Journal of Hydroinformatics

Journal of Hydroinformatics Vol 25 No 1, 113 doi: 10.2166/hydro.2022.106

Integrated machine learning and GIS-based bathtub models to assess the future flood risk in the Kapuas River Delta, Indonesia

Joko Sampurno [a,b,*], Randy Ardianto [c] and Emmanuel Hanert [a,d]

[a] Earth and Life Institute (ELI), Université Catholique de Louvain (UCLouvain), Louvain-la-Neuve, Belgium
[b] Department of Physics, Fakultas MIPA, Universitas Tanjungpura, Pontianak, Indonesia
[c] Pontianak Maritime Meteorological Station (PMMS), Pontianak, Indonesia
[d] Institute of Mechanics, Materials and Civil Engineering (iMMC), Université Catholique de Louvain (UCLouvain), Louvain-la-Neuve, Belgium
*Corresponding author. E-mail: joko.sampurno@uclouvain.be; jokosampurno@physics.untan.ac.id

JS, 0000-0001-9466-426X; RA, 0000-0001-9923-041X; EH, 0000-0002-8359-868X

ABSTRACT

As more and more people live near the sea, future flood risk must be properly assessed for sustainable urban planning and coastal protection. However, this is rarely the case in developing countries where there is a lack of both in-situ data collection and forecasting tools. Here, we consider the case of the Kapuas River Delta (KRD), a data-scarce delta on the west coast of Borneo Island, Indonesia. We assessed future flood risk under three climate change scenarios (RCP2.6, RCP4.5, and RCP8.5). We combined the multiple linear regression and the GIS-based bathtub inundation models to assess the future flood risk. The former model was implemented to model the river's water-level dynamics in the KRD, particularly in Pontianak, under the influence of rainfall changes, surface wind changes, and sea-level rise. The later model created flood maps with inundated areas under a 100-year flood scenario, representing Pontianak's current and future flood extent. We found that about 6.4%–11.9% more buildings and about 6.8%–12.7% more roads will be impacted by a 100-year flood in 2100. Our assessment guides the local water manager in preparing adequate flood mitigation strategies.

Key words: climate change, GIS, data-scarce delta, flood risk, machine learning

HIGHLIGHTS

- The proposed scheme successfully tackled the issues of data scarcity and low computational resources.
- The approach is appropriate for local water managers in developing countries.
- The proposed method combined the simple machine learning and GIS-based bathtub inundation models.
- The scheme is successfully implemented in the Kapuas River Delta.
- The assessment is beneficial for flood mitigation strategies.

1. INTRODUCTION

Climate change has been accelerating at an alarming rate in the last century and is likely to continue in the future (IPCC 2012). Certain regions will experience more intense and frequent rainfalls (Marengo *et al.* 2020), which will increase the flooding risk. In contrast, other regions will experience decreasing rainfall and increasing evaporation, which will accelerate soil's progressive drying, leading to drought (Mukherjee *et al.* 2018). In coastal areas, climate change impacts the frequency and intensity of coastal flooding. The coastal flooding hazard arises through changes in mean sea level and the storminess of the atmosphere that creates storm surges in the first place (Lilai *et al.* 2016; Vousdoukas *et al.* 2016). Extreme events, such as the 100-year flood under the current climate, with future sea-level projections, will occur much more frequently.

Urban areas in deltas are particularly vulnerable to climate change (Ridha *et al.* 2022). These urban areas will face multiple ocean and land threats simultaneously. With the areas' rapid population growth (Bhatta 2010), disaster could impact more people and cause more damage. This risk highlights the importance of preparedness in disaster mitigation (Dinh *et al.* 2012; Chan *et al.* 2021). Communities in disaster-prone cities should properly assess the hazards and adapt to the changing

climate, by for instance mapping their flood hazards and building flood defense systems. Adaptation strategies should be informed by proper risk projections so they can effectively mitigate the impacts (Hallegatte 2009).

The Kapuas River Delta (KRD) is a low-lying marshy delta on the west coast of Borneo Island, Indonesia. Silt deposits cover the delta from the coastline up to about 60 km inland (MacKinnon *et al.* 1996), creating estuarine floodplains for the Kapuas river downstream. Several urban areas lie in this delta and are prone to floods, which could be exacerbated by a combination of storm surges, intense rainfalls, and high discharges. These hazards are further intensified by sea-level rise caused by climate change (Moftakhari *et al.* 2017).

The water levels within the low-lying KRD are influenced by tide and wind surges from Karimata Strait, as well as the river discharges from the Kapuas and the Landak rivers (Sampurno *et al.* 2022). The flood risk here, and in other deltas, is likely to increase in future climate (Kundzewicz *et al.* 2014). The rising sea levels and excessive rainfall in the coming years will cause a severe impact because the KRD is low-lying and densely populated. Flood will be more intensive and frequent in a delta such as this (Lange 2020). A forecasting system is needed to assist local water management in assessing potential flood hazards, mitigating the risks, and planning adequate measures (Ngo *et al.* 2018).

Flood forecasting can leverage water-level modeling based on a machine learning (ML) technique (Ruslan *et al.* 2014; Gallien 2016; Habert *et al.* 2016; Noymanee & Theeramunkong 2019; Nguyen & Chen 2020). ML is based on evidence of relationships manifested in records of input and output data without analyzing the internal structure of the physical process. Using only historical data, an ML model can represent a complex input and output relationship, such as the relationship between water level and its predictor variables (Bishop 2006). The knowledge can subsequently be used to predict future water levels.

This study aims to assess the flood risk of an urban area within the KRD in future climates. We used an ML algorithm, i.e., multiple linear regression (MLR), to predict the water level in the city of Pontianak, the most populated urban area in the KRD. However, since data was scarce, we selected three months of data containing many extreme events (26 flood events) in 2020 as a reference case. The predictors we set comprised sea surface elevation in the river mouth, river discharge, wind velocity, and rainfall. The predicted water levels were used as a proxy to estimate the flood hazards in the city. Next, using a GIS-based bathtub model, we created flooding hazard maps under a 100-year flood scenario for the city with three climate change plots. Lastly, we conducted a risk analysis of these hazards to the nearby infrastructures (buildings and roads). The assessment output is essential for the local water resource managers in Pontianak to mitigate the impacts and create future adaptation strategies.

2. METHODS

2.1. Study area

In this study, we focus our assessment on Pontianak (Figure 1), the capital of the Kalimantan Barat Province, Indonesia, and the largest city in the KRD. Located in a low-lying area, the city has many canals that serve as drainage systems to the Kapuas Kecil River. Meteorologically, Pontianak experiences a bimodal rainfall distribution with two maximum peaks in March and October (Aldrian & Susanto 2003). Located along a tidal river stream, Pontianak is prone to compound inundations, which can occur due to storm surges from the sea, high river discharges from the land, excessive rainfall over the area, or a combination of these events. The flooding hazards may be worsened in the future due to sea-level rise or extreme weather events caused by climate change.

2.2. Data used

To forecast the flood, we used the observed water levels in Pontianak (Figure 2) as a proxy. The Pontianak Maritime Meteorological Station (PMMS) measures the river's water levels to support the Dwikora Port's operations and monitor the city's flood hazards. The data selected in this study were measured hourly for three months, from 1 October 2020 to 31 December 2020, at the PMMS's observation point (−0.020431° S, 109.33852° E). The data captured some flood events when high tides coincided with high river discharges or excessive rainfalls (Ganguli *et al.* 2020). Based on the field observation, PMMS considers that 2.5 m water level is the benchmark where the river begins to overflow the riverbanks.

Since the study area is located in a tidal river, where water levels are influenced by ocean tides, rivers discharges, and weather conditions, we used seven predictors: the sea surface elevation (SSE); the weather variables measured in Pontianak, i.e., precipitation, average wind speed, maximum instantaneous wind speed, and average wind direction; discharges from Kapuas River retrieved at about 50 km upstream from the observation point; and discharges from Landak River retrieved

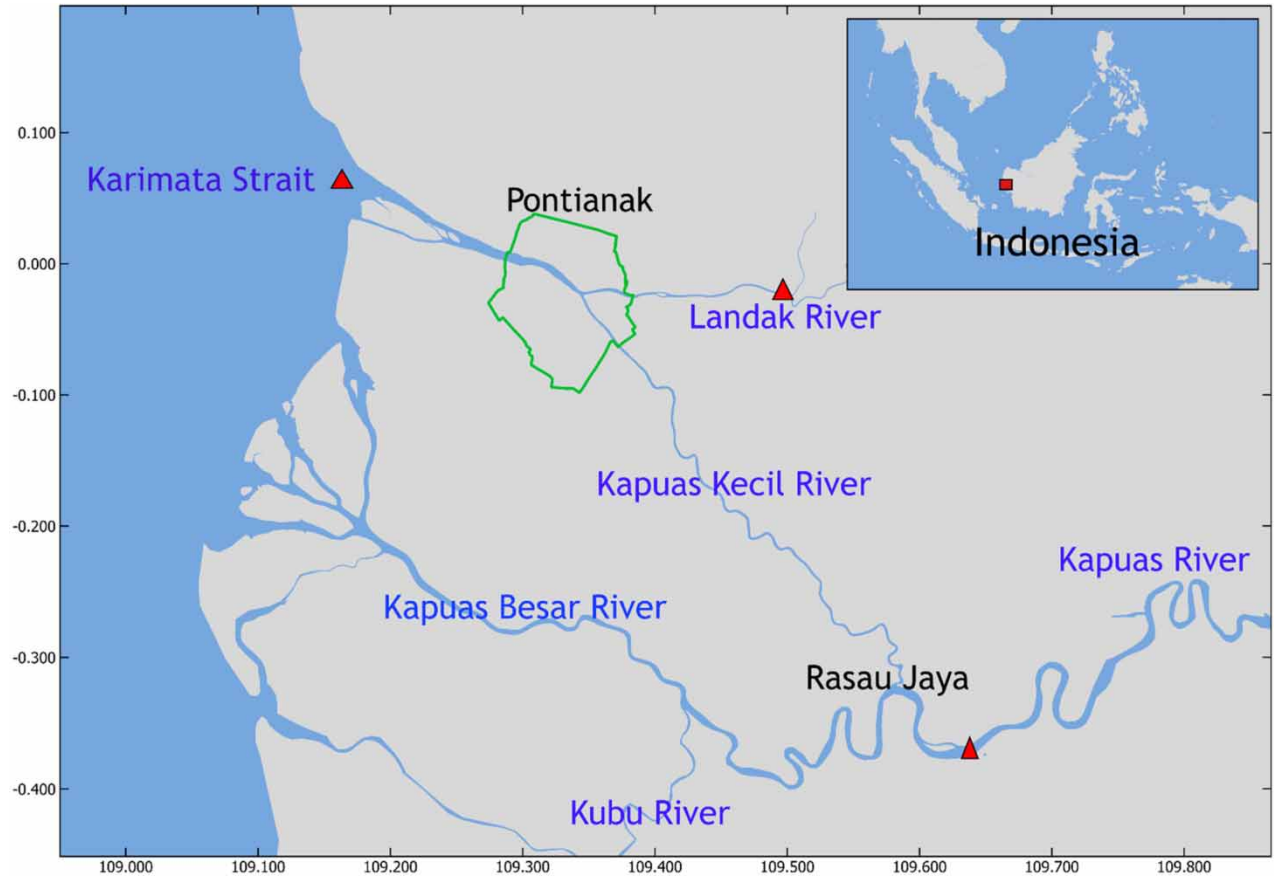

Figure 1 | The region of interest, with a green perimeter marking the city of Pontianak. The red triangles show where the model predictors (sea surface elevation and river discharges) are set, while the red box (in the inset) shows where the region of interest is set in this study. The flow from the Kapuas River's upstream bifurcates in Rasau Jaya, the furthest inland boundary of the KRD area. The main branch (the Kapuas Besar) bifurcates again (the Kubu branch), and the second-largest branch (the Kapuas Kecil) joins the end stream of the Landak River in Pontianak before it debouching into the sea.

at about 18 km upstream from the observation point (Table 1). Before training the model, we split the dataset between 80% for training and 20% for testing purposes.

2.3. Multiple linear regression and GIS-based bathtub model

Since we assume the relationship between the dependent and independent variables used as water level predictors is linear, here we use a well-known ML technique, multiple linear regression (MLR), to model water levels and predict future inundations. MLR aims to model a linear relationship between a set of independent variables and a dependent variable. This method is built upon a least-square algorithm to find the best-fitting model that connects a set of observed data and its predictors. MLR has been successfully applied in hydrology to predict a runoff signature (Zhang *et al.* 2018) and forecast streamflow (Block *et al.* 2009). In the current study, we used the MLR algorithm provided in the RWeka R library (Hornik *et al.* 2008). The MLR model is defined as:

$$H_i = \beta_0 + \beta_1 x_{1,i} + \beta_2 x_{2,i} + \beta_3 x_{3,i} + \beta_4 x_{4,i} + \beta_5 x_{5,i} + \beta_6 x_{6,i} + \beta_7 x_{7,i} + \epsilon \quad (1)$$

where i is the observation index, H_i is the water level at *i-th* time, $x_{j(j=1,\ldots,7)}$ are the predictor variables (see Table 1), β_0 is the water level intercept, $\beta_{j(j=1,\ldots,7)}$ are the slope coefficients, and ε is the residual.

Furthermore, we used a GIS-based bathtub model to create a flood extent map associated with the predicted water levels in the study area. Using the bathtub model, we assume that all areas within a model domain will be inundated if their elevation is

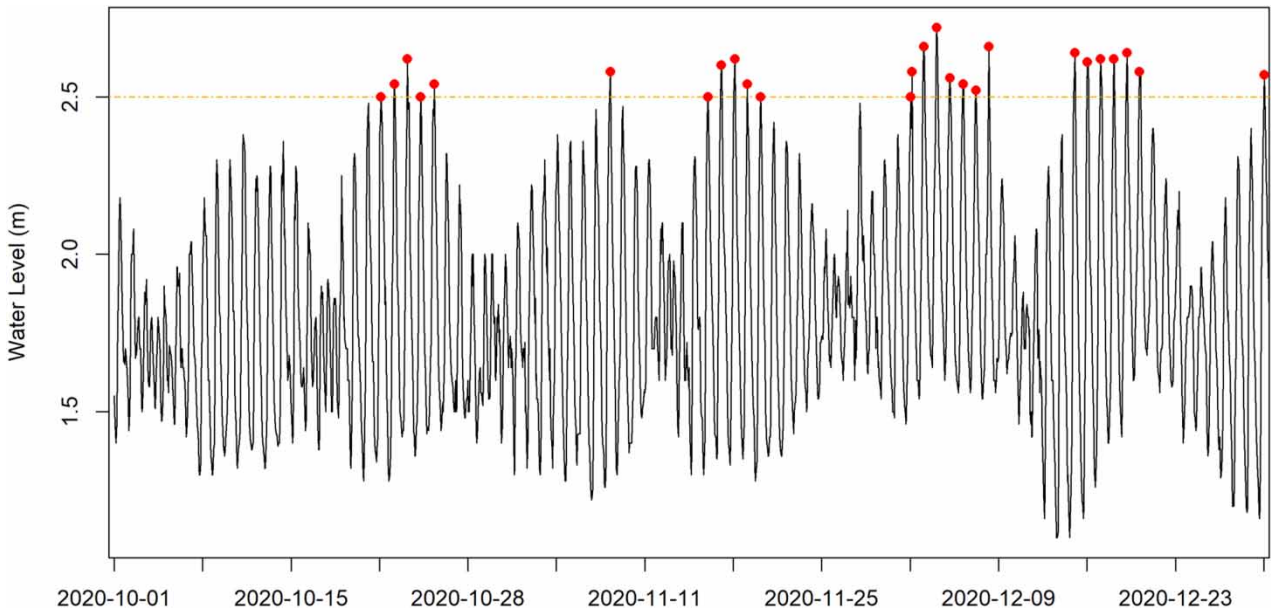

Figure 2 | Hourly riverine water levels in Pontianak, with the orange dash line representing the threshold of flooding, and the red dots representing the peaks of flood events.

Table 1 | Independent variables used as water level predictors

Code	Variable acronym	Description	Source
x_1	SSE	Sea surface elevation retrieved at the river mouth (m)	PUSRIKEL KKP (https://pusriskel.litbang.kkp.go.id/)
x_2	Qkapuas	Hourly discharge of the Kapuas River (m^3/s)	Global Flood Monitoring System (Wu *et al.* 2014)
x_3	Qlandak	Hourly discharge of the Landak River (m^3/s)	Global Flood Monitoring System
x_4	RR	Hourly precipitation in Pontianak (mm)	PMMS
x_5	WSavg	Hourly average wind speed in Pontianak (m/s)	PMMS
x_6	WSmax	Hourly maximum instantaneous wind speed in Pontianak (m/s)	PMMS
x_7	WD	Hourly average wind direction in Pontianak (degree, in the range: 0–360)	PMMS

less than the river's projected water level (Yunus *et al.* 2016). However, since we only focus on floods connected to the river, we consider the hydrology connectivity between the river streams and the land area (Van de Sande *et al.* 2012). Any area within the city will be inundated if a channel connects it to the Kapuas River stream, i.e., city canals. Therefore, using an algorithm created by Wang & Liu (2007), we identify and fill surface depressions in our DEM map that are not connected to the river.

2.4. Predictor selection

We applied the Mutual Information (MI) analysis to evaluate the dependency level of the dependent variables on each predictor because it can assess the correlation between the dependent and the independent variables (Ross 2014). The MI's value will be greater than zero if there is a correlation between the evaluated variables. The stronger the correlation, the greater the

MI's coefficient value is (Kinney & Atwal 2014). The value is determined by the following equation (Choi *et al.* 2020):

$$MI_{(X,Y)} = \sum_{x \in X} \sum_{y \in Y} p(x, y) \log\left(\frac{p(x, y)}{p(x) - p(y)}\right) \tag{2}$$

where X represents the independent variable, Y represents the dependent variable, $p(x)$ is the probability of the independent variable (X), $p(y)$ is the probability of the dependent variable (Y), and $p(x, y)$ is the joint probability distribution of both variables.

Since we employed the MLR algorithm, we also checked the linear regression assumptions between the dependent variables and the predictors to diagnose the model adequacy. Then, we selected the variables that significantly impacted the model and omitted the non-essential ones. To do so, we produced a scatter plot between variables, created histograms of each variable, and calculated the Pearson's correlation coefficients for all pairs of predictor and dependent variables. We also checked the correlation among predictors to detect and avoid multicollinearity. We used the psych library in R to check the linear assumption (Psych: Procedures for Psychological Psychometric and Personality Research 2021). To enhance the prediction skill of the model, we also run a sensitivity analysis before predicting the future flood scenarios impacted by climate change. By doing so, we could evaluate the effect of each parameter on the model's output (Chu 1999). Therefore, we could tune in the model parameters involved in the model-building to improve the prediction results by adjusting only the sensitive parameters (Cacuci 2003). To conduct the sensitivity analysis, we used the konfound R library (Xu *et al.* 2019).

2.5. Metrics for evaluation

Two goodness-of-fit coefficients determine the models' performance: the Root Mean Square Errors (RMSE) and the Nash–Sutcliffe Efficiency (NSE). RMSE is commonly used for regression tasks to measure the accuracy of a predicted variable against an observed variable over an entire dataset (Jackson *et al.* 2019). However, the coefficient was only computed for the maximum values in the dataset to measure how well the model captured the inundation hazards. Meanwhile, NSE was used to assess the 'skill' of the ML models compared to the skill of the observed data's mean to predict an unknown dependent variable (Choi *et al.* 2020).

2.6. Future scenarios

To evaluate the impacts of climate change, we used the projection of sea-level rises, precipitation changes, and surface wind changes in three climate scenarios for Southeast Asia's regional level (Iturbide *et al.* 2021). The scenarios are the low (RCP2.6), the medium (RCP4.5), and the high emission scenarios (RCP8.5). Based on these projections, we created projection datasets (SSE, Precipitation, Wind average, and Wind maximum) for each year into the future (from 2021 to 2100). We then re-ran the ML model to predict future water levels and extract each year's annual maximum water level. Then, we computed the flood frequency for every ten years of data using the Gumbel (1958) distribution curve. Here, we took only the flood frequencies between 2020 and 2030 (as the current state) and between 2090 and 2100 (as the future state). Based on the flood frequency curves, we estimated the 100-year flood level as the annual maximum flood hazard levels in 2020 (current) and 2100 (future hazard state).

2.7. Flood risk analysis

We used a simple bathtub inundation model to assess the inundation risk throughout the city under the 100-year flood level condition (Murdukhayeva *et al.* 2013). We firstly retrieved the DEM of the city of Pontianak from DEMNAS (https://tanahair.indonesia.go.id/demnas/) with 0.27-arcsecond (8.3 m) resolution and corrected it using elevation benchmark points from SRGI (https://srgi.big.go.id/). We then applied another correction to the channels in the DEM by incorporating the channel connectivity map (Figure 3).

Next, based on the elevations in the corrected DEM, we defined wet areas as any points on the map with an elevation lower than the 100-year flood level. Wet areas mean either inundated areas (with an elevation greater than zero) or standing water areas (with an elevation less than zero). We repeated this procedure for all future scenarios.

Next, we conducted a flood risk assessment for further analysis. Here, we identified the infrastructures (buildings and roads) over the study area, retrieved from OpenStreetMap (OpenStreetMap 2020), that may be affected by inundation hazards under the 100-year flood condition in 2100. We analyzed the flood impact on infrastructures using a QGIS plugin: InaSAFE (InaSAFE 2022).

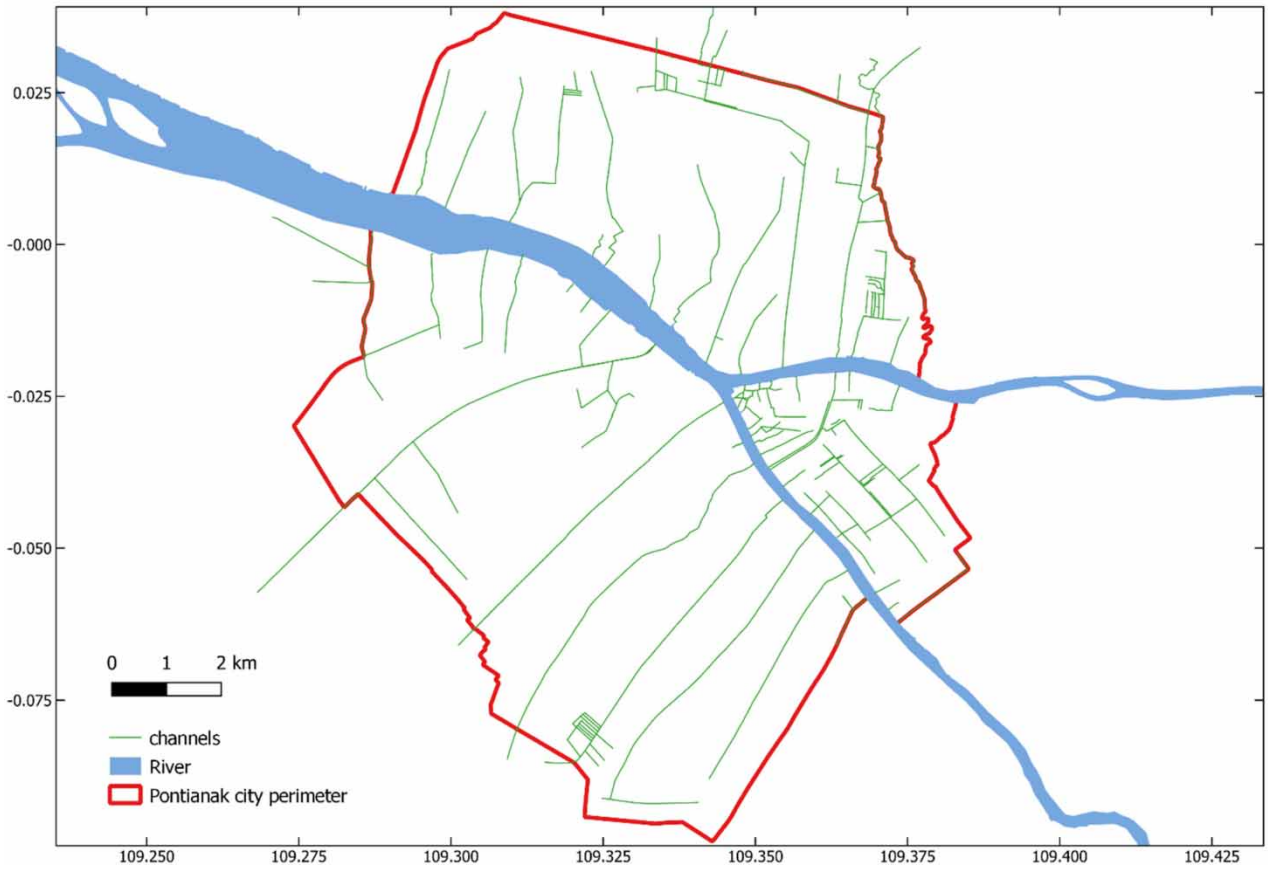

Figure 3 | Channel connectivity within the city of Pontianak (OpenStreetMap, https://planet.osm.org, 2020).

3. RESULTS

3.1. Features selection and model performance

Figure 4 shows the MI coefficient of each independent variable relative to the observed water levels. Overall, all independent variables show a positive MI coefficient, which means they influence the water levels. In particular, two independent variables show a strong relationship, i.e., tidal elevation and precipitation. The rest of the variables show only a moderate relationship.

However, based on the correlation among predictors, two variables (WSmax and WSavg) have significant multicollinearity (Figure 5). The correlation between these two predictors is 0.91. Therefore, we had to omit one of them; otherwise, it would decrease the model's performance. We implemented a sensitivity analysis to determine which variable could be omitted (Table 2). The sensitivity analysis shows that the maximum wind speed (WSmax) contribution to the model was insignificant, so we canceled it and retained the other variables.

We then trained the MLR algorithm using the training dataset to build the model. Next, we assessed its performance on the testing dataset. The model performed well and was stable in the training and testing phases, indicated by the NSE = 0.86 and NSE = 0.88 for the training and testing phases, respectively (Figure 6). In addition, both the training and testing phases show a consistent accuracy (RMSE = 0.12 m). These performance indicators suggest that the model is reliable in predicting water levels (prediction and observation pair points tend to cluster along the 45° line). However, extreme water levels are slightly under-estimated as the regression line has a slope $<45°$.

3.2. Future flood risk analysis

After successfully building the model, we simulated future water levels in the study area under the three climate change scenarios. We then calculated the flood frequency curve of these predicted water levels at the PMMS's observation point within

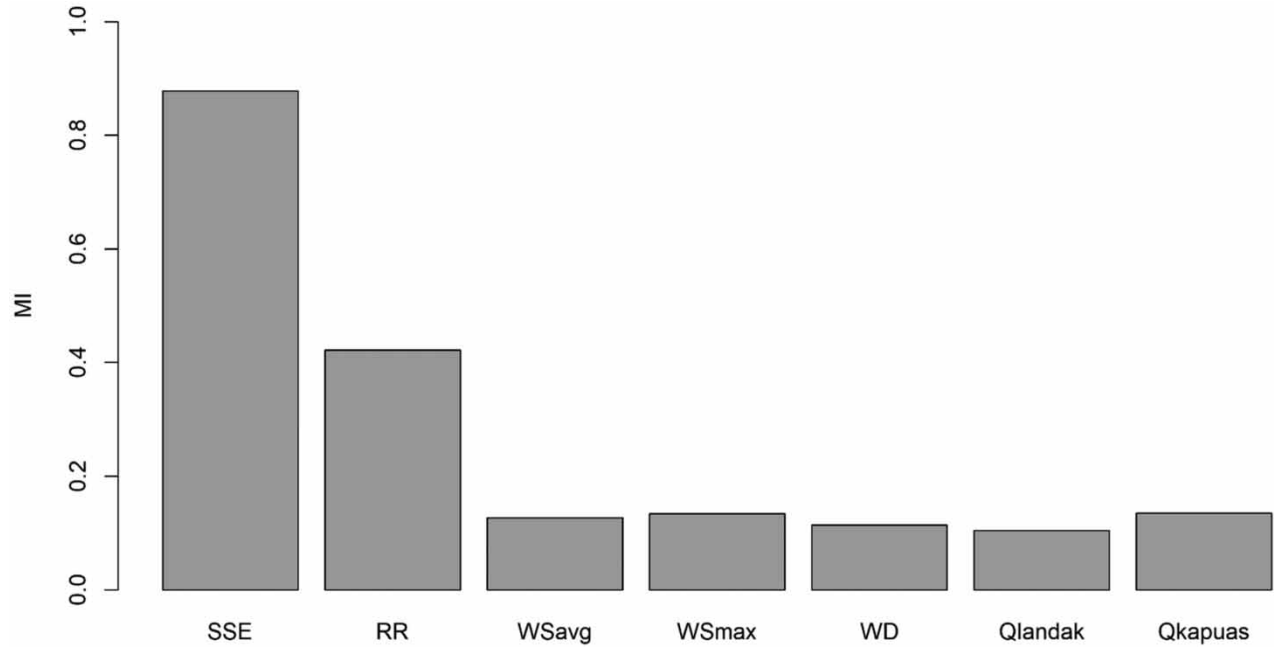

Figure 4 | Mutual information of proposed predictors (Table 2) to the observed water level.

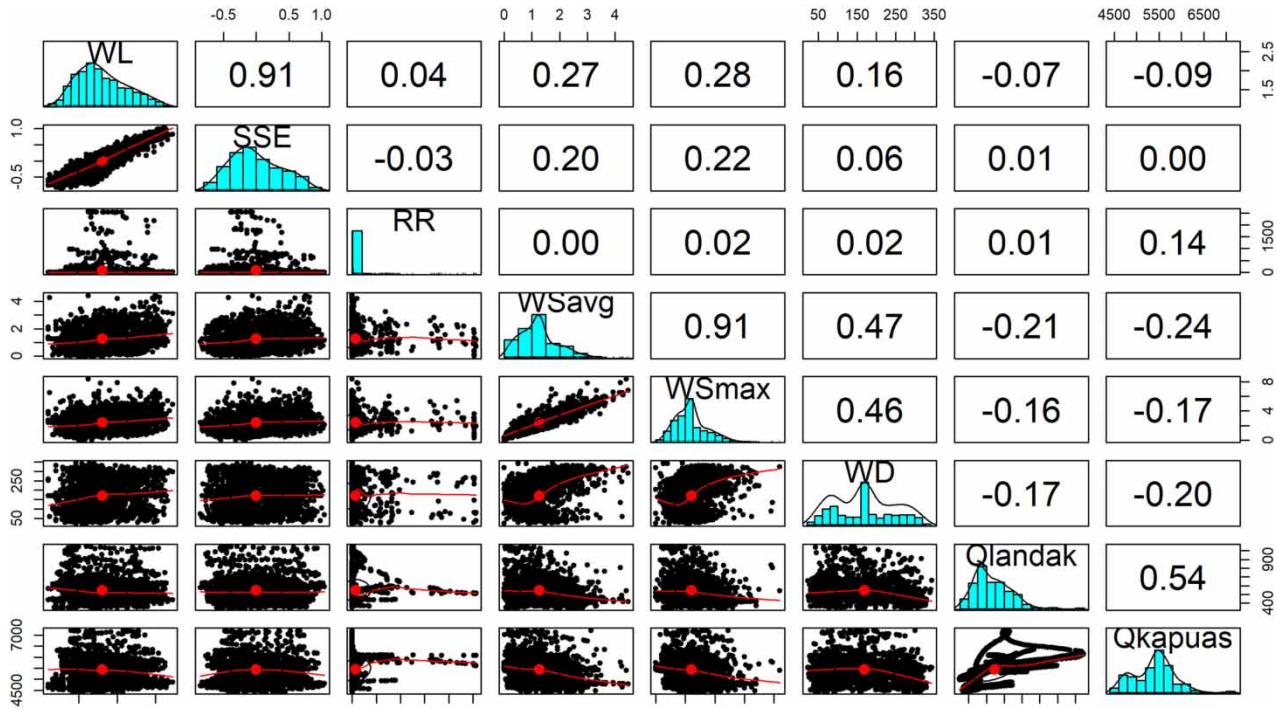

Figure 5 | Linearity assumption evaluation consisting of scatter plots, histograms, and the Pearson correlation coefficient between all proposed variables.

the city (−0.020431° S, 109.33852° E). Figure 7 shows how climate change impacts the 100-year flood level in the study area. Based on the current flood frequency curve, the 100-year flood level is 2.64 m. Under a low emission scenario (RCP2.6), the 100-year flood level will increase by about 0.28 m to reach a 2.92 m water level in 2100. Next, under a medium emission

Table 2 | Sensitivity analysis outputs for the predictors

Coefficients	Estimate slope	Std. Error	t-value	p_test	Significance
SSE	7.50E-01	7.73E-03	97.084	2.0E-16	Significant
RR	8.24E-05	9.42E-06	8.752	2.0E-16	Significant
WSavg	2.66E-02	9.49E-03	2.806	5.0E-03	Significant
WSmax	−9.71E-03	6.08E-03	−1.597	1.1E-01	Not significant
WD	3.30E-04	4.30E-05	7.670	2.8E-14	Significant
Qlandak	−6.78E-05	2.76E-05	−2.455	1.4E-02	Significant
Qkapuas	−4.09E-05	6.82E-06	−6.000	2.4E-09	Significant

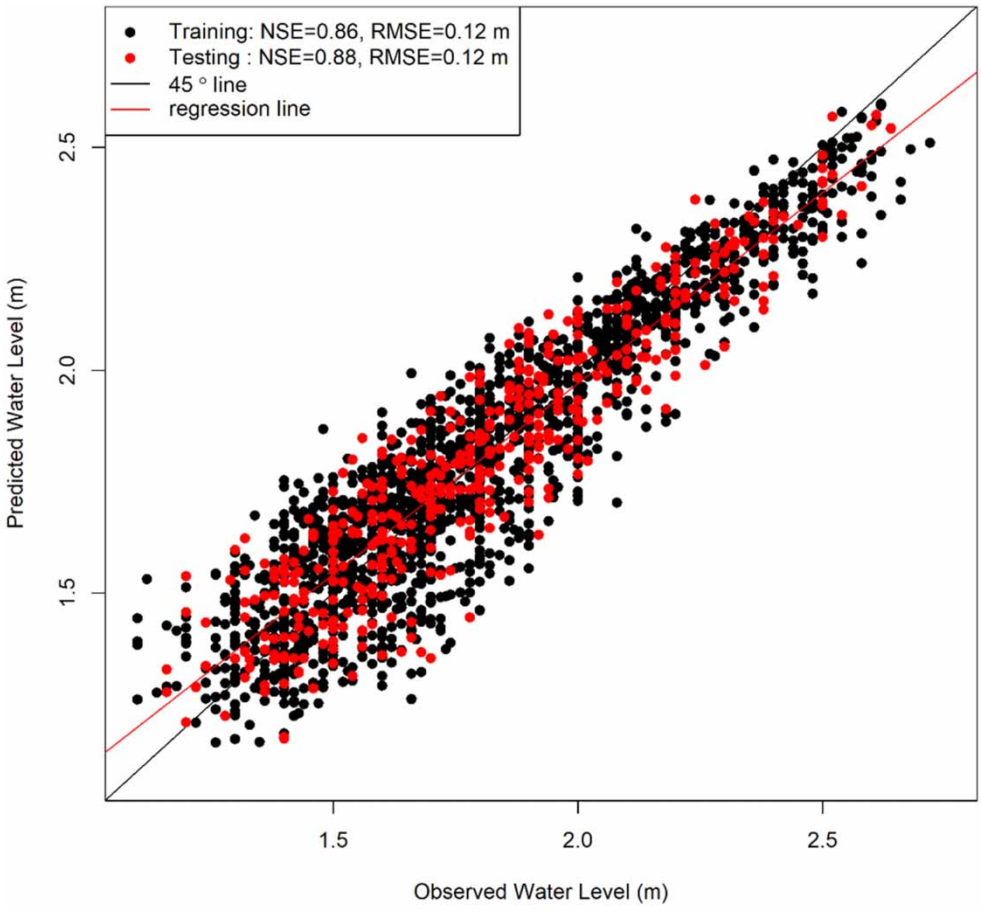

Figure 6 | Model performance during training and testing phases, with a comparison of the regression line orientation and the 45°-line representing the relative quality of the model's prediction.

scenario (RCP4.5), the 100-year flood level in 2100 will increase by about 0.39 m and reach 3.03 m. Lastly, under a high emission scenario (RCP8.5), in 2100, it will increase by 0.7 m–3.34 m.

Using the GIS-based bathtub inundation model, we estimated that the extent of the hazardous area under the 100-year flood level condition is 78.16 km^2 (Figure 8(a)). If the 100-year flood level increases to 2.92 m in 2100 (under RCP2.6), the flooded area will increase to 85.65 km^2 (Figure 8(b)). The flooded area becomes wider in 2100–87.85 km^2 (estimated under RCP4.5, Figure 8(c)) or 93.54 km^2 (estimated under RCP8.5, Figure 8(d)). Next, we investigated how the increase of

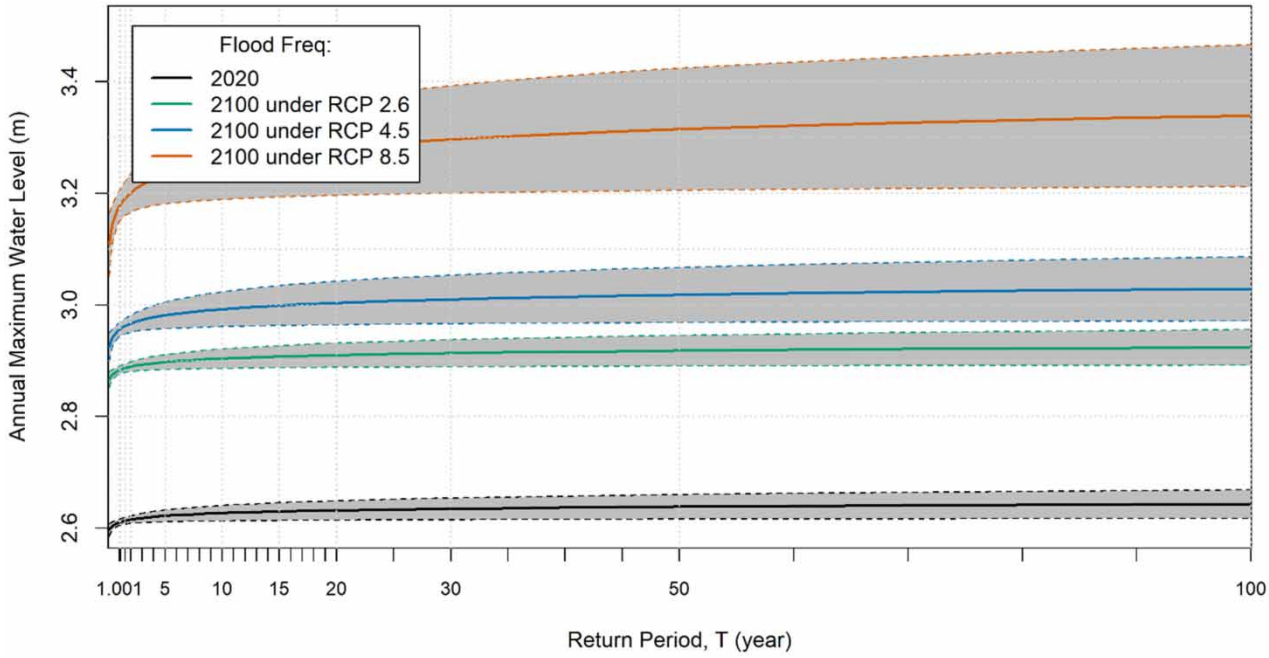

Figure 7 | Flood frequency analysis in Pontianak for 2090–2100 under the three future climate scenarios. Gray shading areas show a 95% confidence interval for each future scenario.

flooded areas impacted by climate change increases the infrastructure's exposure (Table 3). We estimated the number of infrastructures affected by each increase in the flood-hazard level. However, the estimation was only made for the infrastructures located within the city of Pontianak. Our results show that the number of affected buildings in 2100 will increase by about 6, 8, and 12% under the RCP2.6, RCP4.5, and RCP8.5 scenarios. Similarly, the fraction of impacted road length will rise from about 72% (in the current state) to about 78, 80, and 84% in 2100 under the RCP2.6, RCP4.5, and RCP8.5 scenarios, respectively.

4. DISCUSSION

Since the area of interest is low-lying land with a generally low slope ($\leq$ 8%) (Arianti *et al.* 2020) and in almost a natural state (with no dams, dykes, or levees), a 2.64-m flood level (1% annual exceedance probability) already causes an inundation in a significantly large part of the study area (Table 3). The overflow water runs freely into the city through the drainage canals. Therefore, an increase of 100-year flood level of 28 cm, in a low emission scenario (RCP2.6), already affects infrastructures severely (6.4% increase in impacted buildings and 6.8% increase in impacted road lengths). In the medium emission scenario (RCP4.5), while the 100-year flood level will increase by about 39 cm in 2100, the inundation will cause 8.1% more impacted buildings and 8.4% more affected road lengths. Lastly, in the highest emission scenario (RCP8.5), when the 100-year flood level increases by 70 cm from the current state, there will be 11.9% more impacted buildings and 12.7% more affected roads.

Here, we qualitatively categorized the flood risk for each sub-area (district) as high and low regarding the spatial distribution of flood hazards and exposed infrastructures, i.e., buildings and roads (Environment Agency UK 2013). Figure 7 shows that the high-risk area is located in the eastern and western parts of the city. Most of these areas will be inundated under the 100-year flood level condition. Therefore, we can say that the flood hazard in these areas is high. The exposure is also high because there are many infrastructures in these areas, which means they are highly populated or function as, for example, a business center. The high hazards and dense infrastructures made this area a high-risk flood zone.

Meanwhile, the northern part of the city has a lower infrastructure density. The flood hazard in this area is also low due to its higher altitude. A significant portion of this area will not be inundated under all 100-year flood scenarios. Therefore, the flood risk of this zone is categorized as low (Environment Agency UK 2013). Another non-inundated area is the southern part

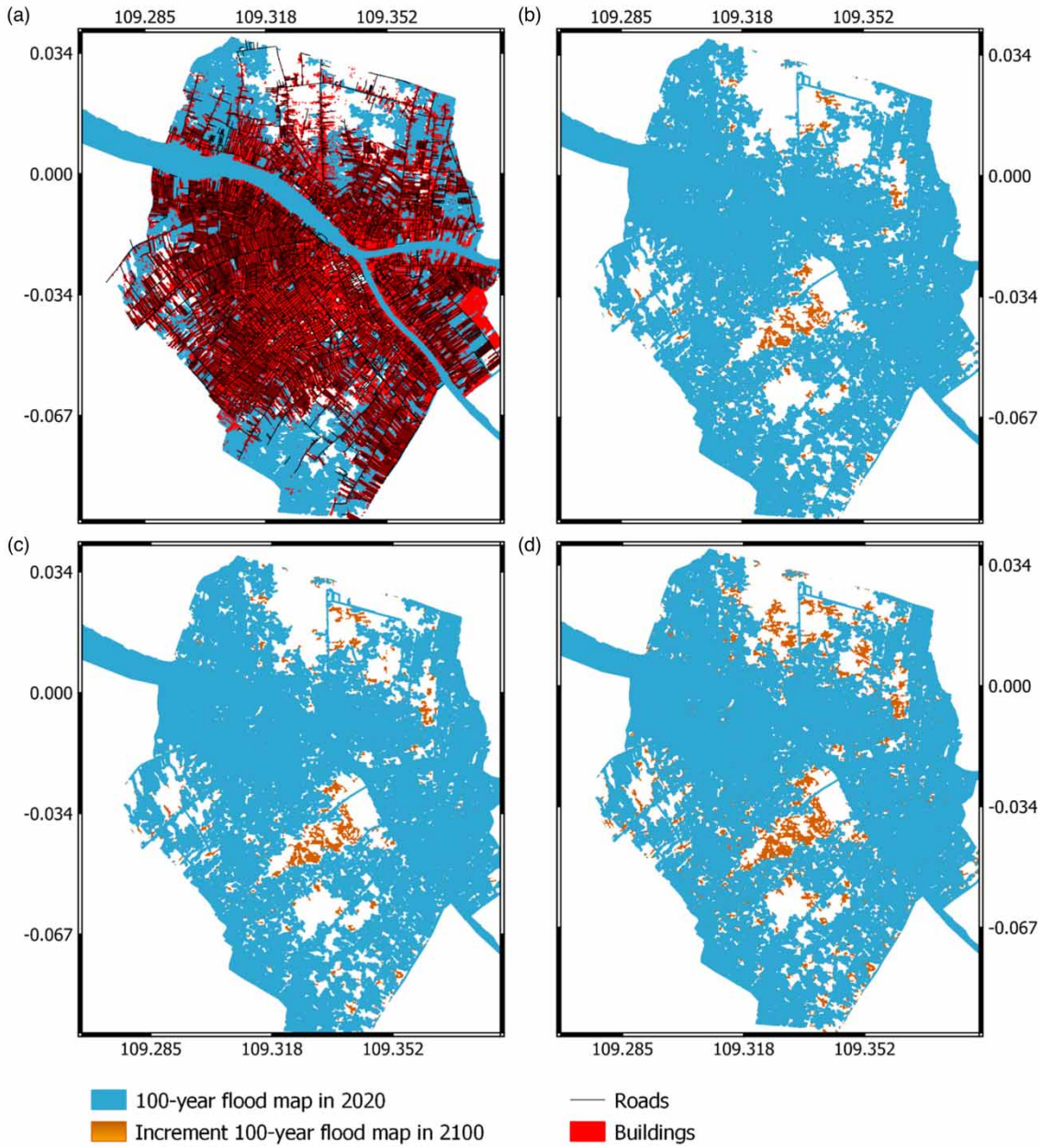

Figure 8 | Flood risk maps in Pontianak under the 100-year flood level condition, comparing the current state (a), under RCP2.6 (b), RCP4.5 (c), and RCP8.5 (d). Note that roads and buildings are only shown in (a), where the blue area (inundated areas) below the red (buildings) is covered (cannot be seen).

Table 3 | Flood risk analysis in the city of Pontianak and the impacted infrastructures

			Impacted Buildings			Impacted Roads		
Scenario	**Flood 100yr (m)**	**Flood Extent Area (km²)**	**NWet**	**NDry**	**%**	**Length_wet (km)**	**Length_dry (km)**	**%**
2020_Current	2.64	78.16	1.16×10^5	3.52×10^4	76.7	1,323	526	71.6
2100_RCP26	2.92	85.65	1.26×10^5	2.56×10^4	83.1	1,448	400	78.4
2100_RCP45	3.03	87.85	1.28×10^5	2.30×10^4	84.8	1,480	369	80.0
2100_RCP85	3.34	93.54	1.34×10^5	1.72×10^4	88.6	1,558	290	84.3

of the joining point of the Landak River and the Kapuas Kecil River, as well as some parts of the city's southwestern area. Even though these areas have dense infrastructures, we still classified the flood risk of these zones as low.

Nevertheless, this study has some limitations. Firstly, we do not account for potential collapse of the Antarctic ice sheet in future scenarios (van de Wal *et al.* 2019). Secondly, the flood analysis is limited to the extent of the city of Pontianak. Hazards and exposure areas outside the city are not included in the assessment. Next, the hazards are classified only in two states: inundated and not inundated, without considering the depth of the flood (Islam & Sado 2000). In addition, the impact assessment on the infrastructure (buildings and roads) only depends on the DEM map (FEMA 2003). The analysis result may differ due to local conditions, such as higher local terrain maps and infrastructure types. Lastly, it is also essential to notice that the MLR model is trained by only three months of data in 2020. The water levels could be higher if we considered other flood events in other years.

Despite the limitations, the local water management or government can use this study's results to mitigate future flood events in the study area. The model's evaluation of climate change impacts can guide the adaptation strategies, such as whether or not it is urgent to adjust the height of flood defense structures along the riverbanks within the city. Next, they can track the hazardous area throughout the city and watch what happens as the water comes from all possible causes. Moreover, other deltas with similar characteristics and limitations can adopt the approach to assess their future flood risk.

5. CONCLUSION

This study successfully assesses future flood risk in the KRD, particularly in Pontianak, using integrated ML and GIS-based bathtub inundation models. We simulated the water level dynamics and quantified the flood frequency curve of the current and future states as modulated by climate change. We created flood maps with potentially inundated areas in 100-year flood (1% annual exceedance probability) under the current and future scenarios. We found that the 100-year flood level in the study area will increase from the current 2.64 m to 2.92 m, 3.03 m, and 3.34 m in 2100 under each future climate scenario (RCP2.6, RCP4.5, and RCP8.5), respectively. These increases correlate to the increment of flood hazard areas over the region of interest. We found that in 2100 more buildings will be exposed (increased by about 6.4%–11.9%), and more roads will be impacted (increased by approximately 6.8%–12.7%) depending on the climate scenario. This assessment benefits the local water managers in preparing adequate mitigation strategies and the city's disaster management plan.

ACKNOWLEDGEMENTS

This work was funded by the Indonesia Endowment Fund for Education (LPDP) under Grant No. 201712220212183.

DATA AVAILABILITY STATEMENT

The relevant data and code for this paper are available at: https://zenodo.org/record/7370291

CONFLICT OF INTEREST

The authors declare there is no conflict.

REFERENCES

Aldrian, E. & Susanto, R. D. 2003 Identification of three dominant rainfall regions within Indonesia and their relationship to sea surface temperature. *Int. J. Climatol.* **23**, 1435–1452. https://doi.org/10.1002/joc.950.

Arianti, I., Soemarno, S., Hasyim, W. & Sulistyono, R. 2020 Determining variable weight of flood vulnerability spatial model in Pontianak City based on Analytic Hierarchy Process (AHP). *Int. J. Humanit. Relig. Soc. Sci.* **4**, 2548–5725.

Bhatta, B. 2010 Causes and consequences of urban growth and sprawl. *Adv. Geogr. Inf. Sci* 17–36. https://doi.org/10.1007/978-3-642-05299-6_2.

Bishop, C. M. 2006 *Pattern Recognition and Machine Learning*. Springer, Singapore, p. 738.

Block, P. J., Filho, F. A. S., Sun, L. & Kwon, H.-H. 2009 A streamflow forecasting framework using multiple climate and hydrological models1. *JAWRA J. Am. Water Resour. Assoc.* **45**, 828–843. https://doi.org/10.1111/J.1752-1688.2009.00327.X.

Cacuci, D. G. 2003 *Sensitivity and Uncertainty Analysis*. Chapman & Hall/CRC Press, Boca Raton.

Chan, F. K. S., Yang, L. E., Scheffran, J., Mitchell, G., Adekola, O., Griffiths, J., Chen, Y., Li, G., Lu, X., Qi, Y., Li, L., Zheng, H. & McDonald, A. 2021 Urban flood risks and emerging challenges in a Chinese delta: the case of the Pearl River Delta. *Environ. Sci. Policy* **122**, 101–115. https://doi.org/10.1016/j.envsci.2021.04.009.

Choi, C., Kim, J., Han, H., Han, D. & Kim, H. S. 2020 Development of water level prediction models using machine learning in wetlands: a case study of Upo Wetland in South Korea. **12**, 93. https://doi.org/10.3390/w12010093.

Chu, P. C. 1999 Two kinds of predictability in the lorenz system. *J. Atmos. Sci.* **56**, 1427–1432.

Dinh, Q., Balica, S., Popescu, I. & Jonoski, A. 2012 Climate change impact on flood hazard, vulnerability and risk of the Long Xuyen Quadrangle in the Mekong Delta. *Int. J. River Basin Manage.* **10**, 103–120. https://doi.org/10.1080/15715124.2012.663383.

Environtment Agency UK 2013 *Northumbria River Basin District: Flood Risk Maps for Rivers and the Sea, Flood Risk Map*. LIT, Bristol, p. 8966.

FEMA 2003 *Multi-Hazard Loss Estimation Methodology: HAZUS-MH, Technical Manual*. US Department of Homeland Security, Washington, DC.

Gallien, T. W. 2016 Validated coastal flood modeling at Imperial Beach, California: comparing total water level, empirical and numerical overtopping methodologies. *Coastal Eng.* **111**, 95–104. https://doi.org/10.1016/j.coastaleng.2016.01.014.

Ganguli, P., Paprotny, D., Hasan, M., Güntner, A. & Merz, B. 2020 Projected changes in compound flood hazard from riverine and coastal floods in Northwestern Europe. *Earth's Future* **8**, e2020EF001752. https://doi.org/10.1029/2020EF001752.

Gumbel, E. J. 1958 *: Statistics of Extremes*. Columbia University Press, New York, p. 375.

Habert, J., Ricci, S., Le Pape, E., Thual, O., Piacentini, A., Goutal, N., Jonville, G. & Rochoux, M. 2016 Reduction of the uncertainties in the water level-discharge relation of a 1D hydraulic model in the context of operational flood forecasting. *J. Hydrol.* **532**, 52–64. https://doi.org/10.1016/j.jhydrol.2015.11.023.

Hallegatte, S. 2009 Strategies to adapt to an uncertain climate change. *Global Environ. Change* **19**, 240–247. https://doi.org/10.1016/j.gloenvcha.2008.12.003.

Hornik, K., Buchta, C. & Zeileis, A. 2008 Open-source machine learning: R meets Weka. *Comput. Stat.* **24**, 225–232. https://doi.org/10.1007/S00180-008-0119-7.

InaSAFE 2022. Available from: http://inasafe.org/ (accessed 26 January 2022).

IPCC. 2012 Managing the risks of extreme events and disasters to advance climate change adaptation. In: *A Special Report of Working Groups I and II of the Intergovernmental Panel on Climate Change* (Field, C. B., Barros, V., Stocker, T. F., Qin, D., Dokken, D. J., Ebi, K. L., Mastrandrea, M. D., Mach, K. J., Plattner, G.-K., Allen, S. K., Tignor, M. & Midgley, P. M., eds). Cambridge University Press, Cambridge, UK.

Islam, M. M. & Sado, K. 2000 Flood hazard assessment in Bangladesh using NOAA AVHRR data with geographical information system. *Hydrol. Processes* **14**, 605–620. https://doi.org/10.1002/(SICI)1099-1085(20000228)14:3 < 605::AID-HYP957 > 3.0.CO;2-L.

Iturbide, M., Fernández, J., Gutiérrez, J. M., Bedia, J., Cimadevilla, E., Díez-Sierra, J., Manzanas, R., Casanueva, A., Baño-Medina, J., Milovac, J., Herrera, S., Cofiño, A. S., San Martín, D., García-Díez, M., Hauser, M., Huard, D. & Yelekci, Ö. 2021 Repository supporting the implementation of FAIR principles in the IPCC-WG1 Atlas. *Zenodo* doi:10.5281/zenodo.3691645. https://doi.org/10.5281/zenodo.5171760.

Jackson, E. K., Roberts, W., Nelsen, B., Williams, G. P., Nelson, E. J. & Ames, D. P. 2019 Introductory overview: error metrics for hydrologic modelling – A review of common practices and an open source library to facilitate use and adoption. *Environ. Modell. Software* **119**, 32–48. https://doi.org/10.1016/j.envsoft.2019.05.001.

Kinney, J. B. & Atwal, G. S. 2014 Equitability, mutual information, and the maximal information coefficient. *Proc. Natl. Acad. Sci.* **111**, 3354–3359. https://doi.org/10.1073/pnas.1309933111.

Kundzewicz, Z. W., Kanae, S., Seneviratne, S. I., Handmer, J., Nicholls, N., Peduzzi, P., Mechler, R., Bouwer, L. M., Arnell, N., Mach, K., Muir-Wood, R., Brakenridge, G. R., Kron, W., Benito, G., Honda, Y., Takahashi, K. & Sherstyukov, B. 2014 Flood risk and climate change: global and regional perspectives. *Hydrol. Sci. J.* **59**, 1–28. https://doi.org/10.1080/02626667.2013.857411.

Lange, J. 2020 *IU Study Shows Coastal Flooding Will Disproportionately Impact River Deltas Populations*. Environmental Resilience Institute Toolkit (ERIT), Bloomington.

Lilai, X., Yuanrong, H., Wei, H. & Shenghui, C. 2016 A multi-dimensional integrated approach to assess flood risks on a coastal city, induced by sea-level rise and storm tides. *Environ. Res. Lett.* **11**, 014001. https://doi.org/10.1088/1748-9326/11/1/014001.

MacKinnon, K., Hatta, G., Mangalik, A. & Halim, H. 1996 *The Ecology of Kalimantan*, Vol. 3. Oxford University Press, Oxford, UK.

Marengo, J. A., Alves, L. M., Ambrizzi, T., Young, A., Barreto, N. J. C. & Ramos, A. M. 2020 Trends in extreme rainfall and hydrogeometeorological disasters in the Metropolitan Area of São Paulo: a review. *Ann. N. Y. Acad. Sci.* **1472**, 5–20. https://doi.org/10.1111/NYAS.14307.

Moftakhari, H. R., Salvadori, G., AghaKouchak, A., Sanders, B. F. & Matthew, R. A. 2017 Compounding effects of sea level rise and fluvial flooding. *Proc. Natl. Acad. Sci. USA.* **114**, 9785–9790. https://doi.org/10.1073/pnas.1620325114.

Mukherjee, S., Mishra, A. & Trenberth, K. E. 2018 Climate change and drought: a perspective on drought indices. *Curr. Clim. Change Rep.* **4**, 145–163. https://doi.org/10.1007/s40641-018-0098-x.

Murdukhayeva, A., August, P., Bradley, M., Labash, C. & Shaw, N. 2013 Assessment of inundation risk from sea level rise and storm surge in Northeastern Coastal National Parks. *J. Coast. Res.* **29**, 1–16. https://doi.org/10.2112/JCOASTRES-D-12-00196.1.

Ngo, H., Pathirana, A., Zevenbergen, C. & Ranasinghe, R. 2018 An effective modelling approach to support probabilistic flood forecasting in coastal cities – case study: Can Tho, Mekong Delta, Vietnam. *J. Mar. Sci. Eng.* **6**, 55. https://doi.org/10.3390/jmse6020055.

Nguyen, D. T. & Chen, S. T. 2020 Real-Time probabilistic flood forecasting using multiple machine learning methods. *Water* **12**, 787. https://doi.org/10.3390/w12030787.

Noymanee, J. & Theeramunkong, T. 2019 Flood forecasting with machine learning technique on hydrological modeling. *Procedia Comput. Sci.* **156**, 377–386. https://doi.org/10.1016/j.procs.2019.08.214.

OpenStreetMap 2020 Planet dump. Available from: https://planet.osm.org: https://www.openstreetmap.org (accessed 20 October 2020).

Psych: Procedures for Psychological, Psychometric, and Personality Research 2021. Available from: https://cran.r-project.org/package=psych (accessed 16 December 2021).

Ridha, T., Ross, A. D. & Mostafavi, A. 2022 Climate change impacts on infrastructure: flood risk perceptions and evaluations of water systems in coastal urban areas. *Int. J. Disaster Risk Reduct.* **73**, 102883. https://doi.org/10.1016/j.ijdrr.2022.102883.

Ross, B. C. 2014 Mutual information between discrete and continuous data sets. *PLoS One* **9**, e87357. https://doi.org/10.1371/journal.pone.0087357.

Ruslan, F. A., Samad, A. M., Zain, Z. M. & Adnan, R. 2014 Flood water level modeling and prediction using NARX neural network: Case study at Kelang river. In: *Proc. - 2014 IEEE 10th Int. Colloq. Signal Process. Its Appl. CSPA*. pp. 204–207, https://doi.org/10.1109/CSPA.2014.6805748.

Sampurno, J., Vallaeys, V., Ardianto, R. & Hanert, E. 2022 Modeling interactions between tides, storm surges, and river discharges in the Kapuas River delta. **19**, 2741–2757. https://doi.org/10.5194/bg-19-2741-2022.

Van de Sande, B., Lansen, J. & Hoyng, C. 2012 Sensitivity of coastal flood risk assessments to digital elevation models. *Water* **4**, 568–579. https://doi.org/10.3390/w4030568.

van de Wal, R. S. W., Zhang, X., Minobe, S., Jevrejeva, S., Riva, R. E. M., Little, C., Richter, K. & Palmer, M. D. 2019 Uncertainties in long-term twenty-first century process-based coastal sea-level projections. *Surv. Geophys.* **406** (40), 1655–1671. https://doi.org/10.1007/S10712-019-09575-3.

Vousdoukas, M. I., Voukouvalas, E., Mentaschi, L., Dottori, F., Giardino, A., Bouziotas, D., Bianchi, A., Salamon, P. & Feyen, L. 2016 Developments in large-scale coastal flood hazard mapping. *Nat. Hazards Earth Syst. Sci.* **16**, 1841–1853. https://doi.org/10.5194/NHESS-16-1841-2016.

Wang, L. & Liu, H. 2007 An efficient method for identifying and filling surface depressions in digital elevation models for hydrologic analysis and modelling. **20**, 193–213. https://doi.org/10.1080/13658810500433453.

Wu, H., Adler, R. F., Tian, Y., Huffman, G. J., Li, H. & Wang, J. 2014 Real-time global flood estimation using satellite-based precipitation and a coupled land surface and routing model. *Water Resour. Res.* **50**, 2693–2717. https://doi.org/10.1002/2013WR014710.

Xu, R., Frank, K. A., Maroulis, S. J. & Rosenberg, J. M. 2019 konfound: command to quantify robustness of causal inferences. *Stata J.* **19**, 523–550. https://doi.org/10.1177/1536867X19874223.

Yunus, A. P., Avtar, R., Kraines, S., Yamamuro, M., Lindberg, F. & Grimmond, C. S. B. 2016 Uncertainties in tidally adjusted estimates of sea level rise flooding (Bathtub Model) for the greater London. *Remote Sens.* **8**, 366. https://doi.org/10.3390/rs8050366.

Zhang, Y., Chiew, F. H. S., Li, M. & Post, D. 2018 Predicting runoff signatures using regression and hydrological modeling approaches. *Water Resour. Res.* **54**, 7859–7878. https://doi.org/10.1029/2018WR023325.

First received 22 July 2022; accepted in revised form 16 November 2022. Available online 25 November 2022