ORIGINAL PAPER

Impact of land cover, rainfall and topography on flood risk in West Java

Rahmawati Rahayu[1] · Simon A. Mathias[2] · Sim Reaney[3,4] · Gianni Vesuviano[5] · Rusmawan Suwarman[6] · Agus M. Ramdhan[7]

Abstract

Flooding represents around 32% of total disasters in Indonesia and disproportionately affects the poorest of communities. The objective of this study was to determine significant statistical differences, in terms of river catchment characteristics, between regions in West Java that reported suffering from flood disasters and those that did not. Catchment characteristics considered included various statistical measures of topography, land-use, soil-type, meteorology and river flow rates. West Java comprises 154 level 9 HydroSHEDS sub-basin regions. We split these regions into those where flood disasters were reported and those where they were not, for the period of 2009 to 2013. Rainfall statistics were derived using the CHIRPS gridded precipitation data package. Statistical estimates of river flow rates, applicable to ungauged catchments, were derived from regionalisation relationships obtained by stepwise linear regression with river flow data from 70 West Javanese gauging stations. We used Kolmogorov–Smirnov tests to identify catchment characteristics that exhibit significant statistical differences between the two sets of regions. Median annual maximum river flow rate (AMRFR) was found to be positively correlated with plantation cover. Reducing plantation land cover from 20 to 10% was found to lead to a modelled 38% reduction in median AMRFR. AMRFR with return periods greater than 10 years were found to be negatively correlated with wetland farming land cover, suggesting that rice paddies play an important role in attenuating extreme river flow events. Nevertheless, the Kolmogorov–Smirnov tests revealed that built land cover is the most important factor defining whether or not an area is likely to report flood disasters in West Java. This is presumably because the more built land cover, the more people available to experience and report flood disasters. Our findings also suggest that more research is needed to understand the important role of plantation cover in aggravating median annual maximum river flow rates and wetland farming cover in mitigating extreme river flow events.

✉ Simon A. Mathias
s.a.mathias@durham.ac.uk

Extended author information available on the last page of the article

1 Introduction

Flooding represents around 32% of total disasters in Indonesia (Yulianto et al. 2015) and disproportionately affects the poorest of communities (Kim and Gim 2020). The impact of fluvial flooding is expected to increase due to increasing rainfall intensity, land subsidence, sedimentation of river channels, rapid population growth and land-use change (Marfai et al. 2015). Historical changes in both land-use and climate have collectively led to increased river flow rates in the region of West Java, with land-use thought to have played a stronger role (Julian Poerbandono and Ward 2014).

Since as early as the 1920s, agricultural activity has represented over 70% of total land cover in Java (Verburg and Bouma 1999). Land cover in West Java can be broadly categorised into six land-use types including water body, built land, dryland farming (mostly agriculture with low irrigation requirements), wetland farming (mostly rice paddies and fish farms), plantation (mostly tea plantation) and forest (including both natural and managed forest) (consider Siswanto and Francés 2019; Ridwansyah et al. 2020; Yulianto et al. 2022). Indonesian spatial planning law prescribes that 30% forest cover is needed in all local government districts to ensure adequate hydrological forest functions (Suprayogo et al. 2020).

Forest cover, as compared to other land cover, provides higher rates of evapotranspiration (including canopy interception loss), higher rates of infiltration and lower rates of surface runoff. A plot scale study in East Java measured runoff coefficients for production forest (both old and young) and arable land to be 0.03 and 0.41, respectively (Suprayogo et al. 2020). Interestingly, hydrological models for West Java often assume runoff coefficients for primary forest, secondary forest and plantation of 0.01, 0.03 and 0.40, respectively (Yulianto et al. 2022).

Hydrological modelling in the Upper Citarum, West Java, suggests that reductions in forest cover, combined with increases in settlement cover, during the period of 1994 to 2014, led to a 15% increase in water yield (Siswanto and Francés 2019). Hydrological modelling in the Cimanuk catchment, West Java, suggests that transformation of bush and plantation to dryland agriculture, during the period of 2011 to 2017, led to a 60% increase in surface runoff (Ridwansyah et al. 2020). However, findings from such studies are strongly dependent on the assumptions and parameter values that define the underlying hydrological models used (Beven and Binley 1992; Beven 2006; Morán-Tejeda et al. 2015; Ekblad and Herman 2021). Furthermore, whether or not increases in river flow lead to problematic flood events is dependent on the damage and disruption likely to be caused (Ferreira and Ghimire 2012). An alternative approach is to directly study historical reports of flood disaster occurrence.

By studying variations in country level "large flood event" reporting data from 56 countries, Bradshaw et al. (2007) concluded that a 10% reduction in natural forest area can lead to between 4 and 28% increase in flood disaster frequency. They derived their results using multiple linear regression considering country area, median average annual precipitation, average slope, total degraded area, natural forest cover, natural forest loss and non-natural forest cover. Their flood frequency data was obtained from the Dartmouth Flood Observatory, which is mostly derived from news and government sources (Brakenridge 2010).

However, revisiting the Bradshaw et al. (2007) analysis, van Dijk et al. (2009) found that the dependence on natural forest cover reduced to a negligible level when population density was additionally considered. Furthermore, they found that population density explained 83% of the variation in reported flood occurrences. Similar country scale

findings were also observed by Ferreira and Ghimire (2012). Indeed, flood events are much more likely to be reported in areas where there are more people to experience and report the events.

Nevertheless, Bhattacharjee and Behera (2018) applied a similar methodology to identify key variables associated with reported flood occurrence at a district level. Specifically they looked at 13 districts in Western Bengal, India, and found that the frequency of reported flood events was strongly negatively correlated with forest cover. Important differences between the study of Bhattacharjee and Behera (2018) and those of Bradshaw et al. (2007), van Dijk et al. (2009) and Ferreira and Ghimire (2012) were that they considered a finer resolution (district as opposed to country scale) and had access to arguably more reliable flood reporting data (obtained directly from the Department of Disaster Management, government of West Bengal, Kolkata). A similar district-scale study in China revealed that frequency of reported flood events was negatively correlated with broadleaf and mixed-tree forest cover but independent of coniferous tree cover (Tembata et al. 2020). The validity of such studies is important because they are likely to affect policy making decisions relating to the preserving and/or extending of forest cover to reduce flood damage in downstream urban environments (Xiao et al. 2022).

A disadvantage of using countries or districts to define individual study areas is that the land within a district does not necessarily represent the hydrological watershed upstream of a given flood event. It would be better to study hydrological units as opposed to district level units, so as to better capture the hydrological contribution area of the flood events being studied.

Another problem concerns the strong correlation between flood reporting frequency and population density. Certainly flood reporting frequency will increase if there are more people to report the floods (van Dijk et al. 2009; Ferreira and Ghimire 2012). An alternative approach is to study statistical differences between areas that report flood disasters and areas that do not over a designated multi-year period. The advantage is that a flood disaster reported at one location in a small village will have the same weight as a flood disaster reported at multiple locations in a much larger town or city. It should then be possible to investigate differences between hydrological regions that report flood disasters and those that do not.

Unfortunately, this kind of dichotomous data is unsuitable for conventional linear regression techniques. However, parameter sensitivity can be explored using a two sample Kolmogorov–Smirnov test (KST). For example, suppose we have forest cover data for 200 river catchments. We can split the catchments into those that reported flood disasters and those that did not. We can then use the KST to determine whether there is a statistically significant difference between forest cover data describing the flood disaster catchments and the remainder of the population. Should such a difference exist, we can use the Kolmogorov–Smirnov statistic (KSS) as a measure of parameter sensitivity.

Now suppose we study a whole range of different factors (forest cover, mean annual rainfall, median annual maximum river flow rate etc.). The KST can be used to identify which parameters show statistically significant differences. Those statistically significant parameters can then be ranked in terms of their KSS. Furthermore, the CDFs for the most important parameters can be directly studied to gain further insights concerning the difference between catchments that report flood disasters and those that did not.

The objective of this study is to identify catchment characteristics that exhibit statistically significant differences in CDFs for river catchments that report flood disasters in West Java, Indonesia (including the government districts of West Java, Jakarta and Banten). Improved understanding about the spatial pattern of flood risk will help inform land use

planning at the national scale, hopefully leading to disaster risk reduction and improved resilience to extreme hydro-meteorological events.

During the period from 2009 to 2013, 601 flood disasters in West Java were reported to the Indonesian National Board for Disaster Management (BNPB 2021). West Java comprises 154 level 9 HydroSHEDS sub-basin regions (Lehner et al. 2008). We split these regions into those where flood disasters were reported and those where they were not. We derive statistical estimates of river flow rates from regionalisation relationships obtained by stepwise linear regression with river flow data from 70 West Javanese river flow gauging stations. We then use KSTs to identify catchment characteristics that exhibit significant statistical differences between the two sets of regions. Catchment characteristics considered include various statistical measures of topography, land-use, soil-type, meteorology and river flow.

The outline of this article is as follows. The various data sources are explained. The methods used for flood frequency analysis, stepwise linear regression and Kolmogorov–Smirnov testing are described. Results from the stepwise linear regression and Kolmogorov–Smirnov testing are presented and discussed. Relevant conclusions are then drawn.

2 Data

2.1 HydroSHEDS data

Digital elevation data, drainage direction data, river network shape-files and level 9 HydroSHEDS sub-basins (standard) shape-files were acquired at 15 arc-second resolution (approximately 500 m), for the West Java region, from Lehner et al. (2008). The digital elevation data, river network shape-files and sub-basin shape-files are presented in Fig. 1a.

HydroSHEDS uses the Pfafstetter coding system (PCS) to generate twelve nested levels of sub-basins (Lehner and Grill 2013), each depicting consistently sized sub-basin polygons at scales ranging from millions (level 1) to tens of square kilometers (level 12) (Linke et al. 2019). Level 9 was selected for this study because it provides catchments of similar size to gauged river catchments in West Java.

The PCS splits each catchment into a set of hydrological regions based on where tributaries intersect with the main river channel (Lehner 2014). The PCS distinguishes between inter-basin regions and tributary basins. Tributary basins collect water only from the within their associated area. Inter-basin regions collect water from within their associated area and also any part of the wider river catchment upstream of the inter-basin region.

The polygons provided by the HydroSHEDS sub-basin package include tributary basins and inter-basin regions. Both sets of polygons are hereafter referred to as "regions". Each region has two areas identified with it, the area of the region and the area of catchment that contributes to the river outlet of the region. For tributary basins, the catchment area is equal to the region area. For inter-basin regions, the catchment area is greater than the region area.

HydroSHEDS delineates West Java into 154 level 9 HydroSHEDS sub-basins. The polygon areas associated with each region range from 57.2 to 879 km^2 with a median value of 230 km^2. The catchment areas associated with each region range from 58.0 to 6710 km^2 with a median value of 298 km^2. For each region we determine the number of flood

a) Elevation (mASL)

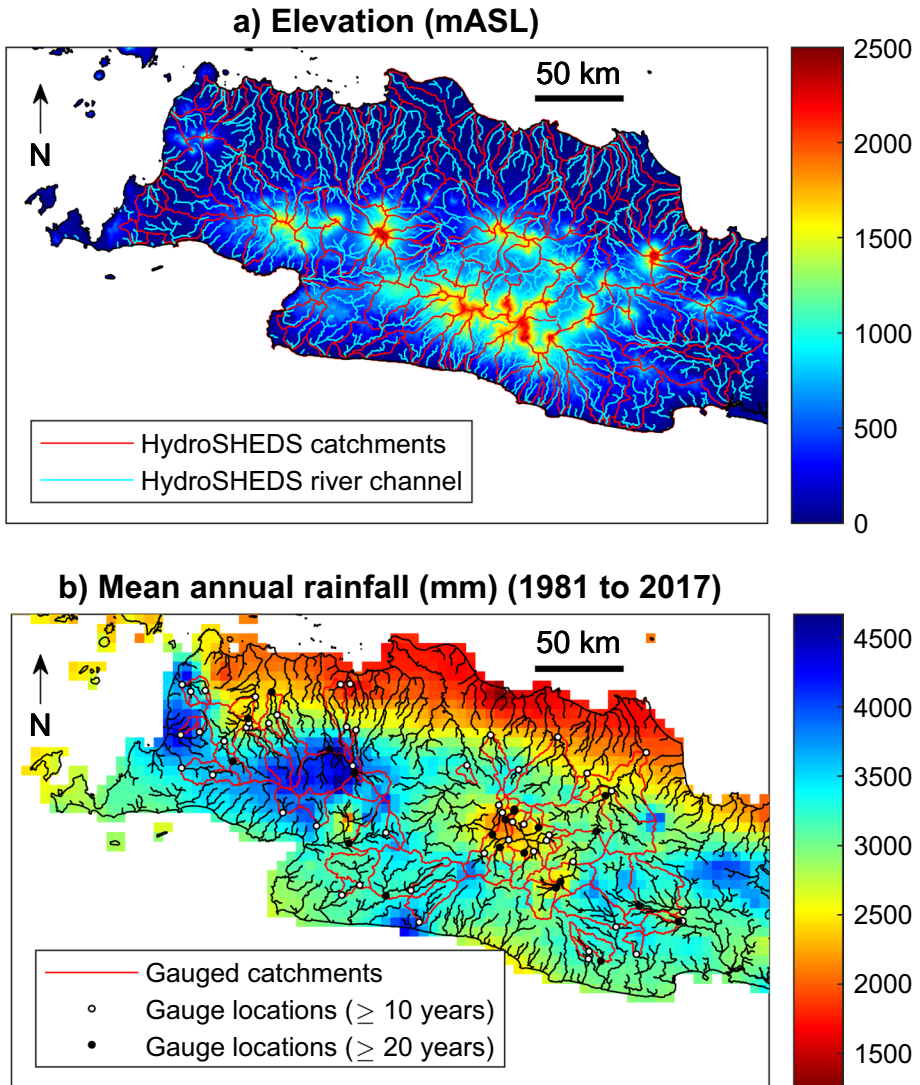# b) Mean annual rainfall (mm) (1981 to 2017)

Fig. 1 Maps of West Java, Indonesia. **a** Shows digital elevation data, river network data and level 9 basins, all acquired from HydroSHEDS (Lehner et al. 2008). **b** Shows mean annual rainfall from CHIRPS (Funk et al. 2015) along with the locations of the river flow gauging stations and their associated catchments

disasters reported within the area associated with the region and catchment characteristics associated with the catchment area upstream of the river outlet of the region.

In addition to the level 9 sub-basins (hereafter referred to as the HydroSHEDS catchments), the catchment areas for the river flow gauging stations were derived from the HydroSHEDS drainage direction data using the D8 algorithm (Jenson and Domingue 1988).

For each of the gauged catchments and HydroSHEDS catchments, we determined the catchment area (km^2), circularity ratio (= 4π catchment area ÷ square of catchment

perimeter), drainage density (= length of river channel network ÷ catchment area) (km^{-1}), P10 elevation (mASL), P50 elevation (mASL), P90 elevation (mASL), P10 slope (m km^{-1}), P50 slope (m km^{-1}) and P90 slope (m km^{-1}). The range and median for the above set of catchment characteristics are given for both the gauged catchments and the HydroSHEDS catchments in Table 1.

The terms P10, P50 and P90 refer to values with a 10%, 50% and 90% probability of non-exceedance within a given catchment. Length of river channel network was obtained from the 15 arc-second HydroSHEDS river network shape-files. The term mASL stands for meters above sea level.

2.2 Flood disaster reports

Records were acquired from all flood disasters reported to the Indonesian National Board for Disaster Management (BNPB 2021) for West Java during the period of 2009 to 2013. This dataset included 601 individual river flooding events, the locations of which are shown on Fig. 2a. Each of these events represents anything from an unplanned flood inundation to a flood event with large-scale property damage and multiple fatalities. The vast majority of events can be seen to be located in the high population areas of Jakarta and Bandung. Nevertheless, many additional events are spread out through many rural regions of West Java. But interestingly, some HydroSHEDS regions reported no flood events during the period of study (compare Figs. 1a and 2a).

2.3 Land-use data

Land-use data in West Java for 2011 were acquired at 1:250,000 from a web-resource managed by the Indonesia Ministry of Environment and Forestry (IMEF 2021). These data were derived from Landsat multispectral satellite data combined with ground check field survey (IMEF 2020). The web-resource distinguishes between fifteen different classifications including water body, built land, dryland farming, dryland farming mixed with shrubs, garden shrubs, mining area, open land, rice field, aquaculture pond, swamps, plantation, primary dryland forest, secondary dryland forest, mangrove forest, industrial forest. The classifications used were broadly derived from those previously described by Di Gregorio et al. (1998).

For our study, we aggregated these fifteen classifications into six simplified land-use types (see Table 2) perceived to have common hydrological functioning attributes, defined as follows:

1. *Water body* includes rivers, reservoirs and lakes.
2. *Built land* includes urban and rural settlements, factories and other built-up infrastructure.
3. *Dryland farming* includes cropland areas requiring low levels of irrigation mainly used for seasonal crops. Here we also include shrubs, mining area and open land. Mining area and open land represent 0.01% and 0.7% of total land cover in West Java, respectively, and are mostly covered by grass and shrubs. The hydrological functioning is therefore thought to be very similar to other dryland farming areas.
4. *Wetland farming* includes agricultural areas that experience both permanent and periodic inundation. These areas are mostly comprised of rice paddies. However, we have chosen to also include aquaculture ponds and swamp, the latter of which represents < 0.001% of
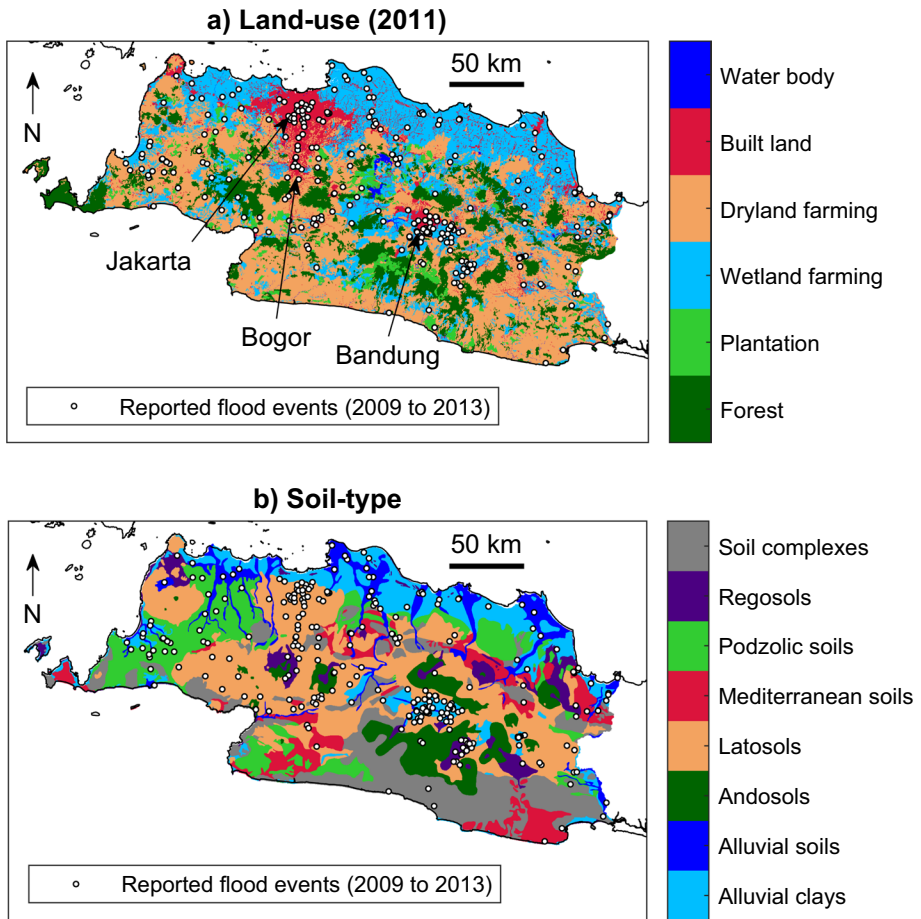
Fig. 2 Maps of West Java, Indonesia. **a** Shows land-use data for 2011 from IMEF (2021). **b** Shows soil-type data from the Soil Research Centre, Ministry of Agroforestry, Indonesia. Both maps also show locations of reported flood disasters from BNPB (2021)

total land cover in West Java. Aquaculture ponds mostly include fish and shrimp ponds, which are commonly integrated within the rice farming system in Java (Nurhayati et al. 2016; Fatimahet al. 2020) and are therefore better considered with rice as a single land-use type from a hydrological functioning perspective.

5. *Plantation* includes dryland farming like practices where crops have an operational life greater than two years, mostly tea plantation.

6. *Forest* includes all dryland and mangrove forest, both natural and managed. Although mangrove forest is likely to have quite different hydrological functioning, it represents < 1% of total land cover in West Java and was therefore not considered as a separate classification in this context. Although mangrove forest is very important for mitigating coastal flooding (Menendez et al. 2018), its role on river flooding is thought to be much less significant.

Figure 2a shows a map for the six land-use classifications in West Java alongside the location of the aforementioned reported flood events. It can be seen that wetland farming is concentrated on the north coast but is also widely distributed within the valleys of the central highlands. Forest and plantation mostly reside in the highland mountain regions.

The percentage land cover for each of the six land-use classifications was determined for each of the gauged catchments as well as the 154 level 9 HydroSHEDS catchments. The range and median values for each classification are given in Table 1.

Land-use distribution varies widely from one hydrological catchment to another. Figure 3 shows the cumulative distribution functions (CDF) for the different land-use classifications, in terms of percentage of land cover, for each of the 154 level 9 HydroSHEDS catchments. Water body represents a very small fraction of land cover for all the catchments studied. Dryland farming predominantly represents the largest fraction. Interestingly, less than 20% of the catchments in West Java satisfy the 30% forest cover government requirement, referred to by Suprayogo et al. (2020).

2.4 Soil-type data

Soil-type data were acquired at 1:250,000 from the Soil Research Centre, Ministry of Agroforestry, Indonesia. Soils in West Java are dominated by alluvial clays, alluvial soils (comprising sand and clay), Andosols, Latosols, Mediterranean soils, podzolic soils, Regosols and soil complexes (a mixture of more than two types of the previously mentioned soils) (Suhardjo and Soepraptohardjo 1982).

The alluvial clays and podzolic soils have high clay content and can be considered relatively low in permeability. The Andosols and Regosols have high sand content and can be considered relatively high in permeability. Other soils should be expected to be more intermediate in permeability. From a hydrological perspective we expect to see higher runoff coefficients in catchments dominated by low permeability soils.

A map of soil-type across West Java is shown in Fig. 2b. Alluvial clays dominate the main flood plains on the North coast. Alluvial soils (which have a higher sand content than the alluvial clays) dominate the main river channels. Andosols and Regosols dominate the highlands around Bandung and Bogor (a smaller city to the south of Jakarta).

The percentage land cover for each of the eight soil-type classifications was determined for each of the gauged catchments as well as the 154 level 9 HydroSHEDS catchments. The range and median values for each classification are given in Table 1.

Fig. 3 Cumulative distribution functions for land use in each of the HydroSHEDS catchments studied. PNE stands for probability of non-exceedance
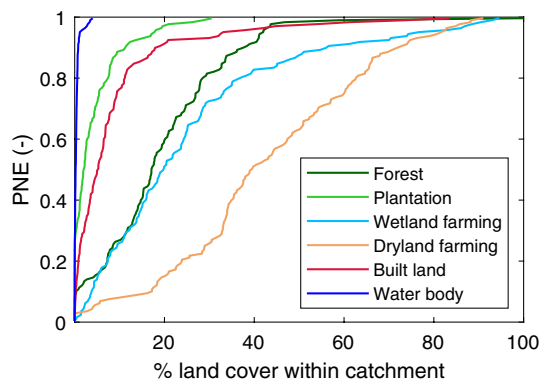
Table 1 Range and median values of catchment characteristics for both the gauged catchments and the HydroSHEDS catchments

	Gauged catchments			HydroSHEDS catchments		
	Min.	Median	Max.	Min.	Median	Max.
Catchment area (km^2)	9.34	211	3400	58	298	6710
Circularity ratio (-)	0.122	0.263	0.385	0.15	0.398	0.747
Drainage density (1/km)	0	0.157	0.3	0.0313	0.177	0.319
P10 elevation (mASL)	5.85	308	1470	0	72	966
P50 elevation (mASL)	13	603	1620	1	361	1340
P90 elevation (mASL)	18	1190	2010	4	854	2000
P10 slope (m km^{-1})	1.99	39.6	226	0	21.7	121
P50 slope (m km^{-1})	7.44	173	442	3.61	133	332
P90 slope (m km^{-1})	18.6	513	1440	10.9	472	930
Forest (%)	0.0	19.8	81.3	0.0	17.1	100.0
Plantation (%)	0.0	1.7	33.4	0.0	2.0	30.5
Wetland farming (%)	0.0	14.2	56.0	0.0	19.7	94.6
Dryland farming (%)	1.4	51.3	85.7	0.0	39.3	91.0
Built land (%)	0.0	4.1	98.6	0.0	5.1	83.5
Water body (%)	0.0	0.0	0.9	0.0	0.2	4.0
Alluvial clays (%)	0.0	0.0	24.7	0.0	3.4	86.0
Alluvial soils (%)	0.0	0.0	34.1	0.0	0.7	41.0
Andosols (%)	0.0	9.3	99.5	0.0	3.8	68.5
Latosols (%)	0.0	38.0	100.0	0.0	29.3	94.3
Mediterranean soils (%)	0.0	0.0	37.3	0.0	9.9	52.9
Podzolic soils (%)	0.0	0.0	93.2	0.0	0.0	88.3
Regosols (%)	0.0	0.0	85.4	0.0	0.7	65.4
Soil complexes (%)	0.0	0.0	100.0	0.1	17.3	100.0
Mean annual evapotranspiration (mm)	1360	1480	1650	987	1520	1800
Mean annual rainfall (mm)	1830	3230	4120	1610	3090	4140
Median annual max. daily rainfall (mm)	45.3	64.4	94	37.7	62.2	112
t for annual max. daily rainfall	0.12	0.148	0.233	0.101	0.144	0.235
t_3 for annual max. daily rainfall	−0.0205	0.202	0.335	0.00875	0.193	0.403
Median annual max. consec. wet days	17	27.3	41	17	28	43
t for annual max. consec. wet days	0.139	0.195	0.258	0.137	0.197	0.248
t_3 for annual max. consec. wet days	−0.0374	0.212	0.391	0.0136	0.224	0.387
Median annual max. river flow, Q_{med} (m^3 s^{-1})	0.605	55.5	1340	13.6	94.9	1760
t for annual max. river flow, Q_t	0.128	0.241	0.461	0.14	0.327	0.548
t_3 for annual max. river flow, Q_{t_3}	0.054	0.25	0.687	0.0134	0.188	0.376

2.5 Weather data

Gridded daily mean temperature, wind speed, relative humidity and incoming shortwave radiation were acquired at 0.25° resolution from the AgMERRA data package (Ruane et al. 2015). These data were used to calculate reference crop potential evapotranspiration

Table 2 Aggregated land-use classification as compared to original land-use classification adopted by IMEF (2021)

Aggregated classification	Original classification
Water body	Water body
Built land	Built land
Dryland farming	Dryland farming
	Dryland farming mixed with shrubs
	Garden shrubs
	Mining area
	Open land
Wetland farming	Rice field
	Aquaculture pond
	Swamps
Plantation	Plantation
Forest	Primary dryland forest
	Secondary dryland forest
	Mangrove forest
	Industrial forest

according to FAO56 (Allen et al. 1998). Gridded daily precipitation data were acquired at 0.05° resolution from the CHIRPS data package (Funk et al. 2015).

The AgMERRA data package was chosen because it currently provides the most comprehensive gridded meteorological dataset (in terms of providing temperature, wind speed, humidity and shortwave radiation) for Southeast Asia. AgMERRA combines reanalysis data, gauged data and satellite date (Ruane et al. 2015). AgMERRA also provides precipitation data. However, the CHIRPS precipitation data package was chosen instead due to its higher spatial resolution. CHIRPS combines gauged data and satellite data (Funk et al. 2015) and has a significant track record of use in Indonesia (e.g., Narulita and Ningrum 2018; Auliyani and Wahyuningrum 2021; Wahyuni et al. 2021).

A map of mean annual rainfall for West Java is shown in Fig. 1b. Rainfall is not strongly controlled by topography and is mostly driven by monsoon and convection processes. Significantly elevated rainfall is observed on the West side, upstream of Jakarta, and this rainfall can be linked to the large cluster of reported flood hazards in Jakarta seen in Fig. 2a. Nevertheless, many reported hazards can also be seen around Bandung, where less rainfall is observed (see Fig. 2a).

2.6 River flow data

Observed daily mean river flow rate data was acquired for 70 river flow gauging stations, scattered across West Java, from the Center for Hydrology and Water Management, which is part of the Indonesia Ministry of Public Works (Balai Besar Wilayah Sungai). Additional details about measurement and quality protocol are presented and discussed by Yuningsih (2019).

The locations of each gauging station along with their associated catchment areas are shown in Fig. 1b (both the open and solid circular markers). Each gauge station record includes at least 10 complete (although not necessarily consecutive) years of daily mean flow rates from within the period of 1980 to 2016.

3 Methods

3.1 Flood frequency analysis

Flood frequency analysis provides a useful method for aggregating the daily river flow time-series into characteristic statistics that describe the extreme nature of the river flows in the gauged catchments.

Following common flood frequency analysis practice (e.g., Lim and Lye 2003; Kjeldsen and Jones 2006; Mulyantari et al. 2011; Badyalina et al. 2021), we determine the median annual maximum flow rates, Q_{med} (m^3 s^{-1}), along with associated t and t_3 sample L-moment ratios (see, Hosking and Wallis 1997, p.28), denoted hereafter as Q_t (-) and Q_{t_3} (-), respectively. The three river flow statistics, Q_{med}, Q_t and Q_{t_3} can be used to parametrise a generalised logistic distribution function, which describes the entire flood frequency curve of the designated river catchment (see, Kjeldsen and Jones 2006). Median values are preferred to mean values in this context to minimize the potential impact of outliers (Kjeldsen and Jones 2006).

3.2 Weather data statistics

Catchment averaged daily potential evapotranspiration and daily rainfall time-series were compiled for each gauged catchment and HydroSHEDS catchment studied. These were then used to derive corresponding values for mean annual potential evapotranspiration, mean annual rainfall, median annual maximum daily rainfall and median annual maximum number of consecutive rainfall days. The sample L-moment ratios, t (similar to coefficient of variation) and t_3 (similar to skewness) were also derived for both annual maximum daily rainfall and annual maximum number of consecutive wet days using equations provided by Hosking and Wallis (1997), p.28. The range and median for the above catchment characteristics are given for both the gauged catchments and the HydroSHEDS catchments in Table 1.

3.3 Step-wise linear regression

The analytical method described in Sect. 3.1 provides values for the river flow statistics, Q_{med}, Q_t and Q_{t_3}, for each of the 70 river flow gauging stations. However, to explore the importance of flood frequency data in determining whether or not a HydroSHED region reports flood disasters, we need to extrapolate the river flow statistics to the 154 level 9 HydroSHEDS sub-basins being studied. This can be achieved by using step-wise linear regression to derive regionalisation relationships, relating river flow statistics to relevant catchment characteristics for the gauged catchments.

Here we adopt a step-wise linear regression approach, previously applied by Mathias et al. (2016) and Ye et al. (2014), whereby additional parameters are incorporated until the so-called Bayesian Information Criterion (BIC) is minimised (i.e., Eq. (12) of Ye et al. (2014)). The procedure can be described as follows:

1. Determine the correlation coefficients of each catchment characteristic listed in Table 1 (with the exception of median annual maximum flow rate) with the flood statistic of concern (Q_{med}, Q_t or Q_{t_3}).

2. Select the catchment characteristic with the highest absolute correlation coefficient and check that the relationship passes a significance test (i.e., that the probability of getting a correlation as large as the observed value by random chance is < 0.05).
3. Develop a linear regression relationship between the selected catchment characteristic plus any previously selected catchment characteristics and the flood statistic of concern.
4. Calculate the BIC.
5. Determine the correlation coefficients of the remaining catchment characteristics with the residuals between the developing regionalization relationship and the flood statistic of concern.
6. Repeat Steps 2 to 4.
7. If the new BIC is less than the previous BIC and the significance test from Step 2 passes, repeat Steps 5 to 7, otherwise consider the previous form of the regionalization relationship to be optimal.

3.4 Kolmogorov Smirnov testing

We use two-sample Kolmogorov–Smirnov tests (KST) to test the null hypothesis that individual catchment characteristics, describing the catchment areas contributing to HydroSHEDS regions with and without reported flood disasters, are from the same statistical distribution. The KST also provides the Kolmogorov–Smirnov statistic (KSS), which measures the maximum absolute difference between the cumulative distribution functions (CDF) for the two data populations. The higher the KSS, the more different the CDFs.

When the KST fails, the difference between the CDFs, for the catchment characteristic under consideration, from the regions where flood disasters were reported and those where they were not, can be considered to be statistically significant. The KSS can then be used to measure the difference between the two CDFs, which represents a relative measure of parameter sensitivity.

We use the KST to identify statistically significant catchment characteristics, in the context of flood disaster occurrence. We then use the KSS to rank the statistically significant catchment characteristics in terms of their sensitivity (high KSS corresponds to high sensitivity) to identify the most important parameters.

4 Results

4.1 Flood frequency regionalisation relationships

4.1.1 Gauges with at least 10 years of data

The regionalisation relationships obtained by stepwise linear regression for the river flow statistics, Q_{med}, Q_t and Q_{t_3}, using river flow data from all 70 river flow gauging stations are listed as follows:

$$Q_{\mathrm{med}} = 0.02655 x_1^{1.078} \exp\left(0.03195 x_2 - 0.0006444 x_3\right) \tag{1}$$

$$Q_t = 0.3683 x_1^{-0.06852} \exp\left(0.9058 x_4\right) \tag{2}$$

$$Q_{t_3} = 0.2907 \exp\left(-0.03976 x_5\right) \tag{3}$$

where x_1 is catchment area (km^2), x_2 is median annual maximum daily rainfall (mm), x_3 is P10 elevation (mASL), x_4 is t_3 for annual maximum daily rainfall (-) and x_5 is percentage land covered by alluvial clays (%).

Plots of observed data against modelled data (using Eqs. (1) to (3)) for each of the 70 gauged stations studied are shown as red open markers in Fig. 4. The correlation between the modelled and observed data for Q_{med} is very high with a correlation coefficient of 0.944. However the correlations for the sample L-moment ratios, Q_t and Q_{t_3}, are much
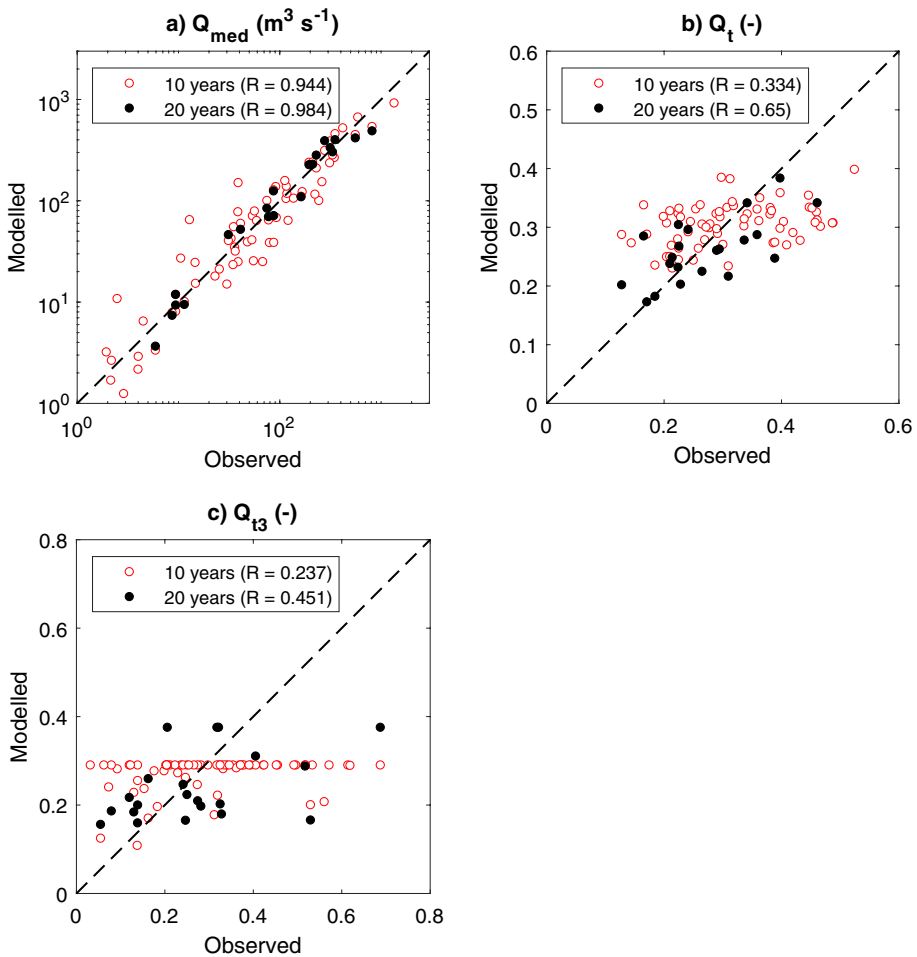
Fig. 4 Comparison of observed and modelled flood statistics. The red open circular markers denote results from Eqs. (1) to (3) based on data from the 70 river flow gauging stations comprising at least 10 full years of data. The black solid circular markers denote results from Eqs. (4) to (6) based on data from the 21 river flow gauging stations comprising at least 20 full years of data. The R values in the legends refer to the correlation coefficients comparing the modelled and observed data. **a** Results for median annual maximum flow rates, Q_{med} (m^3 s^{-1}). **b** Results for sample L-moment ratio, t, for annual maximum flow rates, Q_t (-). **c** Results for sample L-moment ratio, t_3, for annual maximum flow rates, Q_{t3} (-)

lower with correlation coefficients of 0.334 and 0.237, respectively, rendering these relationships of little value.

Note that each of the 70 river flow gauging stations had at least 10 complete years of daily river flow data. However, only 21 of the river flow gauging stations had more than 20 complete years of data. The reason for the higher order statistical relationships (Eqs. (2) to (3)) performing so poorly is arguably due to 10 years of data being insufficient to observe these kind of higher-order phenomena.

4.1.2 Gauges with at least 20 years of data

To gain further insight, we repeated the stepwise linear regression exercise using only river flow gauging stations that had at least 20 complete years of flow data. This included 21 gauging stations in total, the locations of which are shown as black solid circular markers on Fig. 1b. The resulting regression relationships are listed below:

$$Q_{\text{med}} = 0.006962 x_1^{1.129} \exp\left(0.04087 x_2 + 0.04726 x_6\right) \tag{4}$$

$$Q_t = 0.008772 x_2^{1.015} x_7^{0.6397} \tag{5}$$

$$Q_{t_3} = 0.3759 \exp\left(-0.03522 x_8\right) \tag{6}$$

where x_1 and x_2 are as before, x_6 is percentage land cover used by plantation (%), x_7 is catchment circularity (-) and x_8 is percentage land cover used by wetland farming (%).

Plots of observed data against modelled data (using Eqs. (4) to (6)) for these 21 gauging stations are shown as black solid circular markers in Fig. 4. The correlation coefficients comparing the modelled and observed data for all three flood statistics are much higher using Eqs. (4) to (6) as compared to when using Eqs. (1) to (3) (compare the correlation coefficients given in the legends for Fig. 4). To some extent this is because the models are being forced to predict less data. However, the marked improvement in correlation for the two sample L-moment ratios is arguably also due to the longer data records used. Also of interest is that whereas Eqs. (1) to (3) show no dependence on land-use, Eqs. (4) to (6) show dependence on plantation land cover and wetland farming land cover.

Nevertheless, Eq. (1) should be considered a more reliable method for estimating median annual maximum river flow rate for the HydroSHEDS catchments because it utilises a much larger set of observations. Range and median values for the river flow statistics, Q_{med}, Q_t and Q_{t_3}, are shown in Table 1. Values for the gauged catchments are based on observed values. Values for the HydroSHEDS catchments are based on Eqs. (1), (5) and (6).

4.2 Kolmogorov–Smirnov testing

4.2.1 All catchments

Table 3 shows the KSS values for all catchment characteristics for which the KST (as described in Sect. 3.4) failed when considering all 154 of the HydroSHEDS regions studied. Note that catchment characteristics for which the KST passed are not shown in Table 3,

because a passed KST implies that significant statistical differences are not observed when comparing regions with reported flood disasters and regions with none.

The parameters listed in Table 3 have been ranked in order of their KSS values. The higher the KSS value the more sensitive reported flood disaster occurrence is to that parameter. The most important parameter in this context turns out to be built land cover.

Also shown in Table 3 are rank correlation coefficients between each parameter and built land cover (RCCBL). Note that RCCBL is listed as zero where a correlation fails a significance test. Only catchment area, Latosols coverage, and median annual maximum river flow rate are found not to be strongly correlated with built land cover (i.e., with absolute values of RRCCBL < 0.30). CDFs, for both HydroSHEDS regions that reported flood disasters and those that did not, for these three parameters along with built land cover are shown in Fig. 5.

4.2.2 Rural catchments

In an attempt to take out the strong control of built land cover on flood disaster reporting, the KSTs were repeated using only HydroSHEDS regions contributed by catchments with less than 3% built land cover, a threshold commonly used to describe rural catchments
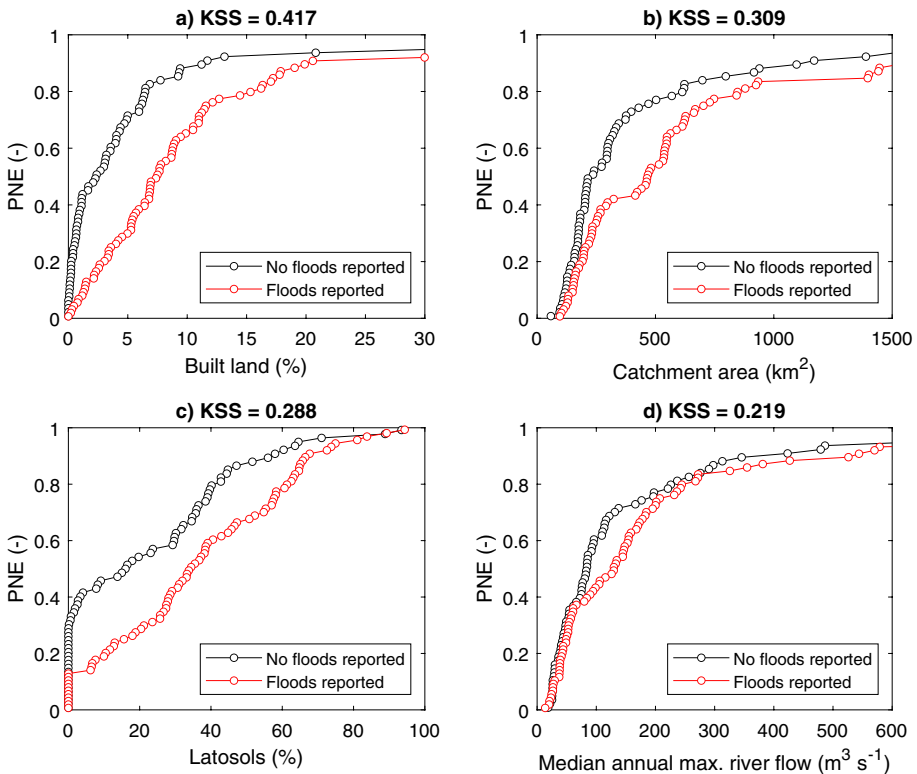

Fig. 5 Important CDFs identified by Kolmogorov–Smirnov tests using all the HydroSHEDS regions. PNE stands for probability of non-exceedance. KSS stands for Kolmogorov–Smirnov statistic

(e.g., MacDonald and Fraser 2014). Of the 154 HydroSHEDS regions studied, 55 satisfied this criterion.

Table 4 shows the KSS values for all catchment characteristics where the KST failed when considering the 55 HydroSHEDS regions considered to be draining rural catchments. The most important parameter in this context turns out to be catchment area followed by median annual maximum river flow rate. RCCBL values for the six identified parameters are also shown in Table 4, from which it can be seen that all of these (with the exception of built land cover) are not strongly correlated with built land cover. Corresponding CDFs for these six parameters are shown in Fig. 6.

5 Discussion

5.1 Flood frequency regionalisation relationships

Although the main objective of this article is to explore catchment controls on flood disaster reporting, the flood frequency regionalisation relationships (described in Sect. 4.1) also warrant further discussion.

Of the relationships obtained using all river flow gauges (i.e., Sect. 4.1.1), only Eq. (1) for the median annual maximum flow rate, Q_{med}, was found to have a reasonable correspondence with the observed data. The expressions for the L-moment ratios, Eqs. (2) and (3), should be dismissed because 10 years of river flow data is arguably inadequate to observe this kind of higher-order phenomena.

Equation (1) shows that Q_{med} increases almost linearly with catchment area, increases exponentially with median annual maximum daily rainfall and decreases exponentially with P10 elevation. This dependence on catchment area and rainfall is very common and is also observed in UK regionalisation relationships (Kjeldsen et al. 2008). The effect of P10 elevation is due to the fact that catchments with low P10 elevation are likely to have outlets closer to the coast and therefore represent some of the larger catchments within the West Java region.

The relationships obtained, using only river flow gauges with greater than 20 years of data (i.e., Eqs. (4), (5) and (6) in Sect. 4.1.2), exhibit much better correspondence with the observed data for all three of river flow statistics studied.

The alternative equation for Q_{med}, Eq. (4) is very similar to Eq. (1) except that P10 elevation is swapped out for plantation land cover. Yulianto et al. (2022) previously hypothesised that runoff coefficient for plantation land cover in West Java might be 40 times greater than that for primary forest. Plantation land cover ranges from 0 to 33.4% with a median of 1.74% for the gauged catchments and from 0 to 30.5% with a median of 2.00% for the HydroSHEDS catchments. Equation (4) suggests that reducing plantation land cover from 2 to 1% leads to a 4.6% reduction in median annual maximum river flow rate. Alternatively, reducing plantation land cover from 20 to 10% leads to a modelled 38% reduction in median annual maximum river flow rate.

Equation (5) for the L-moment ratio, Q_t, suggests that the coefficient of variation for annual maximum flow rate increases with increasing rainfall and increasing catchment circularity. Assuming spatially uniform rainfall, a more circular catchment will have more pathways of the same length and hence more water reaching the outlet at the same time, leading to higher peak flow rates.
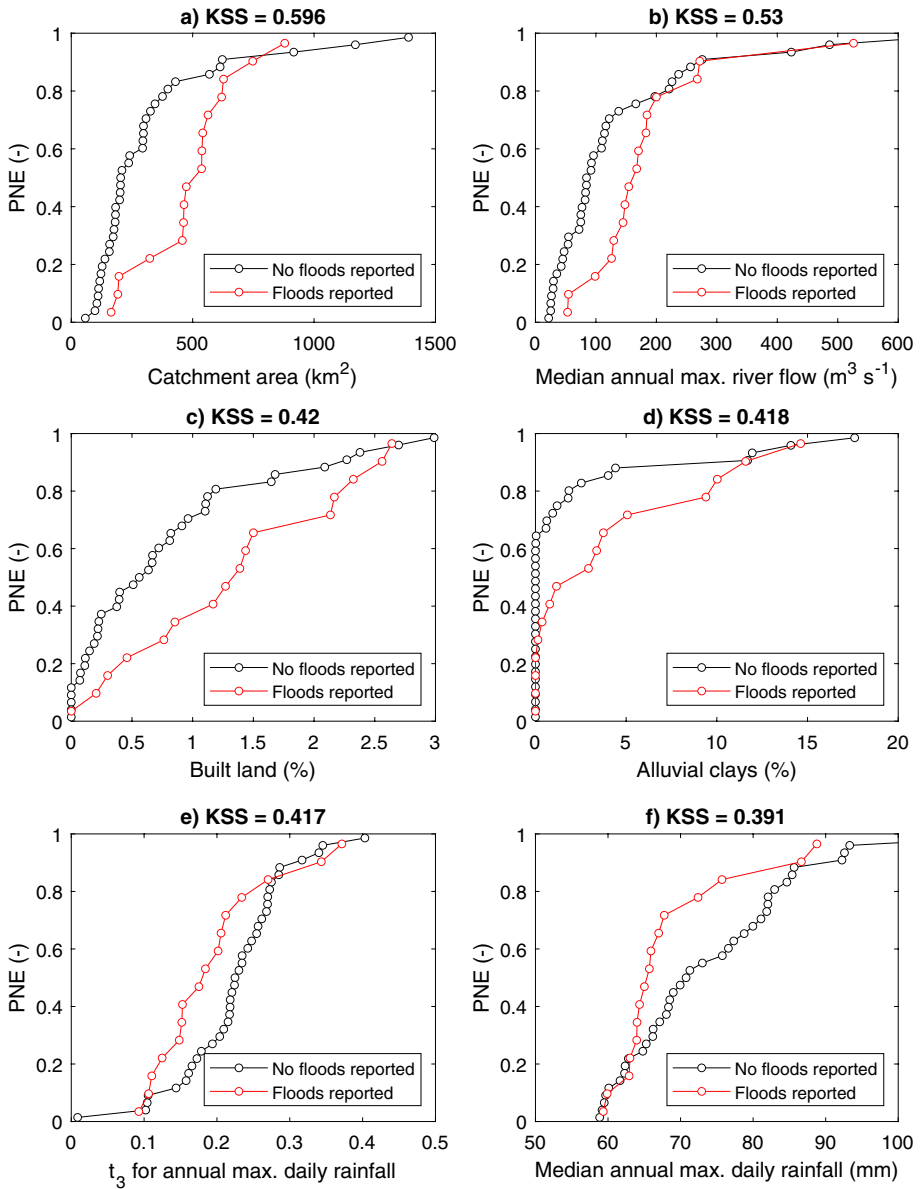
Fig. 6 Important CDFs identified by Kolmogorov–Smirnov tests using only the rural HydroSHEDS regions. PNE stands for probability of non-exceedance. KSS stands for Kolmogorov–Smirnov statistic

Equation (6) for the L-moment ratio, Q_{t3}, suggests that the skewness for annual maximum flow rates decreases with increasing wetland farming land cover. The impact of this effect is better understood by studying how Q_{t3} affects flood frequency.

Figure 7 shows plots of normalised annual maximum river flow rate (normalised by dividing by the median annual maximum river flow rate) against return period for two

Table 3 List of catchment characteristics that failed the Kolmogorov–Smirnov test, when studying all HydroSHEDS regions along with associated Kolmogorov–Smirnov statistics (KSS) and rank correlation coefficients with built land cover (RCCBL)

Catchment characteristic	KSS (-)	RCCBL (–)
Built land (%)	0.417	1.000
Soil complexes (%)	0.384	− 0.548
Alluvial clays (%)	0.317	0.495
Catchment area (km^2)	0.309	0.000
Regosols (%)	0.294	0.493
Latosols (%)	0.288	0.265
Alluvial soils (%)	0.271	0.400
P10 slope (m km^{-1})	0.262	− 0.575
Median annual max. daily rainfall (mm)	0.232	− 0.555
Median annual max. river flow (m^3 s^{-1})	0.219	0.000

That RCCBL has been set to zero where a correlation failed a significance test

Table 4 List of catchment characteristics that failed the Kolmogorov–Smirnov test, when studying only the rural HydroSHEDS regions along with associated Kolmogorov–Smirnov statistics (KSS) and rank correlation coefficients with built land cover (RCCBL)

Catchment characteristic	KSS (–)	RCCBL (–)
Catchment area (km^2)	0.596	0.000
Median annual max. river flow (m^3 s^{-1})	0.530	0.297
Built land (%)	0.420	1.000
Alluvial clays (%)	0.418	0.281
t_3 for annual max. daily rainfall	0.417	0.000
Median annual max. daily rainfall (mm)	0.391	0.000

That RCCBL has been set to zero where a correlation failed a significance test

catchments; one with 0% wetland farming land cover (river Cikapundung at Gandok) and one with 20% wetland farming land cover (river Cimanuk at Tomo). The circular markers are the observed data plotted using a Gringorten plotting position (Shaw 2005, p.314). The solid lines are median and L-moment matched generalised logistic distribution functions (see, Kjeldsen and Jones 2006). The black lines were obtained using the sample L-moments of the observed data. The coloured lines were obtained by determining the L-moments from Eqs. (5) and (6) using the associated catchment characteristics but with wetland farming land cover as specified in the legend.

It can be seen that increasing wetland farming land cover leads to a flattening of the flood frequency curve at higher return periods. An analogous effect was observed in the revitalised UK flood estimation handbook with flood plain extent (Kjeldsen et al. 2008, p.55). This could be because rice paddies are able to attenuate very large flood events better than other land cover types. The flood control value of rice paddies has long been recognised by ecosystem service studies (Natuhara 2013; Saputra and Setiyanto 2021). Rice paddies store rainfall and reduce peak river flows (Huang et al. 2006). They also benefit from being widely distributed across the landscape, offering widespread protection as compared to sparsely located dammed reservoirs (Kim et al. 2006).

Asdak and Supian (2018) previously proposed that maintaining wetland farming land cover in West Java is important to help avoid future increases in the magnitude and
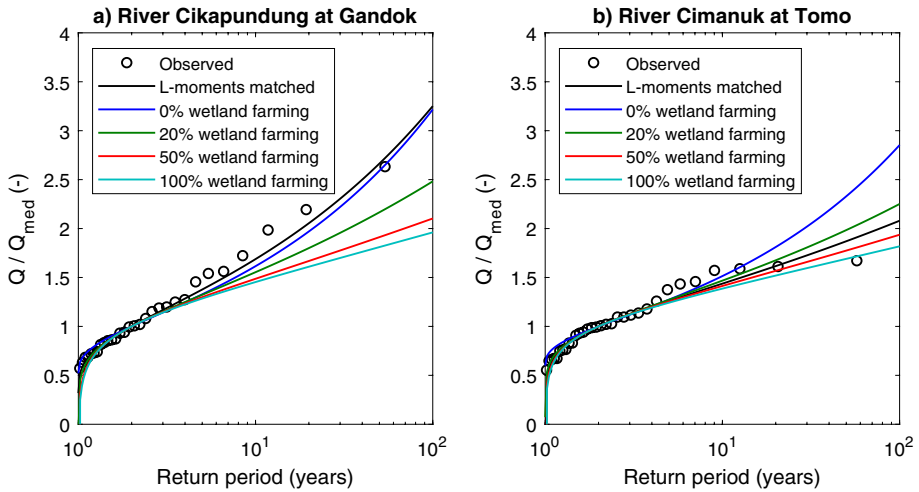
Fig. 7 Comparison of observed and modelled normalised annual maximum river flow rate plotted against return period for two different catchments. The circular markers are observed data plotted using the Gringorten plotting position. The black solid lines are generalised logistic distributions with their median value and L-moments matched to the observed data. The coloured lines are generalised logistic distributions, with t and t_3 obtained from Eqs. (5) and (6), respectively, using catchment characteristics but with wetland farming land cover specified as shown in the legend. **a** Shows data for the river Cikapundung at Gandok, which actually has 0% wetland farming land cover. **b** Shows data for the river Cimanuk at Tomo, which actually has 20% wetland farming land cover

frequency of flooding. Equation (6) provides important empirical evidence to support this point.

5.2 Kolmogorov–Smirnov testing

5.2.1 All catchments

Following removal of catchment characteristics with strong correlation with built land cover, Kolmogorov–Smirnov testing revealed that only built land, catchment area, Latsols cover and median annual maximum river flow exhibit statistically significant differences in CDFs for river catchments that report flood disasters.

The percentage of built land cover on catchments contributing to regions that report flood disasters is generally much larger (see Fig. 5a). A similar finding was observed in Central Java by Handayani et al. (2020). Built land cover is known to lead to higher surface runoff and hence higher river flows. However, this property was not observed in the two stepwise linear regression exercises described in the previous section. A more important factor is probably that the higher the percentage of built land cover, the more people available to experience and report flood disasters (van Dijk et al. 2009; Ferreira and Ghimire 2012).

The second most important parameter is catchment area. Regions that report flood disasters have statistically larger catchment areas (see Fig. 5b). This could be because river flow rate is strongly correlated with catchment area (consider Eqs. (1) and (4)).

Regions that report flood disasters also have statistically higher Latosols coverage (see Fig. 5c). Latosols have only a small correlation with built land cover (see Table 3). However, Latosols are the main soil-type underlying the highly urbanised region between Jakarta and Bogor (see Fig. 2b), which is strongly affected by flooding from the River Cili-wung (Asdak and Supian 2018).

The final parameter concerns median annual maximum river flow rate. However, regions that report flood disasters seem only to experience marginally higher flow rates (see Fig. 5d). Furthermore, median annual maximum river flow rate was calculated from Eq. (1), which has no dependence of land-use cover. Therefore, it can be concluded that land-use cover, with the exception of built land cover, seems not to have an impact on whether HydroSHEDS regions report flood disasters or not, a point supported by van Dijk et al. (2009) and Ferreira and Ghimire (2012).

5.2.2 Rural catchments

When we consider only rural catchments (with < 3% built land cover), built land cover is much less correlated with other important parameters identified by the Kolmogo-rov–Smirnov testing (see Table 4). Similar to for all the catchments, rural catchments that report flood disasters are found to have larger catchment areas, larger median annual maxi-mum river flow and more built land cover (see Figs. 6a and b).

Flood affected (rural) catchments are also found to have more alluvial clay cover and lower t_3 L-moment ratio and median value for annual maximum daily rainfall (see Figs. 6d–f). The importance of clay cover could be due to its low permeability and high runoff coefficient and/or that there is substantial clay cover around the Jakarta and Band-ung cities (see Fig. 2b). An explanation concerning the impact of rainfall statistics could be due to rural communities being more resilient where extreme rainfall is more common.

The general trend in parameters remains the same as when all the catchments are studied, with flood disaster reporting more likely in regions contributed by larger catch-ment areas, larger river flows and more built land. van Dijk et al. (2009) and Ferreira and Ghimire (2012) found that country-scale flood reporting data was strongly correlated with population density (a good proxy for built land cover) and inadequate for investigating the impacts of land-use cover on flood disaster frequency. Bhattacharjee and Behera (2018) and Tembata et al. (2020) observed some success at linking flood disaster reporting to for-est cover at a district scale. However, our findings further support the assertion of van Dijk et al. (2009) and Ferreira and Ghimire (2012), even at a sub-catchment scale.

6 Summary and conclusions

The objective of this article was to determine significant statistical differences, in terms of topography, land-use, soil-type and hydro-meteorology, between river catchments in West Java that reported suffering from flood disasters and those that did not.

Regional analysis using data from river flow gauging stations (focusing on the 21 gauge stations with at least 20 complete years of data) led to stepwise linear regression relation-ships for median and sample L-moment ratios t (similar to coefficient of variation) and t_3 (similar to similar to skewness) of the annual maximum daily mean river flow rate. Median annual maximum flow rates were found to be positively correlated with catchment area, rainfall and plantation cover. Reducing plantation land cover from 20 to 10% was found to

lead to a modelled 38% reduction in median annual maximum river flow rate. This is presumed to be because of the increased runoff coefficient associated with converting forest to plantation.

The t_3 sample L-moment ratios were found to be negatively correlated with wetland farming land cover. Using a generalised logistic function it was shown that increasing wetland farming land cover reduced the extent to which annual maximum flow rates escalate with extreme return periods (i.e., greater than 10 years). This suggests that rice paddies play an important role in attenuating extreme river flow events.

A set of Kolmogorov–Smirnov tests (KST) were performed on 34 topographic, land-use and hydro-meteorological catchment characteristics looking at 154 HydroSHEDS regions across West Java. The aim was to look for statistical differences associated with regions that reported flood disasters during the period from 2009 to 2013 and those that did not. Built land cover was found to be the main factor determining whether a region was likely to report flood disasters. Built land cover was also found to be a significant factor in rural catchments (i.e., with less than 3% built land). Interestingly, built land was not identified as a contributing factor in the aforementioned flood frequency regionalisation study. This would suggest that the importance of built land is more about the fact that the higher the percentage of built land cover, the more people available to experience and report flood disasters.

The KSTs did not identify any other land use factors although annual maximum river flows were found to be slightly larger in flood disaster affected areas, which in turn are partially controlled by plantation and wetland farming cover.

Despite flood frequency (in terms of river flow rate) being strongly controlled by plantation cover and wetland farming cover, our research shows that percentage built land cover is the most important factor defining whether or not an area is likely to report flood disasters in West Java. Our findings also suggest that more research is needed to understand the important role of plantation cover in aggravating median annual maximum river flow rates and wetland farming cover in mitigating extreme river flow events.

Declarations

Conflict of interest The authors have no relevant financial or non-financial interests to disclose.

References

Allen RG, Pereira LS, Raes D, Smith M (1998) Crop evapotranspiration (guidelines for computing crop water requirements), FAO Irrigation and Drainage paper 56. Food and Agriculture Organization of the United Nations

Asdak C, Supian S (2018) Watershed management strategies for flood mitigation: acase study of Jakarta's flooding. Weather and climate extremes 21:117–122

Auliyani D, Wahyuningrum N (2021) Rainfall variability based on the climate hazards group infrared precipitation with station data (CHIRPS) in lesti watershed, Java Island, Indonesia. In: IOP conference series: earth and environmental science Vol 874. No 1. IOP Publishing, p. 012003

Badyalina B, Mokhtar NA, Jan NAM, Hassim NH, Yusop H (2021) Flood frequency analysis using L-moment for Segamat river. Matematika 37(2):47–62

Beven K (2006) A manifesto for the equifinality thesis. J Hydrol 320(1–2):18–36

Beven K, Binley A (1992) The future of distributed models: model calibration and uncertainty prediction. Hydrol Process 6(3):279–298

Bhattacharjee K, Behera B (2018) Does forest cover help prevent flood damage? Empirical evidence from India. Glob Environ Chang 53:78–89

BNPB (2021) Data Informasi Bencana Indonesia. Bidang Pengelolaan Data dan Sistem Informasi (PDSI), Pusat Data Informasi dan Komunikasi Kebencanaan (Pusdatinkom), Badan Nasional Penanggulangan Bencana (BNPB). Indonesia Disaster Management Agency. https://dibi.bnpb.go.id/xdibi Accessed 25 Mar 2022

Bradshaw CJ, Sodhi NS, Peh KSH, Brook BW (2007) Global evidence that deforestation amplifies flood risk and severity in the developing world. Glob Chang Biol 13(11):2379–2395

Brakenridge GR (2010) Global active archive of large flood events. Dartmouth flood observatory, University of Colorado, USA. https://floodobservatory.colorado.edu/Archives/index.html Accessed 12 Sep 2022

Di Gregorio A, Jansen LJM (1998) Land cover classification system (LCCS): classification concepts and user manual. In: Environment and natural resources service, GCP/RAF/287/ITA Africover - East Africa project and soil resources, management and conservation service. FAO, Rome

Ekblad L, Herman JD (2021) Toward data-driven generation and evaluation of model structure for integrated representations of human behavior in water resources systems. Water Resour Res 57(2):e2020WR028148

Fatimah IN, Iskandar J, Partasasmita R (2020) Ethnoecology of paddy-fish integrative farming (minapadi) in Lampegan village, West Java, Indonesia. Biodivers J Biol Divers 21(9):4419–4432

Ferreira S, Ghimire R (2012) Forest cover, socioeconomics, and reported flood frequency in developing countries. Water Resour Res 48(8):W08529

Funk C, Peterson P, Landsfeld M, Pedreros D, Verdin J, Shukla S, Michaelsen J (2015) The climate hazards infrared precipitation with stations-a new environmental record for monitoring extremes. Sci Data 2(1):150066

Handayani W, Chigbu UE, Rudiarto I, Putri IHS (2020) Urbanization and increasing flood risk in the northern coast of central Java-Indonesia: an assessment towards better land use policy and flood management. Land 9(10):343

Hosking JRM, Wallis JR (1997) Regional frequency analysis: an approach based on L-moments. Cambridge University Press, New York

Huang CC, Tsai MH, Lin WT, Ho YF, Tan CH (2006) Multifunctionality of paddy fields in Taiwan. Paddy Water Environ 4(4):199–204

IMEF (2020). Akurasi Data Penutupan Lahan Nasional Tahun 1990-2016. Direkorat Inventarisasi dan Pemantauan Sumber Daya Hutan. Directorate General of Forestry Planning and Environmental Management. Kementerian Lingkungan Hidup dan Kehutanan. Indonesia Ministry of Environment and Forestry (IMEF). Jakarta. https://sigap.menlhk.go.id/sigap-trial/files/download/akurasi-data-penutupan-lahan-nasional-tahun-1990-2016.pdf Accessed 21 Nov 2022

IMEF (2021). WebGIS Kementerian Lingkungan Hidup dan Kehutanan. Indonesia Ministry of Environment and Forestry (IMEF). https://sigap.menlhk.go.id/sigap/ Accessed 21 Nov 2022

Jenson SK, Domingue JO (1988) Extracting topographic structure from digital elevation data for geographic information system analysis. Photogrammetric engineering and remote sensing 54(11):1593–1600

Julian Poerbandono MM, Ward PJ (2014) Assessment of the effects of climate and land cover changes on river discharge and sediment yield, and an adaptive spatial planning in the Jakarta region. Nat Hazards 73(2):507–530

Kim J, Gim THT (2020) Assessment of social vulnerability to floods on Java, Indonesia. Nat Hazards 102(1):101–114

Kim TC, Gim US, Kim JS, Kim DS (2006) The multi-functionality of paddy farming in Korea. Paddy Water Environ 4(4):169–179

Kjeldsen TR, Jones DA, & Bayliss AC (2008) Improving the FEH statistical procedures for flood frequency estimation. Environment Agency. https://assets.publishing.service.gov.uk/media/602e5c0f8fa8f54331b080e6/Improving_the_FEH_Statistical_Procedures_for_Flood_Frequency_Estimation_Technical_Report.pdf Accessed 07 Sep 2022

Kjeldsen TR, Jones DA (2006) Prediction uncertainty in a median-based index flood method using L moments. Water Resour Res 42:W07414

Lehner BH (2014) Global watershed boundaries and sub-basin delineations derived from HydroSHEDS data at 15 second resolution. Technical Documentation Version 1.c (with and without inserted lakes)

Lehner B, Verdin K, Jarvis A (2008) New global hydrography derived from spaceborne elevation data. Eos, Trans Am Geophys Union 89(10):93–94

Lehner B, Grill G (2013) Global river hydrography and network routing: baseline data and new approaches to study the world's large river systems. Hydrol Process 27(15):2171–2186

Lim YH, Lye LM (2003) Regional flood estimation for ungauged basins in Sarawak, Malaysia. Hydrol Sci J 48(1):79–94

Linke S, Lehner B, Ouellet Dallaire C, Ariwi J, Grill G, Anand M, Thieme M (2019) Global hydro-environmental sub-basin and river reach characteristics at high spatial resolution. Sci Data 6(1):1–15

MacDonald DE, Fraser RJ (2014) An improved method for estimating the median annual flood for small ungauged catchments in the United Kingdom. J Flood Risk Manag 7(3):251–264

Marfai MA, Sekaranom AB, Ward P (2015) Community responses and adaptation strategies toward flood hazard in Jakarta, Indonesia. Nat Hazards 75(2):1127–1144

Mathias SA, McIntyre N, Oughton RH (2016) A study of non-linearity in rainfall-runoff response using 120 UK catchments. J Hydrol 540:423–436

Menendez P, Losada IJ, Beck MW, Torres-Ortega S, Espejo A, Narayan S, Lange GM (2018) Valuing the protection services of mangroves at national scale: the Philippines. Ecosyst Serv 34:24–36

Morán-Tejeda E, Zabalza J, Rahman K, Gago-Silva A, López-Moreno JI, Vicente-Serrano S, Beniston M (2015) Hydrological impacts of climate and land-use changes in a mountain watershed: uncertainty estimation based on model comparison. Ecohydrology 8(8):1396–1416

Mulyantari F, Sutanto SJ, Asyantina T (2011) Design flood formula development in ungauged catchments, West Java Indonesia: index flood and L-moment approach. Jurnal Teknik Hidraulik 2(2):141–152

Narulita I, Ningrum W (2018) Extreme flood event analysis in Indonesia based on rainfall intensity and recharge capacity. In: IOP conference series: earth and environmental science Vol. 118, No. 1. IOP Publishing, p 012045

Natuhara Y (2013) Ecosystem services by paddy fields as substitutes of natural wetlands in Japan. Ecolog Eng 56:97–106

Nurhayati A, Lili W, Herawati T, Riyantini I (2016) Derivatif analysis of economic and social aspect of added value minapadi (paddy-fish integrative farming) a case study in the village of Sagaracipta Ciparay sub district, Bandung West Java Province, Indonesia. Aquatic Procedia 7:12–18

Ridwansyah I, Yulianti M, Onodera SI, Shimizu Y, Wibowo H, Fakhrudin M (2020) The impact of land use and climate change on surface runoff and groundwater in Cimanuk watershed, Indonesia. Limnology 21(3):487–498

Ruane AC, Goldberg R, Chryssanthacopoulos J (2015) Climate forcing datasets for agricultural modeling: merged products for gap-filling and historical climate series estimation. Agric For Meteorol 200:233–248

Saputra YH, Setiyanto A (2021) Assessment of the multifunctional role of wetlands in Indonesia: a case study in West Java Province. In: IOP conference series: earth and environmental science Vol. 892, No. 1. IOP Publishing, p 012048

Shaw EM (2005) Hydrology in Practice, 3rd edn. Taylor & Francis

Siswanto SY, Francés F (2019) How land use/land cover changes can affect water, flooding and sedimentation in a tropical watershed: a case study using distributed modeling in the upper Citarum watershed, Indonesia. Environ Earth Sci 78(17):1–15

Suhardjo H, Soepraptohardjo M (1982) Indonesian soil units and subunits for survey and mapping of transmigration areas. Wageningen Agricultural University. https://library.wur.nl/WebQuery/wurpubs/fulltext/304805 Accessed 13 Sep 2022

Suprayogo D, van Noordwijk M, Hairiah K, Meilasari N, Rabbani AL, Ishaq RM, Widianto W (2020) Infiltration-friendly agroforestry land uses on volcanic slopes in the Rejoso Watershed, East Java, Indonesia. Land 9(8):240

Tembata K, Yamamoto Y, Yamamoto M, Matsumoto KI (2020) Don't rely too much on trees: evidence from flood mitigation in China. Sci Total Environ 732:13841138410

van Dijk AI, Van Noordwijk M, Calder IR, Bruijnzeel SL, Schellekens JAAP, Chappell NA (2009) Forest-flood relation still tenuous-comment on "Global evidence that deforestation amplifies flood risk and severity in the developing world'' by CJA Bradshaw, NS Sodi, KS-H. Peh and BW Brook. Glob Chang Biol 15(1):110–115

Verburg PH, Bouma J (1999) Land use change under conditions of high population pressure: the case of Java. Glob Environ Chang 9(4):303–312

Wahyuni S, Sisinggih D, Dewi IAG (2021) Validation of climate hazard group infrared precipitation with station (CHIRPS) data in wonorejo reservoir, Indonesia. In: IOP conference series: earth and environmental science Vol. 930, No. 1. IOP Publishing, p. 012042

Xiao L, Robinson M, O'Connor M (2022) Woodland's role in natural flood management: evidence from catchment studies in Britain and Ireland. Sci Total Environ 813:151877

Ye S, Li HY, Huang M, Ali M, Leng G, Leung LR, Sivapalan M (2014) Regionalization of subsurface stormflow parameters of hydrologic models: derivation from regional analysis of streamflow recession curves. J Hydrol 519:670–682

Yulianto F, Sofan P, Zubaidah A, Sukowati KAD, Pasaribu JM, Khomarudin MR (2015) Detecting areas affected by flood using multi-temporal ALOS PALSAR remotely sensed data in Karawang, West Java, Indonesia. Nat Hazards 77(2):959–985

Yulianto F, Khomarudin MR, Hermawan E, Nugroho NP, Chulafak GA, Nugroho G, Priyanto E (2022) Spatial and temporal distribution of estimated surface runoff caused by land use/land cover changes in the upstream citarum watershed, West Java, Indonesia. J Degrad Min Lands Manag 9(2):3293–3305

Yuningsih SM (2019) Kondisi Kualitas Data Debit Sungai Tahun 2015–2016 di Indonesia. J Sumber Daya Air 15(1):39–54

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Authors and Affiliations

Rahmawati Rahayu[1] · Simon A. Mathias[2] · Sim Reaney[3,4] · Gianni Vesuviano[5] · Rusmawan Suwarman[6] · Agus M. Ramdhan[7]

Rahmawati Rahayu
rahmawati.rahayu@durham.ac.uk

Sim Reaney
sim.reaney@durham.ac.uk

Gianni Vesuviano
giaves@ceh.ac.uk

Rusmawan Suwarman
rusmawan@meteo.itb.ac.id

Agus M. Ramdhan
agusmr@gl.itb.ac.id

[1] Department of Earth Sciences, Durham University, Durham, UK

[2] Department of Engineering, Durham University, Durham, UK

[3] Institute of Hazard, Risk and Resilience, Durham University, Durham, UK

[4] Department of Geography, Durham University, Durham, UK

[5] UK Centre for Ecology and Hydrology, Wallingford, UK

[6] Department of Meteorology, Institut Teknologi Bandung, Bandung, Indonesia

[7] Department of Geology, Institut Teknologi Bandung, Bandung, Indonesia