

Article

Not peer-reviewed version

A Weighted kNN Based Spatial Framework of Flood Inundation Risk for Coastal Tourism – A case study in Zhejiang, China

Shuang Liu , [Nengzhi Tan](#) , [Rui Liu](#) *

Posted Date: 28 August 2023

doi: 10.20944/preprints202308.1811.v1

Keywords: Climate change, Sensitivity analysis, Evaluation accuracy, Hazard, Year return period, Overall accuracy



Preprints.org is a free multidiscipline platform providing preprint service that is dedicated to making early versions of research outputs permanently available and citable. Preprints posted at Preprints.org appear in Web of Science, Crossref, Google Scholar, Scilit, Europe PMC.

Copyright: This is an open access article distributed under the Creative Commons Attribution License which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Article

A Weighted k NN Based Spatial Framework of Flood Inundation Risk for Coastal Tourism – A Case Study in Zhejiang, China

Shuang Liu ¹, Nengzhi Tan ² and Rui Liu ^{3,*}

¹ School of Tourism Management, Huzhou Vocational and Technical College, Huzhou 313000, China; lsiuhuang@163.com

² Huzhou University, Huzhou 313000, China; 02144@zjhu.edu.cn

³ College of Geospatial Information Science and Technology, Capital Normal University, Beijing 100048, China; drruiliu@yeah.net

* Correspondence: drruiliu@yeah.net

Abstract: Flood inundation causes socioeconomic losses for coastal tourism under climate extremes, progressively attracting global attention. Mapping, evaluating, and predicting the flood inundation risk (FIR) is significant for coastal tourism. The study develops a spatial tourism-aimed framework integrating a weighted k -Nearest Neighbors ($WkNN$), Geographic Information Systems, and flood-related spatially environmental criteria such as precipitation, elevation, soil, and drainage systems. These model inputs were standardized and weighted using distance, and integrated into $WkNN$ to infer regional probability and distribution of FIR. Zhejiang province, China, was selected as a case study. The resulting map was pictured to denote the likelihood of the criteria at various risk categories, which was validated by historical Maximum Inundation Extent (MIE) extracted from World Environment Situation Room. The result indicates 80.59% of $WkNN$ results reasonably confirm the MIE. Precipitation and elevation make a negative contribution to high-medium risk, and drainage systems positively alleviate the regional stress of FIR. The results can help stakeholders make suitable strategies to protect coastal tourism, and also weigh $WkNN$ is superior to kNN in FIR assessment. The framework provides a productive way to yield a reliable assessment of FIR and can also be extended to other risk-related environmental studies under climate change.

Keywords: climate change; sensitivity analysis; evaluation accuracy; hazard; year return period; Overall accuracy

1. Introduction

Tourism is one of the fastest-growing economic sectors in coastal areas, which attracts sufficient tourists, and brings socioeconomic benefits to human society [1], due to its convenient transport and natural environment [2]. For instance, it provided 292 million jobs to 10% of the world's workforce, which contributes about 10% of the global GDP in 2016, and may expect to be 11.4% by 2027 [3]. In 2018, a total of about 1.4 billion tourists were recorded globally [4]. However, natural environmental variability would also cause diverse natural disasters [5]. For example, flash floods caused 11 death and forced the evacuation of approximately 4,000 tourists in Jordan in November 2018 [6]. A devastating flash flood caused by heavy rainfall struck Yesanpo, a nature-centered tourist destination near Beijing, leaving over 15,000 visitors trapped overnight in July 2012 [7].

Globally, flood inundation is recognized as one of the most common natural disasters, and triggers catastrophic property damage and even lives, which has been recorded in the past decades [8,9]. Statistically, floods constitute 43% of the total number of natural disasters, and 47% of the amount of weather-related disasters. Floods affected 2.3 billion people and \$662 billion in damage from 1995 to 2015 [10]. Approximately 16,000 lives have been lost in flash floods in China between 2000 to 2018, which occupies 74% of the whole flood-relative mortalities [7]. In 2021, about 400 disastrous events were recorded by the Emergency Event Database, and floods dominated 223 occurrences. The

most severe one is the Henan flood in China, causing 352 deaths, 14.5 million people affected, and \$16.5 billion in economic losses (<https://reliefweb.int/report/world/2021-disasters-numbers>).

Coastal areas not only are the most developed but also extraordinarily flood-prone places since flood frequency and density in these places are higher than others under extreme climates, such as tropic cyclones and typhoons [3,11-13]. In 2006, super typhoon Sang Mei triggered 153 deaths in Wenzhou, Zhejiang province, bringing about RMB 11 billion in direct economic losses [14]. In 2013, flood inundation, which was triggered by typhoons, affected eight million residents, and about RMB 33 billion of straight financial losses in Ningbo, Zhejiang province [15]. Therefore, predicting and understating the potential flood inundation risk (FIR) for tourism in coastal areas is of great importance for regional sustainable development via minimizing possible harm.

Based on the aforementioned, abundant approaches have been used in the flood-tourism field, and make efforts to find suitable ways to mitigate the negative impacts of the flood on tourism. Some used citizen-based activities to improve resilience [16], and local knowledge is efficient for improving the quality of hazard preparation [17,18]. Besides, climate change models connected with socioeconomic data [19], taxable sales records [20], etc., were used in estimating economic losses for tourism. Geographic Information Systems (GIS) embrace the advantage to combine models and gridded datasets, such as environmental raster and socioeconomic features [21]. It is a suitable tool for deriving regional indicators and evaluating their impacts on hotels [22] and properties [23], and spatial accessibility in FIR [24]. GIS was further combined with Remote Sensing (RS), hydrological and hydrodynamic flood simulation models such as FLO-2D [25], and HAZUS-MH [26] to assess flood scenarios for tourism facilities [2,27]. Besides, some comparatively advanced algorithms in Machine Learning have been successfully used in flood risk evaluation as well, such as Bayesian networks [28,29], and the AHP-SA model [21,30]. These methods deeply explored the mechanism of flood disasters by integrating multiple factors, such as rainfall, soil, and river. However, difficulties in modeling FIR for tourism across large areas may be encountered due to model complexity and advanced, professional mathematical knowledge. Besides, the computational cost needs to consider for complex models in long-term spatial data evaluation.

Based on the limitations, k Nearest Neighbors (k NN) was proposed and used to assess FIR for coastal tourism in our previous investigation [31]. The reasonable result demonstrated k NN is a simple but efficient computer algorithm since it has no requirement for the distribution of original data, which makes it faster in the process of data training and calculating. Also, it has been widely used in a few studies, such as the classification of missing data, risk evaluation and prediction [31-33]. While k NN method has some merits, there are some problems needed to be further explored and solved. For example, it is a lazy algorithm and all features with different importance are considered to be equally weighted, which may lead to poor classification performance.

Therefore, this study continually extends our previous k NN-based research investigation for tourism by using distance-weighted methods to improve the evaluation accuracy (EA) and performance of k NN method. Consequently, the aims and innovations of this paper can be summarized as the following points.

1. Improved the performance of k NN algorithm with a distance-weighted method, and demonstrated the weighed k NN (Wk NN) can gain higher accuracy prediction than k NN;
2. Developed and applied Wk NN-based spatial framework with spatial technologies into flood risk assessment for tourism in coastal areas;
3. Due to the limitation of the spatially gridded data, World Environment Situation Room (WESR) was first used in validating flood risk for coastal areas and was demonstrated WESR can be successfully used in flood risk evaluation.

2. Framework Development

2.1. Basic Principle of k NN

The basic principle of k NN assumes that examined objects are similar to k nearby sample neighbors, or at least, they have similar characteristics (please further refer to Liu, Liu and Tan [31])). Based on the basis, there are mainly two steps to classify the categories of the examined objects:

1. Calculate the pairwise distance between examined objects in testing datasets and k nearby sample neighbors in training datasets;
2. Vote the categories of the k nearby samples to confirm the classifications of examined objects.

The distance quantified the similarity between examine-and-sample objects. Usually, the lower the distance, the higher the similarity. Many methods were used in k NN to calculate objects' distance, such as Manhattan Distance [34], Minkowski Distance [35], and Chebyshev Distance [36]. Among them, Euclidean Distance [37] is a popular and frequently used method. It refers to the distance between objects in Euclidean space, which can be described as:

$$d_{kj} = \sqrt{\sum_{i=1}^k (x_k - x_j)^2} \quad (1)$$

Where x_j are the features of examined objects, x_k are the known categories of sample neighbors, and d_{kj} are the distances, and k means the number of nearby neighbors.

K values have a significant impact on the classification results of k NN. Larger k values may cause a complex k NN model and overfitting results, or a simple model and underfitting results in classification [38]. Thus, a proper k value may be between two extremes and should be discussed in model building. Traditionally, k neighbors can be found in training datasets that are nearer to an examined object in testing datasets. The category of a testing dataset will be determined by the following classified decision rules:

$$c_j = \operatorname{argmax} \sum_{x_i \in N_k(x)} I(y_i = c), i = 1, 2, \dots, k \quad (2)$$

Where c_i are the predicted categories, $N_k(x)$ is the k nearby neighbors, I is the indicator function, that is when $y_i = c$, $I=1$, otherwise, $I=0$.

Equations 1 and 2 show that the predicted categories of examined objects are mainly determined by the categories of a majority of k samples. However, the weights or importance between examined objects and neighbors are ignored, which makes the classification accuracy lower. Therefore, distance-based weights can be considered to modify and improve the accuracy of k NN.

2.2 Weighted k NN (WkNN)

Weights refer to the importance or contribution of factors to a system. Many approaches can be engaged in calculating weight, such as entropy methods [39], Analytic Hierarchy Process [30], and Principal component analysis [40]. However, these methods get stuck in the complex process of knowledge and calculation. In the study, k NN's weight can be simply expressed and calculated using an inverse relationship to Euclidean distance (equation 3), which means the larger the distance, the smaller the weight.

$$w_{kj} = \frac{1}{d_{kj}} \quad (3)$$

Then equation (2) can be described as:

$$c_j = \operatorname{argmax} \sum_{x_i \in N_k(x)} I(y_i = c) * w_{kj}, i = 1, 2, \dots, k \quad (4)$$

2.3 Framework Conceptualization

After summarizing similar research investigations, a WkNN-based spatial framework of FIR assessment for coastal tourism is conceptualized and constructed. The framework can be divided into three parts: input collection, model construction, and output classification (Figure 1).

The first module mainly collects spatiotemporal data and flood-related index derivation. The data consists of three spatial branches: climate, environment, and validation data. Several indexes are

derived from the flood-induced factors which range from mean annual rainfall to drainage density. Flood hazard data of different year return periods (YRP) were collected to create maximum inundation extension (MIE) with historical inundated times which verifies the evaluation results of WkNN model.

The second module is the center part of the framework. Following data collection, all spatial indexes are standardized into datasets with four categories: very low, low, medium, and high risk. The standardized datasets within the extent of MIE were divided into two parts: 70% training dataset and 30% testing dataset. KNN and WkNN were employed to calculate the categories of random records from training datasets with nearby k-training datasets. The inferred results were compared with their existing categories in the training dataset, which produces a confusion matrix and Overall Accuracy (OA). WkNN model with the highest OA value will be extended to whole areas. A sensitivity analysis was conducted to explore the relationship between the inputs and outputs of the model.

The third model is for mapping and evaluating the likelihood of FIR and assessing tourism facilities exposed in FIR.

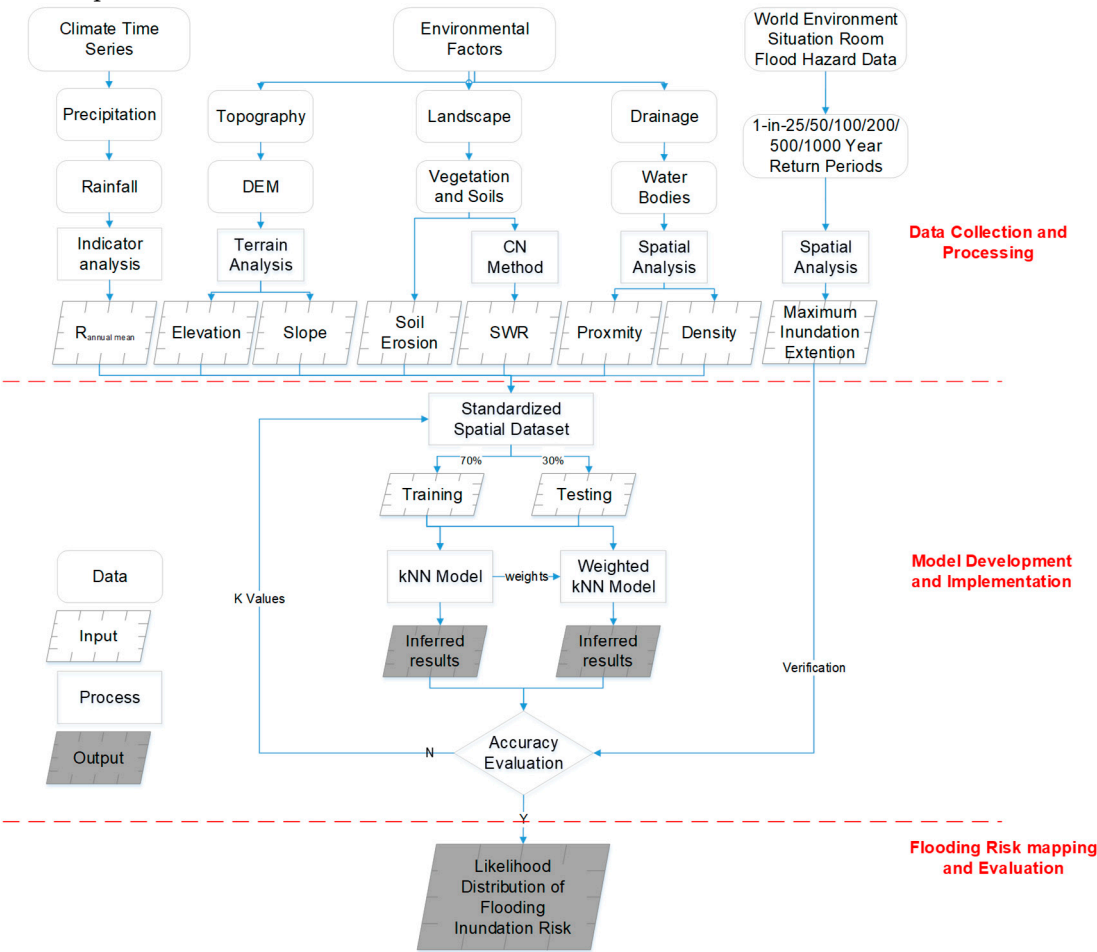


Figure 1. Conceptual framework of WkNN which includes data collection, processing, model construction, validation, accuracy evaluation, and flood risk mapping.

3. Case Study

3.1 Study Area

Zhejiang province (118-122.2°E, 27-31.2°N), China, was selected as a case study. The province sits on the southeast coast of the Yangtze River Delta and at the land-and-sea junction. It faces the East China Sea and slopes from southwest to northeast [15,41]. The superior location and special environment make it spread superior tourism resources (e.g. West Lake) (Figure 2a) over 11 main

cities (Figure 2b), which attracted millions of domestic and foreign tourists every year. In 2014, its tourism income occupies 15.7% (about RMB 630 billion) of the provincial GDP (about RMB 4,015 billion). However, Zhejiang also experienced higher FIR, caused by sea levels, typhoons, and tropical cyclones due to complex environmental conditions under risky climate change. Historical records show that typhoons and tropical cyclones bring heavy rainfall and floods over the study area during the past 60 years from 1950 to 2009, which makes the area suffer direct mean annual losses (about RMB 10 billion), especially in its southeast coastal areas. In 2006, super typhoon Sang Mei triggered 153 deaths in Wenzhou, bringing about RMB 11 billion in direct economic losses [14]. In 2013, typhoon-triggering flood inundation affected eight million residents, and about RMB 33 billion of straight financial losses in Ningbo [15]. Besides, Zhejiang has a 6,500 km coastline and its average sea level has risen 98 mm during the past 30 years and is projected to speed up under extreme climates with destructive potential. All of them severely influence tourism operations, socioeconomic income, and even people's lives [41]. Hence, the continually historical flood damage in the area has underscored an urgent need for assessing FIR to manage flood disasters and promote the stable and sustainable development of Zhejiang's coastal tourism economy.

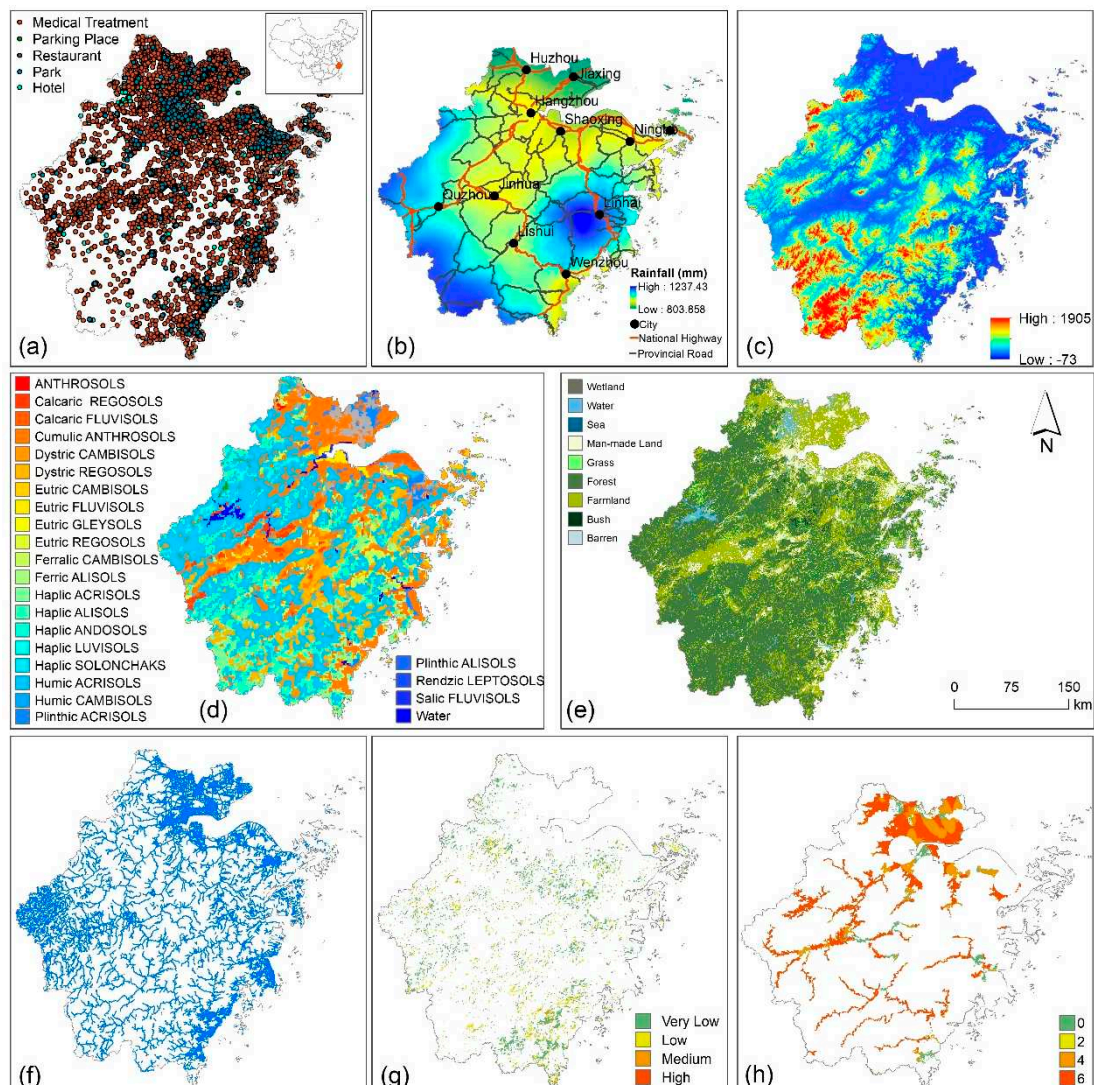


Figure 2. Spatial distribution of collected data. (a) various tourism facilities including parks and hotels, (b) mean annual rainfall (1951-2007), 11 cities and two main levels of roads, (c) Digital Elevation Model at 30 m resolution, (d) soil types and contains, (e) land use and land cover, (f) drainage system, (g) soil erosion and (h) 1-in-50 YRP example of various flood year return period.

3.2 Flood-derived Spatial Data Collection and Processing

Flood risk and evaluation is a comprehensive system, arising from flood hazard, exposure, and disaster-prone environments at a particular location [18]. The criteria from the three parts have been systematically selected and derived in light of their influence on the occurrence and distribution of flood inundation with domain knowledge [21,42,43]. Flood hazard is defined as a deriving factor to trigger FIR, such as extreme rainfall. Exposure refers to the degree or extent to persons, environments, or assets (e.g. tourism facilities) that are likely to be located in flood-prone areas. Other factors, including topography, hydrology, land use, and soil, are defined as disaster-prone environments. They were extracted and standardized via data processing and analysis in GIS environment and worked as spatial inputs for the WkNN-centered framework to reason the FIR in the future.

Rainfall

Rainfall can be divided into two main types: short-period intensive rainfall, and prolonged-extensive rainfall, all of which can cause flood inundation. Due to the limitation of data in some countries, such as China, this study extracted annual mean precipitation from Asian Precipitation - Highly-Resolved Observational Data Integration Towards Evaluation (APHRODITE) between 1951 to 2007 [44] as rainfall indices. APHRODITE dataset has been illustrated to well match the accurate feature of rain belts in China [31,45,46]. The extracted APHRODITE dataset was interpolated into continuous raster rainfall data and clipped within the study area with Inverse Distance Weighting in ArcGIS version 10.4 [47] (Figure 2b, Figure 3a).

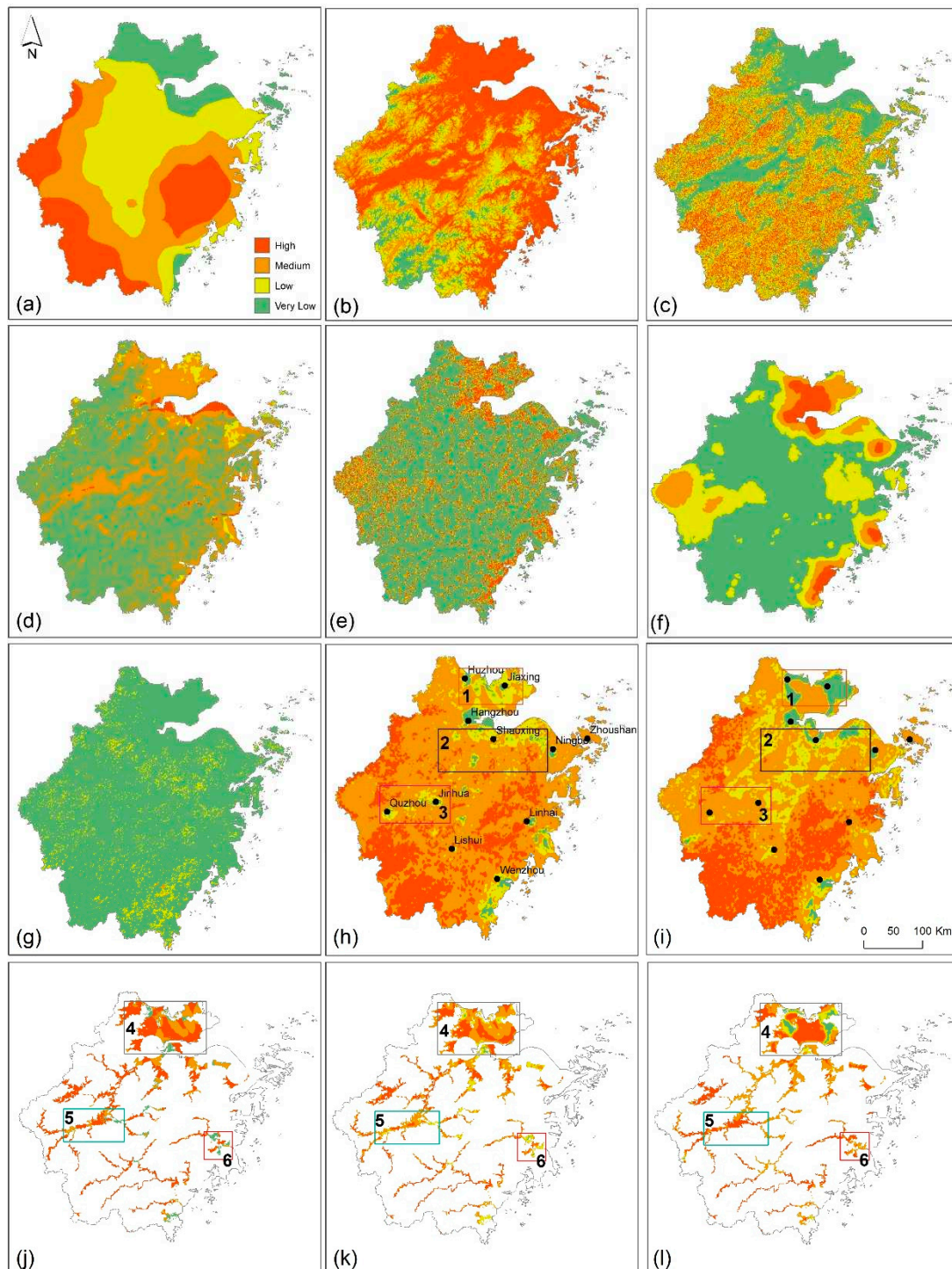


Figure 3. Standard risk evaluation inputs into WkNN model were: (a) mean annual rainfall, (b) elevation, (c) slope, (d) soil water retention, (e) drainage proximity, (f) drainage density, (g) soil erosion, (h) inferred spatial results of WkNN, (i) inferred spatial of kNN; (j) WESR data; and extracted spatial results of WkNN (k) and kNN(L) in WESR extent.

Topographic Features

Topography is the key driver for flood formation and redistribution (Figure 2c). Generally, a lower area has a fairly higher flood risk since it can easily be inundated by surrounding water. In this study, two indices were extracted to represent topography. They are elevation (Figure 3b) and slope (Figure 3c) [48]. Elevation is the height above a fixed reference point, regularly the mean sea level. The area with lower elevation suffers from higher flood inundation, and vice versa [31]. Slope is the

steepness or the degree of incline of a surface. The steeper surface has a lower probability of flood inundation since water easily runs to low-lying land. Both two indices were produced from DEM at 30 m resolution from the United States Geological Survey (<https://earthexplorer.usgs.gov>).

Soil Water Retention (SWR)

Diverse soil types have different water infiltration and storage. The former was determined by soil type and vegetation on the surface. Usually, drier soil needs larger water flows for inundation than moisture or wet ones. Meanwhile, the probability of flood hazards decreases with an increase in soil infiltration [47]. In flood investigations, soil infiltration rates can be reflected by the Hydrologic Soil Group (HSG) which can be further classified into four subgroups from the highest rate (A) to the lowest rate (D). Subgroup A includes sandy characteristics, such as sandy loam, and Subgroup D has clay features, such as silty clay or clay. Subgroup B and subgroup C consist of (silt) loam, and sandy clay loam, respectively.

The soil storage indicates the amount of water present in the soil, which decides the occurrence of flood inundation. The potential maximum Soil Water Retention (SWR) can mirror how much water is in the soil and can be calculated with a hydrological modeling method that is driven by the Soil Conservation Service Curve Number (SCS-CN) [49,50]. The SCS-CN values were together calculated by hydrological features, soil type (Figure 2d), and land use (Figure 2e, Table 1) [51], and referenced from the list of Soil Conservation Service [52]. The CN values can be calculated by intersecting HSG, soil type, and land cover. Based on the CN approach, the SWR (Figure 3d) in cells can be calculated by using:

$$SWR_i = SWR_0 \left(\frac{100}{CN_i} - 1 \right) \quad (5)$$

Where $CN_i \in (0, 100)$ is the CN value of an i th cell, $SWR_0 = 254$ for units of millimeters.

Table 1. CN values under soil type.

Soil Type	A	B	C	D
Farmland	72	82	88	92
Forest	36	60	73	79
Grass	39	61	74	80
Bush	36	60	74	80
Wetland	32	58	72	79
Man-made Land	89	92	94	95
Barren	72	82	88	90
Water	100	100	100	100

Drainage System

A drainage system (Figure 2f) needs to be considered in the evaluation of FIR since it determines the change of flood inundation and easily causes an overflowing flood. Frequently, two main indices of drainage systems influence the distribution of flood inundation on the earth's surface. They are drainage proximity (Figure 3e) and drainage density (Figure 3f). The proximity denotes the distance to the nearest rivers or other water bodies, and the drainage density refers to the length of rivers per unit area. Based on our previous research [31], the areas near drainage networks within 200 m are assigned as high FIR, and the risk level decrease with the increase in distance [53]. The two factors were attained from Global River Database with Multiple Buffer operators and the Line Density function using a 1 km radius in GIS [54,55].

Soil Erosion

Soil erosion refers to the natural process of soil being moved from one location to another by natural environmental factors (e.g. water), and human activities such as deforestation, overgrazing, and improper agricultural practices [56]. Under the same rainfall conditions, areas with severe soil erosion are much more probable to be inundated by flood than areas with well-preserved surface vegetation, since the bare surface has a lower capacity for controlling water. Hence, soil erosion has a greater impact on the form and distribution of flood inundation, which can increase the risk of flood inundation [57]. In this study, soil data at 1 km was accessed from the Geographical Information Monitoring Cloud Platform (<http://www.dsac.cn/>) and processed in ArcGIS (Figure 2g, Figure 3g).

3.2.1 Detection of Maximum Inundation Extent

Remote Sensing (RS) has been applied in many flood-related investigations for many years. Its images are more easily acquired, but they have some restrictions. For instance, Moderate Resolution Imaging Spectroradiometers (MODIS) have lower spatial resolutions (500 m) but are commonly negatively affected by cloud cover, which is tricky to find and may miss out on flash rainfall events. Necessarily, other data sources should be found and replaced RS images in FIR. World Environment Situation Room platform (WESR, <https://wesr.unepgrid.ch/?project=MX-XVK-HPH-OGN-HVE-GGN&language=en>) delivers practicable replacements for FIR. It provides global dynamic data and systematic tools from different sources, as well as visualization tools that enable users to interact with and explore the data online [58]. It also assists users to observe and analyze environmental issues and trends, and to formulate effective environmental policies and protections. WESR products have been illustrated and employed in scientific investigation in data-scarce regions as well as developing countries, which can be extremely helpful to increase the preparedness and awareness of the population and reduce catastrophic impacts [59]. In the flood risk field, WESR provides six-year return periods (YRP) at 1 km resolution. They are 1-in-25, 1-in-50 (Figure 2h), 1-in-100, 1-in-200, 1-in-500, and 1-in-1,000 YRP. All data have been cross-checked with satellite flood footprints from various data sources and shown high accuracy well. The six YRP maps were reassigned if a cell value is greater than 0, it was assigned 1, otherwise it was assigned 0 in R environment. The six reassigned maps were overlaid to derive an inundation frequency map as a maximum inundation extension (MIE, Figure 3j). In MIE, cell values range from 0 to 6 representing very low risk to high risk. High risk means extremely vulnerable and frequent inundation from 1-in-25 to 1-in-1,000 YPR floods. MIE was selected as reference imagery to verify inferred maps derived from *k*NN and *Wk*NN.

3.2.2 Criteria Standardization

The criterion indices have different value ranges, which is difficult for modeling and reasoning FIR. To improve the efficiency of calculation, all the spatial input data were converted into 1 to 4 which represents very low to high risk using specific values (Figure 3a-g). There are many methods to standard the criterion indices, such as domain knowledge and Natural Breaks (Jenks). In the study, all input criteria were scandalized using Natural Breaks (Jenks) method in ArcGIS and R programming since Natural Breaks (Jenks) is based on natural groupings inherent in the data and is well employed in FIR [28,29]. In addition, all spatial datasets were then projected, resampled to 1 km grid cells, clipped to the study area, and registered, so all input grids accurately overlaid with the same projection, cell size, and extent.

4. Result and Discussions

A flood risk map was plotted as a derived result of the spatial evaluation framework for the FIR. The flood risk is divided into four levels: high (red), medium (orange), low (yellow), and very-low risk (green).

4.1 Result Verification

The innovative *Wk*NN-based spatial framework effectively produced the spatial distribution of FIR for the whole study area. To validate the evaluation accuracy (EA) of *Wk*NN model, an accuracy

comparison was conducted between spatially inferred results (Figure 3k) extracted from Figure 3h against MIE (Figure 3j). Overall, 80.59% of WkNN results reasonably confirm the actual MIE where cell values > 0 . Among the matched areas, 80.14%, 90.13%, 65.50%, and 84.14% of the predicted categories in WkNN area (Figure 3k) are well matched with the MIE area (Figure 3j) in high, medium, low, and very-low risk, respectively. It reflects that WkNN results (Figure 3h) are sound and reasonable. The remaining mistakes could be explained by the uncertainties and a little bit of inaccuracy of WESR data in certain areas. Moreover, it should be noted that the predicted risk extent is larger than the WESR data. The reason may be that the extension of WESR data is insufficient, which can be vividly shown in the empty circle in the northeast area, and some areas have no data.

4.2 Sensitivity Analysis

A sensitivity analysis is essential to explore the relationships between inputs and outputs of models, which can picture the performance, structures, and uncertainty of models. For WkNN-based models, the sampling datasets and the k values determine the EA of models and inferred outcomes.

4.2.1 Sensitivity Analysis in Relation to Sampling Times

For exploring the relationship between sampling times and EA, the study continually increased the number of random sampling datasets under an unchanged k value and observed the trend of EA. Two values were exemplified in k values: 5 (blue points in Figure 4A) and 95 (green points in Figure 4A). The formal was selected randomly and the latter roughly equals the square root of training datasets (8,947) [38]. Overall Accuracy (OA) was chosen to evaluate the performance of k NN and WkNN against MIE data. OA refers to the proportion of correct predictions made by models or systems over the total number of predictions [60,61], which can directly reflect EA and is easy to understand and use. Figure 4A shows that the OA of k NN is totally under 0.55 when k equals 5, and around about 0.57 when k equals 95. Under different sampling times (from 1 to 500), the range of both OA is relatively larger, which shows that k NN method is unstable. However, compared with k NN, WkNN method (in red) shows comparative robustness since the OA range of WkNN is aggregated around the medium value (about 0.58), which additionally demonstrates that the predicted performance of WkNN is higher than that of k NN.

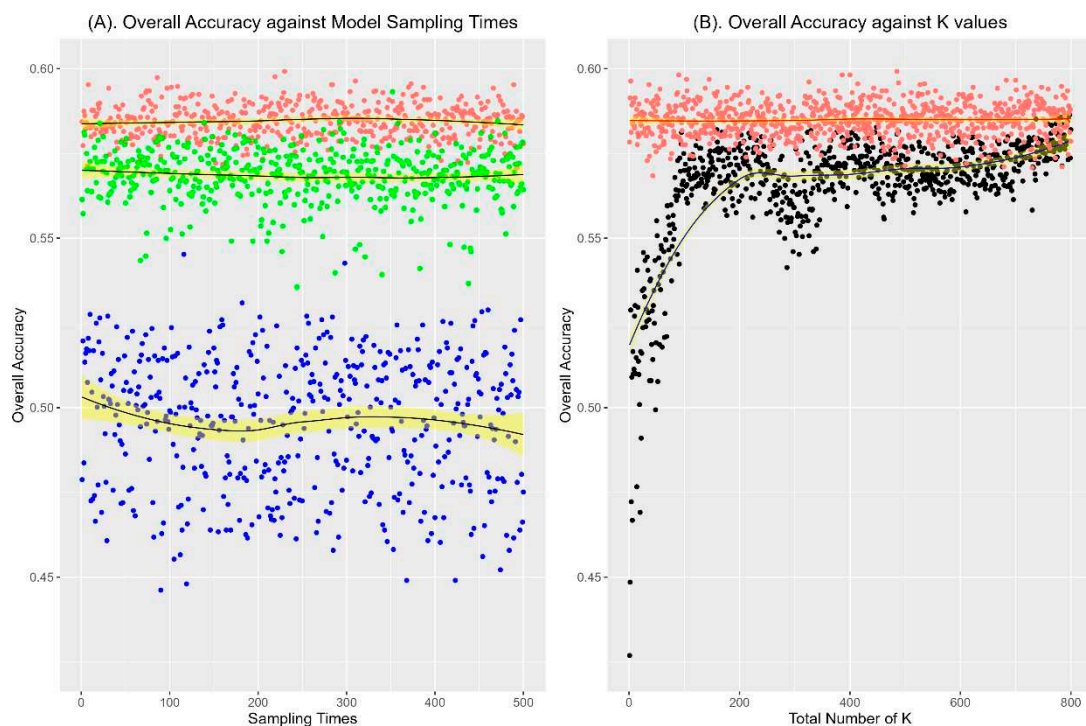


Figure 4. Evaluation accuracy of WkNN and kNN against model sampling times (A), and k values (B).

4.2.2. Sensitivity Analysis in Relation to k Values

K values play a key role in model performance. An appropriate k value determines the robust and predicted results of k NN models. Conversely, inappropriate k values will cause the problem of bias-various tradeoffs [38]. Therefore, the study explored and compared the performance and influence of k values in k NN and WkNN, and selected the optimal k values (the highest value in OA) to infer the FIR for the whole study area. In this study, the range of k values from 1 to 800 covers the square root (95 and 315) of observations (8,947) and the whole study dataset (98,709). The results show that the OA of k NN, which ranges from 0.43 to 0.58, is lower than the OA of WkNN, ranging from 0.57 to 0.60. The trend in EA of WkNN is reasonably stable. Overall, OA increases with the growth of k values, particularly, it shows non-linear increases when k values are between 1 to 200 (Figure 4B). When k values are between 200 and 400, OA presents a declining trend, but after 400, the value of EA increases slowly. All these demonstrate that k values have a significant impact on k NN and WkNN, but the latter performed more robustly.

4.3. Comparison of WkNN with k NN

The evaluation results of WkNN (Figure 3h) using Equation (4) were demonstrated by a comparison with those of a published spatial-based k NN method (Figure 3i) [31] using Equation (1). The comparison shows that WkNN is better than k NN, such as in sampled areas 1 to 3, because their results are more similar to the reference MIE (Figure 3j) in visualization. Also, three areas were sampled in the north (area 4 in a grey rectangle), west (area 5 in a blue rectangle), and southeast (area 6 in a purple rectangle) to compare evaluation accuracy (EA) between WkNN (Figure 3k) and k NN (Figure 3l) against MIE (Figure 3j). Area 4 shows that inferred WkNN's results well match the pattern of MIE with values > 0 . However, k NN's outcomes represented opposite results in some areas, which means high risk in MIE showing low risk or even very low risk in k NN results. These mismatches have also occurred in other regions, such as areas 5 and 6. All these reflect that the WkNN method has higher prediction accuracy than k NN.

4.4. Risk Distribution Analysis

The resulting map (Figure 3h) illustrates the extent and distribution of FIR in the whole study area. The statistics of each flood category from high risk to very low were conducted, which demonstrates about half of the study area is classified as medium-high risk. Around 2.85% of the whole area is covered by high risk, 64.83% is medium risk, 10.8% is low risk and 21.52% is very low risk. High-risk areas can be observed, particularly, in the southwestern area that is affected by elevation (57.63 km²) and precipitation (36.13 km²). The result demonstrates that elevation is the key factor in the form and redistribution of flooding inundation. Besides, high-risk areas are scattered across the eastern part, which is heavily affected by the slope (about 32.77 km²) and SWR (about 24.86 km²). In the northern area, there is also the high-risk area under lower evaluation. Most areas are covered by medium risk in flood inundation, which is larger than high risk, since compared with medium flood risk, high flood risk rarely happens, unless extraordinary climate events and fragile environments, such as extreme rainfall and bare ground. The medium risk is mutually affected by multiple factors over the center area. Elevation (152.55 km²), Precipitation (76.84 km²) contribute greatly to the central part across west to east. In the fringe of the study area, the medium risk is mainly affected by drainage density and proximity. The high-medium area demonstrates that elevation and precipitation are the two extremely important factors for FIR. Slope, erosion, and proximity make a similar contribution to high-medium risk since their areas are 32.77 km² and 73.45 km² in high risk and medium risk, separately. River density and SWR have little contribution to flood high-medium risk. Based on prior experience [43], high density of rivers or areas nearer to the rivers should have high or medium flood risk, but actually, areas have low or very low flood risk in the study. That is mainly because drainage systems carry more flood into the sea and mitigate the stress of flood

inundation pressure for the study area. Besides, larger cities, such as Hangzhou, are mainly located in low or very low-risk areas, which vividly demonstrates that man-made water conservancy facilities play a huge role in protecting socio-economic development and alleviating flood risk.

Figure 5 shows tourism facilities expose to predicted FIR. To highlight tourism facilities in high-medium risk, the facilities with a higher risk level will have a bigger point size. It illustrated that most tourism facilities in coastal cities are at low or very low risk, especially from Hangzhou-centered northern coastal areas to southern Wenzhou areas. However, in most inland areas, some tourism facilities are located in high-medium risk areas of flood inundation, especially hotels (Figure 5a), medical treatment institutions (Figure 5b), and restaurants (Figure 5e). Moreover, it should be noted that the number of medical treatment institutions in high-risk areas is larger than other facilities. It can easily be understood that medical facilities in high-risk areas will play a key significance in effectively mitigating or even preventing the negative impact of flood inundation on tourism facilities and saving lives. Parks (Figure 5c) and parking places (Figure 5d) mainly locate at low or very-low risk since most of them are in urban areas. Besides, the study area has a wonderful road system (Figure 5f) at the national-provincial levels. The majority of roads locates in medium-or-low-risk areas of flood inundation, which not only helps to develop tourism resources but also efficiently carries out disaster relief, and post-disaster reconstruction. Lastly, Figure 3h and Figure 5 demonstrate that tourist facilities and road infrastructure are at a low-risk level in the cities and nearby areas. It illustrates local departments have done a lot of practical and efficient work on disaster prevention and mitigation in coastal flood-prone areas. It can prove that engineering measures play a key role in protecting socioeconomic activities including tourism, which can provide a valuable reference for the vast coastal areas around the world.

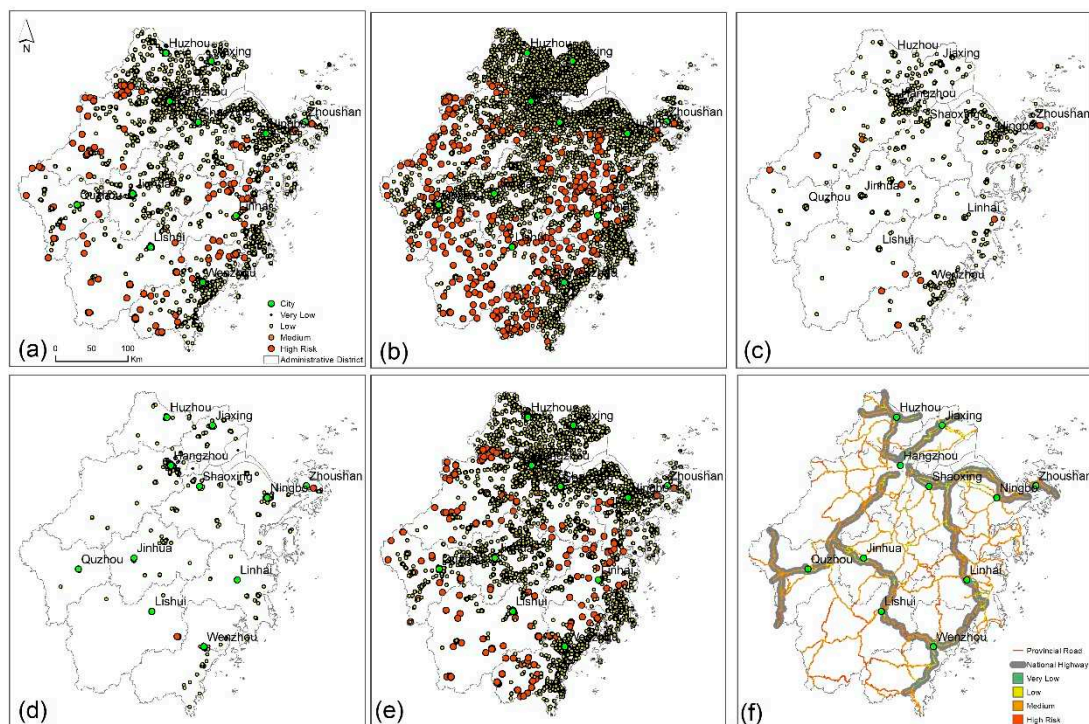


Figure 5. Tourist facilities located in FIR areas, (a) hotels, (b) medical treatment institutions, (c) parks, (d) parking places, (e) Restaurants, and (f) national and provincial roads.

5. Conclusions

This study develops an innovative spatial framework, which integrates WkNN, GIS, and other flood-related indices to infer, map, and evaluate the distribution of flood inundation for tourism. The improved WkNN was developed based on kNN by using the weights method which is inversely proportional to distance in GIS. GIS was used as a spatial tool to derive flood-influenced indices via

collecting and processing the number of spatial factors at multitemporal, and multispatial resolution from different sources. Among flood-related factors, WESR was used as predicted result validation in FIR for tourism for the first time. The WkNN-based framework was effectively carried out in the case study, obtained reasonable outcomes, and further demonstrated WkNN is superior to kNN in flood risk analysis and evaluation accuracy (EA). Meanwhile, k values are still significant parameters for kNN and WkNN. Suitable k values will improve the performance of models in EA. The WkNN outcomes can well match WESR data, which can deliver the fundamentals for flood disaster prevention and mitigation for tourism in a coastal area, and assist decision-makers adopt effective measures for preventing and mitigating the negative impacts of flood disasters.

The innovative spatial framework was programmed and repeatable with GIS and R programming, which can be flexibly used in other disaster-related investigations, and also not limited by the number of model inputs. The evaluation results will make corresponding changes responsive to different input indices. However, there are some limitations the study did not consider. For example, due to the limitation of data sources, the study did not fully use Remote Sensing imagery, such as Synthetic Aperture Radar, in flood risk assessment. Besides, the research did not assess the adverse economic consequences of flooding on the tourism industry. As a further step, we plan to probe deeply into these fields and provide more precise assessments.

Author Contributions: For research articles with several authors, the following statements should be used “Conceptualization, R.L. and S.L.; methodology, R.L.; software, N.T.; validation, R.L., S.L. and N.T.; formal analysis, S.L.; investigation, S.L.; resources, S.L.; data curation, N.T.; writing—original draft preparation, S.L.; writing—review and editing, R.L.; visualization, N.T.; supervision, R.L.; project administration, S.L.; funding acquisition, S.L. All authors have read and agreed to the published version of the manuscript.”

Funding: The study was funded by the Soft Science Research Program of Zhejiang (2023C35072): Research on the Formation Mechanism and Evaluation Strategy of Flood Disaster Risk in Coastal Scenic Areas of Zhejiang Province. The authors are grateful for Huzhou City propaganda ideology and culture young talents propaganda ideology and culture young talents. We are also appreciative for reviewing the article. The anonymous reviewers are acknowledged for their valuable comments.

Acknowledgments: The authors are grateful for reviewing the article. The anonymous reviewers are acknowledged for their valuable comments.

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

References

1. Papageorgiou, M. Coastal and marine tourism: A challenging factor in Marine Spatial Planning. *Ocean & Coastal Management* **2016**, *129*, 44-48, doi:10.1016/j.ocecoaman.2016.05.006.
2. Chen, Y.H.; Kim, J.; Mueller, N. Estimating the economic impact of natural hazards on shared accommodation in coastal tourism destinations. *Journal of Destination Marketing & Management* **2021**, *21*, 100634, doi:10.1016/j.jdmm.2021.100634.
3. Lithgow, D.; Martínez, M.L.; Gallego-Fernández, J.B.; Silva, R.; Ramírez-Vargas, D.L. Exploring the co-occurrence between coastal squeeze and coastal tourism in a changing climate and its consequences. *Tourism Management* **2019**, *74*, 43-54, doi:10.1016/j.tourman.2019.02.005.
4. Perles-Ribes, J.F.; Ramón-Rodríguez, A.B.; Moreno-Izquierdo, L.; Such-Devesa, M.J. Machine learning techniques as a tool for predicting overtourism : The case of Spain. *International Journal of Tourism Research* **2020**, *22*, 825-838, doi:10.1002/jtr.2383.
5. Junqing, H.; Han, T.; Jiawei, H.; Yanting, M.; Xinxiang, J. Impacts of risk perception, disaster knowledge and emotional attachment on tourists' behavioral intentions in Qinling Mountain, China. *Frontiers in Earth Science* **2022**, 932.
6. Morsy, M.; Sayad, T.; Khamees, A.S. Towards instability index development for heavy rainfall events over Egypt and the Eastern Mediterranean. *Meteorology and Atmospheric Physics* **2019**, *132*, 255-272, doi:10.1007/s00703-019-00686-5.
7. Zhang, Y.; Wang, Y.; Chen, Y.; Xu, Y.; Zhang, G.; Lin, Q.; Luo, R. Projection of changes in flash flood occurrence under climate change at tourist attractions. *Journal of Hydrology* **2021**, *595*, doi:10.1016/j.jhydrol.2021.126039.
8. Vitousek, S.; Barnard, P.L.; Fletcher, C.H.; Frazer, N.; Erikson, L.; Storlazzi, C.D. Doubling of coastal flooding frequency within decades due to sea-level rise. *Scientific Reports* **2017**, *7*, 1399, doi:10.1038/s41598-017-01362-7.

9. Lyu, H.M.; Wang, G.F.; Cheng, W.C.; Shen, S.L. Tornado hazards on June 23 in Jiangsu Province, China: preliminary investigation and analysis. *Natural Hazards* **2016**, *85*, 597-604, doi:10.1007/s11069-016-2588-2.
10. Kundzewicz, Z.W.; Su, B.; Wang, Y.; Wang, G.; Wang, G.; Huang, J.; Jiang, T. Flood risk in a range of spatial perspectives – from global to local scales. *Natural Hazards and Earth System Sciences* **2019**, *19*, 1319-1328, doi:10.5194/nhess-19-1319-2019.
11. Sun, S.; Zhai, J.; Li, Y.; Huang, D.; Wang, G. Urban waterlogging risk assessment in well-developed region of Eastern China. *Physics and Chemistry of the Earth, Parts A/B/C* **2020**, *115*, doi:10.1016/j.pce.2019.102824.
12. Fang, J.; Zhang, C.; Fang, J.; Liu, M.; Luan, Y. Increasing exposure to floods in China revealed by nighttime light data and flood susceptibility mapping. *Environmental Research Letters* **2021**, *16*, 104044, doi:10.1088/1748-9326/ac263e.
13. Kumar, M.D.; Tandon, S.; Bassi, N.; Mohanty, P.K.; Kumar, S.; Mohandas, M. A framework for risk-based assessment of urban floods in coastal cities. *Natural Hazards* **2021**, *110*, 2035-2057, doi:10.1007/s11069-021-05024-w.
14. Lu, Y.; Ren, F.; Zhu, W. Risk zoning of typhoon disasters in Zhejiang Province, China. *Natural Hazards and Earth System Sciences* **2018**, *18*, 2921-2932, doi:10.5194/nhess-18-2921-2018.
15. Liang, H.; Zhou, X. Impact of tides and surges on fluvial floods in coastal regions. *Remote Sensing* **2022**, *14*, doi:10.3390/rs14225779.
16. Thaler, T.; Seebauer, S. Bottom-up citizen initiatives in natural hazard management: Why they appear and what they can do? *Environmental Science & Policy* **2019**, *94*, 101-111, doi:10.1016/j.envsci.2018.12.012.
17. Cronin, S.J.; Gaylord, D.R.; Charley, D.; Alloway, B.V.; Wallez, S.; Esau, J.W. Participatory methods of incorporating scientific with traditional knowledge for volcanic hazard management on Ambae Island, Vanuatu. *Bulletin of Volcanology* **2004**, *66*, 652-668, doi:10.1007/s00445-004-0347-9.
18. Lai, X.; Wen, J.; Shan, X.; Shen, L.; Wan, C.; Shao, L.; Wu, Y.; Chen, B.; Li, W. Cost-benefit analysis of local knowledge-based flood adaptation measures: A case study of Datian community in Zhejiang Province, China. *International Journal of Disaster Risk Reduction* **2023**, *87*, 103573, doi:10.1016/j.ijdr.2023.103573.
19. Mendelsohn, R.; Emanuel, K.; Chonabayashi, S.; Bakkensen, L. The impact of climate change on global tropical cyclone damage. *Nature Climate Change* **2012**, *2*, 205-209, doi:10.1038/nclimate1357.
20. Baade, R.A.; Baumann, R.; Matheson, V. Estimating the economic impact of natural and social disasters, with an application to hurricane Katrina. *Urban Studies* **2016**, *44*, 2061-2076, doi:10.1080/00420980701518917.
21. Chen, Y.; Liu, R.; Barrett, D.; Gao, L.; Zhou, M.; Renzullo, L.; Emelyanova, I. A spatial assessment framework for evaluating flood risk under extreme climates. *Science of The Total Environment* **2015**, *538*, 512-523, doi:10.1016/j.scitotenv.2015.08.094.
22. Tsai, C.H.; Chen, C.W. Development of a mechanism for typhoon- and flood-risk assessment and disaster management in the hotel industry – A case study of the Hualien area. *Scandinavian Journal of Hospitality and Tourism* **2011**, *11*, 324-341, doi:10.1080/15022250.2011.601929.
23. Sealy, K.S.; Strobl, E. A hurricane loss risk assessment of coastal properties in the Caribbean: Evidence from the Bahamas. *Ocean & Coastal Management* **2017**, *149*, 42-51, doi:10.1016/j.ocecoaman.2017.09.013.
24. Shi, Y.; Wen, J.; Xi, J.; Xu, H.; Shan, X.; Yao, Q.; Xia, J. A study on spatial accessibility of the urban tourism attraction emergency response under the flood disaster scenario. *Complexity* **2020**, *2020*, 1-9, doi:10.1155/2020/9031751.
25. Erena, S.H.; Worku, H.; De Paola, F. Flood hazard mapping using FLO-2D and local management strategies of Dire Dawa city, Ethiopia. *Journal of Hydrology: Regional Studies* **2018**, *19*, 224-239, doi:10.1016/j.ejrh.2018.09.005.
26. Yildirim, E.; Demir, I. An integrated web framework for HAZUS-MH flood loss estimation analysis. *Natural Hazards* **2019**, *99*, 275-286, doi:10.1007/s11069-019-03738-6.
27. Chen, Y.; Wang, Y.; Zhang, Y.; Luan, Q.; Chen, X. Flash floods, land-use change, and risk dynamics in mountainous tourist areas: A case study of the Yesanpo Scenic Area, Beijing, China. *International Journal of Disaster Risk Reduction* **2020**, *50*, 101873, doi:10.1016/j.ijdr.2020.101873.
28. Liu, R.; Chen, Y.; Wu, J.; Gao, L.; Barrett, D.; Xu, T.; Li, X.; Li, L.; Huang, C.; Yu, J. Integrating entropy-based naïve Bayes and GIS for spatial evaluation of flood hazard. *Risk Analysis* **2017**, *37*, 756-773, doi:10.1111/risa.12698.
29. Liu, R.; Chen, Y.; Wu, J.; Gao, L.; Barrett, D.; Xu, T.; Li, L.; Huang, C.; Yu, J. Assessing spatial likelihood of flooding hazard using naïve Bayes and GIS: a case study in Bowen Basin, Australia. *Stochastic Environmental Research and Risk Assessment* **2015**, *30*, 1575-1590, doi:10.1007/s00477-015-1198-y.
30. Stefanidis, S.; Stathis, D. Assessment of flood hazard based on natural and anthropogenic factors using analytic hierarchy process (AHP). *Natural Hazards* **2013**, *68*, 569-585, doi:10.1007/s11069-013-0639-5.
31. Liu, S.; Liu, R.; Tan, N. A spatial improved-kNN-based flood inundation risk framework for urban tourism under two rainfall scenarios. *Sustainability* **2021**, *13*, 2859, doi:10.3390/su13052859.
32. Liu, K.; Yao, C.; Chen, J.; Li, Z.; Li, Q.; Sun, L. Comparison of three updating models for real time forecasting: a case study of flood forecasting at the middle reaches of the Huai River in East China. *Stochastic Environmental Research and Risk Assessment* **2016**, *31*, 1471-1484, doi:10.1007/s00477-016-1267-x.

33. Liu, K.; Li, Z.; Yao, C.; Chen, J.; Zhang, K.; Saifullah, M. Coupling the k-nearest neighbor procedure with the Kalman filter for real-time updating of the hydraulic model in flood forecasting. *International Journal of Sediment Research* **2016**, *31*, 149-158, doi:10.1016/j.ijsrc.2016.02.002.
34. Cassalho, F.; Beskow, S.; Mello, C.R.; Moura, M.M.; Oliveira, L.F.; Aguiar, M.S. Artificial intelligence for identifying hydrologically homogeneous regions: A state-of-the-art regional flood frequency analysis. *Hydrological Processes* **2019**, *33*, 1101-1116, doi:10.1002/hyp.13388.
35. Aryal, S.; Ting, K.M.; Washio, T.; Haffari, G. A comparative study of data-dependent approaches without learning in measuring similarities of data objects. *Data Mining and Knowledge Discovery* **2019**, *34*, 124-162, doi:10.1007/s10618-019-00660-0.
36. Rodrigues, É.O. Combining Minkowski and Chebyshev: New distance proposal and survey of distance metrics using k-nearest neighbours classifier. *Pattern Recognition Letters* **2018**, *110*, 66-71, doi:10.1016/j.patrec.2018.03.021.
37. Kutylowska, M. K-Nearest Neighbours method as a tool for failure rate prediction. *Periodica Polytechnica Civil Engineering* **2017**, *62*, 318-322, doi:10.3311/PPci.10045.
38. Lantz, B. *Machine learning with R: expert techniques for predictive modeling*; Packt publishing ltd: 2019.
39. Malekinezhad, H.; Sepehri, M.; Pham, Q.B.; Hosseini, S.Z.; Meshram, S.G.; Vojtek, M.; Vojteková, J. Application of entropy weighting method for urban flood hazard mapping. *Acta Geophysica* **2021**, *69*, 841-854, doi:10.1007/s11600-021-00586-6.
40. Rahman, A.S.; Rahman, A. Application of principal component analysis and cluster analysis in regional flood frequency analysis: A case study in New South Wales, Australia. *Water* **2020**, *12*, doi:10.3390/w12030781.
41. Fang, Y.; Yin, J.; Wu, B. Flooding risk assessment of coastal tourist attractions affected by sea level rise and storm surge: a case study in Zhejiang Province, China. *Natural Hazards* **2016**, *84*, 611-624, doi:10.1007/s11069-016-2444-4.
42. Papaioannou, G.; Vasiliades, L.; Loukas, A. Multi-criteria analysis framework for potential flood prone areas mapping. *Water Resources Management* **2015**, *29*, 399-418.
43. Xiao, Y.; Yi, S.; Tang, Z. Integrated flood hazard assessment based on spatial ordered weighted averaging method considering spatial heterogeneity of risk preference. *Science of the Total Environment* **2017**, 599-600, 1034-1046, doi:10.1016/j.scitotenv.2017.04.218.
44. Yatagai, A.; Kamiguchi, K.; Arakawa, O.; Hamada, A.; Yasutomi, N.; Kitoh, A. APHRODITE: constructing a long-term daily gridded precipitation dataset for Asia based on a dense network of rain gauges. *Bulletin of the American Meteorological Society* **2012**, *93*, 1401-1415, doi:10.1175/bams-d-11-00122.1.
45. Tao, G.; Lian, X. Study on progress of the trends and physical causes of extreme precipitation in China during the last 50 years. *Advances in Earth Science* **2014**, *29*, 577.
46. Feng, L.; Hong, W. Characteristics of drought and flood in Zhejiang Province, East China: past and future. *Chinese Geographical Science* **2007**, *17*, 257-264, doi:10.1007/s11769-007-0257-9.
47. Ouma, Y.; Tateishi, R. Urban flood vulnerability and risk mapping using integrated multi-parametric AHP and GIS: methodological overview and case study assessment. *Water* **2014**, *6*, 1515-1545, doi:10.3390/w6061515.
48. Lyu, H.M.; Sun, W.J.; Shen, S.L.; Arulrajah, A. Flood risk assessment in metro systems of mega-cities using a GIS-based modeling approach. *Science of The Total Environment* **2018**, *626*, 1012-1025, doi:10.1016/j.scitotenv.2018.01.138.
49. McCuen, R.H. *A guide to hydrologic analysis using SCS methods*; Prentice-Hall, Inc.: 1982.
50. Vijji, R.; Rajesh Prasanna, P.; Ilangoan, R. GIS based SCS-CN method for estimating runoff in Kundahpalam watershed, Nilgries District, Tamilnadu. *Earth Sciences Research Journal* **2015**, *19*, 59-64.
51. Cao, B.; Yu, L.; Naipal, V.; Ciais, P.; Li, W.; Zhao, Y.; Wei, W.; Chen, D.; Liu, Z.; Gong, P. A 30 m terrace mapping in China using Landsat 8 imagery and digital elevation model based on the Google Earth Engine. *Earth System Science Data* **2021**, *13*, 2437-2456.
52. SCS, Soil Conservation Service. *Urban hydrology for small watersheds* SCS; Engineering Division, Soil Conservation Service, US Department of Agriculture: 1986.
53. Kazakis, N.; Kougias, I.; Patsialis, T. Assessment of flood hazard areas at a regional scale using an index-based approach and Analytical Hierarchy Process: Application in Rhodope-Evros region, Greece. *Science of the Total Environment* **2015**, *538*, 555-563, doi:10.1016/j.scitotenv.2015.08.055.
54. Strahler, A.N. Quantitative geomorphology of drainage basin and channel networks. *Handbook of Applied Hydrology* **1964**.
55. Andreadis, K.M.; Schumann, G.J.P.; Pavelsky, T. A simple global river bankfull width and depth database. *Water Resources Research* **2013**, *49*, 7164-7168, doi:10.1002/wrcr.20440.
56. Park, S.; Oh, C.; Jeon, S.; Jung, H.; Choi, C. Soil erosion risk in Korean watersheds, assessed using the revised universal soil loss equation. *Journal of Hydrology* **2011**, *399*, 263-273, doi:10.1016/j.jhydrol.2011.01.004.
57. Li, X.; Wei, X. Analysis of the relationship between soil erosion risk and surplus floodwater during flood season. *Journal of Hydrologic Engineering* **2014**, *19*, 1294-1311, doi:10.1061/(asce)he.1943-5584.0000912.

58. Jabbour, J.; Caldas, A.; Peduzzi, P. The World Environment Situation Room: rethinking environmental assessment. In Proceedings of the AGU Fall Meeting Abstracts; pp. GH11D-0932.
59. Dottori, F.; Salamon, P.; Bianchi, A.; Alfieri, L.; Hirpa, F.A.; Feyen, L. Development and evaluation of a framework for global flood hazard mapping. *Advances in Water Resources* **2016**, *94*, 87-102, doi:10.1016/j.advwatres.2016.05.002.
60. Foody, G.M. Status of land cover classification accuracy assessment. *Remote Sensing of Environment* **2002**, *80*, 185-201.
61. Liu, C.; Frazier, P.; Kumar, L. Comparative assessment of the measures of thematic classification accuracy. *Remote sensing of environment* **2007**, *107*, 606-616.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.