

Research article

# Flood Risk Analysis Using Spatial Synthetic Population in the Upper Bengawan Solo Watershed, Indonesia

Muhammad Musiyam<sup>1,\*</sup>, Jumadi Jumadi<sup>2,3</sup>, Vidya Nahdhiyatul Fikriyah<sup>2,4</sup>, Heni Masruroh<sup>5</sup>, Ema Dwi Septiyani<sup>2</sup>, Choirul Amin<sup>2</sup>, Hamim Zaky Hadibasyir<sup>2,6</sup>, Farha Sattar<sup>7</sup>, Muhammad Nawaz<sup>8</sup>

<sup>1</sup> Department of Geography Education, Faculty of Education, Universitas Muhammadiyah Surakarta, Surakarta 57162; <sup>2</sup> Department of Geography, Faculty of Geography, Universitas Muhammadiyah Surakarta, Indonesia; <sup>3</sup> INTI International University, 71800 Nilai, Negeri Sembilan, Malaysia; <sup>4</sup> Faculty of Geo information Science and Earth Observation, University of Twente, Netherlands; <sup>5</sup> Department of Geography, Universitas Negeri Malang, Indonesia; <sup>6</sup> School of Built Environment, University of New South Wales, Sydney, New South Wales 2052, Australia; <sup>7</sup> Faculty of Arts and Society, Charles Darwin University, Australia; <sup>8</sup> Department of Geography, National University of Singapore, AS#3-03-11, 1 Arts Link, Kent Ridge, Singapore.

\*Correspondence: mm102@ums.ac.id

Citation:

Musiyam, M., Jumadi, J., Fikriyah, V. N., Masruroh, H., Septiyani, E. D., Amin, H., Hadibasyir, H. Z., Sattar, F., & Nawaz, M. (2025). Flood Risk Analysis Using Spatial Synthetic Population in the Upper Bengawan Solo Watershed, Indonesia. *Forum Geografi*. 39(3), 383-397.

Article history:

Received: 22 November 2025  
Revised: 4 December 2025  
Accepted: 4 December 2025  
Published: 10 December 2025

## Abstract

This study develops a spatial synthetic population (SSP)-based computational model to produce realistic, high-resolution flood-risk maps for the Upper Bengawan Solo Watershed. It combines Global Human Settlement (spatial distribution) with local population statistics (attributes). The SSP is created for flood risk mapping in the Upper Bengawan Solo Watershed (BSH) using a 100 m grid from the Global Human Settlement Layer (GHSL) GHS-POP R2023A. Synthetic individuals are strategically placed around the pixel centre (radius  $\leq 100$  m), and each is assigned demographic attributes (age, gender, education, occupation) validated against official county-level data. Social vulnerability is calculated through weighted aggregation (AHP) across four attributes; individual scores are combined with flood hazard intensity at each location to produce a risk index for each person. Validation shows that (i) the SSP aligns closely with reference statistics: gender and age are nearly identical (MAE  $\approx 0.01-0.02\%$ ), with slight deviations in occupation (MAE 6.52%) and education (MAE 4.89%), (ii) the overall suitability of the SSP compared to GHS counts at pixel samples, and (iii) location plausibility testing using ESRI Sentinel-2 Land Cover (10 m). Results indicate that (i) the SSP aligns well for gender, moderately for education and occupation, but shows significant misalignment in age, (ii) 91.96% of SSP points are in built-up land, suggesting high spatial accuracy. Medium- to high-risk patterns are mainly along the main river corridors and peri-urban areas, while rural non-built zones are mostly low- to medium-risk. These findings suggest that this methodology is scalable, reproducible, and suitable for data limited regions, enabling the production of detail risk maps that can guide mitigation and preparedness efforts.

Keywords: Spatial synthetic population; Global Human Settlement Layer; Flood risk; Upper Bengawan Solo; AHP; Social vulnerability; Flood mitigation.

## 1. Introduction

Floods are the most common hydrometeorological disasters in Indonesia and have significant socio-economic impacts, especially in fast-growing urban and rural areas such as the Upper Bengawan Solo River Basin (Jumadi *et al.*, 2024a; Jumadi *et al.*, 2024b; Purwanto *et al.*, 2023; Sahid, 2024; Tellman *et al.*, 2021). Accurate flood risk mapping requires integrating hazards and vulnerabilities; however, the quality of mapping is often limited by the lack of detailed, population-based raster methods—such as medium-resolution global datasets—help estimate exposure but usually overlook demographic structures and household traits that influence resilience and recovery after disasters. This is where spatial synthetic population (SSP) becomes a breakthrough: it assigns realistic attributes and locations to individual and household units without jeopardizing privacy, thus connecting aggregate data with detailed demographic information for better risk assessment (Agriesti *et al.*, 2022; Chapis & Taillandier, 2019; Jiang *et al.*, 2024; Jumadi *et al.*, 2019; Prédhumeau & Manley, 2023).

Recent research shows rapid progress in creating SSPs at scales from urban to national, emphasizing reproducibility, data openness, and computational efficiency. Earlier studies have focused on building national synthetic populations, highlighting transparent pipelines and statistical alignment between synthetic data and census figures (Prédhumeau & Manley, 2023). Similar endeavours at a larger scale demonstrate that assembling a geographically explicit synthetic population (SSP) is achievable by combining microdata, spatial constraints, and realistic housing placement strategies (Jiang *et al.*, 2024). In mobility planning and public services, assignment and placement techniques for SSPs underscore the importance of matching individual attributes, household structures, and spatial activity patterns to ensure population representations remain accurate and practical (Agriesti *et al.*, 2022). Methodologically, a brief review of synthetic population creation methods identifies Iterative Proportional Fitting/Updating (IPF/IPU), microsimulation, and



Copyright: © 2025 by the authors. Submitted for possible open access publication under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

machine learning as the primary approaches, each suited to varying data availability and research goals (Balac & Hörl, 2021; Bigi *et al.*, 2024; Burger *et al.*, 2017; De Mooij *et al.*, 2024; Jiang *et al.*, 2022; Jumadi *et al.*, 2018, 2019).

In the context of hydrometeorological disasters, the main challenge is not only in representing the number of exposed individuals but also in accounting for their demographic structure, employment, and socio-economic conditions, which influence vulnerability. Many studies use social vulnerability indices (e.g., derivatives of SoVI) to summarize sensitivity and adaptive capacity at the community level; however, at the village level in developing countries, data gaps are common, requiring the synthesis of indicators from incomplete distributions (Chakraborty *et al.*, 2023; Cutter *et al.*, 2003; Cutter & Finch, 2008; Jumadi *et al.*, 2020). At the same time, flood hazard information can be analyzed based on several parameters by combining indicators such as slope gradient, rainfall, drainage density, soil moisture, land use/land cover (LULC), elevation, distance to rivers, normalized difference vegetation index (NDVI), curvature, flow accumulation, and topographic wetness index (TWI); these indicators can be calibrated with historical event records and/or remotely sensed rainfall if available (Jumadi *et al.*, 2024a; Jumadi *et al.*, 2024b; Musiyam *et al.*, 2020).

Although global population grid maps support macro-level analysis, several studies show that flood exposure results are sensitive to the choice of population datasets and their distribution schemes. These findings highlight the importance of methodological transparency in selecting and creating population representations, as minor differences in inputs can significantly affect risk estimates (Abdelkareem & Mansour, 2023; Arnell & Gosling, 2016; Chakraborty *et al.*, 2023; Diriba *et al.*, 2024; Jumadi, *et al.*, 2024). This methodological awareness is especially relevant for floods, as demographic structures and settlement patterns can determine who is exposed to flooding, where flooding occurs, and its severity.

In Thailand, for instance, flood risk assessments have become increasingly important as the population ages rapidly (Waiyasuri *et al.*, 2021; Sawangnate *et al.*, 2022; Waiyasuri *et al.*, 2023). On the other hand, informal settlements and regional terrain variations pose a high risk of flooding in the Philippines (Usamah *et al.*, 2014). Similar challenges are also evident in other coastal regions, such as peninsular Malaysia (Maqtan *et al.*, 2025), highlighting the need to consider population context in flood risk mapping.

In Indonesia, including the Upper Bengawan Solo River Basin, there are practical gaps in three areas. First, the lack of microdata at the village level limits the use of IPF/IPU; as a result, many studies depend on an aggregate population raster that may be biased in areas with mixed settlements and open land. Second, the spatial distribution of populations often ignores the actual settlement structure (building footprints and road networks), so exposure at the unit level (buildings and households) is not accurately reflected. Third, combining SSP with composite multi indicator hazard and social vulnerability indices in a reproducible format is not yet a common practice at the village level, making cross regional and temporal comparison difficult (Agriesti *et al.*, 2022; Chapuis & Taillandier, 2019; Jiang *et al.*, 2024; Prédhumeau & Manley, 2023).

This research addresses the gap identified above by developing an SSP model that utilizes the GHS-POP R2023A (Pesaresi *et al.*, 2024; Schiavina *et al.*, 2023) using lightweight computing. It incorporates directed stochastic allocation around settlement centers (based on GHS-POP pixels) to match local spatial patterns. Population attributes such as age, gender, education, occupation, and income are synthesized from official frequency distributions, with their margins enforced for consistency across districts. Flood hazard is modelled using eleven measurable risk factors (slope, rainfall, drainage density, soil moisture, land use/land cover, elevation, distance to river, NDVI, curvature, flow accumulation, TWI) through ordinal reclassification and normalized weighted overlay.

Meanwhile, social vulnerability is summarized into a composite index based on tested demographic, social, and economic indicators. Risk is then calculated as a composite or multiplicative function of hazard and vulnerability at an operational resolution. This framework is built on best practices for SSP generation—ensuring transparency and evaluating statistical congruence—while remaining adaptable to regions with data limitations. This design produces an individual-based flood risk map that more accurately reflects local demographic and socio-economic conditions, making it ready for stakeholders in the Upper Bengawan Solo region to prioritize mitigation and risk communication. Finally, the rest of the paper is organized as follows: the Research Methods section explains the construction of the SSP, the composite hazard and vulnerability, and the validation processes. The Results and Discussion section displays the risk maps and validation outcomes. The Conclusion section discusses contributions, limitations, and future research directions.

## 2. Research Methods

The study develops SSP modelling, which uses demographic data, gender, education, and occupation, with randomized spatial allocation centered on GHS-POP pixels to determine individual distributions. The accuracy of the SSP is validated against official statistics (using MAE and RMSE metrics), spatial fit with population grids, and placement tests on built-up land based on ESRI Sentinel 2 Land Cover. The social vulnerability index (V) is calculated as a linear weighted sum using AHP. Flood risk (R) is derived by multiplying the standardized hazard (H) by V, then classified into five categories using natural breaks (Jenks). The Getis-Ord Gi hotspot analysis identifies significant risk clusters, which are then interpreted in relation to river corridors and urban areas. All steps are performed in a geographic information system (GIS) environment with a reproducible workflow (Figure 1).

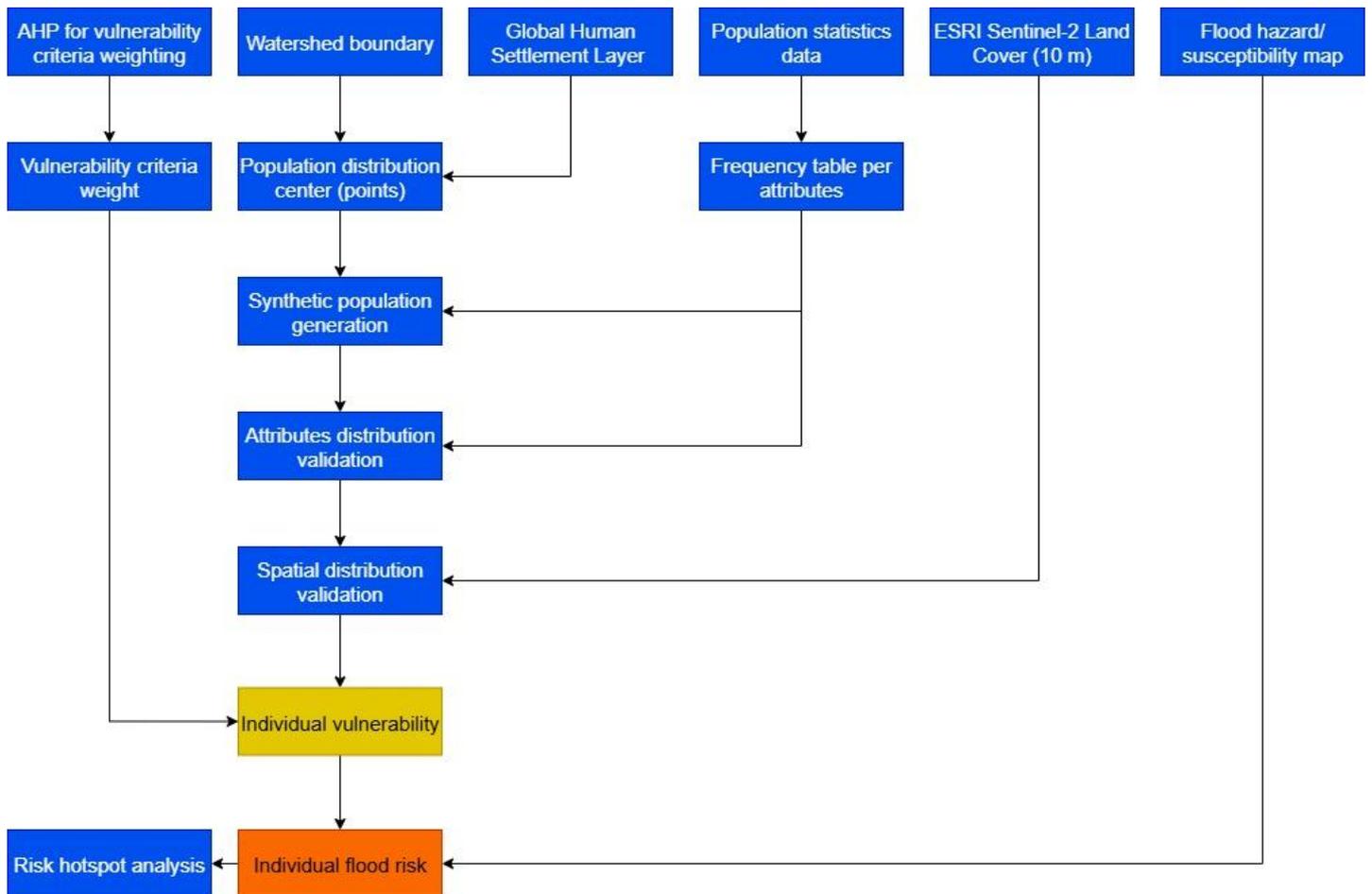


Figure 1. Research Framework.

### 2.1. Study Area

The study area is the Upper Bengawan Solo Watershed (BSH), the upstream part of the Bengawan Solo system, the largest river basin on Java Island (Figure 2). Geographically, this area is located in Central Java and covers the upstream region, supplied by the Merapi–Merbabu volcanic complex to the west and the Lawu to the east. Generally, the area's terrain is dominated by relatively flat plains; however, the northeastern and northwestern sections feature undulating to rolling terrain near the mountain foothills. The hydrological segmentation classifies BSH as one of the three main sub-watersheds within the Bengawan Solo system. The study boundaries are defined by hydrological (watershed) boundaries based on the digital elevation model (DEM) and align with administrative borders (Absori *et al.*, 2023; Anna *et al.*, 2023; Anna & Priyana, 2015; Santhyami *et al.*, 2025).

The topographic conditions and river networks primarily influence flood dynamics in BSH. Past studies have shown varying vulnerability to flooding: areas around Surakarta (the urban core) tend to be more susceptible, while certain regions in Wonogiri, Karanganyar, and Boyolali are comparatively less at risk. This pattern aligns with the morphological gradient (flat versus undulating or steep terrain), drainage density, and land use/cover structure along the main river corridor. Variations in vulnerability among districts highlight the need for locally tailored management,

especially during the rainy season when increased rainfall strains urban drainage systems (Jumadi *et al.*, 2025). Administratively, BSH includes at least six districts/cities: Boyolali, Surakarta, Karanganyar, Sukoharjo, Klaten, and Wonogiri. The Surakarta urban area acts as a hub for population and economic activity, leading to higher exposure of residents and property in this region (Santhyami *et al.*, 2025).

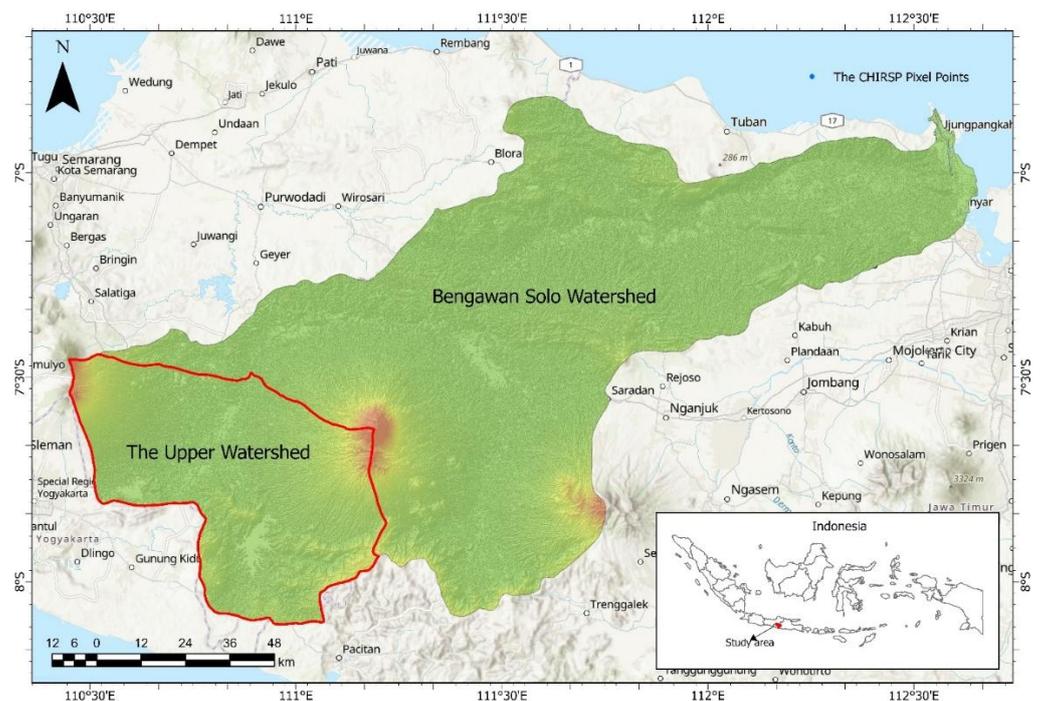


Figure 2. Study Area.

## 2.2. Data and Sources

### 1. Population Centers (GHS-POP R2023A)

The population center dataset is derived from the latest Global Human Settlement Layer (GHSL) release (GHS global population grid multitemporal, GHS-POP R2023A) (Pesaresi *et al.*, 2024; Schiavina *et al.*, 2023). The population pixel values from GHSL, typically at 100 m resolution, were converted into point centers using Raster to Point in ArcGIS Pro. Each point retains a center\_id, pixel center coordinates, and population count (the number of residents in that pixel). These points are used as references to generate synthetic individuals within a 100 m radius.

### 2. Attribute Frequency Tables (Age, Gender, Education, Occupation)

Four tables display the proportions of demographic and social categories at the watershed level, which combines four districts or cities, for age, gender, education, and occupation. For age, the labels are given as age ranges; during synthesis, these are converted to specific ages by sampling integers within each range. For gender, education, and occupation, category selection is weighted by frequency. Substantive rules are applied to maintain social realism, such as automatically assigning individuals under 15 years old to the occupation 0 (no occupation).

### 3. Flood Hazard Data Based on RS–GIS

The hazard index (Figure 3) is compiled from physical-environmental parameters derived from open-access spatial data (Jumadi *et al.*, 2024a; Jumadi *et al.*, 2024b). The core framework uses eleven parameters: elevation (EI), slope (SI), flow accumulation (FA), distance to river (DR), drainage density (DD), topographic wetness index (TWI), curvature (Cu), land cover/use (LULC), NDVI, rainfall (Rf), and soil moisture (SM). The primary data sources include SRTM 30 m (EI, SI, FA, DR, DD, TWI, Cu), Sentinel-2 10 m (LULC, NDVI), GPM v6 (Rf), and SMAP (SM). All parameters are projected to the working CRS, standardized in spatial resolution, and then reclassified to an ordinal scale of 1–5 (very low – very high) with local thematic thresholds. Weights among parameters are explicitly assigned (e.g., EI/SI/FA/DR as major controlling factors), and the weighted overlay results in a hazard surface H. Physically, low or flat locations and those near rivers tend to experience flow accumulation and low flow velocity; high NDVI values indicate vegetative cover that may delay runoff; TWI and SM capture soil saturation; flat or concave curvature leads to pooling (see further explanations in Jumadi *et al.*, 2024a; Jumadi *et al.*, 2024b).

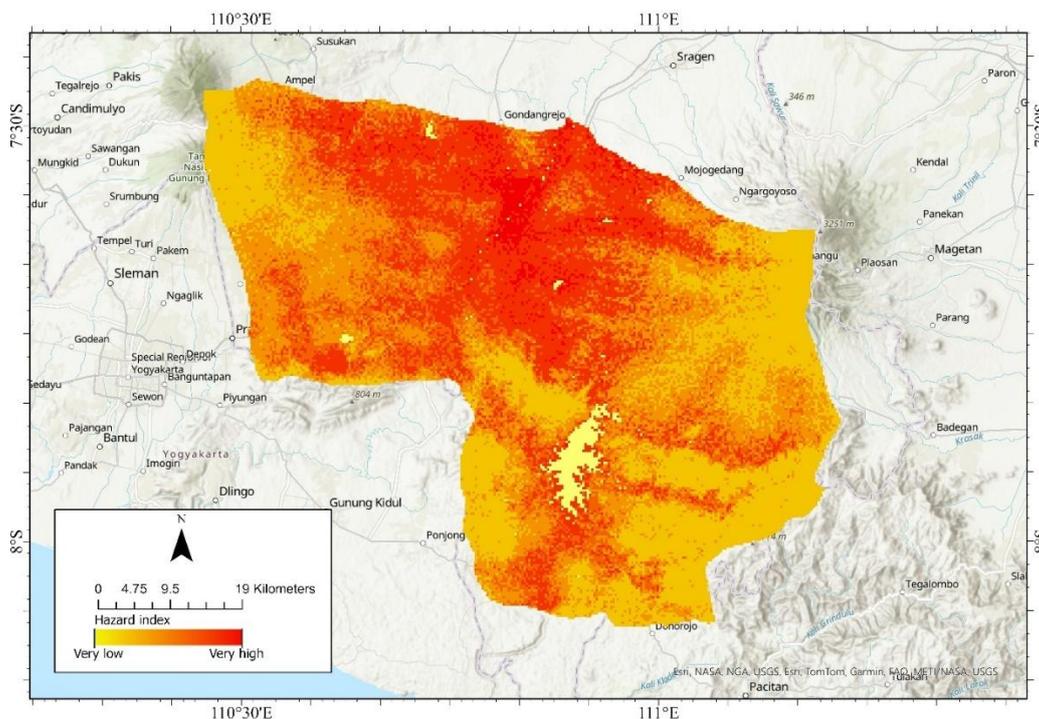


Figure 3. Flood hazard map (source: Jumadi *et al.*, 2024a; Jumadi *et al.*, 2024b).

### 3. Supporting Data

The research uses administrative boundaries (districts), river networks, watershed boundaries, and ESRI Sentinel 2 Land Cover. Administrative boundaries are essential for reporting, especially when comparing results with different regions. The river network helps interpret the results. ESRI Sentinel-2 land cover is needed to validate the distribution of SSP.

The characteristics and sources of all datasets are summarised in Table 1.

Table 1. Data Description and Sources.

No	Data	Description	Source	Derived data
1	Population	Global Human Settlement Layer (GHS-POP R2023A). Pixel at 100 m resolution	Earth Engine Data Catalogue	Population center_id, coordinates, and count (number of residents)
2	Demographic and social categories	Regency /city statistics data	Central Bureau of Statistics	Age, gender, education, occupation
3	Images	Sentinel-2 at 10 m resolution	ESA	LULC, NDVI
4	DEM	Shuttle Radar Topography Mission (SRTM) at 30 m resolution	USGS	Elevation, slope, flow accumulation, distance to river, drainage density, TWI, and curvature
5	Rainfall	Global Precipitation Measurement (GPM) v6 at 10 km resolution	NASA	Rainfall
6	Soil moisture	NASA-USDA Enhanced Soil Moisture Active Passive (SMAP) at 10 km resolution	NASA	Soil moisture
7	Land cover	Based on the ESA Sentinel-2 image at 10 m resolution	ESRI	Land cover

### 2.3. Data Pre-processing

#### 1. Spatial Harmonization

Preprocessing starts with aligning the coordinate system and analysis grid so that all layers function properly without distortion. All data is projected to UTM 49S (EPSG:32749) as a standard CRS to ensure consistency in distance and area measurements. The extent is defined by the mask of the Upper Solo River Basin (BSH), which is then used as a clipping boundary for all raster and vector data.

#### 2. Extraction of GHSL Population Data

The GHSL population raster is converted into points using ArcGIS Pro (Raster to Point). These points are clipped by the watershed mask and assigned a unique ID (center\_id). The population attribute (number of people per pixel) is cleaned by removing negative or empty values. Points with a population of zero are discarded to improve computational efficiency. If there is an off by

one issue at the watershed boundary where GHSL cells cross it, a small adjustment is applied to the analysis grid. All points are stored in the working CRS and exported to CSV for use in the Python pipeline (SSP Algorithm).

## 2.4. Spatial Synthetic Population (SSP) Development

### 1. Principles and Assumptions

The development of SSP aims to represent the population at the individual level in a synthetic way while maintaining overall consistency with official statistics and spatial accuracy regarding settlement areas. The main principles are: (i) aggregate consistency—matching the number of individuals generated at each center point to the population count from GHS-POP; (ii) micro-spatial realism—dispersing individuals around the pixel center within a fixed radius (100 m) with a denser radial distribution at the center; (iii) attribute consistency—sampling age, gender, education, and occupation based on weighted standardized frequency tables; and (iv) substantive rules—to avoid unrealistic attribute combinations (e.g., children working), such as age < 15 → education = 0, occupation = 0. Other key assumptions include: (a) category frequencies reflect the population makeup at the watershed level; (b) micro-spatial dispersion within a 100 m radius sufficiently represents the intrapixel variability observed in GHS-POP; (c) ages within intervals are assumed uniform unless additional information is available.

### 2. Spatial Algorithm (Individual Placement)

For each center point (cx, cy) with a population n, generate n individual positions using radial jitter, where the angle theta is uniformly distributed between 0 and 2π, and the radius r is drawn from a Beta(2, 2,5) distribution scaled by RADIUS\_M. This creates a denser clustering around the center. Offsets are calculated as dx = r\* cos (theta) and dy = r\* sin (theta).

### 3. Demographic-Social Attribute Sampling

Individual attributes are assigned through weighted random sampling based on frequency tables. For age, the ID label can indicate a range (e.g., “a–b”), a single value, or “a+.” The range “a–b” is converted into specific ages by selecting a random integer within [a, b]; labels “a+” are limited to [a, a+4] to avoid an infinite tail; single values are used directly. For gender, education, and occupation, categories are selected based on their relative probability (frequency divided by total). After age is assigned, the rule age < 15 is applied to disable education and occupation (set to “0”).

### 4. Computational Implementation

The implementation uses Python on Google Colab Pro (High-RAM). It reads population centers from DBF files with dbfread and converts data types to numeric. The source code for SSP generation is available at <https://doi.org/10.17605/OSF.IO/2KEJQ>.

### 5. Validation of Aggregate Consistency and Spatial Distribution

Validation involves three layers. (i) Aggregate count: the number of SSP per point must match the GHSL count; a difference greater than 0 indicates anomalies (e.g., non integer population values or casting issues). (ii) Attribute distribution: calculate the proportion of results for age, gender, education, and occupation, then compare these with frequency tables using MAE and RMSE (Equations 1 and 2), where y is the observed proportion and ŷ is SSP proportion in category i, and n is the number of categories. (iii) Compatibility of the spatial distribution of SSP with the built-up land positions according to ESRI Sentinel-2 Land Cover.

$$MAE = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i| \tag{1}$$

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2} \tag{2}$$

## 2.5. Vulnerability Analysis

### 1. Determination of Scores per Attribute Class

The method for calculating scores per class follows the schema you established in the table. To ensure consistency and facilitate multicriteria aggregation, all scores are expressed within the range [0, 1]. If an initial score column falls outside [0, 1], linear normalization can be applied

without altering the actual ordering:  $s_i(\text{norm}) = (s_i - \min s_i) / (\max s_i - \min s_i)$ . Table 2 illustrates this basic schema.

**Table 2.** Criteria and Scoring of Vulnerability.

No	Criteria	Category	Description	Vulnerability Score
1	Age	0-10	extremely vulnerable	1
		11-20	strongly vulnerable	0.85
		21-30	vulnerable	0.7
		31-40	less vulnerable	0.4
		41-50	vulnerable	0.7
		51-60	strongly vulnerable	0.85
		>60	extremely vulnerable	1
2	Education	Elementary School	extremely vulnerable	1
		Yunior High School	strongly vulnerable	0.85
		Senior High School	vulnerable	0.7
		University Graduate	less vulnerable	0.4
3	Occupation	Informal Worker	extremely vulnerable	1
		Self-employed/entrepreneur	strongly vulnerable	0.85
		Employee	vulnerable	0.7
4	Gender	Man	less vulnerable	0.4
		Woman	vulnerable	0.7

Regarding age groups, the young (0–10 years) and elderly (>60 years) are assigned the highest scores due to mobility limitations, dependency, and increased support needs during emergency responses. The middle working age group (31–40 years) receives a lower score because of its relatively high responsiveness. This mapping also assumes that higher levels of education are linked to greater risk literacy, better access to information, and improved ability to use services—thereby reducing vulnerability scores. Jobs with low income stability or high physical exposure are assigned higher scores (e.g., category 3) because of increased economic and operational risks during flood disruptions. From a gender perspective, women are considered more vulnerable than men.

2. Determination of Vulnerability Parameter Weights

The established core preferences are: Age (very important), Occupation (important), Education (somewhat important), Gender (less important) (Table 3). These preferences are linked to the Saaty scale as anchor vectors  $s = [7, 5, 3, 1]$ . The paired comparison matrix  $A = [a_{ij}]$  is created as anchor ratios, where  $a_{ij} = s_i/s_j$ , ensuring the matrix's transitivity and mathematical consistency (Table 4).

**Table 3.** Summary of Saaty Scale.

Value	Meaning
1	Equally important
3	A little more important
5	More/stronger important
7	Very strong important
9	Extreme importance
2,4,6,8	Intermediate value
1/x	Opposite preference

**Table 4.** Pairwise comparison matrix A.

	Age	Work	Education	Gender
Age	1.0	1.4	2.333	7.0
Occupation	0.714	1.0	1.667	5.0
Education	0.429	0.6	1.000	3.0
Gender	0.143	0.2	0.333	1.0

The AHP weights  $w$  are derived from the normalized principal eigenvector of matrix A (Table 5). When constructing anchor ratios, the maximum eigenvalue becomes  $\lambda_{\max} = n$ ; as a result, the

consistency index and ratio are:  $CI = (\lambda_{max} - n)/(n - 1) = 0$  and  $CR = CI/RI = 0$  (with RI for  $n = 4$  being 0.90).

**Table 5.** AHP Weights Between Attributes.

Attribute	Weight ( $w_i$ )
Age	0.4375
Occupation	0.3125
Education	0.1875
Gender	0.0625
Total	1.0

### 3. Determination of Vulnerability Index

Vulnerability (V) is the tendency of a population to experience adverse impacts when exposed to hazards, influenced by demographic characteristics and socio-economic conditions. In this context, V does not measure the likelihood of floods (hazard domain), but rather the internal capacity of individuals to anticipate, absorb, and recover. The scope of indicators is limited to attributes directly available from the SSP—namely, age, gender, education level, and employment status—where  $S_i$  is the score for the  $i$ -th attribute category and  $W_i$  is the weight for the  $i$ -th attribute (Equation 3).

$$V = \sum_{i=1}^m S_i W_i \tag{3}$$

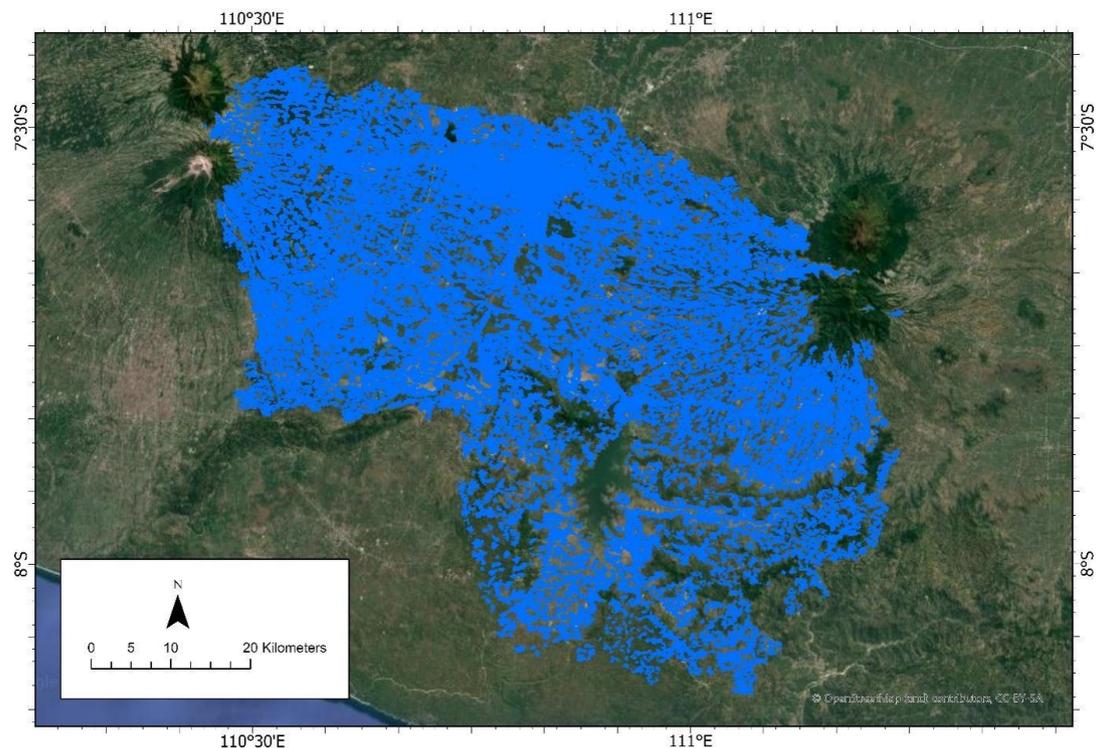
### 2.6. Risk Analysis (R)

Risk analysis combines hazards (H) and vulnerabilities (V) to evaluate the potential impact of flooding on populations. We use a complementary approach: (1) Relative Risk Index:  $R = H \times V$ , with H and V scaled from 0 to 1.

## 3. Results and Discussion

### 3.1. Spatial Synthetic Population (SSP)

The SSP modeling result is shown in Figure 4. The resulting SSP includes 4,459,936 individual entities geolocated on a 100 m grid in the Upper Bengawan Solo watershed, derived from GHS-POP R2023. Each individual is assigned directional jitter within a radius of  $\leq 100$  m of the pixel center and is assigned attributes such as age, sex, education, and occupation, which are calibrated against official district or city level statistics via marginal alignment (IPF) until they are consistent.



**Figure 4.** Spatial Synthetic Population Map.

### 3.2. Attributes Validation

The results of attribute validation are shown in Table 5 and Figure 5. Based on these results, the errors for each attribute are as follows: education (MAE 4.89%; RMSE 5.19%), occupation (MAE 6.52%; RMSE 6.93%), gender (MAE 0.0103%; RMSE 0.0103%), and age (MAE 0.0137%; RMSE 0.0163%). Notably, gender and age are nearly identical to the observed values (a difference of about 0.01), highlighting the success of mapping age bins and maintaining consistent category labels. This alignment is important because these attributes are often key factors in flood vulnerability indices (such as the proportion of children and the elderly as vulnerable groups); therefore, the stability of their values helps prevent errors from affecting the overall risk results.

Table 6. Error Metrics.

Attribute	MAE (%)	RMSE (%)
Education	4.892	5.186
Occupation	6.522	6.926
Gender	0.010	0.010
Age	0.014	0.016

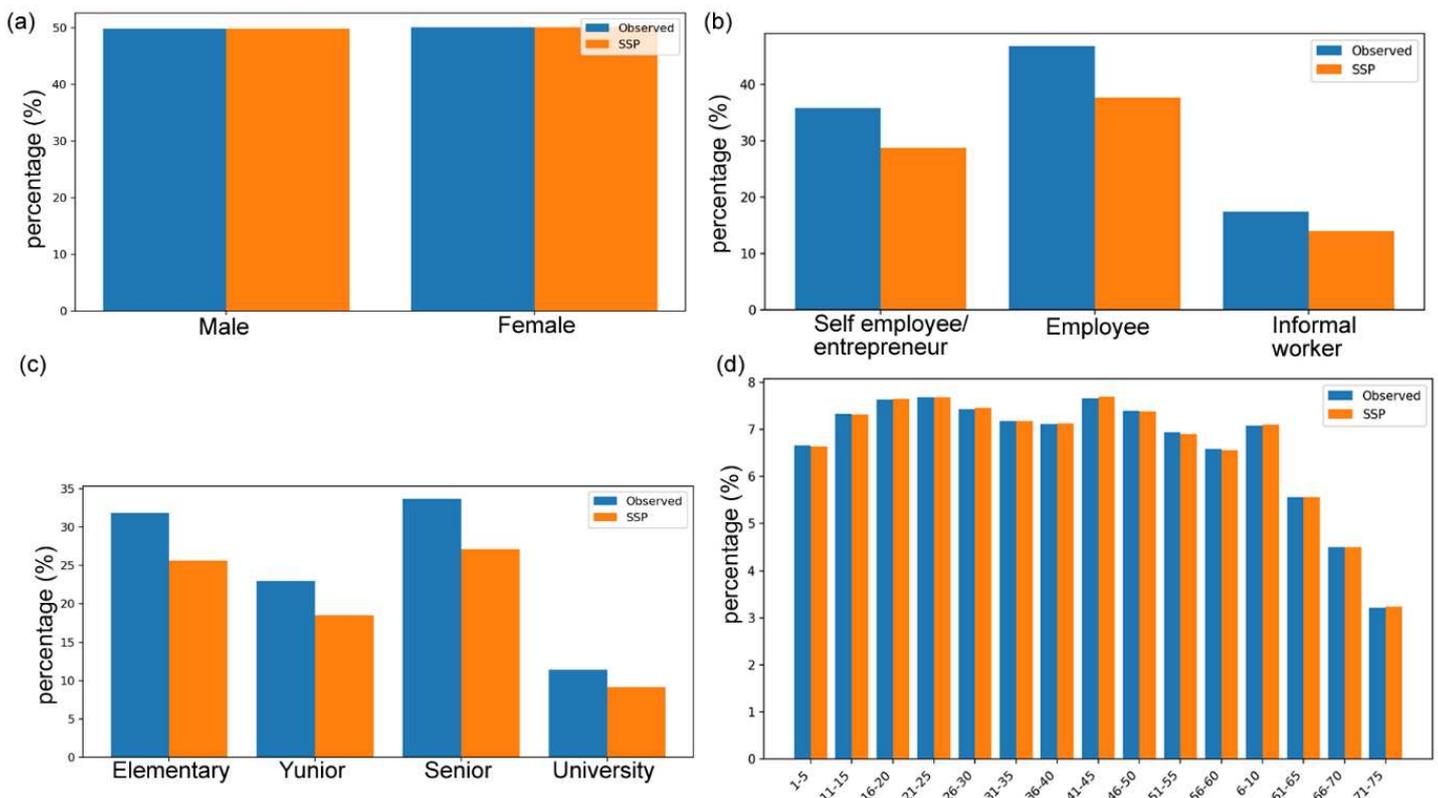


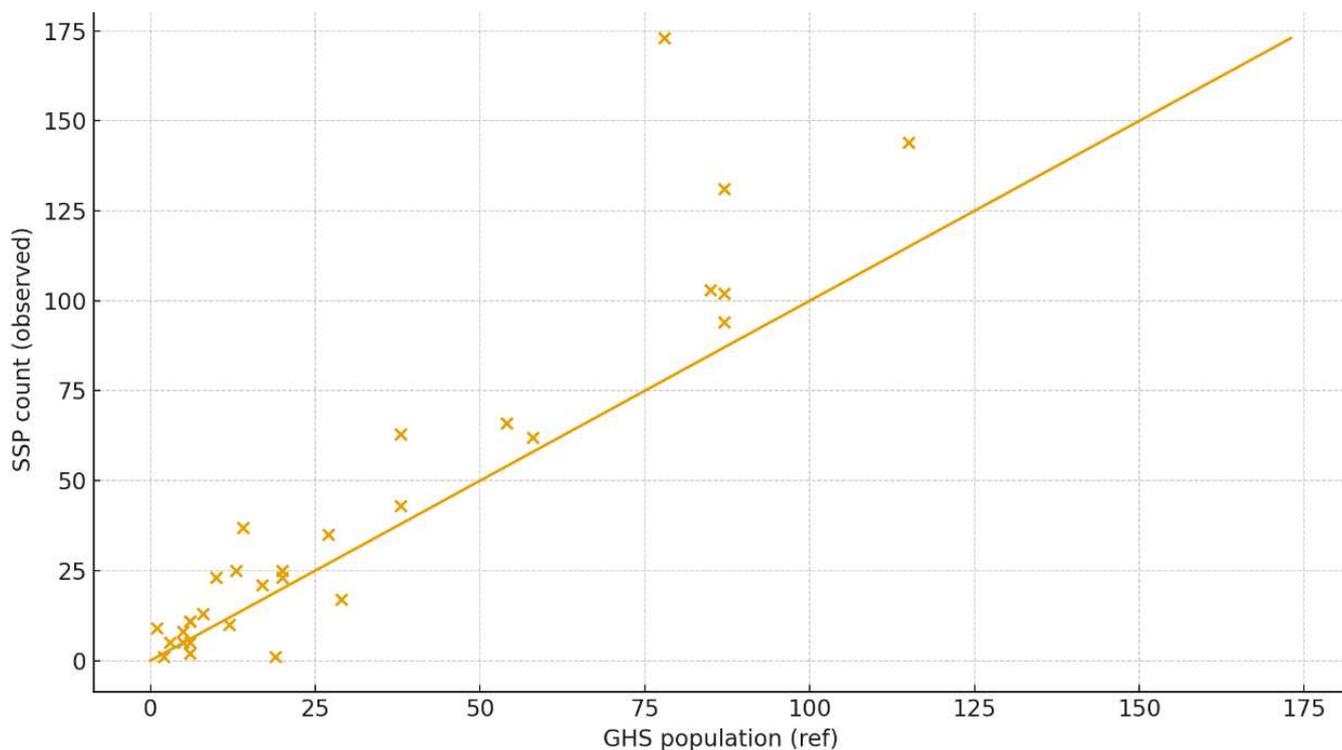
Figure 5. Comparison of Population Distribution by: (a) Gender, (b) Occupation, (c) Education, (d) Age.

Regarding education and occupation, the MAE of 4.9–6.5 reflects different dynamics: the SSP distribution closely matches the observed distribution (the pattern of comparisons between categories is preserved), but there is an offset in some high frequency categories. In practice, this often happens when category ontologies are not perfectly aligned, such as combining High School/Vocational School into a single source and distinguishing High School from Vocational School in another. In these cases, a few percentage points difference in one category is usually offset by a comparable difference in a nearby category.

### 3.3. Spatial Error Analysis

The SSP validation results against GHS-POP for 30 sample pixels show a plausible pattern (Figure 6). The relationship between the two is strong ( $r = 0.935$ ), yet there is a systematic overcount with a bias of +10.4 people per pixel (MAE 12.93; RMSE 22.22), and a 95% agreement limit that remains wide (-28.74 to 49.54), indicating significant residual uncertainty in some pixels. Structurally, the regression slope is less than 1, and increasing error at higher density bin points to heteroskedasticity and the potential for spatial “bleeding” caused by jitter points extending beyond

pixel limits and non mass preserving allocations per pixel. The SSP has successfully captured the density gradient but needs scale calibration, and jitter restrictions within pixel polygons and allocations hinder consistency at 100 m. With these adjustments, especially for aggregations  $\geq 300$  m, the accuracy is expected to improve, but expanding the sample will be necessary to stabilize the metrics and reduce the Level of Acceptance range.



**Figure 6.** Comparison between GHS and SSP Pixel Values (n = 30).

Additionally, the distribution of SSP locations compared to ESRI Sentinel-2 Land Cover (10 m) shows high consistency: out of a total of 4,235,296 points, 91.95% (3,894,372) are on built-up land and 8.05% (340,924) on non-built-up areas. Treating this as an estimate of binomial proportions, its statistical uncertainty is minimal (95% confidence interval approximately 91.92–91.98%), indicating a strong “location accuracy” signal for the SSP. Substantively, the roughly 8% off-built-up figure remains reasonable given (i) thematic errors in land cover maps (such as commission/omission of built classes, especially in rural residential areas mixed with vegetation), (ii) the impact of the 10 m boundary—misregistration and mixed pixels at settlement edges, and (iii) jitter or bleeding of SSP points near class boundaries. These results strongly support the conclusion that the placement of SSP aligns with built-up spaces; the remaining 8% likely results from class boundary issues and thematic uncertainties, which could be minimized through small buffering, datum/registration alignment, or the use of building footprints in the test zone.

### 3.4. Population Vulnerability Analysis based on SSP

The demographic vulnerability analysis indicates a very “low” profile (Figures 7 and 8). The average weighted population is 0.2160 (scale 0–1), with 99.63% of the population categorized as Very Low (44.43%) and Low (55.20%), while only 0.37% falls into the Moderate category. The AHP weights are consistent (CR=0), confirming that age is the primary factor ( $w=0.5333$ ), followed by occupation (0.2667), education (0.1333), and gender (0.0667).

Consequently, pockets of vulnerability mainly develop in areas with higher proportions of children and the elderly, or in areas dominated by low adaptive capacity jobs (informal/daily workers) and low education levels, even though their overall contribution is smaller than their share of the population. Spatially, the impact is that vulnerability hotspots are limited but operationally significant: although they account for only 0.37% of the population, the intersection of moderate pockets with high hazard zones (H) could prioritize interventions such as increasing household preparedness, improving access to emergency services, and strengthening social networks. These results highlight that most regions possess reasonable demographic resilience reserves; however, risk management must continue to focus on micro locations with vulnerable age groups and job/education vulnerabilities to deploy mitigation resources effectively.

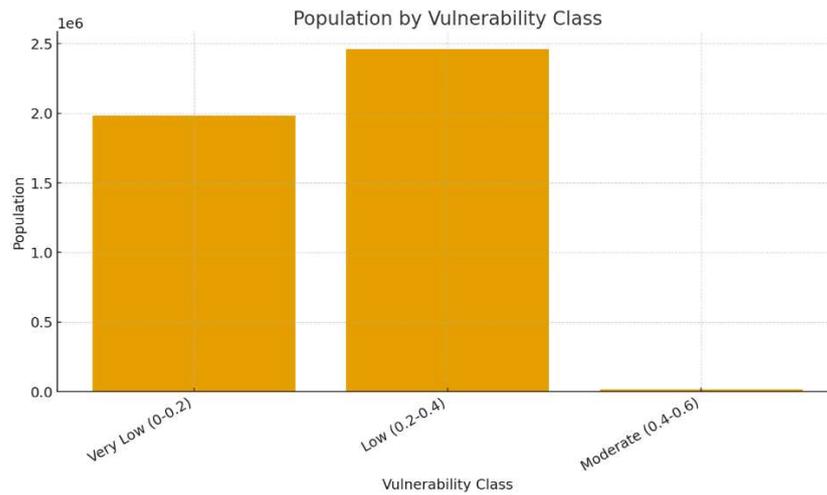


Figure 7. Distribution of Population by Vulnerability Class.

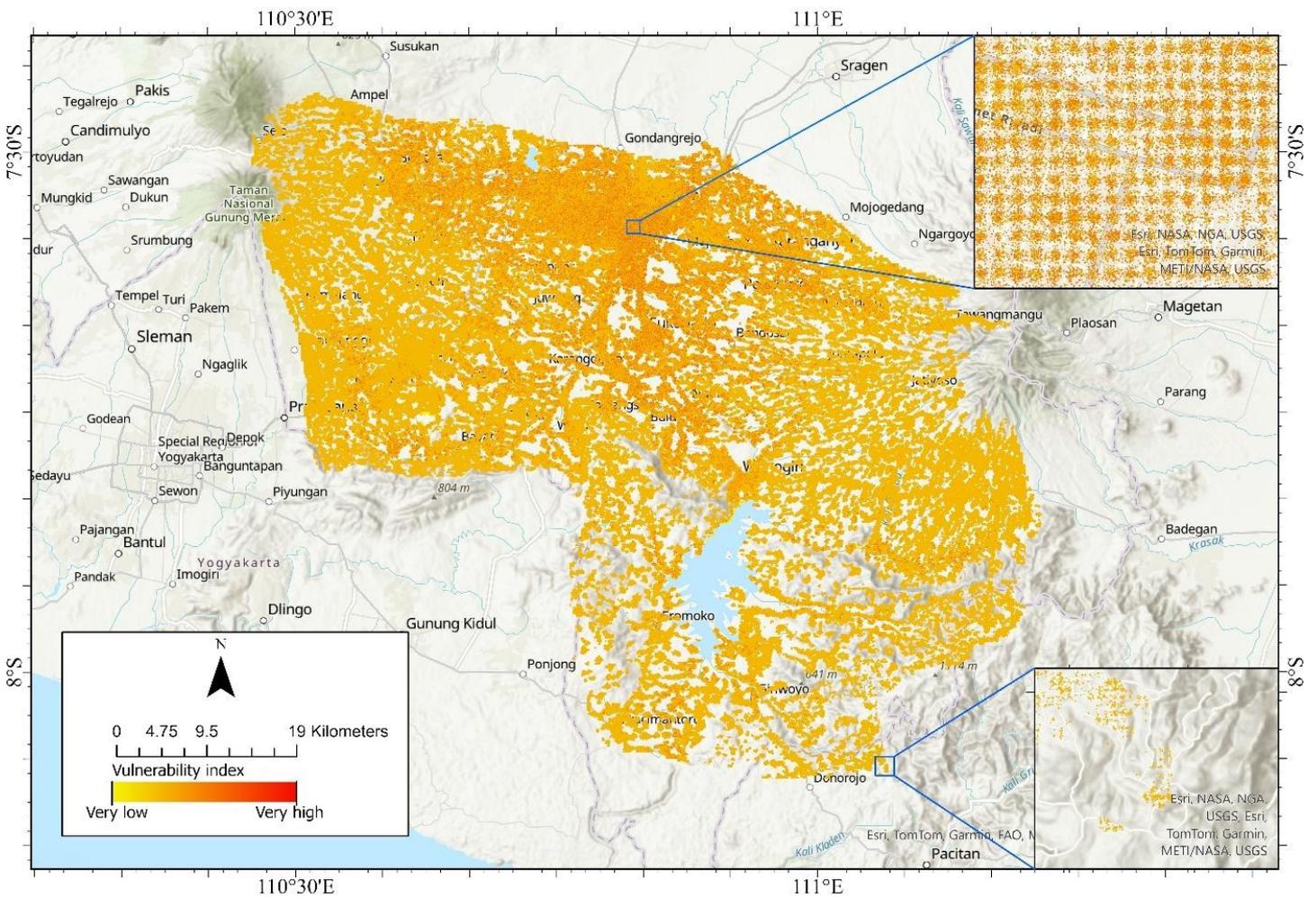


Figure 8. Spatial Distribution of Population by Vulnerability Class.

### 3.4. Risk

The distribution of the population's flood risk based on SSP (approximately 4.23 million entities) is heavily concentrated in the Low Risk category (3,391,760; 80.21%), with a substantial portion in the Moderate Risk category (830,469; 19.64%). Meanwhile, the tail distribution is nearly zero for the Very Low (5,026; 0.12%) and High (1,602; 0.04%) categories, indicating that overall exposure is primarily in the low risk class. However, about 832,071 ( $\pm 19.68\%$ ) of the population falls into the moderate high risk category, which should be the focus of priority interventions, such as improvements to early warning systems, micro infrastructure protection, and spatial planning in hotspots (see Figure 9).

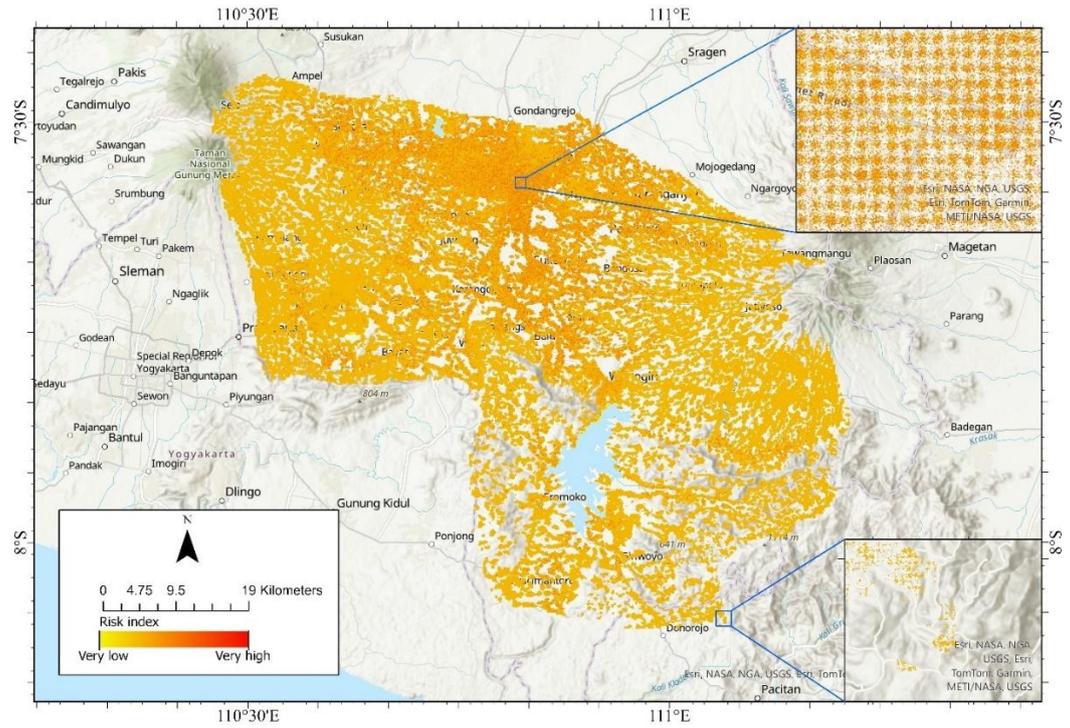


Figure 9. Spatial Distribution of Population by Risk Index.

The Getis–Ord  $G_i^*$  analysis reveals that risk hotspots form significant clusters along river network corridors and in urban core areas (Figure 10). The directional pattern along water pathways indicates the dominance of river flooding in floodplains, while the clustering in urban zones reflects pluvial flooding caused by high impervious surfaces and limited drainage capacity; both are worsened by high exposure (population/asset density) and local vulnerabilities (informal settlements, fragile buildings, service inequality). Therefore, the interaction of  $H \times V$  explains the consistently positive z-score of  $G_i$  observed in riverbanks and urban pockets. Operationally, this involves addressing setbacks and restoring floodplain connectivity, improving drainage systems and green-blue infrastructure in cities, and prioritizing early warning systems and protecting critical facilities in hotspots. To ensure the accuracy of these conclusions, sensitivity testing should be performed on spatial neighbor weighting schemes (including river network based methods), along with corrections for multiple testing and cross validation using historical flood records.

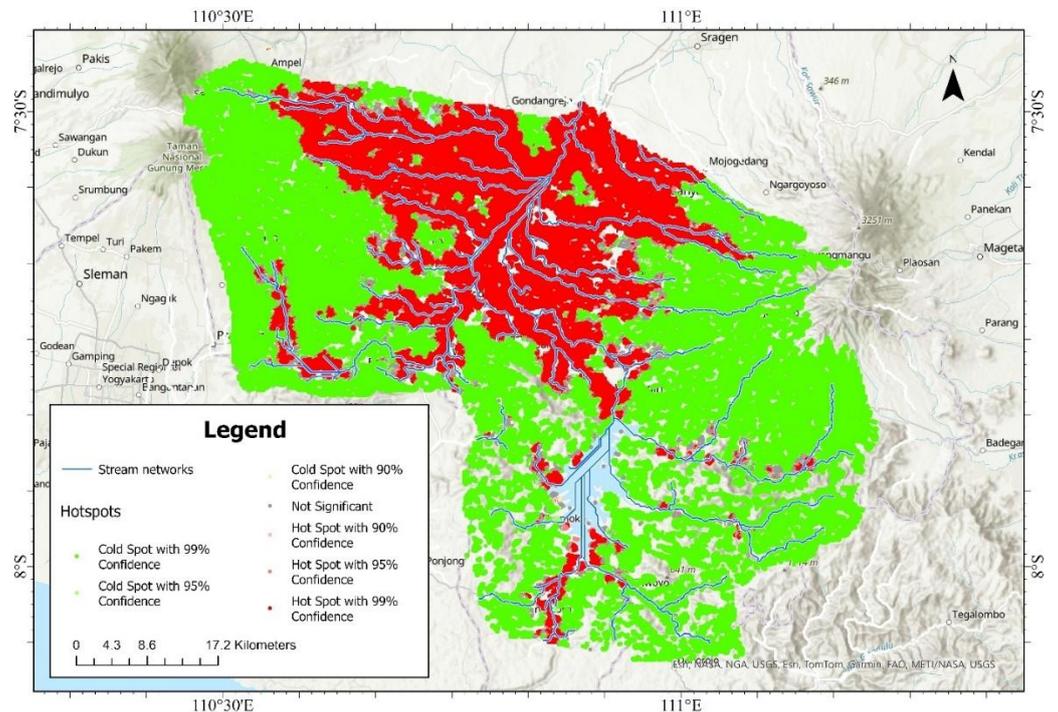


Figure 10. Risk Hotspots.

### 3.4. Discussion

The metrics findings show that the synthetic population distribution (SSP) has been accurately reconstructed regarding attribute spread and spatial distribution, although some residual errors remain in several categories. In high resolution population mapping, harmonizing category definitions and maintaining consistent classification schemes are essential steps to reduce artifact related residual bias (Rubinyi *et al.*, 2022).

Substantively, age and gender attributes exhibit practical errors of approximately 0.01% ( $\approx 0.01\%$ ). This confirms the success of age bin mapping, the consistency of category labels, and the normalization of proportions. High accuracy in these two attributes is essential because the proportions of children and the elderly often influence response capacity and evacuation needs, while demographic structure also impacts risk projection by incorporating population dynamics and climate change (Shu *et al.*, 2023). Therefore, the propagation of errors to the overall vulnerability index should be minimal. Conversely, in education and employment, a mean absolute error (MAE) of around 5–7 percentage points was observed. While this is generally considered a good fit for categorical shape, some dominant classes still show offset levels. From a policy standpoint, such deviations tend to affect thematic evaluations—such as post flood job training programs or school level disaster literacy initiatives—more than they impact spatial risk rankings between zones, provided these two attribute weights do not dominate the composite index.

From a utilization standpoint, current performance is sufficient for operational risk mapping at a resolution of  $\geq 100$  m (or village level). The research on composite based flood risk assessment models indicates that aggregate outcomes are fairly resilient to minor variations in less sensitive components, as long as the most sensitive factors—in this case, age and gender—are accurately measured. However, it is also important to note that fluctuations in population datasets alone have been shown to alter flood exposure estimates across different regions (Li *et al.*, 2024; Zhang *et al.*, 2025; Zhang *et al.*, 2025). Consequently, documenting and monitoring residual patterns remain essential for quality control, especially in urban and peri urban areas, which often display greater social and economic diversity.

Regarding sources of bias, three main factors should be noted. First, misalignment between category lexicons (observed vs. SSP) could cause discrepancies of several percentage points in large categories. Second, imperfections in conditioning employment attributes on demographic predictors and regional contexts (e.g., age, education, and urban rural typologies) may lead to small but systematic distortions. Third, rounding in reference data can mathematically amplify differences in high frequency classes without changing the overall shape of the distribution. Two low cost steps are recommended that do not alter the SSP architecture: (i) post synthesis ranking or Iterative Proportional Fitting (IPF) in each territorial unit to "attach" margins to the targets; this method is common and effectively reduces the mean absolute error (MAE) of categories to  $< 3$  percentage points (Nejad *et al.*, 2021), and (ii) one to one category ontology harmonization (consistently merging or crossing "similar" classes), as recommended in high resolution synthetic population mapping (Rubinyi *et al.*, 2022). Both steps improve accuracy without sacrificing spatial consistency or computational efficiency.

As an additional measure, two more tests are recommended. First, a sensitivity analysis of the weight assigned to the vulnerability index: assessing the stability of spatial rankings across reasonable variations in educational and employment weights will confirm the robustness of the results. Second, constructing a residual map (observed – SSP, in %) for each main category can help identify spatial bias patterns—such as residual concentration in urban cores versus peri urban areas—that may influence the next ranking stage. This can also help distinguish between methodological bias and actual demographic heterogeneity (Zhang *et al.*, 2025). In relation to medium to long term planning, combining demographic structure accuracy with these light balancing mechanisms is also important for a risk projection framework that considers future population dynamics and climate related hazard scenarios (Shu *et al.*, 2023). Overall, these results support the conclusion that the SSP has adequately captured key demographic structures for area prioritization and operational level mitigation planning, with room for improvement through IPF/ranking and category harmonization. This approach aligns with industry best practices and offers a way to increase precision without adding computational complexity or model architecture overhead (Nejad *et al.*, 2021; Rubinyi *et al.*, 2022; Shu *et al.*, 2023; Zhang *et al.*, 2025; Zhang *et al.*, 2025).

### 4. Conclusion

This study presents the development of SSP for flood risk analysis in the Upper Bengawan Solo Watershed. The developed SSP is valid in both a distributional and a spatial sense, with some

suggestions for improvement. For attributes, the fit to reference data is very high for sex and age (MAE=0.0103; RMSE=0.0103 and MAE=0.0137; RMSE=0.0163). Education (MAE 4.89; RMSE 5.19) and employment (MAE 6.52; RMSE 6.93) preserve a similar distribution shape to the observed but show an offset in some high frequency categories. In spatial validation results, the SSP-GHS relationship is strong ( $r=0.935$ ), but there is an overcount of +10.4 people per pixel (MAE 12.93; RMSE 22.22) and wide 95% Limits of Agreement (-28.74 to 49.54), indicating heteroskedasticity and possible issues from jitter and allocations not being mass preserving per 100 m. Limiting jitter within pixels, using conservative allocations, and aggregating at  $\geq 300$  m are expected to improve accuracy. Positional validation against ESRI Sentinel-2 shows that 91.95% of points are within built-up areas (CI95% $\approx$ 91.92–91.98%). In comparison, about 8% outside built-up areas is likely due to thematic or edge uncertainty, which can be reduced through buffering, improved registration, or more precise building footprint definitions. The risk of residents is mainly in Low (80.21%) and Moderate (19.64%) categories. Getis-Ord Gi analysis confirms risk hotspots tend to run along river corridors and cluster in urban centers—reflecting the combined effects of river flood and pluvial processes—with operational implications: setback planning, floodplain restoration, improving drainage and green blue infrastructure, strengthening early warning systems, and protecting critical facilities, alongside spatial weight sensitivity tests, corrections for multiple testing, and cross validation against historical floods. In short, the research enables fine scale flood risk assessment, identifying hotspots for targeted interventions such as urban planning, drainage improvements, and early warning systems. Validated and robust, the SSP framework offers policymakers a practical tool to enhance resilience, optimize resources, and mitigate the societal impacts of rising flood hazards.

### Acknowledgements

The author would like to express sincere gratitude to all parties who contributed to the research and completion of this manuscript, including mentors, funding supporters, and those who provided essential data and information. Their assistance has been invaluable to this work.

### Author Contributions

**Conceptualization:** Musiyam, M., Jumadi, J., Fikriyah, V. N., Masrurroh, H., Septiyani, E. D., Amin, H., Hadibasyir, H. Z., Sattar, F., & Nawaz, M.; **methodology:** Musiyam, M., & Jumadi, J.; **investigation:** Musiyam, M., Fikriyah, V. N., & Nawaz, M.; **writing—original draft preparation:** Musiyam, M., Jumadi, J., Septiyani, E. D., Amin, H., Hadibasyir, H. Z., Sattar, F., & Nawaz, M.; **writing—review and editing:** Musiyam, M., & Nawaz, M.; **visualization:** Musiyam, M. All authors have read and agreed to the published version of the manuscript.

### Conflict of interest

All authors declare that they have no conflicts of interest.

### Data availability

Data is available upon Request.

### Funding

This research was funded by the Matching Fund Program of Universitas Muhammadiyah Surakarta (DRPPS) and Universitas Negeri Malang MoA Number 5.2.7/UN32.14/KS/2024 and 48/C.3-III/LRI/II/2024 through RKN Scheme.

### References

- Abdelkareem, M., & Mansour, A. M. (2023). Risk assessment and management of vulnerable areas to flash flood hazards in arid regions using remote sensing and GIS-based knowledge-driven techniques. *Natural Hazards*, 117(3), 2269–2295. doi: 10.1007/s11069-023-05942-x
- Absori, A., Hernanda, T., Fitrijadi, A., Wardiono, K., & Budiono, A. (2023). Analysis of the Issues on Bengawan Solo River Basin Management Policies. *WSEAS Transactions on Environment and Development*, 19, 25–32.
- Agriesti, S., Roncoli, C., & Nahmias-Biran, B. (2022). Assignment of a synthetic population for activity-based modeling employing publicly available data. *ISPRS International Journal of Geo-Information*, 11(2), 148.
- Anna, A. N., & Priyana, Y. (2015). Kajian biofisik lahan untuk penilaian kerentanan banjir di DAS Bengawan Solo Hulu. *Jurnal Pengelolaan Sumberdaya Alam dan Lingkungan*, 8(1), 96–108. doi: 10.29244/jpsl.8.1.96-108
- Anna, A. N., Priyana, Y., Rudiyanto, & Fikriyah, V. N. (2023). Water resources management model based area (A case study in some areas of Bengawan Solo watershed). *Proceeding Of International Summit On Education, Technology, And Humanity 2021*, 2727(1), 050027.
- Arnell, N. W., & Gosling, S. N. (2016). The impacts of climate change on river flood risk at the global scale. *Climatic Change*, 134(3), 387–401. doi: 10.1007/s10584-014-1084-5
- Balac, M., & Hörl, S. (2021). *Synthetic population for the state of California based on open-data: Examples of San Francisco Bay area and San Diego County*. 100th Annual Meeting of the Transportation Research Board (TRB). Retrieved From <https://hal.science/hal-03208848/>
- Bigi, F., Rashidi, T. H., & Viti, F. (2024). Synthetic Population: A Reliable Framework for Analysis for Agent-Based Modeling in Mobility. *Transportation Research Record: Journal of the Transportation Research Board*, 2678(11), 1–15. doi: 10.1177/03611981241239656
- Burger, A., Oz, T., Crooks, A., & Kennedy, W. G. (2017). Generation of Realistic Mega-City Populations and Social Networks for Agent-Based Modeling. *Proceedings of the 2017 International Conference of The Computational Social Science Society of the Americas*, 1–7. doi: 10.1145/3145574.3145593
- Chakraborty, L., Thistlethwaite, J., Scott, D., Henstra, D., Minano, A., & Rus, H. (2023). Assessing social vulnerability and identifying spatial hotspots of flood risk to inform socially just flood management policy. *Risk Analysis*, 43(5), 1058–1078. doi: 10.1111/risa.13978
- Chapuis, K., & Taillandier, P. (2019). *A brief review of synthetic population generation practices in agent-based social simulation*. Submitted to SSC2019, Social Simulation Conference. Retrieved From [https://www.researchgate.net/profile/Kevin-Chapuis/publication/335601121\\_A\\_brief\\_review\\_of\\_synthetic\\_population\\_generation\\_practices\\_in\\_agent-based\\_social\\_simulation/links/5d6f58e592851c1b3b279b9a/A-brief-review-of-synthetic-population-generation-practices-in-agent-based-social-simulation.pdf](https://www.researchgate.net/profile/Kevin-Chapuis/publication/335601121_A_brief_review_of_synthetic_population_generation_practices_in_agent-based_social_simulation/links/5d6f58e592851c1b3b279b9a/A-brief-review-of-synthetic-population-generation-practices-in-agent-based-social-simulation.pdf)
- Cutter, S. L., Boruff, B. J., & Shirley, W. L. (2003). Social Vulnerability to Environmental Hazards\*. *Social Science Quarterly*, 84(2), Article 2. doi: 10.1111/1540-6237.8402002
- Cutter, S. L., & Finch, C. (2008). Temporal and spatial changes in social vulnerability to natural hazards. *Proceedings of the National Academy of Sciences*, 105(7), 2301–2306. doi: 10.1073/pnas.0710375105
- De Mooij, J., Sonnenschein, T., Pellegrino, M., Dastani, M., Ettema, D., Logan, B., & Versteegen, J. A. (2024). GenSynthPop: Generating a spatially explicit synthetic population of individuals and households from aggregated data. *Autonomous Agents and Multi-Agent Systems*, 38(2), 48. doi: 10.1007/s10458-024-09680-7
- Diriba, D., Takele, T., Karuppannan, S., & Husein, M. (2024). Flood hazard analysis and risk assessment using remote sensing, GIS, and AHP techniques: A case study of the Gidabo Watershed, main Ethiopian Rift, Ethiopia. *Geomatics, Natural Hazards and Risk*, 15(1), 2361813. doi: 10.1080/19475705.2024.2361813
- Jiang, N., Crooks, A. T., Kavak, H., Burger, A., & Kennedy, W. G. (2022). A method to create a synthetic population with social networks for geographically-explicit agent-based models. *Computational Urban Science*, 2(1), 7. doi: 10.1007/s43762-022-00034-1
- Jiang, N., Yin, F., Wang, B., & Crooks, A. T. (2024). A Large-Scale Geographically Explicit Synthetic Population with Social Networks for the United States. *Scientific Data*, 11(1), 1204.
- Jumadi, J., Carver, S. J., & Quincey, D. J. (2019). An agent-based evaluation of varying evacuation scenarios in Merapi: Simultaneous and staged. *Geosciences*, 9(7), 317.

- Jumadi, J., Danardono, D., Priyono, K. D., Roziaty, E., Masruroh, H., Rohman, A., Amin, C., Hadibasyir, H. Z., Fikriyah, V. N., Nawaz, M., Sattar, F., & Lotfata, A. (2024). Utilizing Open Access Spatial Data for Flood Risk Mapping: A Case Study in the Upper Solo Watershed. *Geoplanning: Journal of Geomatics and Planning*, 11(2), 189–204. doi: 10.14710/geoplanning.11.2.189-204
- Jumadi, J., Heppenstall, A., Malleson, N., Carver, S., Quincey, D., & Manville, V. (2018). Modelling individual evacuation decisions during natural disasters: A case study of volcanic crisis in Merapi, Indonesia. *Geosciences*, 8(6), Article 6.
- Jumadi, J., Malleson, N., Carver, S., & Quincey, D. (2020). Estimating spatio-temporal risks from volcanic eruptions using an agent-based model. *Journal of Artificial Societies and Social Simulation*, 23(2) 2).
- Jumadi, J., Sari, D. N., Umrotun, U., Musiyam, M., Nurmantyo, C., Muhammad, S. F., & Ibrahim, M. H. (2024). Remote Sensing and GIS-Driven Model for Flood Susceptibility Assessment in the Upper Solo River Watershed. *Geographia Technica*, 19(2).
- Jumadi, J., Danardono, D., Roziaty, E., Ulinuha, A., Supari, S., Choy, L. K., Sattar, F., & Nawaz, M. (2025). AI-Driven Ensemble Learning for Spatio-Temporal Rainfall Prediction in the Bengawan Solo River Watershed, Indonesia. *Sustainability*, 17(20), 9281. doi: 10.3390/su17209281
- Li, Y., Gao, J., Yin, J., Liu, L., Zhang, C., & Wu, S. (2024). Flood Risk Assessment of Areas under Urbanization in Chongqing, China, by Integrating Multi-Models. *Remote Sensing*, 16(2), 2. doi: 10.3390/rs16020219
- Maqtan, R., Othman, F., Jaafar, W. Z. W., & Elshafie, A. (2025). Flash flood data compilation and risk insights in Peninsular Malaysia: a data-driven analysis (2014–2019). *Natural Hazards*, 1-27.
- Musiyam, M., Jumadi, J., Wibowo, Y. A., Widiyatmoko, W., & Hafida, S. H. N. (2020). Analysis of Flood-Affected Areas Due to Extreme Weather In Pacitan, Indonesia. *Geomate Journal*, 19(75), 27–34.
- Nejad, M. M., Erdogan, S., & Cirillo, C. (2021). A statistical approach to small area synthetic population generation as a basis for carless evacuation planning. *Journal of Transport Geography*, 90, 102902.
- Pesaresi, M., Schiavina, M., Politis, P., Freire, S., Krasnodębska, K., Uhl, J. H., Carioli, A., Corbane, C., Dijkstra, L., Florio, P., Friedrich, H. K., Gao, J., Leyk, S., Lu, L., Maffeni, L., Mari-Rivero, I., Melchiorri, M., Syrris, V., Van Den Hoek, J., & Kemper, T. (2024). Advances on the Global Human Settlement Layer by joint assessment of Earth Observation and population survey data. *International Journal of Digital Earth*, 17(1), 2390454. doi: 10.1080/17538947.2024.2390454
- Prédhumeau, M., & Manley, E. (2023). A synthetic population for agent-based modelling in Canada. *Scientific Data*, 10(1), 148.
- Purwanto, A., Andrasgoro, D., Eviliyanto, E., Rustam, R., Ibrahim, M. H., & Rohman, A. (2023). Validating the GIS-based flood susceptibility model using Synthetic Aperture Radar (SAR) data in sengah Temila watershed, Landak Regency, Indonesia. *Forum Geografi*, 36(2), 185–201. doi: 10.23917/forgeo.v36i2.16368
- Rubinyi, S., Verschuur, J., Goldblatt, R., Gussenbauer, J., Kowarik, A., Mannix, J., Bottoms, B., & Hall, J. (2022). High-resolution synthetic population mapping for quantifying disparities in disaster impacts: An application in the Bangladesh Coastal Zone. *Frontiers in Environmental Science*, 10, 1033579.
- Sahid, S. (2024). Enhancing Digital Elevation Model Accuracy for Flood Modelling—A Case Study of the Ciberes River in Cirebon, Indonesia. *Forum Geografi*, 38(1), 40–56. doi: 10.23917/forgeo.v38i1.1839
- Santhyami, S., Jumadi, J., Priyono, K. D., Rahayu, T., Sari, D. N., Othman, M., & Rudi, R. (2025). Spatio-Temporal Modeling for the Analysis of Hydrological Drought and its Impact on Rice Production in the Upper Bengawan Solo Basin, Central Java, Indonesia. *Geographia Technica*, 20(2).
- Sawangnate, C., Chaisri, B., & Kittipongvises, S. (2022). Flood hazard mapping and flood preparedness literacy of the elderly population residing in Bangkok, Thailand. *Water*, 14(8), 1268.
- Schiavina, M., Freire, S., & MacManus, K. (2023). *GHS-POP R2023A-GHS population grid multitemporal (1975–2030)*. Joint Research Centre Data Catalogue.
- Shu, E. G., Porter, J. R., Hauer, M. E., Sandoval Olascoaga, S., Gourevitch, J., Wilson, B., Pope, M., Melecio-Vazquez, D., & Kearns, E. (2023). Integrating climate change induced flood risk into future population projections. *Nature Communications*, 14(1), 7870.
- Tellman, B., Sullivan, J. A., Kuhn, C., Kettner, A. J., Doyle, C. S., Brakenridge, G. R., Erickson, T. A., layback, D. A. (2021). Satellite imaging reveals increased proportion of population exposed to floods. *Nature*, 596(7870), 80–86. doi: 10.1038/s41586-021-03695-w
- Usamah, M., Handmer, J., Mitchell, D., & Ahmed, I. (2014). Can the vulnerable be resilient? Co-existence of vulnerability and disaster resilience: Informal settlements in the Philippines. *International journal of disaster risk reduction*, 10, 178-189.
- Waiyasusri, K., Wetchayont, P., Tananonchai, A., & Suwanmajo, D. (2023). Flood susceptibility mapping using logistic regression analysis in Lam Khan Chu watershed, Chaiyaphum Province, Thailand. *Geography, Environment, Sustainability*, 16(2), 41-56.
- Waiyasusri, K., Kulpanich, N., Worachairungreung, M., Sae-Ngow, P., & Chaysmithikul, P. (2021). Flood Prone Risk Area Analysis During 2005-2019 In Lam Se Bok Watershed, Ubon Ratchathani Province, Thailand. *Geographia Technica*, 16(1).
- Zhang, M., Fu, X., Liu, S., & Zhang, C. (2025). Integrating Remote Sensing and Machine Learning for Actionable Flood Risk Assessment: Multi-Scenario Projection in the Ili River Basin in China Under Climate Change. *Remote Sensing*, 17(7), 1189.
- Zhang, Q., Li, C., Wen, D., Kang, J., Chen, T., Zhang, B., Hu, Y., Yin, J., & Slater, L. (2025). Global South shows higher urban flood exposures than the Global North under current and future scenarios. *Communications Earth & Environment*, 6(1), 594.