ORIGINAL ARTICLE

An integrated fluvial and flash pluvial model using 2D high-resolution sub-grid and particle swarm optimization-based random forest approaches in GIS

Hossein Mojaddadi Rizeei[1] · Biswajeet Pradhan[1] · Maryam Adel Saharkhiz[1]

Abstract

Two types of flooding, namely fluvial flood (FF) and pluvial flash flood (PFF), exist in tropical cities located close to permanent rivers, where extreme precipitation intensity occurs. Although several methods are available for assessment of FF, however, PFF has received minimal attention from the researchers. Studies rarely presented joint FF and PFF hazards. Therefore, the current study not only aims to evaluate probability and hazards for FF and PFF independently but also implements combined FF with PFF probabilistic inundation analysis. First, an integrated model was developed to analyze probability using fully distributed geographic information system (GIS)-based algorithms. These methods were performed on Damansara River Catchment in Kuala Lumpur, because yearly monsoon triggers FFs and simultaneously coincides with heavy local rainfalls. A hydraulic 2D high-resolution sub-grid model of Hydrologic Engineering Center River Analysis System was performed to simulate FF probability and hazard. Nine significant contributing parameters were trained with PFF inventory by GIS-based random forest (RF) model and each RF parameter was optimized by particle swarm optimization algorithm (PSO) to model the PFF probabilistic hazard. Finally, PFF was combined with FF probabilities to discover the impact and contribution of each type of urban flood hazard. This study is the first attempt to model PFF hazard using GIS and physical-based PSO–RF model and combined FF and PFF probabilistic map. The results provide detailed flood information for urban managers to smartly equip infrastructures, such as highways, roads, and sewage network.

Keywords Fluvial flooding · Pluvial flash flooding · Remote sensing · Random forest · PSO · GIS

Introduction

Flooding is one of the most frequent natural disasters; thus analysis, forecasting, and modeling flood at various temporal scenarios and spatial scales bear significance [1]. Particularly, tropical countries have received significant attention in flood mitigation plans because of frequent occurrences of urban floods. Malaysia suffers from frequent flood events especially during the monsoon period [2]. Despite that, flood events are generally unanticipated to a high certain extent, they can be governed using recent and precise deterministic and probabilistic flood modeling to predict flood and decrease the amount of damages or losses [3].

Simulated flood inundation and flood plain system can provide significant information, benefitting probability and emergency cases to mitigate loss and damages to human lives and properties [4]. A significant portion of the urban flood damages occur especially in dense populations and areas of concentrated urban infrastructures [5].

Different types of flooding, namely fluvial flood (FF) and pluvial flash flood (PFF), typically occur in urban regions located next to rivers [6]. In tropical cities, such as Kuala Lumpur with many impervious urban surfaces that prevent ground infiltration and presence of many permanent rivers flowing inside the city, both types of floods may possibly occur. Such situation is aggravated at cities in river deltas, wherein high intensity of precipitation limits drainage capacity, and long duration of precipitation in upstream areas causes additional fluvial inundation [7,8].

PFF is considered quick flooding caused by heavy raining. Within a short time, high-precipitation intensity results in impervious surface without sufficient infiltration. High-

✉ Biswajeet Pradhan
Biswajeet24@gmail.com; Biswajeet.Pradhan@uts.edu.au

1   School of Systems, Management and Leadership, Faculty of Engineering and IT, University of Technology Sydney, Sydney, NSW, Australia

precipitation intensity in this condition is typically rapid and occurs within a few minutes to some hours after raining [9].

FF exists when excessive precipitation over an extended period causes a river to exceed its capacity. Inundation flooding occurs once river water overflows over river banks [10]. Recently, as a result of rapid urbanization in tropical cities, PFF and FF prediction, flood probability, flood hazard and risk assessment, and operational flood mitigation preparation became a critical task in flood management and water resource planning [11].

Despite huge effects of urban flood by different inundation types, assessment of urban flood hazard is generally limited to simply one type of flooding (e.g., FF). An increasing number of urban FF analyses are provided in the literature, predominantly with advance computational capabilities and specific direction toward flood probability, hazard, and risk models [10–15]. Characterizing river behavior and fluvial floodplains with high potential for FFs can help specialists develop management approaches for overflow mitigation; these approaches include creating water control structures (dike and levee implementations) and facilitating disaster preparedness to handle situations before, during, and after flood occurrence [16].

To date, 2D hydraulic models are possibly the most well-known models to extend accurate flood mapping and flood hazard analysis projects [17]. Chang et al. (2010) combined a 1D hydraulic model using Hydrologic Engineering Center River Analysis System (HEC-RAS) with a travel forecast model to assess urban flooding on roads and highway network systems. However, the previous literature mostly focused on regional scale or large basin level only, with less courtesy on PFF at the urban level [18]. Given the lack of high spatial and temporal resolution data, studies were usually unable to explicitly deliver thorough information from 2D surface flooding on urban infrastructures [18].

Previous studies rarely discussed probabilistic analyses of pluvial floods. Natural parameters are the most effective contributing factors associated with pluvial flood probability; these factors include land use (LU) types, magnitude of intensity precipitation, geomorphology, altitude, surface slope, and hydrological characteristics and should not be overlooked [3]. Though understanding PFF risk in urban is highly on demand, to date, a few studies struggled to model the potential area of PFFs in urban infrastructures. This difficulty may be due to lack of sufficient observations and precise data, struggle in systematic PFF modeling dynamics, and complexity of quantifying inundation model by spatial heterogeneity caused by rainfalls [19]. Basically, flood modeling mostly relies on water-level gauging and meteorological station historical data, which feature some drawbacks. Stations are not well distributed among study areas and are placed on river sites which cannot record the amount of surface water outside rivers; lack of sufficient

number of stations used as historical inventory data also limits PFF probability and surface runoff simulation studies [20]. Intensity–duration–frequency curves are one of the gauging station-based traditional approaches for quantifying probability of rainfall events. However, stochastic, probabilistic, and spatial rainfall simulators were also established to accurately define possible precipitation spatial coverage and intensity and probability of pluvial occurrence instead of traditional approaches [21–23].

Recently, hydrological studies for determining flood mitigation used machine learning (ML) based geographic information system (GIS) and hydraulic models [24]. Accuracy and precision of ML approaches were previously tested for flood probability by numerous researchers [25,26]. Decision tree (DT), artificial neural network, and logistic regression algorithms are some examples of ML models, and they are capable of modeling flooding probability and hazards [27]. Although some ML models can produce acceptable results, they still feature some special weak points that require improvement [28].

Therefore, in the current study, a comprehensive methodology was proposed by combining FF and PFF probability modeling to quantify impacts of PFF and FF in urban areas. To anticipate PFF probability, coupled GIS-based random forest (RF) with particle swarm optimization (PSO) methods were implemented based on sufficient number of recorded historical PFF events. Additionally, 2D high-resolution sub-grid (2D-HRS) hydraulic model was performed to engage FF inundation probability. Finally, we combined PFF with FF models to discover impact and contribution of each type to urban flood hazard in Damansara catchment. The city center of Damansara, Malaysia was selected; in this area, residential buildings and urban infrastructures are prone to both FF and PFF events. The main objectives are as follows: (a) developing a sample RF-PSO physical model to describe PFF probability regions using recorded historical events (flash flood inventory) and maximum recorded rainfall intensity; (b) implementing hydraulic 2D-HRS approach to model FF probability and flood inundation depth measurement; and (c) integrating two different inundation probabilistic models to quantify probable type of flood events that may threaten urban inhabitants. Therefore, the approach was used as source for flood mitigation planning in Damansara City, where FFs and high-intensity rainfalls naturally occur at the same period and are not easily distinguished. Consequently, the developed model was utilized as a plan for flood hazard and risk assessments in these study areas.

Study area and dataset

Recently, rapid urban development at river watershed basin led to high runoff while causing increases in flood magnitude and frequency [29]. In the past decades, Damansara catch-
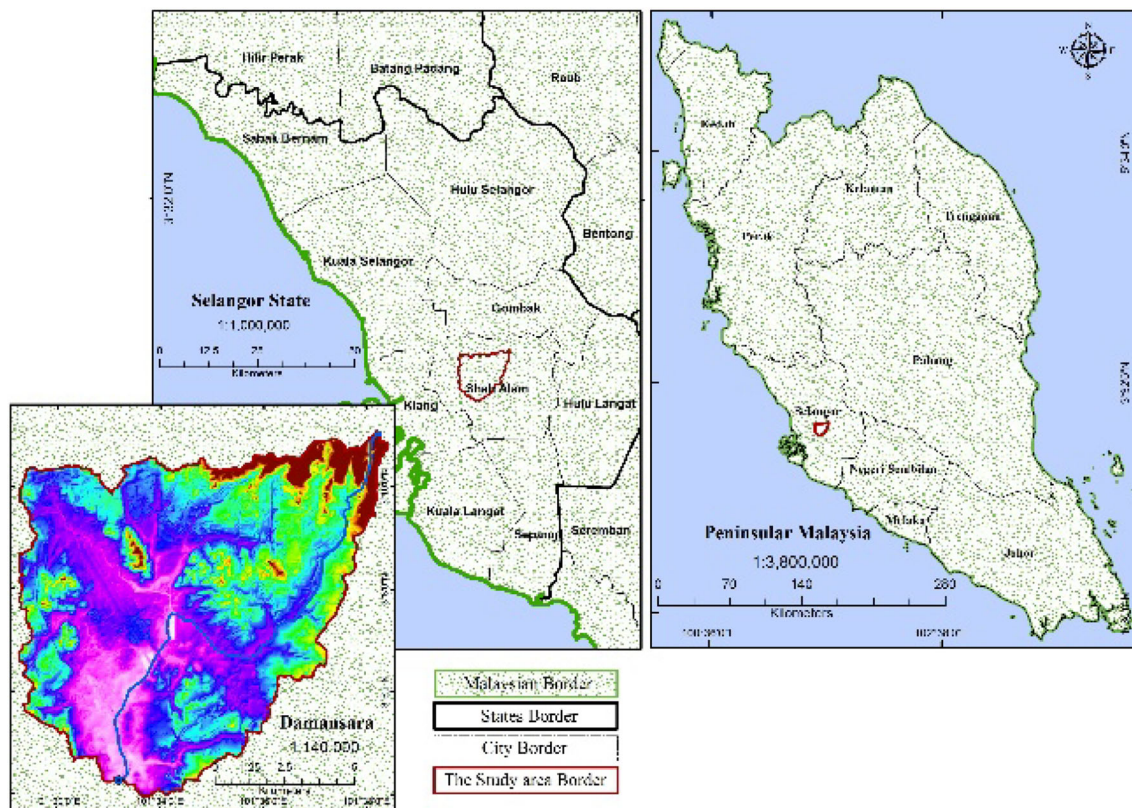
Fig. 1 Location of the study area

ment experienced different types of flood events, usually between November and February, because of the monsoon season [30]. Considering that this area is an urban environment (e.g., residential and commercial buildings and highways) and includes a part of Klang river watershed basin with some permanent rivers, it is prone to both PFF and FF.

The Damansara River catchment is located in Kuala Lumpur, Selangor, Malaysia. The study area is situated at 3°8′45.6″ latitude and 101°32′27.24″ longitude. This catchment measures almost 117 km^2 and considered a small watershed, (see Fig. 1).

Hydro-geomorphological characteristics of the study area, such as basin slope, area, and length, were calculated using GIS spatial analyst tools (Table 1).

Digital elevation model (DEM) was used for data analysis. DEM was extracted from interferometric synthetic aperture radar (InSAR) images at a pixel size of 5 × 5 m. High-resolution WorldView-3 satellite imagery was processed to extract land use (LUs) of the study area. Both satellite images were captured in 2015.

Precipitation and streamflow data were provided by the Department of Irrigation and Drainage, Selangor meteorological rain-gauge stations. In this study, 15-year-meteorological data, including hourly precipitation and hourly streamflow at stations, were investigated among 11

Table 1 Hydro-geomorphological details for Damansara River catchment

Hydrological characteristics	Values	Units
Basin slope	0.07	Degree
Maximum flow distance	22,222.66	Meter
Stream centroid to outlet	8426.41	Meter
Maximum stream length	21,473.58	Meter
Maximum stream slope	0.001	Degree
Basin length	15,625.50	Meter
Basin shape factor	2.09	Index
Basin perimeter	71,440.35	Meter
Basin average elevation	37.86	Meter
CENTROID X	397,595.17	Longitude
CENTROID Y	346,340.79	Latitude
Basin LA County lateral	65.00	Index
Basin LA County basin hydrograph	2.00	Index

rainfall stations and four gauging stations inside or nearby the Damansara River catchment. Table 2 and Fig. 2 illustrate availability and locations of rainfall stations, respectively.

Table 2 Streamflow and rainfall stations

No.	Rainfall station code	Streamflow and water level station code	Station name	Latitude	Longitude	Started	Ended
1	3115084	3115404	KG Melayu Subang	3.15006	101.541	2/1/2009	Present
2	3115083	3115403	Taman Mayang	3.11228	101.596	1/6/2003	Present
4	–	3015435	BATU 3 (TIGA)	3.08	101.55	23/1/2002	Present
5	3015083	3015490	TTDI jaya Shahalam	3.09694	101.554	18/2/2002	Present
7	3014080	–	Bandar Kelang	3.04806	101.445	23/1/2002	Present
8	3015001	–	Puchong Drop	3.01953	101.597	27/9/2003	Present
9	3015084	–	Taman Sri Muda	3.03772	101.535	23/1/2002	Present
10	3016080	–	Gandhi	3.07644	101.621	29/9/2005	Present
11	3114087	–	Bukit lanjan Subang	3.18308	101.612	18/2/2003	Present
12	3016077	–	Penchala (jalan 222)	3.09653	101.634	9/1/2004	Present
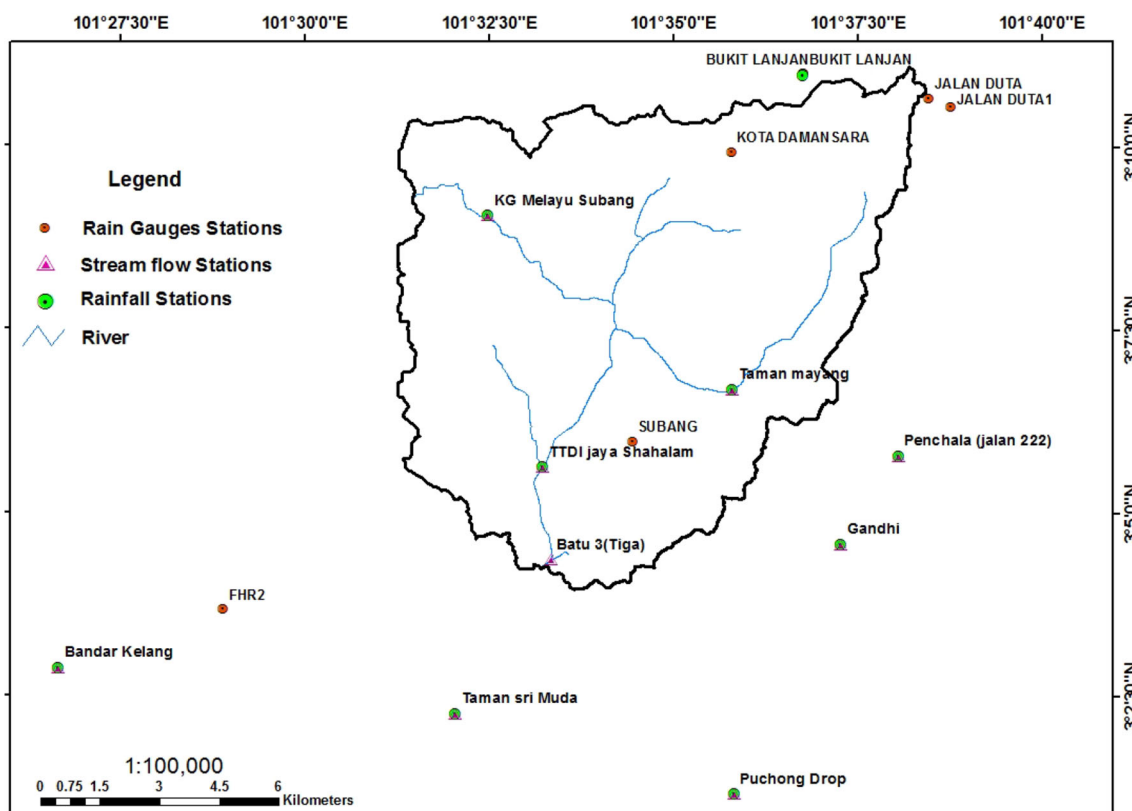
Fig. 2 Distribution of streamflow, rainfall, and rain gauges stations in and around the Damansara River catchment

Methodology

Overall, the applied method can be divided in three parts; PFF simulation, FF simulation, and combined FF and PFF which is shown in details in Fig. 3.

Preprocessing of geo-statistical GIS-based approach

Geo-statistical GIS-based probability model analyzes and transforms dependent input factors with independent param-

eters into a unique output layer using proper computed weighting, interpolating, data mining, and qualitative techniques [31]. Considering that PFF disaster follows a non-linear concept because of the complexity of morphology and climate dynamics, land cover, rainfall intensity and triggering factors, sufficient and precise condition factors are needed to run the probability model [3]. In this model, optimal regression was developed between dependent (conditioning factors) and independent parameters (pluvial inventory

Fig. 3 Overall flowchart of current study
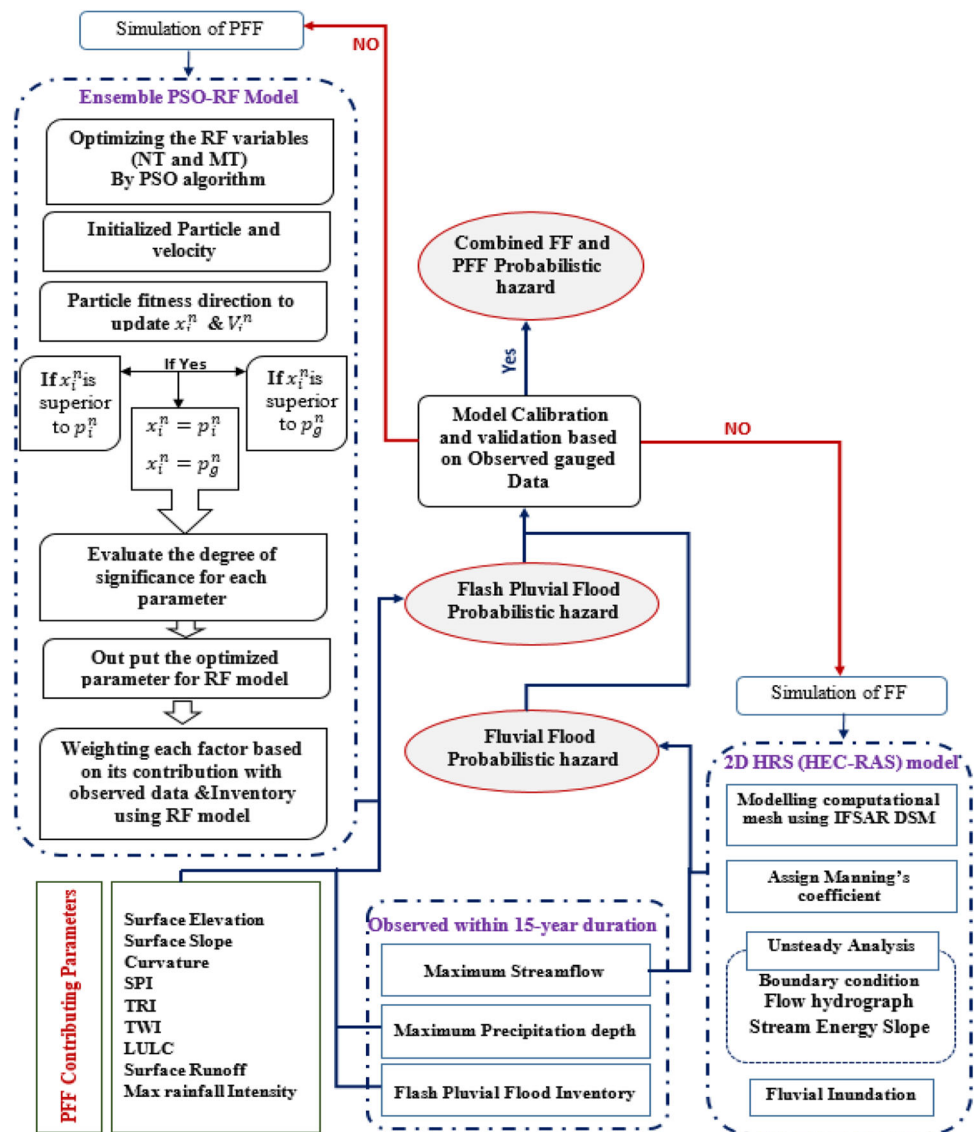
records). Then, each parameter obtains its own weight using RF algorithm to model PFF probability map.

Inventory of historical flood events

To evaluate flood probability analysis in the catchment, last flood events were examined and analyzed [32]. Therefore, inventory information reflects the most essential parts for predicting probable flood occurrence; such information can signify multiple historical events within a certain return period in a specific region [33]. For this research, flood inventory was created by mapping single PFF locations, wherein excess water was recorded by field observation and surveying. In other words, PFF locations and conditions were recorded within 4 h after high-intensity rainfall. In general, 68 different events, which were far from the river bank, were recorded since 2002 until the present (15-year interval) in the

study area. Additionally, locations of maximum precipitation for 1 h storms were extracted from rainfall stations.

PFF inventory map was then separated into 70% training and 30% validation [3], as shown in Fig. 4. Training-flooded locations (47 out of 68 points) were randomly selected. PFF probability model was run based on training events and validated based on testing events. Basically, the model was developed using two sets of value, namely 0 and 1. Zero specifies absence of PFF events, whereas 1 shows presence of PFF events. Similarly, an equal number of points (47 out of 68) were selected as non-flooded areas, wherein any PFF occurrence since 2002 was not recorded and assigned with a value of 0. The remaining observed PFF events (21 points) were utilized for model validation.
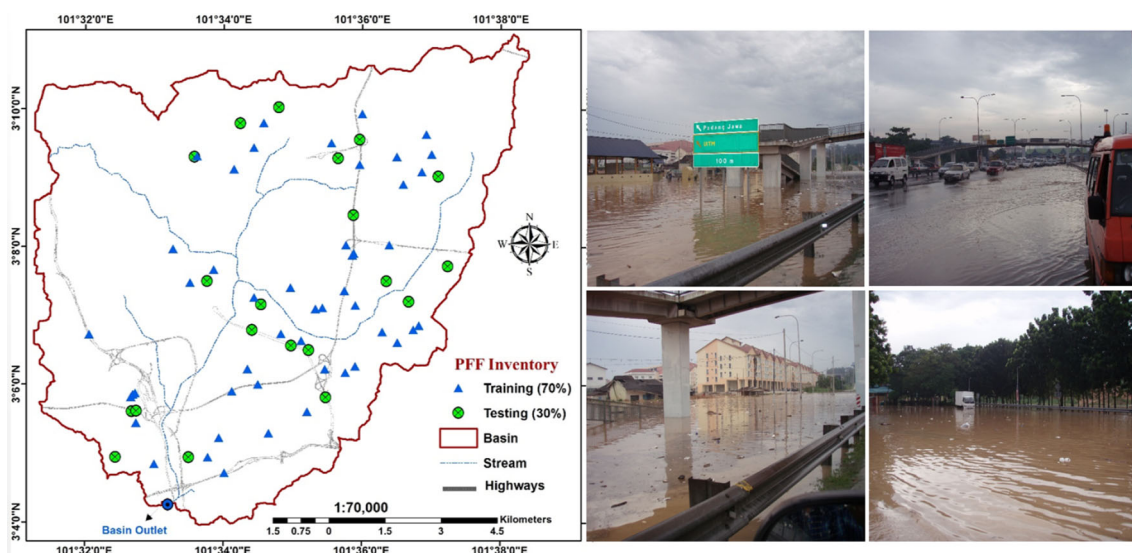
Fig. 4 Historical inventory events of PFF

Flood-conditioning parameters

Some contributing available parameters that were intended for inventory information and influenced PFF occurrence were named as "conditioning factors" [25]. The correlation between conditioning factors with flood occurrence was examined to perform probability analysis.

Building a PFF probability assessment model requires a set of training parameters [3]. Precision of the model can be influenced by accuracy of conditioning parameter. Thirteen conditioning factors were tested for FF and were considered the most significant [30]. The analysis indicated that soil, geology, aspect, and sediment transportation indexes are insignificant contributors to flood risk analysis.

Thus, in the current study, related PFF conditioning factor layers contributing to probable PFF comprised the following: curvature, stream power index (SPI), topographic roughness index (TRI), topographic wetness index (TWI), digital surface elevation, surface slope, surface runoff, maximum precipitation intensity, and LU/land cover (LULC).

Surface elevation

Elevation is one of the most significant parameters in flood analysis (Fig. 5a), and occurrence of PFF in highly elevated areas is nearly impossible [34]. Water flows from highly elevated areas toward lower regions. Consequently, probability of any type of flood event is naturally high in low altitude or flat terrains. Rather than digital terrain model (DTM), digital surface model (DSM) must be considered in calculating altitude parameter because of the urban pattern of this study area, wherein high-rise buildings and other facilities act as flood obstacles. However, other topographical factors related
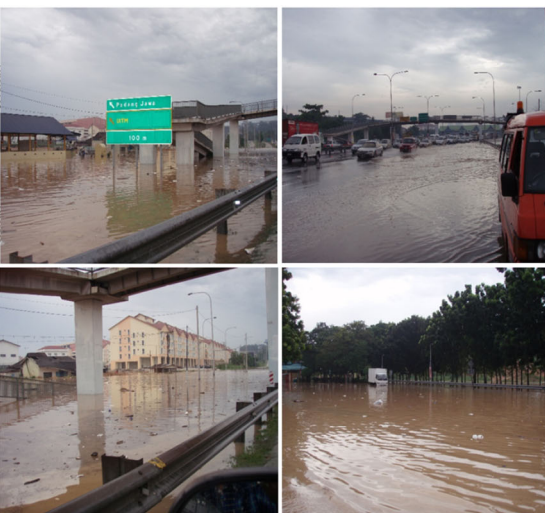
to flood occurrence were derived from DTM. Thus, a highly precise DTM presents a significant basic datum [35].

Surface slope

Slope is another topographical factor regarded as an important parameter in hydrology [30] because of its effects on runoff accumulation and velocity of excess rainfall. An increase in slope degree decreases time for surface infiltration. Subsequently, a large amount of water enters drainage networks and causes flood (Fig. 5b).

Curvature

Curvature also contributes significantly to PFF model physically; it ranges originally from negative to positive values in raster and must be classified into three classes. Positive values were converted into convex areas. Negative values were then grouped into concave areas, and pixels with zero value were assigned to flat regions. Basically, concave and flat regions are prone to flooding (Fig. 5c).
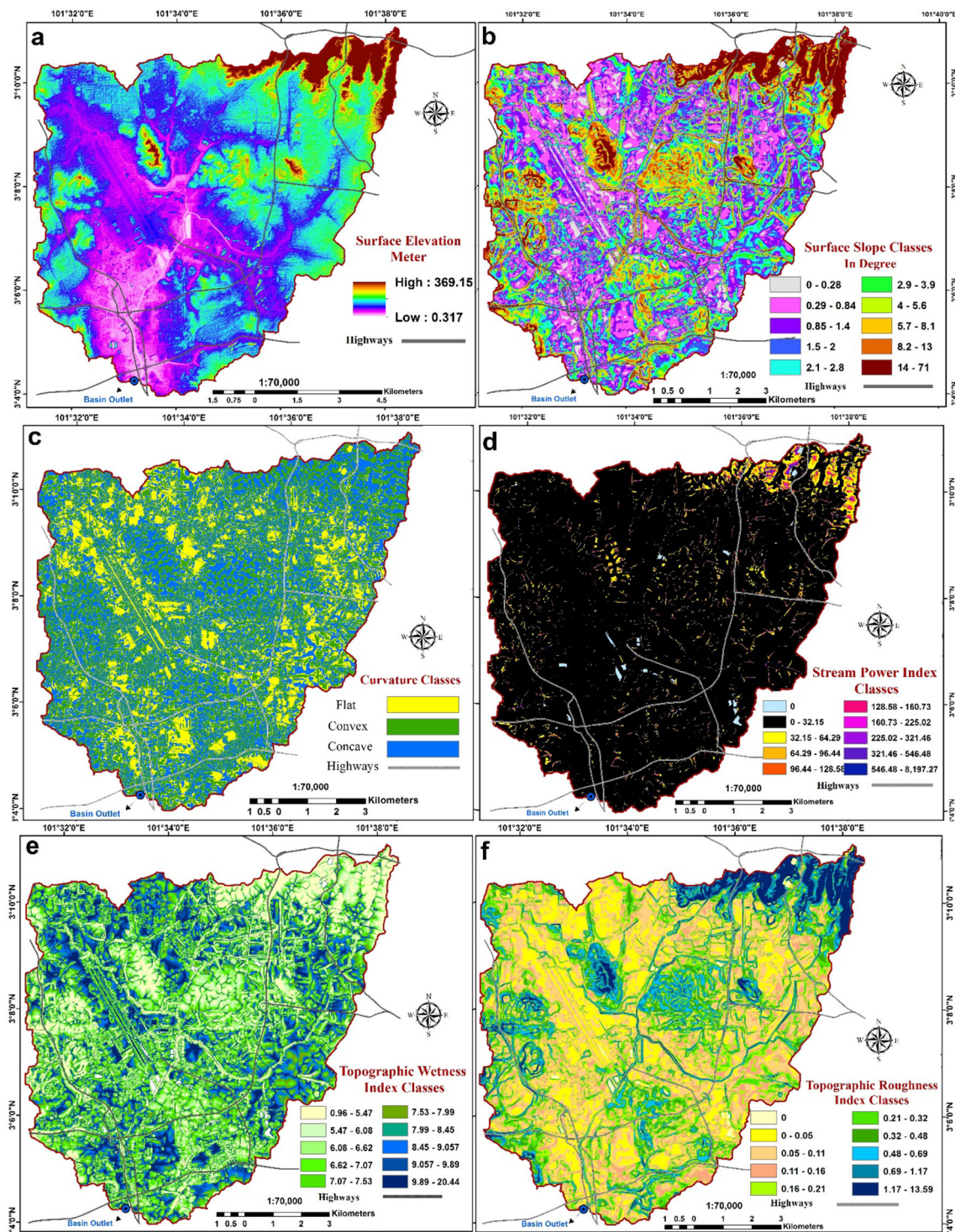
Hydrological indices

SPI and TWI are water-related parameters that are calculated using the following formulas [36]:

$$SPI = As \tan \beta, \tag{1}$$

$$TWI = \ln \left(\frac{As}{\tan \beta} \right), \tag{2}$$

Fig. 5 contributing factors to PFF: **a** surface elevation, **b** surface slope, **c** curvature, **d** SPI, **e** TWI, and **f** TRI. Continued PFF-contributing factors: **g** surface runoff, **h** LU, and **i** maximum rainfall intensity

where As represents catchment area or flow accumulation ($m^2 \, m^{-1}$), and β refers to local slope gradient measured in degrees.

SPI indicates erosive power of water flow (Fig. 5d). TWI represents effects of topography on runoff generation and amount of flow accumulation at any location in the river catchment [36], as shown in Fig. 5e. Accuracy of a topographic index can be estimated with regard to grid spacing and terrain roughness by comparing topographic index surface with respect to reference data.
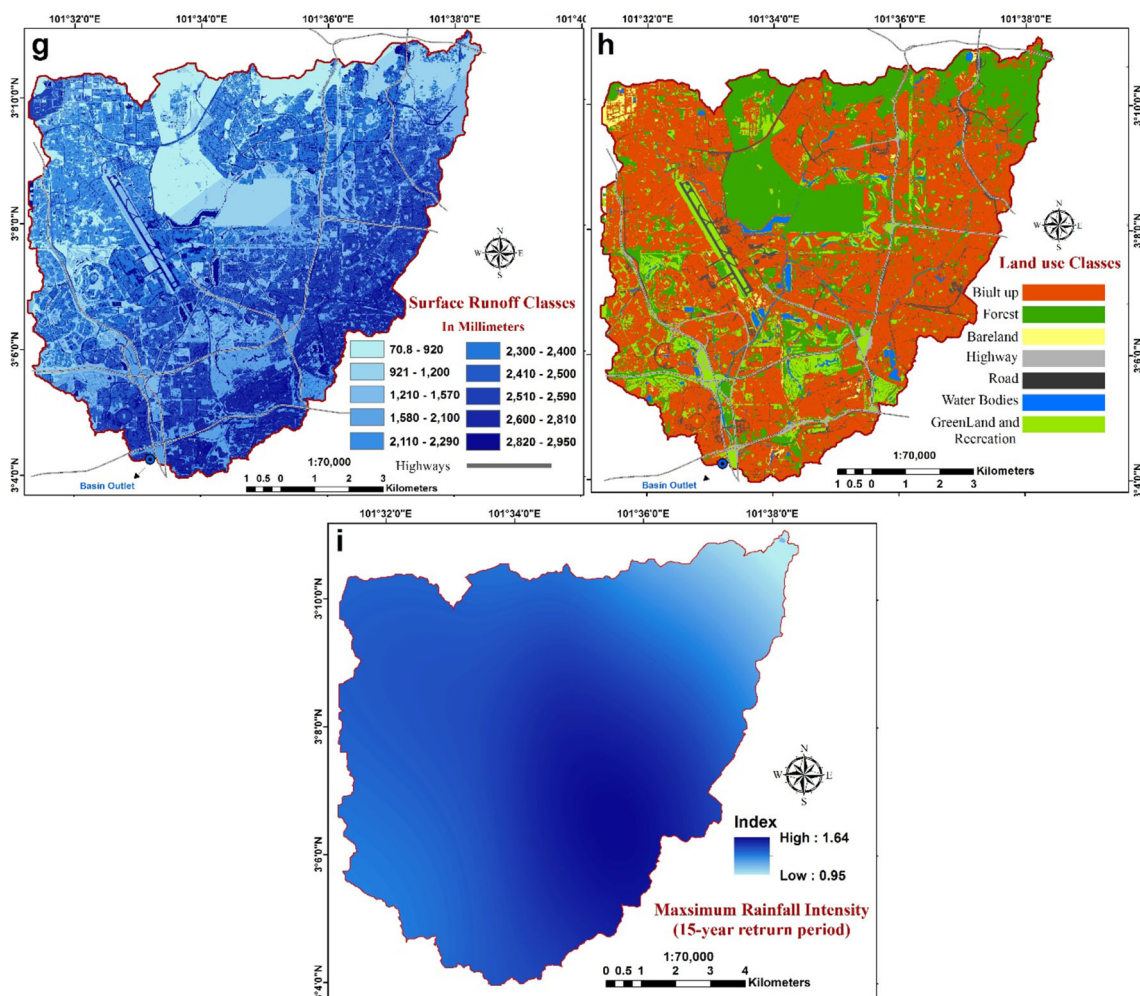
Fig. 5 continued

TRI is another morphological parameter widely used in flood analysis and calculated using the following equation:

$$\text{TRI} = \sqrt{\text{Abs}\left(\max^2 - \min^2\right)}, \tag{3}$$

where max and min represent the largest and smallest values of cells in nine rectangular neighborhoods of altitude, respectively (Fig. 5f).

Surface runoff

Soil capacity is fully saturated by water throughout land, and water flow exceeds limits required for surface runoff (Fig. 5g). This parameter was estimated using an empirical equation called Soil Conservation Service curve number method [37,38]:

$$S = \frac{1000}{\text{CN}} - 10. \tag{4}$$

Thus, *S* was calculated to generate curve number (CN) map. In generating CN map index, soil hydrologic groups from soil map and LU classes were combined in antecedent moisture condition scheme:

$$Q = \frac{(P - 0.2s)}{P + 0.8s}, \tag{5}$$

where *Q* refers to direct runoff (mm); *P* represents accumulated rainfall (mm); *S* refers to potential maximum soil retention (mm), and CN is the curve number.

Land use (LU)

LU types are also primary-related factors that strongly contribute to flooding. A detailed understanding of LUs bear extreme significance for environmental and natural hazards [39]. Vegetated areas are less prone to flooding because of the negative correlation between flood events and vegetation density. However, urban areas are typically composed of

impermeable surfaces and bare lands, which increase storm water runoff. Therefore, considering the importance of this factor, high-resolution image obtained from the WorldView-3 satellite was used to extract an LULC map [40]. The WorldView-3 satellite is a high-spectral- and high-resolution satellite imagery. This satellite features 31 cm panchromatic resolution, 1.24 m multispectral resolution, and 3.7 m short-wave infrared resolution captured in 2015 (Fig. 5h).

Rainfall intensity

Rainfall intensity is the most important factor affecting flash floods [41]. Intensity and frequency of rainfall display equal importance in evaluating high-magnitude floods in specific basins. Basically, under heavy rainfall during a limited time, surface soils are under full saturation condition. Thus, further rainfall fail to penetrate the ground and are converted as excess runoff. In urban regions, maximum intensity of precipitation causes failure of sewage network systems to drain out runoff at streets. Thus, this phenomenon results in PFF in impervious infrastructures [42].

In this study, maximum rainfall event for 1-h storm among 15-year data was selected in each station. Then, related intensities were extracted by dividing maximum rainfall depth at 60 min into nine metrological stations. Using inverse distance weightage (IDW) model, maximum intensity of each station was interpolated and extended spatially to the entire catchment area [30] (Fig. 5i).

PFF probability assessment using GIS and physical-based model

Machine-learning algorithms, such as RF, demonstrated excellent performance on many environmental applications and is used for modeling natural resource phenomena [43]. RF method was used to predict probability of PFF model in this study. This model evaluates the relationship between each conditioning factor with historical inventory events to forecast future PFF-prone areas [44]. RF is an ensemble learning method for optimization, classification, and regression run by building a large set of DTs at training time, resulting in formation of a class, which is the mode among entire classes (classification) or calculated weight for each dependent factor with respect to their contribution in independent events or achieved mean prediction of individual trees [45].

In the RF model, each tree is built using a deterministic method by selecting a random dataset of variables and a random sample among training data [46]. Basically, to gain ideal results using RF, three factors of this method were optimized, namely (a) "n tree" (NT), which represents the number of regression trees developed based on observation bootstrap sample, (b) "m try" (MT), which refers to the number of var-

ious predictors examined at each node, and (c) "node size" (NS), which corresponds to the minimum size of terminal nodes of trees. Degree of significance of predictor was measured by calculating percent increase in root mean square error (RMSE).

Particle swarm optimization algorithm

PSO method is a known computation method [47] and is extracted from well-known complex adaptive system. This method was initially motivated by consistency of bird's activity; then, Kennedy and Eberhart introduced basic implementation of PSO on swarm intelligence [47]. PSO describes the solution of each optimization issue as a bird, which searches one space "particle". In general, PSO is modified to a class of unsystematic particles used to explore ideal answers using iterative techniques [40]. Thus, for each iteration, particles communicate by tracing excesses of position and velocity. This mentioned behavior of ith particles is expressed mathematically (Eq. 6) [48] as follows:

$$\begin{cases} V_i^{n+1} = t.V_i^n + c_1 \times r_1 \times \left(p_i^n - x_i^n \right) \\ \qquad + c_2 \times r_2 \times \left(p_g^n - x_i^n \right), \\ x_i^{n+1} = x_i^n + V_i^n \end{cases} \quad (6)$$

where $i = 1, 2, \ldots, K$, K defines the entire number of particles, and n shows existing iteration number. t stands for inertia weight; p_i^n represents distinct ideal position for ith particle, and p_g^n describes the best location of total particles at nth iteration. c_1 and c_2 refer to learning elements; $r1$ and $r2$ show sample random numbers fluctuating from 0 to 1. V_i^n and x_i^n are defined as existing locations of ith particles and velocity, respectively. V_i^{n+1} 1 stands for updated velocity, whereas x_i^{n+1} represents place of ith particle at $n + 1$ iteration.

To validate and evaluate the accuracy of PSO model, RMSE was used for each practice [49]. Primary population with 20 practices is created. The population size is selected with respect to number of particles versus RMSE.

$$(\text{OPT}_i)2n\text{RMSE} = \text{SQRT}\left[\left(\frac{1}{n} \right) \times (\text{TAG}_i - \text{OPT}_i)^2 \right] \quad (7)$$

where, n stands as total number of samples in training dataset or the validation dataset. TAG_i is the goal values of the training dataset; OPT_i is output values from the PFF model.

The lower RMSE indicates good fitness. Position and velocity of entire particles can be updated using Eq. (6) and RMSE for their new position are calculated to select the finest one. In this stage, lowest RMSE determines the final position of swarm.

مدينة الملك عبدالعزيز
KACST للعلوم والتقنية

∅ Springer

Table 3 Step by step PSO-RF technique

Input	Possible range of random forest training and testing parameters
1	Producing primary particles included in NT and MT for RF model
	Defining PSO factors, such as total number of iterations, highest iteration number, first-particle swarm site, unsystematic flight velocity, and learning factors
	For each particle, completing training procedures for Stages 3 to 9
2	Selecting training and testing trials
3	Assigning number of iteration $n = 20$
4	Implementing 20 times cross-validation over training samples and computing average of cross-validation precision regarding *NT* and *MT*
5	Assessing fitness by average cross-validation precision (output of Stage 4)
6	Adjusting local and global ideal solution regarding fitness assessment result
7	Transferring each particle to a new location x_i^n with certain velocity V_i^n based on Eq. (6)
8	Up to this iteration, comparing local ideal result of ith particle p_i^n with new location x_i^n. However, good results may be the new p_i^n by $n + 1$ iteration. The same approach was applied on p_g^n, which is considered a global ideal result for all particles until n iteration
9	Repeating Step 3 when maximum iterations met optimal solution. Otherwise, the technique proceeded to Step 10
10	Stopping training when examined iteration reached optimal cross-validation precision to escape any over-training issues
11	Forming RF model over verification samples with respect to its optimal parameters (*NT and MT*) obtained from Stage 10
Output	Optimized RF parameters

Ensemble of PSO-RF model

To enhance performance of RF model, optimal values for NT, MT, and NS parameters were selected. Based on the last position of the swarm, the optimal consequent and antecedent factors were extracted to derive the optimal RF model. Ensemble PSO algorithm with RF model can efficiently solve the aforementioned issue. Table 3 presents the steps of this combination, which was implemented in MATLAB.

After assigning optimal parameters in RF using PSO algorithm, sequence of predictor significance (weightage) was derived. Then, each parameter was obtained its own degree and overlaid using spatial analyst tools. Degree of significance of predictor was measured by calculating percent increase in RMSE.

In Damansara catchment, a sewerage and sewer network system include open channels and underground pipes, which are designed for average excess rainfalls. However, for this study, capacity of sewage system was neglected in the proposed PFF simulations.

Hazard analysis of FF and PFF inundation

A hazard is a possibly damaging physical event or phenomenon that may cause environmental degradation, property damage, and loss of life within a specified period. PFF probability can be transformed into PFF hazard inundation depth by multiplying it with hazardous triggering factor, which may be an extreme rainfall event in this study (Eq. 8):

$$H_{pff} = f\left(P_{pff}, T_{mp}\right), \tag{8}$$

where H_{pff} indicates hazard probability; P_{pff} indicates PFF probability obtained from the coupled RF-PSO model, and T_{mp} refers to hazardous triggering layer. Rainfall was assumed as one of the primary triggering factors of flood occurrence over a study area, resulting in extreme events, such as flooding and overflowing; it also contributed to preparation of flood hazard maps.

T_{Mp} is defined as maximum precipitation depth from each meteorological station and was selected within a 15-year-return time period. These distributed data were interpolated by IDW method to extend over the entire urban catchment area. As a result of applying Eq. (17), inundation depth of PFF probability was quantified. Hazard was high when PFF inundation depth reached more than 20 cm [50].

FF probability and hazard assessment using 2D-HRS inundation analysis

Recent mainstream 2D models were developed with integration of computational hydraulics and numerical methods with rapid advances in information technology and graphical user interface design [51]. Basically, 2D models signify floodplain flow as a 2D aspect, which indicates inundation probability, by calculating water depth as a third dimension

where it reflects the hazard of inundated FF [39]. Majority of methods explain 2D shallow water calculations by momentum and mass conservation in a surface terrain obtainable by applying Navier–Stokes equations [18]

$$\text{Conservation of mass} \quad \frac{\partial h}{\partial t} + \frac{\partial (hu)}{\partial x} + \frac{\partial (hv)}{\partial y} = 0, \quad (9)$$

$$\text{Conservation of momentum} \quad \frac{\partial (hu)}{\partial t} + \frac{\partial}{\partial x}\left(hu^2 + \frac{1}{2}gh^2\right)$$

$$+\frac{\partial (huv)}{\partial y} = 0, \quad (10)$$

$$\frac{\partial (hv)}{\partial t} + \frac{\partial (huv)}{\partial x} + \frac{\partial}{\partial y}\left(hv^2 + \frac{1}{2}gh^2\right) = 0, \quad (11)$$

where x and y represent spatial dimensions of plane, and 2D vector $(v; u)$ describes average horizontal velocity across vertical column.

To solve these equations, u, v, and h over space and time were estimated. Several numerical structures were developed for this algebraic approximation. With regard to numerical discretization schemes, 2D methods are considered distributed models and may be classified into finite volume, finite difference, and finite elements [52]. In terms of spatial characteristics, these models can use either structured mesh, unstructured mesh, or flexible mesh [52].

Unlike traditional 1D models (e.g., Saint–Venant equations), computational cells need not feature flat bottoms, and cell edges need not indicate a straight line with a single height. Instead, each cell face and computational cell follow details of basic terrain morphology. This kind of inundation model is titled the HRS model [53].

HRS model uses detailed underlying structured mesh of sub-grid in conjunction with finite elements powered by high-resolution DEM to develop precise hydraulic and geometric property [30]. Two-dimensional HRS model can be established in HEC-RAS and can run preprocessing for flood areas, analyzing cell faces into hydraulic property tables.

Sub-grid resolution principals

Using wetting and drying algorithms, for any identified bathymetry $h(x, y)$, a detailed explanation of flow domain, which is capable for arbitrary subgrid resolution, can be described as ancillary *porosity* function $p(x, y, z)$ demarcated by the following:

$$p (x, y, z)$$
$$= \begin{cases} 1 \text{ if } h (x, y) + z > 0, \\ 0 \text{ otherwise} \end{cases} (x, y) \in \Omega, -\infty < z < \infty,$$
$$(12)$$

where horizontal integral estimated at $z = n_i^n$ inside individual polygon is specified as follows:

$$P_i \left(n_i^n\right) = \int_{\Omega_i} p(x, y, n_i^n)\mathrm{d}x\mathrm{d}y, \quad (13)$$

where free surface area is signified. Equation (11) indicates that $pi (n_i^n)$ is non-decreasing, non-negative, and restricted. Explicitly, $0 \leq pi (n_i^n) \leq Pi$. Remarkably, once $pi (n_i^n) = 0$, ith polygon becomes dry; at time $pi (n_i^n) = Pi$, it become wet, and at $0 < pi (n_i^n) < Pi$, ith polygon *partially* becomes wet. Additionally, at each single point inside the ith polygon, water depth is assumed by the following:

$$H \left(x, y, n_i^n\right) = \int_{-\infty}^{n_i^n} p (x, y, z) \, \mathrm{d}z$$
$$= MAX[0, h (x, y) + n_i^n]. \quad (14)$$

Consequently, $H(x, y, n_i^n) \geq 0$ and strict incongruence recognize a wet point. The wet region inside the ith polygon is computed using the following:

$$\Omega_i^n = \left\{ (x, y) \in \Omega_i : H \left(x, y, n_i^n\right) > 0 \right\}. \quad (15)$$

Volume of water inside the ith polygon is defined either as a surface vertical integral or a horizontal integral for overall water depth calculated by the following equation:

$$V_i \left(n_i^n\right) = \int_{-\infty}^{n_i^n} p_i (z) \, \mathrm{d}z = \int_{\Omega_i} H(x, y, n_i^n)\mathrm{d}x \, \mathrm{d}y. \quad (16)$$

Thus, considering that $pi (z)$ is non-decreasing and non-negative, one features $Vi (n_i^n) \geq 0$, and strict inequality essentially indicates $pi (n_i^n) > 0$. Non-negative cell-averaged water depth is well defined as follows:

$$H_i^n = V_i(n_i^n)/p_i. \quad (17)$$

Lastly, by indicating $x(s)$ and $y(s)$, parameters synchronize single points in the jth edge, linking two points recognized by $s = s_j^1$ and $s = s_j^2$ parameters for an identified constant value, where n_i^n represents the level along jth edge. The resultant wet cross-section zone is described as follows:

$$A_i \left(n_i^n\right) = \int_{s_j^1}^{s_j^2} H(x (s) , y (s) , n_i^n)\mathrm{d}s. \quad (18)$$

Therefore, non-negative edge-averaged water depth can be described as $Hn = A j (n_i^n)/\lambda j$.

Basically, the 2D flow area is considered the boundary for which 2D computations occur and is following mostly the ridge of basin catchment (Fig. 6).

A detailed LU dataset was used for surface roughness analysis. LU map was extracted from WorldView-3 satellite
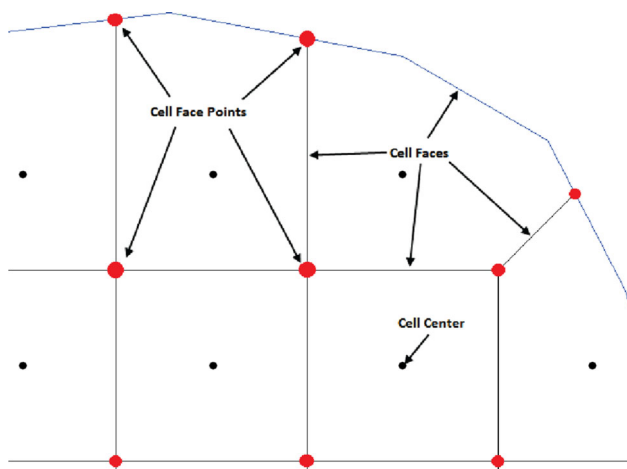
Fig. 6 2D-HRS modeling computational mesh terminology

Table 4 Optimal values of Manning's n

Surface type	Manning's n values	
	Range	Calibrated
Concrete/Asphalt	0.01–0.013	0.010
Bare Sand/land	0.01–0.06	0.02
Bare Clay-Loam (Eroded)	0.012–0.033	0.023
Graveled surface	0.012–0.03	0.018
Short Grass	0.10–0.20	0.16
Light Turf/Grass	0.15–0.25	0.19
Lawns/Green lands	0.02–0.30	0.21
Dense Turf	0.30–0.40	0.34
Dense shrubbery and forest	0.35–0.50	0.41

imagery using object-based support vector machine algorithm. Seven LU classes were detected, namely highways, bare lands, forest, built-up area, green lands and recreation area, roads, and water bodies. Then, averaged Manning's n values were assigned to each LU class [2], as shown in Table 4. After comparison between simulated fluvial inundation depth and observed water level, optimal Manning's n values for each class were figured out to calibrate FF simulation (Fig. 11).

Flow hydrograph analysis

Flow hydrograph was calculated to divert a streamflow to the 2D flow area. Some requirements were needed for this analysis: (a) flow hydrograph calculated by flow (Q)/time (t) and (b) energy slope of stream defined by degree of stream slope. For computing normal depth, energy slope from stream flow rate along the boundary condition line was calculated for each computational time period. Energy grade line slope was defined at the downstream boundary.

Four gauging stations located inside the Damansara River catchment recorded water level and streamflow since 2002, as shown in Fig. 1. In this research, hourly recorded streamflow (2002–2017) was used for unsteady analysis to model maximum FF inundation probability.

Boundary conditions in 2D-HRS model were also extracted from the KG Melayu Subang and Taman Mayang Kratie gauging stations to signify the upper boundary, and Batu Tiga defines downstream of the Damansara watershed basin. These boundary conditions were linked with probability scenarios to achieve reliable hazard analysis [54].

DSM is converted into triangulated irregular network format for the next step of unsteady analysis. Additional 135 cross-sections were engaged in Damansara River, and they can enhance performance and calibration of 2D-HRS model [18]. Then, river centerline, river bank, and flow path were also derived.

Combined FF and PFF probability analysis

Degree of dependency is an important issue for joint fluvial and pluvial probability analysis. Two basic statements should be considered in integration of probabilities: (a) dependence and (b) coincidence. Dependence assumes a functional correlation between two types of floods. In other words, FF and PFF may influence each other either in terms of magnitude or probability of occurrence. However, coincidence is not about any variable relationship but describes percentage chances that FF and PFF occur simultaneously.

Although cause and effect of these two types of flood differ, they may be similar in seasonal occurrence and feature initial triggering factors (extreme intensity of rainfall). Thus, FF and PFF occasions are not absolutely independent of each other. For the first step of joint hazard analysis, combined probability was quantified by single probabilities of incidences. For instance, for a certain pixel, when annual FF probability is 0.5 (50%), and annual probability of PFF occurrence reaches 0.5 (50%), then combined probability totals 0.25 (25%). Consequently, probability of occurrence of combined pluvial–fluvial flooding can be illustrated as follows:

$$P(ff, pff) = f\{(P(ff) \times P(pff) \times P(c)\} \qquad (19)$$

where $P(c)$ represents probability of coincidence for both flooding. $P(c)$ is valued by usual duration of monsoon flooding season and common length of FF events [55] suggested 0.2 as a value $P(co)$ coefficient in tropical regions with almost 80 days of flood season and 6 days of peak FF. In this study, the authors also applied the same value because of climate similarity.

Additionally, these two flood probability maps were standardized into a common dimensionless scale before they were combined given that the scales of their data differ from

each other [30]. The following equation was used for standardization:

$$X_{ij} = \frac{X_j - X_{ij}}{X_{\max - j} - X_{\min - j}},\qquad(20)$$

where, X_{ij} represents standardized score for ith alternative and jth attribute; X_{ij} refers to the raw score, and $[X_{\max - j} - X_{\min - j}]$ stand for maximum and minimum probability values for jth attributes, respectively.

Accuracy assessment, calibration, and validation of applied models

This procedure was performed to evaluate efficiency and precision of derived results. For fluvial hydraulic model (2D-HRS), calibration was conducted on the 7th July 2011, during one of the fluvial events. Simulated inundation depth for each hour was calibrated by hourly water level and observed at three gauging stations using linear regression method. After successful calibration, the model was validated in another period (28/12/2010) with dissimilar estimated magnitudes using RMS deviation (RMSD) approach [42]. RMSD is commonly used on a cell-by-cell basis for evaluating difference in water depths between observed and simulated data [56] and can be measured as follows:

$$\text{RMSD} = \sqrt{\frac{\sum_{i=0}^{n} \left(d_i^s - d_i^r\right)^2}{n}},\qquad(21)$$

where d_j^s and d_i^s represent simulated and referenced water depths, respectively, and n refers to total number of wet cells.

Distributed PFF GIS-based model was validated using receiver operating characteristic (ROC) curve. This method calculated the area under the curve (AUC), which is widely used in numerous studies to estimate performance of probability modeling [25]. In this validation method, 30% of observed historical inventory, which was not involved in training PFF model, was used to test model accuracy. The curve was produced by plotting accumulative percentage of simulated PFF prone regions (from maximum to minimum probability) and accumulative percentage of historical PFF events. ROC statistic varies between 1, for a perfect fit, and 0 once no overlapped inundation value exists.

Result and discussion

Simulated probability and hazard PFF results

Correlation of each conditioning factor with dependent pluvial inventories was developed by assembling PSO-RF model, as shown in Fig. 7.
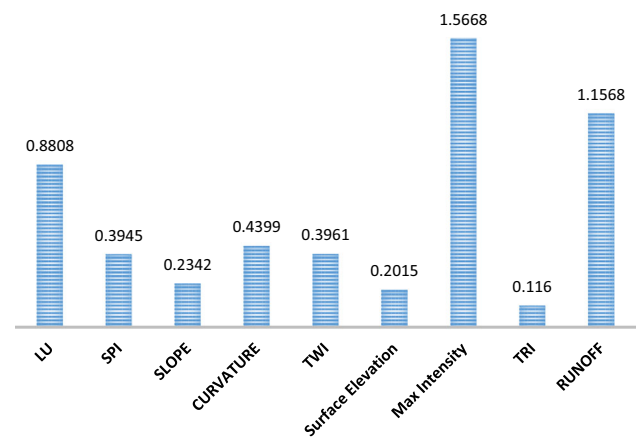

Fig. 7 Optimized weightage of parameters extracted by PSO-RF model

TRI, surface elevation, and slope achieved less significant weight among other conditioning parameters, reaching 0.116, 0.2015, and 0.2342, respectively. PFF probability was not influenced by these factors. Thus, these factors do not play significant roles in PFF prediction. SPI, curvature, and TWI gained moderate weights (approximately 0.4), showing fluctuations in these index values affecting probability of PFF occurrence. The most significant parameters highly contributing in PFF probability include maximum intensity, surface runoff, and LU factors, which gained 1.5668, 1.1568, and 0.8808 weightage, respectively.

As mentioned previously, PSO-RF was used to generate a PFF probability map in GIS environment. To optimize RF variables, NT values were examined from 500 to 9000, whereas MT was tested from 1 to 20 using PSO algorithm. Optimal values for NT and MT totaled 2500 and 17, respectively. Node size was defined as one. PSO was rapidly accomplished and significantly improved the model compared with stand-alone models, such as RF. PSO can produce generalization error, making assembled PSO-RF model as one of the successful MLs and statistical methods [57]. Achieved correlation coefficient and mean absolute error reached 0.86 and 0.071, respectively. Figures 8 and 9 show results of PFF probability and hazard, respectively.

A strong relationship exists between maximum of rainfall density classes with probability of PFF. Basically, when magnitude of precipitation was high within 60-min duration, probability of PFF occurrence was also high. Finally, sensitivity analysis showed that each type of LU features susceptibility to PFF. LU features affect water flow velocity and infiltration. For example, miner roads, water bodies, highways, and bare lands are highly prone to PFF unlike forest, green lands, and buildings.

Probability of PFF occurrence ranged from 0 to 0.99 for a 15-year return period using coupled PSO-RF distributed model (Fig. 8). Highest probability of PFF event was located
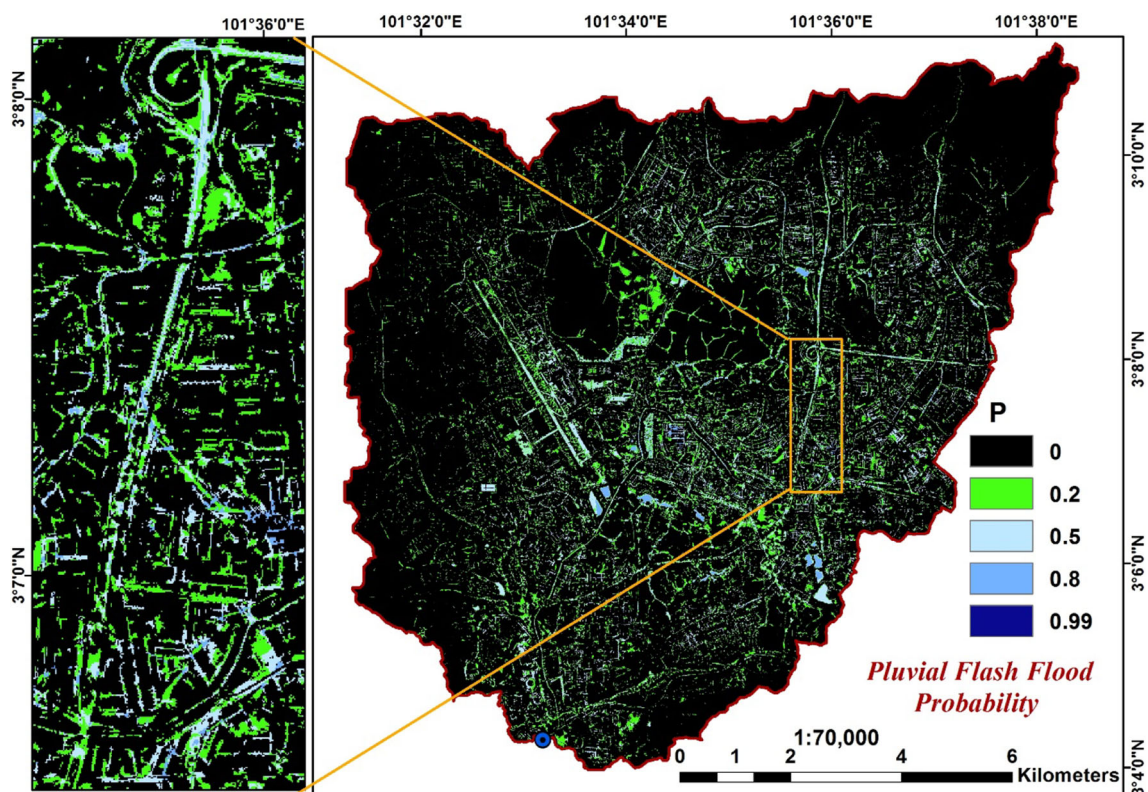
Fig. 8 PFF probability map using GIS-based PSO-RF model

at scattered pond areas, central and eastern miner roads, and highways. However, low degree of PFF events was observed in either forest, green lands or high-rise buildings particularly at west and north of the study area. During maximum level of precipitation intensity, excess water flow spread out over roads and highways, causing a wide range of road closures.

In this study, all aforementioned factors were classified into different classes according to natural break method and were examined by 130 individual points from simulated PFF probability map. The model was performed by sensitivity analysis with 12-times iteration with different training sample selections to test stability of accuracy. Sensitivity analysis determines how different values of an independent variable impact a particular dependent variable [58].

In all iterations, simulation accuracy was almost similar, showing significantly less amount of uncertainties. Sensitivity of PFF probability index was extracted for each condition factor. Lowest class of altitude obtains the highest value, which is illustrative of the highest correlation of this class with flood occurrence. This result may be due to natural behavior of flooding, which occurs mostly in flat regions instead in highly elevated place. No meaningful difference was observed between PFF probability value and slope diversity. Possibly, flash floods occurred at steep impervious areas unlike FF, which mostly occurs at low-inclined lands. For curvature factor, flat and concaved areas expectedly achieved the

highest correlation. High value of SPI factor gained less probable value, indicating that low-power streams are prone to flooding. However, high-value class of TWI factor obtained high-sensitivity value, implying possible wetting. In TRI factors, class ranges of 0–0.05 obtained the highest probable value, showing that in smooth regions, such as plain or airport band, occurrence of PFF is higher than rough area and supports accuracy of distributed model.

To quantify PFF-probable areas, inundation depth should be considered. Magnitude of flooding depth can illustrate the level of hazard for PFF events. Hazardous areas were affected by high magnitude of inundation depth. PFF inundation depths, which were divided into five classes according to natural break approach, spanned 0–2 and 2–10 cm classes over the main part of catchment (Fig. 9). Ponding areas are the most hazardous place when extreme rainfall events occurs in the study area. Bands of Subang airport and some parts of New Klang Valley Expressway are located in hazard zones because of their simulated inundation depth. Visually, more inundation depths (above 10 cm) are observed in highways and road networks, mostly due to their impervious surface, flat topography, and concentration of high-rainfall intensity. However, this type of flash flood steadily recedes through sewage and drainage networks.
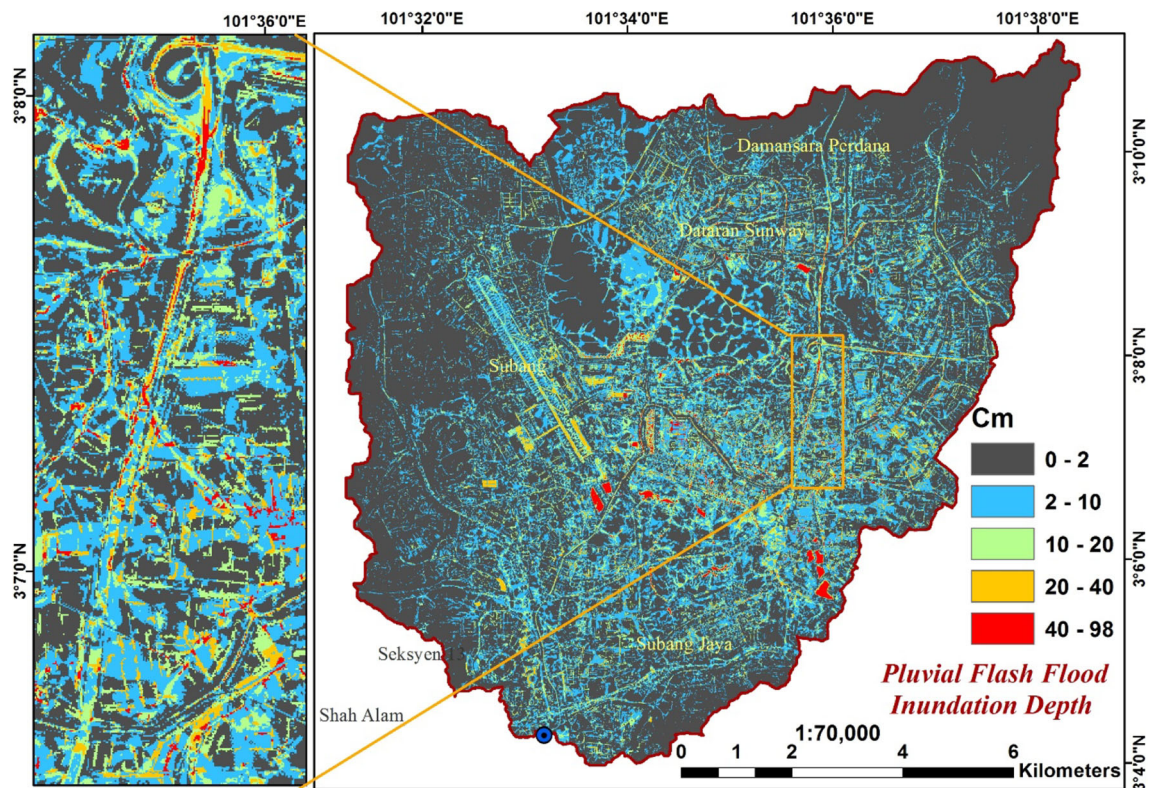
Fig. 9 PFF inundation depth hazard map using GIS-based PSO-RF model

Simulated probability and hazard FF results

Figure 10a demonstrates probabilistic FF map for a 15-year return period using hydraulic 2D-HRS model. Inundation map significantly shows flooding inundation pathways. Map values range from 0 to 0.99 and are classified into five intervals based on natural break method. At the center of the basin and along the main river, some lakes exist, and they are very prone to FF (from 0.8 to 0.9). Extended area from river bank is also located at 0.5 probable area, indicating 50% possibility of flooding within a 15-year return period. However, wide areas located in low elevated lands were predicted to be influenced by river inundation of 20% probability. Subang, Sunway, and Subang Jaya districts are threatened by FF hazard (Fig. 10b). Based on FF hazard assessment, scattered ponding central areas and water bodies located on stream direction represent the most hazardous areas with simulated 1–5 m inundation depth. Expectedly, flood probability is high when inundation model shows any significant depth.

Calibration and validation assessment

Comparison of results of simulated fluvial inundation depth with observed water level showed that 2D-HRS model logically simulated inundation amount. Sensitivity analysis was performed by 21 iterations to achieve optimal Manning's n

factor which results in synchronized simulated FF probability with observed events. Simply, a small number of under prediction was observed at the over-bank zones; this phenomenon may be due to infiltration, depression, and surface roughness dynamics. Then, hydraulic model was optimized in terms of assigned Manning's n and unsteady analysis, and assessment was repeated for another period of validation (Table 5 and Fig. 11).

The same approach was performed for PFF probabilistic inundation map, wherein simulated scenario was calibrated and validated based on observed maximum precipitation depth over three rainfall stations located in Damansara catchment (Table 5 and Fig. 12).

R-squared and RMSD statistical findings showed satisfactory and significant results for simulated FF and PFF, respectively.

Additional validation assessment was also implemented on simulated PFF probabilistic inundation to ensure high performance of new GIS-based PSO-RF model. To verify the model, 30% (21 points) of observed historical PFF inventory not involved in training model were overlaid with simulated PFF to measure accuracy of occurrence.

Performance of ROC was completely satisfactory in forecasting natural disaster occurrence short of any bias. When calculated AUC was close to 1.0, this value shows constancy and precision of applied model. A sharp curve signifies

Fig. 10 **a** Maximum FF
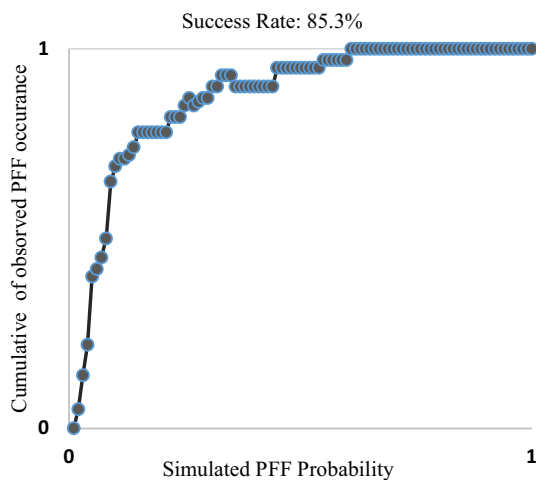probability map and **b**
maximum FF inundation depth
hazard map
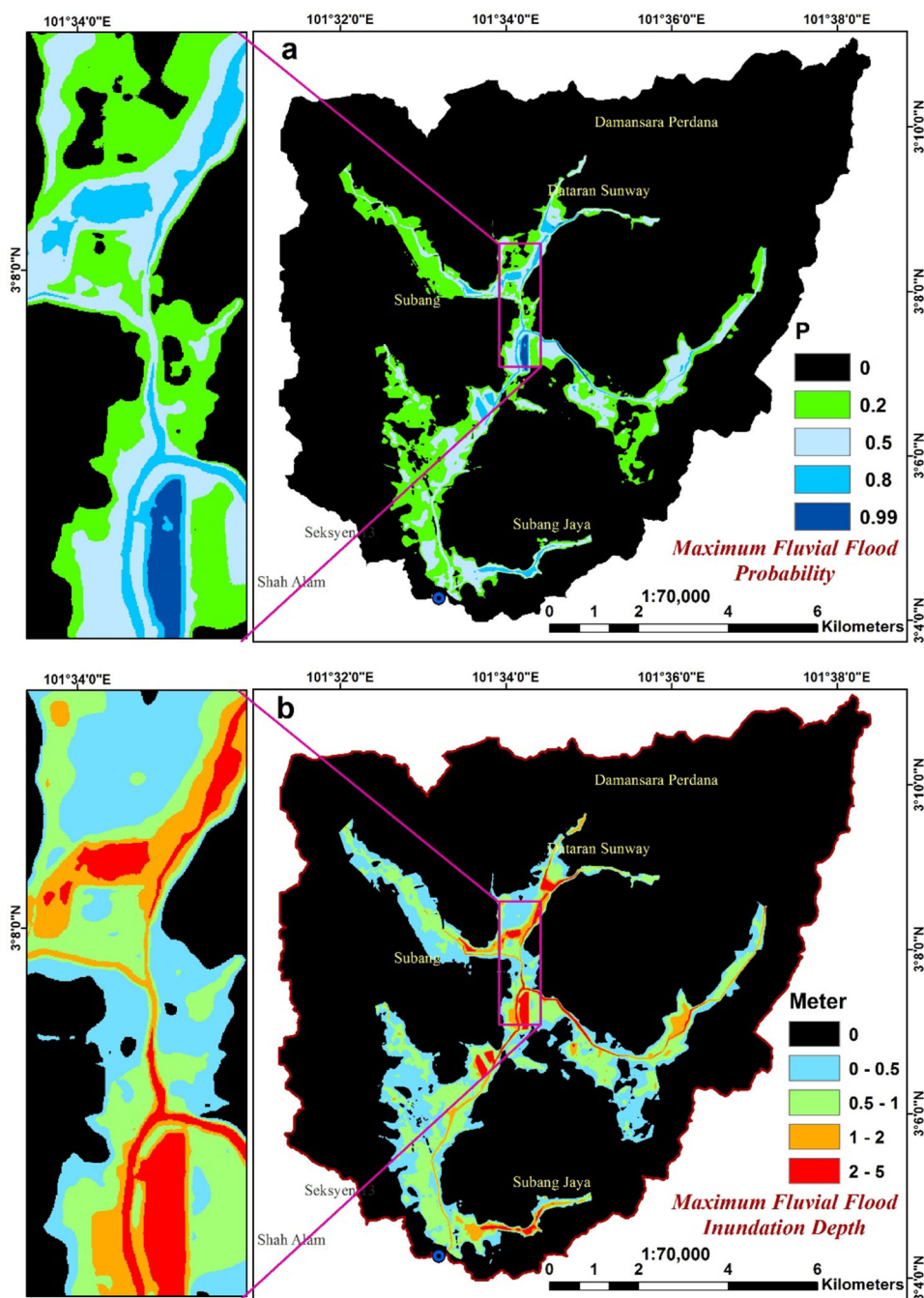
Table 5 Calibration and
validation results for FF and
PFF probabilities with observed
data at metrological stations

Flood type	Statistical methods		Water level stations			Entire basin
			Mayang	Subang	Batu-3	
Fluvial	Calibration	R^2	0.846	0.718	0.701	0.75
	Validation	RMSD	1.874	1.194	0.453	1.17
Flood type	Statistical methods		Rainfall stations			Entire Basin
			Mayang	Subang	Shahalam	
Pluvial	Calibration	R^2	0.908	0.825	0.842	0.86
	Validation	RMSD	2.235	8.105	5.469	5.79

Fig. 11 Compared simulated FF inundation depths with observed water level depths over three gauged stations on the 7th of July 2011
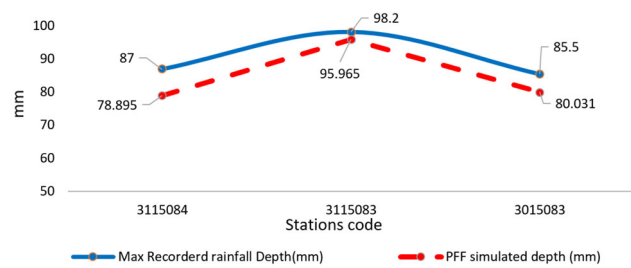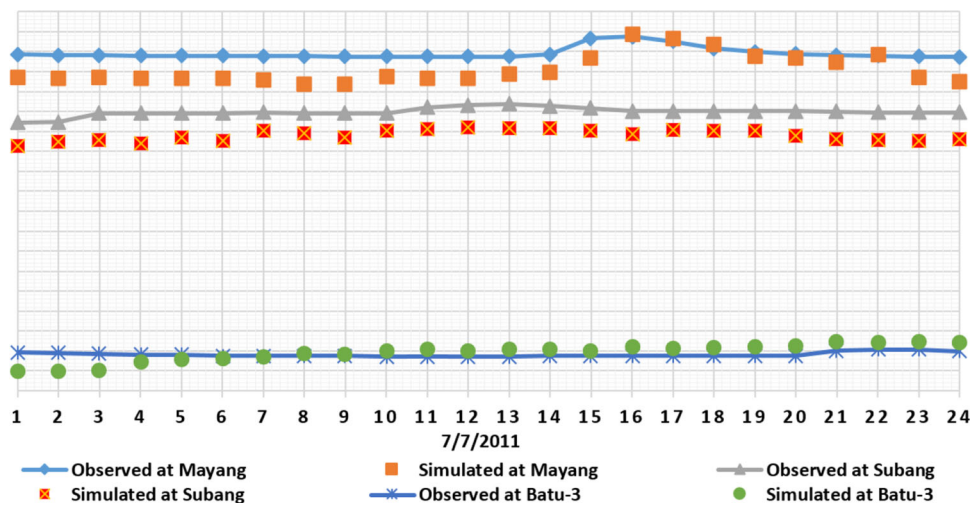
Fig. 12 Compared simulated PFF inundation depth with observed precipitation depth over three rainfall stations

Fig. 13 ROC accuracy assessment of GIS-based PFF probability map

high number of observed PFF points falling into most prone flood zones. ROC statistical model was performed to evaluate whether these points were located at high-probability zones. Success rate of the model was proven by over 85% accuracy, which is considered significant (Fig. 13).

Combined PFF and FF probabilistic

Joint FF and PFF probabilistic maps (Fig. 14) synthesize characteristics of individual probability of each flood type. In general, this method shows the same different inundation hot zones at deep depths and flow of streamline with fluvial behavior, as perfected by PFF with spatially distributed but shallow inundation. Technically, FF is a hazardous phenomenon caused by deep inundations, whereas PFF occurs frequently and plays a widespread spatial role with no chance of FF occurrence.

Probabilities of both types of flood are simultaneously much lower than that of individual event of each type (see Fig. 14). Results were classified into eight classes based on quantile method to visualize most possibilities (Fig. 15). More than 1339 ha of catchment is threatened by 1% probability of combined FF and PFF occurrence in a given return period. However, total areas ranging from 11 to 20% probable to both FF and PFF measure less than 3 ha.

Conclusion

Related PFF conditioning factors contributing to PFF probabilistic hazard assessment were extracted as follows: curvature, SPI, TRI, TWI, DSM, surface slope, surface runoff, maximum precipitation intensity, and LULC. Significant contribution parameters were successfully trained with PFF inventory by coupling GIS-based RF with PSO models. 2D-HRS hydraulic model was designed and calibrated to determine FF probability and hazards. This model uses a structured mesh (sub-grid) in conjunction with finite elements powered by high resolution (5×5 m) DEM extracted from InSAR. Hourly streamflow from three gauge stations were applied for unsteady analysis since 2002. Finally, we successfully combined PFF with FF probabilities to esti-
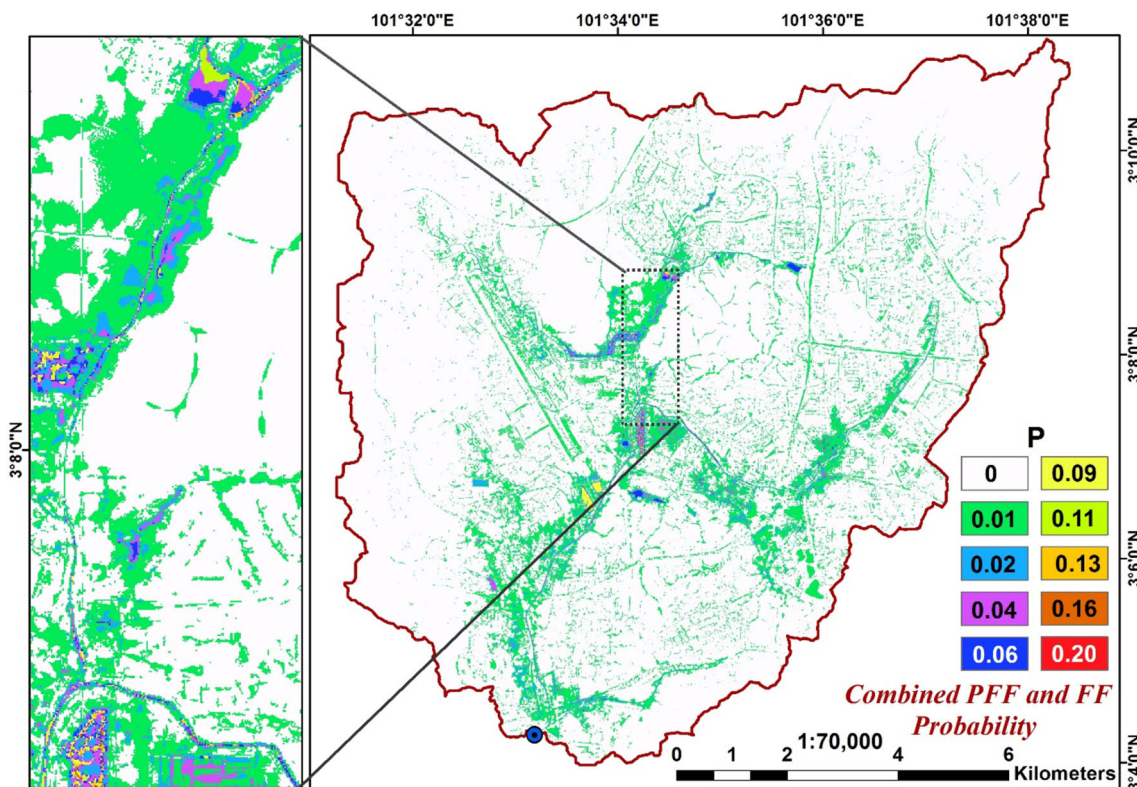
مدينة الملك عبدالعزيز
KACST للعلوم والتقنية

Springer

Fig. 14 Combined PFF and FF probabilistic model
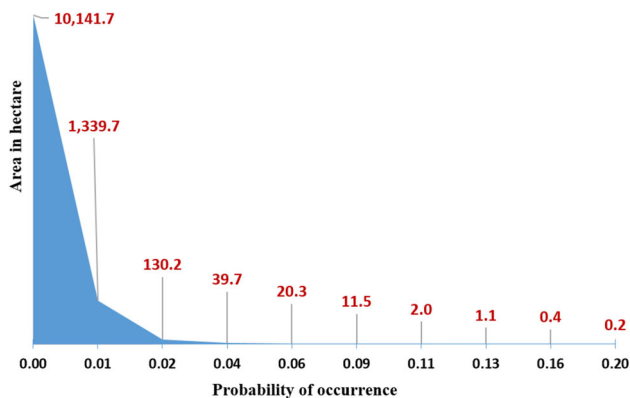
Fig. 15 Distribution of different combined FF and PFF probability classes

of rainfalls naturally occur at the same season and are not easily distinguished.

For further studies, we suggest improving the model in the following aspects: (a) using laser scanning data (e.g., LiDAR imagery), which features fine resolution, (b) considering sewer capacity (sewage network) in pluvial inundation modeling, and (c) using this study to attain risk analysis.

mate impacts and contributions of each type to urban flood hazard in Damansara catchment. R^2 and RMSD statistical assessments showed satisfactory and significant results for simulated FF (0.755 and 1.174) and PFF (0.858 and 5.790), respectively. Success rate of ROC method reached 85.3% accuracy. Sensitivity analysis between simulated PFF, and contributing factors obtained a reasonable cross-correlation. The presented approach can be utilized for flood mitigation planning for Damansara City, where FFs and high intensity

References

1. Smemoe CM, Nelson EJ, Zundel AK, Miller AW (2007) Demonstrating floodplain uncertainty using flood probability Maps1
2. Teng J, Jakeman AJ, Vaze J, Croke BFW, Dutta D, Kim S (2017) Flood inundation modelling: a review of methods, recent advances and uncertainty analysis. Environ Model Softw 90:201–216

3. Tehrany MS, Pradhan B, Jebur MN (2013) Spatial prediction of flood susceptible areas using rule based decision tree (DT) and a novel ensemble bivariate and multivariate statistical models in GIS. J Hydrol 504:69–79

4. Chang LC, Shen HY, Wang YF, Huang JY, Lin YT (2010) Clustering-based hybrid inundation model for forecasting flood inundation depths. J Hydrol 385(1):257–268

5. Falter D, Dung NV, Vorogushyn S, Schröter K, Hundecha Y, Kreibich H, Apel H, Theisselmann F, Merz B (2016) Continuous, large-scale simulation model for flood risk assessments: proof-of-concept. J Flood Risk Manag 9(1):3–21

6. Renaud FG, Syvitski JPM, Sebesvari Z, Werners SE, Kremer H, Kuenzer C, Ramesh R, Jeuken A, Friedrich J (2013) Tipping from the holocene to the anthropocene: how threatened are major world deltas? Curr Opin Environ Sustain 5:644–654. https://doi.org/10.1016/j.cosust.2013.11.007

7. Syvitski JPM, Higgins S (2012) Going under. The world's sinking deltas. New Sci 216:40–43. https://doi.org/10.1016/S0262-4079(12)63083-8

8. Naulin JP, Payrastre O, Gaume E (2013) Spatially distributed flood forecasting in flash flood prone areas: application to road network supervision in Southern France. J Hydrol 486:88–99

9. Hall JW, Sayers PB, Dawson RJ (2005) National-scale assessment of current and future flood risk in England and Wales. Nat Hazard 36:147–164

10. Apel H, Merz B, Thieken AH (2008) Quantification of uncertainties in flood risk assessments. Int J River Basin Manag 6:149–162

11. Merz B, Thieken AH (2005) Separating natural and epistemic uncertainty in flood frequency analysis. J Hydrol 309:114–132. https://doi.org/10.1016/j.jhydrol.2004.11.015

12. McMillan HK, Brasington J (2008) End-to-end flood risk assessment: a coupled model cascade with uncertainty estimation. Water Resour Res 44:W03419. https://doi.org/10.1029/2007wr005995

13. Arrighi C, Brugioni M, Castelli F, Franceschini S, Mazzanti B (2013) Urban micro-scale flood risk estimation with parsimonious hydraulic modelling and census data. Nat Hazard Earth Syst Sci 13:1375–1391. https://doi.org/10.5194/nhess-13-1375-2013

14. Falter D, Schröter K, Dung NV, Vorogushyn S, Kreibich H, Hundecha Y, Apel H, Merz B (2015) Spatially coherent flood risk assessment based on long-term continuous simulation with a coupled model chain. J Hydrol 524:182–193. https://doi.org/10.1016/j.jhydrol.2015.02.021

15. Neelz S, Pender G (2009) Desktop Review of 2D Hydraulic Modelling Packages. DEFRA/Environment Agency, UK. http://evidence.environment-agency.gov.uk/

16. Sarhadi A, Soltani S, Modarres R (2012) Probabilistic flood inundation mapping of ungauged rivers: linking GIS techniques and frequency analysis. J Hydrol 458:68–86 (363–370). Thomas Telford Ltd

17. Suarez P, Anderson W, Mahal V, Lakshmanan T (2005) Impacts of flooding and climate change on urban transportation: a system wide performance assessment of the Boston Metro Area. Transp Res D 10:231–244

18. Xu CY (2002) Hydrologic models. Department of Earth Sciences and Hydrology, Uppsala University, Uppsala

19. Yin J, Yu DP, Wilby R (2016) Modelling the impact of land subsidence on urban pluvial flooding: a case study downtown Shanghai. Chin Sci Total Environ 544:744–753

20. Burton A, Kilsby CG, Fowler HJ, Cowpertwait PSP, O'Connell PE (2008) RainSim: a spatial-temporal stochastic rainfall modelling system. Environ Model Softw 23:1356–1369. https://doi.org/10.1016/j.envsoft.2008.04.003

21. Hundecha Y, Pahlow M, Schumann A (2009) Modeling of daily precipitation at multiple locations using a mixture of distributions to characterize the extremes. Water Resour Res 45:W12412. https://doi.org/10.1029/2008wr007453

22. Nuswantoro R, Diermanse F, Molkenthin F (2014) Probabilistic flood hazard maps for Jakarta derived from a stochastic rainstorm generator. J Flood Risk Manage (in press). https://doi.org/10.1111/jfr3.12114

23. Schumann G, Matgen P, Hoffmann L, Hostache R, Pappenberger F, Pfister L (2007) Deriving distributed roughness values from satellite radar data for flood inundation modelling. J Hydrol 344(1):96–111

24. Liu YB, De Smedt F (2005) Flood modeling for complex terrain using GIS and remote sensed information. Water Resour Manag 19(5):605–624

25. Tehrany MS, Pradhan B, Jebur MN (2014) Flood susceptibility mapping using a novel ensemble weights-of-evidence and support vector machine models in GIS. J Hydrol 512:332–343

26. Kia MB, Pirasteh S, Pradhan B, Mahmud AR, Sulaiman WNA, Moradi A (2012) An artificial neural network model for flood simulation using GIS: Johor River Basin, Malaysia. Environ Earth Sci 67(1):251–264

27. Lamovec P, Veljanovski T, Mikoš M, Oštir K (2013) Detecting flooded areas with machine learning techniques: case study of the Selška Sora river flash flood in September 2007. J Appl Remote Sens 7(1):073564–073564

28. Botzen WJW, Aerts JCJH, van den Bergh JCJM (2012) Individual preferences for reducing flood risk to near zero through elevation. Mitig Adapt Strateg Glob Chang 2:229–244. https://doi.org/10.1007/s11027-012-9359-5

29. Chang CK, Ab-Ghani A, Abdullah R, Zakaria NA (2008) Sediment transport modeling for Kulim River—a case study. J Hydrol Environ Res 2(1):47–59

30. Mojaddadi H, Pradhan B, Nampak H, Ahmad N, Ghazali AHB (2017) Ensemble machine-learning-based geospatial approach for flood risk assessment using multi-sensor remote-sensing data and GIS. Geomatics, Natural Hazards and Risk, pp 1–23

31. Meyer V, Scheuer S, Haase D (2009) A multicriteria approach for flood risk mapping exemplified at the Mulde River, Germany. Nat Hazards 48(1):17–39

32. Manandhar B (2010) Flood plain analysis and risk assessment of Lothar Khola. Doctoral dissertation, Tribhuvan University

33. Bui DT, Bui QT, Nguyen QP, Pradhan B, Nampak H, Trinh PT (2017) A hybrid artificial intelligence approach using GIS-based neural-fuzzy inference system and particle swarm optimization for forest fire susceptibility modeling at a tropical area. Agric For Meteorol 233:32–44

34. Pradhan B (2009) Groundwater potential zonation for basaltic watersheds using satellite remote sensing data and GIS techniques. Open Geosci. https://doi.org/10.2478/v10085-009-0008-5

35. Gokceoglu C, Sonmez H, Nefeslioglu HA, Duman TY, Can T (2005) The 17 March 2005 Kuzulu landslide (Sivas, Turkey) and landslide-susceptibility map of its near vicinity. Eng Geol 81(1):65–83. https://doi.org/10.1016/j.enggeo.2005.07.011

36. Hossein MR, Maryam AS, Pradhan B, Ahmad N (2016) Soil erosion prediction based on land cover dynamics at the Semenyih watershed in Malaysia using LTM and USLE models. Geocarto Int. https://doi.org/10.1080/10106049.2015.1120354

37. Apel H, Trepat OM, Hung NN, Chinh DT, Merz B, Dung NV (2016) Combined fluvial and pluvial urban flood hazard analysis: concept development and application to Can Tho city, Mekong Delta, Vietnam. Nat Hazard Earth Syst Sci 16(4):941–961

38. Rizeei HM, Shafri HZ, Mohamoud MA, Pradhan B, Kalantar B (2018) Oil palm counting and age estimation from WorldView-3 imagery and LiDAR data using an integrated OBIA height model and regression analysis. J Sens

39. Roberts S, Nielsen O, Gray D, Sexton J, Davies G (2015) ANUGA user manual. Commonwealth of Australia (Geoscience Australia) and the Australian National University, p. 127

40. Cheng S, Lu H, Lei X, Shi Y (2018) A quarter century of particle swarm optimization. Complex Intell Syst. https://doi.org/10.1007/s40747-018-0071-2

41. Sahoo SN, Sreeja P (2016) Relationship between peak rainfall intensity (PRI) and maximum flood depth (MFD) in an urban catchment of Northeast India. Nat Hazards 83(3):1527–1544

42. Mishra SK, Singh V (2013) Soil conservation service curve number (SCS-CN) methodology, vol 42. Springer, New York

43. Clerici N, Weissteiner CJ, Gerard F (2012) Exploring the use of MODIS NDVI-based phenology indicators for classifying forest general habitat categories. Remote Sens 4:1781–1803 (Remote Sens 5, 2854 (2013))

44. Breiman L (2001) Random forests. Mach Learn 45(1):5–32

45. Liaw A, Wiener M (2002) Classification and regression by random forest. R news 2(3):18–22

46. Vincenzi S, Zucchetta M, Franzoi P, Pellizzato M, Pranovi F, De Leo GA, Torricelli P (2011) Application of a random Forest algorithm to predict spatial distribution of the potential yield of Ruditapes philippinarum in the Venice lagoon, Italy. Ecol Model 222(8):1471–1478

47. Kennedy J, Eberhart R (1995) Particle swarm optimization. In Proceedings of the IEEE International of First Conference on Neural Networks, Perth, pp. 1942–1948

48. Kennedy J (2011) Particle swarm optimization. Encyclopedia of machine learning. Springer, Berlin, pp 760–766

49. Rizeei HM, Pradhan B, Saharkhiz MA (2018) Surface runoff prediction regarding LULC and climate dynamics using coupled LTM, optimized ARIMA, and GIS-based SCS-CN models in tropical region. Arab J Geosci 11(3):53

50. Faghih M, Mirzaei M, Adamowski J, Lee J, El-Shafie A (2017) Uncertainty estimation in flood inundation mapping: an application of non-parametric bootstrapping. River Research and Applications

51. Lamb R, Crossley M, Waller S (2009) A fast two-dimensional floodplain inundation model. In Proceedings of the institution of civil engineers-water management, Vol 162(6)

52. Yin J, Yu D, Wilby R (2016) Modelling the impact of land subsidence on urban pluvial flooding: a case study of downtown Shanghai, China. Sci Total Environ 544:744–753

53. Casulli V (2009) A high-resolution wetting and drying algorithm for free-surface hydrodynamics. Int J Numer Meth Fluids 60(4):391–408

54. Yu D, Lane SN (2011) Interaction between subgrid-scale resolution, feature representation and grid-scale resolution in flood inundation modelling. Hydrol Process 25:36–53

55. Sarita Mahapatra, Alok Kumar Jagadev, Bighnaraj Nai (2011) Performance evaluation of PSO based classifier for classification of multidimensional data with variation of PSO parameters in knowledge discovery database. Int J Adv Sci Technol 34

56. Merz B, Hall J, Disse M, Schumann A (2010) Fluvial flood risk management in a changing world. Nat Hazard Earth Syst Sci 10(3):509

57. Dung NV, Merz B, Bárdossy A, Apel H (2015) Handling uncertainty in bivariate quantile estimation—an application to flood hazard analysis in the Mekong Delta. J Hydrol 527:704–717. https://doi.org/10.1016/j.jhydrol.2015.05.033

58. Peters TJ, Richards SH, Bankhead CR, Ades AE, Sterne JAC (2003) Comparison of methods for analyzing cluster randomized trials: an example involving a factorial design. Int J Epidemiol 32(5):840–846