and social sensing–based disaster studies, together with the absence of integrated analytical frameworks, underscore the importance of developing a comprehensive sensing–integration–application pipeline. Building on insights from previous studies, the proposed approach advances the state of the art by positioning multi-source data integration as a knowledge creation process that enhances urban resilience. The next chapter builds on this foundation by introducing the methodological framework underpinning the dissertation's three experiments.

Chapter 3 Methodological Framework

3.1 Research Framework Overview

Urban flood estimation and assessment require analytical frameworks capable of capturing both the physical mechanisms of hazard development and the behavioral responses of affected populations[116]. Traditional disaster models generally rely on hydrological parameters or remote sensing imagery, but these approaches often struggle to represent rapid changes in human activity during extreme events[134]. In contrast, the increasing availability of social sensing data, such as mobile GPS traces and real-time Twitter posts, provides an opportunity to observe how residents respond to emerging flood conditions[140]. This dissertation therefore adopts a methodological framework that integrates physical sensing and social sensing into a unified analytical system for urban flood estimation.

The overall research framework is structured around three interconnected stages: sensing, integration, and application. These stages are illustrated in Figure 3.1, which provides a conceptual overview of how heterogeneous information is processed throughout the study.

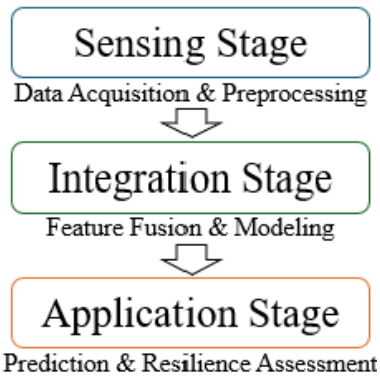


Figure 3.1: Conceptual workflow of the sensing–integration–application framework

The methodological structure of this dissertation follows a progressive and cumulative logic. Experiment I examines GPS data to determine whether social sensing alone can reveal temporal and spatial anomalies associated with flooding. This establishes the behavioral dimension of disaster monitoring. Experiment II builds on this foundation by integrating SAR backscatter, DEM-derived terrain indices, and population mobility data within a Random Forest model to evaluate how multi-source inputs enhance flood-extent estimation. Experiment III further expands the integration by incorporating real-time Twitter information and increasing spatial resolution to a fine-grid scale, enabling a more detailed understanding of urban flood dynamics and near-real-time estimation capability.

This framework also reflects the principles of Knowledge Science. Through the integration of physical observations and human behavioral signals, raw data is gradually

transformed into spatial knowledge that can support urban disaster management. The sensing stage corresponds to data collection, the integration stage corresponds to information structuring and knowledge generation, and the application stage provides decision-relevant insights that improve urban resilience.

In summary, the methodological framework presented in this chapter establishes the foundation for the analyses that follow. It integrates heterogeneous data, bridges multiple analytical approaches, and supports the overall goal of developing a multi-source, knowledge-driven system for urban flood estimation and assessment.

3.2 Data Sources and Features

This dissertation integrates multiple heterogeneous datasets derived from remote sensing, topographic and environmental databases, and social sensing platforms. These datasets collectively form the foundational feature set for flood estimation and assessment at multiple spatial scales. Table 3.1 provides an overview of all data sources, while the following subsections describe their features and relevance in detail.

The selection of datasets in this dissertation was not based on data availability alone. Instead, each dataset was deliberately chosen to represent a specific mechanism relevant to flood detection and impact assessment. Remote sensing data captures physical surface water expansion, terrain indices represent hydrological controls on flood formation, and social sensing data reflect human behavioral and perceptual responses to flooding. This theory-driven selection ensures that the integrated dataset provides a comprehensive representation of urban flood processes.

From the physical perspective, SAR backscatter decreases over water surfaces due to specular reflection, a principle derived from electromagnetic scattering theory. Multi-orbit SAR further reduces layover and shadowing, enabling more reliable detection in urban settings. In this study, flooding is therefore defined as abnormal surface water expansion associated with river overflow during the disaster period, distinguished through temporal comparison between pre-event and event-period observations and validated using official inundation records.

From the hydrological perspective, terrain-based indices such as TWI, SPI, elevation, and slope represent the structural controls governing water accumulation and flow pathways. These geospatial indicators encode the physical processes that shape the spatial distribution of flood extent.

From the behavioral perspective, GPS mobility patterns are explained by behavioral adaptation theory, which states that individuals adjust movement in response to perceived risk and accessibility constraints. Mobility reductions therefore function as measurable indicators of exposure and disruption.

Similarly, crisis communication theory suggests that disaster-related messaging on social media increases where local impacts are felt. Twitter activity thus reflects real-time situational awareness, collective stress, and localized disaster perception, providing information that physical sensors cannot capture. Together, these theoretical foundations

justify the integrated use of physical, hydrological, and behavioral sensing for urban flood estimation.

3.2.1. Remote Sensing Data

Remote sensing datasets constitute the physical sensing layer within the multi-source integration framework. These datasets provide spatially continuous, weather-independent observations of surface conditions during the 2019 Nagano flood event and form the physical foundation of the estimation model.

(a) Sentinel-1 SAR (Ascending and Descending Orbits)

Sentinel-1A and Sentinel-1B IW-mode GRDH images (approximately 5 × 20 m resolution) were acquired for pre-flood and co-flood periods. Both ascending and descending orbits were utilized to obtain multi-angle backscatter observations, thereby reducing layover effects and enhancing the detectability of inundated surfaces. This multi-orbit configuration improves estimation reliability in complex urban environments by capturing scattering differences arising from varying look geometries.

Preprocessing included radiometric calibration, thermal noise removal, orbit correction, speckle filtering, and terrain correction. The resulting calibrated backscatter coefficients serve as core inputs for pixel-level and grid-level estimation in Chapters 5 and 6.

(b) ALOS-2 SAR

L-band ALOS-2 imagery provides longer-wavelength scattering information that penetrates vegetation more effectively and offers higher coherence stability. These characteristics improve flood detection in areas with mixed land cover where shorter-wavelength sensors may underestimate inundation. When used, ALOS-2 images undergo standard radiometric and geometric preprocessing to derive calibrated backscatter features.

(c) Remote Sensing Feature Rationale

The selection of SAR is grounded in electromagnetic scattering theory: water surfaces produce low backscatter because of specular reflection, making SAR highly sensitive to inundation even under cloud or rain conditions. L-band (ALOS-2) and C-band (Sentinel-1) provide complementary responses—L-band enhances detection in vegetated zones, while C-band is more sensitive to shallow urban flooding.

Optical imagery was deliberately excluded due to persistent cloud cover during the typhoon event, and alternative SAR datasets lacked either temporal frequency or spatial coverage. Thus, the combined use of Sentinel-1 and ALOS-2 represents the most theoretically sound and operationally suitable remote sensing configuration for flood extent estimation in Nagano City.

3.2.2. Topographic and Environmental Data

Topographic and environmental features describe the underlying physical landscape that governs water accumulation, flow paths, and exposure conditions. These features enrich machine-learning models by embedding hydrological and structural information about the study area.

(a) Digital Elevation Model (DEM)

A 10 m DEM is used to derive elevation and slope. Elevation influences gravitational flow direction and helps identify low-lying flood-prone areas. Slope affects runoff speed and drainage potential.

(b) Terrain Indices (TWI and SPI)

Hydrological indices computed from the DEM include Topographic Wetness Index (TWI): Indicates potential moisture accumulation and surface saturation. Stream Power Index (SPI): Represents erosive or hydraulic force related to concentrated flow. Both indices are used as inputs in Chapters 5 and 6 to improve terrain-driven flood estimation.

(c) River and Road Network Data

River centerlines and major road networks were used to derive distance-based indicators. Proximity to the Chikuma River serves as an important indicator of flood susceptibility, while distance to road networks helps characterize potential mobility constraints and accessibility during the flood.

(d) Infrastructure Layers

Spatial datasets of public shelters, hospitals, and public service facilities were incorporated. Distance-to-infrastructure variables were calculated for each grid cell at both coarse (500 m) and fine (70 m) resolutions. These indicators support the interpretation of flood-extent estimation results by linking inundated zones with human-system exposure and vulnerability.

3.2.3. Social Sensing Data

Social sensing datasets provide real-time or near-real-time indicators of behavioral responses and public information exchange during the flood. These datasets capture dimensions of disaster impact—risk perception, accessibility disruption, and situational awareness—that are not observable through physical sensing alone.

(a) GPS Mobility Data (Agoop)

The GPS dataset consists of anonymized mobile device locations recorded before, during, and after the flood. Three key features were derived: Population mobility density, aggregated to 500 m grids (Chapter 4) and 70 m grids (Chapter 6). Temporal anomaly indicators, quantifying deviations from pre-flood mobility baselines. Movement-based behavioral signatures, reflecting evacuation dynamics, constrained mobility, or localized disruptions.

From a theoretical perspective, GPS mobility is grounded in behavioral adaptation theory: individuals modify their movements in response to perceived risk and accessibility constraints. Thus, mobility disruptions constitute measurable behavioral proxies of flood impact.

(b) Twitter Geotagged Data

A total of 1,520 geolocated tweets from October 12–13, 2019 were collected. After natural language processing and keyword extraction, several features were derived: Keyword frequency distributions, indicating shifts in public attention and hazard perception. Fine-scale spatial tweet density, mapped onto 70 m grids. Temporal keyword variation reflects evolving situational awareness.

Twitter activity corresponds to crisis communication theory, which posits that communication intensifies in locations experiencing direct impacts or heightened risk perception. Thus, Twitter provides complementary evidence of local experiences, self-reported impacts, and community-level information flow.

(c) Social Sensing Feature Rationale

GPS and Twitter were selected because they provide two complementary behavioral dimensions:

GPS → objective mobility disruption (exposure, accessibility loss, behavioral adaptation)

Twitter → subjective perception and communication (awareness, distress, self-reporting)

Other platforms (e.g., Facebook, Instagram, LINE) were excluded due to insufficient geolocation accuracy, data access restrictions, or low disaster-related activity during the event.

Thus, GPS and Twitter together offer a theoretically grounded, operationally feasible representation of human-centered disaster signals.

(d) Role of Social Sensing Within Multi-Source Fusion

While SAR captures the physical distribution of floodwater, GPS and Twitter reveal human behavioral dynamics that evolve alongside physical processes. Integrating physical and behavioral indicators enables a more complete understanding of flood progression and supports estimation models that reflect both environmental conditions and human experience.

Table 3.1: Summary of Feature Variables Used in the Thesis

Category	Feature Name	Type	Unit / Resolution	Source	Role in Model
Remote sensing	SAR backscatter (VV)	Continuous	5 × 20 m	Sentinel-1	Detect floodwater extent
	Elevation	Continuous	10 m	GSI DEM	Terrain baseline
	Slope	Continuous	Degree	DEM-derived	Surface gradient
	Topographic Wetness Index (TWI)	Continuous	–	DEM-derived	Water accumulation potential

	Stream Power Index (SPI)	Continuous	–	DEM-derived	Flow energy indicator
Social sensing	GPS population density	Continuous	500/70 m grid	Agoop	Proxy for human activity
	Tweet frequency index (TFI)	Count	70 m grid	Twitter API	Public perception indicator
Auxiliary	Distance to river	Continuous	m	MLIT	Flood exposure
	Distance to shelter	Continuous	m	MLIT	Accessibility and resilience
	Land-use class	Categorical	–	MLIT	Surface type and vulnerability

3.3 Multi-Source Data Integration Strategy

The integration of remote sensing, topographic, and social sensing data forms the core methodological foundation of this dissertation. Because each dataset represents a different dimension of the disaster process, the analytical framework requires a consistent strategy to align, transform, and harmonize heterogeneous information into a unified spatial and temporal structure. The integration process ensures that physical sensing, social behavior, and real-time information streams can be analyzed together within machine-learning models at multiple spatial resolutions.

Remote sensing data, including Sentinel-1 SAR imagery and DEM-derived terrain variables, provide objective and continuous measurements of the physical environment during the flood. SAR backscatter reveals the presence of surface water and moisture patterns under all-weather conditions, while elevation, slope, TWI, and SPI describe the hydrological and geomorphological setting that governs flood pathways. These physical sensing datasets form the environmental foundation of the integrated model and serve as baseline indicators of flood occurrence.

Social sensing data enriches the physical models by incorporating human behavioral and perceptual dimensions. GPS mobility data capture spatial variations in population flow before, during, and after the flood event. Changes in movement density, reductions in mobility, or concentrated patterns of evacuation indicate how residents respond to hazard conditions. Meanwhile, geolocated Twitter messages provide real-time expressions of public awareness, distress signals, and first-hand observations. Through natural language processing, keywords related to flooding are extracted and aggregated to represent collective situational awareness. These social sensing sources introduce behavioral features that cannot be derived from remote sensing alone, thereby improving the model's ability to identify socially relevant flood impacts.

To integrate these heterogeneous datasets, a unified spatial framework based on grid

cells is applied. Depending on the analytical purpose and the characteristics of each dataset, different grid resolutions are employed. A 500 m grid is used in Experiment I to detect large-scale population mobility anomalies, as coarse grids reduce noise in individual GPS movements and highlight broader behavioral trends that correspond to flood progression at the city scale. Experiment II adopts an intermediate resolution of 100 m, which balances spatial detail with computational stability when integrating SAR backscatter and terrain features with population data. Experiment III uses a 70 m grid to support fine-resolution urban flood estimation, as this scale preserves local heterogeneity in residential, commercial, and transportation zones and aligns more closely with the spatial patterns of inundation observable in high-density built environments. Each grid cell serves as a common spatial unit onto which all physical and social features are mapped, allowing datasets with different native resolutions, coordinate systems, and sampling patterns to become comparable. The use of multiple grid sizes is therefore not arbitrary but reflects the analytical needs of each experiment: coarse grids to reveal broad behavioral patterns, mid-scale grids for multi-source model integration, and fine-scale grids for detailed urban estimation. The grid-based approach eliminates sampling biases inherent in point-based observations, enables aggregation without losing critical spatial variation, and provides standardized inputs for machine-learning models across all experimental stages.

The spatial integration process includes several key steps. All datasets are first reprojected to a common coordinate system to ensure geometric consistency. Remote sensing features are calculated at the pixel level and spatially joined to the grid cells using area-weighted or mean aggregation. GPS data are filtered based on accuracy, speed thresholds, and temporal segmentation, then converted into continuous mobility-density surfaces using kernel density estimation. These density values are assigned to grid cells according to their spatial location. Twitter data undergo language preprocessing and keyword extraction before being aggregated by the administrative or grid unit in which each geotagged message is located. The entire feature set is standardized to ensure numerical consistency across variables and to prevent scale-dependent bias during machine-learning training.

This multi-source integration strategy creates a unified representation of the disaster environment, one that simultaneously reflects physical hazard conditions, population behavior, and collective public perception. By aligning remote sensing, topographic variables, GPS mobility patterns, and social media indicators within a common grid structure, the framework enables a comprehensive and scalable approach to urban flood estimation. This integrated spatial knowledge base supports the subsequent modeling in Chapters Four, Five, and Six and ensures that the machine-learning models can draw on the full complexity of multi-source information.

3.4 Machine Learning Modeling Framework

3.4.1. Choice of Random Forest

The Random Forest (RF) classifier serves as the primary analytical model across this dissertation because it is highly suitable for integrating heterogeneous geospatial and social-sensing data. RF can handle complex nonlinear relationships, which are intrinsic to urban flood processes where hydrological, geomorphological, and behavioral factors interact in non-additive ways. Its ensemble structure enhances robustness against noise and outliers, a critical advantage when working with GPS mobility traces or social media data that inherently contain variability. Furthermore, RF accommodates multi-dimensional feature spaces without requiring strict distributional assumptions, making it particularly appropriate for combining SAR backscatter, terrain indices, distance-based environmental variables, and human mobility indicators. The model also produces feature-importance measures that support interpretability, allowing physical and behavioral determinants of flooding to be examined transparently.

To address model selection concerns, Support Vector Machine (SVM) models were implemented as comparative baselines in preliminary experiments. While SVM achieved reasonable classification performance, RF demonstrated more stable results and clearer interpretability when integrating multi-source physical and social data. Therefore, RF was adopted as the central modeling framework of this dissertation due to its superior robustness, integration capacity, and explanatory potential for multi-source flood estimation.

3.4.2. Feature Preparation

All experiments in this dissertation follow a unified strategy for feature construction to ensure comparable and standardized model inputs. Physical sensing features include SAR backscatter coefficients derived from Sentinel-1 ascending and descending orbits, which provide water-sensitive scattering information essential for flood detection. Topographic and hydrological indicators, including elevation, slope, the Topographic Wetness Index (TWI), and the Stream Power Index (SPI), describe terrain structure and the likelihood of surface water accumulation or concentrated flow. Environmental context is represented by distance metrics measuring the proximity of each grid cell to rivers, major roads, shelters, hospitals, and public service facilities. Social sensing features further expand the model's descriptive capacity. GPS density is computed from kernel-based processing of mobile-phone location traces to quantify the magnitude and spatial distribution of population movement before and during the flood. Twitter-derived variables, produced through natural language processing, represent the frequency of flood-related keywords and provide supplementary real-time evidence of disaster perception and reporting. After preprocessing, filtering, and spatial alignment, all features are standardized onto the unified grid system to form the multi-source input matrix for RF modeling.

3.4.3. Feature Preparation

The Random Forest model is trained following a consistent procedure across the different experimental settings in Chapters 4, 5, and 6. The dataset is divided into training and testing subsets using an 80/20 ratio to ensure stable model learning while maintaining sufficient samples for independent validation. Parameter optimization is performed through grid search, in which combinations of decision-tree depth, the entropy-based splitting criterion, and the number of estimators are systematically evaluated to identify the configuration that yields the highest predictive performance. After the optimal parameters are determined, the final RF model is trained on the full training data and tested using the held-out samples. Model performance is assessed using the confusion matrix, which summarizes the distribution of true positives, false positives, true negatives, and false negatives. Based on this matrix, Accuracy, Precision, Recall, and F1-score are calculated to provide a comprehensive understanding of estimation quality under varying class conditions. In addition, feature-importance metrics are computed to quantify the contribution of each physical, environmental, and social variable, forming an essential link between the model's computational behavior and the knowledge-science interpretation of flood phenomena.

3.5 Experimental Design Logic

The overall experimental design of this dissertation follows a progressive structure that moves from disaster sensing to multi-source data integration, and finally to urban-scale application. This logical progression reflects the conceptual framework of the thesis, where social sensing data, remote sensing data, and machine learning are gradually combined to deepen the understanding and estimation of urban flood dynamics. The three experiments, although presented in separate chapters, form a coherent analytical sequence in which each stage builds directly upon the findings and limitations of the previous one.

The first experiment investigates whether human mobility data can serve as an effective indicator of disaster occurrence. Using coarse 500-m grids, GPS data from mobile phones were analyzed to detect temporal and spatial anomalies during the 2019 Nagano flood. This analysis revealed significant changes in population movement patterns that were closely synchronized with rainfall peaks, river water-level rises, and official inundation timelines. The experiment therefore establishes social sensing—specifically GPS mobility data—as a valid behavioral signal capable of capturing the evolution of a flood event. This behavioral insight forms the conceptual foundation for incorporating social sensing into the subsequent flood-extent estimation models developed in later chapters.

Building on these findings, the second experiment focuses on the development of a Random Forest framework for flood-extent estimation using multi-source geospatial data. This experiment contains two major components. The first examines the contribution of GPS information within the Random Forest model by comparing a model using only

physical environmental features (SAR backscatter, elevation, slope, and terrain indices) with a model that additionally incorporates GPS-derived population density change. The inclusion of GPS significantly improves the spatial coherence and accuracy of the estimated inundation distribution, demonstrating that behavioral signals can meaningfully complement physical flood indicators. The second component evaluates the value of integrating ascending and descending Sentinel-1 SAR data. Because the two viewing geometries capture different backscattering characteristics in complex urban landscapes, their combined use reduces misclassification caused by layover, shadowing, and geometric distortions. Results show that dual-orbit SAR inputs improve the model's ability to delineate flood boundaries, particularly in built-up areas. Together, these two components illustrate how both social sensing and enhanced physical sensing contribute independently and complementarily to the accuracy of machine-learning–based flood estimation.

The third experiment advances the research framework from regional analysis to fine-scale urban estimation. Using 70-m grids, this experiment integrates georeferenced Twitter messages with GPS, SAR, DEM, and infrastructure layers to construct a high-resolution Random Forest model capable of capturing real-time flood dynamics. Through natural language processing, flood-related Twitter keywords provide direct evidence of public perception and localized disaster reporting, while GPS mobility anomalies reflect behavioral reactions. When fused with the physical sensing layers, these social indicators significantly reduce false positives and yield a more precise representation of the actual inundation distribution. This experiment therefore demonstrates the practical potential of multi-source data integration for near-real-time urban flood mapping, completing the sensing → integration → application cycle established in Chapter 3.

Taken together, the three experiments form an escalating methodological pathway: Experiment I validates social sensing as a behavioral disaster signal; Experiment II embeds these signals into a machine-learning framework together with enhanced SAR inputs; and Experiment III operationalizes a fine-grid, multi-source estimation system suitable for urban disaster response. To visually support this progression, Figure 3.2 illustrates the comparative structure, data flows, and analytical outputs of the three experiments.

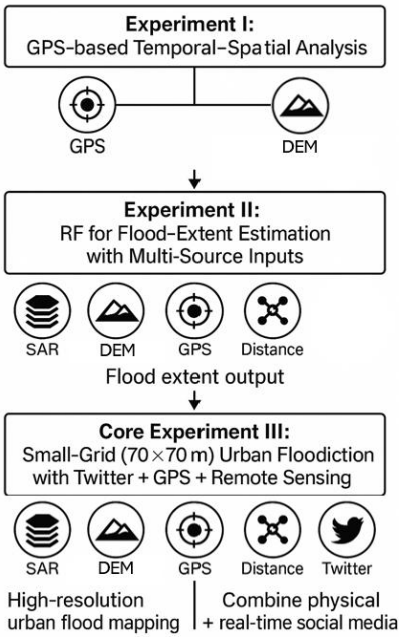


Figure 3.2: Overall Experimental Workflow and Progressive Research Structure

3.6 Integration into a Knowledge Science Framework

The methodological framework presented in this dissertation extends beyond data processing and machine learning estimation. It is fundamentally positioned within the broader discipline of Knowledge Science, which seeks to understand how data, information, and human cognition interact to generate actionable knowledge. In the context of urban disaster management, this dissertation conceptualizes flood estimation not merely as a technical modeling task, but as a knowledge-creation process that emerges through the integration of multi-source sensing modalities, analytical models, and human behavioral responses.

(a) From Data to Information to Knowledge

The multi-source datasets used in this research—remote sensing (SAR, DEM), environmental variables (TWI, SPI, distances), and social sensing inputs (GPS mobility, Twitter posts)—represent the "data" layer. These raw observations require significant preprocessing, spatial alignment, and feature engineering before they become interpretable "information." Through Random Forest modeling, these processed indicators are transformed into "knowledge" about flood mechanisms, susceptibility patterns, and human behavioral responses.

This transformation aligns with the knowledge-creation concepts in the SECI model, where externalized observational data become integrated and interpreted through iterative cycles of combination and internalization. The proposed framework demonstrates how machine learning acts as the combination phase, merging heterogeneous information into a unified analytical output, while spatial interpretation, feature importance analysis, and

cross-experiment comparison represent the internalization of new knowledge about disaster dynamics.

(b) Collective Sensing and Knowledge Integration

A central theoretical contribution of this dissertation is the concept of collective sensing, defined as the integration of physical sensors and human sensors into a multi-layered perception system for urban disasters. Satellite SAR imagery provides consistent physical measurements, GPS reveals population mobility patterns, and Twitter messages reflect real-time human cognition and communication. Individually, these sources reveal partial aspects of the disaster process; collectively, they form a socio-technical sensing ecosystem capable of detecting, interpreting, and validating disaster phenomena.

The integration of these data sources operationalizes a form of knowledge integrational process in which different types of information (physical, behavioral, semantic) are reconciled to achieve higher levels of understanding. This aligns with the Knowledge Science view that resilience depends not only on physical robustness but also on the city's information-processing capabilities, including its ability to sense, interpret, and learn from extreme events.

(c) Role of Machine Learning in Knowledge Creation

Random Forest modeling, as the central analytical tool of this dissertation, plays a crucial role in organizing multi-source information into a coherent knowledge structure. The inherent interpretability of RF—particularly through its feature importance metrics—reveals how different environmental and social variables contribute to flood occurrence. This transparency transforms the model from a purely estimation tool into an explanatory mechanism, enabling insights into disaster causation and vulnerability to emerge through data-driven learning. For example, the consistently high importance of elevation and slope confirms their foundational role in physical flood susceptibility, while the varying contributions of GPS and Twitter highlight the behavioral dimension of disaster impacts.

(d) Resilience as a Knowledge-Based System

Through the integration of remote sensing and social sensing, this dissertation argues that urban resilience is fundamentally a knowledge-based construct. Physical resilience—such as embankment strength or drainage capacity—is only one dimension. Information resilience, defined as the capacity of a city to sense, interpret, and respond to hazards in a timely manner, is equally critical. The collective sensing approach demonstrates that real-time population movement and social communication provide key cognitive signals that enhance situational awareness during disasters. Thus, resilience emerges not simply from infrastructure, but from the dynamic interplay among physical conditions, behavioral patterns, and shared knowledge.

(e) Three Experiments as a Knowledge-Creation Cycle

The structure of the dissertation—Experiment I (GPS anomaly analysis), Experiment II (RF flood estimation), and Experiment III (fine-grid multi-source fusion)—follows a progressive knowledge-creation cycle:

Sensing: Experiment I identifies how flood dynamics manifest in human mobility

anomalies using GPS data.

Integration: Experiment II develops a unified modeling framework that integrates physical and social information into a cohesive analytical model.

Application: Experiment III advances to fine-grid, high-resolution flood estimation that synthesizes physical sensing, behavioral sensing, and semantic information (Twitter), representing the full maturation of knowledge integration.

Together, these experiments illustrate how knowledge evolves from raw signals to structured understanding, ultimately supporting resilient decision-making in urban environments. The cumulative outcome is a socio-technical system for disaster estimation that aligns with the epistemological principles of Knowledge Science and the practical needs of resilient city management.

3.7 Summary

This chapter established the methodological foundation of the dissertation by presenting a unified framework for multi-source data integration, machine learning modeling, and knowledge-based interpretation for urban flood estimation and assessment. The chapter first introduced the overall conceptual structure linking sensing, integration, and application, which defines how raw observational data are transformed into actionable knowledge in urban disaster contexts. A comprehensive description of all data sources—remote sensing, topographic and environmental variables, and social sensing datasets—clarified the heterogeneous nature of the information involved and the complementary roles each dataset plays within the analytical system.

The chapter then detailed the integration strategy used to align these disparate datasets into a consistent spatial framework. This included preprocessing procedures, grid-based spatial harmonization at multiple resolutions, feature engineering, and normalization, forming the basis for multi-source comparability and enabling robust machine learning analysis. The Random Forest modeling framework was described as the core analytical method, emphasizing its advantages in handling nonlinear relationships, multi-source features, and interpretability through feature importance. Training procedures, parameter tuning, and evaluation metrics were also outlined to ensure methodological consistency across subsequent experiments.

A key component of this chapter was the articulation of the experimental design logic. The dissertation's three experiments were positioned as progressive stages in a knowledge creation cycle: Experiment I explores GPS-based behavioral anomalies during flood events; Experiment II integrates physical and behavioral sensing using Random Forest to evaluate the added value of GPS and SAR; and Experiment III advances to a fine-grid, multi-source model that incorporates Twitter data for high-resolution urban flood estimation. This layered design demonstrates how analytical complexity and data richness increase across experiments, reflecting a systematic progression from sensing individual modalities to integrating them and finally applying them in a detailed urban-scale model.

Finally, the chapter connected the methodological framework to the principles of

Knowledge Science. By showing how multi-source sensing data transition from raw observations to structured knowledge, and how collective sensing emerges from the fusion of physical and social information, the chapter positioned the dissertation within a broader epistemological context. This integration underscores the central argument that urban resilience depends not only on physical conditions but also on information responsiveness, interpretive capability, and collective learning.

Together, these methodological components provide the necessary groundwork for the empirical analyses presented in Chapters 4 through 6. They ensure that the subsequent experiments are not isolated case studies, but integral parts of a cohesive framework aimed at advancing both the practical and theoretical understanding of urban disaster estimation.

Chapter 4 Experiment I: GPS-based Temporal–Spatial Analysis

This chapter is based on the following published paper: Yifan Yang, Naoki Ohira and Hideomi Gokon, “Time and Spatial Analysis of Flood Disaster based on GPS Data: A Case Study of Flood Disaster in Nagano City in 2019,” Journal of Social Safety Science, vol. 44, pp. 121–130, 2024. Some descriptions have been reorganized and expanded to align with the structure and objectives of this dissertation.

This chapter presents the first empirical experiment, which examines whether mobile phone GPS data can capture temporal and spatial mobility anomalies associated with the 2019 Typhoon Hagibis flood in Nagano City. As discussed in Chapter 3, social sensing provides behavioral information that complements physical sensing data, offering a human-centered perspective on disaster dynamics. Before integrating GPS into multi-source machine learning models in later chapters, it is necessary to verify its independent ability to reflect flood-related population responses.

Experiment I analyze GPS records across a 500 m grid and compare movement patterns during the flood weekend with those from a normal weekend. Temporal changes are used to identify abnormal decreases caused by heavy rainfall and abnormal increases linked to evacuation. Spatial clustering of these anomalies reveals areas most affected by inundation.

The findings confirm that GPS data exhibit clear abnormal patterns during the flood period and can effectively indicate both the timing and the approximate locations of flood impacts. These results establish the behavioral foundation for Experiments II and III, where GPS is integrated with SAR, DEM, and Twitter data for enhanced flood estimation. Thus, Experiment I serves as the initial “sensing layer” of the multi-source framework developed in this dissertation.

4.1 Objectives

This experiment aims to explore the temporal and spatial dynamics of human mobility during the 2019 Typhoon Hagibis flood in Nagano City using large-scale GPS location data. The purpose of this analysis is to examine whether population movement patterns recorded by mobile devices can serve as reliable indicators for detecting flood-affected areas and the timing of disaster impact. By quantitatively analyzing temporal variations and spatial clustering of GPS data, this study seeks to evaluate the feasibility of using social sensing data as a complementary source for disaster observation and validation.

The specific objectives of this experiment are as follows:

(a) To analyze the temporal variation of population activity before, during, and after the flood event, using GPS records aggregated in 6-hour intervals across a three-day period (October 12–14, 2019). This allows the identification of behavioral changes

corresponding to different stages of the disaster.

(b) To detect spatial anomalies in GPS data distribution within Nagano City's lowland administrative districts (below 330 m elevation), where flood inundation was most severe. By comparing GPS data volumes across time intervals, areas experiencing significant reductions in data density can be inferred as flood-affected zones or evacuation regions.

(c) To establish a grid-based spatial framework that provides statistical stability while maintaining sufficient spatial resolution. Through testing multiple grid sizes (100 m, 300 m, and 500 m), the experiment determines the optimal grid dimension—500 m—for analyzing urban-scale population movements during disasters.

(d) To validate the relationship between GPS-derived anomalies and actual flood conditions, by comparing spatial distributions of abnormal population reductions with known inundation areas along the Chikuma River. This comparison verifies the capability of GPS data to reflect real-world disaster impacts.

(e) To provide empirical evidence for subsequent model development, demonstrating that social sensing data such as GPS can effectively capture flood dynamics. The findings from this experiment form the foundational basis for the integration of SAR and GPS data in Experiment II (Chapter 5), where the flood extent estimation model is constructed and validated.

In summary, this experiment establishes the feasibility of using anonymized mobile location data to monitor and interpret disaster evolution at fine temporal scales. By revealing how population mobility correlates with flood occurrence and spatial extent, the study bridges human behavioral data with physical flood phenomena, contributing to the broader goal of developing a data-driven resilient city model.

4.2 Study Area & Data

4.2.1. Study Area

The study area of Experiment I is Nagano City, located in central Honshu, Japan (Figure 4.1). The city is situated along the Chikuma River, the longest river in Japan, which flows through the northern part of Nagano Prefecture. With a population of approximately 370,000 and a high concentration of residential, commercial, and agricultural zones distributed along the river basin, Nagano City represents a typical urban–riverine flood-prone environment.

Figure 4.1: Location of Nagano City

During Typhoon Hagibis in October 2019, Nagano experienced one of the most destructive flood events in recent years. On 12 October, heavy rainfall caused the collapse of the Chikuma River embankment near the Hoyasu district, leading to rapid inundation of extensive low-lying areas. According to the Ministry of Land, Infrastructure, Transport, and Tourism (MLIT), more than 1,400 hectares were flooded, including densely populated neighborhoods and critical transportation facilities.

To ensure analytical consistency and to capture the area's most directly affected by the disaster, this study focuses on 189 administrative districts located below 330 meters in elevation, representing the primary floodplain of the Chikuma River. These low-elevation districts correspond closely with the official inundation zone and contain the majority of observed mobility disruptions during the flood.

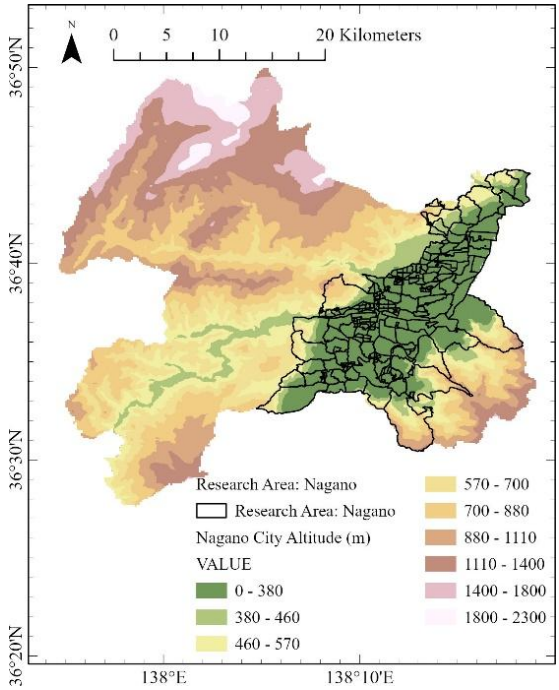


Figure 4.2: Elevation map of Nagano City highlighting the 189 administrative districts below 330 m

Figure 4.2 illustrates the elevation distribution of Nagano City and delineates the boundaries of the 189 selected administrative districts. The lowland regions below 330 m form a continuous floodplain extending along the Chikuma River, confirming that this elevation threshold effectively isolates the flood-prone area.

In addition to elevation-based district selection, administrative boundary data and official flood extent maps issued by MLIT were incorporated into the study. These datasets, shown in Figure 4.3, provide accurate delineation of inundated zones and serve as essential references for validating whether observed GPS activity anomalies align with the physical flood footprint.

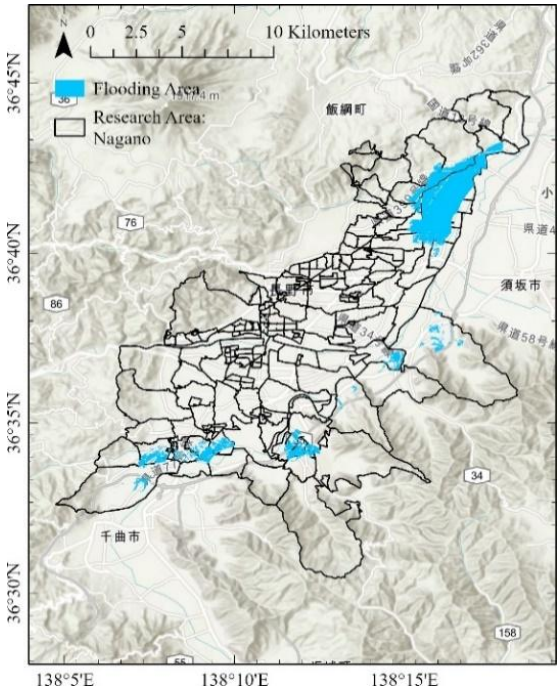


Figure 4.3: Official flood extent map of Nagano City based on MLIT post-disaster assessment

Together, these geographic and hydrological datasets establish a precise spatial framework for Experiment I, enabling a rigorous examination of the relationship between flood dynamics and population mobility as captured through large-scale GPS data.

4.2.2. Data Sources

The dataset used in this experiment consists of large-scale anonymized mobile phone GPS location data, supplemented by official environmental datasets to contextualize the flood event.

(a) GPS Location Data (Social Sensing Source)

Provider: Agoop Corporation, Japan

Data type: Anonymized mobile device GPS logs containing latitude, longitude, and timestamp information.

Observation period: October 12–14, 2019 (covering pre-flood, during-flood, and post-flood phases).

Temporal resolution: Approximately one record per user per minute under normal mobility conditions.

Spatial coverage: Entire Nagano City area, aggregated into 500 m × 500 m grid cells.

Data volume: Approximately several million GPS points across three days.

Privacy assurance: All records were anonymized and aggregated prior to analysis in accordance with Agoop's data protection guidelines.

Figure 4.4 illustrates the hourly variation of GPS data counts from 12–14 October 2019. A sharp decline is observed during the peak flooding period, which supports the use of GPS activity as an indicator of population mobility disruptions. GPS data provides a high-resolution representation of population movement and activity patterns. Because location signals are generated primarily when devices change position, notable reductions in GPS counts may reflect restricted mobility due to evacuation, road closure, or inundation. This makes GPS an effective proxy for detecting abnormal human behavior during disasters.

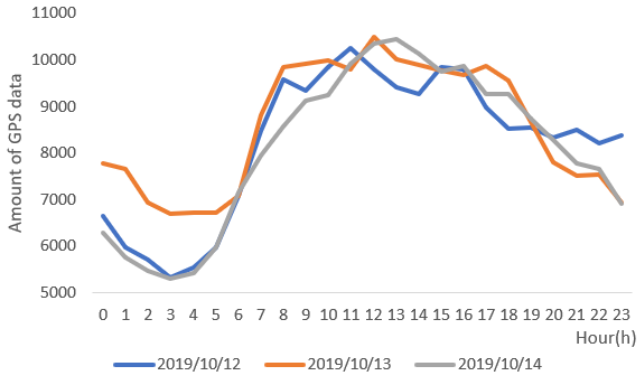


Figure 4.4: Hourly variation of GPS data counts from 12–14 October 2019

(b) Meteorological and Hydrological Data

Source: Ministry of Land, Infrastructure, Transport, and Tourism (MLIT).

Variables: Precipitation records and river water level data for Nagano City.

Observation period: October 12–14, 2019.

These data sets enable identification of the flood peak and facilitate comparison between physical flood dynamics and anomalies observed in the GPS data.

(c) Topographic and Administrative Data

Elevation data: 10 m-resolution DEM provided by the Geospatial Information Authority of Japan (GSI).

Administrative boundaries: Municipal divisions of Nagano City, used to define analysis units.

Flood inundation reference: Official post-disaster flood map published by MLIT (2019) for validation of flood-affected areas.

4.2.3. Data Preprocessing

To analyze temporal–spatial variations in population activity and to prepare the GPS dataset for subsequent multi-source integration, a series of preprocessing steps were performed as follows.

(a) Temporal Segmentation:

The three-day observation period (12–14 October 2019) was divided into twelve six-hour intervals, denoted as D_1–D_{12}. This duration provides sufficient temporal granularity

to capture sudden reductions in population movement that corresponds to the peak flooding period. The segmentation enables consistent comparison across pre-flood, during-flood, and post-flood phases. Figure 4.5 presents the spatial distribution of GPS points on 12, 13, and 14 October. Differences in data density across days highlight the need for an aggregated grid structure.

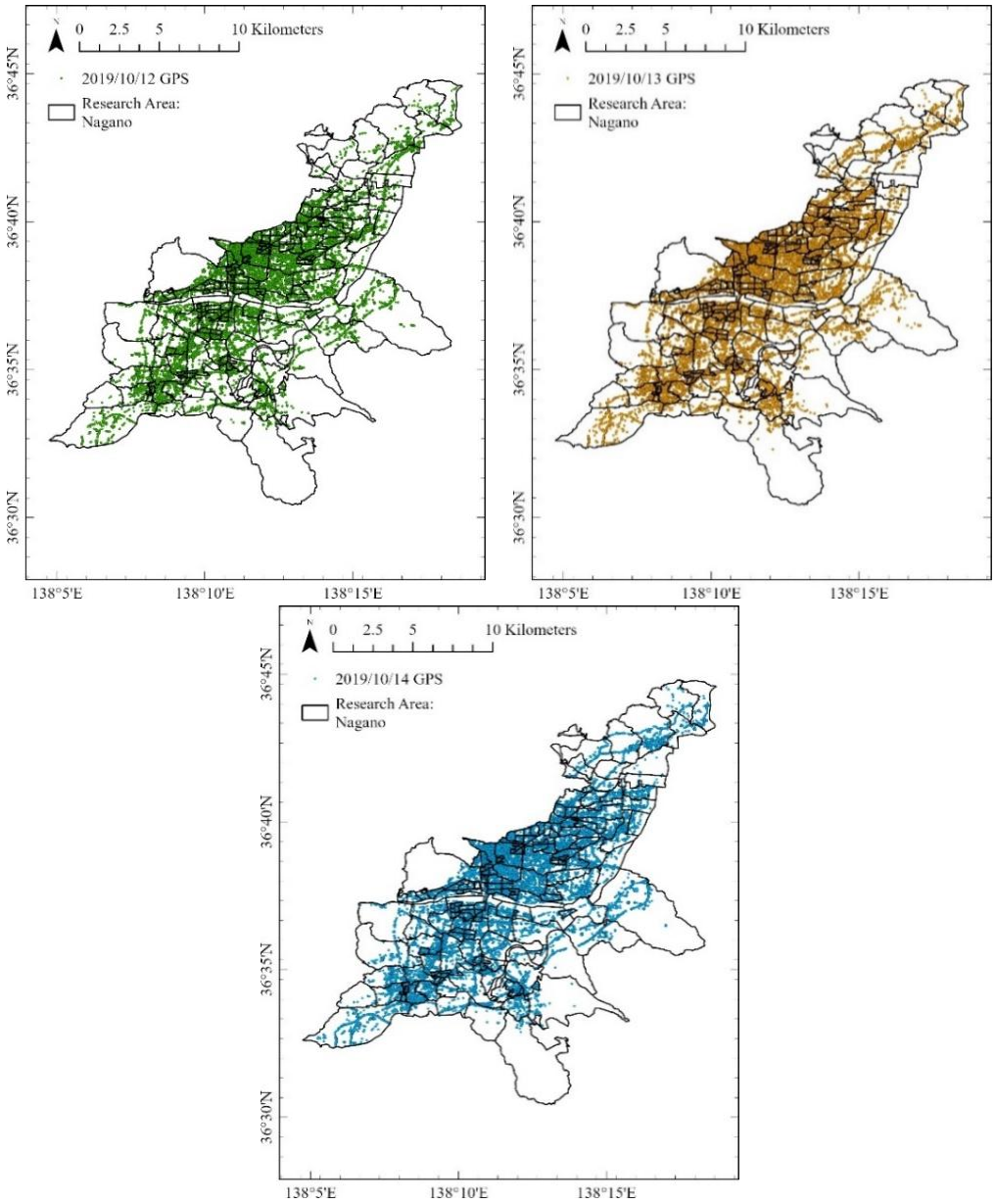


Figure 4.5: GPS data distribution map

(b) Spatial Gridding:

The study area was initially divided into grids of 100 m, 300 m, and 500 m to determine the appropriate spatial unit. Because GPS counts during the lowest-activity period (03:00–04:00 on 14 October) fell below the number of 100 m and 300 m grids, the 500 m × 500 m resolution was selected to ensure statistical stability. The final grid structure

consisted of 1,222 cells, providing a balance between spatial detail and data reliability.

(c) Data Cleaning:

Erroneous or duplicate GPS records—such as points outside the study boundary or identical consecutive timestamps—were removed. Nighttime stationary points were retained only when relevant for capturing evacuation behavior or sheltering patterns. This filtering ensured that the aggregated data accurately reflected meaningful population movement rather than noise.

(d) Data Aggregation:

GPS points were aggregated by grid cell and time interval, producing a three-dimensional spatiotemporal dataset. This data cube forms the analytical base for: identifying temporal anomalies (e.g., sudden drops in activity); mapping spatial clusters of reduced mobility; and linking human behavioral responses to the physical flood timeline.

Overall, the preprocessed GPS dataset provides high-resolution temporal and spatial information characterizing population activity before, during, and after the 2019 Nagano flood. When combined with hydrological and topographic datasets, these GPS-derived patterns enable a detailed examination of urban behavioral responses to flooding. This dataset forms the analytical foundation for validating the effectiveness of social sensing in disaster monitoring and serves as a critical precursor to the multi-source data fusion modeling conducted in Experiment II (Chapter 5).

4.3 Methods

This section describes the methodology used to analyze the temporal and spatial distribution of GPS data during the 2019 Typhoon Hagibis flood in Nagano City. The aim is to extract patterns of population movement and identify potential anomalies that correspond to the flood occurrence period. The method includes three key components: temporal segmentation, spatial grid construction, and spatiotemporal anomaly detection.

4.3.1. Temporal Segmentation

The observation period for this analysis was set from October 12 to October 14, 2019, covering the entire duration of the flood event.

To capture detailed temporal variations in population activity, the three-day period (totaling 72 hours) was divided into 12 equal time intervals, each lasting six hours.

Let

$$D = \{D_1, D_2, \dots, D_{12}\} \quad (1)$$

denote the complete GPS dataset, where each subset D_i represents the GPS data points recorded within a specific six-hour interval. Table 4.1 shows the amount of GPS data points in each classified dataset.

Table 4.1: Amount of GPS data in dataset

Dataset	Amount of GPS data	Dataset	Amount of GPS data
D_1	34538	**D_7**	60705
D_2	53717	**D_8**	48939
D_3	56220	**D_9**	34295
D_4	49757	**D_{10}**	52113
D_5	42735	**D_{11}**	59942
D_6	56446	**D_{12}**	48669

This segmentation provides sufficient temporal resolution to identify when flood-induced population changes occurred, while ensuring that each interval contains a statistically stable number of GPS points. The segmentation also aligns with the temporal progression of rainfall and river level rise reported by the Ministry of Land, Infrastructure, Transport, and Tourism (MLIT).

4.3.2. Spatial Grid Construction

To analyze spatial variations, the study area was divided into a uniform grid structure using ArcGIS Pro.

Three grid sizes—100 m, 300 m, and 500 m—were tested to determine an appropriate balance between spatial resolution and data density.

Hourly statistical analyses showed that the smallest number of GPS records occurred between 03:00–04:00 on October 14, totaling 5,331 points. This count was smaller than the number of 100 m and 300 m grid cells, resulting in excessive data sparsity. To maintain statistical reliability while avoiding over-fragmentation, the 500 m grid size was adopted as the standard spatial analysis unit.

The final study domain contained 1,222 grid cells, each serving as a spatial unit for aggregating GPS points and calculating activity intensity during each time interval.

4.3.3. Data Processing and Cleaning

The raw GPS dataset was preprocessed as follows:

(a) Noise removal: Outliers with implausible coordinates (e.g., outside Nagano City boundaries) were excluded.

(b) Temporal alignment: All timestamps were standardized to Japan Standard Time (JST, UTC+9).

(c) Duplicate filtering: Repeated records generated by static users were merged to

prevent overcounting.

(d) Aggregation: For each time interval and each 500 m grid cell, the total number of GPS data points was counted, forming a spatiotemporal data cube $N(x, y, t)$.

This cube structure allowed for simultaneous visualization of spatial patterns and their temporal evolution throughout the disaster period.

4.3.4. Data Overlay and Grid-Based Aggregation

To integrate population movement data with spatial reference units, a grid-based spatial join was performed in ArcGIS Pro. For each six-hour interval D_t, the GPS points were overlaid onto the 500 m × 500 m grid cells, and the "Join Count" was computed as the number of GPS records contained in each cell. This procedure is defined as:

$$C = \text{Spatial Join}\,(D, R) \tag{2}$$

where Drepresents the GPS dataset, Rdenotes the grid layer, and $C$is the resulting count of GPS points in each grid cell.

In addition, the official flood extent map published by the Ministry of Land, Infrastructure, Transport and Tourism (MLIT) was overlaid onto the same grid system to generate a binary flood–non-flood classification for each cell. This dataset serves as the reference for subsequent analyses of flood timing and spatial impact.

4.3.5. Overall Temporal–Spatial Analysis of Flood Impact

After aggregating GPS counts into 12 six-hour intervals, a comprehensive temporal–spatial analysis was conducted to characterize flood-related anomalies. The goal was to identify: the date and time interval during which the flood occurred, the onset and termination of mobility disturbance, and the approximate spatial extent of abnormal activity changes.

This analysis forms the basis for detailed temporal segmentation and anomaly detection in the following sections.

4.3.6. Flood Area Detection Through Grid-Based Anomaly Analysis

To determine the timing of flood occurrence, hourly GPS counts from October 12–14 was compared against normal mobility patterns observed before and after the disaster. GPS datasets from: September 29–October 5 (week before the flood), and October 20–26 (week after the flood), were treated as *normal-condition data*, allowing the construction of hourly baseline mobility profiles.

Because population activity varies through hours, weekdays, and weekends, a normal

distribution was modeled for each hourly time slot. Let x_i denote the GPS count of the i-th normal date. The normal distribution is defined as:

$$f(x \mid \mu, \sigma^2) = \frac{1}{\sqrt{2\Pi\sigma^2}} \cdot e^{-\frac{(x-\sigma)^2}{2\sigma^2}} \quad (3)$$

The hourly mean and standard deviation under normal conditions were computed as:

$$\mu = \frac{\sum_{i=1}^{N} x_i}{N} \quad (4)$$

$$\sigma = \sqrt{\frac{\sum_{i=1}^{N}(x_i - \mu)^2}{N}} \quad (5)$$

Following the 3σ rule, 99.7% of normal-condition data are expected to fall within:

$$\mu - 3\sigma \leq X_i \leq \mu + 3\sigma \quad (6)$$

The anomaly indicator is assigned as:

$$X_i^{\backslash *} = \begin{cases} 0, \mu - 3\sigma \leq Xi \leq \mu + 3\sigma \\ 1, otherwise \end{cases} \quad (7)$$

where $X_i^{\backslash *}$ is the hourly GPS count on the abnormal date (October 12–14).

This process identifies the precise hours during which population mobility deviated significantly from normal patterns, marking the temporal boundaries of the flood impact.

4.3.7. Flood Time Analysis Using Normal Distribution Modeling

To estimate the spatial extent of flood-affected zones, GPS counts for each grid cell were subjected to a normal distribution analysis similar to the temporal model.

Let $z_{i,j}$ denote the GPS count for grid j during hour i under normal conditions. The normal distribution for each grid is:

$$f(z \mid \mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(z-\mu)^2}{2\sigma^2}} \quad (8)$$

Grid-level means and standard deviations were computed as:

$$\mu = \frac{\sum_{i=1}^{N} z_j}{N} \quad (9)$$

$$\sigma = \sqrt{\frac{\sum_{i=1}^{N}\left(z_{i,j} - \mu\right)^2}{N}} \quad (10)$$

During the detected flood period (from Section 4.3.X), the total GPS count for each grid was calculated as:

$$Z_j = \sum_{i=a}^{b} Z_{i,j} \quad (a \leq i \leq b) \quad (11)$$

Flood-induced anomalies were identified by comparing Z_j with the normal-condition

distribution using:

$$Z_j^{\backslash *} = \begin{cases} 0, \mu - 3\sigma \le Z_j \le \mu + 3\sigma \\ 1, otherwise \end{cases} \quad (12)$$

Cells with $Z_j^{\backslash *} = 1$ were interpreted as areas experiencing abnormal reductions in population activity, corresponding to potential flooding.

To improve interpretability, government-designated emergency shelter locations active between October 12–14 were incorporated into the spatial analysis. Evacuation-related mobility patterns—such as increased activity near shelters or reduced movement within flooded areas—provided additional evidence supporting anomaly interpretation.

4.4 Results & Discussion

This section presents the analytical results of the temporal–spatial population dynamics during the 2019 Nagano City flood, derived from the aggregated GPS dataset. The discussion interprets the observed patterns in relation to the timing and spatial extent of the flood and examines the potential of GPS data as a proxy indicator for flood impact assessment.

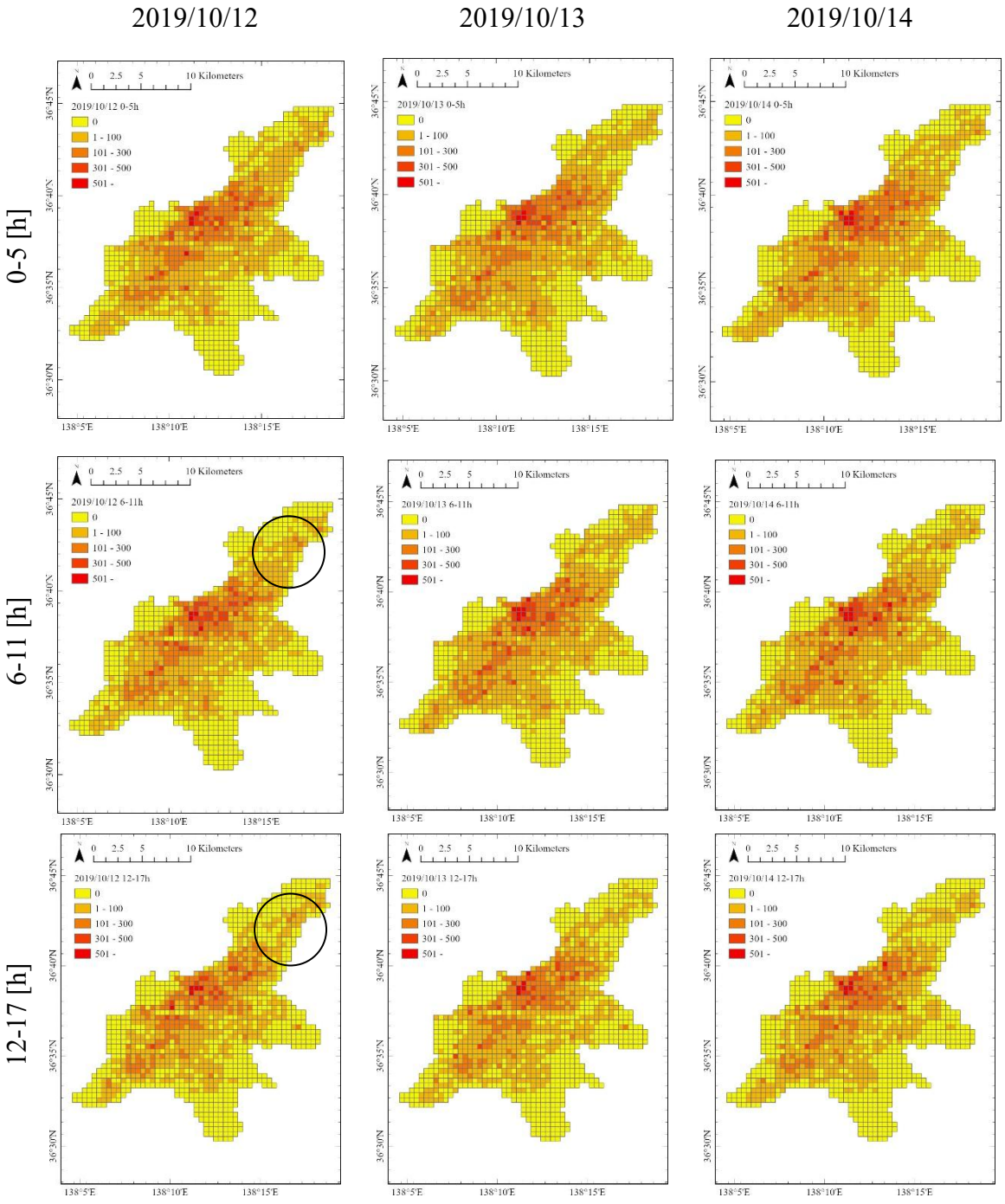
4.4.1. Overall Variation in GPS Activity

GPS data aggregated across the 12 six-hour intervals revealed distinct temporal (Figure 4.6) –spatial anomalies associated with the 2019 flood event. Although the overall GPS activity in Nagano City remained generally stable throughout the three-day observation period, localized declines were evident in areas later confirmed as inundated.

Comparing the GPS distribution between 06:00–11:00 and 12:00–17:00 on October 12 showed a pronounced reduction in GPS activity in the northeastern portion of the city. This decline persisted through October 13, reaching its lowest level during the early morning hours, and began returning toward normal conditions by October 14.

Similarly, during the 17:00–23:00 intervals of October 12–14, the lower-right portion of the study area exhibited a sharp drop in GPS counts on the 12th and 13th, followed by recovery on the 14th. This pattern is consistent with population immobility due to inundation followed by post-flood return movements.

These results—combined with the temporal distribution of GPS count statistics (Figure 4.4) and the official MLIT inundation map (Figure 4.3)—confirm that the flood induced clear anomalies in population mobility in both specific times and geographic zones. The most significant impacts occurred between October 12 and 13, which became the focus of subsequent time-specific and spatial analyses.


2019/10/12
2019/10/13
2019/10/14
0-5 [h]
6-11 [h]
12-17 [h]
0 2.5 5 10 Kilometers
2019/10/12 0-5h
2019/10/13 0-5h
2019/10/14 0-5h
2019/10/12 6-11h
2019/10/13 6-11h
2019/10/14 6-11h
2019/10/12 12-17h
2019/10/13 12-17h
2019/10/14 12-17h
0
1 - 100
101 - 300
301 - 500
501 -
36°45'N
36°40'N
36°35'N
36°30'N
138°5'E
138°10'E
138°15'E

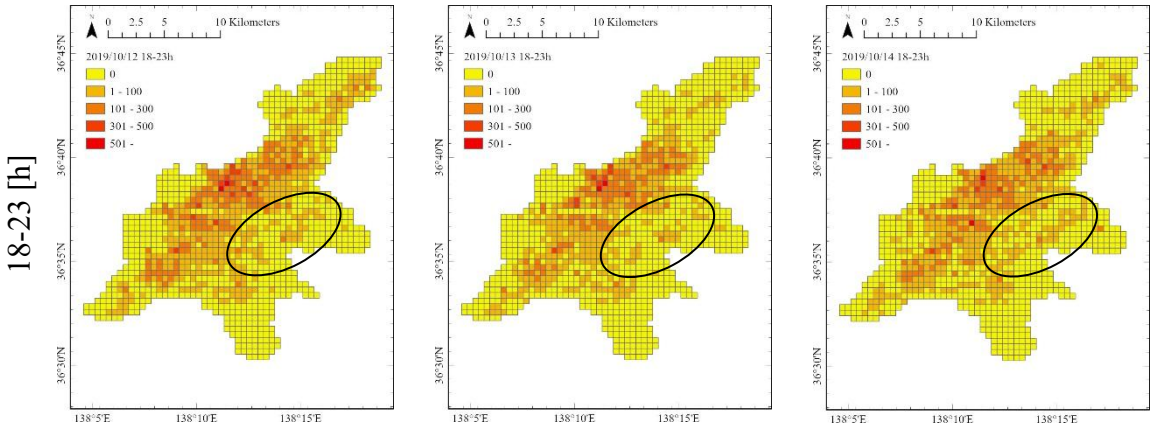


Figure 4.6: 2019/10/12-2019/10/14 GPS Data Variation Chart for Nagano City for 12 Time Intervals

4.4.2. Time Variation and Flood Timing Detection

Figure 4.7 shows the expected hourly distribution of GPS activity under normal weekend conditions, expressed as the interval $\mu \pm 3\sigma$. Two major deviations were observed:

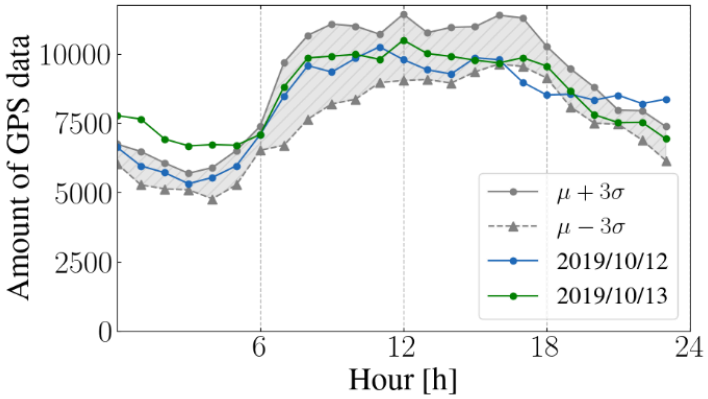


Figure 4.7: 12th and 13th GPS data per hour with normal ranges

(a) 16:00–18:30 on October 12: GPS activity was significantly below the normal range. This coincides with the 38.5 mm rainfall peak, suggesting suppressed population movement due to heavy precipitation.

(b)22:30 on October 12 to 05:00 on October 13: GPS activity sharply exceeded normal levels, with the largest deviation (>1000 points) at 01:00. This corresponds closely to the rapid rise of the Chikuma River water level, which began increasing around 17:00 and peaked at 02:00. The surge in GPS activity reflects nighttime evacuations, displacement movements, or emergency responses.

The strong temporal alignment between abnormal GPS behavior and hydrological indicators confirms that GPS mobility data is sensitive to flood dynamics. Based on this analysis, the interval 22:30–05:00 was identified as the principal flood impact period used

for spatial anomaly detection.

4.4.3. Spatial Variation and Flood-Affected Areas

Using the anomaly detection model described in Section 4.3, grid-based spatial analysis identified 175 grid cells with abnormal GPS activity during the flood period. These cells were classified into 52 isolated anomalous grids (30%), and 123 adjacent anomalous grids (70%), forming coherent spatial clusters.

Figure 4.8 shows that most clustered anomalies were concentrated in low-lying districts along the Chikuma River, including Hoyasu, Akimoto, and Tsuchiura—areas verified as flooded on MLIT maps.

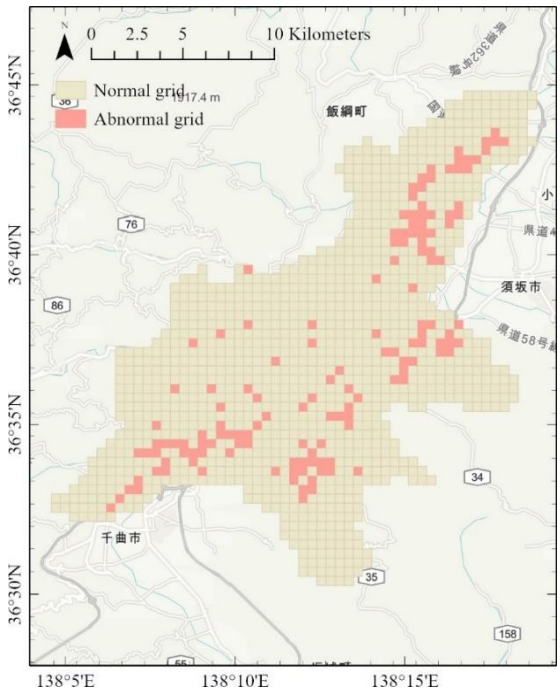


Figure 4.8: Results and distribution of abnormal grids

When overlaying the anomalous grids with flood extent, river networks, and shelters activated between October 12–14 (Figure 4.9), several patterns emerged: Adjacent anomalous grids strongly overlapped or bordered official flood zones. Many anomalous areas were located along major evacuation routes or near designated shelters. Movement anomalies were especially concentrated at the transition boundary between flooded and non-flooded regions, reflecting evacuation flows, road closures, and shelter-seeking behavior.

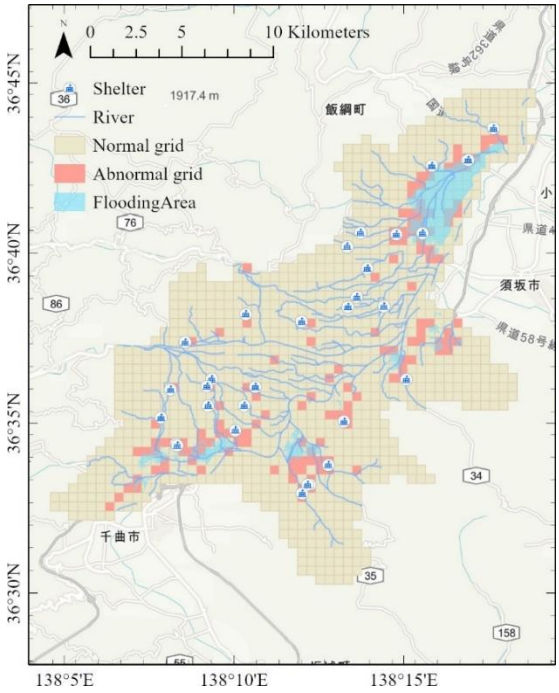


Figure 4.9: Overlay analysis of abnormal grids

These findings demonstrate that GPS-based anomaly detection effectively delineates flood-impacted areas, both in terms of direct inundation and broader mobility disruptions.

4.4.4. Integrated Interpretation

The results confirm that large-scale GPS mobility data is a reliable indicator of flood timing, spatial extent, and human behavioral response to disaster conditions. Temporal anomalies corresponded closely with rainfall intensity and river water level changes, while spatial anomalies aligned strongly with official inundation maps and known evacuation patterns.

Key findings include:

(a) Temporal sensitivity: GPS activity captured the onset, peak, and decline of flood impacts with high precision.

(b) Spatial accuracy: Clustered GPS anomalies overlapped flooded areas with strong consistency.

(c) Behavioral insights: Increased nighttime GPS activity revealed evacuation and emergency mobility patterns that are not observable in physical sensors alone.

These results validate the potential of population mobility data as a social-sensing tool for disaster detection and provide a foundational component for the multi-source RF modeling developed in Chapter 5.

4.5 Implications for Multi-Source Flood Modeling

The findings of Experiment I demonstrate that large-scale GPS mobility data provides robust temporal and spatial indicators of flood impact, revealing both the timing of the hazard and the geographic extent of population disruption. Through anomaly detection, grid-based aggregation, and validation against official inundation records, the experiment confirmed that social sensing can effectively capture disaster dynamics even in the absence of physical observations. This outcome establishes an essential foundation for the multi-source integration framework developed in the subsequent chapters.

More specifically, three insights derived from Chapter 4 directly inform the design and modeling strategies of Chapter 5:

(a) Temporal sensitivity of mobility data

The precise identification of flood onset and peak hours confirms that human behavioral responses exhibit rapid, measurable changes during hazard events. This supports the integration of GPS-derived indicators into the feature space of machine learning models for flood extent estimation.

(b) Spatial correspondence with inundation patterns

The strong spatial overlap between anomalous GPS grids and MLIT-documented flood zones provides empirical evidence that population movement patterns reflect ground conditions. This justifies the incorporation of GPS density as a complementary layer to SAR backscatter and terrain features when constructing Random Forest models in Chapter 5.

(c) Behavioral–physical linkages

The alignment of mobility anomalies with rainfall intensity, river water level, and shelter locations demonstrates that social sensing captures both direct impacts (inundation) and indirect responses (evacuation, avoidance behavior). This dual sensitivity enhances the potential of GPS data to improve the detection of flood boundaries, especially in areas where SAR may be limited by speckle noise, urban geometry, or mixed land cover.

Building on these insights, Chapter 5 advances from anomaly identification to quantitative flood-extent estimation, moving from descriptive social sensing analysis to a structured machine-learning framework. In Chapter 5, Random Forest models are constructed to evaluate how GPS-derived features enhance flood-extent estimation when combined with multi-orbit SAR backscatter and terrain indices. This progression—from detecting behavioral anomalies (Chapter 4) to estimating physical inundation patterns (Chapter 5)—marks a critical step in developing an integrated, multi-source approach to urban disaster assessment.

Thus, Experiment I serves not only as an independent validation of social sensing for disaster monitoring but also as a conceptual and empirical foundation for the multi-source flood estimation models developed in Chapter 5.

Chapter 5 Experiment II: RF for Flood-Extent Estimation with Multi-Source Inputs

This chapter is primarily derived from the following conference and journal manuscripts authored by the candidate: Yifan Yang, Naoki Ohira and Hideomi Gokon, “Model Construction and Evaluation of Flood Area Estimation Based on SAR and GPS Data,” IGARSS 2024, Athens, Greece. Yifan Yang, Hideomi Gokon and Wataru Takeuchi, “Flood Extent Prediction Using Remote Sensing Data Based on Machine Learning and Dempster-Shafer Theory: A Case Study of Nagano City,” IGARSS 2025, Brisbane, Australia. The content has been substantially revised, expanded, and integrated to form the second stage of the dissertation’s multi-source estimation framework.

This chapter presents the second stage of the dissertation’s multi-source disaster estimation framework by developing Random Forest (RF) models for flood-extent estimation in Nagano City. Building on the findings of Chapter 4—which demonstrated that GPS mobility anomalies reliably reflect the temporal and spatial progression of the 2019 flood—this chapter examines how behavioral and physical sensing data can be systematically integrated into a supervised machine-learning model.

Experiment II consists of two complementary RF-based analyses. The first evaluates the contribution of social sensing by comparing model performance with and without GPS population-density features. The second assesses the value of physical remote-sensing inputs by integrating Sentinel-1 ascending and descending orbit SAR backscatter with topographic variables such as elevation, slope, TWI, and SPI. Together, these two analyses represent the core of the “integration” stage in the Sensing–Integration–Application framework established in Chapter 3.

The RF models developed in this chapter operate at the urban scale and use a unified spatial grid to ensure comparability across data sources. Model accuracy, feature importance, and spatial reliability are evaluated against the official inundation map published by MLIT. The combined results demonstrate how multi-source integration enhances flood-extent estimation beyond what can be achieved using physical or behavioral data alone.

By clarifying the respective roles of GPS and SAR in machine-learning-based flood modeling, Chapter 5 forms a methodological bridge between the coarse-grid behavioral analysis of Chapter 4 and the fine-grid, multi-modal urban estimation model presented in Chapter 6.

5.1 Objectives and Study Design

Building upon the findings of Experiment I (Chapter 4), which demonstrated the feasibility of detecting flood-induced mobility anomalies using GPS-based social sensing, Chapter 5 advances the analysis toward quantitative flood-extent estimation through multi-source data integration. The overarching purpose of this chapter is to evaluate how the combination of remote sensing and social sensing within a Random Forest (RF) framework enhances the accuracy, robustness, and spatial coherence of flood mapping in urban environments. Whereas Experiment I focused on behavioral indicators of disaster impact, the experiments in this chapter directly examine how physical and behavioral datasets contribute to classification performance when fused into a machine-learning model.

Accordingly, the objectives of this chapter are fourfold. First, to develop RF-based flood-extent estimation models that incorporate diverse geospatial indicators, including SAR backscatter, DEM-derived terrain features, and GPS population density. This allows the construction of a unified machine-learning framework reflecting both environmental conditions and human behavioral responses during the 2019 Nagano flood.

Second, to compare two distinct multi-source integration schemes: Experiment II-A, which evaluates the incremental value of social sensing by comparing RF models with and without GPS population density; and Experiment II-B, which examines the contribution of multi-orbit SAR (ascending + descending) for improving physical-feature representation in complex urban landscapes.

Third, to quantify how different data sources—physical or behavioral—reduce misclassification, improve delineation of flood boundaries, and enhance the spatial stability of estimations. These comparisons clarify the respective roles of remote sensing and social sensing within urban disaster modeling.

Finally, to validate the estimated flood extents against the official inundation map of the Ministry of Land, Infrastructure, Transport and Tourism (MLIT), thereby assessing the external validity and practical applicability of the proposed RF models.

The study design for this chapter follows a structured multi-stage workflow, illustrated in Figure 5.1, and shared across both Experiment II-A and Experiment II-B.

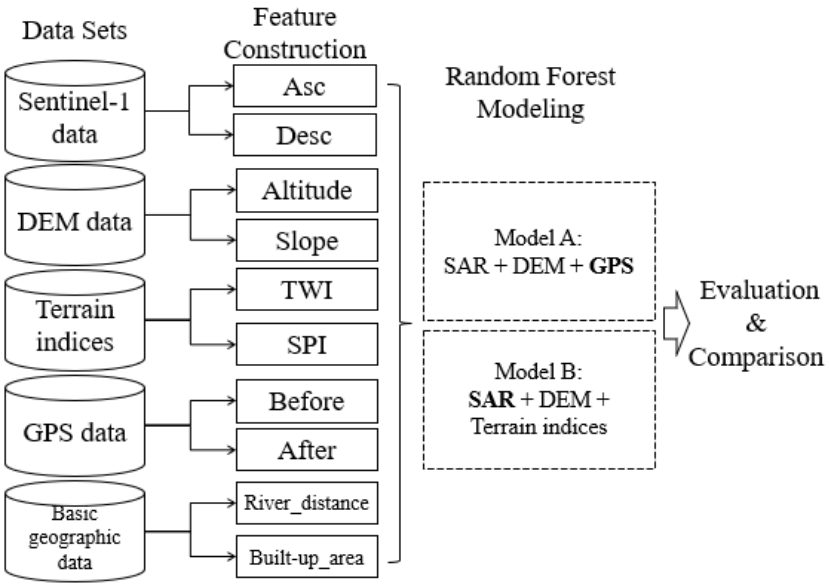


Figure 5.1: Overall workflow of Chapter 5

(a) Data Preparation: Multi-source datasets—including Sentinel-1 SAR (ascending and descending), DEM, terrain indices (TWI, SPI), distance-related variables, and GPS density—are acquired and preprocessed through radiometric calibration, terrain correction, gridding, and normalization.

(b) Feature Construction: Physical and behavioral variables are resampled to a unified spatial grid and combined into feature vectors for each pixel or grid cell.

(c) Model Construction: Two Random Forest models are trained: one incorporating GPS social sensing (Experiment II-A), and one incorporating multi-orbit SAR without GPS (Experiment II-B). Parameter tuning is conducted using grid search.

(d) Evaluation and Comparison: The models are evaluated using accuracy, confusion matrix components, and spatial overlay with MLIT flood maps. Comparative analyses highlight the relative contributions of social sensing and multi-orbit SAR.

(e) Synthesis: Findings are integrated to determine how different data sources enhance model performance and how multi-source fusion contributes to the construction of a resilient and data-driven urban flood assessment framework.

Through these two experiments, Chapter 5 serves as a methodological bridge between behavioral anomaly detection (Chapter 4) and fine-resolution resilient urban modeling (Chapter 6), demonstrating how multi-source data integration can progressively enhance flood estimation accuracy from human mobility patterns to fully spatialized inundation mapping.

5.2 Experiment II-A: RF Flood-Extent Estimation with

GPS-Integrated Multi-Source Inputs

5.2.1. Objectives

The objective of Experiment II-A is to evaluate the contribution of GPS-based social sensing data to Random Forest (RF) flood-extent estimation. While SAR backscatter and topographic features provide physical evidence of inundation, they cannot capture the behavioral disruptions that occur during flood events. Building on the findings of Experiment I, which demonstrated that population mobility anomalies reliably correspond to flood-affected areas, this experiment investigates whether integrating GPS population density into an RF model can improve flood detection accuracy, particularly in residential and urban districts where physical flood signals may be weak or ambiguous.

Specifically, this experiment aims to:

(a) Quantify the value added by social sensing by comparing a physical-only RF model with a multi-source RF model incorporating GPS population density.

(b) Evaluate whether GPS-derived behavioral information reduces omission errors, especially in flood-affected neighborhoods where SAR backscatter alone may not detect water.

(c) Assess the contribution of GPS to spatial coherence and boundary definition in flood mapping, thereby validating the complementary role of social sensing in remote sensing–based disaster monitoring.

Through this analysis, Experiment II-A positions GPS data not as a replacement for physical indicators but as a critical complementary layer that captures human behavioral responses during urban flood emergencies.

5.2.2. Study Area

This experiment focuses on the urban floodplain of Nagano City, Japan, affected severely by Typhoon Hagibis in October 2019. The area shown in Figure 5.2 is located along the Chikuma River and consists of densely populated residential districts, agricultural land, and transportation corridors. Following the same spatial definition used in Experiment I, grid cells below 330 m elevation were extracted as the primary analytical domain to ensure consistency between GPS-based mobility anomaly detection and multi-source flood-mapping experiments.

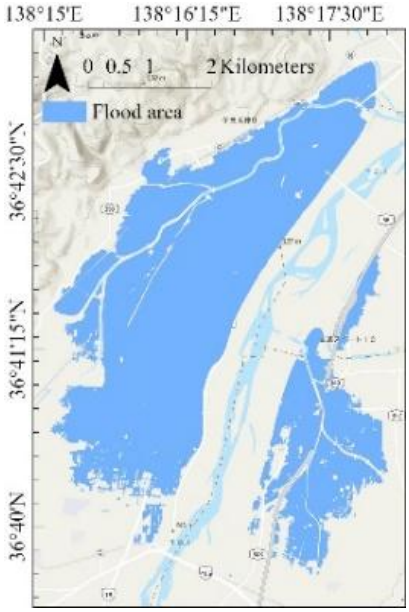


Figure 5.2: Study area map of Nagano City & MLIT inundation map

The Ministry of Land, Infrastructure, Transport and Tourism (MLIT) provides the official inundation map used as the reference dataset for model validation.

This experiment integrates multiple geospatial datasets, covering both physical remote sensing observations and human activity–based social sensing data. All datasets were preprocessed and harmonized into a common spatial reference and grid system before modeling.

5.2.3. Study Data Sets

This experiment integrates multiple data sources, combining remote sensing and social sensing to build a flood-extent estimation model.

(a) Sentinel-1 data

Sentinel-1 data with a resolution of 5×20 m were acquired using the S1B satellite's high-resolution Ground Range Detection (GRDH), interpolated Wide Interferometric (IW) mode, Single View Complexity (SLC) detection, and validated processing level VV polarization. The S1B satellite's acquisition period was 12 days, and the pre-flood data of 5 October 2019 and the post-flood data of 17 October 2019 were selected. As shown in Figure 5.3, the backscattering coefficient for each pixel was obtained after pre-processing. As shown in Figure 5.4, the backscattering coefficient is used to determine the area of the water, and areas with low values are the area of the water.

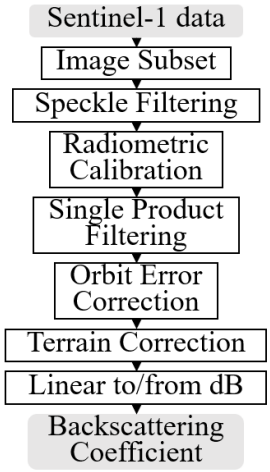
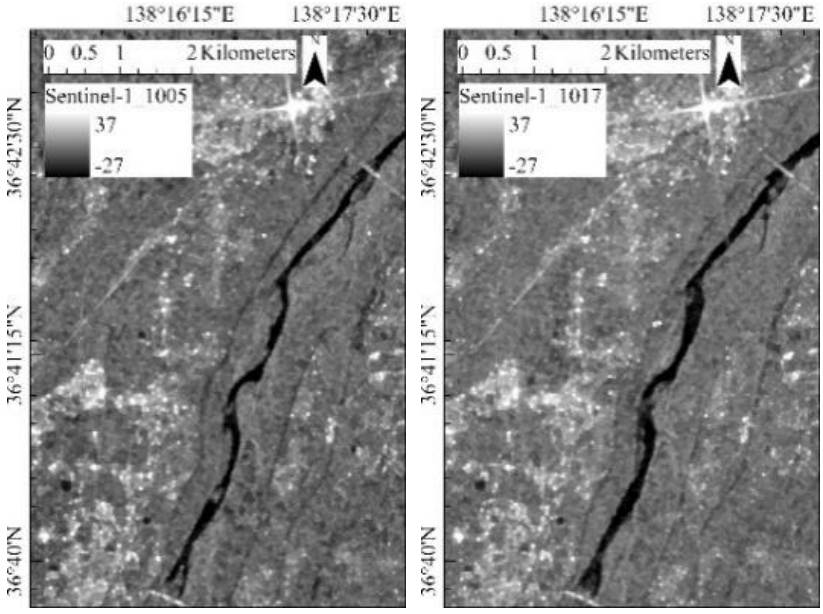


Figure 5.3: Sentinel-1 data preprocessing flow



(a) pre-flood (b) post-flood

Figure 5.4: Sentinel-1 data backscattering coefficient distribution

(b) DEM data

Digital Elevation Model (DEM) data provides a detailed description of surface elevation that can be analyzed to obtain key topographic features. DEM data with a resolution of 10m*10m was selected. Detailed elevation and slope information was obtained by calculating the study area as shown in Figure 5.5 and Figure 5.6.

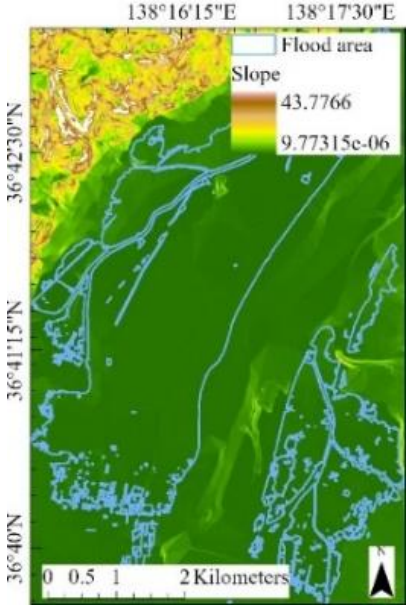


Figure 5.5: Slope distribution

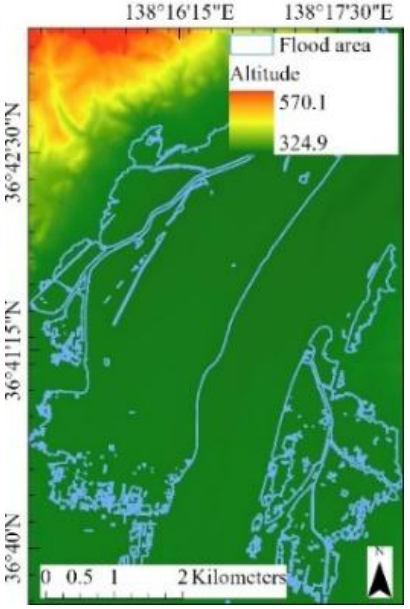
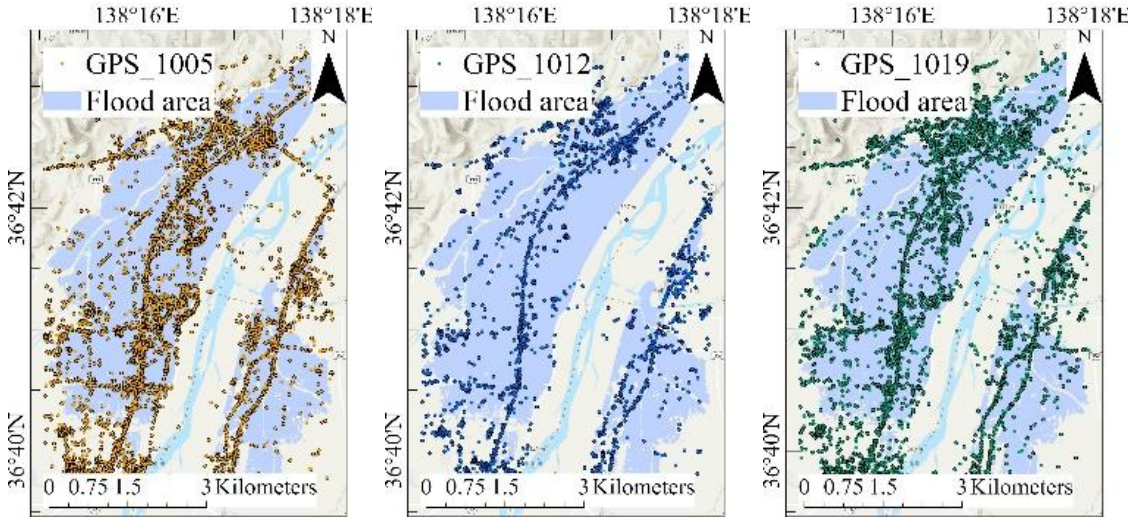


Figure 5.6: Altitude distribution

(c) GPS data

Mobile phone GPS location data provided by Agoop coop.. This data includes time, latitude and longitude and other data that records the location of mobile phone users at different times. The flooding occurred on the 12th and 13th of October 2019, which were weekends. As the population movement patterns are different between weekends and weekdays, GPS data from the weekend before the flood (October 5th and 6th), during the flood (October 12th and 13th) and after the flood (October 19th and 20th) are used to accurately estimate the area of the flood. Figure 5.7 shows the distribution of GPS data before (October 5th), during (October 12th) and after (October 19th) the flood.



(a) October 5th (b) October 12th (c) October 19th

Figure 5.7: GPS data distribution

(d) Basic geographic data

Basic geographic data related to the flooded areas, including administrative boundary data, land use data, river and flood data, were downloaded from the website of the Ministry of Land, Infrastructure, Transport and Tourism (MLIT) of Japan.

5.2.4. Methodology

In this experiment, two Random Forest (RF) models were developed to estimate flood extent at the grid-cell level in Nagano City. Prior to modeling, all datasets were harmonized to a uniform spatial grid of 70 m × 70 m to ensure consistent feature resolution across inputs. The Feature Construction process included the following key steps:

(a) Grid search parameter tuning

To optimize the performance of the Random Forest model, a grid search method is used to determine the parameters. The grid search method selects the best combination of parameters that maximizes the performance of the model by searching through a predefined grid of parameters. The parameters include the tree splitting criterion ($Criterion$), the number of trees ($n_{estimators}$), the maximum depth of the tree (max_{depth})

and the seed value of the tree ($random_{state}$).

(b) Model construction

After determining the optimal parameters, two flood area estimation models were constructed using the Random Forest algorithm, one with GPS data and the other without GPS data. The models were built using the 'RandomForestClassifier' implementation in the Scikit-Learn library. 80% of the data is used to train the model and 20% of the data is used for testing. The model training method is as follows:

$$Model_{with_{GPS}} = RandomForestClassifier\left(n_{estimators}, max_{depth}\right) \quad (13)$$

$$Model_{without_{GPS}} = RandomForestClassifier\left(n_{estimators}, max_{depth}\right) \quad (14)$$

$Model_{with\_GPS}$ and $Model_{without\_GPS}$ refer to the model with and without GPS data respectively.

(c) Importance of features

Using the Gini importance approach, the cumulative effect of each feature in reducing model impurity was examined to determine the relative importance of each feature in the two models. As shown in Figure 5.8, higher GINI importance values indicate that the corresponding feature makes a more significant contribution to the classification performance of the model.

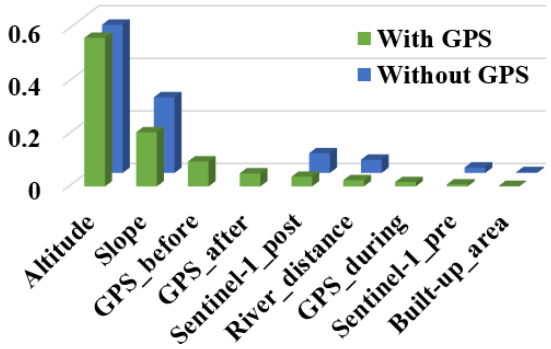


Figure 5.8: Results of the importance of features

(d) Model accuracy analysis

The two models were applied separately to the test data to obtain the confusion matrix shown in Table 5.1.

Table 5.1: Confusion matrix explanation

	Predicted Positive	Predicted Negative
Actual Positive	True Positive (TP)	False Negative (FN)
Actual Negative	False Positive (FP)	True Negative (TN)

The accuracy, precision, recall and F1 score of the model were calculated based on the confusion matrix to determine the accuracy of the model. The calculation formula and results are shown in Table 5.2.

Table 5.2: Model accuracy formula and results

	Formula	With GPS	Without GPS
Accuracy	$\frac{TP+TN}{TP+TN+PF+FN}$	0.815	0.799
Precision	$\frac{TP}{TP+FP}$	0.750	0.722
Recall	$\frac{TP}{TP+FP}$	0.870	0.885
F1 Score	$\frac{2\times Precision\times Recall}{Precision+Recall}$	0.806	0.796

5.2.5. Results and Discussion

(a) General analysis of the model results

According to Table 5.2, the accuracy of the flood area estimation model with GPS data is 0.815, and the accuracy of the model without GPS data is 0.799, with a difference of about 0.016. Although the difference is very small, it shows that GPS data is useful for flood area estimation.

Looking closely at the percentage of results of the two models for flood area estimation, Figure 5.9 shows that the model with GPS data performs better in FP estimation with a difference of 0.021 compared to the model without GPS data. However, the model with GPS data has a lower FN estimation of 0.007 compared to the model without GPS data. This is evidence that the use of GPS data as an explanatory variable has a positive effect on the accuracy of the model.

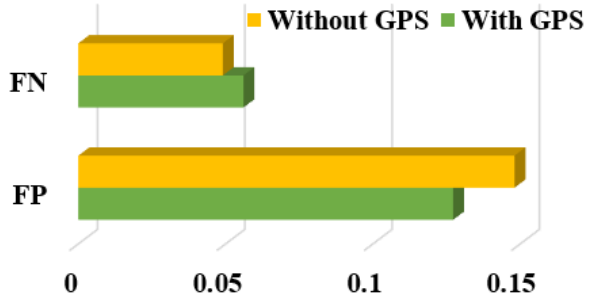


Figure 5.9: Model FN data and FP data statistics

(b) Analysis of the impact of GPS data

The flood areas estimated by the two models without and with GPS data are shown in Figure 5.10 and Figure 5.11 respectively. Compared with Figure 5.2, it can be seen that the two models estimate most of the area in the middle of the embankment on both sides of the Chikuma River as the area where flooding occurs. According to Fig. 11, the model that includes GPS data has a higher accuracy in judging this area than the model that does not include GPS data.

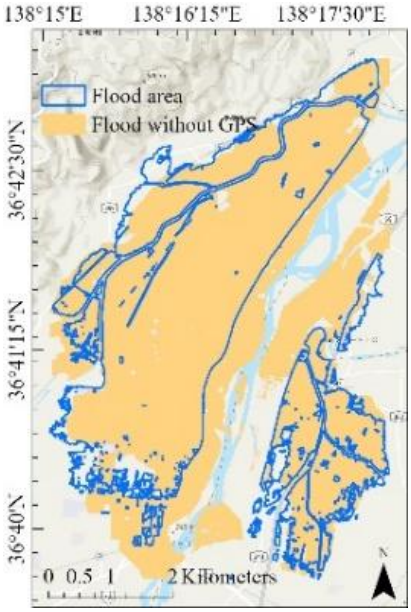


Figure 5.10: Model estimation results without GPS data

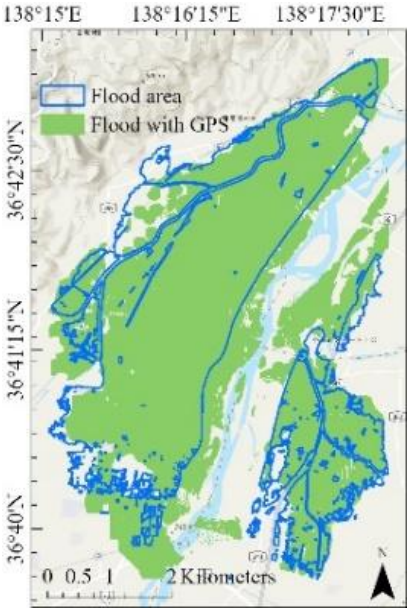


Figure 5.11: Model estimation results with GPS data

This result is mainly due to the high importance of elevation and slope in the estimation model, and the elevation and slope in this area are higher than the surrounding area. The use of GPS data slightly improves the accuracy of this area of the model, showing that GPS data played a positive role in improving the accuracy of the estimation area. However, it is also necessary to optimize the processing of GPS data.

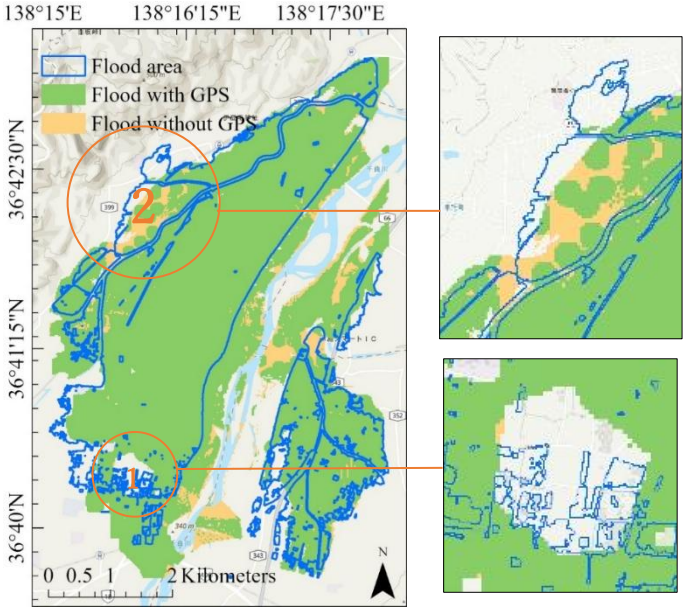


Figure 5.12: Flood area estimated by models

Area 2 in Figure 5.12 was also not estimated as a flood area due to its elevation and slope. However, the model using GPS data has a high error rate in this area. This area is

agricultural land in the unbuilt area and was determined to be a non-flood area due to the low amount of GPS data in this area and the lack of GPS data records for the non-flood area. This suggests that the model using GPS data is more suitable for estimating areas with a certain amount of GPS data, particularly built-up areas where people live or are active. This finding suggests that when using GPS data in models, data adequacy and regional characteristics need to be considered to further improve model accuracy.

(c) Model Generalization

The data used in this experiment, are generalized SAR data, GPS data, DEM data, etc., which makes the proposed flood area estimation model generalizable. However, the performance of the model may vary in different geographical areas or flood events. Therefore, in practical applications, the applicability of the model should be considered, and the appropriate data should be selected according to the specific situation.

In this experiment,, only generic SAR data, GPS data, DEM data, and basic geographic data were used for modelling flood area estimation. By comparing the two models, one with and one without GPS data, it was found that GPS data is useful for improving the accuracy of flood area estimation. Future research can explore more correlations between SAR and GPS data with flooding, optimize model parameters and use more data sources to improve the accuracy and usefulness of flood area estimation. This is important for improving flood disaster monitoring and response capabilities.

5.3 Experiment II-B: RF Flood-Extent Estimation Using Multi-Orbit SAR and Terrain Indices

5.3.1. Objectives

The objective of Experiment II-B is to evaluate the effectiveness of using multi-orbit Sentinel-1 SAR data (ascending + descending) as the primary remote sensing input for Random Forest–based flood extent estimation.

Unlike Experiment II-A, which examined the contribution of GPS-based social sensing, this experiment focuses exclusively on physical remote sensing variables and investigates whether combining different viewing geometries from S1A (descending) and S1B (ascending) can enhance the detection of floodwater in urban areas.

Because ascending and descending orbits observe the land surface from different azimuth directions, their σ^0 backscatter responses vary depending on surface roughness, building orientation, vegetation structure, and water inundation. By integrating both orbits into a unified RF model, this experiment aims to determine whether dual-orbit SAR inputs can improve spatial completeness and reduce misclassification in complex urban environments.

5.3.2. Study Area and Data

The study area is the urban area of Nagano City, Japan, as shown in Figure 5.13. The flood disaster was caused by the collapse of the banks of the Chikuma River, so the study area is defined as the area within 3 kilometers on both sides of the river. To eliminate potential confounding factors, the area within the riverbanks was excluded, and the actual flood extent within the study area is approximately 26.1 % of the total study area.

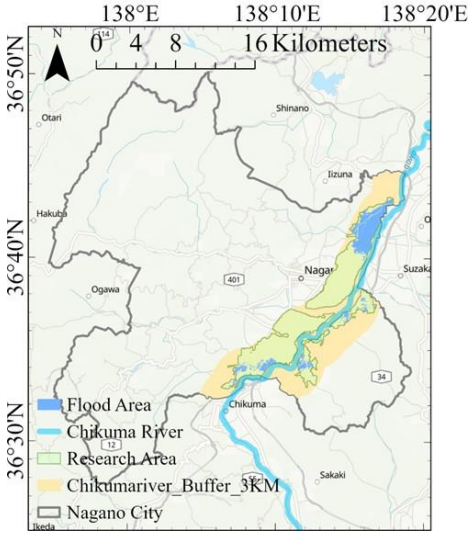


Figure 5.13: Study Area

(a) Topographic Data

The topographic data was obtained from the Digital Elevation Model (DEM) and calculated to obtain Elevation (Figure 5.14) and Slope (Figure 5.15) data.

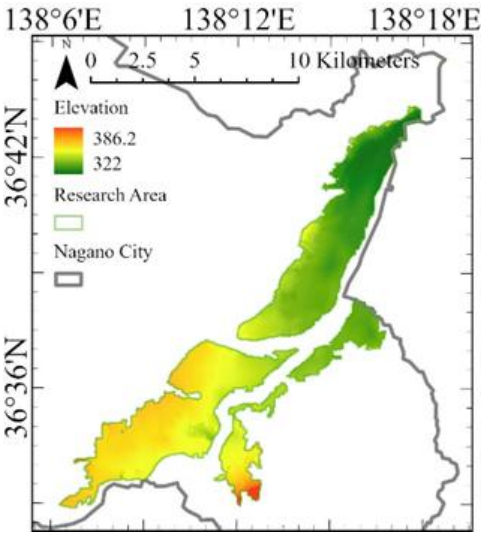


Figure 5.14: Elevation Data

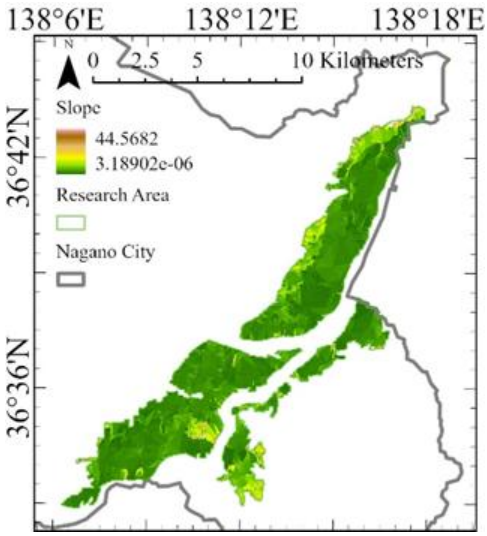


Figure 5.15: Slope Data

(b) Sentinel-1 Data

The Sentinel-1 data used in this study consists of IW-mode satellite imagery with a

resolution of 5×20 meters (GRDH). This study utilizes descending orbit data from the S1A satellite on October 11, 2019, and ascending orbit data from the S1B satellite on October 12, 2019. After preprocessing this data, the backscatter coefficient was extracted for each pixel, as shown in Figure 5.16 and Figure 5.17 Areas with low backscatter coefficients in the water surface areas.

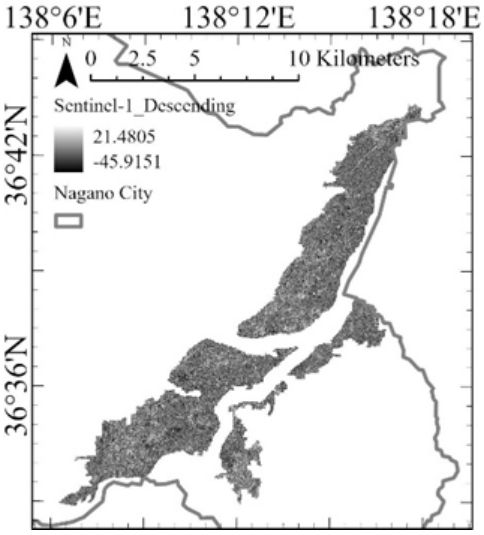


Figure 5.16: Sentinel-1 Descending Orbit Data

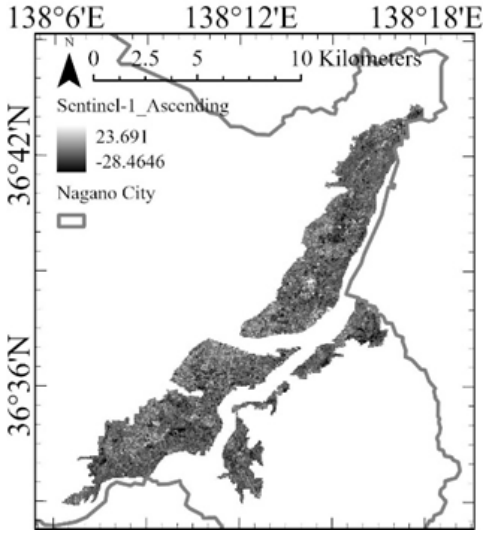


Figure 5.17: Sentinel-1 Ascending Orbit Data

(c) Terrain Indices Data

Based on the DEM data, the Topographic Wetness Index (TWI) and the Stream Power Index (SPI) were calculated, as shown in Figure 5.18 and Figures 5.19. TWI is an index that reflects the potential for surface water accumulation and can reveal moisture accumulation and drainage conditions within the region. The SPI is used to evaluate the hydraulic energy of river areas by measuring the kinetic energy of water flow within the region. These indices can indicate the potential for flooding. The formulas used to calculate TWI and SPI are as follows:

$$TWI = \ln\left(\frac{FA}{\tan(Slope)}\right) \tag{15}$$

$$SPI = \frac{FA \times \left(\frac{\sin Slope}{100}\right)}{1 + FA} \tag{16}$$

where FA is the flow accumulation as calculated by ArcGIS and Slope is the slope as calculated above.

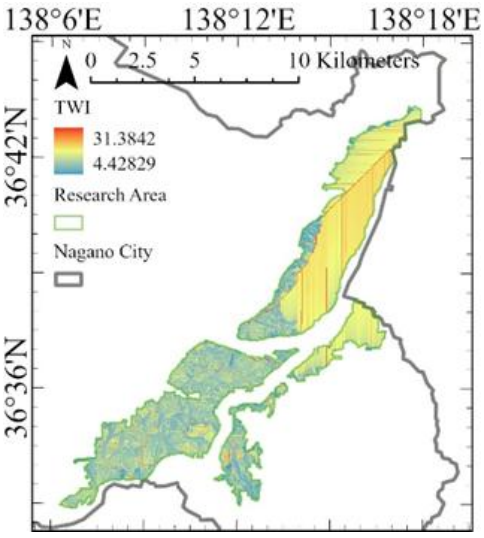


Figure 5.18: TWI Value Data

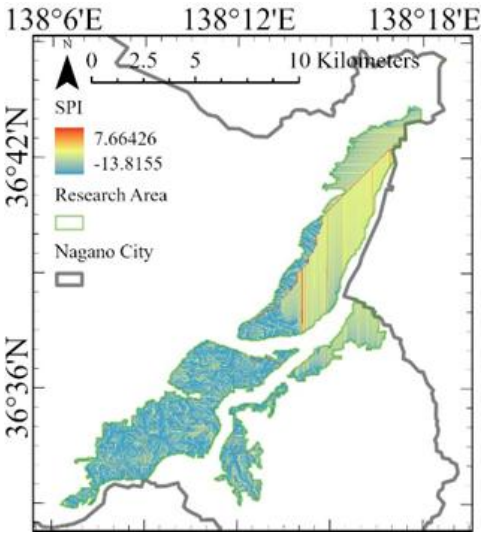


Figure 5.19: SPI Value Data

(d) Distance Data: Considering the spatial distance characteristics of flood occurrence, the distance from each location to the Chikuma River and major roads were calculated as indicators.

5.3.3. Methodology

Based on the multi-orbit SAR and terrain-related datasets introduced in Section 5.3.2, this experiment develops a Random Forest (RF) classification model to estimate flood extent within the urban area of Nagano City. The methodology consists of integrating ascending and descending Sentinel-1 backscatter coefficients with DEM-derived hydrological indicators and distance features, followed by supervised model training and accuracy evaluation.

(a) Feature Construction

All geospatial variables were resampled into a common spatial grid to ensure pixel-level alignment across datasets. The final feature set included: σ^0 backscatter coefficients from descending (S1A) and ascending (S1B) orbits; Elevation and Slope derived from DEM; TWI and SPI indices representing moisture accumulation and stream power; Distance to the Chikuma River and distance to major roads, capturing spatial proximity to likely inundation areas.

The use of both ascending and descending SAR observations provides complementary viewing geometries with different incidence angles and look directions. This multi-angle configuration helps reduce geometric distortions such as layover and shadowing, which are common in urban and riverine environments, and improves the robustness and spatial consistency of surface water detection compared with single-orbit observations. By integrating multi-orbit backscatter features with terrain and proximity variables, the feature set captures both the physical inundation characteristics and the underlying topographic controls on flood extent.

Each pixel was linked to its corresponding ground-truth label (flooded / non-flooded) using the MLIT official inundation map. The use of multi-orbit SAR enables complementary observation angles, reducing uncertainty caused by layover or shadowing effects that affect single-orbit flood detection.

(b) Random Forest Model Construction

A Random Forest classifier was implemented using Python's scikit-learn library. The model aggregates estimations from multiple decision trees to generate a robust classification. Following established procedures, 80% of the labeled dataset was randomly selected for training, while the remaining 20% was used for validation. Model parameters such as the number of trees, maximum tree depth, and splitting criteria were optimized using a grid search strategy.

The RF estimation for each pixel X is calculated through ensemble averaging:

$$RF(X) = \frac{1}{n}\sum_{i=1}^{n} T_i(X) \tag{17}$$

where $T_i(X)$denotes the estimation produced by the i^{th}decision tree and $n$is the total number of trees.

(c) Model Evaluation

The classification results were evaluated using the confusion matrix, consisting of True Positives (TP), False Positives (FP), False Negatives (FN), and True Negatives (TN). Overall accuracy was calculated using:

$$Accuracy = \frac{TP + TN}{TP + TN + PF + FN} \tag{18}$$

This accuracy assessment is supplemented by visual comparisons between predicted flood maps and MLIT inundation data, enabling a spatial evaluation of model reliability.

5.3.4. Results and Discussion

The Random Forest (RF) model was applied to the multi-source dataset consisting of ascending and descending Sentinel-1 backscatter coefficients, terrain attributes, hydrological indices, and distance features described in Section 5.3.2. The following subsections present the spatial estimation results, model accuracy, and the interpretation of classification performance.

(a) Flood Extent Estimation Results

The spatial distribution of the RF-predicted flood extent is shown in Figure 5.20, where each pixel is classified as either flooded or non-flooded. Overall, the model successfully delineates the major inundation zones along the Chikuma River that were recorded in the MLIT official flood map. In particular, the large flooded area in the eastern portion of the study area is clearly identified, demonstrating the model's sensitivity to strong hydrological and backscatter signatures typical of extensive inundation.

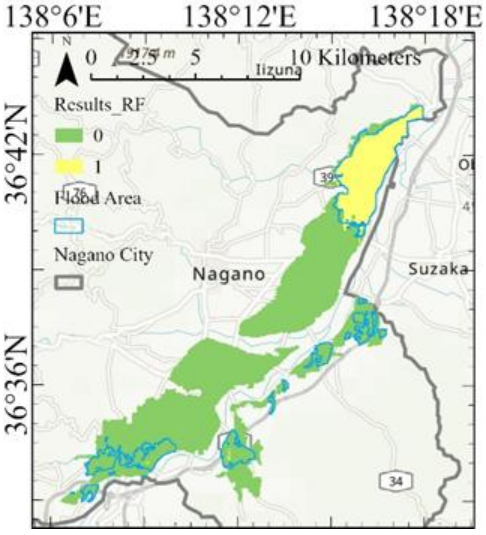


Figure 5.20: RF-predicted result

However, the RF model shows reduced performance when detecting small and spatially fragmented flood patches located in downstream or peripheral low-lying zones. These areas often exhibit subtle spectral or topographic changes that resemble non-flooded pixels, leading to misclassification. This limitation is consistent with previous studies, where RF tends to favor larger homogeneous clusters due to the ensemble nature of tree-based learning.

Despite this, the overall spatial agreement between the estimated and reference flood extent remains high, suggesting that integrating multi-orbit SAR with DEM-derived hydrological features provides a strong foundation for flood mapping in urban riverine environments. The visually coherent flood shapes in the RF output indicate that multi-angle SAR backscatter (ascending + descending) contributes to reducing geometric distortions and improves spatial consistency in the mapped inundation areas.

(b) Model Accuracy Evaluation

Quantitative evaluation of the RF classification results was performed using the confusion matrix. Computed accuracy is shown in Table 5.3 and summarized as follows:

Table 5.3: Confusion Matrix Results

	True (Actual data)	False (Actual data)
Positive (Estimated)	0.143	0.005
Negative (Estimated)	0.735	0.117

Overall Accuracy: 0.878

True Positive (TP): moderate; underestimates small flooded patches.

True Negative (TN): high; effectively identifies non-flooded areas.

False Positive (FP): limited; only small clusters incorrectly labeled as flooded.

False Negative (FN): more frequent, particularly in small scattered inundation zones.

The relatively high TN value indicates that the model robustly identifies non-flooded

areas, consistent with the strong topographic separation between riverine lowlands and elevated urban districts. The comparatively lower TP value reflects the model's difficulty in detecting small-scale or shallow flood regions, where SAR backscatter contrast may be weak and where hydrological indices are less distinctive.

Nevertheless, the overall accuracy of 0.878 demonstrates that the RF model performs reliably for large-scale flood mapping and achieves performance comparable to reported values in the literature for flood detection using Sentinel-1 data.

(c) Interpretation of Model Behavior

The model's estimation characteristics can be interpreted from both the nature of the input data and the behavior of ensemble tree models:

Influence of Multi-Orbit SAR: The inclusion of both ascending and descending orbit images mitigates the effect of radar shadowing and layover, contributing to more stable flood delineation in the major inundation zones.

Role of Terrain and Hydrological Features: Elevation, slope, TWI, and SPI play an essential role in guiding the model toward hydrologically plausible flood shapes. Areas with high moisture accumulation potential (high TWI) and low slope are more likely to be estimated correctly.

Challenges in Small Flood Areas: Fragmented water bodies or shallow inundation frequently produce backscatter values similar to wet soil or asphalted urban areas, which explains the higher FN rate.

General Model Stability: The balanced combination of SAR backscatter and hydrological features ensures that estimations are not dominated by a single data source, contributing to overall model robustness.

(d) Summary

The RF model demonstrates strong capability in estimating the spatial patterns of flood extent in Nagano City following the 2019 Typhoon Hagibis disaster. The model effectively captures the major inundation zones and achieves high overall accuracy, confirming the value of integrating multi-orbit SAR and terrain-based hydrological features. Although the detection of small-scale flood patches remains a challenge, the results indicate that the RF classifier provides a reliable and interpretable approach to flood mapping and forms a key methodological component of the multi-source data integration framework established in this dissertation.

5.4 Summary of Experiment II

Experiment II examined how multi-source geospatial information—specifically remote sensing data, terrain attributes, and social sensing indicators—can be integrated within a Random Forest framework to improve flood-extent estimation. The chapter presented two complementary analyses. The first experiment evaluated the contribution of GPS population data to RF-based flood estimation, while the second experiment assessed the estimation improvements achieved by integrating ascending and descending

Sentinel-1 backscatter with hydrological and topographic parameters. Together, these analyses provide an empirically grounded understanding of how different data modalities enhance model accuracy and spatial reliability.

The GPS-based RF model demonstrated that incorporating human mobility information strengthens flood detection in urbanized districts where physical signals alone may be ambiguous. Increased population density, reduced movement, and localized behavioral anomalies contributed to refining flood estimations in areas where SAR backscatter is affected by built-up surfaces or shallow inundation. The results confirmed that integrating social sensing into machine-learning models improves both classification accuracy and spatial coherence, offering a valuable complement to traditional remote sensing approaches.

In the second analysis, the use of multi-orbit SAR—combining ascending and descending Sentinel-1 imagery—proved essential for capturing the geometric complexity of the flood event along the Chikuma River. The RF classifier, trained on dual-orbit backscatter and enhanced with terrain-derived indices such as elevation, slope, TWI, and SPI, yielded high overall accuracy and successfully mapped major inundation zones. Although small, fragmented flood patches remained challenging due to subtle backscatter differences, the model demonstrated strong capability in delineating large, contiguous flood regions. These findings reinforce the importance of combining physical environmental factors with multi-angle radar observations to overcome limitations caused by layover, shadowing, and surface heterogeneity.

Across both experimental settings, several consistent themes emerged.

First, the Random Forest model exhibited robustness and interpretability, with stable Feature importance rankings that highlight the key roles of elevation, slope, and SAR backscatter. Second, the integration of additional data sources—whether social sensing or multi-orbit SAR—produced measurable improvements in estimation accuracy, confirming the central premise of this dissertation: multi-source data fusion enhances the reliability of urban flood assessment. Third, both experiments demonstrated strong agreement between the model-estimated flood extent and official inundation maps, validating the practicality of the proposed framework for real-world disaster management.

Overall, Experiment II establishes the effectiveness of machine-learning-based multi-source integration for urban flood modeling. It bridges the gap between physical sensing, human behavioral indicators, and data-driven estimation techniques. The insights gained here form the methodological foundation for the final experiment in Chapter 6, where remote sensing and social sensing are jointly integrated within a fine-resolution small-grid model to advance toward a unified framework for resilient city assessment.

Chapter 6 Core Experiment III: Small-Grid (70×70 m) Urban Flood Estimation with Twitter + GPS + Remote Sensing

This chapter is adapted from the following peer-reviewed journal article: Yifan Yang, Naoki Ohira and Hideomi Gokon, "Small-Grid Urban Flood Prediction Model Using Twitter Data and Population GPS Data: An Example of the 2019 Nagano City Flood," Progress in Disaster Science, vol. 24, 2024. Additional methodological explanations, data descriptions, and discussions have been added to better fit the dissertation framework.

The previous chapters (Chapters 4 and 5) established the methodological and empirical foundations for multi-source flood analysis in Nagano City. Experiment I (Chapter 4) demonstrated that large-scale GPS mobility data exhibit clear temporal and spatial anomalies during the 2019 flood, confirming their potential as social-sensing indicators of population response. Experiment II (Chapter 5) expanded this foundation by integrating physical sensing (Sentinel-1 SAR and DEM-derived terrain indices) with GPS data in a Random Forest framework, proving that social-sensing information enhances flood-extent classification accuracy and spatial consistency.

Experiment III, presented in this chapter, constitutes the central contribution of this doctoral thesis. Building upon the insights from Experiments I and II, this experiment advances to a fine-resolution, 70 m × 70 m small-grid estimation model that integrates georeferenced Twitter posts, GPS mobility data, SAR backscatter, DEM information, and urban infrastructure features. Unlike the previous chapters that utilized coarser spatial units (500 m grid in Experiment I; mixed pixel/grid-level RF models in Experiment II), this experiment focuses on urban-scale flood dynamics, enabling high-resolution detection of inundation patterns in residential, commercial, and transportation zones.

The integration of real-time social media (Twitter) with GPS-derived human mobility indicators provides complementary behavioral insights that physical sensing alone cannot capture—particularly in densely built environments where SAR backscatter may exhibit ambiguity due to surface roughness, urban structures, or shallow flooding. This chapter operationalizes the multi-source fusion concept by incorporating (a) natural language processing to extract flood-related Twitter information, (b) spatial clustering of GPS population density changes, and (c) environmental indicators derived from SAR and DEM. The resulting model provides a more comprehensive representation of flood impacts, capturing both physical inundation and human behavioral responses during the disaster.

This experiment represents the culmination of the thesis framework "Sensing → Integration → Application." By fusing multi-source data at a fine spatial resolution, the study advances toward a practical, scalable, and near-real-time flood-monitoring system.

The small-grid model not only improves estimation accuracy but also contributes to the broader goal of enhancing urban resilience assessment, demonstrating how combined physical and social sensing can support decision-making in disaster management.

6.1 Objective

The objective of Experiment III is to develop a fine-resolution, small-grid flood estimation model that integrates multi-source physical and social sensing data to represent urban-scale disaster dynamics with high spatial precision. Building upon the findings from Experiments I and II, this experiment advances the analytical framework from coarse-grid anomaly detection and medium-resolution Random Forest modeling to a 70 m × 70 m urban-grid model, enabling detailed assessment of inundation patterns within densely built environments.

The primary goal of this experiment is to evaluate how social media–derived disaster information can complement physical remote sensing indicators in estimating flood extent. To achieve this, georeferenced Twitter posts related to flood impacts are combined with GPS-based population mobility data, Sentinel-1 SAR backscatter values, DEM-derived terrain parameters, and proximity to urban infrastructure. Through this integration, the experiment seeks to demonstrate that behavioral signals observed in real time can enhance or refine estimations derived from physical sensing alone, particularly in complex urban settings where radar responses may be influenced by surface roughness, land cover heterogeneity, or shallow flooding.

Specifically, the objectives of Experiment III are as follows:

(a) To construct a high-resolution (70 m × 70 m) flood estimation model that integrates remote sensing (SAR, DEM), social sensing (GPS), and social media (Twitter) information into a unified feature space suitable for fine-grid analysis.

(b) To quantify the estimative value of geospatially filtered Twitter messages by incorporating them as supplementary behavioral indicators, thereby assessing whether real-time public reporting improves grid-level flood detection.

(c) To evaluate the spatial complementarity between physical and social sensing variables, particularly in areas where SAR backscatter alone may be insufficient (e.g., residential areas with mixed land cover, urban roads, shallow inundation zones).

(d) To validate the resulting model against official inundation records by comparing estimation performance, spatial agreement, and grid-level accuracy, and to identify conditions where multi-source data provide meaningful improvement.

(c) To establish the small-grid model as a core component of the dissertation's "Sensing → Integration → Application" framework, demonstrating the feasibility of near-real-time urban flood mapping for resilient city assessment.

By achieving these objectives, Experiment III provides the highest-resolution representation of flood impacts within this dissertation and serves as a practical demonstration of the usefulness of integrating remote sensing, social sensing, and social

media analytics within a data-driven urban resilience framework.

6.2 Study Area and Data

6.2.1. Study Area

The study area selected for Experiment III corresponds to the urban corridor along the Chikuma River in Nagano City, Japan, one of the region's most severely affected during Typhoon Hagibis in October 2019. According to the inundation records released by the Ministry of Land, Infrastructure, Transport and Tourism (MLIT), several flood zones were distributed along the Chikuma River, with one area exhibiting the largest continuous flood extent. As shown in Figure 6.1, this area was therefore chosen as the primary study region to minimize the influence of confounding flood events in surrounding districts.

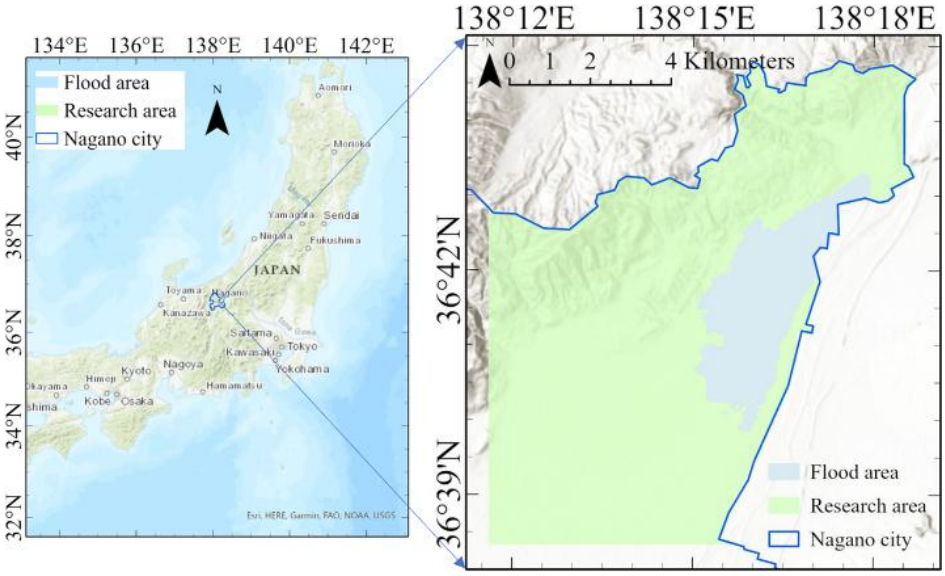


Figure 6.1: Study Area

To ensure balanced comparison between flooded and non-flooded zones, the largest non-flooded areas unaffected by other flood extents were also included. In total, the flooded area accounts for approximately one-fifth of the entire study region. The study area was further segmented into 70 m × 70 m grid cells, allowing fine-scale integration of physical, demographic, and social media indicators.

6.2.2. Twitter Data

To capture real-time public responses to the flood event, georeferenced Twitter posts related to flooding were collected for the period 12–13 October 2019, corresponding to the peak flooding hours in Nagano City. The tweets were filtered using flood-related keywords and spatially restricted to the Chikuma River corridor.

Figure 6.2 presents the hourly tweet count, which shows a pronounced increase in

posting activity during the flood onset and peak inundation period. These temporal spikes indicate heightened public reporting and situational awareness, making Twitter a valuable supplemental information source for flood detection and urban impact assessment.

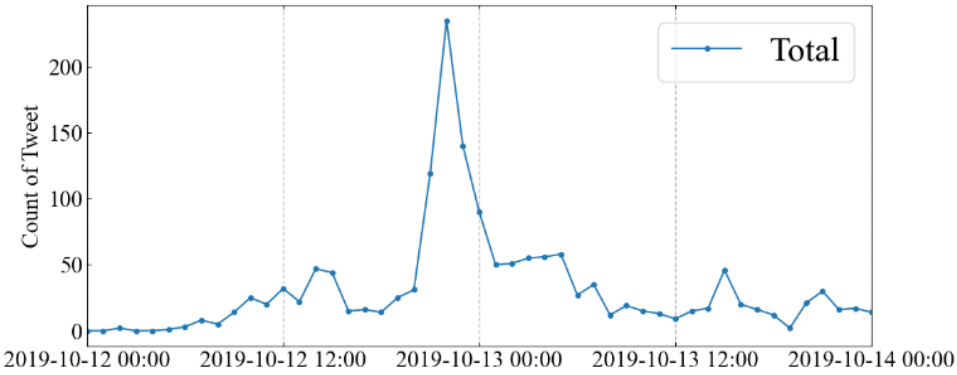


Figure 6.2: Hourly distribution of flood-related tweets during 12–13 October 2019

After preprocessing, the georeferenced tweets were mapped to the 70 m grid cells to generate a tweet-density feature, which later contributed to the fine-grid urban flood estimation model.

6.2.3. GPS Mobility Data

Mobile-phone GPS data were obtained from Agoop Corporation for two periods: Flood period: 12–13 October 2019; Reference period: 5–6 October 2019 (one week prior).

Figure 6.3 illustrates the spatial distribution of GPS points on 12 October 2019, showing dense mobility patterns in urbanized zones around central Nagano.

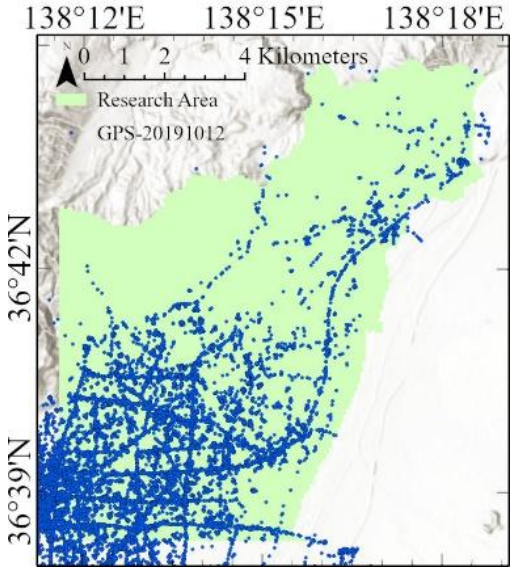


Figure 6.3: Spatial distribution of GPS data on 12 October 2019

The collection of GPS data is based on signals automatically transmitted from mobile phones to base stations during movement and at fixed intervals during inactivity. Thus, GPS point density reflects both population concentration and mobility intensity. During flood events, anomalies in GPS activity—such as abrupt increases due to evacuation movement or decreases due to immobilization—provide critical behavioral indicators of flood impact.

In this experiment, GPS data were aggregated to 70 m grids to ensure consistency with other spatial datasets and to support fine-scale multi-source modeling.

6.2.4. Flood-Related Data

To comprehensively model and interpret flood conditions, several environmental and infrastructural datasets were incorporated.

(a) Geographical data

The flood-related geographic data used in this study include elevation, slope, river, road, land use, and land cover data.

Elevation and slope data were calculated for the study area using Digital Elevation Model (DEM) data with a resolution of 10 meters. Geographic data on rivers and roads were also obtained to provide a more comprehensive understanding of the influence of geography on flood shape.

Land use data downloaded from Japan's Ministry of Land, Infrastructure, Transport and Tourism (MLIT) included 17 categories divided into building and non-building land based on the actual situation in Nagano City.

In addition, Synthetic Aperture Radar (SAR) remote sensing data with high-resolution, all-day observations were used to improve the accuracy of the land data. Specifically, Sentinel-1 data acquired at 8:00 am on 5 October before the flood and at 8:00 am on 12 October at the time of the flood, S1B satellite, IW mode, and GRDH level were used.

(b) Infrastructure data

In the event of a flood disaster, people often seek areas such as shelters to ensure their safety. In Japan, these shelters often include sturdy buildings such as schools, gymnasiums and community centers, which are highly resistant to disasters and can provide enough space to accommodate large numbers of people. Therefore, the infrastructure data of interest in this study includes data on flood-related shelters, data on public service facilities, and data on hospitals.

A comprehensive analysis of these infrastructure data will help to estimat the impact of flooding on the city in a more comprehensive way.

6.2.5. Data Summary

All data sources were harmonized into the unified 70 m × 70 m grid system, allowing integration of:

Physical indicators (SAR, DEM, slope, river proximity)

Behavioral indicators (GPS mobility)
Social media indicators (Twitter posts)
Infrastructure indicators (shelters, public facilities)
Official MLIT flood extent (ground truth)

These multi-dimensional datasets constitute the full input space for Experiment III's fine-grid flood estimation model.

6.3 Methodology

The methodological framework of Experiment III follows a multi-stage process that integrates social media information, GPS-based population mobility, and remote sensing–derived environmental variables into a unified fine-grid flood estimation model. This section presents the complete workflow, beginning with the processing of Twitter and GPS data, followed by the derivation of physical and infrastructural variables, and culminating in the construction, optimization, and evaluation of Random Forest–based models. The overall workflow is summarized in Figure 6.4, which outlines how heterogeneous datasets were transformed into grid-level features and fused within the machine learning framework.

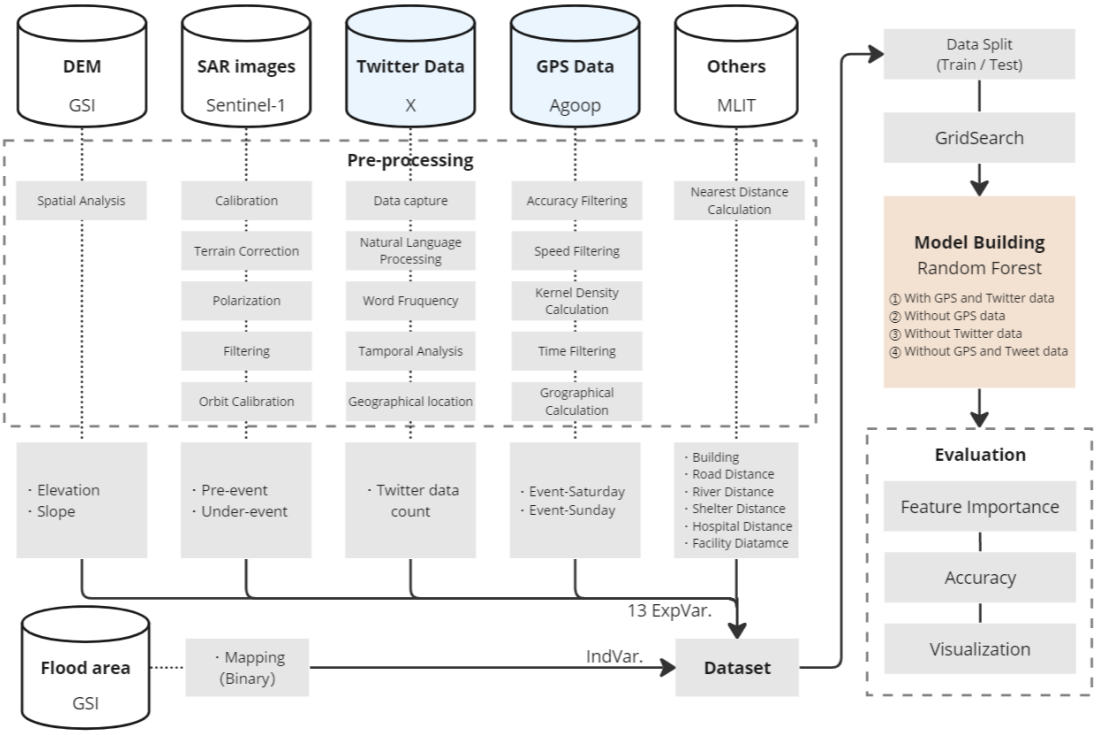


Figure 6.4: Workflow of Experiment III

6.3.1. Processing of Twitter Data

A total of 1,520 geolocated tweets posted between October 12 and 13, 2019 were collected within Nagano City. The primary objective of the Twitter preprocessing stage

was to extract flood-related content and identify spatiotemporal patterns in social media reactions during the flood event. To filter relevant information, natural language processing techniques were applied to all tweets, with an emphasis on identifying frequently occurring nouns through Noun Term Frequency (Noun TF). The Noun TF metric quantifies the relative proportion of each noun term in the entire corpus:

$$Noun\ TF(t, D) = \frac{\text{Total occurrences of term } t \text{ in all tweets } D}{\text{Total number of all noun terms}} \quad (19)$$

Using this method, the top ten high-frequency nouns were extracted (Figure 6.5), revealing terms closely associated with flooding, such as "water," "river," "evacuation," and "danger." Keywords irrelevant to the flood—such as "Nagano City," which appeared excessively due to being part of user profile or location metadata—were removed.

Temporal analysis of keyword usage (Figure 6.6) showed that keyword frequency fluctuated sharply between 12:00 on October 12 and 12:00 on October 13, corresponding directly to the known timeline of the flood. In this study, such temporal variations are interpreted as reflecting different phases of public communication during the disaster, rather than direct changes in flood intensity. An increase in keyword frequency indicates heightened situational awareness and information sharing as flood impacts emerge, while a subsequent decrease may reflect evacuation behavior, power outages, reduced communication activity, or shifts in public attention during peak flooding conditions.

Therefore, the increase and decrease in Twitter activity are understood as changes in communication behavior across disaster phases, rather than as a direct proxy for flood severity. This confirmed that Twitter activity was sensitive to disaster evolution and could supplement physical sensing datasets.

Spatially, tweets containing flood-relevant keywords were aggregated by administrative district and mapped across Nagano City (Figure 6.7), highlighting concentrations of posts near riverbanks and inundated regions.

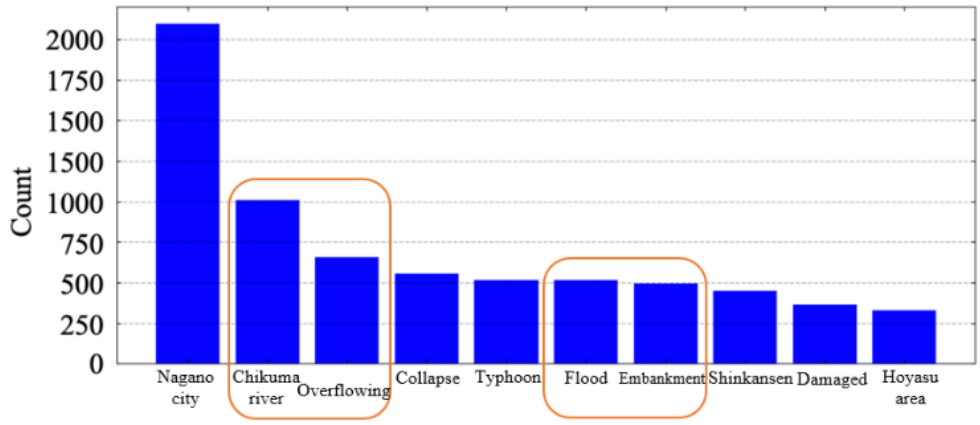


Figure 6.5: Twitter Keyword Statistics

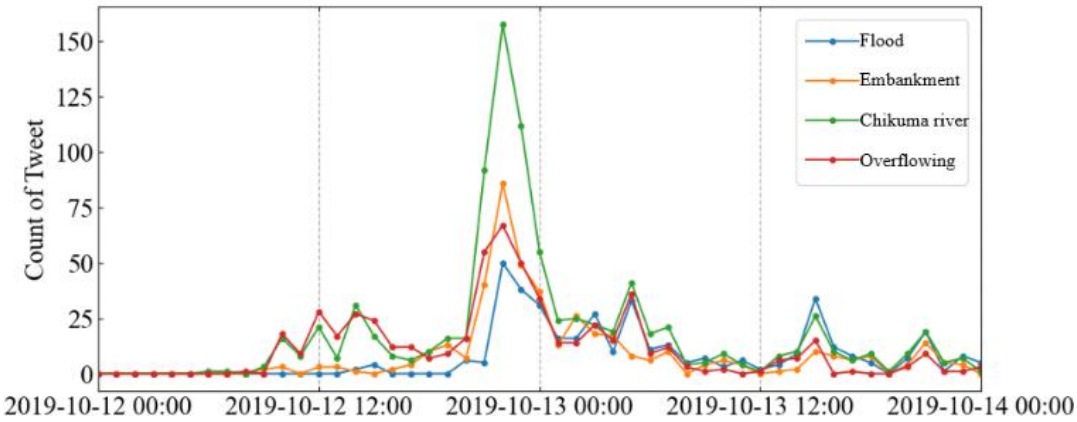


Figure 6.6: Distribution of keywords over time

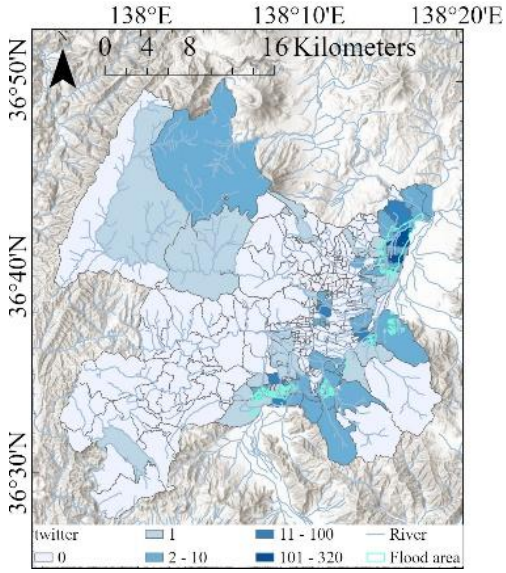


Figure 6.7: Spatial distribution of Twitter data

6.3.2. Processing of GPS Data

GPS data for two periods—October 12–13 (flood dates) and October 5–6 (baseline weekend)—were obtained from Agoop Corporation. Data accuracy, expressed as location uncertainty in meters, showed that 68% of points had an accuracy worse than 70 meters. To ensure spatial reliability, all GPS points with <70 m accuracy were removed. Additionally, records representing highway travel were excluded based on speed thresholds: GPS points with speeds exceeding 13.89 m/s (equivalent to the minimum highway speed of 50 km/h) were filtered out.

To quantify population movement intensity, kernel density estimation (KDE) was applied to the cleaned GPS dataset using a 70 m bandwidth, consistent with the spatial resolution of the 70 m × 70 m analytical grid. The KDE formula is:

$$Kernel\ density(x,y) = \sum_{i=1}^{N} K\left(\frac{\|(x,y) - (x_i,y_i)\|}{h}\right) \quad (20)$$

where K is the kernel function and h is the bandwidth. A total of 15,799 grid cells (70 m × 70 m) were constructed (Figure 6.6), and KDE values were assigned to each grid cell.

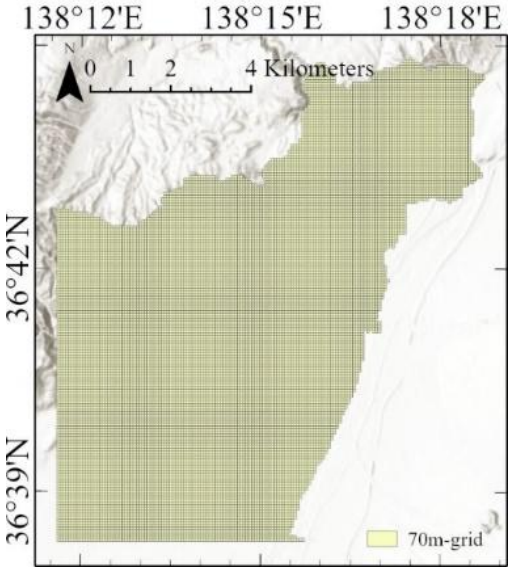


Figure 6.8: The 70m x 70m grid of the study area

Because Twitter analysis indicated that the flood impact period occurred between 12:00 on October 12 and 12:00 on October 13, GPS mobility anomalies were defined as significant deviations from normal movement patterns during this period. These anomalies were quantified by subtracting baseline population densities (October 5–6) from the corresponding densities during the flood-impact period. This approach isolates flood-induced mobility changes that arise from disrupted accessibility, evacuation, or sheltering behavior, thereby enhancing sensitivity to localized disruptions associated with flooding.

6.3.3. Processing of Flood-Related Physical and Infrastructure Data

Flood-related variables were computed for each 70 × 70 m grid cell using the datasets summarized in Table 6.1. Physical variables included elevation, slope, land cover, SAR backscatter intensity, and distances to nearest rivers and roads; infrastructural variables included distances to shelters, public service facilities, and hospitals.

Table 6.1: Methods for calculation of flood-related data

Data	Method	Explanation
Elevation	Spatial join	Average elevation
Slope	Spatial join	Average Slope
Building land	Spatial join	Percentage of building land

Land cover (backscatter coefficient)	Spatial join	Average backscatter coefficient
River	Near	Distance to the nearest river
Roads	Near	Distance to the nearest road
Shelter	Near	Distance to the nearest shelter
Public Service Facilities	Near	Distance to the nearest public service facility
Hospitals	Near	Distance to the nearest hospital

Backscatter Coefficient Calculation: Sentinel-1 SAR images (pre-flood and during-flood acquisitions) were preprocessed following standard procedures—thermal noise removal, radiometric calibration, orthorectification, topographic correction, and filtering. The derived backscatter coefficients (Figure 6.9) capture surface reflectance characteristics, with markedly lower values indicating water-covered areas. These coefficients are critical indicators in flood models because floodwater greatly attenuates SAR returns.

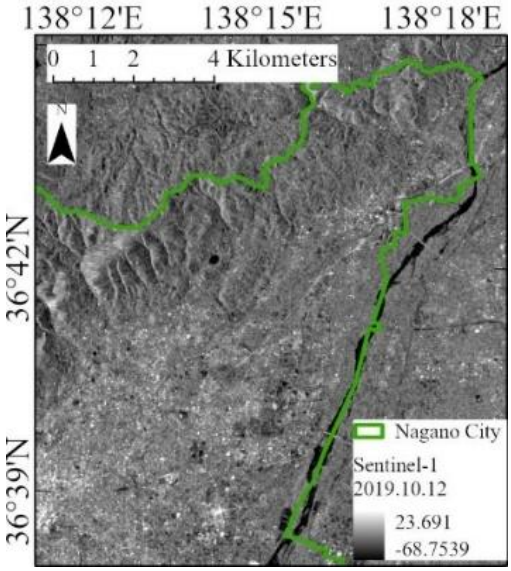


Figure 6.9: Distribution of backward scattering coefficient – 2019.10.12

Other Physical and Infrastructure Variables: Using Spatial Join operations, grid-level averages were computed for elevation, slope, land-cover backscatter values, and percentage of building land. The mean for each grid was calculated as:

$$Average\ Value(i) = \frac{1}{N_i}\sum_{j=1}^{N_i} Value_j \tag{21}$$

Distances from each grid cell to the nearest river, road, shelter, public service facility, and hospital were computed using the Near tool:

$$Nearest\ Distance(i, Feature_k) = min_j Distance(Grid_i, Feature_k)_j \quad (22)$$

The spatial patterns of these flood-related variables are shown in Figure 6.10.

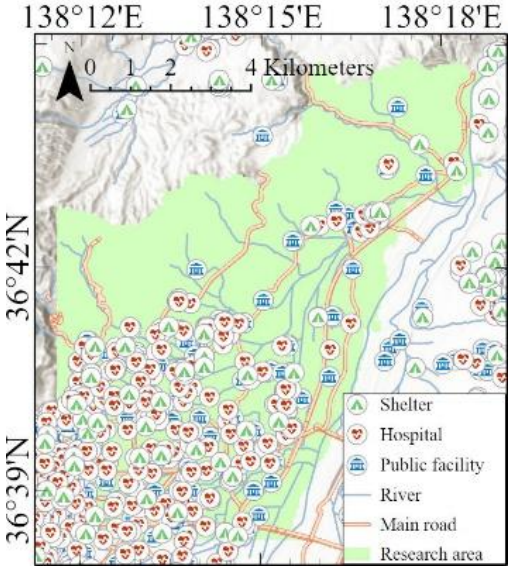


Figure 6.10: Distribution of flood-related data

6.3.4. Construction of the Estimation Model

Random Forest (RF) modeling was employed to predict flood extent under different data-integration scenarios. Four models were constructed:

(a) GPS + Twitter model (With Twitter and GPS data model),

(b) GPS-only model (Without Twitter data model),

(c) Twitter-only model (Without GPS data model),

(d) Baseline physical model (Without Twitter and GPS data model).

All models used the same hyperparameters tuned using grid search applied to the GPS + Twitter model. The optimal parameters were: entropy criterion, maximum tree depth of 4, 100 trees, and random_state = 77. The RF model is formulated as:

$$RF(X) = \frac{1}{n}\sum_{i=1}^{n} T_i(X) \quad (23)$$

where T_i denotes the i-th decision tree.

Eighty percent of grid samples were used for training and twenty percent for testing. After training, feature importance scores were computed based on Gini impurity reduction:

$$Feature\ Importance = \frac{Sum\ of\ Gini\ Importance}{Total\ Gini\ Importance} \quad (24)$$

This analysis revealed the relative contribution of social sensing, social media, and physical features in explaining flood extent.

6.3.5. Accuracy Evaluation

Model performance was evaluated using confusion matrices and four standard metrics—accuracy, precision, recall, and F1 score. The confusion matrix structure is same as Table5.2.

These evaluation metrics provided a comprehensive assessment of how different data combinations influenced estimation sensitivity, false positives, and overall model robustness. The results of these metrics are presented in Section 6.4.

6.4 Evaluation and Results

6.4.1. Feature Importance Analysis

Figure 6.11 presents the feature importance scores for the four Random Forest models using different combinations of data sources: (1) model with both Twitter and GPS data, (2) model without Twitter data, (3) model without GPS data, and (4) baseline model without Twitter and GPS data. Across all four configurations, elevation and slope consistently emerge as the two most important features, reflecting the dominant role of topographic controls in determining flood susceptibility in the Chikuma River floodplain.

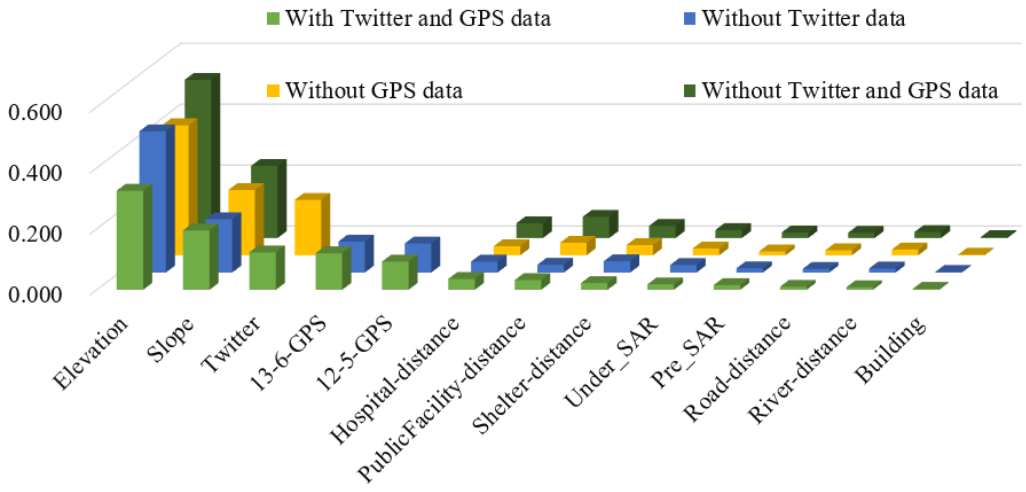


Figure 6.11: Feature importance of 4 models

In the full model that integrates both Twitter and GPS, elevation and slope jointly account for a substantial proportion of the total importance, but their dominance is moderated by the contributions of social sensing and infrastructure-related features. When Twitter data are removed, the importance of elevation increases markedly, while slope importance decreases and the distance to shelters becomes more influential. This indicates that without Twitter information, the model tends to rely more heavily on terrain and static infrastructure proximity to infer potential flood areas. When GPS data are removed, the importance of elevation again increases, and both slope and Twitter data gain slightly more weight, suggesting that Twitter partially compensates for the absence

of mobility-based information.

In the baseline model without both Twitter and GPS data, elevation alone contributes more than half of the total feature importance, and the relative weights of slope, distance to hospitals, public service facilities, and shelters also increase. In this configuration, the model effectively becomes a purely physical–geographical classifier, relying on static conditions rather than real-time or behavior-based signals. These results highlight that, while topography is a fundamental driver of flooding, the inclusion of social sensing and social media features redistributes importance across variables and reduces excessive dependence on elevation alone.

6.4.2. Model Accuracy and Confusion Matrix Results

The confusion matrix components for the four models—True Positive (TP), False Positive (FP), False Negative (FN), and True Negative (TN)—are summarized in Figure 6.12. When both Twitter and GPS data are incorporated, the model exhibits the highest TP and lowest FP, indicating that it is able to correctly identify flooded areas while minimizing the misclassification of non-flooded cells as flooded. In contrast, the model without Twitter and GPS data yields the lowest TP and highest FP, although its TN is relatively high and FN relatively low. This pattern implies that the baseline model classifies a large proportion of the study area as flooded, achieving many correct detections but at the expense of substantial overestimation.

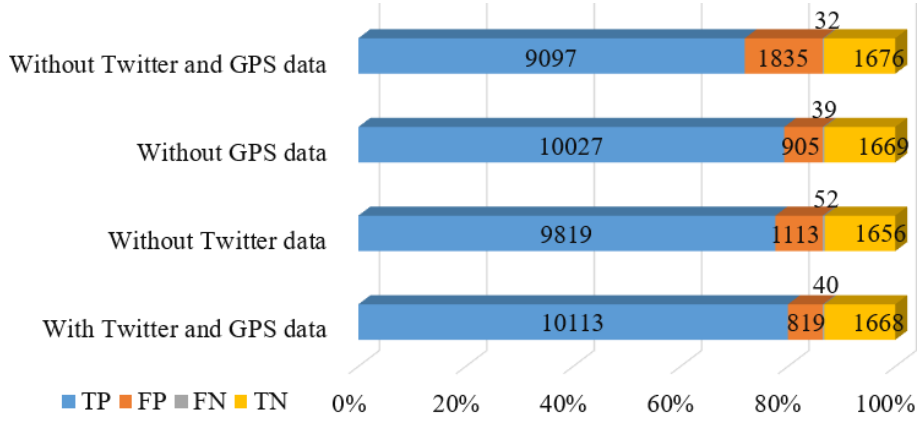


Figure 6.12: Confusion matrices result of 4 models

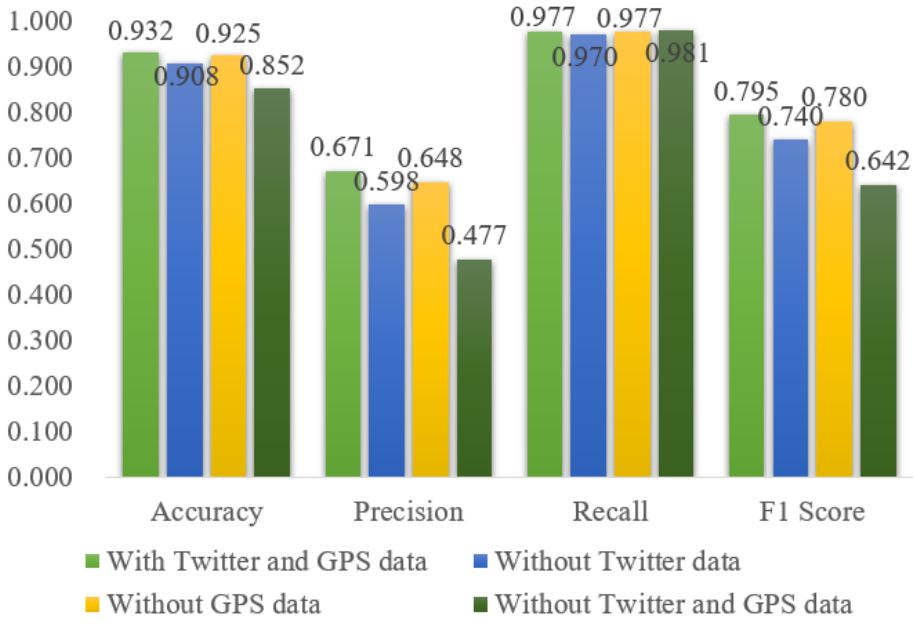


Figure 6.13: Accuracy evaluation indicator results of 4 models

Figure 6.13 reports the four-accuracy metrics—Accuracy, Precision, Recall, and F1 Score—for all model configurations. The integrated model with Twitter and GPS data achieves the highest overall accuracy, with a value of approximately 0.932, followed by 0.925 for the model without Twitter and 0.908 for the model without GPS, while the baseline model without both data sources shows the lowest accuracy (0.852). The recall of the model without GPS data is the highest among all, indicating that this configuration is most sensitive to detecting actual flooded cells, but at a cost of increased false positives.

The baseline model without Twitter and GPS data exhibits the lowest precision, meaning that a large proportion of cells estimated as flooded are not actually inundated. In contrast, introducing both social media and social sensing data substantially improves precision while maintaining high recall, thereby enhancing the model's practical suitability for operational flood mapping.

6.4.3. Spatial Patterns of Flood Estimation

Figures 6.14–6.17 compare the spatial distribution of estimated flood extent with the official inundation map for each model. In all four configurations, the main flooded zone along the Chikuma River is successfully identified, and the core inundation area is consistently covered. However, in every case the estimated flood extent exceeds the actual mapped flood area, with overestimation primarily occurring in the southeastern and northeastern portions of the study region.

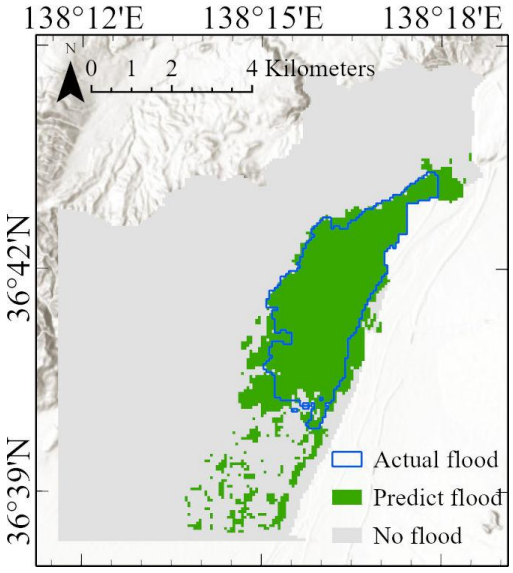


Figure 6.14: Flood estimation model result with twitter data and GPS data

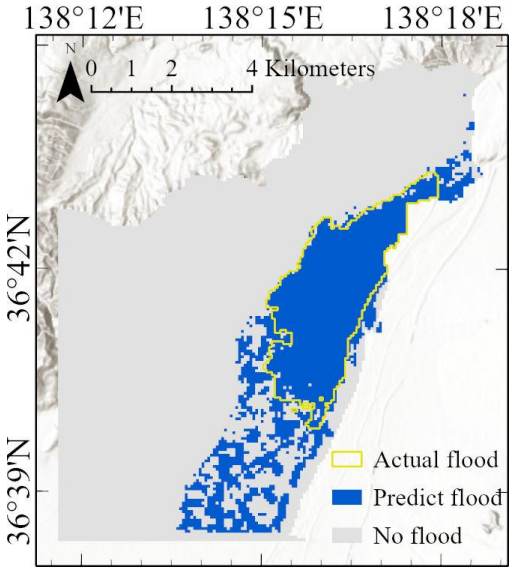


Figure 6.15: Flood estimation model result without twitter data

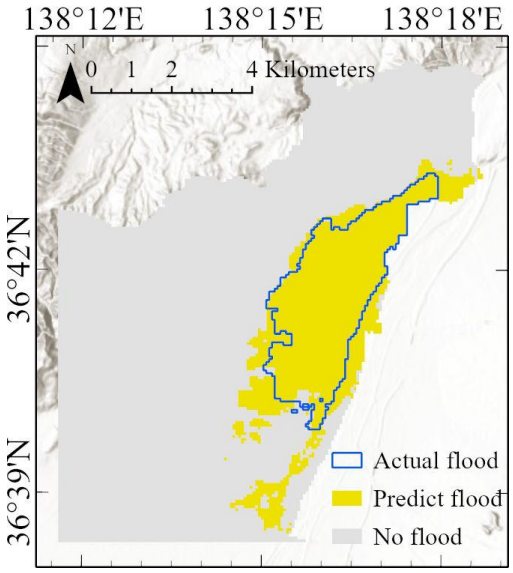


Figure 6.16: Flood estimation model result without GPS data

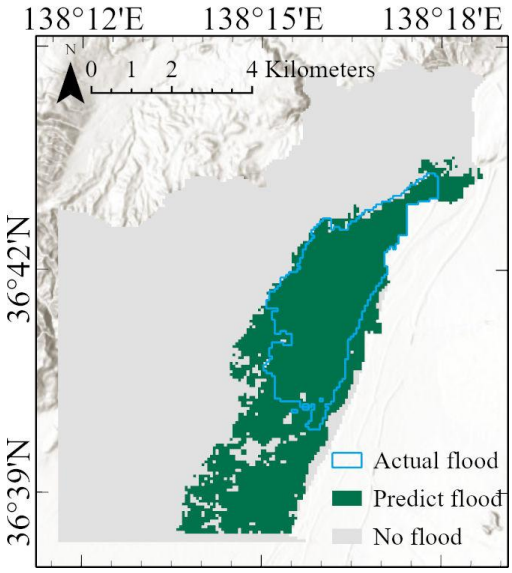


Figure 6.17: Flood estimation model result without twitter data and GPS data

The model integrating both Twitter and GPS data yields the most compact and spatially concentrated flood extent, with comparatively fewer overestimated areas. In contrast, the models that exclude Twitter data tend to estimate larger flood zones that extend well beyond the actual inundation boundary. In particular, the model without GPS data estimates almost the entire southeastern sector of the study area as flooded, revealing the tendency of purely terrain- and Twitter-based inference to misclassify low-lying regions even when no strong behavioral or social-media evidence of flooding exists.

These spatial patterns confirm that while all models are able to capture the major flooded region, the incorporation of both Twitter and GPS data plays a critical role in

constraining the spatial spread of estimated floods, thereby improving the realism and operational applicability of flood-extent mapping.

6.5 Discussion

6.5.1. Comprehensive Analysis of Model Performance

The Random Forest models constructed in Experiment III were tuned through grid search optimization and applied to the flood-affected region along the Chikuma River. The Feature importance results demonstrate that elevation and slope consistently dominate all configurations, underscoring the strong influence of topographic gradients on flood distribution in Nagano City. As shown in Figure 6.11 (Feature importance results), the steep terrain in the northwest and the low-lying floodplain in the southeast shape the model's estimations, with lower elevations and gentle slopes strongly associated with observed inundation.

In the absence of Twitter and GPS data, the model relies almost exclusively on physical and infrastructural factors such as elevation, slope, distance to shelters, and proximity to roads. While this configuration still identifies the broad flood-prone zone, the lack of real-time human-sensing information results in substantial false positives (FP) and artificially expanded flood boundaries, reflecting the model's reduced ability to capture the temporal dynamics of flooding.

When GPS data are excluded, the model compensates by increasing reliance on elevation, slope, and Twitter information. Conversely, when Twitter data are absent, the importance of elevation further increases, indicating that Twitter contributes unique contextual information not captured by mobility patterns alone. The decline in GPS activity during the flood—likely driven by heavy rainfall and stay-at-home behavior—reduces the distinctiveness of GPS anomalies, particularly in commercial districts. In contrast, Twitter posts filtered through natural language processing preserve explicit flood-related semantics, giving Twitter a stronger contribution to estimative accuracy.

In summary, integrating both GPS and Twitter data enhances the real-time responsiveness of the model. The comparison across different data configurations reveals the unique role of each social-sensing source and provides insights for refining multi-source flood estimation frameworks.

6.5.2. Interpretation of Estimation Patterns

All four Random Forest models successfully capture the major inundation zone along the Chikuma River, with accuracies of 0.932, 0.908, 0.925, and 0.852 for the four respective data configurations. However, all models slightly overestimate the flood extent, particularly toward the southwest portion of the study area. This pattern is consistent with the high importance of elevation and slope, as regions with low elevation or gentle terrain are inherently more susceptible to misclassification as flooded.

Figure 6.18 clearly illustrates that incorporating both Twitter and GPS data markedly reduces false positives. These two data sources are particularly valuable in densely populated areas, where mobility patterns and social media activity provide strong behavioral signatures during disasters. In contrast, sparsely populated regions have limited GPS activity and negligible Twitter presence, which may lead the full-data model to misclassify them as non-flooded. Interestingly, in such low-density areas, the baseline model without Twitter and GPS data sometimes performs better because it relies purely on terrain-based physical characteristics.

A more detailed comparison of the models without GPS and without Twitter is shown in Figure 6.19. The model without Twitter but with GPS estimates the black-circled area as flooded, reflecting significant GPS activity reduction due to rainfall-related movement changes. Conversely, the orange-circled areas are better estimated when Twitter data are included, demonstrating that Twitter provides clear, location-specific flood confirmation that GPS alone cannot offer.

Further evidence is provided in Figures 6.20 and 6.21. Figure 6.20 shows that GPS data can detect abnormal population movements; however, these anomalies reflect disruptions in physical mobility and accessibility and are not uniquely attributable to flooding. Such mobility changes may also be influenced by non-flood-related factors, including rainfall intensity, transportation conditions, or time-of-day effects. In contrast, Figure 6.21 demonstrates that Twitter data, after keyword filtering, captures intentional human communication related to flood experiences. Because Twitter posts are generated through conscious reporting of perceived impacts, they provide more event-specific and spatially precise indications of flooding.

From a mechanistic perspective, GPS data represent physical constraints imposed by environmental and infrastructural conditions, while Twitter data represent semantic confirmation of disaster perception. Therefore, GPS data offers sensitivity to physical disruption, whereas Twitter data provide specificity regarding flood-related impacts. By integrating both sources, the proposed framework links physical accessibility changes with human-reported flood information, yielding more balanced and reliable flood estimation.

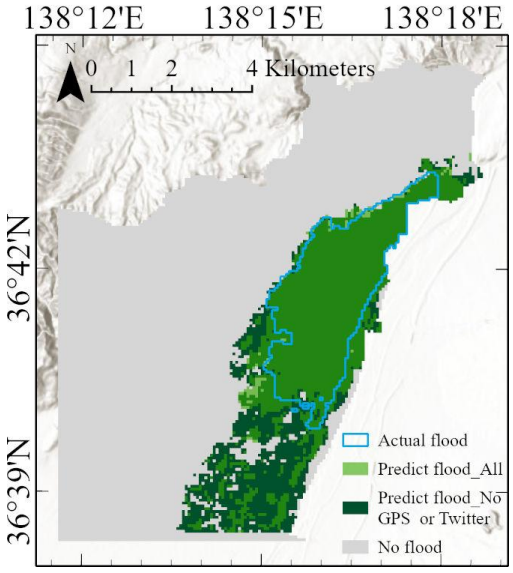


Figure 6.18: Flood estimation model results with and without Twitter data and GPS data

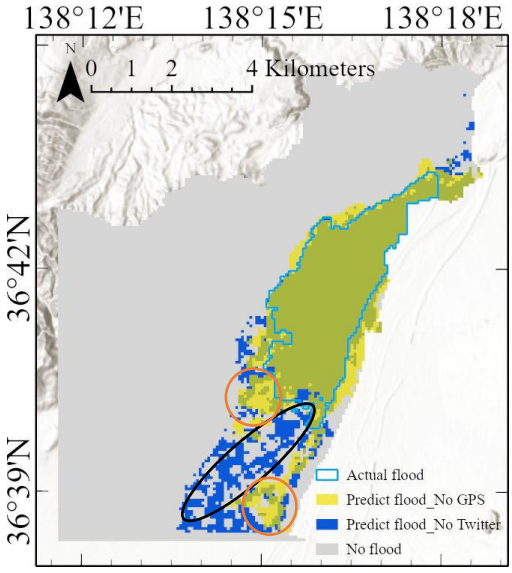


Figure 6.19: Flood estimation model results without Twitter data in comparison to without GPS data

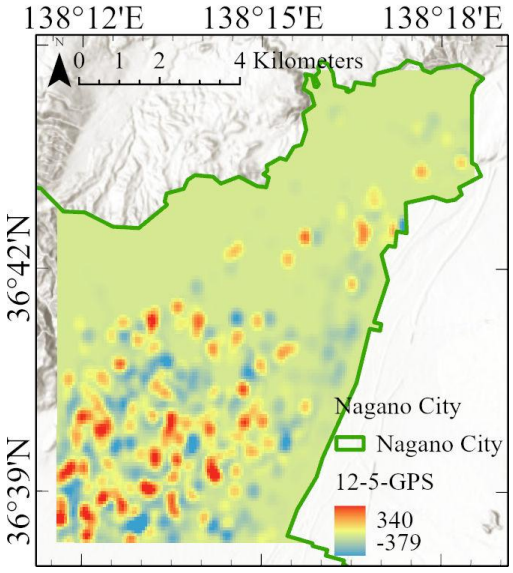


Figure 6.20: Distribution of GPS data

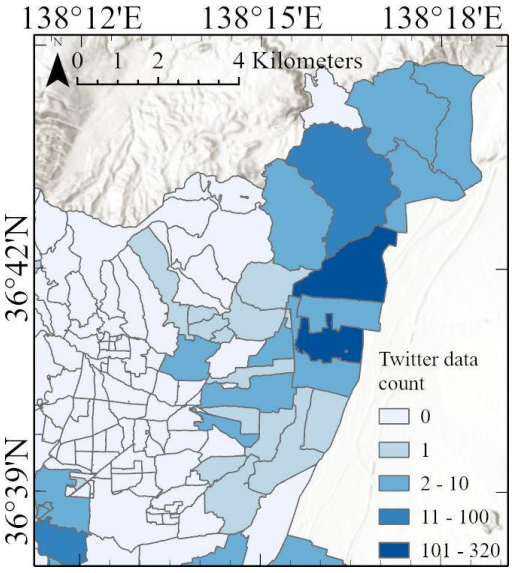


Figure 6.21: Distribution of the count of Twitter data

6.5.3. Limitations and Recommendations

Although the proposed flood estimation model demonstrates strong potential for near-real-time flood assessment, several limitations must be acknowledged. First, the reference flood extent used for model validation relies on Sentinel-1 SAR imagery, which cannot be obtained instantaneously during disasters. This temporal gap affects the model's ability to support fully real-time decision making. Second, during severe flood events, communication disruptions or infrastructure damage may reduce GPS and Twitter data

availability, weakening the effectiveness of social-sensing inputs.

To enhance the operational applicability of the model, collaboration with government agencies is essential. Public-sector organizations have access to a wider range of real-time data, including hydrological sensors, monitoring networks, emergency reports, and historical hazard databases. Integrating these resources could substantially improve model robustness and reduce uncertainty. Additionally, incorporating complementary remote sensing sources—such as ALOS-2, optical imagery, UAV observations, or high-frequency CubeSat data—could further strengthen model performance under diverse conditions.

Future research should also explore more advanced data-fusion strategies, such as weighted ensemble learning or probabilistic reasoning frameworks, to better integrate heterogeneous data sources and mitigate temporal discrepancies. These enhancements would contribute to more reliable, scalable, and timely flood estimation systems, strengthening urban disaster resilience.

Chapter 7 Discussion: Toward a Data-Driven Resilient City Model

7.1 Overview

This chapter synthesizes the findings from the three experiments conducted in this dissertation and discusses their broader implications for urban disaster estimation, assessment, and resilience. Chapters 4–6 collectively demonstrated the value of multi-source data—remote sensing, GPS-based mobility information, and georeferenced Twitter posts—in capturing the physical and behavioral dynamics of flood events in Nagano City. Building on these results, this chapter examines methodological contributions, evaluates the integration of physical and social sensing, and articulates how these findings advance the development of a data-driven resilient city model. The chapter concludes with reflections on practical implications, limitations, and opportunities for future research.

7.2 Key Discoveries and Theoretical Contributions

The integration of multi-source data throughout this dissertation contributes not only to advancements in flood estimation but also to the broader domain of Knowledge Science (KS) by demonstrating how heterogeneous physical and behavioral information can be transformed into a unified system of knowledge creation. Conceptually, the dissertation proposes a knowledge-oriented perspective in which disaster understanding arises from the interaction of environmental processes and human behavioral responses.

By showing that GPS mobility and Twitter communication function as interpretable indicators of population behavior, the study extends KS by positioning human activity patterns as meaningful knowledge sources rather than peripheral data signals. This framework reframes flood estimation as a process of sensing, interpreting, and integrating diverse forms of knowledge — physical, hydrological, behavioral, and perceptual — thereby contributing a conceptual advance to KS in the context of urban disaster research.

Methodologically, the dissertation introduces a structured Sensing–Integration–Application framework that organizes multi-source data into a coherent analytical progression. This methodological architecture demonstrates how estimation models can be scaled from coarse-grained behavioral analysis to fine-grid, urban-scale flood estimation. The development of multi-source Random Forest models further contributes to KS by showing how physical and behavioral indicators can be fused into interpretable, evidence-based analytical systems. The incorporation of feature importance analysis enhances methodological transparency and produces explanatory knowledge about the mechanisms that influence flood extent estimation. These innovations collectively strengthen the methodological foundation of multi-source disaster modeling and illustrate how KS principles of knowledge generation and interpretation can be embedded into computational workflows.

The empirical findings of the dissertation reveal several important discoveries. First, human mobility disruptions encode behavioral signatures of flood impact. Rather than aligning strictly with the timing of physical inundation, GPS anomalies reflect perceived risk, accessibility constraints, and exposure to hazard zones. This demonstrates that behavioral sensing offers early, fine-grained insights that complement physical measurements. Second, social sensing significantly improves the spatial coherence and accuracy of flood extent estimation. The integration of GPS and Twitter reduces false positives, sharpens inundation boundaries, and reveals information not captured by SAR or terrain features alone. Third, the fine-grid model shows that multi-scale integration enables near–real-time monitoring by combining social media content with mobility patterns and SAR-derived indicators. These discoveries collectively illustrate how behavioral and physical sensing jointly inform the understanding of urban flood dynamics.

The theoretical foundations of the proposed framework further validate the scientific rationale for multi-source integration. SAR backscatter characteristics are grounded in electromagnetic scattering theory, while terrain indices such as elevation, TWI, and SPI reflect hydrological pathways of water accumulation. Behavioral adaptation and accessibility theories explain why GPS mobility declines under hazardous conditions, and crisis communication theory supports the use of Twitter as an indicator of public awareness and localized impact. Together, these theoretical underpinnings justify the integration of physical, hydrological, and behavioral data as a unified knowledge system.

Comparisons with existing approaches demonstrate that the proposed framework not only improves estimation performance but also generates richer interpretive value. Traditional flood estimation methods—such as hydrodynamic simulations or SAR-only

detection—lack the ability to incorporate human behavioral response or provide real-time awareness. The multi-source fusion approach developed in this dissertation produces more operationally relevant, context-aware estimations aligned with the analytical spirit of Knowledge Science.

Overall, the contributions of this dissertation demonstrate that flood estimation is not simply a technical classification problem, but a dynamic knowledge-generation process shaped by the interplay of environmental conditions and human behavior. By integrating physical sensing with social sensing and geospatial information, the dissertation forms a coherent knowledge system that advances both disaster science and Knowledge Science. This perspective highlights how multi-source fusion can transform flood estimation into a more holistic, interpretable, and resilience-oriented process that supports informed decision-making in urban disaster management.

7.3 Integration of Multi-Source Data for Urban Flood Assessment

The integration of multi-source data throughout this dissertation not only advances methodological innovation in urban flood estimation but also embodies a broader knowledge-creation process central to Knowledge Science. Across Experiments I–III, heterogeneous information—ranging from remote sensing (SAR, DEM), geospatial indices (TWI, SPI, distance metrics), GPS-based mobility patterns, and Twitter-derived public communication—was systematically transformed into actionable knowledge for disaster assessment. This progression illustrates how complex urban hazards can be understood through a cyclical interplay between sensing, interpretation, integration, and application.

At its core, the fusion of remote sensing and social sensing data represents a conceptual shift from traditional, physically oriented disaster modeling to a human-centered, systemic, and knowledge-driven paradigm. Each data type contributes differently to this epistemic ecosystem. SAR imagery provides consistent physical observations under all-weather conditions but lacks insight into human impacts. DEM-derived indices capture inherent topographic susceptibility but remain static in time. GPS data, by contrast, offer dynamic behavioral signals that reflect population movement and distress during floods, while Twitter data supply real-time narratives that reveal human perceptions, awareness, and localized incident reporting. When integrated, these multi-dimensional datasets form a richer and more coherent representation of flood phenomena than any single data source could achieve.

Within the perspective of Knowledge Science, this integration process aligns closely with the SECI model of knowledge creation. Social sensing data (GPS, Twitter) originates as tacit, collective human behavior and perception, which are transformed into explicit knowledge through spatial aggregation, NLP filtering, anomaly detection, and machine

learning. Remote sensing data, already explicit, become contextualized when fused with social indicators, forming new combinations of knowledge that would not emerge within isolated analytical domains. The resulting models—Random Forest-based flood estimation and fine-grid multi-source fusion, represent a form of internalization, where abstract patterns are converted into operational tools that can guide real-world decision-making. This cyclic interaction across sensing modalities mirrors the continuous expansion of knowledge through shared, collective interpretation.

Moreover, the dissertation advances the concept of an urban sensing ecosystem, in which physical sensors (satellites, digital elevation systems) and human sensors (citizens with smartphones, social media users) jointly construct situational awareness. This expands the traditional boundaries of disaster science by recognizing that resilient cities are not only determined by their physical infrastructure but also by their ability to continuously sense, interpret, and adapt to rapidly evolving hazards. The knowledge generated from multi-source fusion is therefore both technical, in terms of estimation accuracy, and social, in terms of capturing human experience, exposure, and response.

The integration also provides a theoretical contribution to the field of urban resilience research. The findings suggest that resilience should be understood not only as the capacity to withstand disruptions but also as the ability to leverage diverse information streams for rapid learning and adaptation. Multi-source fusion improves accuracy (as seen in Chapter 5 and Chapter 6), enhances spatial resolution, and enables near-real-time situational awareness. These outcomes demonstrate that resilience emerges from the interaction between environmental sensing, social behavior, and technology-enabled knowledge creation.

Finally, this integrative framework bridges the gap between data-driven engineering models and the human-centered epistemology of Knowledge Science. By transforming heterogeneous datasets into a unified spatial–behavioral knowledge base, the dissertation offers a novel approach to constructing actionable disaster intelligence. It positions multi-source fusion not merely as a technical procedure, but as a knowledge-creation cycle, where sensing, reasoning, and decision support are interconnected stages contributing to continuous urban learning.

In summary, the multi-source data integration framework developed in this dissertation serves as both methodological and conceptual advancement. It demonstrates how urban disaster estimation can evolve from isolated data processing into a comprehensive knowledge system—one that supports real-time monitoring, enhances resilience assessment, and strengthens the capacity of cities to understand and respond to complex hazards.

7.4 Knowledge Creation Through the Integration Process

The process of integrating multi-source data throughout this dissertation not only enhanced flood estimation accuracy but also functioned as a knowledge creation cycle in

the sense described by Knowledge Science. Rather than treating each dataset—SAR, DEM, GPS, or Twitter—as isolated informational units, the analytical framework of this research synthesizes them into a collective knowledge structure that evolves through iterative sensing, interpretation, and model refinement.

In Experiment I (Chapter 4), the discovery that GPS anomalies could reveal the timing and progression of flood events represents the initial stage of knowledge creation. Here, raw human mobility data—often dismissed as noise in traditional hazard analysis—was reinterpreted as meaningful behavioral evidence. This shift from data to insight reflects the conceptualization stage of knowledge formation, where implicit patterns in human activity become explicit indicators of disaster dynamics.

Experiment II (Chapter 5) transformed these insights into an interdisciplinary knowledge system by embedding behavioral information (GPS population density) into the physical modeling environment (SAR backscatter and DEM terrain indices). This integration reflects the combination phase of knowledge creation: disparate domains—remote sensing physics, hydrological susceptibility, and human behavioral signals—were fused into a unified explanatory framework. The Random Forest model, serving as the computational embodiment of this synthesis, operationalized how environmental factors and human responses jointly determine flood extent. This stage does not merely combine data; it generates a new understanding of how floods propagate and how populations react, forming a richer conceptual model than either domain could provide alone.

Experiment III (Chapter 6) extended the knowledge creation cycle by incorporating real-time social communication. Twitter posts, once processed through natural language techniques, provided context-specific human perceptions and on-site reports. When combined with GPS mobility and SAR-derived inundation indicators, these data enabled the construction of a fine-grid (70 m × 70 m) estimation model capable of capturing the micro-scale spatial variability of flood impacts. This stage represents internalization: the model assimilates heterogeneous data into a coherent urban-scale representation of hazard progression, enabling actionable insights such as identifying high-risk zones, understanding evacuation patterns, and recognizing where disaster communication is most active.

Through these three experiments, knowledge creation occurred through an iterative and cumulative process. Each stage added new layers of meaning: GPS data revealed behavioral signatures; SAR and DEM grounded these signatures in physical processes; Twitter data contextualized them within real-time human awareness and communication patterns. Through this continuous integration, the research transitioned from simple data analysis to a comprehensive, knowledge-driven system of urban flood assessment.

This knowledge creation process holds broader implications for urban resilience research. It demonstrates that effective disaster estimation requires the interaction of physical sensing and social sensing, not as parallel streams but as mutually reinforcing forms of knowledge. It also illustrates that resilience is not only a matter of infrastructure or hazard physics, but of collective cognition—how cities perceive, communicate, and

respond to risk. Ultimately, the integration framework developed in this dissertation can be interpreted as a knowledge creation model, where sensing (Ch.4–6 data collection), integration (multi-source fusion), and application (estimation models and spatial risk interpretation) form a continuous cycle that strengthens urban disaster preparedness.

7.5 Theoretical Contributions to Urban Resilience

The findings of this dissertation contribute to the theoretical advancement of urban resilience by reframing how disaster-related information is sensed, interpreted, and operationalized in complex urban environments. Traditional resilience theory has emphasized physical robustness, redundancy of infrastructure, or the adaptive capacity of institutions. While these dimensions remain foundational, the results from Chapters 4–6 demonstrate that resilience must also be conceptualized as an information–behavior system embedded within the urban fabric—a system in which human mobility, collective awareness, and environmental sensing interact dynamically during disasters.

A major theoretical contribution lies in demonstrating that urban resilience is not solely an infrastructural attribute but a data-responsive capability. Experiment I (Chapter 4) showed that population mobility patterns, captured through GPS data, reflect immediate behavioral adjustments to flood hazards. These responses—reduced movement during heavy rainfall, nighttime anomalies during evacuation hours—represent a form of emergent resilience that originates from human adaptation rather than engineered systems. This finding suggests that resilience theory must account for bottom-up behavioral signals as measurable, quantifiable indicators of urban coping capacity.

Experiment II (Chapter 5) further advances theoretical understanding by illustrating that resilience emerges from the interaction between physical exposure and human response. The integration of SAR backscatter, DEM terrain indices, and GPS population density into a single Random Forest model revealed that flood extent is best explained when both environmental vulnerability (e.g., low elevation, flat terrain) and behavioral reactivity (e.g., population displacement) are considered simultaneously. This dual-perspective modeling supports a more holistic conceptualization of resilience in which environmental processes and social actions are co-evolving components of the same system.

Experiment III (Chapter 6) expands this theoretical framework by demonstrating the value of information flows—specifically social media communication—as an integral part of urban resilience. The georeferenced distribution of Twitter posts provided real-time awareness signals, revealing where residents witnessed flooding, shared warnings, or sought resources. When combined with GPS and remote sensing data in a fine-grid estimation model, these communication signals contributed to more accurate and spatially coherent flood mapping. This illustrates that resilience is not merely the capacity to physically withstand a disaster but the city's ability to collectively sense, process, and circulate information during crisis.

Together, the findings across the three experiments introduce a refined understanding of urban resilience defined by three interdependent dimensions:

(a) Physical Resilience: represented by terrain vulnerability, hydrological behavior, and the physical propagation of floodwaters (SAR + DEM).

(b) Behavioral Resilience: represented by population movement, evacuation patterns, and adaptive mobility (GPS).

(c)Information Resilience: represented by communication flows, situational awareness, and knowledge sharing (Twitter).

This tri-dimensional model reflects a shift from viewing resilience as static robustness toward understanding it as a dynamic, knowledge-driven system. The fusion framework developed in this dissertation shows that cities become resilient not only through strong infrastructure but through integrated sensing—physical, behavioral, and informational—that allows for timely detection, accurate estimation, and coordinated response.

Finally, this research contributes theoretically by positioning multi-source sensing as a core mechanism of resilience emergence. The "Sensing → Integration → Application" cycle demonstrates how resilience can be conceptualized as a continuous process of knowledge creation, where new information from diverse domains is assimilated into decision-making structures. This conceptualization aligns with modern resilience theories that emphasize learning, adaptation, and information flows as critical determinants of a city's ability to cope with environmental hazards.

In summary, this dissertation advances the theoretical landscape of urban resilience by demonstrating that resilience is fundamentally an information–behavior–environment system, where human activity, communication patterns, and physical hazard dynamics co-produce the city's ability to withstand and recover from disasters.

7.6 Practical Implications for Urban Planning and Disaster Management

The results of this dissertation provide several practical implications for urban planning, emergency response, and the development of data-driven disaster-management systems. By systematically integrating remote sensing, social sensing, and social media data, the proposed framework demonstrates a feasible pathway toward building operational early-warning and real-time monitoring mechanisms in urban environments.

First, the findings highlight the operational value of population mobility data for disaster response. Experiment I (Chapter 4) revealed that GPS-derived mobility anomalies were strongly aligned with the timing and spatial distribution of the actual flood event. This suggests that large-scale mobility datasets can serve as a real-time behavioral sensor that reflects population distress, evacuation tendencies, and disruption of daily routines. For emergency managers, this offers a valuable supplementary indicator to traditional physical hazard monitoring. In practice, continuous monitoring of mobility

anomalies could support early detection of emerging hazards, validation of risk alerts, and assessment of evacuation effectiveness.

Second, the fusion of physical hazard indicators and human-behavior data enhances the accuracy of flood mapping. Experiment II (Chapter 5) demonstrated that integrating SAR backscatter, DEM-derived terrain metrics, and GPS population density significantly improved flood-extent estimation. The comparison between RF-only-physical and RF-multi-source models clearly showed that social-sensing inputs reduce false positives and sharpen the boundaries of estimated inundation. For urban planners and hydrological modelers, this reinforces the importance of combining human-sensing and physical-sensing information, especially in complex urban landscapes where hydrodynamics interacts closely with population exposure. Such multi-source models can be incorporated into rapid-damage assessments, post-disaster mapping workflows, and the validation of flood-risk zoning maps.

Third, the fine-grid estimation model using Twitter and GPS data provides a new method for real-time situational awareness. Experiment III (Chapter 6) demonstrated that Twitter data—after NLP-based filtering—captures real-time, human-reported information about flooding conditions. When combined with GPS mobility and traditional geospatial datasets in a 70 m × 70 m grid, the resulting model delivered higher estimation accuracy and more spatially coherent results compared to models without social media input. This suggests that social media, when properly processed, can serve as a critical component of real-time disaster intelligence systems. Emergency-response agencies can exploit such data to identify hotspots of concern, prioritize rescue operations, and monitor hazard evolution from the perspective of affected communities.

Fourth, the dissertation offers a replicable framework for developing data-driven resilience tools. The three experiments collectively form a scalable methodological pipeline—from coarse-grid anomaly detection to fine-grid multi-source estimation—that can be adapted to other hazards such as landslides, heatwaves, or earthquakes. For policymakers and municipal governments, the proposed "Sensing → Integration → Application" workflow provides a structured pathway for operationalizing smart-resilience strategies, enabling cities to harness the sensing capabilities already embedded in everyday technologies such as mobile phones and social media platforms.

Fifth, the findings underscore the need for institutional cooperation and data-sharing mechanisms. The practical implementation of these models requires timely access to mobile GPS records, SAR imagery, administrative boundaries, infrastructure layers, and social media streams. Establishing cross-sectoral partnerships among governmental agencies, telecom operators, satellite organizations, and academic institutions is therefore essential. This cooperation not only enhances the accuracy of disaster monitoring but also accelerates the transformation of research outputs into actionable policy tools. The dissertation highlights that resilience is not only a technical goal but also an institutional process requiring coordinated governance and a supportive data ecosystem.

Finally, the research provides evidence to support community-centered disaster

planning. By demonstrating that human behavioral signals (GPS) and public communication activities (Twitter) improve estimation accuracy, the dissertation suggests that residents are active participants in the sensing and reporting of disasters. Encouraging public engagement, improving risk communication channels, and incorporating citizen-generated data into early-warning systems can strengthen community preparedness and empower residents as partners in urban resilience.

7.7 Data-Driven Resilient City Model

The findings of this dissertation collectively support the development of a *data-driven resilient city model*, in which multi-source sensing, data integration, and estimation-based analytics form an interconnected system for understanding and managing urban flood impacts. Unlike traditional disaster frameworks that rely primarily on physical measurements or engineering resilience perspectives, the proposed model emphasizes the synergy between physical sensing and human behavioral sensing, reflecting a contemporary shift toward information-driven urban resilience. This model emerges organically from the progression of Experiments I to III, each of which expands the city's capacity to perceive, interpret, and respond to disaster conditions.

At the foundational level, the model begins with a multi-source sensing layer that captures both environmental processes and population dynamics. Remote sensing data, including SAR, DEM, terrain indices, and distance-to-infrastructure variables, provide objective and spatially continuous observations of flood-prone areas. Social sensing data, derived from GPS mobility and Twitter messages, complement these physical observations by revealing how residents perceive, experience, and react to emerging hazards. Together, these datasets form a richer and more nuanced sensing system in which physical and social dimensions of disaster unfold simultaneously.

Building upon this sensing foundation, the second layer of the framework integrates heterogeneous data into a coherent knowledge base. Through machine learning, spatial alignment, natural language processing, kernel density analysis, and feature engineering, raw observations are transformed into interpretable information about hazard extent, exposure, and behavioral response. This layer reflects the Knowledge Science principle of converting data into information and information into actionable knowledge. The Random Forest model, in particular, functions as a knowledge integration engine: it merges physical and social variables, identifies their relative contributions, and uncovers hidden relationships that describe how floods develop in urban environments.

The third layer of the framework focuses on estimation modeling and real-time assessment. Across different spatial scales—from the 500 m grids used to detect mobility anomalies, to the multi-source RF flood extent estimation, to the 70 m fine-grid urban estimation—the framework demonstrates how increasing data richness enhances the resolution and practical relevance of estimation outputs. The outputs evolve from identifying abnormal mobility patterns, to estimating flood extent from combined SAR

and GPS data, to generating near-real-time urban flood maps using Twitter and GPS information. Through this multi-scalar progression, the model becomes not only more accurate but also more closely aligned with the operational needs of urban disaster management.

The final layer of the model embeds these estimation insights into resilience-oriented decision-making. By capturing population responses alongside physical hazard patterns, the model supports real-time situational awareness, helps identify priority zones, and highlights vulnerable infrastructure. Moreover, because it incorporates information flow and collective sensing, the framework expands the concept of resilience beyond physical robustness. It emphasizes information responsiveness, adaptive behavior, and the capacity of communities and institutions to learn from and react to dynamic disaster conditions. This perspective aligns with contemporary urban resilience theories and positions information as a central determinant of a city's ability to withstand and recover from extreme events.

Altogether, the data-driven resilient city model proposed in this dissertation bridges remote sensing science, social sensing analytics, machine learning, and Knowledge Science. It offers a unified conceptual structure in which sensing, integration, and application form a continuous knowledge creation cycle, enabling cities to transition from reactive disaster response to proactive, intelligence-driven resilience management. This framework constitutes a key theoretical contribution of the thesis, demonstrating how multi-source data integration can support more informed, adaptive, and human-centered approaches to urban disaster estimation and resilience assessment.

7.8 Summary

This chapter presented a comprehensive synthesis of the methodological, theoretical, and practical insights generated through the multi-source flood estimation framework developed in this dissertation. Rather than treating the three experiments as independent components, the discussion highlighted how the integration of remote sensing, social sensing, and geospatial information forms a unified knowledge system for understanding and estimating flood processes in Nagano City.

A central conclusion is that integrating heterogeneous data sources allows flood-related phenomena to be captured with greater precision and timeliness. Physical sensing data such as SAR backscatter, DEM-based terrain metrics, and distance indicators offered stable and spatially continuous representations of environmental conditions. Yet, these variables alone could not represent the rapid behavioral changes that occur during urban flood events. Social sensing data, including GPS mobility patterns and geotagged Twitter communications, filled this gap by offering real-time and behaviorally meaningful signals. When combined, these data streams complemented one another and produced a more comprehensive, multi-dimensional representation of the disaster.

From the perspective of Knowledge Science, the multi-source integration framework

can be understood as an urban “sensing ecosystem” in which satellites, mobile devices, and social media platforms collectively function as distributed sensors. Through processes of data transformation, feature extraction, and machine-learning–based modeling, raw observations were converted into structured information and eventually into actionable knowledge. This aligns with knowledge creation theories that emphasize the movement from data to information to shared understanding. By embedding human behavioral signals alongside physical hazard indicators, the framework reframes flood estimation as a knowledge-driven process rather than a purely technical computation.

Ultimately, this chapter underscored that integrating remote sensing and social sensing is not simply a methodological improvement but a conceptual advancement for urban disaster research. It supports real-time and adaptive disaster management, aligns with the principles of resilient and smart cities, and demonstrates how diverse data streams can be merged to generate more accurate, interpretable, and operationally meaningful assessments. In this sense, multi-source data integration provides a critical foundation for developing next-generation urban disaster estimation and resilience-enhancement systems.

Chapter 8 Conclusion

8.1 Overall Conclusion

This dissertation developed a comprehensive multi-source framework for urban flood estimation and assessment, integrating remote sensing, geographic information, and social sensing into a unified analytical system. The work was motivated by the limitations of traditional disaster analysis methods, which rely predominantly on physical observations and often overlook the behavioral dynamics that shape disaster progression in urban environments. To address these gaps, the study combined Synthetic Aperture Radar (SAR), Digital Elevation Model (DEM) data, terrain indices such as TWI and SPI, GPS-based population mobility, and Twitter-derived information to examine the 2019 Typhoon Hagibis flood in Nagano City, Japan.

A key contribution of the dissertation lies in the sequential structure of the three experiments, which collectively form an integrated knowledge-building process. Experiment I established the behavioral sensing foundation by demonstrating that GPS mobility anomalies correspond to both the timing and spatial distribution of flood impacts. This confirmed that human mobility patterns can serve as meaningful indicators of hazard exposure and public response. Experiment II progressed from sensing to integration by incorporating GPS mobility with SAR and topographic variables in a Random Forest model. The results showed that combining behavioral and physical indicators enhances estimation accuracy and spatial coherence. Experiment III further advanced this approach by incorporating Twitter-derived information and applying the integrated model at a fine spatial resolution. This experiment demonstrated the feasibility of near real-time urban flood estimation and underscored the value of multi-source fusion for operational monitoring.

Through this stepwise development—from identifying behavioral signatures, to embedding them in a hybrid physical–social model, to operationalizing the model at an urban scale—the dissertation demonstrates how each experiment builds upon the previous one. Rather than functioning as independent analyses, the three experiments work together to form a cohesive system for understanding and estimating flood impacts.

Compared with conventional flood estimation approaches that rely solely on satellite remote sensing or terrain-based hydrological indicators, the proposed framework offers several clear advantages. SAR-only methods are effective for detecting surface water but often suffer from misclassification in complex urban environments, particularly over road networks and built-up areas. DEM- or hydrodynamic-based approaches, while physically grounded, typically require extensive parameterization and are limited in their ability to provide timely assessments immediately after a disaster. By contrast, the proposed approach integrates social sensing data that capture human behavioral responses, enabling the identification and correction of flood misdetections that cannot be resolved through

physical sensing alone. As demonstrated in this study, incorporating GPS mobility and Twitter information reduces false positives, improves spatial coherence, and enhances robustness in urban flood estimation.

The integrated modeling results show substantial improvements in delineating inundated areas, reducing false positives, and capturing population responses during the disaster. These findings highlight the importance of human-centered data sources, particularly in dense urban environments where physical indicators alone may be insufficient. The research demonstrates that multi-source data fusion effectively links physical hazard processes with behavioral responses, providing a foundation for knowledge-informed urban disaster monitoring.

In conclusion, this dissertation offers conceptual and practical contributions to data-driven resilient city analysis. It emphasizes that robust urban flood estimation requires not only remote sensing-based observations but also the integration of real-time behavioral and communication patterns. The multi-source fusion framework developed here provides a scalable and transferable methodology capable of supporting early detection, spatially detailed distribution mapping, and post-event assessment. More broadly, the study illustrates how Knowledge Science can guide the development of analytical systems that transform diverse data sources into interpretable, actionable insights for resilience-oriented decision-making.

8.2 Limitations and Future Directions

Despite the demonstrated effectiveness of multi-source data integration, several limitations remain that present opportunities for future research.

(a) Temporal and data availability constraints

Although social sensing data enhances the responsiveness of flood estimation, neither satellite remote sensing nor social sensing data can provide fully real-time observations. SAR imagery is constrained by satellite revisit cycles, such as Sentinel-1's acquisition schedule, which may not coincide with the peak timing of fast-onset flood events. Similarly, GPS mobility and Twitter data are subject to delays related to data access, processing, and platform availability. During severe floods, power outages or communication disruptions may further reduce data availability, resulting in spatial or temporal gaps. Consequently, the proposed framework should be regarded as supporting near–real-time flood estimation rather than instantaneous monitoring. Nevertheless, the results demonstrate that integrating physical and behavioral data substantially reduces information gaps during the post-event information blackout period, offering timely situational awareness for emergency response.

(b) Uneven spatial distribution of social sensing data

GPS and Twitter data are more abundant in densely populated areas and commercial districts, resulting in higher estimation accuracy in urban cores but comparatively weaker performance in rural or sparsely populated zones. This spatial imbalance can cause models to underestimate flood extent in low-population regions, where human behavioral

signals are limited. Future approaches may incorporate additional data sources such as LINE messages, vehicle trajectory logs, or passive mobile network data to compensate for this bias.

(c) Lack of physical hydrodynamic modelling

The models developed in this dissertation are data-driven and do not incorporate hydrodynamic simulations or rainfall-runoff relationships. As a result, although the models identify flood patterns effectively, they do not simulate water depth, flow pathways, or flood propagation mechanisms. Integrating hydrodynamic models with social sensing–based indicators would enhance both interpretability and the robustness of flood extent estimation.

(d) Generalizability across hazards and regions

The framework was validated using a single urban flood event in Nagano City. Although the methods are theoretically transferable, performance may vary in regions with different urban morphology, population structures, or communication behaviors. Future research could apply the framework to multi-hazard scenarios (e.g., earthquakes, landslides, compound disasters) and different countries to validate scalability and robustness.

Building on the findings of this dissertation, several promising research directions emerge:

(a) Real-time data ingestion and automated models: Developing streaming pipelines for Twitter and mobile network data, combined with cloud-based SAR preprocessing, would enable near-real-time flood mapping.

(b) Deep learning and multimodal fusion: Integrating convolutional neural networks (CNNs) for SAR image processing and transformer-based models for social media text analysis could improve spatial feature extraction and semantic understanding of public communication during disasters.

(c) Integration with hydrodynamic simulations: Coupling machine learning models with physics-based flood models would support more accurate estimations of flood depth, velocity, and impact mechanisms.

(d) Smart city and IoT sensor integration: Future models may incorporate river gauges, CCTV imagery, road sensors, and volunteer reporting systems to build a holistic “urban sensing ecosystem.”

(e) Urban resilience assessment: The fine-grid model developed in Chapter 6 can be extended toward long-term resilience evaluations, including evacuation simulations, infrastructure vulnerability mapping, and adaptive planning for climate change.

Bibliography

[1] C. Wasko, R. Nathan, L. Stein, and D. O'Shea, "Evidence of shorter more extreme rainfalls and increased flood variability under climate change," *J. Hydrol.*, vol. 603, p. 126994, Dec. 2021, doi: 10.1016/j.jhydrol.2021.126994.

[2] D.-L. Zhang, "Rapid urbanization and more extreme rainfall events," *Sci. Bull.*, vol. 65, no. 7, pp. 516–518, Apr. 2020, doi: 10.1016/j.scib.2020.02.002.

[3] T. Gudiyangada Nachappa, S. Tavakkoli Piralilou, K. Gholamnia, O. Ghorbanzadeh, O. Rahmati, and T. Blaschke, "Flood susceptibility mapping with machine learning, multi-criteria decision analysis and ensemble using Dempster Shafer Theory," *J. Hydrol.*, vol. 590, p. 125275, Nov. 2020, doi: 10.1016/j.jhydrol.2020.125275.

[4] Y. Wang, Z. Fang, H. Hong, and L. Peng, "Flood susceptibility mapping using convolutional neural network frameworks," *J. Hydrol.*, vol. 582, p. 124482, Mar. 2020, doi: 10.1016/j.jhydrol.2019.124482.

[5] G. D. Bathrellos, H. D. Skilodimou, K. Soukis, and E. Koskeridou, "Temporal and Spatial Analysis of Flood Occurrences in the Drainage Basin of Pinios River (Thessaly, Central Greece)," *Land*, vol. 7, no. 3, p. 106, Sept. 2018, doi: 10.3390/land7030106.

[6] Z. Kundzewicz, B. Su, Y. Wang, J. Xia, J. Huang, and T. Jiang, "Flood risk and its reduction in China," *Adv. Water Resour.*, vol. 130, pp. 37–45, Aug. 2019, doi: 10.1016/j.advwatres.2019.05.020.

[7] G. Zhao, B. Pang, Z. Xu, J. Yue, and T. Tu, "Mapping flood susceptibility in mountainous areas on a national scale in China," *Sci. Total Environ.*, vol. 615, pp. 1133–1142, Feb. 2018, doi: 10.1016/j.scitotenv.2017.10.037.

[8] S. H. Mahmoud and T. Y. Gan, "Urbanization and climate change implications in flood risk management: Developing an efficient decision support system for flood susceptibility mapping," *Sci. Total Environ.*, vol. 636, pp. 152–167, Sept. 2018, doi: 10.1016/j.scitotenv.2018.04.282.

[9] G. Ren *et al.*, "Historical and recent change in extreme climate over East Asia," *Clim. Change*, vol. 168, no. 3, p. 22, Oct. 2021, doi: 10.1007/s10584-021-03227-5.

[10] H. Tsuguti, N. Seino, H. Kawase, Y. Imada, T. Nakaegawa, and I. Takayabu, "Meteorological overview and mesoscale characteristics of the Heavy Rain Event of July 2018 in Japan," *Landslides*, vol. 16, no. 2, pp. 363–371, Feb. 2019, doi:

10.1007/s10346-018-1098-6.

[11] L. Moya, E. Mas, and S. Koshimura, “Learning from the 2018 Western Japan Heavy Rains to Detect Floods during the 2019 Hagibis Typhoon,” *Remote Sens.*, vol. 12, no. 14, p. 2244, Jan. 2020, doi: 10.3390/rs12142244.

[12] Y. Kanani-Sadat, R. Arabsheibani, F. Karimipour, and M. Nasseri, “A new approach to flood susceptibility assessment in data-scarce and ungauged regions based on GIS-based hybrid multi criteria decision-making method,” *J. Hydrol.*, vol. 572, pp. 17–31, May 2019, doi: 10.1016/j.jhydrol.2019.02.034.

[13] M. Sakurai and Y. Murayama, “Information technologies and disaster management – Benefits and issues -,” *Prog. Disaster Sci.*, vol. 2, p. 100012, July 2019, doi: 10.1016/j.pdisas.2019.100012.

[14] M. Ishiwatari, “Institutional Coordination of Disaster Management: Engaging National and Local Governments in Japan,” *Nat. Hazards Rev.*, vol. 22, no. 1, p. 04020059, Feb. 2021, doi: 10.1061/(ASCE)NH.1527-6996.0000423.

[15] I. H. Sawalha, “A contemporary perspective on the disaster management cycle,” *Foresight*, vol. 22, no. 4, pp. 469–482, May 2020, doi: 10.1108/FS-11-2019-0097.

[16] E. Dodangeh *et al.*, “Integrated machine learning methods with resampling algorithms for flood susceptibility prediction,” *Sci. Total Environ.*, vol. 705, p. 135983, Feb. 2020, doi: 10.1016/j.scitotenv.2019.135983.

[17] Z. Tang, H. Zhang, S. Yi, and Y. Xiao, “Assessment of flood susceptible areas using spatially explicit, probabilistic multi-criteria decision analysis,” *J. Hydrol.*, vol. 558, pp. 144–158, Mar. 2018, doi: 10.1016/j.jhydrol.2018.01.033.

[18] K. Shi, X. Peng, H. Lu, Y. Zhu, and Z. Niu, “Application of Social Sensors in Natural Disasters Emergency Management: A Review,” *IEEE Trans. Comput. Soc. Syst.*, vol. 10, no. 6, pp. 3143–3158, Dec. 2023, doi: 10.1109/TCSS.2022.3211552.

[19] H. S. Munawar, A. W. A. Hammad, and S. T. Waller, “Remote Sensing Methods for Flood Prediction: A Review,” *Sensors*, vol. 22, no. 3, p. 960, Jan. 2022, doi: 10.3390/s22030960.

[20] V. Pekar, J. Binner, H. Najafi, C. Hale, and V. Schmidt, “Early detection of heterogeneous disaster events using social media,” *J. Assoc. Inf. Sci. Technol.*, vol. 71, no. 1, pp. 43–54, 2020, doi: 10.1002/asi.24208.

[21] M. Ren, Z. Zhang, J. Zhang, and L. Mora, “Understanding the Use of Heterogenous Data in Tackling Urban Flooding: An Integrative Literature Review,” *Water*, vol. 14, no. 14, p. 2160, Jan. 2022, doi: 10.3390/w14142160.

[22] R. I. Ogie, R. J. Clarke, H. Forehead, and P. Perez, “Crowdsourced social media data for disaster management: Lessons from the PetaJakarta.org project,” *Comput.*

Environ. Urban Syst., vol. 73, pp. 108–117, Jan. 2019, doi: 10.1016/j.compenvurbsys.2018.09.002.

[23] S. A. Shah, D. Z. Seker, M. M. Rathore, S. Hameed, S. Ben Yahia, and D. Draheim, "Towards Disaster Resilient Smart Cities: Can Internet of Things and Big Data Analytics Be the Game Changers?," *IEEE Access*, vol. 7, pp. 91885–91903, 2019, doi: 10.1109/ACCESS.2019.2928233.

[24] V. Kumar, K. V. Sharma, T. Caloiero, D. J. Mehta, and K. Singh, "Comprehensive Overview of Flood Modeling Approaches: A Review of Recent Advances," *Hydrology*, vol. 10, no. 7, p. 141, July 2023, doi: 10.3390/hydrology10070141.

[25] A. Mosavi, P. Ozturk, and K. Chau, "Flood Prediction Using Machine Learning Models: Literature Review," *Water*, vol. 10, no. 11, p. 1536, Nov. 2018, doi: 10.3390/w10111536.

[26] B. Jamali, R. Löwe, P. M. Bach, C. Urich, K. Arnbjerg-Nielsen, and A. Deletic, "A rapid urban flood inundation and damage assessment model," *J. Hydrol.*, vol. 564, pp. 1085–1098, Sept. 2018, doi: 10.1016/j.jhydrol.2018.07.064.

[27] "Near-real-time non-obstructed flood inundation mapping using synthetic aperture radar," *Remote Sens. Environ.*, vol. 221, pp. 302–315, Feb. 2019, doi: 10.1016/j.rse.2018.11.008.

[28] T. Yang *et al.*, "Evaluation and machine learning improvement of global hydrological model-based flood simulations," *Environ. Res. Lett.*, vol. 14, no. 11, p. 114027, Nov. 2019, doi: 10.1088/1748-9326/ab4d5e.

[29] A. Tarpanelli, A. C. Mondini, and S. Camici, "Effectiveness of Sentinel-1 and Sentinel-2 for flood detection assessment in Europe," *Nat. Hazards Earth Syst. Sci.*, vol. 22, no. 8, pp. 2473–2489, Aug. 2022, doi: 10.5194/nhess-22-2473-2022.

[30] A. Matsuki and M. Hatayama, "Identification of issues in disaster response to flooding, focusing on the time continuity between residents' evacuation and rescue activities," *Int. J. Disaster Risk Reduct.*, vol. 95, p. 103841, Sept. 2023, doi: 10.1016/j.ijdrr.2023.103841.

[31] M. Ouyang, S. Kotsuki, Y. Ito, and T. Tokunaga, "Employment of hydraulic model and social media data for flood hazard assessment in an urban city," *J. Hydrol. Reg. Stud.*, vol. 44, p. 101261, Dec. 2022, doi: 10.1016/j.ejrh.2022.101261.

[32] J. A. de Bruijn, H. de Moel, B. Jongman, M. C. de Ruiter, J. Wagemaker, and J. C. J. H. Aerts, "A global database of historic and real-time flood events based on social media," *Sci. Data*, vol. 6, no. 1, p. 311, Dec. 2019, doi: 10.1038/s41597-019-0326-9.

[33] R. Metulini and M. Carpita, "Modeling and forecasting traffic flows with mobile

phone big data in flooding risk areas to support a data-driven decision making," *Ann. Oper. Res.*, vol. 342, no. 3, pp. 1629–1654, Nov. 2024, doi: 10.1007/s10479-023-05195-8.

[34] Z. Li, C. Wang, C. T. Emrich, and D. Guo, "A novel approach to leveraging social media for rapid flood mapping: a case study of the 2015 South Carolina floods," *Cartogr. Geogr. Inf. Sci.*, vol. 45, no. 2, pp. 97–110, Mar. 2018, doi: 10.1080/15230406.2016.1271356.

[35] Q. Wei, H. Zhang, Y. Chen, Y. Xie, H. Yin, and Z. Xu, "City scale urban flooding risk assessment using multi-source data and machine learning approach," *J. Hydrol.*, vol. 651, p. 132626, Apr. 2025, doi: 10.1016/j.jhydrol.2024.132626.

[36] B. Zhao *et al.*, "Flood inundation monitoring using multi-source satellite imagery: a knowledge transfer strategy for heterogeneous image change detection," *Remote Sens. Environ.*, vol. 314, p. 114373, Dec. 2024, doi: 10.1016/j.rse.2024.114373.

[37] D. Amitrano, G. Di Martino, A. Di Simone, and P. Imperatore, "Flood Detection with SAR: A Review of Techniques and Datasets," *Remote Sens.*, vol. 16, no. 4, p. 656, Jan. 2024, doi: 10.3390/rs16040656.

[38] M. Muthusamy, M. R. Casado, D. Butler, and P. Leinster, "Understanding the effects of Digital Elevation Model resolution in urban fluvial flood modelling," *J. Hydrol.*, vol. 596, p. 126088, May 2021, doi: 10.1016/j.jhydrol.2021.126088.

[39] Y. Li, F. B. Osei, T. Hu, and A. Stein, "Urban flood susceptibility mapping based on social media data in Chengdu city, China," *Sustain. Cities Soc.*, vol. 88, p. 104307, Jan. 2023, doi: 10.1016/j.scs.2022.104307.

[40] D. T. Bulti and B. G. Abebe, "A review of flood modeling methods for urban pluvial flood application," *Model. Earth Syst. Environ.*, vol. 6, no. 3, pp. 1293–1302, Sept. 2020, doi: 10.1007/s40808-020-00803-z.

[41] T. H. Assumpção, I. Popescu, A. Jonoski, and D. P. Solomatine, "Citizen observations contributing to flood modelling: opportunities and challenges," *Hydrol. Earth Syst. Sci.*, vol. 22, no. 2, pp. 1473–1489, Feb. 2018, doi: 10.5194/hess-22-1473-2018.

[42] P. Luo *et al.*, "Urban flood numerical simulation: Research, methods and future perspectives," *Environ. Model. Softw.*, vol. 156, p. 105478, Oct. 2022, doi: 10.1016/j.envsoft.2022.105478.

[43] Y. Wang, A. S. Chen, G. Fu, S. Djordjević, C. Zhang, and D. A. Savić, "An integrated framework for high-resolution urban flood modelling considering multiple information sources and urban features," *Environ. Model. Softw.*, vol. 107, pp. 85–95, Sept. 2018, doi: 10.1016/j.envsoft.2018.06.010.

[44] L. Goodarzi, M. E. Banihabib, and A. Roozbahani, "A decision-making model for flood warning system based on ensemble forecasts," *J. Hydrol.*, vol. 573, pp. 207–219, June 2019, doi: 10.1016/j.jhydrol.2019.03.040.

[45] J.-M. Cariolet, M. Vuillet, and Y. Diab, "Mapping urban resilience to disasters – A review," *Sustain. Cities Soc.*, vol. 51, p. 101746, Nov. 2019, doi: 10.1016/j.scs.2019.101746.

[46] Y. Liu, Y. Jiang, and Y. Deng, "A framework for urban flood resilience assessment: Integrating multidimensional indicators and dynamic adaptation strategies," *J. Environ. Manage.*, vol. 392, p. 126841, Sept. 2025, doi: 10.1016/j.jenvman.2025.126841.

[47] X. Wang and Y.-M. Shen, "A Framework of Dependence Modeling and Evaluation System for Compound Flood Events," *Water Resour. Res.*, vol. 59, no. 8, p. e2023WR034718, 2023, doi: 10.1029/2023WR034718.

[48] Y. Yang, N. Ohira, and H. Gokon, "Small-grid urban flood prediction model using Twitter data and population GPS data - an example of the 2019 Nagano city flood," *Prog. Disaster Sci.*, vol. 24, p. 100385, Dec. 2024, doi: 10.1016/j.pdisas.2024.100385.

[49] F. Yavari, S. A. Salehi Neyshabouri, J. Yazdi, A. Molajou, and A. Brysiewicz, "A Novel Framework for Urban Flood damage Assessment," *Water Resour. Manag.*, vol. 36, no. 6, pp. 1991–2011, Apr. 2022, doi: 10.1007/s11269-022-03122-3.

[50] "Urbanization and climate change: Insights from eco-hydrological diagnostics," *Sci. Total Environ.*, vol. 647, pp. 29–36, Jan. 2019, doi: 10.1016/j.scitotenv.2018.07.319.

[51] M. Wang *et al.*, "Assessing urban flooding risk in response to climate change and urbanization based on shared socio-economic pathways," *Sci. Total Environ.*, vol. 880, p. 163470, July 2023, doi: 10.1016/j.scitotenv.2023.163470.

[52] A. Shadmehri Toosi, G. H. Calbimonte, H. Nouri, and S. Alaghmand, "River basin-scale flood hazard assessment using a modified multi-criteria decision analysis approach: A case study," *J. Hydrol.*, vol. 574, pp. 660–671, July 2019, doi: 10.1016/j.jhydrol.2019.04.072.

[53] S. Ohtsuka, Y. Sato, T. Yoshikawa, T. Sugii, T. Kodaka, and K. Maeda, "Levee damage and revetment erosion by the 2019 Typhoon Hagibis in the Chikuma River, Japan," *Soils Found.*, vol. 61, no. 4, pp. 1172–1188, Aug. 2021, doi: 10.1016/j.sandf.2021.05.009.

[54] M. Higashino and H. G. Stefan, "Variability and change of precipitation and flood discharge in a Japanese river basin," *J. Hydrol. Reg. Stud.*, vol. 21, pp. 68–79, Feb.

2019, doi: 10.1016/j.ejrh.2018.12.003.

[55] B. R. Rosenzweig *et al.*, "The Value of Urban Flood Modeling," *Earths Future*, vol. 9, no. 1, p. e2020EF001739, 2021, doi: 10.1029/2020EF001739.

[56] C. Richert, K. Erdlenbruch, and F. Grelot, "The impact of flood management policies on individual adaptation actions: Insights from a French case study," *Ecol. Econ.*, vol. 165, p. 106387, Nov. 2019, doi: 10.1016/j.ecolecon.2019.106387.

[57] M. Moradzadeh and M. Ahmadi, "Unraveling the interplay of human decisions and flood risk: An agent-based modeling approach," *Int. J. Disaster Risk Reduct.*, vol. 107, p. 104486, June 2024, doi: 10.1016/j.ijdrr.2024.104486.

[58] W. G. Namara, T. A. Damisse, and F. G. Tufa, "Application of HEC-RAS and HEC-GeoRAS model for Flood Inundation Mapping, the case of Awash Bello Flood Plain, Upper Awash River Basin, Oromiya Regional State, Ethiopia," *Model. Earth Syst. Environ.*, vol. 8, no. 2, pp. 1449–1460, June 2022, doi: 10.1007/s40808-021-01166-9.

[59] X. Zhang, A. Kang, X. Lei, and H. Wang, "Urban drainage efficiency evaluation and flood simulation using integrated SWMM and terrain structural analysis," *Sci. Total Environ.*, vol. 957, p. 177442, Dec. 2024, doi: 10.1016/j.scitotenv.2024.177442.

[60] J. Li, B. Zhang, Y. Li, and H. Li, "Simulation of Rain Garden Effects in Urbanized Area Based on Mike Flood," *Water*, vol. 10, no. 7, p. 860, July 2018, doi: 10.3390/w10070860.

[61] C. Hu, J. Xia, D. She, Z. Song, Y. Zhang, and S. Hong, "A new urban hydrological model considering various land covers for flood simulation," *J. Hydrol.*, vol. 603, p. 126833, Dec. 2021, doi: 10.1016/j.jhydrol.2021.126833.

[62] X. Ren, Z. Guo, Q. Li, Z. Zhou, X. Chen, and J. Liao, "Integrating hydrological and economic assessments of soil infiltration enhancement for urban flood resilience," *Sci. Total Environ.*, vol. 1000, p. 180383, Oct. 2025, doi: 10.1016/j.scitotenv.2025.180383.

[63] X. Zhou, W. Ma, W. Echizenya, and D. Yamazaki, "The uncertainty of flood frequency analyses in hydrodynamic model simulations," *Nat. Hazards Earth Syst. Sci.*, vol. 21, no. 3, pp. 1071–1085, Mar. 2021, doi: 10.5194/nhess-21-1071-2021.

[64] X. Ming, Q. Liang, X. Xia, D. Li, and H. J. Fowler, "Real-Time Flood Forecasting Based on a High-Performance 2-D Hydrodynamic Model and Numerical Weather Predictions," *Water Resour. Res.*, vol. 56, no. 7, p. e2019WR025583, 2020, doi: 10.1029/2019WR025583.

[65] A. Khoshkonesh, R. Nazari, M. R. Nikoo, and M. Karimi, "Enhancing flood risk

assessment in urban areas by integrating hydrodynamic models and machine learning techniques," *Sci. Total Environ.*, vol. 952, p. 175859, Nov. 2024, doi: 10.1016/j.scitotenv.2024.175859.

[66] "Towards better flood risk management: Assessing flood risk and investigating the potential mechanism based on machine learning models," *J. Environ. Manage.*, vol. 293, p. 112810, Sept. 2021, doi: 10.1016/j.jenvman.2021.112810.

[67] A. Domeneghetti, G. J.-P. Schumann, and A. Tarpanelli, "Preface: Remote Sensing for Flood Mapping and Monitoring of Flood Dynamics," *Remote Sens.*, vol. 11, no. 8, p. 943, Jan. 2019, doi: 10.3390/rs11080943.

[68] E. Hamidi, B. G. Peter, D. F. Muñoz, H. Moftakhari, and H. Moradkhani, "Fast Flood Extent Monitoring With SAR Change Detection Using Google Earth Engine," *IEEE Trans. Geosci. Remote Sens.*, vol. 61, pp. 1–19, 2023, doi: 10.1109/TGRS.2023.3240097.

[69] A. Dasgupta, S. Grimaldi, R. A. A. J. Ramsankaran, V. R. N. Pauwels, and J. P. Walker, "Towards operational SAR-based flood mapping using neuro-fuzzy texture-based approaches," *Remote Sens. Environ.*, vol. 215, pp. 313–329, Sept. 2018, doi: 10.1016/j.rse.2018.06.019.

[70] T. McCormack, J. Campanyà, and O. Naughton, "A methodology for mapping annual flood extent using multi-temporal Sentinel-1 imagery," *Remote Sens. Environ.*, vol. 282, p. 113273, Dec. 2022, doi: 10.1016/j.rse.2022.113273.

[71] E. Ferrentino *et al.*, "On the Combination of Dual-Polarization Sentinel-1 Ascending/Descending Orbiting Passes to Estimate Damage Due to the 2016 Central Italy Earthquake," *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.*, vol. 15, pp. 9509–9518, 2022, doi: 10.1109/JSTARS.2022.3217889.

[72] L. Yan, J. Wang, and D. Shao, "Glacier Mass Balance in the Manas River Using Ascending and Descending Pass of Sentinel 1A/1B Data and SRTM DEM," *Remote Sens.*, vol. 14, no. 6, p. 1506, Jan. 2022, doi: 10.3390/rs14061506.

[73] X. Zhu, H. Guo, and J. J. Huang, "Urban flood susceptibility mapping using remote sensing, social sensing and an ensemble machine learning model," *Sustain. Cities Soc.*, vol. 108, p. 105508, Aug. 2024, doi: 10.1016/j.scs.2024.105508.

[74] M. Singha *et al.*, "Identifying floods and flood-affected paddy rice fields in Bangladesh based on Sentinel-1 imagery and Google Earth Engine," *ISPRS J. Photogramm. Remote Sens.*, vol. 166, pp. 278–293, Aug. 2020, doi: 10.1016/j.isprsjprs.2020.06.011.

[75] X. Zheng, D. R. Maidment, D. G. Tarboton, Y. Y. Liu, and P. Passalacqua, "GeoFlood: Large-Scale Flood Inundation Mapping Based on High-Resolution

Terrain Analysis," *Water Resour. Res.*, vol. 54, no. 12, p. 10,013-10,033, 2018, doi: 10.1029/2018WR023457.

[76] K. Khosravi *et al.*, "A comparative assessment of flood susceptibility modeling using Multi-Criteria Decision-Making Analysis and Machine Learning Methods," *J. Hydrol.*, vol. 573, pp. 311–323, June 2019, doi: 10.1016/j.jhydrol.2019.03.073.

[77] L. Liuzzo, V. Sammartano, and G. Freni, "Comparison between Different Distributed Methods for Flood Susceptibility Mapping," *Water Resour. Manag.*, vol. 33, no. 9, pp. 3155–3173, July 2019, doi: 10.1007/s11269-019-02293-w.

[78] M. Avand, H. Moradi, and M. R. lasboyee, "Using machine learning models, remote sensing, and GIS to investigate the effects of changing climates and land uses on flood probability," *J. Hydrol.*, vol. 595, p. 125663, Apr. 2021, doi: 10.1016/j.jhydrol.2020.125663.

[79] B. Mester, K. Frieler, and J. Schewe, "Human displacements, fatalities, and economic damages linked to remotely observed floods," *Sci. Data*, vol. 10, no. 1, p. 482, July 2023, doi: 10.1038/s41597-023-02376-9.

[80] J. Breznik, K. Oštir, and G. Rak, "The potential of Sentinel-1 imagery for flood event detection: A satellite vs. hydraulic model comparison," *J. Hydrol.*, vol. 651, p. 132587, Apr. 2025, doi: 10.1016/j.jhydrol.2024.132587.

[81] D. Munasinghe, R. P. de M. Frasson, C. H. David, M. Bonnema, G. Schumann, and G. R. Brakenridge, "A multi-sensor approach for increased measurements of floods and their societal impacts from space," *Commun. Earth Environ.*, vol. 4, no. 1, p. 462, Dec. 2023, doi: 10.1038/s43247-023-01129-1.

[82] Y. You, J. Cao, and W. Zhou, "A Survey of Change Detection Methods Based on Remote Sensing Images for Multi-Source and Multi-Objective Scenarios," *Remote Sens.*, vol. 12, no. 15, p. 2460, Jan. 2020, doi: 10.3390/rs12152460.

[83] J. C. J. H. Aerts *et al.*, "Integrating human behaviour dynamics into flood disaster risk assessment," *Nat. Clim. Change*, vol. 8, no. 3, pp. 193–199, Mar. 2018, doi: 10.1038/s41558-018-0085-1.

[84] M. T. Rashid, D. Y. Zhang, and D. Wang, "SocialDrone: An Integrated Social Media and Drone Sensing System for Reliable Disaster Response," in *IEEE INFOCOM 2020 - IEEE Conference on Computer Communications*, July 2020, pp. 218–227. doi: 10.1109/INFOCOM41043.2020.9155522.

[85] T. Yabe, K. Tsubouchi, and Y. Sekimoto, "CityFlowFragility: Measuring the Fragility of People Flow in Cities to Disasters using GPS Data Collected from Smartphones," *Proc ACM Interact Mob Wearable Ubiquitous Technol*, vol. 1, no. 3, p. 117:1-117:17, Sept. 2017, doi: 10.1145/3130982.

[86] T. Yabe, N. K. W. Jones, P. S. C. Rao, M. C. Gonzalez, and S. V. Ukkusuri, "Mobile phone location data for disasters: A review from natural hazards and epidemics," *Comput. Environ. Urban Syst.*, vol. 94, p. 101777, June 2022, doi: 10.1016/j.compenvurbsys.2022.101777.

[87] "An Analysis of Factors Influencing Disaster Mobility Using Location Data from Smartphones: Case Study of Western Japan Flooding." Accessed: Nov. 16, 2025. [Online]. Available: https://www.jstage.jst.go.jp/article/jdr/14/6/14_903/_article/-char/ja/

[88] X. Chaoxu, N. Gaozhong, F. Xiwei, Z. Junxue, and P. Xiaoke, "Research on the application of mobile phone location signal data in earthquake emergency work: A case study of Jiuzhaigou earthquake," *PLOS ONE*, vol. 14, no. 4, p. e0215361, Apr. 2019, doi: 10.1371/journal.pone.0215361.

[89] S. P. Cumbane and G. Gidófalvi, "Spatial Distribution of Displaced Population Estimated Using Mobile Phone Data to Support Disaster Response Activities," *ISPRS Int. J. Geo-Inf.*, vol. 10, no. 6, p. 421, June 2021, doi: 10.3390/ijgi10060421.

[90] L. Alexander, S. Jiang, M. Murga, and M. C. González, "Origin–destination trips by purpose and time of day inferred from mobile phone data," *Transp. Res. Part C Emerg. Technol.*, vol. 58, pp. 240–250, Sept. 2015, doi: 10.1016/j.trc.2015.02.018.

[91] L. Zou, N. S. N. Lam, H. Cai, and Y. Qiang, "Mining Twitter Data for Improved Understanding of Disaster Resilience," *Ann. Am. Assoc. Geogr.*, vol. 108, no. 5, pp. 1422–1441, Sept. 2018, doi: 10.1080/24694452.2017.1421897.

[92] V. V. Mihunov, N. S. N. Lam, L. Zou, Z. Wang, and K. Wang, "Use of Twitter in disaster rescue: lessons learned from Hurricane Harvey," *Int. J. Digit. Earth*, vol. 13, no. 12, pp. 1454–1466, Dec. 2020, doi: 10.1080/17538947.2020.1729879.

[93] M. Martínez-Rojas, M. del C. Pardo-Ferreira, and J. C. Rubio-Romero, "Twitter as a tool for the management and analysis of emergency situations: A systematic literature review," *Int. J. Inf. Manag.*, vol. 43, pp. 196–208, Dec. 2018, doi: 10.1016/j.ijinfomgt.2018.07.008.

[94] Md. A. Islam, F. Rabbi, and N. U. I. Hossain, "Performance evaluation of NLP and CNN models for disaster detection using social media data," *Soc. Netw. Anal. Min.*, vol. 14, no. 1, p. 213, Nov. 2024, doi: 10.1007/s13278-024-01374-y.

[95] S. Shan, F. Zhao, Y. Wei, and M. Liu, "Disaster management 2.0: A real-time disaster damage assessment model based on mobile social media data—A case study of Weibo (Chinese Twitter)," *Saf. Sci.*, vol. 115, pp. 393–413, June 2019, doi: 10.1016/j.ssci.2019.02.029.

[96] Y. Feng, X. Huang, and M. Sester, "Extraction and analysis of natural disaster-

related VGI from social media: review, opportunities and challenges," *Int. J. Geogr. Inf. Sci.*, vol. 36, no. 7, pp. 1275–1316, July 2022, doi: 10.1080/13658816.2022.2048835.

[97] "Spatial biases in crowdsourced data: Social media content attention concentrates on populous areas in disasters," *Comput. Environ. Urban Syst.*, vol. 83, p. 101514, Sept. 2020, doi: 10.1016/j.compenvurbsys.2020.101514.

[98] M. Löchner *et al.*, "Case Study on Privacy-Aware Social Media Data Processing in Disaster Management," *ISPRS Int. J. Geo-Inf.*, vol. 9, no. 12, p. 709, Dec. 2020, doi: 10.3390/ijgi9120709.

[99] M. Jamali, A. Nejat, S. Ghosh, F. Jin, and G. Cao, "Social media data and post-disaster recovery," *Int. J. Inf. Manag.*, vol. 44, pp. 25–37, Feb. 2019, doi: 10.1016/j.ijinfomgt.2018.09.005.

[100] S. Soomro *et al.*, "How effective is twitter (X) social media data for urban flood management?," *J. Hydrol.*, vol. 634, p. 131129, May 2024, doi: 10.1016/j.jhydrol.2024.131129.

[101] K. Rudra, N. Ganguly, P. Goyal, and S. Ghosh, "Extracting and Summarizing Situational Information from the Twitter Social Media during Disasters," *ACM Trans Web*, vol. 12, no. 3, p. 17:1-17:35, July 2018, doi: 10.1145/3178541.

[102] C. S. Renschler and Z. Wang, "Multi-source data fusion and modeling to assess and communicate complex flood dynamics to support decision-making for downstream areas of dams: The 2011 hurricane irene and schoharie creek floods, NY," *Int. J. Appl. Earth Obs. Geoinformation*, vol. 62, pp. 157–173, Oct. 2017, doi: 10.1016/j.jag.2017.06.002.

[103] Z. Wu, Y. Chen, X. Zheng, S. Huang, C. Duan, and P. Wang, "A novel framework for evidence-based assessment of flood resilience integrating multi-source evidence: A case study of the Yangtze River Economic Belt, China," *Ecol. Indic.*, vol. 167, p. 112705, Oct. 2024, doi: 10.1016/j.ecolind.2024.112705.

[104] A. Maranzoni, M. D'Oria, and C. Rizzo, "Quantitative flood hazard assessment methods: A review," *J. Flood Risk Manag.*, vol. 16, no. 1, p. e12855, 2023, doi: 10.1111/jfr3.12855.

[105] T. P. Pangali Sharma, J. Zhang, U. A. Koju, S. Zhang, Y. Bai, and M. K. Suwal, "Review of flood disaster studies in Nepal: A remote sensing perspective," *Int. J. Disaster Risk Reduct.*, vol. 34, pp. 18–27, Mar. 2019, doi: 10.1016/j.ijdrr.2018.11.022.

[106] T. H. Nguyen *et al.*, "Improvement of Flood Extent Representation With Remote Sensing Data and Data Assimilation," *IEEE Trans. Geosci. Remote Sens.*, vol. 60,

pp. 1–22, 2022, doi: 10.1109/TGRS.2022.3147429.

[107] R. Sadiq, Z. Akhtar, M. Imran, and F. Ofli, “Integrating remote sensing and social sensing for flood mapping,” *Remote Sens. Appl. Soc. Environ.*, vol. 25, p. 100697, Jan. 2022, doi: 10.1016/j.rsase.2022.100697.

[108] M. Bersi, H. Saibi, K. Abdelrahman, M. S. Fnais, and M. Saber, “Remote sensing and climatic data for flood impact assessment in Al-Ain (UAE),” *Sci. Rep.*, vol. 15, no. 1, p. 26182, July 2025, doi: 10.1038/s41598-025-12234-w.

[109] P. Hansana, X. Guo, S. Zhang, X. Kang, and S. Li, “Flood Analysis Using Multi-Scale Remote Sensing Observations in Laos,” *Remote Sens.*, vol. 15, no. 12, p. 3166, Jan. 2023, doi: 10.3390/rs15123166.

[110] J. Liang, D. Liu, L. Feng, and K. Huang, “Rapid Probabilistic Inundation Mapping Using Local Thresholds and Sentinel-1 SAR Data on Google Earth Engine,” *Remote Sens.*, vol. 17, no. 10, p. 1747, Jan. 2025, doi: 10.3390/rs17101747.

[111] M. Li, R. Zhang, and K. Liu, “A New Marine Disaster Assessment Model Combining Bayesian Network with Information Diffusion,” *J. Mar. Sci. Eng.*, vol. 9, no. 6, p. 640, June 2021, doi: 10.3390/jmse9060640.

[112] C. Bethuel *et al.*, “Applying the Dempster–Shafer Fusion Theory to Combine Independent Land-Use Maps: A Case Study on the Mapping of Oil Palm Plantations in Sumatra, Indonesia,” *Remote Sens.*, vol. 17, no. 2, p. 234, Jan. 2025, doi: 10.3390/rs17020234.

[113] S. Hitouri *et al.*, “Flood Susceptibility Mapping Using SAR Data and Machine Learning Algorithms in a Small Watershed in Northwestern Morocco,” *Remote Sens.*, vol. 16, no. 5, p. 858, Jan. 2024, doi: 10.3390/rs16050858.

[114] H. Khodaei, F. Nasiri Saleh, A. Nobakht Dalir, and E. Zarei, “Future flood susceptibility mapping under climate and land use change,” *Sci. Rep.*, vol. 15, no. 1, p. 12394, Apr. 2025, doi: 10.1038/s41598-025-97008-0.

[115] A. Arora, P. Durga G, M. Pandey, and A. Arabameri, “Machine learning model optimization for flood susceptibility zonation over the Kosi megafan, Himalayan foreland basin, India,” *Sci. Rep.*, vol. 15, no. 1, p. 32757, Sept. 2025, doi: 10.1038/s41598-025-07403-w.

[116] D. Wagenaar *et al.*, “Invited perspectives: How machine learning will change flood risk and impact assessment,” *Nat. Hazards Earth Syst. Sci.*, vol. 20, no. 4, pp. 1149–1161, Apr. 2020, doi: 10.5194/nhess-20-1149-2020.

[117] M. Belgiu and L. Drăguţ, “Random forest in remote sensing: A review of applications and future directions,” *ISPRS J. Photogramm. Remote Sens.*, vol. 114, pp. 24–31, Apr. 2016, doi: 10.1016/j.isprsjprs.2016.01.011.

[118] V. F. Rodriguez-Galiano, B. Ghimire, J. Rogan, M. Chica-Olmo, and J. P. Rigol-Sanchez, “An assessment of the effectiveness of a random forest classifier for land-cover classification,” *ISPRS J. Photogramm. Remote Sens.*, vol. 67, pp. 93–104, Jan. 2012, doi: 10.1016/j.isprsjprs.2011.11.002.

[119] Z. Liu, T. Felton, and A. Mostafavi, “Interpretable machine learning for predicting urban flash flood hotspots using intertwined land and built-environment features,” *Comput. Environ. Urban Syst.*, vol. 110, p. 102096, June 2024, doi: 10.1016/j.compenvurbsys.2024.102096.

[120] A. Shastry, E. Carter, B. Coltin, R. Sleeter, S. McMichael, and J. Eggleston, “Mapping floods from remote sensing data and quantifying the effects of surface obstruction by clouds and vegetation,” *Remote Sens. Environ.*, vol. 291, p. 113556, June 2023, doi: 10.1016/j.rse.2023.113556.

[121] J. S. Navarro, R. Zhuang, C. Albertini, and S. Manfreda, “Mapping flood susceptibility using Random Forest exploiting satellite observations and geomorphic features,” *Sci. Total Environ.*, vol. 1002, p. 180592, Nov. 2025, doi: 10.1016/j.scitotenv.2025.180592.

[122] T. Wang, H. Wang, Z. Wang, and J. Huang, “Dynamic risk assessment of urban flood disasters based on functional area division—A case study in Shenzhen, China,” *J. Environ. Manage.*, vol. 345, p. 118787, Nov. 2023, doi: 10.1016/j.jenvman.2023.118787.

[123] T. Schempp, H. Zhang, A. Schmidt, M. Hong, and R. Akerkar, “A framework to integrate social media and authoritative data for disaster relief detection and distribution optimization,” *Int. J. Disaster Risk Reduct.*, vol. 39, p. 101143, Oct. 2019, doi: 10.1016/j.ijdrr.2019.101143.

[124] J. Ahern, “From *fail-safe* to *safe-to-fail*: Sustainability and resilience in the new urban world,” *Landsc. Urban Plan.*, vol. 100, no. 4, pp. 341–343, Apr. 2011, doi: 10.1016/j.landurbplan.2011.02.021.

[125] F. Berkes and H. Ross, “Community Resilience: Toward an Integrated Approach,” *Soc. Nat. Resour.*, vol. 26, no. 1, pp. 5–20, Jan. 2013, doi: 10.1080/08941920.2012.736605.

[126] J. Rowley, “The wisdom hierarchy: representations of the DIKW hierarchy,” *J. Inf. Sci.*, vol. 33, no. 2, pp. 163–180, Apr. 2007, doi: 10.1177/0165551506070706.

[127] S. Meerow, J. P. Newell, and M. Stults, “Defining urban resilience: A review,” *Landsc. Urban Plan.*, vol. 147, pp. 38–49, Mar. 2016, doi: 10.1016/j.landurbplan.2015.11.011.

[128] S. Carpenter, B. Walker, J. M. Anderies, and N. Abel, “From Metaphor to

Measurement: Resilience of What to What?," *Ecosystems*, vol. 4, no. 8, pp. 765–781, Dec. 2001, doi: 10.1007/s10021-001-0045-9.

[129] M. Bruneau *et al.*, "A Framework to Quantitatively Assess and Enhance the Seismic Resilience of Communities," *Earthq. Spectra*, vol. 19, no. 4, pp. 733–752, Nov. 2003, doi: 10.1193/1.1623497.

[130] C. Folke, "Resilience: The emergence of a perspective for social–ecological systems analyses," *Glob. Environ. Change*, vol. 16, no. 3, pp. 253–267, Aug. 2006, doi: 10.1016/j.gloenvcha.2006.04.002.

[131] J. Teng, A. J. Jakeman, J. Vaze, B. F. W. Croke, D. Dutta, and S. Kim, "Flood inundation modelling: A review of methods, recent advances and uncertainty analysis," *Environ. Model. Softw.*, vol. 90, pp. 201–216, Apr. 2017, doi: 10.1016/j.envsoft.2017.01.006.

[132] J. Fohringer, D. Dransch, H. Kreibich, and K. Schröter, "Social media as an information source for rapid flood inundation mapping," *Nat. Hazards Earth Syst. Sci.*, vol. 15, no. 12, pp. 2725–2738, Dec. 2015, doi: 10.5194/nhess-15-2725-2015.

[133] I. Nonaka and G. von Krogh, "Tacit knowledge and knowledge conversion: Controversy and advancement in organizational knowledge creation theory," *Organ. Sci.*, vol. 20, no. 3, pp. 635–652, 2009, doi: 10.1287/orsc.1080.0412.

[134] J. Yin, A. Lampert, M. Cameron, B. Robinson, and R. Power, "Using Social Media to Enhance Emergency Situation Awareness," *IEEE Intell. Syst.*, vol. 27, no. 6, pp. 52–59, Nov. 2012, doi: 10.1109/MIS.2012.6.

[135] M. Frické, "The knowledge pyramid: a critique of the DIKW hierarchy," *J. Inf. Sci.*, vol. 35, no. 2, pp. 131–142, Apr. 2009, doi: 10.1177/0165551508094050.

[136] I. Nonaka and R. Toyama, "The knowledge-creating theory revisited: knowledge creation as a synthesizing process," *Knowl. Manag. Res. Pract.*, vol. 1, no. 1, pp. 2–10, July 2003, doi: 10.1057/palgrave.kmrp.8500001.

[137] "Social Sensing: A New Approach to Understanding Our Socioeconomic Environments: Annals of the American Association of Geographers: Vol 105 , No 3 - Get Access." Accessed: Nov. 16, 2025. [Online]. Available: https://www.tandfonline.com/doi/full/10.1080/00045608.2015.1018773

[138] B. T. Haworth, "Implications of Volunteered Geographic Information for Disaster Management and GIScience: A More Complex World of Volunteered Geography," *Ann. Am. Assoc. Geogr.*, vol. 108, no. 1, pp. 226–240, Jan. 2018, doi: 10.1080/24694452.2017.1321979.

[139] F. R. Stevens, A. E. Gaughan, C. Linard, and A. J. Tatem, "Disaggregating Census Data for Population Mapping Using Random Forests with Remotely-Sensed and

Ancillary Data," *PLOS ONE*, vol. 10, no. 2, p. e0107042, Feb. 2015, doi: 10.1371/journal.pone.0107042.

[140] "Rapid assessment of disaster damage using social media activity | Science Advances." Accessed: Nov. 16, 2025. [Online]. Available: https://www.science.org/doi/10.1126/sciadv.1500779

[141] K. Tierney, "Disaster Governance: Social, Political, and Economic Dimensions," *Annu. Rev. Environ. Resour.*, vol. 37, no. Volume 37, 2012, pp. 341–363, Nov. 2012, doi: 10.1146/annurev-environ-020911-095618.

[142] T. McDaniels, S. Chang, D. Cole, J. Mikawoz, and H. Longstaff, "Fostering resilience to extreme events within infrastructure systems: Characterizing decision contexts for mitigation and adaptation," *Glob. Environ. Change*, vol. 18, no. 2, pp. 310–318, May 2008, doi: 10.1016/j.gloenvcha.2008.03.001.

Publications

International Journals

[1] YANG Yifan, OHIRA Naoki, GOKON Hideomi: “Small-Grid Urban Flood Prediction Model Using Twitter Data and Population GPS Data- An example of the 2019 Nagano City Flood”, Progress in Disaster Science, vol. 24, p. 100385, Dec. 2024, doi: 10.1016/j.pdisas.2024.100385.

Domestic Journals

[2] YANG Yifan, OHIRA Naoki, GOKON Hideomi: “Time and Spatial Analysis of Flood Disaster based on GPS Data: A Case Study of Flood Disaster in Nagano City in 2019”, Journal of Social Safety Science, vol. 44, pp. 121–130, 2024, doi: 10.11314/jisss.44.121.

International Conference Proceedings

[3] YANG Yifan, GOKON Hideomi, TAKEUCHI Wataru: “Flood Extent Prediction Using Remote Sensing Data Based on Machine Learning and Dempster-Shafer Theory: A Case Study of Nagano City”, Proceedings of the IGARSS 2025 - 2025 IEEE International Geoscience and Remote Sensing Symposium, pp. 4517-4521, doi: 10.1109/IGARSS55030.2025.11242843.

[4] YANG Yifan, OHIRA Naoki, GOKON Hideomi: “Model construction and evaluation of flood area estimation based on SAR and GPS data”, Proceedings of the IGARSS 2024 - 2024 IEEE International Geoscience and Remote Sensing Symposium, pp. 1362-1365, doi: 10.1109/IGARSS53475.2024.10642419.

International Conferences

[5] YANG Yifan, GOKON Hideomi, TAKEUCHI Wataru: “Flood Extent Prediction

Using Remote Sensing Data Based on Machine Learning and Dempster-Shafer Theory: A Case Study of Nagano City" IEEE International Geoscience and Remote Sensing Symposium, Brisbane, Australia, 2025.8.3- 8.8.

[6] YANG Yifan, GOKON Hideomi, TAKEUCHI Wataru: "Estimation of Ground Deformation in Suzu City with ALOS-2/PALSAR-2 Data", The 3rd International Symposium on One Health One World, Putrajaya, Malaysia, 2024.12.10-12.12.

[7] YANG Yifan, OHIRA Naoki, GOKON Hideomi: "Model Construction and Evaluation of Flood Area Estimation Based on SAR and GPS Data," IGARSS 2024 - 2024 IEEE International Geoscience and Remote Sensing Symposium, Athens, Greece, 2024.7.7– 7.12.

[8] YANG Yifan, OHIRA Naoki, GOKON Hideomi: "Small-Grid Urban Flood Prediction Model Using Twitter Data and Population GPS Data- An example of the 2019 Nagano City Flood", Big Data for Disaster Response and Management in Asia and the Pacific, Sendai, Japan, 2024.2.15-2.17.

Awards

Best Paper Presenter Award (Gold), The 3rd International Symposium on One Health One World (Putrajaya, Malaysia, December 10–12, 2024), awarded for "Estimation of Ground Deformation in Suzu City with ALOS-2/PALSAR-2 Data."