INVESTIGATING THE RELATIONSHIP BETWEEN THE URBAN HEAT ISLAND EFFECT AND SHORT-DURATION EXTREME RAINFALL IN KUALA LUMPUR

TAN YAN KAI

UNIVERSITI TUNKU ABDUL RAHMAN

INVESTIGATING THE RELATIONSHIP BETWEEN THE URBAN HEAT ISLAND EFFECT AND SHORT-DURATION EXTREME RAINFALL IN KUALA LUMPUR

TAN YAN KAI

A project report submitted in partial fulfilment of the requirements for the award of Bachelor of Civil Engineering with Honours

Lee Kong Chian Faculty of Engineering and Science
Universiti Tunku Abdul Rahman

May 2025

DECLARATION

I hereby declare that this project report is based on my original work except for citations and quotations which have been duly acknowledged. I also declare that it has not been previously and concurrently submitted for any other degree or award at UTAR or other institutions.

Name : TAN YAN KAI

ID No. : 2003386

Date : 10 May 2025

COPYRIGHT STATEMENT

This final year project report is submitted in partial fulfilment of the requirements for the degree of Bachelor of Civil Engineering with Honours at Universiti Tunku Abdul Rahman (UTAR). This final year project report represents the work of the author, except where due acknowledgement has been made in the text. No part of this final year project report may be reproduced, stored, or transmitted in any form or by any means, whether electronic, mechanical, photocopying, recording, or otherwise, without the prior written permission of the author or UTAR, in accordance with UTAR's Intellectual Property Policy.

ACKNOWLEDGEMENTS

I would like to express my heartfelt gratitude to Ir. Prof. Dr. Huang Yuk Feng, my research supervisor, and Ts. Dr. Chin Ren Jie, my research co-supervisor, for their invaluable advice, unwavering guidance and immense patience throughout the development of this research.

My sincere thanks also go to the faculty members and staff of the Lee Kong Chian Faculty of Engineering and Science, especially the Department of Civil Engineering, for fostering a supportive and pleasant working environment during my years at UTAR.

I am deeply grateful to all my beloved friends for their constant support, encouragement and assistance, without which this research would not have been possible.

Lastly, I wish to extend my deepest appreciation to my dearest family members for their unwavering support and financial assistance, which have been instrumental to the success of this research project.

ABSTRACT

Urbanisation significantly alters land surface characteristics, leading to the intensification of the urban heat island (UHI) effect, which may influence the short-duration extreme rainfall. This study investigates the relationship between UHI intensity and short-duration extreme rainfall in Kuala Lumpur through an integrated remote sensing, machine learning and statistical approach. Landsat imagery from 2007, 2015 and 2023 was used to analyse spatiotemporal changes in land use and land cover (LULC) and to estimate land surface temperature (LST). LULC classification was performed using Support Vector Machine (SVM) and Random Forest (RF) algorithms, while LST was estimated using the Single Channel (SC) algorithm and surface urban heat island intensity (SUHII) was subsequently derived from the LST data. Hourly rainfall data exceeding the 99th percentile from 2007 to 2023 were used to assess spatiotemporal variation, diurnal distribution and trends. Statistical relationships between SUHII and hourly extreme rainfall were examined using the coefficient of determination (R^2) and Kendall's Tau (τ). Results show that SVM consistently outperformed RF in terms of overall accuracy and kappa coefficient across all study years. Built-up areas and SUHII both exhibited a net increase, particularly in northern Kuala Lumpur, likely due to intense urbanisation and industrial activities. The number of hourly extreme rainfall events also increased, especially during late afternoon and evening hours. However, the mean intensity of extreme rainfall events remained relatively stable. Correlation analysis identified moderate, statistically significant relationships between the annual SUHII and the annual total number of hourly extreme rainfall events at four of nine stations (R^2 = 0.2530 - 0.3088; τ = 0.3616 - 0.4593; $p < 0.05$). These findings suggest that urban-induced heating may contribute to enhanced localised convective rainfall. It is recommended that UHI mitigation measures, such as green infrastructure and climate-sensitive urban planning, be prioritised to manage future rainfall-related flood risks in urban environments.

Keywords: Urban Heat Island; Land Surface Temperature; Land Use and Land Cover; Remote Sensing; Machine Learning; Short-Duration Extreme Rainfall; Rainfall Analysis; Relationship

Subject Area: TA170-171 Environmental Engineering

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF SYMBOLS / ABBREVIATIONS

τ	Kendall's tau correlation coefficient
R^2	Coefficient of determination
H_0	Null hypothesis
H_1	Alternative hypothesis
α	Significance level
r_1	Lag-1 serial correlation
r	Pearson's correlation coefficient
ρ	Spearman's Rho Correlation Coefficient
P_0	Observed agreement
P_1	Expected agreement
L_λ	TOA spectral radiance, $W/(m^2 \cdot sr \cdot \mu m)$
$LMAX_\lambda$	Spectral radiance that is scaled to $Q_{cal\,max}$, $W/(m^2 \cdot sr \cdot \mu m)$
$LMIN_\lambda$	Spectral radiance that is scaled to $Q_{cal\,min}$, $W/(m^2 \cdot sr \cdot \mu m)$
Q_{cal}	Quantised calibrated pixel value in DNs
$Q_{cal\,max}$	Maximum quantised calibrated pixel value corresponding to $LMAX_\lambda$
$Q_{cal\,min}$	Minimum quantised calibrated pixel value corresponding to $LMIN_\lambda$
BT	TOA brightness temperature, °C
K_1	A constant value for Band 10 thermal conversion
K_2	A constant value for Band 10 thermal conversion
M_L	Radiance multiplicative scaling factor for Band 10
A_L	Radiance additive scaling factor for Band 10
P_V	Proportion of vegetation
$\in$	Land surface emissivity
λ	Wavelength of emitted radiance
c	Velocity of light
h	Planck's constant
σ	Boltzmann constant

LST_i	Land surface temperature at any pixel, °C
$LST_{vegetation}$	Mean land surface temperature of vegetation covers, °C
P_i	The rainfall estimate at the station i
P_x	Rainfall measurement at neighbouring station x
n	The number of neighbouring station / The number of tied group / Sample size / Total number of observations
t	The power parameter that influences the weight of neighbouring measurements
N	Sample size
S	MK test statistic
x_j, x_i	Consecutive data points
Var(S)	Variance of S
t_k	The number of ties within the k^{th} group
Z	The standardised test statistic
h	Hurst coefficient
ρ_l	The autocorrelation function of lag l for a specified h
T_i	Slopes
Q	Sen's slope
SS_{res}	Residual sum of squares
SS_{tot}	Total sum of squares
y_i	Observed values
$\hat{y}_i$	Predicted values
$\bar{y}$	Mean of observed values
z_{score}	Z-score
C	Number of concordant pairs
D	Number of discordant pairs
LULC	Land use and land cover
UHI	Urban heat island
LST	Land surface temperature
SUHII	Surface urban heat island intensity
RF	Random forest
SVM	Support vector machine
IDW	Inverse distance weighting

AVHRR	Advanced Very High Resolution Radiometer
MODIS	Moderate Resolution Imaging Spectroradiometer
ETM+	Landsat's Enhanced Thematic Mapper Plus
TM	Thematic Mapper
OLI	Operational Land Imager
USGS	United States Geological Survey
GEE	Google Earth Engine
NDBI	Normalised Difference Built-up Index
NDVI	Normalised Difference Vegetation Index
NDBSI	Normalised Difference Bare Soil Index
NDWI	Normalised Difference Water Index
DT	Decision Tree
KNN	K-Nearest Neighbour
NIR	Near-infrared
R	Red
SWIR	Shortwave infrared
GE	Google Earth
ANN	Artificial Neural Networks
CART	Classification and Regression Tree
SC	Single Channel
SW	Split Window
ASTER	Advanced Spaceborne Thermal Emission and Reflection Radiometer
BI	Built-up intensity
MAE	Mean absolute error
RMSE	Root mean square error
APHRODITE	Asian Precipitation Highly Resolved Observational Data Integration Towards Evaluation
MK	Mann-Kendall
MMK	Modified Mann-Kendall
ITA	Innovative Trend Analysis
TSA	Theil-Sen Approach
STD	Standard deviation

MSE	Mean squared error
CI	Concordance Index
AAPRE	Average Absolute Percent Relative Error
DID	Department of Irrigation and Drainage
PA	Producer’s accuracy
UA	User’s accuracy
OA	Overall accuracy
DN	Digital number
TOA	Top of atmosphere

LIST OF APPENDICES

CHAPTER 1

INTRODUCTION

1.1 General Introduction

Urbanisation has contributed significantly to improving quality of life and reducing social inequalities by enhancing access to employment, education, and essential services. However, rapid and often unplanned urban growth also accelerates land use and land cover (LULC) changes, particularly in previously undeveloped areas. Such changes can lead to significant environmental consequences, notably the urban heat island (UHI) effect, where urban regions experience higher temperatures than their surrounding rural regions (Huang et al., 2023). This localised warming not only increases energy demand for cooling but also degrades air quality, raising the risk of respiratory and other health issues (Li, Zha, & Wang, 2020).

The severity of the UHI effect largely depends on urban composition and population density. Urban development typically replaces natural landscapes like water bodies and vegetation with buildings and infrastructure. These impervious surfaces, which have high thermal capacity and low albedo, absorb and store more heat, especially in city centres (Qin et al., 2024; Rao, Tassinari, & Torreggiani, 2023). Additionally, as urban populations increase, so do anthropogenic heat emissions from sources such as transportation, combustion processes, and electrical appliances (Mirabi & Davies, 2022). To mitigate UHI effects, a range of strategies have been proposed. For example, the use of smart appliances and highly reflective building materials can enhance energy efficiency in buildings. Additionally, incorporating green spaces and green roofs helps reduce urban heat stress and supports sustainable urban planning (Qin et al., 2024; Tahooni, Kakroodi and Kiavarz, 2023).

According to the Clausius-Clapeyron scaling, the intensity of extreme rainfall is anticipated to enhance by roughly 7% for every 1 ℃ rise in temperature. Numerous studies have identified the UHI effect as a contributing factor to increased convective rainfall, particularly during the late afternoon. During the daytime, heated air rises into the atmosphere, while cooler air moves to the urban areas and is heated again by the urban surfaces. In humid

tropical climates, this rising air condenses and releases latent heat, warming the atmosphere further and strengthening low-level convergence, ultimately enhancing precipitation (Li et al., 2020; Whitford et al., 2023; Siswato, Schrier and Hurk, 2022).

This study investigates the role of the UHI effect in amplifying short-duration extreme rainfall, focusing on rapidly urbanising areas such as Kuala Lumpur. Accelerated urbanisation alters LULC, elevates surface temperatures and influences local microclimates. This research integrated machine learning and remote sensing to classify LULC and employed detailed computations to estimate land surface temperature (LST), a key parameter for assessing surface urban heat island intensity (SUHII). Ground-based rainfall data were also utilised to analyse extreme rainfall patterns. By examining the relationship between UHI and short-duration extreme precipitation, this research aims to provide insights that are valuable for both scientific understanding and urban disaster risk management.

1.2 Importance of the Study

This research is critically important across several domains, including urban flood risk mitigation, sustainable urban planning and scientific contribution to UHI-rainfall dynamics. Understanding how the UHI effect contributes to short-duration extreme rainfall is essential for improving flood risk management in rapidly urbanising cities like Kuala Lumpur. Intense, short-duration rainfall events can quickly overwhelm urban drainage systems, resulting in flash floods. This study helps identify areas affected by UHI that may be more susceptible to such events. Furthermore, by quantifying changes in LULC, LST and SUHII, this research provides data-driven insights to support climate-responsive urban planning. Planners and policymakers can use these findings to implement climate-sensitive zoning, promote green infrastructure and regulate land conversion to minimise UHI impacts. Finally, there is still limited empirical research directly linking SUHII to extreme rainfall, especially using statistical tools like Kendall's Tau correlation (τ) and coefficient of determination (R^2) in tropical urban settings. This study adds to the scientific literature by applying robust analytical methods to explore the relationship between UHI and precipitation extremes.

1.3 Problem Statement

Over the past two decades, natural disasters have impacted more than four billion people worldwide and caused substantial economic losses. These problems are particularly severe in the developing world, where hundreds of thousands of lives are lost each year, due to insufficient funding, low public awareness and insufficient advanced disaster preparedness and control systems (Mizutori and Guha-Sapir, 2020; Watson, Gayer and Connolly, 2007). Among all natural hazards, floods are the most frequent, accounting for 3254 recorded events, approximately 44% of global disasters between 2000 and 2019. Asia has been disproportionately affected, with 41% of its landmass experiencing floods and 93% of the world's flood victims residing in the region (Mizutori and Guha-Sapir, 2020).

In Malaysia, urban centres like Kuala Lumpur are increasingly facing more intense and frequent flood events (Sapawi et al., 2023). Emerging studies suggest that this trend is closely linked to the UHI effect, which is exacerbated by rapid LULC changes due to urbanisation. These changes lead to elevated land surface temperatures and altered microclimatic conditions, potentially intensifying convective activity and extreme rainfall events.

Despite this, there is limited research that quantitatively investigates the relationship between SUHII and the short-duration extreme rainfall, particularly in tropical urban environments. Existing studies often lack high-resolution spatial and temporal analysis and underuse advanced tools such as machine learning, remote sensing and robust statistical methods.

This gap necessitates investigating the relationship between the UHI effect and short-duration extreme rainfall to support effective flood management and sustainable urban development strategies in Kuala Lumpur. Therefore, this study aims to examine this relationship by leveraging advanced machine learning, remote sensing and statistical methods.

1.4 Aim and Objectives

The primary aim of this research is to analyse the impacts of the UHI effect on the short-duration extreme rainfall in Kuala Lumpur. To effectively achieve this aim, several objectives are provided in the following:

1. To assess spatiotemporal variation in land use and land cover (LULC) through the analysis of Landsat satellite imagery, utilising Support Vector Machine (SVM) and Random Forest (RF) classifiers.
2. To estimate surface urban heat island intensity (SUHII) by quantifying land surface temperature (LST) derived from Landsat thermal imagery.
3. To analyse the spatiotemporal variation, diurnal distribution and trends in hourly extreme rainfall.
4. To examine the relationship between SUHII and hourly extreme rainfall using coefficient of determination (R^2) and Kendall's Tau correlation (τ).

1.5 Scope and Limitation of the Study

This study investigates the relationship between the UHI effect and short-duration extreme rainfall in Kuala Lumpur from 2007 to 2023. It utilises Landsat satellite imagery (2007, 2015, 2023) for LULC and LST analysis. Machine learning algorithms, including RF and SVM, are applied for supervised classification and the accuracy is evaluated using the Kappa coefficient and standard accuracy metrics. LST is quantified using established retrieval algorithms, while SUHII is derived from LST data. Ground-based rainfall data recorded at 15-minute intervals are obtained and processed to extract the 99th percentile threshold for identifying hourly extreme rainfall events. Missing rainfall values are interpolated using the inverse distance weighting (IDW) method and the Theissen polygons are generated to delineate rainfall station influence zones. Rainfall metrics, including the total number of hourly extreme rainfall events, the total of hourly extreme rainfall, and the mean of hourly extreme rainfall, are analysed for spatiotemporal variation, diurnal distribution, and trend patterns. The statistical relationship between SUHII and hourly extreme rainfall is evaluated using the coefficient of determination (R^2) and Kendall's Tau correlation (τ).

Several limitations are present in this research. Firstly, Landsat satellite imagery is susceptible to data gaps caused by heavy cloud cover, restricting the LULC, LST and SUHII analyses to only three years: 2007, 2015 and 2023. Secondly, cloud contamination in satellite images may affect the accuracy of LULC classification using machine learning algorithms.

Additionally, the performance of these algorithms is highly sensitive to parameter selection and data quality, potentially introducing uncertainty into the classification results. Moreover, although the rainfall dataset spans 17 years, missing data are filled using the IDW interpolation method, which may introduce spatial uncertainty. The limited availability of rainfall stations within the Kuala Lumpur region also constrains the spatial scope and resolution of the analysis. Finally, the correlation analysis focuses solely on surface temperature and rainfall intensity, without incorporating other atmospheric variables that could also influence precipitation behaviour.

1.6 Contribution of the Study

This study makes several important contributions to the fields of urban climatology, hydrology and sustainable urban planning. Firstly, it enhances the understanding of UHI–rainfall dynamics in tropical cities by analysing the relationship between SUHII and hourly extreme rainfall in a rapidly urbanising tropical environment. This addresses a critical knowledge gap, as few empirical studies have explored this relationship using long-term, high-resolution satellite imagery and ground-based rainfall data in Southeast Asia. Secondly, the study integrates machine learning algorithms (RF and SVM), remote sensing techniques and geostatistical methods to generate accurate and detailed LULC and LST maps. This comprehensive approach enhances the spatiotemporal precision of SUHII estimation. Thirdly, by identifying areas susceptible to intensified rainfall associated with UHI effects, the study offers practical insights to support urban flood mitigation strategies. The findings contribute to the development of climate-sensitive infrastructure planning and land use regulations that strengthen urban resilience. Lastly, this research presents a robust analytical framework for linking satellite-derived surface temperature metrics with high-frequency rainfall data using both Kendall's Tau and the coefficient of determination. This dual-method approach improves the reliability and interpretability of results for future climate impact assessments.

1.7 Outline of the Report

The structure of this report is organised into five comprehensive chapters, each detailing a critical component of the research:

Chapter 1 introduces the study by providing a general overview of urbanisation, the UHI effect and short-duration extreme rainfall. It also clearly defines the aim and objectives, the scope and limitations and potential contributions of the study.

Chapter 2 presents an in-depth literature review based on credible and up-to-date sources. It covers key areas such as satellite imagery, LULC classification using machine learning algorithms, LST and SUHII estimation, statistical analyses for rainfall and the relationship between UHI and short-duration extreme rainfall, and other relevant topics, ensuring a solid foundation for innovative and impactful research.

Chapter 3 details the research methodology and workflow. This includes a flowchart of the study framework, a description of the study area, data acquisition methods, and procedures for data processing and analysis, encompassing both remote sensing and statistical techniques.

Chapter 4 presents the results and provides a comprehensive discussion supported by appropriate justifications. Visual aids such as figures and tables are incorporated to enhance clarity and facilitate understanding of key findings.

Chapter 5 concludes the report by summarising the key findings, assessing the achievement of the research objectives and providing practical recommendations for future studies and potential real-world applications.

CHAPTER 2

LITERATURE REVIEW

2.1 Introduction

A comprehensive literature review is conducted to examine current trends and relevant methodologies for investigating the impact of the UHI effect on short-duration extreme rainfall. The review is based on a systematic search using keywords such as “urban heat island,” “land surface temperature,” “land use land cover classification,” “remote sensing,” “machine learning,” “short-duration extreme rainfall,” “rainfall analysis,” “statistical method,” “relationship,” and so on. The literature encompasses a wide range of approaches adopted by researchers, including satellite image acquisition and processing, LULC classification, LST and SUHII estimation, rainfall data collection and processing, rainfall trend analysis, and the evaluation of UHI's influence on short-duration extreme rainfall.

2.2 Remote Sensing for LULC and LST Data

Remote sensing is an advanced data collection technique that captures the energy radiated from matters on the Earth’s surface using satellites. In the study of UHI, LST is crucial for illustrating the heat exchange between the Earth and the air just above it, owing to the consistent dynamics of LST and air temperature (Reiners, Sobrino and Kuenzer, 2023). The importance of LULCC in UHI research is highlighted by Phan et al. (2024), suggesting that the urban thermal environment is affected by landscape metrics including the configuration, size, complexity and shape of LULC. Remote sensing of LST and LULCC offers multiple benefits such as efficient and continuous monitoring of spatiotemporal information (Gyimah, 2023), long-term data storage, dynamic observations of climate changes, and high temporal and spatial resolutions, which are not typically provided by weather stations (Shi et al., 2021).

Remote sensing captures LST and LULCC imagery via sensors mounted on satellites. Over the decades, numerous sensors with varying resolutions have been launched for Earth observation missions, including

Landsat's Thematic Mapper (TM), Enhanced Thematic Mapper Plus (ETM+), and Operational Land Imager (OLI); the Advanced Very High Resolution Radiometer (AVHRR); the Visible Infrared Imaging Radiometer Suite; Along Track Scanning Radiometers; the Advanced Spaceborne Thermal Emission and Reflection Radiometer (ASTER); and the Moderate Resolution Imaging Spectroradiometer (MODIS). Since Landsat and AVHRR were launched in the 1980s, they offer longer time series information compared to other satellites (Reiners, Sobrino and Kuenzer, 2023). According to Phan et al. (2024), satellite data from MODIS and Landsat are preferred by the majority of researchers, while other sensors, due to their shortages in temporal or spatial resolutions, serve as complementary data sources to address limitations faced by the primary sensors.

The quality of remotely sensed imagery is susceptible to several factors, including sensor angle, complexity of LULC pixels, variation in urban surface and atmospheric attenuation or cloud cover. The first three limitations can be effectively tackled by using high spatial resolutions satellites like Landsat to produce finer images (Shi et al., 2021). The cloud cover issue is usually mitigated by either adopting Passive-Microwave-derived-LST or regressing on-site LST with inputs like the elevation, albedo and land use indices. The temporal aggregation method can be employed for polar-orbiting satellites that possess a consistent revisit period. It is not recommended to use cloud-free images for LST measurement as they can lead to ambiguous results when comparing to cloudy scenarios (Reiners, Sobrino and Kuenzer, 2023).

2.2.1 Landsat

The Landsat series comprises eight satellites operating in sun-synchronous orbits, each equipped with high spatial resolution sensors. For instance, Landsat 4 and 5 are equipped with 120 m TM sensors, Landsat 7 with 60 m ETM+ sensors, and Landsat 8 and 9 with 100 m Thermal Infrared (TIR) sensors. Although Landsat series has a comparatively long revisit period of 16 days, the demand for its data has increased due to the complexity of urban landscapes. Landsat's high spatial resolution sensors are effective in identifying LST in both urban and non-urban regions (Reiners, Sobrino, and Kuenzer, 2023).

In the UHI research conducted by Huang et al. (2023), Landsat 8 satellite images on 16 November 2013 and 24 December 2021 were obtained from the United States Geological Survey (USGS) Earth Explorer online platform. These dates represent the oldest and the newest data captured by Landsat 8 for the KL area during the study period. Landsat 8 imagery was processed using QGIS to obtain LST and comparisons between urban LST and the reference area LST were made to identify the UHI impacts.

Rao, Tassinari and Torreggiani (2023) extracted the temporal data during the summer months from Landsat 5 (1991-2011) and Landsat 8 (2013-2021) to compute LULC indices and LST for both urban and non-urban regions. Over the 31-year period, data from 1992, 2000, 2006 and 2012 were missing due to problems like extreme weather, cloudy conditions and data unavailability. Google Earth Engine (GEE) was selected to analyse the abundant satellite images using a cloud platform in conjunction with spectral bands and proper algorithm combinations.

Satellite data remotely sensed by the 30 m OLI and TIR sensors on Landsat 8 for every July from 2014 to 2021 were used by Rendana et al. (2023) to compute LST and classify land use in the Hulu Langat area. The data were retrieved from the USGS Earth Explorer online platform and processed using ArcGIS Ver.10.

Remote sensing images from 2006 to 2021 were gathered by Tanoori, Soltani and Modiri (2024) from the Landsat series to construct land use maps and produce LST for analysing the effect of LULCC on LST. ArcGIS 10.2.1, ENVI 5.3.1 and Fragstats were applied for managing the satellite data whilst Python was used for data manipulation and modelling. Radiometric calibration and FLAASH atmospheric correction were performed to reduce light and atmospheric influences on the image quality, ensuring accurate LST estimations.

In the study by Al-Taei, Alesheikh and Boloorani (2023), the data collection was carried out by retrieving Landsat 7 (2000-2012) and Landsat 8 (2013-2022) imagery. Annual satellite imagery was then processed to acquire necessary variables, such as spectral bands, land use indices, and textural features. To ensure high quality analysis, the median image with less than 50%

cloud cover between April and September of each year was chosen. Data lost in Landsat 7 imagery were amended before extracting features.

2.2.2 MODIS

The MODIS sensors are onboard polar-orbiting satellites, offering outstanding daily temporal resolution by compromising on spatial resolution (1 km). Compared to Landsat, MODIS is more suitable for time series studies on both local and global scales due to its large swath width and high temporal resolution (Reiners, Sobrino and Kuenzer, 2023).

According to the methodology applied by Li, Zha and Wang (2020), global LST data with temporal coverage at 8-day intervals and a spatial resolution of 1 km were collected from the MODIS V6 MYD11A2 product, including imagery captured both during the day and at night from 2003 to 2013.

In the study by Zhou et al. (2014), satellite data from 2003 to 2011 were acquired from the Aqua MODIS V5 MYD11A2 product, which offered 8-day temporal coverage and a spatial resolution of 1 km. By applying the generalised split-window algorithm, these data were obtained from cloud-free observations with 99% confidence that captured at 1:30 and 13:30 local solar time. Notably, MODIS V5 offered higher accuracy in LST analysis in most cases.

LST data from the Terra MODIS MOD11A2 product were retrieved by Moazzam, Kim and Lee (2024), for the period from 2003 to 2020, observed daily at 10.30 am and 10.30 pm. The 8-day MODIS LST product was preferred over the daily LST product to mitigate the impact of data loss due to heavy cloud cover and to minimise the computational workload for data processing. The generalised split-window algorithm was also employed in the LST retrieval process, such as cloud removal and adjustments for radiation, water vapor and temperature, improving the accuracy of LST analysis.

MODIS LST data spanning from 2008 to 2018 were obtained by Tang et al. (2022) from the MOD11A1 dataset. Similar to the datasets mentioned above, the MODIS MOD11A1 product offers the same spatial and temporal resolution. Data processing was carried out using the MODIS

Reprojection Tool, which involved raster clipping, projection conversion to WGS-84 and data conversion to TIFF format.

Additionally, daily LST satellite imagery with a 1200m spatial resolution from 2000 to 2022 was acquired by Rees, Baidy and Belenok (2024) from the MODIS V6.1 MOD11A1 product. These data were collected during the day and night with surface temperature bands and quality control assessments. The split-window algorithm was utilised to produce cloud-free LST pixel values. In areas with overlapping pixels, values are averaged with appropriate weights applied to the overlaps. A script was created using GEE and the study area shapefile with the relevant period was filtered from the MODIS/006/MOD11A1 LST_Day_1km dataset to compute LST values.

2.3 LULC Classification Using Machine Learning Algorithms

In the LULC classification, various land use indices and machine learning algorithms are utilised to improve both efficiency and accuracy. This section commences with a discovery of commonly applied land use indices, such as the Normalised Difference Built-up Index (NDBI), Normalised Difference Bare Soil Index (NDBSI), Normalised Difference Water Index (NDWI) and Normalised Difference Vegetation Index (NDVI). These indices highlight specific land features through the use of multi-spectral bands. Multiple machine learning algorithms, including K-Nearest Neighbour (KNN), Decision Tree (DT), SVM, and RF, have been proven effective in classifying LULC. The application and performance of these algorithms are clarified by case studies, providing strong justifications for selecting appropriate methods for this research.

Machine learning is a derivative of artificial intelligence, leveraging input data and statistical algorithms to emulate human learning processes and execute sophisticated missions without explicit orders. Machine learning can be categorised into four main learning methods, with supervised and unsupervised learning being the most commonly applied in UHI studies. In supervised learning, the training process requires labelled datasets, consisting of both input data and corresponding output data, to analyse and recognise relationships between them. The finalised model can be utilised to address classification and regression problems. Conversely, unsupervised learning

excels in data aggregation and clustering which do not require the labelled data for training. It can organise unlabelled datasets into distinct groups via the identification of patterns, similarities and structures among variables (Ahorloo et al., 2024).

The process of LULC classification follows a structured methodology that typically involves seven distinct stages as identified by Lu and Weng (2007): choosing the appropriate remotely sensed data, defining the classification system and selecting training samples, pre-processing the data, extracting and selecting pertinent features, applying the classification method, executing post-classification processing and assessing accuracy. This methodical sequence guarantees consistency and enhances the reliability of land cover information derived from satellite imagery.

2.3.1 Normalised Difference Vegetation Index (NDVI)

NDVI is a common indicator of green vegetation cover and relies on data captured by remote sensing. The concept of NDVI is based on the absorption of sunlight in the photosynthetically active radiation range and the reflection of near-infrared radiation. NDVI is computed by dividing the difference between the near-infrared (NIR) band and red (R) band by their sum (Vilcins et al., 2022). NDVI values represent green vegetation density, where a value of 1 indicates the densest vegetation and -1 represents extreme bareness with only water, rock and sand. Values between 0.2-0.3 correspond to grassland while 0.6-0.8 represent rainforests (Gascon et al., 2016).

Several shortcomings of NDVI have limited the accuracy of detecting green cover. NDVI is only capable of detecting the presence of green vegetation rather than distinguishing between different flora species. In addition, it cannot precisely measure the coverage area of withered plants and bare land. The discrepancies between the spatial resolution of NDVI and study area may lead to inaccuracies. Lastly, NDVI cannot provide information on whether the vegetation locates in public or private areas (Vilcins et al., 2022).

2.3.2 Normalised Difference Built-up Index (NDBI)

By using satellite data, NDBI detects the density of built-up areas based on their distinct spectral response. Built-up surfaces reflect more shortwave

infrared (SWIR) radiation than NIR. NDBI is computed as the ratio of the difference between SWIR and NIR bands to the sum of these two bands. NDBI values range between -1 and 1, where positive values typically represent built-up areas, while negative values correspond to non-built-up regions. (Dammayatri, Susantoro and Wikantika, 2023).

2.3.3 Normalised Difference Water Index (NDWI)

NDWI is applied to indicate the presence of open water bodies in the satellite data. It distinguishes water bodies by leveraging the characteristic of water that tends to strongly absorb NIR compared to other materials such as vegetation and soil. In other words, water bodies can be detected by satellite sensors due to its low NIR reflectance. NDWI is formulated as the ratio of the difference between GREEN and NIR bands to the sum of these two bands. GREEN band is used to capture reflected green light from water bodies, enhancing water feature detection. NIR band can easily detect and exclude both soil and vegetation, highlighting the water feature in the satellite imagery. The presence of water bodies results in the positive NDWI value with the maximum of 1 while values from 0 to -1 indicate that the area is occupied by vegetation and soil. Even though NDWI is widely utilised in the investigation of water quality, it can barely differentiate between chlorophyll α and suspended solids in water bodies (McFeeters, 1996).

2.3.4 Normalised Difference Bare Soil Index (NDBSI)

NDBSI is a measure of the bareness of an area, taking advantage of RED, BLUE, NIR and SWIR wavelengths captured by remote sensing. It was firstly designed for the forest management to inspect forest density. The NIR and BLUE bands are selected to detect vegetation cover while the RED and SWIR bands distinguish the soil mineral from other matters. When the NDBSI value is closer to 1, a larger scale of bare soil and impervious layers exist in the study area. Conversely, the density of vegetation and pervious layers increases with a NDBSI value closer to -1 (Polovina et al., 2024).

2.3.5 Supervised and Unsupervised Classification

Supervised classification represents an analyst-directed approach to LULC mapping where human expertise plays a central role in defining spectral signatures for land cover classes. This technique employs a three-phase process: training (identification of representative pixel samples), signature extraction (statistical characterisation of spectral values) and classification (assignment of all pixels to defined classes). Through this methodology, the analyst maintains control over category definition and classification outcomes. Several algorithms have been developed for supervised classification, each with distinct mathematical foundations, such as SVM, DT and maximum likelihood classification (Madariya, Pandey and Sharma, 2022).

Unsupervised classification represents an automated approach to image segmentation where algorithms identify natural spectral groupings without prior definition of land cover categories. This process involves spectral clustering using statistical techniques, followed by analyst interpretation to assign meaningful land cover labels to the resulting clusters. This approach operates independently of training data, relying instead on the spectral properties of image pixels to determine class boundaries. Two primary algorithms dominate unsupervised classification applications, including Iterative Self-Organizing Data Analysis Technique and K-means (Madariya, Pandey and Sharma, 2022).

In the study by Mohd Hasmadi, Pakhriazad and Shahrin (2009) in Selangor, Malaysia, supervised classification achieved an overall accuracy of 90.28% with a Kappa coefficient of 0.86, indicating strong agreement between the classified map and ground truth data. In contrast, unsupervised classification, which clusters image pixels based solely on their spectral properties without prior training, yielded lower overall accuracy at 80.56% and a Kappa of 0.73. The higher performance of the supervised method is attributed to its ability to incorporate prior knowledge and carefully selected training samples, which help the classifier distinguish between classes that may have overlapping spectral signatures.

A second case study from Afghanistan compared both classification approaches over multiple time periods (2000, 2013, and 2020) using Landsat imagery. The supervised model, implemented with advanced machine learning

techniques such as SVM, consistently achieved higher overall accuracies of 94.23%, 90.09%, and 88.18% for the respective years, compared to unsupervised methods, which reported accuracies around 89%, 82.5%, and 84.26%. These differences underscore that supervised classification tends to be more reliable, particularly in heterogeneous and complex landscapes, because it leverages expert knowledge and carefully delineated training areas (Doost and Yaseen, 2023).

2.3.6 Support Vector Machine (SVM)

As noted by Awad and Khanna (2015), SVM is a commonly employed supervised machine learning model for applications like weather forecasting and image classification, due to its powerful generalisation capability and ability to achieve optimal global solutions. SVM operates on a simple principle: it determines an optimal hyperplane that perfectly separates different classes with maximum margins in a n-dimensional space. For problems involving non-linearly separable data, kernel functions are applied to obtain higher-dimensional data, facilitating the classification process of a linear separator between each class.

According to Al Kafy et al. (2021), LULC classification of Landsat satellite data from 1999, 2009, and 2019 was conducted using a SVM algorithm embedded in ENVI 5.3. The classification focused on four major LULC categories: urban areas, bare soils, vegetation cover, and water bodies. The SWM algorithm adopted the kernel type of radial basis function with certain parameters, including a penalty parameter of 120, a gamma value of 0.07, a classification probability threshold of 0.05 and a pyramid level that remained zero. Areal and spectral profiles of the imagery were analysed to acquire additional training data and background information. LULC maps were produced by taking approximately 45 samples for each year's LULC category. The performance of the SVM-based LULC classification was evaluated based on four metrics: kappa statistics, overall accuracy, user accuracy and producer accuracy. These evaluations were based on 200 and 150 randomly selected ground data and Google Earth (GE) images for each map. Consequently, the results indicated that the SVM algorithm was highly

effective in LULC classification, achieving overall accuracy rates of 86.04%, 84.82% and 87.23% for the three years, respectively.

Similar procedures were adopted by Edan, Maarouf and Hasson (2021) to classify LULC using satellite images from 2000, 2010 and 2020. Prior to the LULC classification, Landsat 8 OLI images (bands 1 to 7) and Landsat 5 TM images (bands 1 to 5 and 7) were processed through band composition. LULC maps were generated by collecting roughly 200 signatures for each LULC class in each year. To assess the performance of the SVM algorithm, the four assessments metrics were adopted, alongside 300 ground-truthing data obtained from global positioning systems and GE. As a result, the SVM algorithm demonstrated high accuracy in LULC classification, attaining a kappa coefficient above 0.83 and an overall accuracy exceeding 0.94 across all studied years.

2.3.7 K-Nearest Neighbours (KNN)

KNN is a simple and flexible supervised algorithm suitable for both regression and classification tasks. In contrast to conventional regression models that define relationships between input and output variables using mathematical formulae, KNN simplifies this process due to its non-parametric characteristic. It finds the “k” nearest data points or neighbours considering criteria such as Euclidean distances to make a prediction. The average values of nearest neighbours are computed to predict the new data point in regression models while the classification of the new data point is based on the majority class among its neighbours (Ran et al., 2024; Verma et al., 2023).

In the study by Athukorala and Murayama (2021), the KNN algorithm was used to perform LULC classification in R software. Landsat 5 (bands 2-4) and Landsat 8 (bands 3-5) imagery were prepared as inputs for LULC map analysis. The classification was based on four LULC groups including bare land, water, green space and impervious surface. Automatic sampling with 400 points per year facilitated the production of high quality LULC maps. After the initial classification, amendments for classified LULC maps were conducted, such as the hybrid classification method and majority filter, to rectify random misclassifications like the salt and pepper noise. KNN-classified LULC maps were observed to be the most accurate among other

algorithms, including ANN, SVM, and RF, achieving an accuracy score of 90% or higher.

2.3.8 Decision Tree (DT)

DT is a non-parametric supervised learning model that made up of edges and nodes, systematically and continuously sorting data into different classes based on certain criteria. The operational process commences from the root node splitting data into edges based on the computation of entropy and information gain or the Gini index. Subsequent test nodes receive incoming edges (input) from previous nodes and continue this splitting process until reaching the leaves node which provides the final decision. One of the advantages of DT is its ability to process large datasets for classification and regression models within a short period of time, obtaining high accuracy results. DT is also a flexible solution used to address linear or non-linear models with little to no data normalization needed. Lastly, DT has a simple structure that can be interpreted easily compared to other algorithms like Artificial Neural Networks (ANN). However, DT tends to overfit when dealing with complex relationships between features and attenuation of DT performance could also result from variations in the dataset (Talekar and Agrawal, 2020).

Four LULC groups, such as built-up, vegetation, cropland and water were used by Mohammad et al. (2022) for LULC classification spanning from 1995 to 2020 at five-year intervals. The Classification and Regression Tree (CART), a subset of the DT algorithm, was used to create LULC maps via GEE while multiple Landsat imagery bands were applied as nodes for classification. To identify the four LULC features, various band combinations of Landsat imagery were acquired from over 200 training samples for the respective features. The background GE images were employed as reference data to examine the training points. The obtained sample points were separated into training sets (70%) and validation sets (30%) using random sampling approaches. Two reliable image classification assessment methods, including the confusion matrix and Kappa statistics, were applied to measure the CART performance. The results indicated that the CART algorithm was highly accurate for LULC classification, with a mean accuracy score of 0.91 over the study periods.

2.3.9 Random Forest (RF)

According to Fawagreh, Gaber and Elyan (2014), RF is an ensemble learning algorithm that integrates multiple DT using the random selection of features and Breiman's bagging sampling method. Each DT is generated from repetitively chosen samples of a training set, which helps minimize variance in the dataset. The predictions for the class label of a given instance are simultaneously executed and a vote is cast for the predicted class label by each DT. The final output is determined by selecting the class label that obtains the highest votes among all trees. RF is a popular choice for both classification and regression due to its superior performance in handling intricate relationships between variables and capability against overfitting. In addition, RF is relatively simpler, faster and more precise compared to other boosting or bagging models.

The LULC classification of Landsat data into different land use types was executed using the RF algorithm through the GEE platform. The land use types included dense vegetation, bare land, water, sparse vegetation and urban areas. The image pixels were carefully digitised for each class to train the RF algorithm. A minimum of 50 training samples per class were created, resulting in a total of 1045 training samples proportionally allocated based on the frequency of occurrence of each LULC category in the study area. 100 decision trees were employed to form the RF model and 376 more samples were generated to examine the RF performance. The classification process did not rely simply on the spectral bands from Landsat imagery, but also leveraged various land use indices, such as Band Ratio for Built-up Area, Index-Based Built-up Index, Normalised Difference Water Index and so on, to precisely distinguish between similar classes. The Kappa coefficient, overall accuracy, producer's accuracy and user's accuracy were used to assess the LULC maps (Rees, Baidy and Belenok, 2024).

The RF algorithm was also adopted by Liu, An and Ming (2024) to classify LULC maps. By using GEE, the surface reflectance data were retrieved from Landsat imagery. A median algorithm was used, combining all images into a single composite and calculating the median value for each pixel. Additional layers, such as NDBI, NDVI, Normalised Difference Bareness

Index and Digital Elevation Model (DEM), were mosaicked to add details and texture to the classification. A total of 796 training data were gathered to train the RF model in classifying the land use into five categories, encompassing bare land, water bodies, forests, built-up areas and cropland. To examine its performance, a random sampling approach was applied to collect 200 samples for each LULC group. Consequently, the RF model demonstrated high accuracy in LULC classification, achieving an overall accuracy of more than 0.86 and Kappa coefficients exceeding 0.83.

2.4 LST Estimation

LST is a fundamental parameter for understanding surface and atmosphere interactions, urban heat island effects and various climatic processes. Retrieving LST from thermal infrared satellite data has evolved into a vital research area, with several algorithms developed to meet diverse environmental conditions and sensor specifications. Two prominent methods include the Single Channel (SC) and the Split Window (SW) algorithms.

2.4.1 SC Algorithm

The SC method has emerged as a practical and efficient approach for retrieving LST from thermal infrared satellite data, particularly Landsat-8 TIRS Band 10. Unlike the SW techniques, the SC method requires fewer input parameters, primarily the land surface emissivity, the effective band wavelength and atmospheric water vapor content. This reduced dependency simplifies the retrieval process, especially in areas where detailed atmospheric profiles are unavailable (Garcia and D**í**az, 2021; Cristóbal et al., 2018).

García and Díaz (2021) compared several LST algorithms and found that the SC approach delivered robust and reliable LST estimates in a highly polluted urban environment in Granada, Spain. Their findings highlighted that SC algorithms are particularly effective in urban settings, offering consistent performance across multiple temporal scales.

Building on this, an improved single-channel method was proposed by Cristóbal et al. (2018), which incorporated not only water vapor content but also near-surface air temperature into the retrieval process. By integrating near-surface air temperature, the improved SC method reduced uncertainties

inherent in atmospheric correction, yielding an overall error of approximately 1 K and a bias of -0.5 K when compared to on-site observations. This enhancement demonstrated a significant performance improvement over earlier models that relied solely on water vapor.

Furthermore, Maithani et al. (2022) applied the SC algorithm to Landsat thermal datasets for the years 2000, 2010 and 2019 in an urban context in India. Their approach, which served as a precursor to machine learning simulations of future urban temperature patterns, yielded an average root mean square error (RMSE) of 1.9 ℃ during summer and 1.9 ℃ during winter when validated against downscaled MODIS LST products. These results underscore the SC method's operational reliability and its suitability for long-term urban climate studies.

2.4.2 SW Algorithm

The SW algorithm is a widely used approach for retrieving LST from remote sensing data, particularly by leveraging two adjacent thermal infrared channels. One key feature is its ability to mitigate atmospheric effects, especially water vapor absorption, by using a nonlinear combination of brightness temperatures measured at wavelengths around 11 and 12 μm. For instance, in a notable case study using Landsat 8 TIRS data, the algorithm was developed using an extensive simulation dataset (over 350,000 scenarios) that varied atmospheric conditions, land surface emissivity and true LST. The study divided column water vapor into several sub-ranges and derived tailored coefficients for each range, resulting in an LST RMSE of less than 1.0 K (Du et al., 2015).

The radiance-based split-window algorithm (RBSWA) used with MODIS data demonstrated notable improvements over conventional brightness temperature-based methods. Simulation analyses reported an RMSE of only 0.5 K with an improvement of about 0.3 K compared to traditional methods. When applied to real MODIS data covering the continental United States, the RBSWA produced a mean RMSE of 1.33 K and validation against Surface Radiation Budget and ASTER LST products further confirmed its reliability (Wang et al., 2019).

An investigation of SW algorithms for Landsat 9 TIRS-2 data further highlights that while different candidate methods exist, those that incorporate adjustments for atmospheric water vapor and land surface emissivity tend to reduce overall bias and error by up to 1.30 K and 1.0 K, respectively (Su, Meng and Sun, 2024).

2.5 SUHII Estimation

The SUHII was used to measure the magnitude of UHI by subtracting the average urban LST with the average rural LST (Li, Zha and Wang, 2020; Sun et al., 2018). Several steps were taken by Zhou et al. (2014) to define the urban and rural areas. Firstly, a 1 km x 1 km moving window approach was applied to create a built-up intensity (BI) map, which was classified into high-BI and low-BI regions using a 50% threshold. Then, the high-BI regions were combined and surrounded by a 2 km buffer zone that encompassed the nearest and scattered high-BI patches within the urban boundary. The area inside the buffer zone was treated as an urban area while the buffer zone with an equivalent area to the urban area, excluding waterbodies, was designated as a rural area.

According to Huang et al. (2023) and Rendana et al. (2023), the UHI estimation was performed by calculating the ratio of the difference in LST between urban and reference areas to the LST of the reference area. The study areas were then classified into five different classes based on the UHI value, including very weak UHI effect (< 0.07), weak UHI effect (0.08 - 0.12), moderate UHI effect (0.13 - 0.17), strong UHI effect (0.18 - 0.20) and very strong UHI effect (> 0.21) (Huang et al., 2023).

Two different methods were applied by Rao, Tassinari and Torreggiani (2023) to measure UHI. The SUHII Type 1 method, also employed by Rees, Baidy and Belenok (2024), computed the difference between the LST of the research area and the mean LST of green spaces. The Type 2 method calculated the difference between the study area LST and predefined UHI threshold. Unlike the Type 1 method, a threshold value was defined in the Type 2 method to identify UHI and non-UHI regions without considering the influence of LULC on temperature changes across the study area. Positive SUHII values indicated UHI presence, with categories ranging from low

SUHII (0 - 2.0), moderate SUHII (2.0 - 4.0), high SUHII (4.0 - 6.0), very high SUHII (6.0 - 8.0) to extremely high SUHII (> 8). Statistical comparisons, including standard deviation, RMSE and mean absolute error (MAE), showed that Type 1 was more precise for the estimation of SUHII in areas with noticeable anthropogenic activities as it emphasised distinguishing land use, while Type 2 provided a broader but less precise estimation.

Due to varying weather conditions throughout the year, the comparison of several satellite images within the same year to assess a region's thermal properties may yield inaccurate results. Thus, a normalised method was employed to quantify the UHI across different seasons within the same year, accounting for atmospheric fluctuations by finding a ratio of the difference between LST and mean LST to the standard deviation (Al Kafy et al., 2021; Rahaman et al., 2022).

2.6 Acquisition and Pre-Processing of Short-Duration Extreme Rainfall Data

High-resolution hourly rainfall data from 1960 to 2020 were collected by Yan et al. (2024) from various sources to analyse rainfall patterns in the Great Bay Area. For instance, Integrated Multi-satellite Retrievals for Global Precipitation Measurement Mission Final Run Version 07, Multi-Source Weighted-Ensemble Precipitation product, ERA5-Land reanalysis data of the European Centre for Medium-Range Weather Forecast and local weather stations. Despite adjustments made to merge data from different sources, discrepancies were still detected when comparing gridded rainfall data to actual hourly rainfall observations. This problem was addressed by applying a Random Forest-based Merging Procedure to enhance data accuracy. Then, maximum rainfall intensities over multiple time durations were retrieved for each year. "Rainfall extremes" were considered as heavy rainfall with intensities exceeding the 98.75^{th} and 99.7^{th} percentiles.

In addition, temperature and precipitation data from 1998 to 2015 were acquired by Oh et al. (2022) from the Asian Precipitation Highly Resolved Observational Data Integration Towards Evaluation of water resource (APHRODITE) project. This data was captured by approximately 5000 to 12000 weather stations that provide daily high-resolution (0.25°)

climate information across Asia. It has been widely recognised that APHRODITE data is well-suited for understanding Asian monsoon rainfall variability and performing rainfall forecasting. Rainfall estimates at 3-hour intervals were also extracted from the Tropical Rainfall Measuring Mission (TRMM) products to evaluate precipitation changes.

A Malaysian hourly rainfall dataset, combined and checked for quality as part of the Global Sub-Daily Rainfall dataset was employed. Rainfall data from 1981 to 2011 were extracted from rain gauge stations around Kuala Lumpur with over 80% data availability. To ensure the independence of precipitation events, only the highest hourly rainfall intensity for each day was selected. Hourly rainfall intensities were added up to obtain daily rainfall intensities. Instead of using wet-hour/day data, the top 5% (95^{th} percentile) average hourly/daily rainfall intensities were identified for trend analysis, as increases in rainfall intensity do not strongly correlate with increases in wet-day percentiles (Li, et al., 2020).

To explore the relationship between urbanisation and extreme rainfall in Paris and Shanghai, the top 1% (99th percentile) and top 0.28% (99.72nd percentile) of hourly and daily rainfall throughout the year were analysed. The most extreme 1% of hours or days were represented by the 99^{th} percentile, while the most extreme 24 hours in a year were represented by the 99.72^{nd} percentile. The intensity, frequency and total amount of rainfall exceeding these percentiles were observed (Steensen et al., 2022).

In the study by Li, Zha and Wang (2020), rainfall data from global megacities with daily precipitation greater than 50 mm were excluded from the correlation estimation. This is because such heavy rainfall events generally result from extreme weather conditions like low vortex, typhoons and cyclones, hampering the observation of UHI impacts on precipitation. Mamun, Salleh, and Noor (2018) defined short-duration rainfall in Klang Valley, Malaysia, as precipitation lasting between 15 minutes and 18 hours, but not exceeding 24 hours.

2.7 Inverse Distance Weighting (IDW)

IDW is one of the most widely applied spatial interpolation techniques in hydrology and meteorology, particularly for estimating rainfall at ungauged

locations. Introduced by Robertson (1967), IDW operates on the principle that points nearer to the prediction location exert a stronger influence on the estimated value than those farther away. This method computes a weighted average of observed values, with weights inversely proportional to a power of the distance from the target location (Tiwari, Jha and Sivakumar, 2019; Benmoshe, 2025).

Despite its simplicity, IDW continues to demonstrate strong performance in real-world applications. For example, Wimala et al. (2025) evaluated multiple interpolation methods using rainfall data from 20 automatic telemetering stations and found that IDW produced the lowest MAE and RMSE, making it the most accurate method in their study region. While Kriging and Co-Kriging captured spatial variability more effectively in theory, they introduced greater estimation errors when data were limited, highlighting their reliance on a denser station network. Mremi et al. (2025) also utilised IDW in GIS to model spatial rainfall distribution, affirming its practical utility in data-scarce environments. Similarly, Fung et al. (2022) evaluated IDW as a univariate spatial interpolation method in Peninsular Malaysia and found that IDW with a power of two ($p = 2$) outperformed Ordinary Kriging, yielding a lower MAE (67.8), RMSE (91.2), and higher R^2 (0.540). However, the study also noted that IDW was outperformed by the multivariate Multi-scale Geographical Weighted Regression method, which accounted for additional variables like elevation and location, offering better spatial accuracy by adjusting spatial bandwidths.

2.8 Rainfall Analysis

Rainfall analysis plays a crucial role in understanding climate variability, especially in hydroclimatic studies. Among various statistical methods available, the Mann-Kendall (MK) and Modified Mann-Kendall (MMK) test, Sen’s slope estimator, Innovative Trend Analysis (ITA) and spatiotemporal comparative analysis are widely used due to their robustness and effectiveness. These methods enable researchers to assess the presence, direction and magnitude of long-term trends in rainfall data across various temporal and spatial scales.

2.8.1 Mann-Kendall (MK) and Modified Mann-Kendall (MMK) Tests

The MK test is widely adopted non-parametric approach to assess monotonic trends in time series data without requiring the data to follow a specific distribution. Its non-parametric nature makes it particularly suitable for analysing climate extremes, as it can handle missing data and is less sensitive to outliers. However, the original MK test can be significantly influenced by autocorrelation in the data, which may lead to overestimation of trend significance. To address this limitation, the MMK test, proposed by Hamed and Ramachandra Rao (1998), incorporates adjustments to account for serial correlation by modifying the variance of the test statistic. This is achieved through techniques such as pre-whitening, block bootstrapping and variance correction. The test compares the null hypothesis (H_0), which posits no trend, against the alternative hypothesis (H_1), suggesting a significant increasing or decreasing trend. Statistical significance is typically evaluated at a 95% confidence level ($\alpha = 0.05$). A trend is considered significant when the p-value is less than 0.05. The Kendall rank correlation coefficient (τ), which ranges from -1 to +1, indicates the direction and strength of the trend. Positive values suggest an increasing trend, while negative values indicate a decreasing trend. Nevertheless, recent studies emphasise that even modified versions of the MK test may not fully eliminate the impact of long-term autocorrelation, highlighting the importance of careful methodological choice when interpreting trend significance (Sharma et al., 2024; Ng et al., 2024; Mekuria, Demissie and Feyessa, 2025; Miniandi et al., 2024).

2.8.2 Sen's Slope Estimator

In addition to the MK test, the Sen's slope estimator (Sen, 1968) is commonly used to quantify the magnitude of a trend. It is a non-parametric method that calculates the median of all pairwise slope values in the dataset, providing a robust estimate that is less sensitive to outliers. Sen's slope is typically applied in conjunction with the MK test, as it offers a more precise measure of trend magnitude once a significant trend has been identified. This combined approach has been adopted in various recent studies, such as Kenabatho (2025), Sharma et al. (2024), Ng et al. (2024), Mekuria, Demissie and Feyessa (2025), and Miniandi et al. (2024). While other trend detection methods, such as

Spearman's Rho, linear regression, and wavelet analysis, are also utilised in the literature, the MK test coupled with Sen's slope estimator remains the preferred choice in hydroclimatic studies due to its robustness, minimal data assumptions and effectiveness in handling non-normally distributed or incomplete data (Mekuria, Demissie, and Feyessa, 2025).

2.8.3 Innovative Trend Analysis (ITA)

Sen (2012) suggested the ITA approach to measure trends in hydro-meteorological time series data, including precipitation and temperature. ITA visualises data trends based on graphical and non-parametric method, comparing two halves of the dataset. This method was adopted by Ahmed et al. (2022), Doiphode and Swami (2024), Wang et al. (2020) and Deopa et al. (2024) for the temporal rainfall study. Precipitation data were separated into two equal halves and assigned to both horizontal and vertical axes in ascending order. The time series shows an increasing trend when the majority of data points fall above the 1:1 line and vice versa. No significant trend is expected if data points are evenly scattered along the 1:1 line.

In the case of without a clear monotonic trend, Wang et al. (2020) categorised the data into low, medium and high rainfall classes, utilising 10^{th} and 90^{th} percentiles as boundary lines. The results of the ITA are assessed by both the MK test and the Theil-Sen Approach (TSA). The MK test detects the sign and significance of trends, while the TSA investigates the magnitude of trends and is more resilient to outliers. The presence of the serial correlation in a time series could interfere with the trend test results. Hence, lag-1 serial correlation (r_1) was computed for the time series data. The series is pre-whitened before employing trend analysis methods, only if r_1 is significant.

2.8.4 Spatiotemporal Comparative Analysis

To evaluate temporal and spatial changes in extreme rainfall, Miniandi et al. (2024) divided the analysis into two periods: 2000-2010 (early period) and 2013-2022 (recent period). Spatial distribution maps were generated for both percentiles across the two time periods, revealing a noticeable expansion in the coverage and intensity of extreme rainfall across Kuala Lumpur. These

comparisons demonstrated a spatial shift and intensification of rainfall extremes in tandem with urban expansion over time.

Lee et al. (2023) conducted a spatiotemporal analysis of probability precipitation in South Korea across four periods (1981-2020, 1991-2020, 2001-2020, 2011-2020) using data from 61 rainfall stations, focusing on probability rainfall, intensity and duration for 10-, 20-, 50-, and 100-year return periods and 1-, 2-, 6-, and 24-hour durations. Spatial distributions, mapped via kriging interpolation in ArcGIS, revealed that the southeast inland area consistently exhibited lower probability precipitation and intensity anomalies, while the northeast, northwest and south coast regions showed higher values. Decadal rainfall analysis, focusing on nine representative stations in the northern (Ganghwa, Daegwallyeong and Seoul), southwest inland (Jeongeup, Gwangju and Namwon) and southeast inland (Miryang, Daegu and Yeongcheon) regions, showed that the 2011-2020 decade had notably lower annual rainfall in the southwest and northern inland areas (absolute differences > 115 mm) compared to 2001-2010 but stable precipitation in the southeast inland area (differences ≤ 15 mm). These findings, supported by box plots and anomaly maps, highlight a strong correlation between annual precipitation trends and spatiotemporal variations in probability rainfall factors.

The rainfall analysis in the Ceyhan River basin, Türkiye, by Darabi et al. (2023), spanning 1975-2014. This study utilised data from 15 meteorological stations to compare rainfall patterns across different periods of the year, segmented into four decades (1975-1984, 1985-1994, 1995-2004, 2005-2014) and analysed over annual and monthly scenarios. The findings highlight significant intra-annual and decadal variability, with pronounced regional differences and notable drought events in 1992-1996 and 2007-2010 impacting the basin's water resources.

2.9 Statistical Methods for Analysing the Relationship between the UHI Effect and Short-Duration Extreme Rainfall

Research examining the relationship between the UHI effect and short-duration extreme rainfall remains limited. To address this gap, this section reviews key statistical methods that have been used in related studies to

analyse associations between climatic and environmental variables. By referring to existing literature involving comparable parameters, such as LST, precipitation intensity and urban characteristics, this section highlights suitable correlation and regression techniques for investigating the UHI–rainfall relationship.

2.9.1 Pearson's Correlation Coefficient (r)

Pearson's correlation coefficient is a statistical method used to assess the direction and strength of a linear relationship between two consecutive variables. Spanning from -1 to +1, values closer to the extremes indicate stronger relationships, and r = 0 represents no linear correlation. The strength of the association increases as the scatter of data points decreases, approaching a straight line as the coefficient nears ±1. However, proper inference requires that the data meet several assumptions: both variables should be continuous, jointly normally distributed and drawn from a representative sample. Additionally, there should be no extreme outliers, as these can significantly distort the correlation results (Schober, Boer, and Schwarte, 2018; Berman, 2016).

In environmental and urban climate studies, Pearson's correlation has been widely applied to assess the relationships between the SUHII and climate variables such as air temperature, precipitation and vegetation activity. The standard deviation (STD) of SUHII over time was used as a metric for its stability, with a lower STD indicating stronger resilience to climatic drivers. Moreover, spatial correlations between SUHII variability (STD) and precipitation were analysed across 145 cities to assess regional sensitivity (Li, Zha, and Wang, 2020).

Further extending its application, Xu (2025) used Pearson's correlation alongside a random forest model to capture both linear and nonlinear relationships between green space morphological characteristics and UHI intensity across different seasons.

2.9.2 Kendall's Tau Correlation Coefficient (τ)

Kendall's Tau (τ) is a classic non-parametric statistical method, introduced by Kendall (1938), that evaluates the monotonic relationship between two

variables using ranked data. Its values range from -1 to +1, where a value of 0 indicates no correlation, positive values signify a direct relationship and negative values indicate an inverse relationship (Cheng et al., 2021). Unlike Pearson's correlation coefficient, Kendall's Tau does not require assumptions of linearity or normally distributed data, making it particularly suitable for non-linear or non-normally distributed datasets. Its main advantages include distribution independence, sensitivity to monotonic trends and resilience to outliers (Miniandi, 2024). This robustness is supported by simulation studies from Puth, Neuhäuser, and Ruxton (2015), which found that while both Kendall's Tau and Spearman's Rho perform similarly to Pearson's correlation in non-normal distributions, Kendall's Tau tends to produce narrower confidence intervals, especially when there are no tied values, making it statistically preferable in such cases. However, Spearman's Rho may outperform Kendall's Tau in terms of coverage accuracy when ties are present.

Comparative efficiency studies, such as those by Croux and Dehon (2010), also show that Kendall's Tau exhibits greater robustness to extreme and correlation outliers, with lower mean squared error (MSE) across a variety of data contamination scenarios compared to both Spearman's Rho and Pearson's correlation. In practical applications, Kendall's Tau has been widely used in climate and environmental studies. For instance, Nath et al. (2023) used it to analyse the relationship between summer temperature and monsoon rainfall, while Omer et al. (2020) applied it to examine the statistical dependence between water scarcity indicators and hydroclimatic variables. Similarly, Miniandi (2024) employed Kendall's Tau to assess the association between LST and extreme rainfall, emphasising its utility in urban climate research. Overall, Kendall's Tau remains a reliable and robust method for evaluating correlations in complex, real-world environmental datasets.

2.9.3 Spearman's Rho Correlation Coefficient (ρ)

Spearman's rank correlation coefficient, commonly referred to as Spearman's Rho (ρ), is a non-parametric method employed to evaluate the direction and strength of a monotonic relationship between two variables. This method is useful when the data do not meet the assumptions required for Pearson's correlation, such as normality or linearity. Unlike Pearson's product-moment

correlation coefficient, Spearman's Rho does not require a linear relationship or interval-level measurement; it can be applied to ordinal data and is robust in situations where the association is monotonic but non-linear. It spans from -1 to +1, with values near ±1 indicating stronger associations, while a value of 0 suggests no correlation. While a significant Spearman's coefficient indicates a statistically detectable monotonic relationship, it should be interpreted cautiously, as significance does not always imply a strong or meaningful relationship, especially in the practical or linear sense (Jan and Toasz, 2011; Cheng et al., 2021).

In applied research, Spearman's Rho has been used across various environmental and climate-related studies. For instance, Aucahuasi-Almidon, Cabrera-Carranza, and Garate-Quispe (2024) employed it to analyse the relationship between deforestation and climate variables in the southern Peruvian Amazon. Similarly, Sharma et al. (2016) used Spearman's Rho to examine the correlation between rainfall and temperature in the eastern region of India, demonstrating its utility in assessing hydroclimatic interactions where non-linearities and ordinal data are present.

2.9.4 Coefficient of Determination (R^2)

The coefficient of determination (R^2) is a commonly applied statistical metric that quantifies the goodness-of-fit of regression models by indicating how well the predicted values match the observed data. Ranging from 0 to 1, an R^2 value of 1 represents a perfect fit, whereas a value of 0 implies that the model fails to account for any variation in the dependent variable (Motegaonkar and Kashid, 2024). The concept of the coefficient of determination as the proportion of "variance explained" by the independent variables makes it an intuitive and unitless measure, allowing for easy comparison across models and studies, similar to standardised effect size metrics (Nakagawa and Schielzeth, 2012).

In environmental and UHI-related research, the coefficient of determination has been extensively employed to assess the strength of predictive relationships between climatic or surface parameters and urban heat metrics. For example, Pande et al. (2024) performed linear regression analysis using mean values of rainfall, NDVI and LST to determine R^2 values that reflected the strength of associations between these environmental factors.

Similarly, Fetene (2025) developed regression models using variables such as NDVI, NDBI, temperature, rainfall, relative humidity and wind speed to predict LST and UHI intensity, with R^2 used to assess model fit and p-values (threshold 0.05) used to test statistical significance. Additionally, Hussien et al. (2023) applied the coefficient of determination to evaluate the seasonal and annual relationships between NDVI and various climatic variables over a 26-year period, highlighting its utility in long-term trend analysis and environmental monitoring.

2.10 Classification Metrics

Classification metrics are important tools for assessing the performance of machine learning algorithms in LULC classification. Among these, the Kappa coefficient, overall accuracy, producer's accuracy and user's accuracy are widely employed.

2.10.1 Kappa Coefficient

The Kappa coefficient measures the level of agreement between a classification outcome and a reference dataset, while accounting for chance agreement. It is computed using the confusion matrix, which tabulates the classified outputs against the actual ground-truth observations. A Kappa value of 1 signifies perfect agreement, whereas values close to 0 indicate agreement comparable to random chance. Kappa values are categorised as follows: 0.81 - 1.00 represents almost perfect agreement; 0.61 - 0.80 indicates substantial agreement; 0.41 - 0.60 suggests moderate agreement; 0.21 - 0.40 corresponds to fair agreement; 0.00 - 0.20 implies slight agreement; and values below 0.00 indicate poor agreement. A study achieving 81.7% overall accuracy with a Kappa of 0.722 demonstrated substantial agreement, though individual classes like barren land showed high omission errors (73.3%). This highlights Kappa's utility in identifying systemic classification weaknesses (Rwanga and Ndambuki, 2017). The recent incorporation of machine learning techniques with Landsat imagery has further enhanced classification accuracy. In one study from Brahmani-Dwarka Interfluve, five classifiers, including RF, SVM, CART, Gradient Boosted Trees and Minimum Distance, were compared,

demonstrating Kappa's relevance even within sophisticated classification frameworks (Mandal, 2024).

While overall accuracy measures raw correctness, it often fails to account for class imbalances. For example, a hypothetical 90% accuracy might mask poor performance in rare classes if common classes dominate. Kappa improves on this by discounting chance agreement, making it particularly useful in heterogeneous landscapes. However, some critics argue that the "chance correction" provided by Kappa is unnecessary since overall accuracy already reflects observable agreement. They point out that Kappa's reliance on prevalence distributions can hinder cross-study comparisons, as it assumes independent class distributions between the classified and reference data, a condition rarely met in practice. Consequently, when rare classes are present, the inflated expected chance agreement can lower the Kappa value even if the classification is largely accurate. For instance, a class that covers only 1% of an area would need nearly flawless classification to achieve a high Kappa, potentially skewing the results. Additionally, Kappa does not differentiate between systematic errors and random noise, which can limit its diagnostic usefulness (Rwanga and Ndambuki, 2017; Foody, 2020).

2.10.2 Accuracy Metrics

The confusion matrix, also known as the error matrix, forms the basis for accuracy metrics by comparing classified pixels with ground reference data. In this matrix, the diagonal elements indicate correct classifications, while the off-diagonal values represent errors: commission errors (pixels incorrectly assigned to a class) are found in the rows, and omission errors (pixels that truly belong to a class but were not classified as such) are found in the columns (Dash et al., 2023).

Overall accuracy gives a broad snapshot of performance, yet it can conceal significant class-specific problems. For example, a Nigerian LULC study reported an overall accuracy of 81.7%, but this figure masked a severe 73.3% omission error for barren land. This case illustrates how overall accuracy may be misleading in imbalanced landscapes, where dominant classes, such as agriculture covering 65% of the area, can disproportionately skew the results (Rwanga and Ndambuki, 2017).

Producer's accuracy measures the probability that a reference sample is correctly classified, thereby highlighting omission errors, while user's accuracy assesses the likelihood that a classified sample truly matches the reference data, pinpointing commission errors (Rwanga and Ndambuki, 2017).

2.11 Regression Metrics

Regression metrics are commonly used to evaluate the predictive accuracy and reliability of statistical models. A summary of these metrics, compiled by the author based on information from Tanoori, Soltani, and Modiri (2024), Mohammad et al. (2022), and Jedox (2025), is presented in Table 2.1.

Table 2.1: The Summary of Regression Metrics.

Type	Descriptions
Root Mean Squared Error (RMSE)	• It is commonly applied in regression models by measuring the average difference between the estimated values generated by a model and the observed values obtained from a sensor. • Squaring residuals amplifies the impact of large errors, making RMSE ideal for applications where catastrophic failures must be avoided. • Models with higher accuracy result in lower RMSE value.
Concordance Index (CI)	• It is typically employed in tasks involving ranking systems. • It evaluates a model's capability to precisely rank the predicted values. • A 90% CI of [0.72, 0.78] for R^2 indicates the model explains 72 -

Table 2.1 (Continued)

	78% of variance with 90% confidence.
Average Absolute Percent Relative Error (AAPRE)	• It evaluates a model's predictive performance. • It expresses the average difference between the observed and estimated values in percentage form. • Models with higher accuracy result in lower AAPRE percentage.
R-squared (R^2)	• It assesses the model's fit by measuring how much of the variation in the outcome can be explained by the input variables. • A high R-squared value implies a strong association between the inputs and the predicted outcome.
Mean Absolute Error (MAE)	• It assesses the mean of the absolute errors between the estimated and observed values. • It is less sensitive to outliers than RMSE, making it preferable for datasets with heavy-tailed error distributions. • Models with greater accuracy result in lower MAE value.

2.12 Summary

In summary, this literature review evaluates and compares various research methods to identify the most appropriate approaches for investigating the relationship between the UHI effect and short-duration extreme rainfall. The review highlights the characteristics and applications of Landsat and MODIS imagery in LULC and LST analysis. It also covers the concepts and case studies of LULC classification using machine learning algorithms, such as SVM, KNN, DT and RF, along with the role of various land use indices in improving classification accuracy. For LST estimation, two algorithm types, SC and SW, are discussed, while multiple formulae adopted by other researchers are presented for SUHII estimation. Methods for extracting short-duration extreme rainfall are also reviewed, including diverse data sources and definitions of extreme events. Among spatial interpolation methods, IDW stands out for its balance between performance and simplicity, outperforming more complex methods in many cases. In terms of rainfall analysis, methods such as the MMK test, Sen's slope, ITA and spatiotemporal comparative analysis are reviewed for their suitability to different analytical objectives. For relationship analysis, four key statistical approaches, Pearson's correlation, Kendall's Tau, Spearman's rho, and coefficient of determination, are evaluated. Kendall's Tau shows the highest resilience to outliers, while the coefficient of determination combined with linear regression enables both prediction and assessment of relationship strength. Finally, the review examines classification metrics and regression metrics, which are essential for evaluating the performance of machine learning models and the reliability of statistical analyses.

CHAPTER 3

METHODOLOGY AND WORK PLAN

3.1 Introduction

This study employed a structured methodology to ensure analytical rigor and reliability. Kuala Lumpur, Malaysia, was selected as the study area due to its dense urbanisation and susceptibility to the UHI effect. Satellite imagery from Landsat 5 (2007) and Landsat 8 (2015 and 2023) was used, along with ground-based rainfall data from 2007 to 2023 acquired from the Department of Irrigation and Drainage (DID) Malaysia.

LULC classification was conducted using SVM and RF algorithms, categorising land into built-up areas, water bodies, vegetation, and bare soil. Classification accuracy was assessed using Kappa coefficients and accuracy metrics. LST was estimated using the SC algorithm for both satellite datasets. SUHII was calculated by subtracting the mean LST of vegetated areas from the LST of each pixel.

Rainfall data were pre-processed to extract the 99th percentile of hourly extreme rainfall from the original 15-minute records. Missing data were interpolated using the IDW method and Theissen polygons were generated based on rainfall station locations. Analyses were performed to assess the spatiotemporal variation, diurnal distribution and trends in three rainfall metrics: the total number of hourly extreme rainfall events, the total of hourly extreme rainfall and the mean of hourly extreme rainfall.

To explore the relationship between SUHII and rainfall extremes, both the coefficient of determination (R^2) and Kendall's Tau correlation (τ) were applied. SUHII values for 2007, 2015 and 2023 were interpolated to derive annual estimates for each station from 2007 to 2023. Rainfall metrics were aggregated annually to maintain consistency in statistical analysis. The overall methodological workflow is illustrated in Figure 3.1.

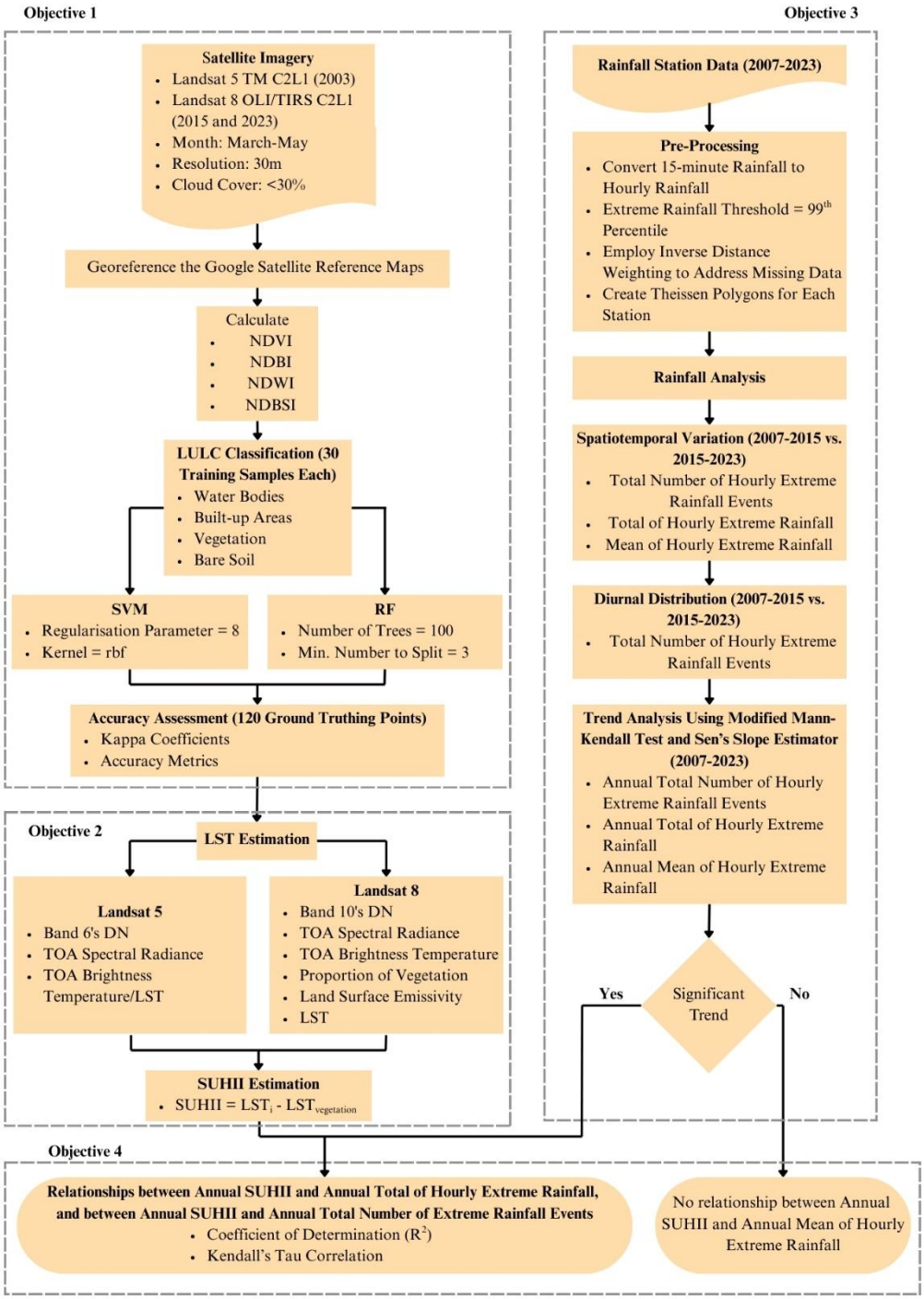


Figure 3.1: The Flowchart of Methodology.

3.2 Study Area

Kuala Lumpur, the capital of Malaysia, spans roughly 243 km² and features densely urbanised metropolitan areas distributed across the city. Geographically, it is located in the central part of the west coast of Peninsular Malaysia and forms part of the Klang Valley conurbation. Kuala Lumpur

serves as the country's economic hub, supporting a population of around 1.8 million and hosting numerous commercial activities. The city's economy is driven by diverse sectors including finance, construction, commerce, science and technology, transportation, and manufacturing, contributing approximately 15.9% to Malaysia's gross domestic product in 2023.

Situated near the equator, Kuala Lumpur experiences a tropical climate characterised by high humidity, frequent rainfall and minimal seasonal temperature variation throughout the year. Average temperatures range from about 33.0 °C during the day to 24.6 °C at night, with the months from March to August typically being the hottest. Kuala Lumpur receives significant daily rainfall, ranging from 5.29 mm to a peak of 12.24 mm. The city also maintains a high average relative humidity of 83% throughout the year (WorldData.info, 2024). Rapid urbanisation since the 1980s has intensified the UHI effect in Kuala Lumpur, making afternoon thunderstorms and flash floods prevalent. Figure 3.2 shows the location of Kuala Lumpur in Malaysia.

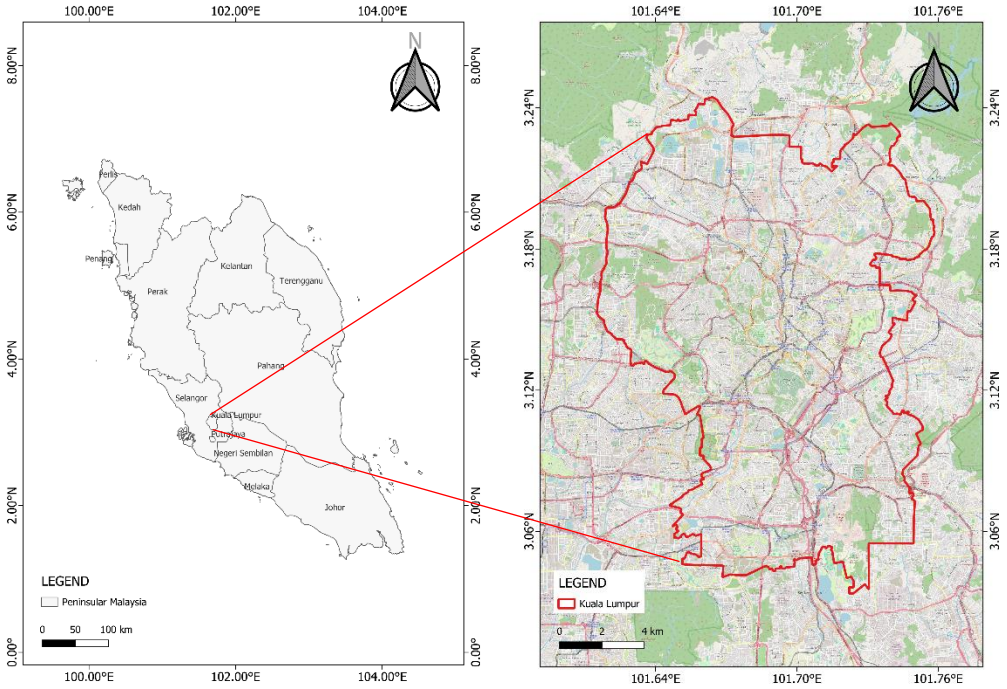


Figure 3.2: The Location of Kuala Lumpur in Malaysia.

3.3 Data Collection

Before beginning the analysis, data collection is essential to acquire the most representative datasets from reliable sources, thereby minimising computational costs and uncertainties in the results. A study period from 2007 to 2023 with an 8-year interval was purposely selected to observe the effects of urbanisation on the UHI phenomenon and to provide a longer period for analysing short-duration extreme rainfall events.

3.3.1 Satellite Imagery

Three satellite images from Landsat 5 (2007) and Landsat 8 (2015 and 2023) were downloaded from the USGS Earth Explorer website. The satellite data were selected from the months of March to August, as these represent the hottest period of the year. Images with less than 30% cloud cover were chosen to minimise the impact of weather conditions on LULC classification and LST estimation. The spatial resolution of the multi-spectral bands used was 30 m, while the thermal bands were resampled from 120 m to 30 m to ensure both types of bands are compatible with each other. The satellite imagery served as training data for machine learning algorithms in QGIS to classify LULC. The specifications of the extracted satellite imagery are outlined in Table 3.1.

Table 3.1: The Specifications of the Extracted Satellite Imagery.

Product ID	Date Captured	Sensor	Resolution	Cloud Cover
LT05_L1TP_127058_20070524_20200830_02_T1	24-05-2007	Landsat 5 TM C2 L1	30 m	<30%
LC08_L1TP_127058_20150327_20200909_02_T1	27-03-2015	Landsat 8 OLI/TIRS C2 L1		
LC08_L1TP_127058_20230317_20230324_02_T1	17-03-2023	Landsat 8 OLI/TIRS C2 L1		

3.3.2 Rainfall Data

15-minute local rain gauge measurements from 2007 to 2023 were obtained from the DID Malaysia. Site-measured rainfall data were selected for their higher precision. Owing to data availability constraints, only nine stations located within or near the study area were included in this research. Table 3.2 lists the selected rainfall stations, while Figure 3.3 illustrates their locations.

Table 3.2: Rainfall Stations.

No.	Station	Latitude	Longitude
1	Kolam Takungan Batu (0231391RF)	3.2185	101.6819
2	Taman Ehsan At Kepong W. Persekutuan (0231441RF)	3.21822	101.6313
3	Pusat Penyelidekan At Jps Ampang Selangor (0231351RF)	3.15489	101.7487
4	Km 10 Ulu Kelang At Uk Height (0231401RF)	3.18228	101.7597
5	Ibu Bekalan Km. 11 At Gombak W. Persekutuan (0230721RF)	3.23856	101.7122
6	Empangan Genting Klang At W. Persekutuan (0230631RF)	3.2361	101.7528
7	I/pejabat Jps Malaysia At W. Persekutuan (0230641RF)	3.15567	101.6818
8	Jln. Sg. Udang At Segambut (0231381RF)	3.19331	101.658
9	Ldg. Edinburgh Site 2 At W. Persekutuan (0230651RF)	3.1833	101.6333

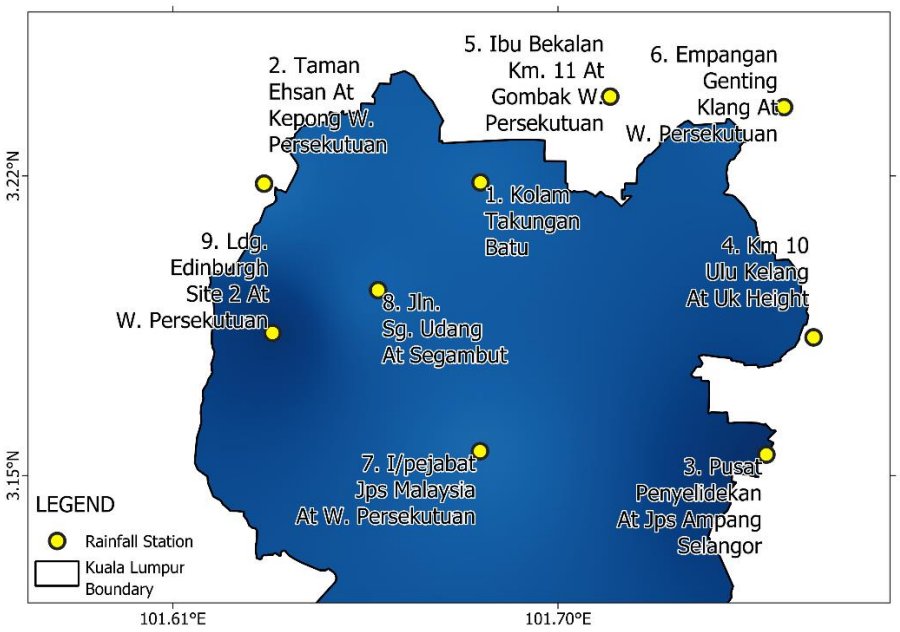


Figure 3.3: Locations of Rainfall Stations.

3.4 LULC Classification

In this section, two different machine learning algorithms, including SVM and RF, were adopted to classify the Landsat imagery in 2007, 2015 and 2023 into different LULC types using QGIS. Based on accuracy assessment, the LULC maps with the higher accuracy were selected for subsequent analyses.

3.4.1 Georeferencing

Georeferenced maps were created using high-resolution satellite images of the study area in 2007, 2015 and 2023, obtained via Google Earth Pro. A total of 50 ground control points were uniformly distributed across the images to ensure accurate georeferencing. A transformation type of “Polynomial 3” and a target coordinate reference system of “EPSG:32647 – WGS84 / UTM zone 47N” were used.

3.4.2 Land Use Indices

Various land use indices, such as NDVI (Equation 3.1), NDBI (Equation 3.2), NDWI (Equation 3.3) and NDBSI (Equation 3.4), were used to improve classification accuracy.

$$\mathrm{NDVI} = \frac{\mathrm{NIR} - \mathrm{RED}}{\mathrm{NIR} + \mathrm{RED}} \tag{3.1}$$

$$\mathrm{NDBI} = \frac{\mathrm{SWIR1} - \mathrm{NIR}}{\mathrm{SWIR1} + \mathrm{NIR}} \tag{3.2}$$

$$\mathrm{NDWI} = \frac{\mathrm{GREEN} - \mathrm{NIR}}{\mathrm{GREEN} + \mathrm{NIR}} \tag{3.3}$$

$$\mathrm{NDBSI} = \frac{(\mathrm{RED} + \mathrm{SWIR1}) - (\mathrm{BLUE} + \mathrm{NIR})}{(\mathrm{RED} + \mathrm{SWIR1}) + (\mathrm{BLUE} + \mathrm{NIR})} \tag{3.4}$$

where

NDVI = Normalised Difference Vegetation Index

NDBI = Normalised Difference Built-up Index

NDWI = Normalised Difference Water Index

NDBSI = Normalised Difference Bare Soil Index

NIR = Near-infrared (Band 4 for Landsat 5 and Band 5 for Landsat 8)

RED = Red (Band 3 for Landsat 5 and Band 4 for Landsat 8)

SWIR1 = Shortwave Infrared 1 (Band 5 for Landsat 5 and Band 6 for Landsat 8)

GREEN = Green (Band 2 for Landsat 5 and Band 3 for Landsat 8)

BLUE = Blue (Band 1 for Landsat 5 and Band 2 for Landsat 8)

3.4.3 Classification of LULC Maps with SVM and RF

The Landsat raster images were clipped to the study area and classified using SVM and RF models into four categories, encompassing built-up areas (roads, commercial, residential and industrial areas) water bodies (lakes, rivers, seas and canals), vegetation (forests, agriculture areas, green lands and green plants) and bare land (sand, fallow and vacant land). This classification was executed using the Semi-Automatic Classification Plugin in QGIS. Image pixels were properly labelled for each class to train the algorithms for supervised classification. A total of 30 training samples per class were generated for each

year, with all samples proportionally distributed based on the occurrence rate of the LULC classes.

For the SVM model, a radial basis function kernel with a regularisation parameter of 8.0 was employed.

The RF model, consisting of 100 decision trees, was utilised for the classification process, with the minimum number to split set to three.

3.4.4 Accuracy Assessment

To evaluate the accuracy of SVM- and RF-based LULC maps, producer's accuracy (PA), user's accuracy (UA), overall accuracy (OA) and the Kappa statistic were calculated, as shown in Equations 3.5 to 3.8. The assessment used 30 randomly selected ground-truth data points for each category, derived from the georeferenced maps.

$$\text{Producer's Accuracy} = \frac{\text{Correctly Classified Pixels of a Category}}{\text{Total Reference Pixels for that Category}} \times 100 \quad (3.5)$$

$$\text{User's Accuracy} = \frac{\text{Correctly Classified Pixels of a Category}}{\text{Total Classified Pixels for that Category}} \times 100 \quad (3.6)$$

$$\text{Overall Accuracy} = \frac{\text{Total Correctly Classified Pixels}}{\text{Total Pixels}} \times 100 \quad (3.7)$$

$$\text{Kappa Coefficient} = \frac{p_0 - p_e}{1 - p_e} \quad (3.8)$$

where

$$p_0 = \text{Observed Agreement} = \frac{\text{Total Correctly Classified Pixels}}{\text{Total Pixels}}$$

$$p_e = \text{Expected Agreement} = \sum \frac{\text{Column Total x Row Total}}{\text{Total Pixels}}$$

3.5 LST Estimation

The LST estimation was performed in QGIS by analysing thermal radiance values associated with different LULC types using specific formulae. The SC algorithm was applied to the thermal bands of both Landsat 5 and Landsat 8 imagery.

3.5.1 Extraction of LST Data from Landsat 5

The first step involved converting the digital number (DN) of Band 6 to Top of Atmosphere (TOA) spectral radiance, standardising the radiometric scale across all satellite images using Equation 3.9. Input parameters were obtained from the MTL file of the satellite image.

$$L_\lambda = \left(\frac{LMAX_\lambda - LMIN_\lambda}{Q_{cal\,max} - Q_{cal\,min}}\right)(Q_{cal} - Q_{cal\,min}) + LMIN_\lambda \qquad (3.9)$$

where

L_λ = TOA spectral radiance, W/(m^2 · sr · μm)

$LMAX_\lambda$ = spectral radiance that is scaled to $Q_{cal\,max}$, W/(m^2 · sr · μm)

$LMIN_\lambda$ = spectral radiance that is scaled to $Q_{cal\,min}$, W/(m^2 · sr · μm)

Q_{cal} = quantised calibrated pixel value in DNs

$Q_{cal\,max}$ = maximum quantised calibrated pixel value corresponding to $LMAX_\lambda$

$Q_{cal\,min}$ = minimum quantised calibrated pixel value corresponding to $LMIN_\lambda$

Secondly, the TOA spectral radiance was converted to the TOA brightness temperature using Equation 3.10.

$$BT = \left(\frac{K_2}{\ln\left(\frac{K_1}{L_\lambda} + 1\right)}\right) - 273.15 \qquad (3.10)$$

where

BT = TOA brightness temperature, °C

L_λ = TOA spectral radiance, W/(m^2 · sr · μm)

K_1= A constant value for Band 6 / Band 10 thermal conversion (K1_CONSTANT_BAND_6 / K1_CONSTANT_BAND_10)

K_2= A constant value for Band 10 thermal conversion
(K2_CONSTANT_BAND_10 / K2_CONSTANT_BAND_10)

3.5.2 Extraction of LST Data from Landsat 8

Firstly, the Band 10 DN was converted into the TOA spectral radiance by applying Equation 3.11. The radiance scaling factors were retrieved from the MTL file of the satellite image.

$$L_\lambda = M_L \cdot Q_{cal} + A_L \quad (3.11)$$

where
L_λ = TOA spectral radiance, W/($m^2 \cdot sr \cdot \mu m$)
M_L = Radiance multiplicative scaling factor for Band 10
(RADIANCE_MULT_BAND_10)
Q_{cal} = Level 1 pixel value in DN
A_L = Radiance additive scaling factor for Band 10
(RADIANCE_ADD_BAND_10)

Similar to Landsat 5, the TOA brightness temperature was calculated in the second step by using Equation 3.10. An emissivity correction was conducted in the following step, involving the NDVI approach. The method used was the same as Equation 3.1, which required both near-infrared (NIR) and red (R) DN values.

Then, the proportion of vegetation (P_V) was computed from Equation 3.12, using the DN values from the NDVI image and their maximum and minimum values.

$$P_V = \left(\frac{NDVI - NDVI_{min}}{NDVI_{max} - NDVI_{min}}\right)^2 \quad (3.12)$$

Subsequently, the P_V value and a correction value of 0.986 were applied in Equation 3.13 to calculate the land surface emissivity.

$$\epsilon = 0.004 \times P_V + 0.986 \quad (3.13)$$

Lastly, the LST of Landsat 8 for Bands 10 was computed using Equations 3.14 and 3.15.

$$LST = \frac{BT}{\left\{1 + \left[\frac{\lambda BT}{\rho}\right] \ln \epsilon\right\}} \quad (3.14)$$

$$\rho = h\frac{c}{\sigma} \quad (3.15)$$

where

LST = Land surface temperature, °C

BT = TOA brightness temperature, °C

ϵ = Land surface emissivity

λ = Wavelength of emitted radiance = 10.985 x 10^{-6} m

c = Velocity of light = 2.998 x 10^{8} m/s

h = Planck's constant = 6.626 x 10^{-34} Js

σ = Boltzmann constant = 1.38 x 10^{-23} J/K

ρ = 1.438 x 10^{-2} mK

3.6 SUHII Estimation

To measure the UHI effect within the study area, the SUHII formula was calculated using Equation 3.16, which determines the difference between the LST at each pixel and the mean LST of vegetation-covered areas. Table 3.3 summarises the categories of UHI based on SUHII values, ranging from none (SUHII ≤ 0), low (0 < SUHII ≤ 2.0), moderate (2.0 < SUHII ≤ 4.0), high (4.0 < SUHII ≤ 6.0), very high (6.0 < SUHII ≤ 8.0) to extremely high intensity (SUHII > 8.0). This approach effectively captures UHI by accounting for the impact of different land use types on LST.

$$SUHII = LST_i - LST_{vegetation} \quad (3.16)$$

where

SUHII = Surface urban heat island intensity

LST_i = Land surface temperature at any pixel, °C

$LST_{vegetation}$ = Mean land surface temperature of vegetation covers, °C

Table 3.3: The Categories of UHI Based on SUHII values.

SUHII Value Range	Category
SUHII ≤ 0	None
0 < SUHII ≤ 2.0	Low UHI
2.0 < SUHII ≤ 4.0	Moderate UHI
4.0 < SUHII ≤ 6.0	High UHI
6.0 < SUHII ≤ 8.0	Very high UHI
SUHII > 8.0	Extremely high UHI

3.7 Rainfall Analysis

A detailed rainfall analysis was carried out using Python and QGIS to examine variations in rainfall patterns from 2007 to 2023. Rainfall data from nine selected stations were retrieved and pre-processed prior to analyses.

3.7.1 Pre-Processing

Data pre-processing was essential to facilitate the analytical process and enhance the reliability of the results. The IDW interpolation method, shown in Equation 3.17, was employed to estimate missing values based on rainfall measurements from nearby stations within the same period. As there is no theoretical basis for selecting the power parameter, a value of t = 2 was used in most cases.

$$P_i = \frac{\sum_{x=1}^{n} \left[\frac{1}{d_x^{\,t}} \times P_x \right]}{\sum_{x=1}^{n} \frac{1}{d_x^{\,t}}} \tag{3.17}$$

where

P_i = The rainfall estimates at the station i

P_x = Rainfall measurement at neighbouring station x

d_x = The distance between the station i to the station x

n = The number of neighbouring stations

t = The power parameter that influences the weight of neighbouring measurements = 2

To enhance interpretability and computational efficiency, the original 15-minute data were aggregated into hourly rainfall data. The 99th percentile threshold was applied to differentiate extreme rainfall events from others. To analyse the spatial distribution of rainfall impacts, Thiessen polygons (Voronoi polygons) were generated, defining distinct impact areas for each station and establishing nine study areas.

3.7.2 Spatiotemporal Variation of Hourly Extreme Rainfall

Due to the high variability in Malaysia's annual rainfall patterns, the dataset was divided into two study periods: 2007-01-01 to 2015-07-01 and 2015-07-01 to 2023-12-31 (hereinafter referred to as 2007-2015 and 2015-2023) to enable comparative analyses of hourly extreme rainfall. Three rainfall parameters were included, such as the total number of hourly extreme rainfall events, the total of hourly extreme rainfall and the mean of hourly extreme rainfall. This comparison aimed to highlight changes in the actual values of these parameters at the nine selected stations across the two periods.

The spatial distribution of rainfall characteristics was visualised using the IDW interpolation method, with a distance coefficient of 2 and a pixel size of 0.00027° in both the x and y directions.

3.7.3 Diurnal Distribution of Hourly Extreme Rainfall

To further examine variations in hourly rainfall patterns, a diurnal distribution analysis was conducted on the total number of hourly extreme rainfall events across the nine stations for the periods 2007-2015 and 2015-2023. Rainfall data were categorised by hour of the day and clustered column charts were generated for each station. This analysis provided clearer insights into the timing of extreme rainfall occurrences, supporting the investigation of UHI impacts on short-duration extreme rainfall.

3.7.4 MMK Test and Sen's Slope Estimator

The MMK test and Sen's slope estimator were employed to analyse the trend and rate of change in hourly extreme rainfall from 2007 to 2023. The hourly extreme rainfall data were grouped on an annual basis. The MMK test was selected for its ability to handle non-normally distributed time series and its robustness against the effects of autocorrelation.

The MK test statistic S was computed using Equation 3.18, which assesses the difference in ranks between sequential data points.

$$S = \sum_{i=1}^{N-1} \sum_{j=i+1}^{N} \operatorname{sgn}(x_j - x_i) \tag{3.18}$$

N indicates the sample size, while x_j and x_i represent consecutive data points within the time series. Equation 3.19 was applied to obtain the statistic $\operatorname{sgn}(x_j - x_i)$.

$$\operatorname{sgn}(x_j - x_i) = \begin{cases} 1 \text{ if } (x_j - x_i) > 0 \\ 0 \text{ if } (x_j - x_i) = 0 \\ -1 \text{ if } (x_j - x_i) < 0 \end{cases} \tag{3.19}$$

Since the rainfall dataset is assumed to be independently and identically distributed, the mean of S is equal to zero. The variance of S, denoted as, Var(S), was computed using Equation 3.20.

$$\operatorname{Var}(S) = \frac{1}{18}\left[N(N-1)(2N+5) - \sum_{k=1}^{n} t_k(t_k-1)\,(2t_k+5)\right] \tag{3.20}$$

In this context, n refers to the number of tied groups and t_k indicates the number of ties within the k^{th} group. Equation 3.21 presents the standardised test statistic Z, which accounts for the presence of tied data.

$$Z = \begin{cases} \dfrac{S-1}{\sqrt{Var(S)}} \text{ if } S > 0 \\ 0 \text{ if } S = 0 \\ \dfrac{S+1}{\sqrt{Var(S)}} \text{ if } S < 0 \end{cases} \quad (3.21)$$

A positive Z value indicates an upward trend, while a negative Z value suggests a downward trend. A trend is considered statistically significant at the 0.05 level if the absolute value of Z exceeds 1.96. In the MMK test, the Hurst coefficient h and the autocorrelation function were incorporated to correct for autocorrelation effects on the variance of S, as shown in Equations 3.22 and 3.23.

$$C_N(h) = \left[\rho_{|j-i|}\right] for\ i = 1{:}N; j = 1{:}N \quad (3.22)$$

$$\rho_l = \frac{1}{2}(|l+1|^{2h} - 2|l|^{2h} + |l-1|^{2h}) \quad (3.23)$$

The degree of self-similarity in a time series is quantified by h, while the autocorrelation function of lag l for a specified h is represented by ρ_l. The value of h was determined using the maximum likelihood estimation method. Its significance was evaluated by comparing it to the expected mean and standard deviation under the assumption of h = 0.5, following a normal distribution. If h is deemed significant, the variance of S is adjusted using a bias-corrected estimate based on the specified h, as presented in Equation 3.24, replacing the original Var(S) in Equation 3.21.

$$\mathrm{Var}(S)^{h'} = \sum_{i<j}\sum_{k<l}\frac{2}{\pi}\sin^{-1}\left(\frac{\rho|j-i| - \rho|i-l| - \rho|j-k| + \rho|i-k|}{\sqrt{(2-2\rho|i-j|)2-2|k-l|)}}\right) \quad (3.24)$$

The Sen's slope estimator was then utilised following the MMK test to assess the rate of change. The slopes T_i for all consecutive pairs of data

were calculated using Equation 3.25 and the Sen's slope Q was derived as the median of these slopes, determined by Equation 3.26.

$$T_i = \frac{x_j - x_k}{j - k}, \qquad i = 1, 2, 3 \ldots N, \qquad j > k \tag{3.25}$$

$$Q = \begin{cases} T_{\frac{N+1}{2}} \; if \; N \; is \; odd \\ \frac{1}{2}\left(T_{\frac{N}{2}} + T_{\frac{N+2}{2}}\right) if \; N \; is \; even \end{cases} \tag{3.26}$$

3.8 Relationships between UHI and Short-Duration Extreme Rainfall

Both the coefficient of determination (R^2) and Kendall's Tau correlation (τ) were employed to evaluate the relationships between SUHII and rainfall parameters. The coefficient of determination assumes a linear relationship, aligning with the Clausius-Clapeyron scaling, which suggests a proportional increase in rainfall with rising temperature. In contrast, Kendall's Tau assumes a monotonic relationship and is better-suited for environmental data, which often violate the assumptions of normality and linearity. Moreover, Kendall's Tau is robust to outliers, enhancing its reliability in real-world climatic datasets. The combined use of the coefficient of determination and Kendall's Tau enables a more comprehensive analysis by capturing both linear and consistent directional associations, regardless of data distribution. This dual approach enhances the reliability and interpretability of the findings.

Both datasets were pre-processed prior to the relationship analysis to ensure consistency in the statistical analysis. SUHII data for the years 2007, 2015 and 2023 were interpolated to estimate annual SUHII values from 2007 to 2023 at each station, while hourly extreme rainfall data were aggregated on an annual basis.

3.8.1 Coefficient of Determination (R^2)

The coefficient of determination is a statistical approach used to evaluate how well a regression model explains the variation in a dependent variable based on the independent variable. R^2 spans from 0 to 1, where $R^2 = 0$ denotes no linear relationship and $R^2 = 1$ represents a perfect fit.

Equation 3.27 presents the formula for the coefficient of determination, while Equation 3.28 shows the Fisher's Z-transformation (z-score) utilised to evaluate the statistical significance of the coefficient.

$$R^2 = 1 - \frac{SS_{res}}{SS_{tot}} \tag{3.27}$$

where

$SS_{res} = \sum(y_i - \hat{y}_i)^2 =$ Residual sum of squares

$SS_{tot} = \sum(y_i - \bar{y})^2 =$ Total sum of squares

$y_i =$ Observed values

$\hat{y}_i =$ Predicted values

$\bar{y} =$ Mean of observed values

$$z_{score} = \frac{1}{2}\ln\left(\frac{1+r}{1-r}\right) \cdot \sqrt{n-3} \tag{3.28}$$

where

n = Sample size

$r = \sqrt{R^2}$

By comparing to the z-score to the standard normal distribution, the null hypothesis of no significant relationship is rejected if $|z_{score}| > 1.96$, implying that the relationship is statistically significant at the 0.05 significance level.

3.8.2 Kendall's Tau Correlation (τ)

Kendall's Tau is a non-parametric statistic that measures the direction and strength of association between two ranked variables, ranging from -1 to +1. A value of -1 suggests perfect disagreement, +1 represents perfect agreement and zero indicates no correlation. Equation 3.29 shows the formula for Kendall's Tau correlation.

$$\tau = \frac{2(C - D)}{n(n-1)} \tag{3.29}$$

where

C = Number of concordant pairs

D = Number of discordant pairs

n = Total number of observations

A concordant pair occurs when observations (x_i, y_i) and (x_j, y_j) maintain a consistent order, such that $x_i > x_j$ and $y_i > y_j$, or $x_i < x_j$ and $y_i < y_j$. A discordant pair occurs when the observations follow a different order. Equation 3.30 presents the formula for the z-score.

$$z_{score} = 3\tau \cdot \frac{\sqrt{n(n-1)}}{\sqrt{2(2n+5)}} \tag{3.30}$$

Similarly, the statistical significance of the correlation is determined by comparing the z-score to the standard normal distribution; if $|z_{score}| > 1.96$, the null hypothesis of no significant relationship is rejected at the 0.05 significance level.

3.9 Summary

A systematic methodology was adopted to achieve the objectives of this study. Landsat satellite imagery (Landsat 5 for 2007 and Landsat 8 for 2015 and 2023) and 15-minute rainfall data from 2007 to 2023, sourced from the DID Malaysia, formed the core datasets. LULC classification was performed using SVM and RF algorithms, and accuracy was evaluated using Kappa coefficients and accuracy metrics. LST was estimated via the SC algorithm, and SUHII was derived by subtracting the mean LST of vegetated areas from the LST of each pixel.

Rainfall data were processed to extract the 99th percentile of hourly extreme rainfall from 15-minute records. Missing values were interpolated using the IDW method, and Theissen polygons were generated to define station influence zones. Three rainfall metrics were analysed: the total number of hourly extreme rainfall events, the total of hourly extreme rainfall and the mean of hourly extreme rainfall. These metrics were assessed in terms of the spatiotemporal variation, diurnal distribution and trends.

Both the coefficient of determination (R^2) and Kendall's Tau (τ) correlation were applied to assess the relationship between SUHII and short-duration extreme rainfall. SUHII data for 2007, 2015, and 2023 were interpolated to produce annual estimates across the study period, while rainfall metrics were aggregated annually to ensure consistency in statistical analysis.

CHAPTER 4

RESULTS AND DISCUSSION

4.1 Introduction

In this chapter, the comprehensive results of the research are presented and critically discussed with appropriate justifications. The first section outlines the spatiotemporal variations in historical LULC maps (2007, 2015 and 2023) across Kuala Lumpur, classified using SVM and RF algorithms. The performance of the classifications is evaluated through Kappa coefficients and accuracy metrics. The second section examines the historical variation of LST across Kuala Lumpur during the same period, supported by detailed analyses of LST changes. The third section presents the evolution of historical SUHII across major townships, accompanied by multiple statistical assessments. Subsequently, the chapter discusses the findings of various rainfall analyses, including the spatiotemporal variation, diurnal distributions and trends of hourly extreme rainfall events. Finally, the relationships between UHI and short-duration extreme rainfall are explored using the coefficient of determination (R^2) and Kendall's Tau correlation (τ).

4.2 Mapping Historical LULC

Landsat 5 and Landsat 8 imagery from 2007, 2015 and 2023 were processed using SVM and RF classification algorithms in QGIS to generate LULC maps for Kuala Lumpur. Each LULC class was trained with 30 sample points, and classification accuracy was evaluated using OA and kappa coefficients based on 120 ground-truth points. Figures 4.1 and 4.2 illustrate the LULC maps generated by SVM and RF for the respective years.

Tables 4.1 and 4.2 summarise the accuracy assessment of LULC classification using SVM and RF. The diagonal values in the confusion matrix represent the proportion of correctly classified pixels. Across all three study years, SVM consistently outperformed RF in terms of OA and kappa coefficient, achieving accuracy scores of 94.86%, 93.54% and 94.04%, with corresponding kappa coefficients of 0.9128, 0.8620 and 0.8766. In contrast, RF recorded slightly lower accuracies of 94.64%, 90.57% and 93.53%, with

kappa values of 0.9081, 0.7946 and 0.8758. Given its higher classification accuracy, the SVM-based LULC maps were selected for the subsequent areal change analysis.

Despite the overall high accuracy, water bodies and bare soil exhibited the lowest PA, ranging from 45.39% to 81.31%. This variability is likely due to the presence of cloud cover and shadows in the satellite imagery, which compound the challenges faced by machine learning algorithms in distinguishing these classes from built-up areas. Additionally, the limited spatial resolution of the satellite imagery increases the likelihood of human errors in creating training samples for the supervised classification. In 2023, the UA for bare soil fell below 60% due to misclassification, with portions of bare soil incorrectly identified as built-up areas and vegetation.

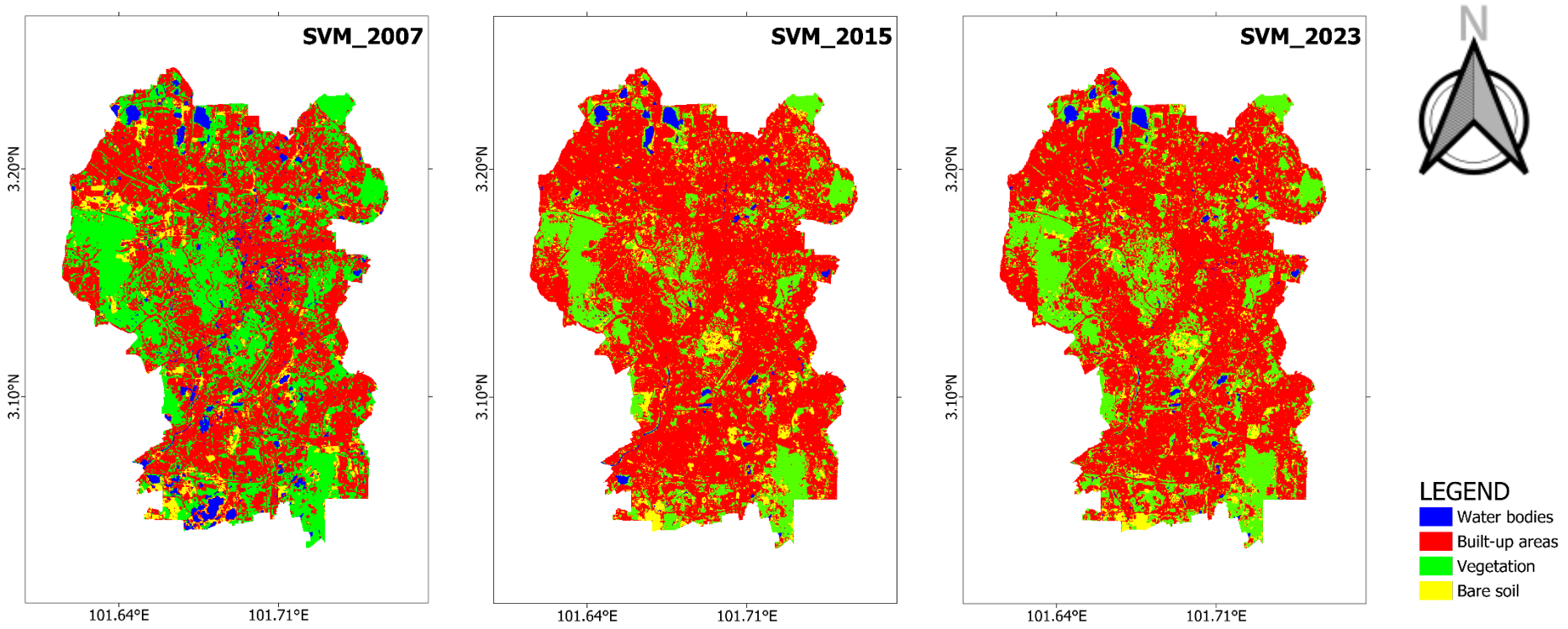


Figure 4.1: Classification of LULC in Kuala Lumpur Using SVM Algorithm in 2007, 2015 and 2023.

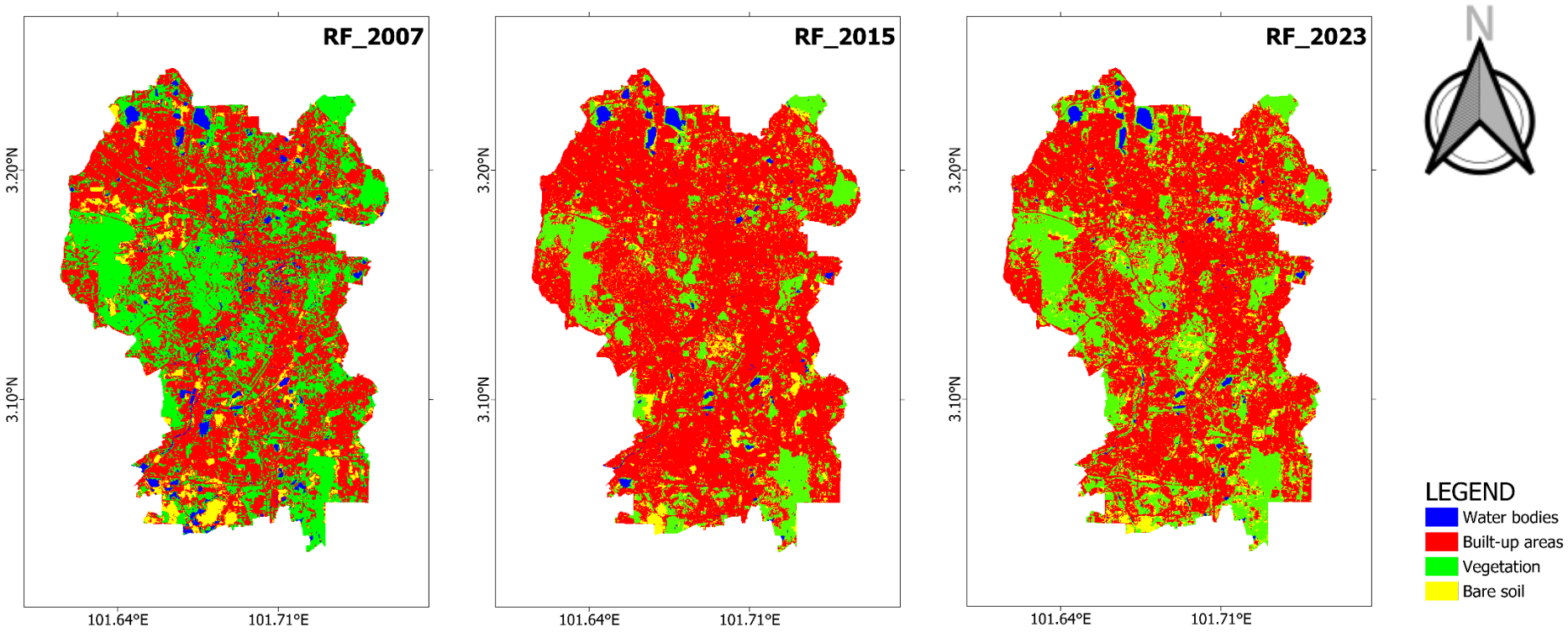


Figure 4.2: Classification of LULC in Kuala Lumpur Using RF Algorithm in 2007, 2015 and 2023.

Table 4.1: Accuracy Assessment of SVM-Based LULC Classification in 2007, 2015 and 2023.

SVM – 2007						
LULC Class	**Water Bodies**	**Built-up Areas**	**Vegetation**	**Bare Soil**	**Total**	**UA (%)**
Water Bodies	0.0384	0.0004	0.0004	0.0002	0.0394	97.6077
Built-up Areas	0.0268	0.4964	0.0067	0.0067	0.5366	92.5000
Vegetation	0.0000	0.0027	0.3603	0.0054	0.3684	97.8022
Bare Soil	0.0011	0.0011	0.0000	0.0535	0.0557	96.1538
Total	0.0663	0.5006	0.3673	0.0658	1.0000	
PA (%)	57.9172	99.1720	98.0713	81.3098		
OA (%)	94.8573					
Kappa	0.9128					
SVM - 2015						
LULC Class	**Water Bodies**	**Built-up Areas**	**Vegetation**	**Bare Soil**	**Total**	**UA (%)**
Water Bodies	0.0142	0.0000	0.0000	0.0000	0.0142	99.7015
Built-up Areas	0.0082	0.6656	0.0192	0.0192	0.7122	93.4615
Vegetation	0.0007	0.0011	0.1963	0.0011	0.1992	98.5619
Bare Soil	0.0000	0.0092	0.0059	0.0593	0.0744	79.6610
Total	0.0231	0.6759	0.2214	0.0796	1.0000	
PA (%)	61.5403	98.4693	88.6827	74.4647		
OA (%)	93.5400					
Kappa	0.8620					
SVM - 2023						
LULC Class	**Water Bodies**	**Built-up Areas**	**Vegetation**	**Bare Soil**	**Total**	**UA (%)**
Water Bodies	0.0131	0.0006	0.0000	0.0000	0.0137	95.6386
Built-up Areas	0.0062	0.6484	0.0082	0.0103	0.6731	96.3303
Vegetation	0.0005	0.0037	0.2413	0.0041	0.2496	96.6972
Bare Soil	0.0007	0.0137	0.0116	0.0377	0.0637	59.1398
Total	0.0204	0.6664	0.2612	0.0521	1.0000	
PA (%)	64.1150	97.3053	92.3887	72.3267		
OA (%)	94.0435					
Kappa	0.8766					

Table 4.2: Accuracy Assessment of RF-Based LULC Classification in 2007, 2015 and 2023.

RF - 2007						
LULC Class	**Water Bodies**	**Built-up Areas**	**Vegetation**	**Bare Soil**	**Total**	**UA (%)**
Water Bodies	0.0309	0.0000	0.0003	0.0001	0.0313	98.5714
Built-up Areas	0.0065	0.4923	0.0194	0.0194	0.5376	91.5663
Vegetation	0.0000	0.0028	0.3673	0.0028	0.3729	98.5130
Bare Soil	0.0011	0.0011	0.0000	0.0559	0.0581	96.1538
Total	0.0385	0.4962	0.3870	0.0782	1.0000	
PA (%)	80.2809	99.2161	94.9015	71.4267		
OA (%)	94.6427					
Kappa	0.9081					
RF - 2015						
LULC Class	**Water Bodies**	**Built-up Areas**	**Vegetation**	**Bare Soil**	**Total**	**UA (%)**
Water Bodies	0.0159	0.0000	0.0001	0.0001	0.0161	98.8131
Built-up Areas	0.0106	0.6599	0.0452	0.0266	0.7423	88.8889
Vegetation	0.0006	0.0012	0.1748	0.0010	0.1776	98.4462
Bare Soil	0.0000	0.0064	0.0024	0.0551	0.0639	86.1635
Total	0.0271	0.6675	0.2225	0.0828	1.0000	
PA (%)	58.5487	98.8585	78.5451	66.5635		
OA (%)	90.5708					
Kappa	0.7946					
RF - 2023						
Land Use Class	**Water Bodies**	**Built-up Areas**	**Vegetation**	**Bare Soil**	**Total**	**UA (%)**
Water Bodies	0.0137	0.0005	0.0000	0.0000	0.0142	96.1415
Built-up Areas	0.0114	0.6102	0.0000	0.0095	0.6311	96.6767
Vegetation	0.0015	0.0030	0.2634	0.0049	0.2728	96.5580
Bare Soil	0.0036	0.0169	0.0133	0.0480	0.0818	58.6957
Total	0.0302	0.6306	0.2767	0.0625	1.0000	
PA (%)	45.3926	96.7643	95.1807	76.8316		
OA (%)	93.5303					
Kappa	0.8758					

Table 4.3 presents the areal changes in LULC classes based on SVM classification. Over the 16-year period, water bodies shrank by 2.57%, reducing from 9.56 km² to 3.32 km². Built-up areas and bare soil expanded significantly between 2007 and 2015 but experienced contraction from 2015 to 2023, leading to net increases of 13.64% and 0.81%, respectively. Meanwhile, vegetation cover saw a sharp decline of 41.11 km² from 2007 to 2015 but partially recovered by 12.22 km² over the subsequent eight years, resulting in a total loss of 11.88%.

The observed trends align with Malaysia's economic and urban development trajectory. In the 2000s and 2010s, the Malaysian government pursued extensive residential and infrastructure projects, including the Kuala Lumpur Structure Plan 2020, to drive economic growth. This development led to the rapid expansion of built-up areas and bare soil, often at the expense of natural landscapes. However, after 2015, urban expansion slowed due to development saturation and the economic impacts of the COVID-19 pandemic. Consequently, vegetation cover began to recover significantly under multiple urban afforestation initiatives, including the Taman Tugu Urban Forest Park, the Greening KL Program and the Wangsa Maju Zero-Carbon Township Initiative.

Table 4.3: Changes in LULC Areas in Kuala Lumpur Based on SVM Classification from 2007 to 2023.

LULC Class (SVM)	**Area in Different Year (km^2)**			**Changes in Area from 2007 to 2023 (%)**
	2007	**2015**	**2023**	
Water Bodies	9.56	3.46	3.32	-2.57
Built-up Areas	130.44	173.09	163.60	13.64
Vegetation	89.53	48.42	60.64	-11.88
Bare Soil	13.52	18.08	15.48	0.81

4.3 Variation of Historical LST

LST across Kuala Lumpur was retrieved from the thermal bands of Landsat 5 and Landsat 8 for 2007, 2015 and 2023. The DN values were converted to temperature in Celsius through detailed calculations in QGIS. The variation in LST for these years is visualised in Figure 4.3. From 2007 to 2015, most areas of Kuala Lumpur experienced significant warming. Between 2015 and 2023, the northern region continued to heat up, while the central and southern regions exhibited signs of cooling.

Table 4.4 presents the changes in LST across Kuala Lumpur from 2007 to 2023. Over this period, both minimum and maximum LST increased by 43.71% (from 16.10 ℃ to 23.14 ℃) and 10.10% (from 32.46 ℃ to 35.73 ℃), respectively. The mean LST rose sharply from 26.60 ℃ in 2007 to 28.74 ℃ in 2015, followed by a slight decrease to 28.72 ℃ in 2023, resulting in a net increase of 7.99% over the study period.

Table 4.4: Changes in LST Across Kuala Lumpur from 2007 to 2023.

Year	**LST (°C)**		
	Minimum	**Maximum**	**Mean**
2007	16.10	32.46	26.60
2015	19.78	35.06	28.74
2023	23.14	35.73	28.72
Changes in LST from 2007 to 2023 (%)	43.71	10.10	7.99

Table 4.5 summarises the distribution of different LST ranges across Kuala Lumpur. In all three study years, the LST range covering the largest area was between 25 °C and 30 °C, though its coverage declined from 78.90% (191.76 km²) in 2007 to 67.94% (165.12 km²) in 2023. The area experiencing temperatures above 30 °C expanded significantly, from just 1.16% (2.83 km²) in 2007 to 27.60% (67.08 km²) in 2015, and slightly further to 27.85% (67.68 km²) in 2023. This highlights a rapid warming trend, particularly between 2007 and 2015. Meanwhile, the proportion of the area with LST ≤ 20 °C decreased from 0.54% (1.32 km²) in 2007 to zero in both 2015 and 2023, indicating a substantial reduction in cooler zones.

Table 4.5: Distribution of Different LST Ranges Across Kuala Lumpur in 2007, 2015 and 2023.

Year	LST Ranges in °C							
	LST ≤ 20		**20 < LST ≤ 25**		**25 < LST ≤ 30**		**LST > 30**	
	Area (km^2)	**Area (%)**	**Area (km^2)**	**Area (%)**	**Area (km^2)**	**Area (%)**	**Area (km^2)**	**Area (%)**
2007	1.32	0.54	47.14	19.39	191.76	78.90	2.83	1.16
2015	0.00	0.00	9.55	3.93	166.41	68.47	67.08	27.60
2023	0.00	0.00	10.24	4.22	165.12	67.94	67.68	27.85

As discussed in the previous section, large areas of water bodies and vegetation were replaced by impervious built-up surfaces and bare soil between 2007 and 2015. The clustering of these artificial land uses intensified heat retention and contributed to a widespread rise in LST, affecting most areas, especially developed ones. However, between 2015 and 2023, some central and southern regions experienced an increase in vegetation cover, which helped mitigate urban heating, leading to a slight reduction in mean LST. The areas with LST above 30 °C were predominantly concentrated in the northern regions, where industrial and commercial activity was extensive. Notably, light industrial and commercial zones such as Kepong Entrepreneurs Park, Kepong Industrial Park, MARA Industrial Area, Setapak Industrial Area, and The Parc Factory Outlets contributed significantly to localised warming.

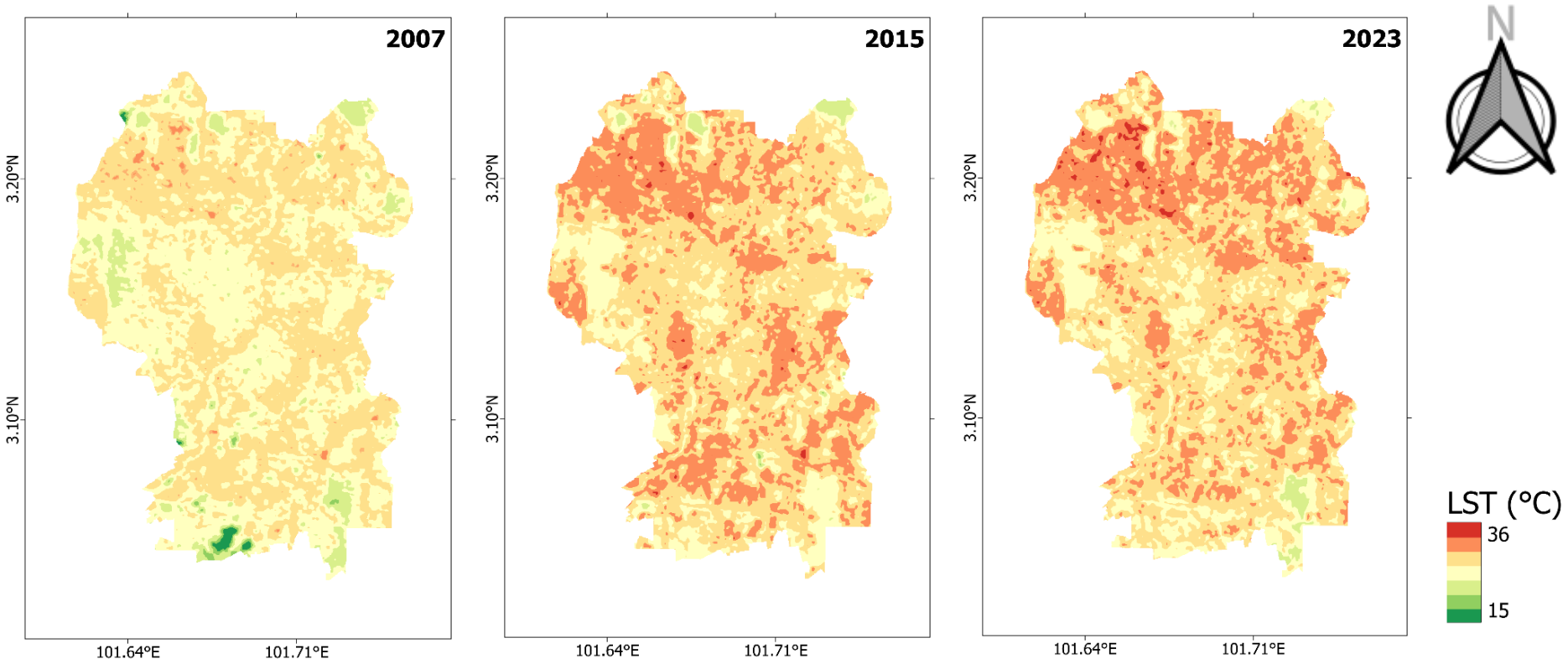


Figure 4.3: Variation of LST Across Kuala Lumpur in 2007, 2015 and 2023.

To evaluate the impact of LULC on LST, the mean LST for different LULC classes is illustrated in Figure 4.4. Throughout the study period, water bodies and vegetation consistently recorded the lowest and second-lowest mean LST values, increasing only slightly from 24.29 °C to 25.20 °C and from 25.64 °C to 26.74 °C, respectively. In contrast, built-up areas experienced a sharp rise in mean LST, increasing from 27.54 °C in 2007 to 29.49 °C in 2015 and peaking at 29.58 °C in 2023. Similarly, bare soil exhibited a notable temperature increase, rising from 25.75 °C in 2007 to 28.10 °C in 2015 and reaching 28.14 °C in 2023, making it the second hottest land cover type.

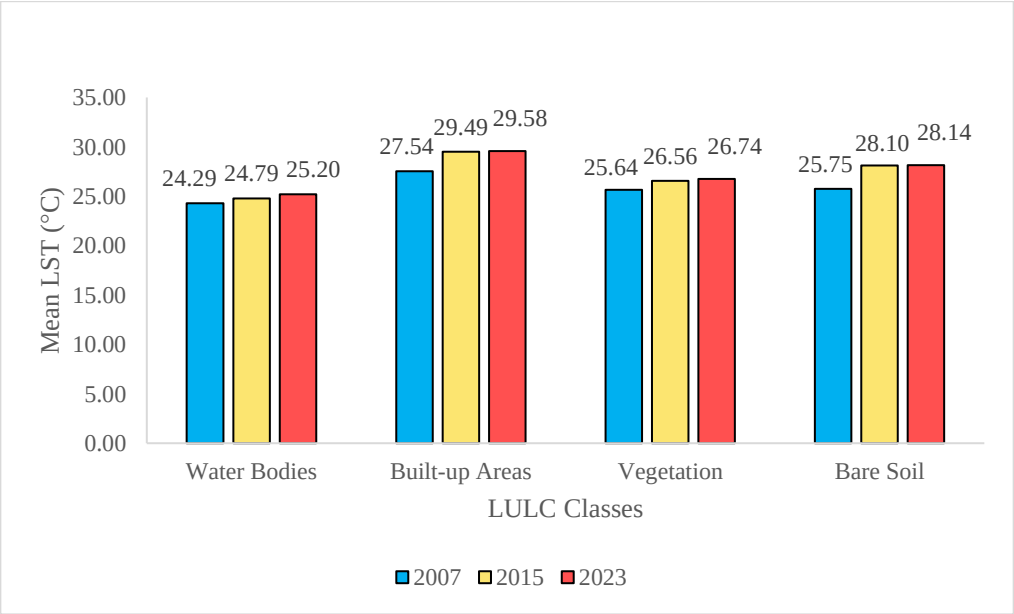


Figure 4.4: Mean LST for Different LULC Classes in 2007, 2015 and 2023.

These observations suggest that LST variation is primarily driven by the biophysical interactions between LULC types and surface thermal properties. LULC alterations directly disrupt the surface energy balance by changing how solar radiation is absorbed, reflected and released. Built-up areas, characterised by impervious materials such as asphalt and concrete, possess high thermal capacity and low albedo, resulting in greater heat absorption and retention. This absorbed heat is subsequently re-emitted, raising LST. Additionally, built-up areas are associated with anthropogenic heat sources from residential, commercial, and industrial activities, further amplifying urban temperatures (Patel, Indraganti, and Jawarneh, 2024;

Tahooni, Kakroodi, and Kiavarz, 2023). Due to their low moisture content and lack of vegetation, bare soil areas exhibit similar thermal properties to built-up areas, efficiently absorbing and storing heat over extended periods (Mahata et al., 2024; Agrawal, Pandey, and Tiwari, 2023). Conversely, water bodies and vegetated areas provide significant cooling effects through evapotranspiration and shading, contributing to the cool island effect (Jia et al., 2024; Wu et al., 2021; Tahooni, Kakroodi, and Kiavarz, 2023).

4.4 Variation of Historical SUHII

The SUHII indicator was used to assess the impact of UHI in Kuala Lumpur by subtracting the mean LST of vegetated areas from the LST of individual pixels. SUHII values for 2007, 2015 and 2023 were calculated using QGIS and the results are illustrated in Figure 4.5. In 2007, most urban areas in Kuala Lumpur experienced low (yellow) to moderate (gold) UHI, while early signs of high (orange) UHI were observed in northern regions such as Kepong and Segambut. By 2015, rapid urban expansion and population growth caused significant UHI intensification, with multiple locations, including Kepong, Segambut, Sentul, Semarak, Bukit Bintang, Pudu, Bangsar, Cheras, Taman OUG and Sri Petaling, experiencing high (orange) to very high (red) UHI over large areas. In 2023, UHI impacts in the central and southern regions were partially mitigated by urban afforestation initiatives, reducing UHI levels to low (yellow) and moderate (gold). However, commercial and industrial growth in the northern region led to continued UHI intensification, with expansions of high (orange) and very high (red) UHI areas. Throughout the study period, locations with water bodies and dense vegetation, such as Bukit Kiara, Taman Tugu, and the Sungai Besi Forest Reserve, consistently exhibited none (blue/light green) UHI due to their strong cooling effects. Bukit Jalil was the only township in Kuala Lumpur that consistently exhibited minimal or no UHI impact.

Table 4.6 presents the statistical changes in SUHII across Kuala Lumpur from 2007 to 2023. The minimum SUHII remained negative but increased by 62.29% from -9.54 to -3.60, reflecting a decline in the cooling effect of vegetated and water-covered areas due to urbanisation. The maximum SUHII increased by 31.99%, rising from 6.82 to 9.00 over the 16-

year period, primarily due to industrial expansion in the northern region. Meanwhile, the mean SUHII rose by 107.37%, from 0.96 in 2007 to 2.18 in 2015, before slightly declining to 1.98 in 2023. This scenario suggests that while urbanisation remains a dominant factor in UHI intensification, recent mitigation efforts have had some positive impact.

Table 4.7 summarises the spatial distribution of different SUHII levels across Kuala Lumpur. In 2007, areas with none and low UHI were the most dominant, but by 2015, there was a significant reduction in these categories, accompanied by an increase in moderate and high UHI. The none UHI category decreased drastically from 66.34 km² (27.31%) in 2007 to 33.77 km² (13.90%) in 2015, before slightly recovering to 40.66 km² (16.74%) in 2023. Similarly, low UHI declined from 110.15 km² (45.35%) to 71.49 km² (29.43%) over the same period, before increasing slightly to 76.42 km² (31.46%) in 2023. Conversely, moderate and high UHI peaked in 2015, with moderate UHI expanding from 63.63 km² (26.19%) in 2007 to 93.47 km² (38.48%) in 2015, before declining to 87.10 km² (35.86%) in 2023. High UHI increased sharply from 2.81 km² (1.16%) in 2007 to 42.18 km² (17.37%) in 2015, before contracting to 35.97 km² (14.81%) in 2023. Although very high UHI remained relatively small in coverage, it showed a gradual increase, from 0.02 km² (0.01%) in 2007 to 1.97 km² (0.81%) in 2015, and further to 2.81 km² (1.16%) in 2023. The extremely high UHI category was absent in 2007 and exhibited only minor increases, expanding by 0.01 km² every eight years.

The overall trend from 2007 to 2015 shows a significant increase in UHI levels across all categories except none and low UHI, reinforcing the dominant role of urbanisation in UHI intensification. However, between 2015 and 2023, a divergent trend emerged, where all UHI levels increased in coverage except for moderate and high UHI. This shift can be attributed to the success of urban landscaping and afforestation efforts in the central and southern regions, which helped mitigate UHI impacts. In contrast, the northern region continued to experience UHI intensification, driven by rapid commercial and industrial expansion.

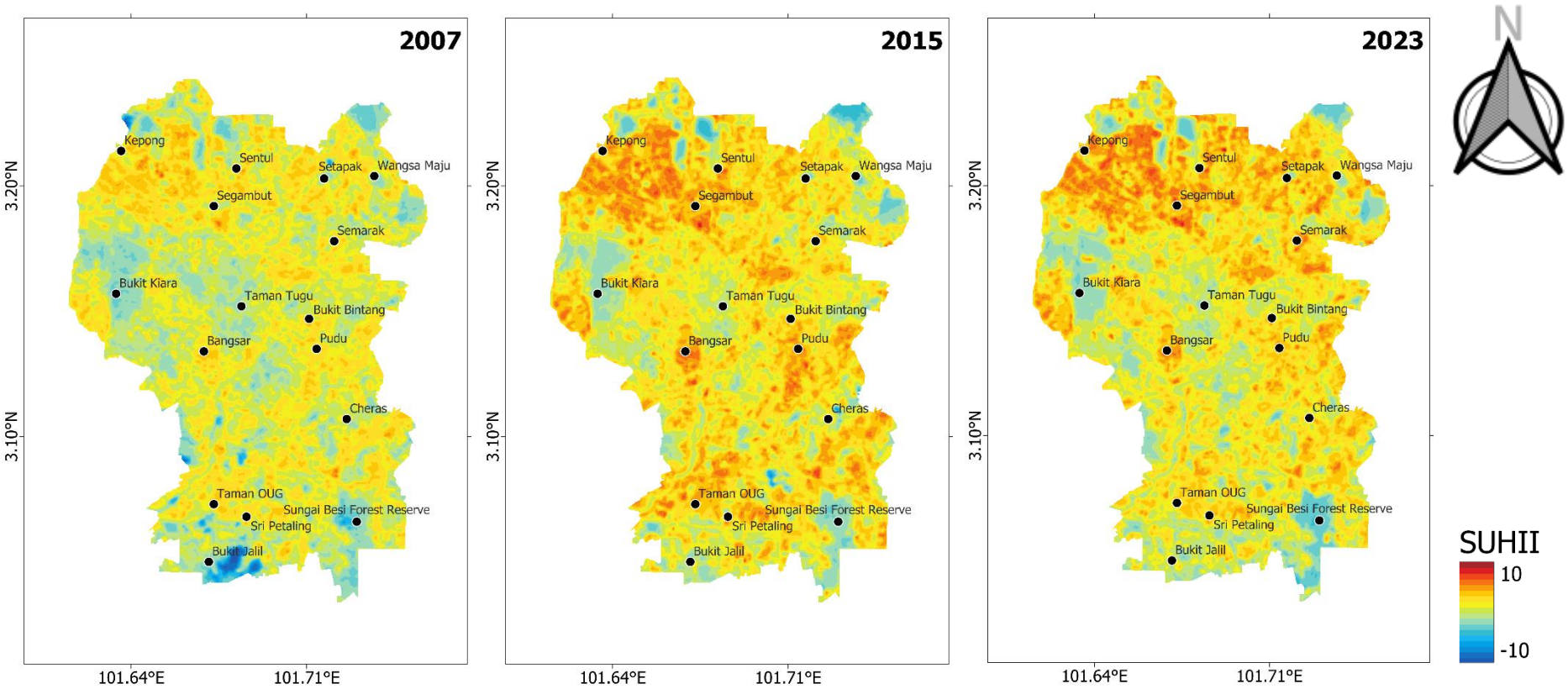


Figure 4.5: Variation of SUHII Across Kuala Lumpur in 2007, 2015 and 2023.

Table 4.6: Changes in SUHII Across Kuala Lumpur from 2007 to 2023.

Year	SUHII		
	Minimum	**Maximum**	**Mean**
2007	-9.54	6.82	0.96
2015	-6.77	8.50	2.18
2023	-3.60	9.00	1.98
Change in SUHII from 2007 to 2023 (%)	62.29	31.99	107.37

Table 4.7: Distribution of Different SUHII Ranges Across Kuala Lumpur in 2007, 2015 and 2023.

Year	SUHII Ranges											
	None		**Low**		**Moderate**		**High**		**Very High**		**Extremely High**	
	SUHII ≤ 0		**0 < SUHII ≤ 2.0**		**2.0 < SUHII ≤ 4.0**		**4.0 < SUHII ≤ 6.0**		**6.0 < SUHII ≤ 8.0**		**SUHII > 8.0**	
	Area (km^2)	**Area (%)**	**Area (km^2)**	**Area (%)**	**Area (km^2)**	**Area (%)**	**Area (km^2)**	**Area (%)**	**Area (km^2)**	**Area (%)**	**Area (km^2)**	**Area (%)**
2007	66.34	27.31	110.15	45.35	63.62	26.19	2.81	1.16	0.02	0.01	0.00	0.00
2015	33.77	13.90	71.49	29.43	93.47	38.48	42.18	17.37	1.97	0.81	0.01	0.01
2023	40.66	16.74	76.42	31.46	87.10	35.86	35.97	14.81	2.81	1.16	0.02	0.01

4.5 Rainfall Analysis

Rainfall data from nine selected stations covering the period 2007-2023 were processed using Python and QGIS to conduct a comprehensive rainfall analysis. The IDW interpolation method was applied to address missing data by estimating values based on rainfall measurements from nearby stations within the same period. The original 15-minute data were aggregated into hourly rainfall data. Extreme rainfall events were identified using the 99th percentile threshold, with the corresponding threshold values provided in Appendix A. Thiessen polygons were created to define the spatial extent of influence for each rainfall station, resulting in the delineation of nine distinct study areas.

Multiple analyses were conducted, focusing on the spatiotemporal variation, diurnal distribution and trends of hourly extreme rainfall to explore variations in rainfall patterns, including the total number of hourly extreme rainfall events, the total of hourly extreme rainfall and the mean of hourly extreme rainfall during the study period.

4.5.1 Spatiotemporal Variation of Hourly Extreme Rainfall

Given the high variability in Malaysia's annual rainfall patterns, the dataset was segmented into two study periods (2007 - 2015 and 2015 - 2023) to facilitate comparative analyses of hourly extreme rainfall.

Table 4.8 and Figure 4.6 show the total number of hourly extreme rainfall events across the study areas, revealing a notable increase at all stations. The Kolam Takungan Batu Station recorded the lowest increase, with an additional 83 events (11.81%), while the Pusat Penyelidekan At Jps Ampang Selangor Station experienced the highest rise, with 152 more hourly extreme rainfall events (22.72%). Other stations also exhibited significant increases, with the Ldg. Edinburgh Site 2 At W. Persekutuan Station recording 142 additional events (21.07%) and the Km 10 Ulu Kelang At UK Height Station observing an increase of 118 events (17.28%).

Table 4.8: Total Number of Hourly Extreme Rainfall Events at Selected Stations for the Periods 2007-2015 and 2015-2023.

No.	Station	Total Number of Hourly Extreme Rainfall Events		Change (%)
		2007-2015	2015-2023	
1	**Kolam Takungan Batu (0231391RF)**	703	786	11.81
2	**Taman Ehsan At Kepong W. Persekutuan (0231441RF)**	680	783	15.15
3	**Pusat Penyelidekan At Jps Ampang Selangor (0231351RF)**	669	821	22.72
4	**Km 10 Ulu Kelang At Uk Height (0231401RF)**	683	801	17.28
5	**Ibu Bekalan Km. 11 At Gombak W. Persekutuan (0230721RF)**	688	797	15.84
6	**Empangan Genting Klang At W. Persekutuan (0230631RF)**	703	787	11.95
7	**I/pejabat Jps Malaysia At W. Persekutuan (0230641RF)**	685	784	14.45
8	**Jln. Sg. Udang At Segambut (0231381RF)**	702	787	12.11
9	**Ldg. Edinburgh Site 2 At W. Persekutuan (0230651RF)**	674	816	21.07

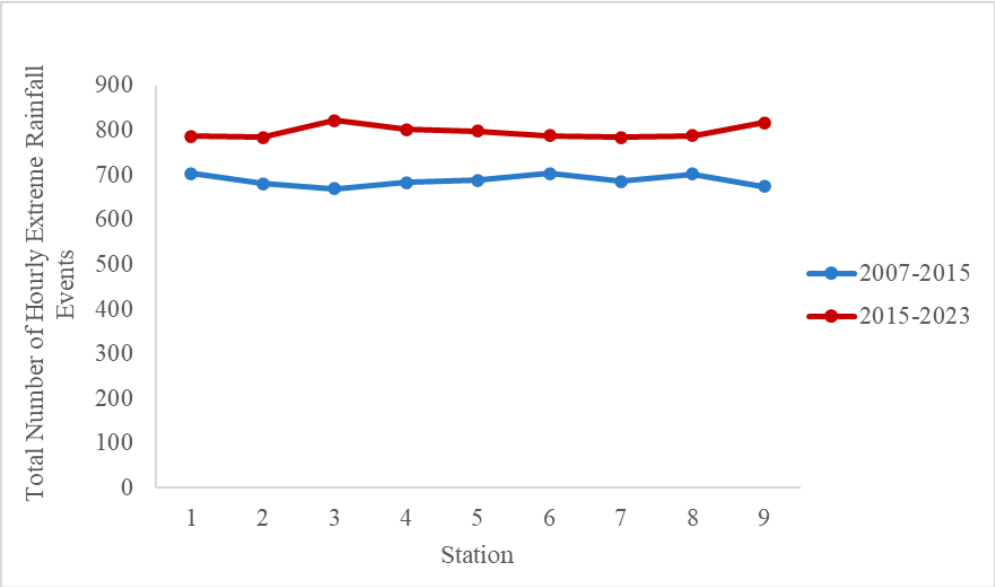


Figure 4.6: Total Number of Hourly Extreme Rainfall Events at Selected Stations for the Periods 2007-2015 and 2015-2023.

The spatial distribution for the total number of hourly extreme rainfall events, depicted in Figure 4.7, further supports this shift. The deepening of colour tones in the 2015 - 2023 map suggests a widespread intensification of hourly extreme rainfall occurrences across all study areas. During the 2007 - 2015 period, the total number of hourly extreme rainfall events ranged from 669 at the Pusat Penyelidekan At Jps Ampang Selangor Station to 703 at both the Kolam Takungan Batu and Empangan Genting Klang At W. Persekutuan Stations. During the 2015 - 2023 period, all study areas recorded an increase in the total number of events, ranging from 783 events at the Taman Ehsan At Kepong W. Persekutuan Station to 821 events at the Pusat Penyelidekan At Jps Ampang Selangor Station.

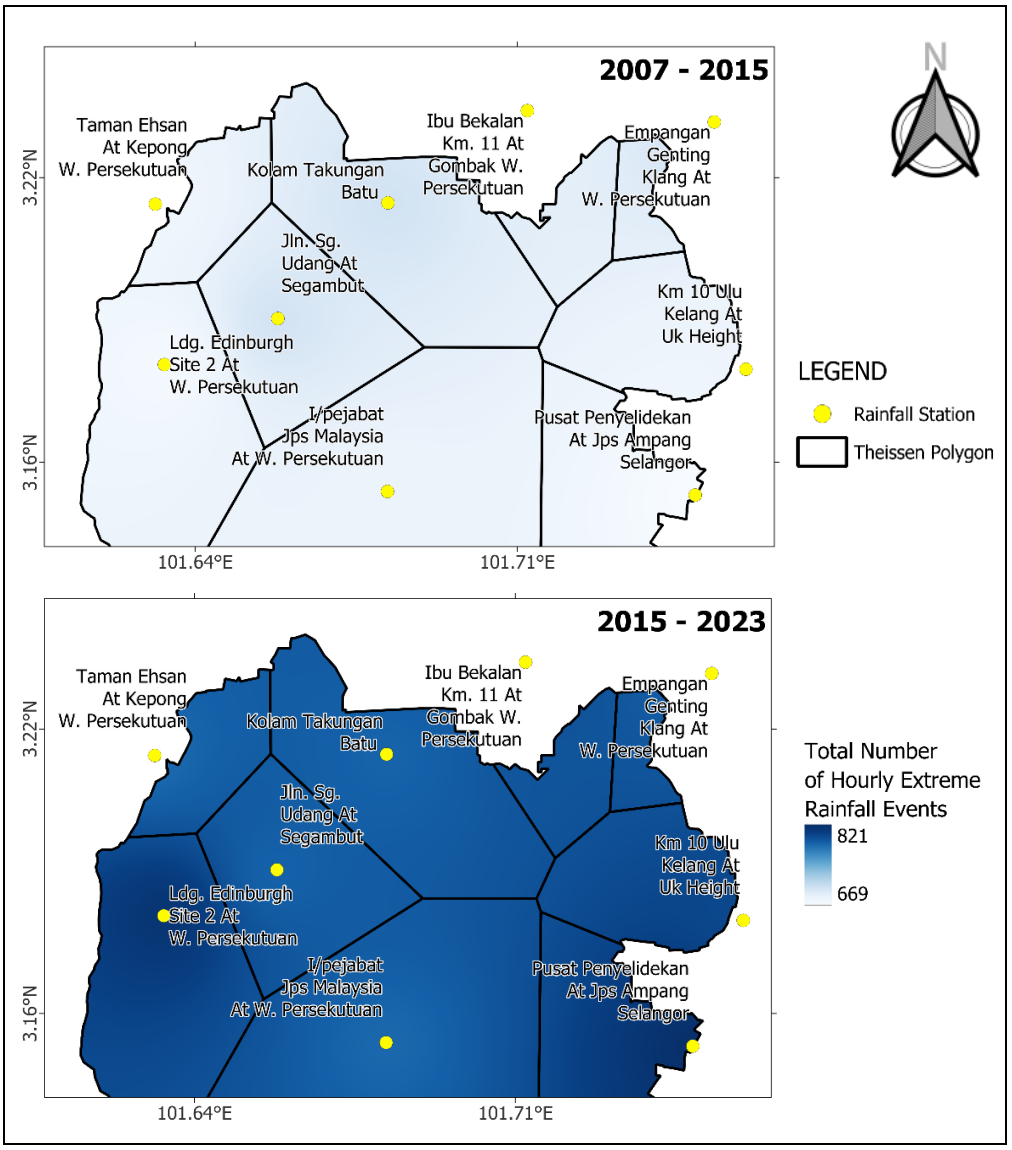


Figure 4.7: Spatial Distribution for the Total Number of Hourly Extreme Rainfall Events Across the Study Areas for the Periods 2007-2015 and 2015-2023.

The total of hourly extreme rainfall for both study periods at different stations is summarised in Table 4.9 and Figure 4.8. A consistent increase was observed at all stations, with increments ranging from 880.7 mm (6.10%) at the Empangan Genting Klang At W. Persekutuan Station to 3579.88 mm (23.90%) at the Pusat Penyelidekan at JPS Ampang Selangor Station. Notably, four stations, including Taman Ehsan At Kepong W. Persekutuan, Pusat Penyelidekan At Jps Ampang Selangor, Km 10 Ulu Kelang At Uk Height and Ldg. Edinburgh Site 2 At W. Persekutuan, experienced substantial increases of

at least 2514.18 mm (18.06%), while the remaining stations showed moderate increases of up to 2307.52 mm (14.68%).

Table 4.9: Total of Hourly Extreme Rainfall (mm) at Selected Stations for the Periods 2007-2015 and 2015-2023.

No.	Station	Total of Hourly Extreme Rainfall (mm)		Change (%)
		2007-2015	2015-2023	
1	**Kolam Takungan Batu (0231391RF)**	15104.73	17290.30	14.47
2	**Taman Ehsan At Kepong W. Persekutuan (0231441RF)**	13918.48	16432.66	18.06
3	**Pusat Penyelidekan At Jps Ampang Selangor (0231351RF)**	14978.92	18558.80	23.90
4	**Km 10 Ulu Kelang At Uk Height (0231401RF)**	15195.40	18338.30	20.68
5	**Ibu Bekalan Km. 11 At Gombak W. Persekutuan (0230721RF)**	14590.20	16159.90	10.76
6	**Empangan Genting Klang At W. Persekutuan (0230631RF)**	14446.20	15326.90	6.10
7	**I/pejabat Jps Malaysia At W. Persekutuan (0230641RF)**	15718.74	18026.26	14.68
8	**Jln. Sg. Udang At Segambut (0231381RF)**	15450.87	17498.90	13.26
9	**Ldg. Edinburgh Site 2 At W. Persekutuan (0230651RF)**	15674.92	19157.95	22.22

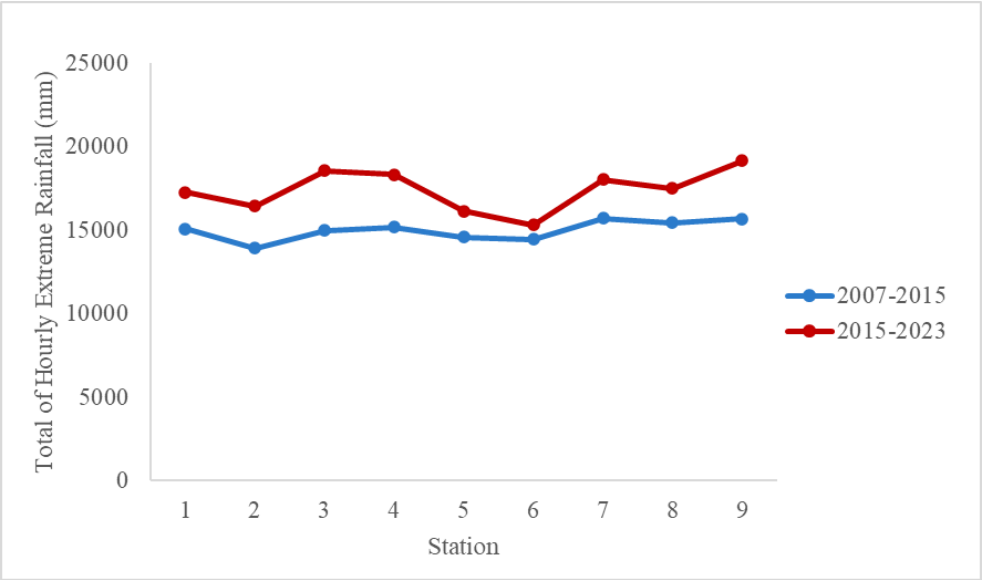


Figure 4.8: Total of Hourly Extreme Rainfall (mm) at Selected Stations for the Periods 2007-2015 and 2015-2023.

The spatial distribution for the total of hourly extreme rainfall, illustrated in Figure 4.9, highlights a steady increase across all study areas. During the 2007-2015 period, the total of hourly extreme rainfall ranged from 13952.83 mm to 15718.72 mm, with the I/Pejabat Jps Malaysia At W. Persekutuan Station recording the highest value, followed by the Ldg. Edinburgh Site 2 At W. Persekutuan and Jln. Sg. Udang At Segambut Stations. In the second study period (2015 - 2023), rainfall increased substantially, ranging from 15727.89 mm to 19157.95 mm, with the Ldg. Edinburgh Site 2 At W. Persekutuan Station experiencing the highest total of hourly extreme rainfall, followed by the Pusat Penyelidekan At JPS Ampang Selangor and Km 10 Ulu Kelang At Uk Height Stations.

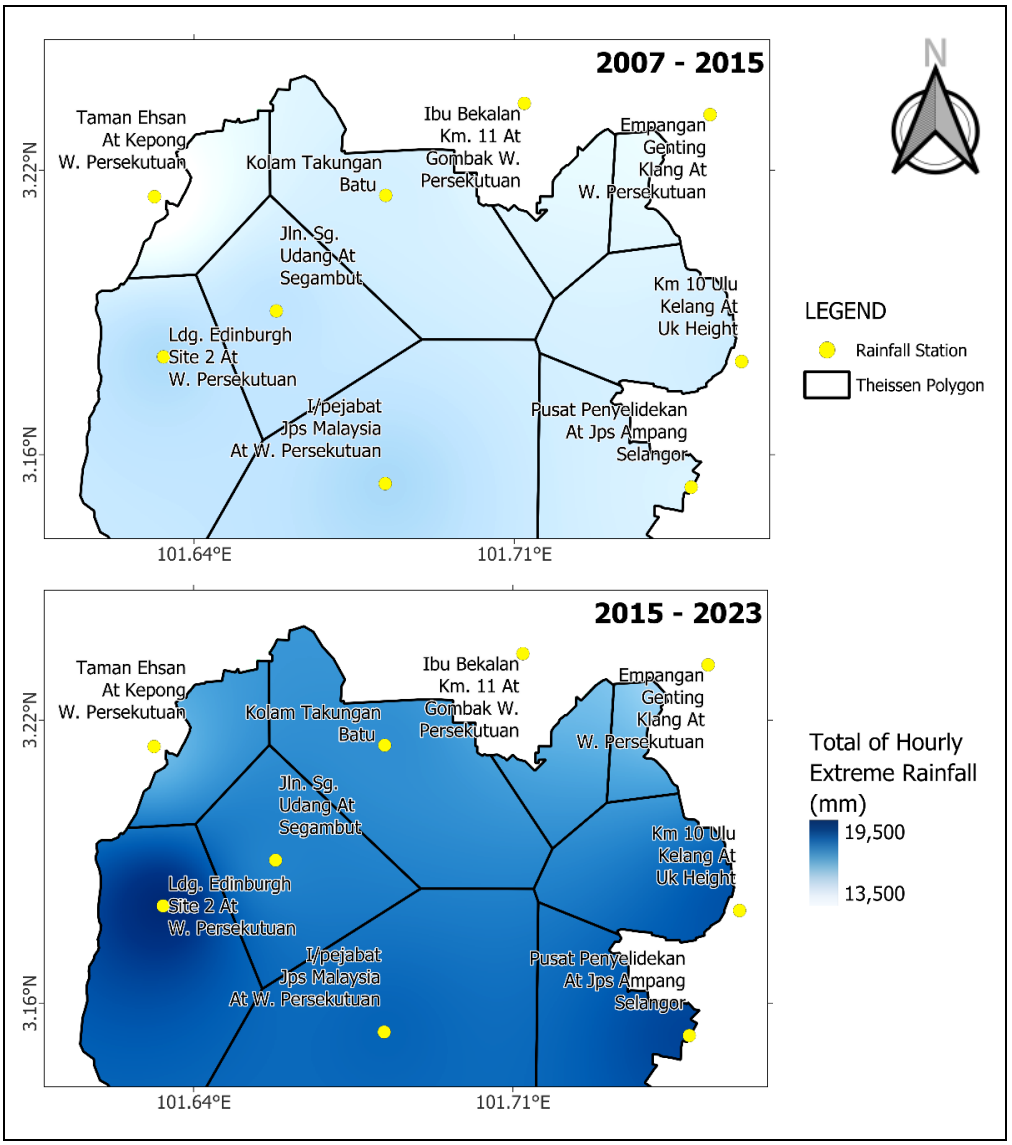


Figure 4.9: Spatial Distribution for the Total of Hourly Extreme Rainfall (mm) across the Study Areas for the Periods 2007-2015 and 2015-2023.

Table 4.10 and Figure 4.10 present the mean of hourly extreme rainfall at different stations for the two time periods. Unlike the total of hourly extreme rainfall, the mean of hourly extreme rainfall did not exhibit significant changes across stations. Two stations, Ibu Bekalan Km. 11 At Gombak W. Persekutuan and Empangan Genting Klang At W. Persekutuan, showed a slight decrease in the mean of hourly extreme rainfall by 0.98 mm (-4.38%) and 1.07 mm (-5.21%), respectively, while the other seven stations recorded slight increases ranging from 0.04 mm (0.17%) to 0.64 mm (2.88%).

Table 4.10: Mean of Hourly Extreme Rainfall (mm) at Selected Stations for the Periods 2007-2015 and 2015-2023.

No.	Station	Mean of Hourly Extreme Rainfall (mm)		Change (%)
		2007-2015	2015-2023	
1	**Kolam Takungan Batu (0231391RF)**	21.49	22	2.37
2	**Taman Ehsan At Kepong W. Persekutuan (0231441RF)**	20.47	20.99	2.54
3	**Pusat Penyelidekan At Jps Ampang Selangor (0231351RF)**	22.39	22.61	0.98
4	**Km 10 Ulu Kelang At Uk Height (0231401RF)**	22.25	22.89	2.88
5	**Ibu Bekalan Km. 11 At Gombak W. Persekutuan (0230721RF)**	21.21	20.28	-4.38
6	**Empangan Genting Klang At W. Persekutuan (0230631RF)**	20.55	19.48	-5.21
7	**I/pejabat Jps Malaysia At W. Persekutuan (0230641RF)**	22.95	22.99	0.17
8	**Jln. Sg. Udang At Segambut (0231381RF)**	22.01	22.23	1.00
9	**Ldg. Edinburgh Site 2 At W. Persekutuan (0230651RF)**	23.26	23.48	0.95

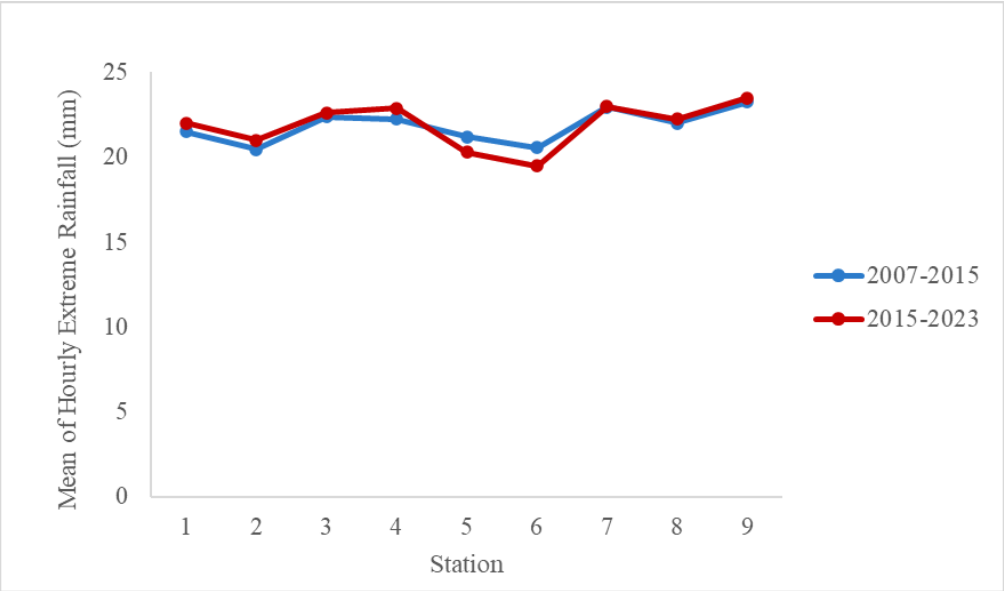


Figure 4.10: Mean of Hourly Extreme Rainfall (mm) at Selected Stations for the Periods 2007-2015 and 2015-2023.

The spatial distribution for the mean of hourly extreme rainfall, illustrated in Figure 4.11, indicates that all stations recorded comparable mean hourly extreme rainfall during both periods, ranging from 19.48 mm to 23.48 mm.

The observed increases in both the total number of hourly extreme rainfall events and the total of hourly extreme rainfall, despite minimal changes in the mean of hourly extreme rainfall, suggest a shift in rainfall patterns from 2007 to 2023. This shift is characterised by a rise in the frequency of hourly extreme events, rather than an intensification of individual events. The relatively stable mean values indicate that the increase in the total of hourly extreme rainfall was primarily driven by the growing number of extreme rainfall occurrences, rather than changes in the magnitude of each event.

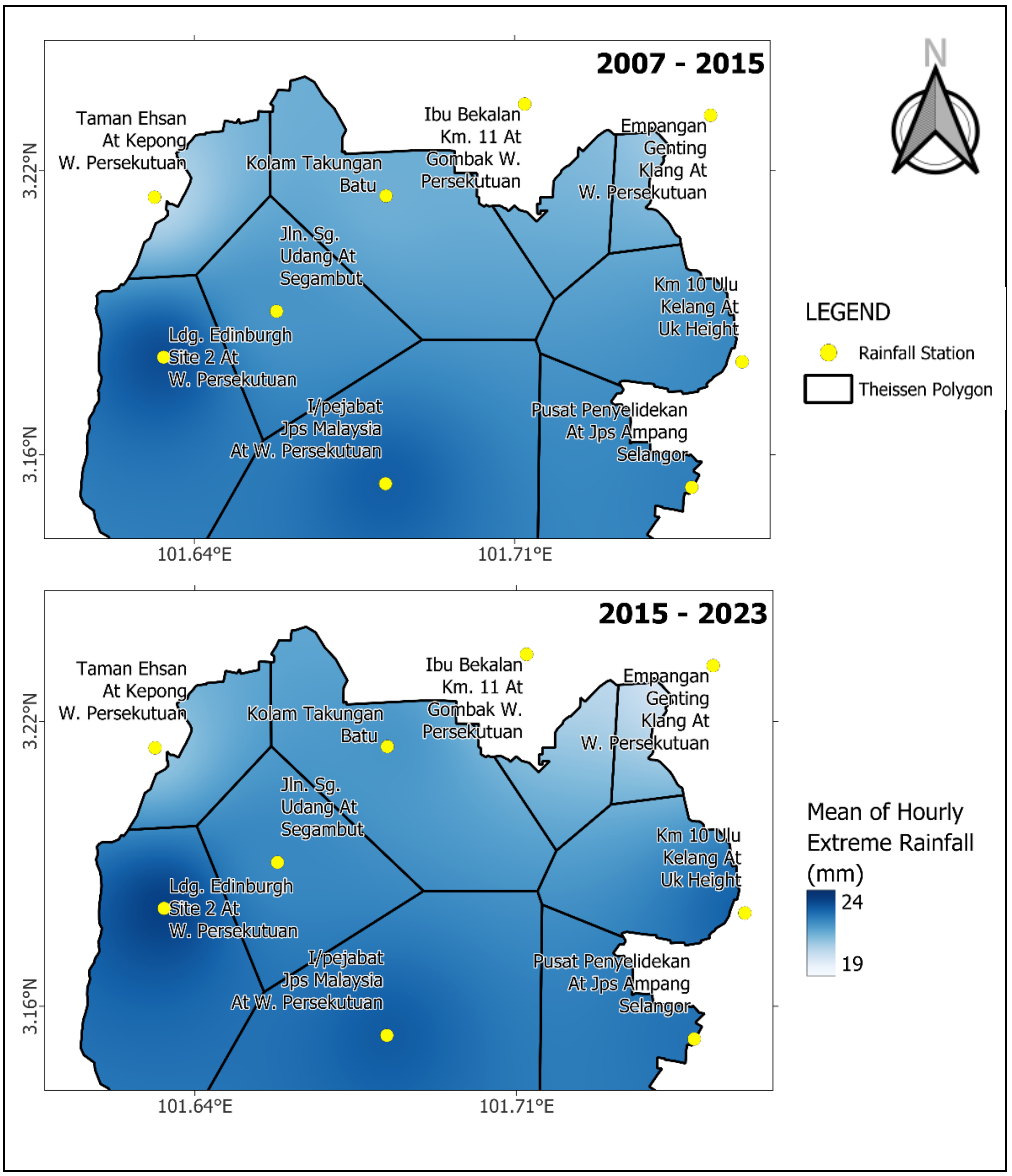


Figure 4.11: Spatial Distribution for the Mean of Hourly Extreme Rainfall (mm) across the Study Areas for the Periods 2007-2015 and 2015-2023.

4.5.2 Diurnal Distribution of the Total Number of Hourly Extreme Rainfall Events

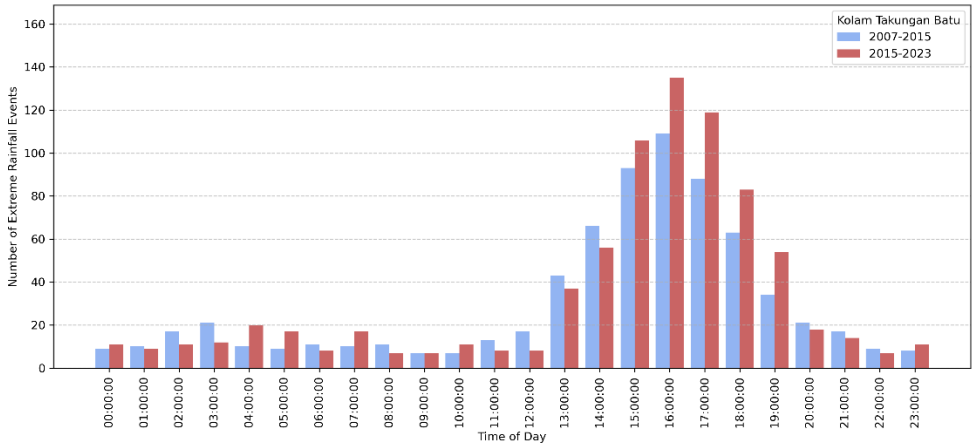
To gain deeper insight into variations in hourly rainfall patterns, the diurnal distribution for the total number of hourly extreme rainfall events across various stations for the periods 2007 - 2015 and 2015 - 2023 is illustrated in Figure 4.12. The peak occurrence of extreme rainfall in both periods was observed between 16:00 and 16:59, with 988 and 1145 events recorded, respectively. Generally, extreme rainfall events were concentrated during the

afternoon and evening hours at all stations, with a notable intensification observed during the 2015 - 2023 period, particularly between 16:00 and 19:59. The most significant increase occurred during 17:00 - 17:59, with an additional 237 events, marking a 27.69% rise compared to the earlier period.

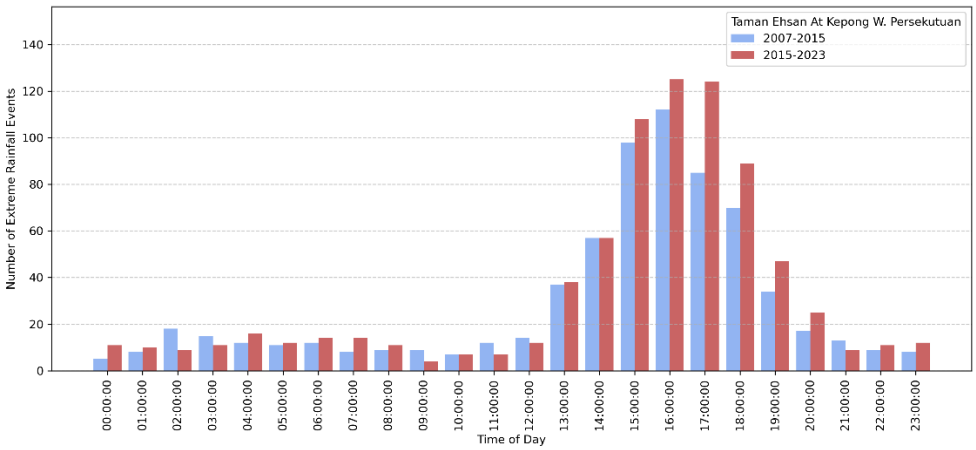
The Kolam Takungan Batu and Ibu Bekalan Km. 11 At Gombak W. Persekutuan Stations exhibited increases in the number of extreme rainfall events during 15:00 - 18:59, with increments ranging from 13 to 31 events and 4 to 26 events, respectively. The Pusat Penyelidekan At Jps Ampang Selangor Station recorded the highest cumulative increase, with 124 events occurring between 14:00 and 19:59. This was followed closely by I/Pejabat Jps Malaysia At W. Persekutuan with 123 additional events during 16:00 - 18:59 and Ldg. Edinburgh Site 2 At W. Persekutuan with 122 additional events during 15:00 - 21:59.

At the Taman Ehsan At Kepong W. Persekutuan Station, the number of extreme rainfall events rose by 102 events between 15:00 and 19:59, with the highest single-hour increase of 39 events occurring at 17:00 - 17:59. The Km 10 Ulu Kelang At UK Heights Station saw increases ranging from 4 to 28 events during the 14:00 - 19:59 window. Empangan Genting Klang At W. Persekutuan and Jln. Sg. Udang At Segambut Stations recorded the smallest total increases, both below 100 events, during the critical 16:00 - 19:59 period.

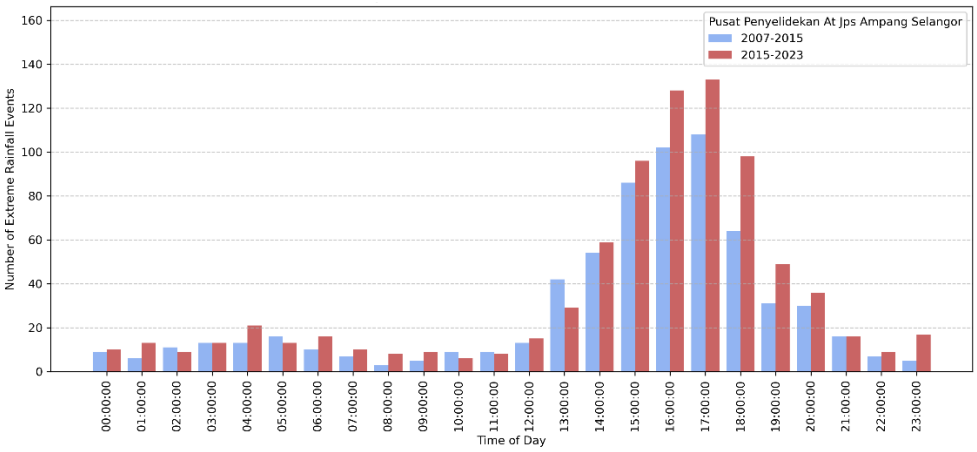
These patterns align with the peak UHI effect, which typically occurs in the late afternoon, suggesting a possible link between urban heat buildup and the triggering of convective rainfall (Li et al., 2016).



(a)

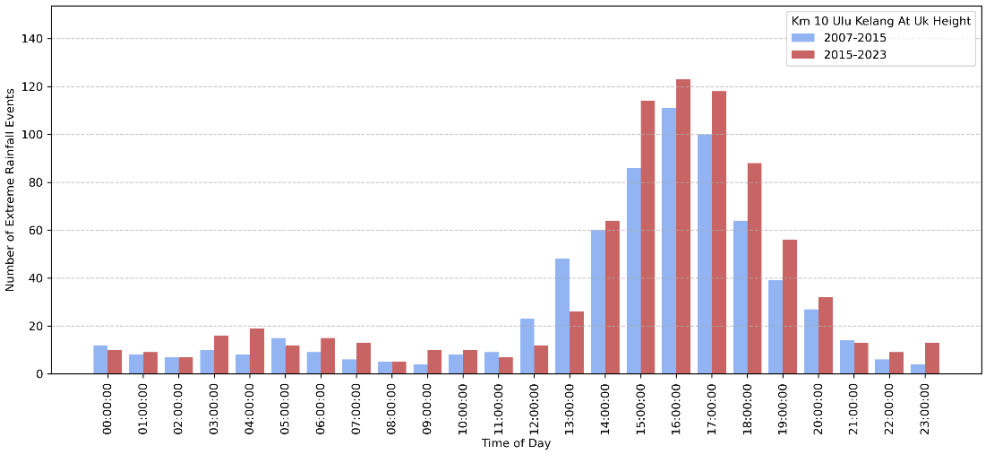


(b)

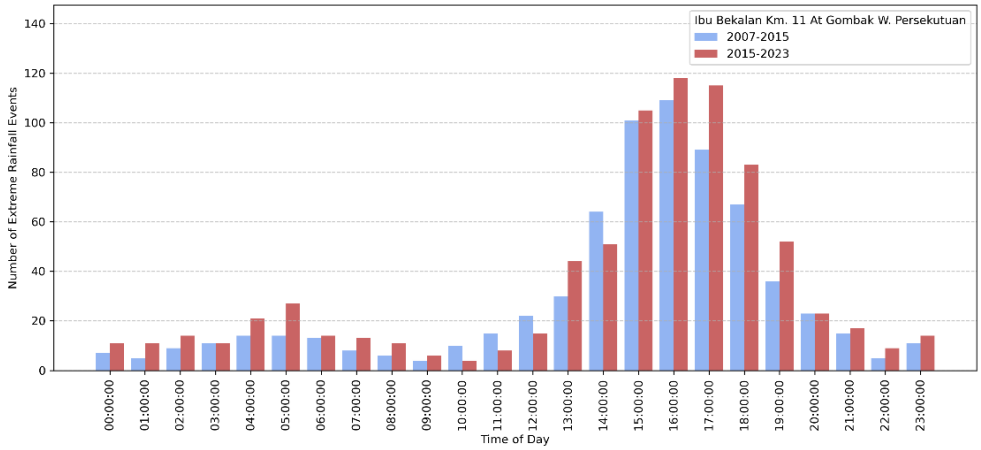


(c)

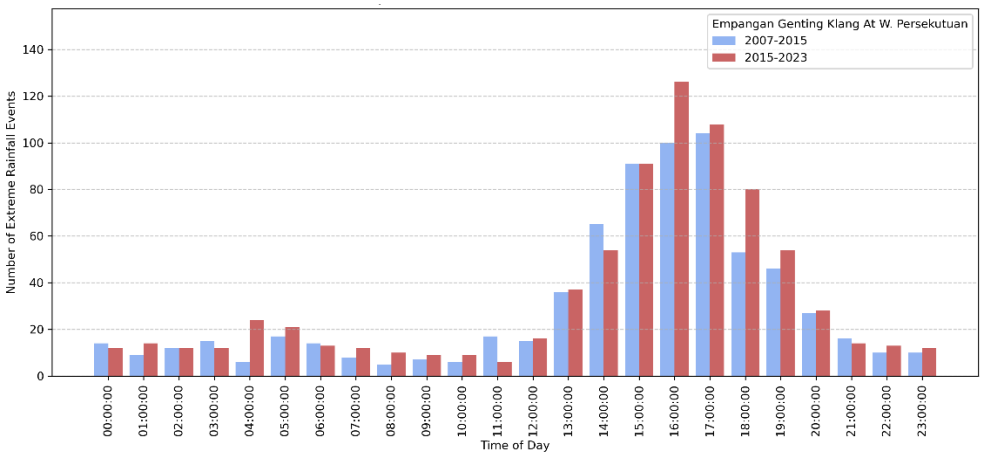
Figure 4.12: Diurnal Distribution for the Total Number of Hourly Extreme Rainfall Events at (a) Kolam Takungan Batu Station, (b) Taman Ehsan At Kepong W. Persekutuan Station and (c) Pusat Penyelidekan At Jps Ampang Selangor Station for the Periods 2007-2015 and 2015-2023.



(d)

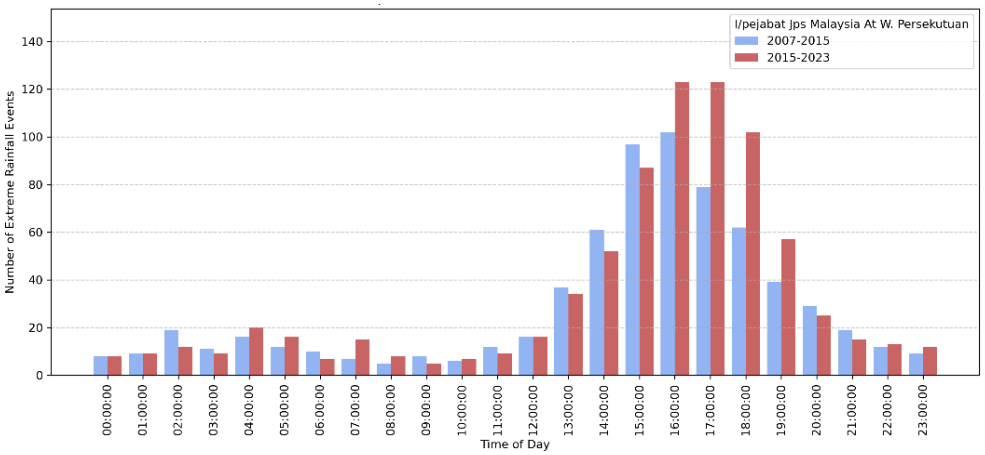


(e)

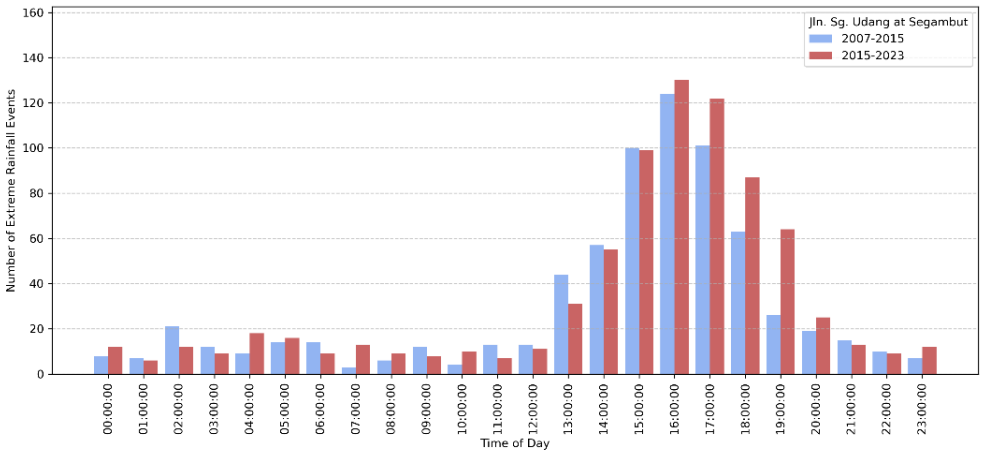


(f)

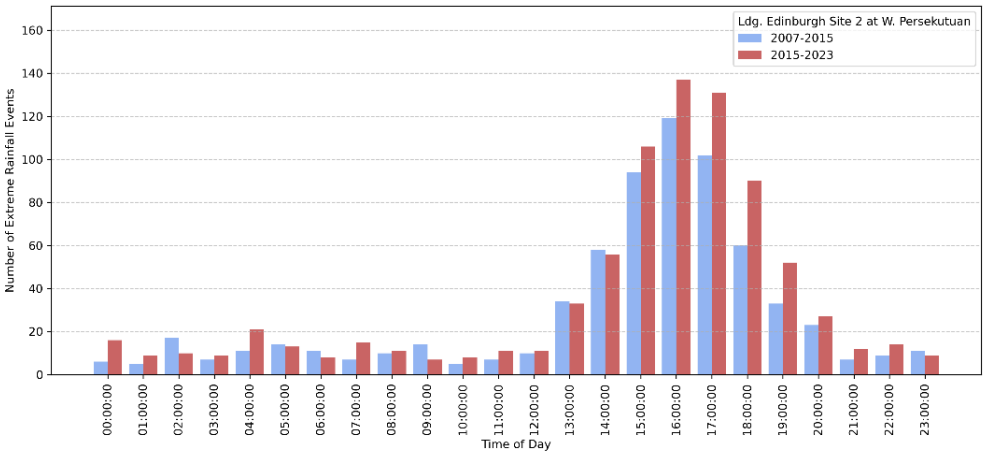
Figure 4.12: Diurnal Distribution for the Total Number of Hourly Extreme Rainfall Events at (d) Km 10 Ulu Kelang At Uk Height Station, (e) Ibu Bekalan Km. 11 At Gombak W. Persekutuan Station and (f) Empangan Genting Klang At W. Persekutuan Station for the Periods 2007-2015 and 2015-2023.



(g)



(h)



(i)

Figure 4.12: Diurnal Distribution for the Total Number of Hourly Extreme Rainfall Events at (g) I/pejabat Jps Malaysia At W. Persekutuan Station, (h) Jln. Sg. Udang At Segambut Station and (i) Ldg. Edinburgh Site 2 At W. Persekutuan Station for the Periods 2007-2015 and 2015-2023.

4.5.3 Trends and Sen's Slope for the Hourly Extreme Rainfall

Using the MMK test and Sen's slope estimator, the trends and rates of change in the annual total number of hourly extreme rainfall events were analysed, as shown in Table 4.11 and Figure 4.13. Among the nine stations, seven exhibited statistically significant positive trends in the annual total number of hourly extreme rainfall events at the 95% confidence level. The station with the highest statistically significant increase was the Pusat Penyelidekan At Jps Ampang Selangor, recording a rise of 2.08 events/year, followed by the Km 10 Ulu Kelang At Uk Height with 1.89 events/year, and the Ldg. Edinburgh Site 2 At W. Persekutuan with 1.85 events/year.

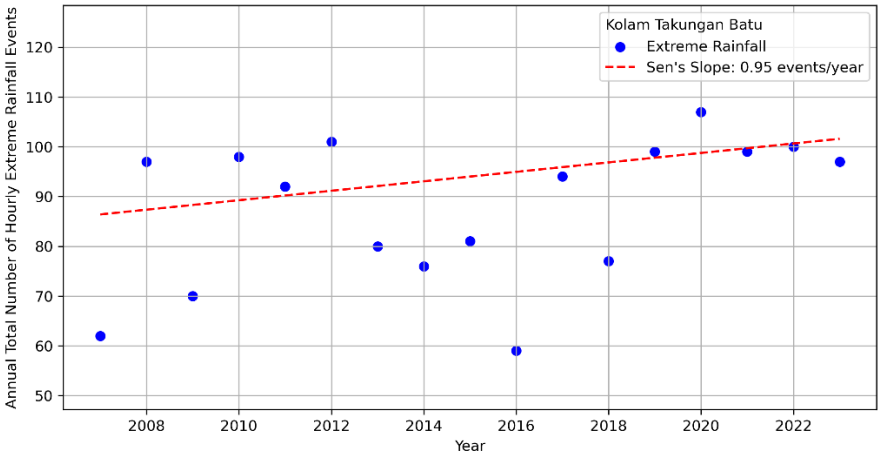
Jln. Sg. Udang At Segambut and Kolam Takungan Batu were the only stations that exhibited non-statistically significant positive trends, with the lowest magnitudes of 0.82 events/year and 0.95 events/year, respectively, suggesting more modest increases likely influenced by high interannual variability.

The widespread presence of statistically significant upward trends in the annual total number of hourly extreme rainfall events across most stations reinforces the indication of increasingly frequent extreme rainfall patterns, particularly in the northern regions of Kuala Lumpur. This phenomenon may be attributed to increasing urbanisation and the associated UHI effects.

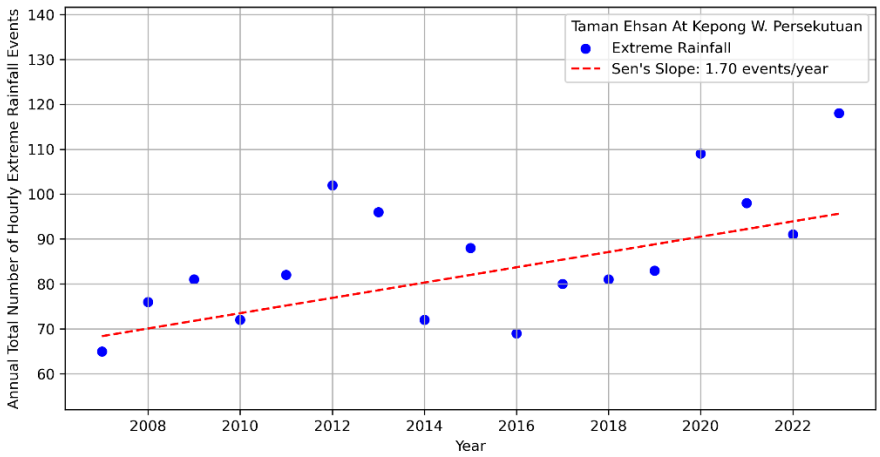
Table 4.11: Trends and Sen's Slope for the Annual Total Number of Hourly Extreme Rainfall Events at Selected Stations from 2007 to 2023.

No.	Station	Trend	Sen's Slope (events/year)
1	**Kolam Takungan Batu (0231391RF)**	Positive	0.95
2	**Taman Ehsan At Kepong W. Persekutuan (0231441RF)**	Positive*	1.70
3	**Pusat Penyelidekan At Jps Ampang Selangor (0231351RF)**	Positive*	2.08
4	**Km 10 Ulu Kelang At Uk Height (0231401RF)**	Positive*	1.89
5	**Ibu Bekalan Km. 11 At Gombak W. Persekutuan (0230721RF)**	Positive*	1.46
6	**Empangan Genting Klang At W. Persekutuan (0230631RF)**	Positive*	1.37
7	**I/pejabat Jps Malaysia At W. Persekutuan (0230641RF)**	Positive*	1.72
8	**Jln. Sg. Udang At Segambut (0231381RF)**	Positive	0.82
9	**Ldg. Edinburgh Site 2 At W. Persekutuan (0230651RF)**	Positive*	1.85

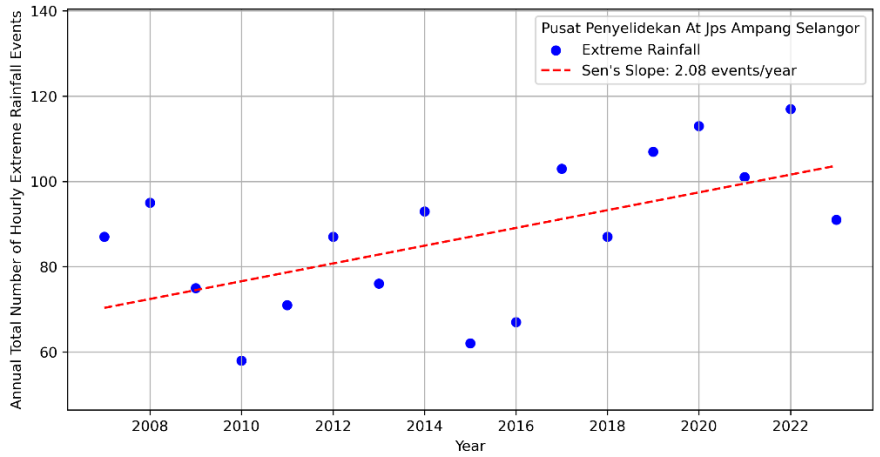
Note: An asterisk (*) indicates statistical significance at 95% confidence level.



(a)

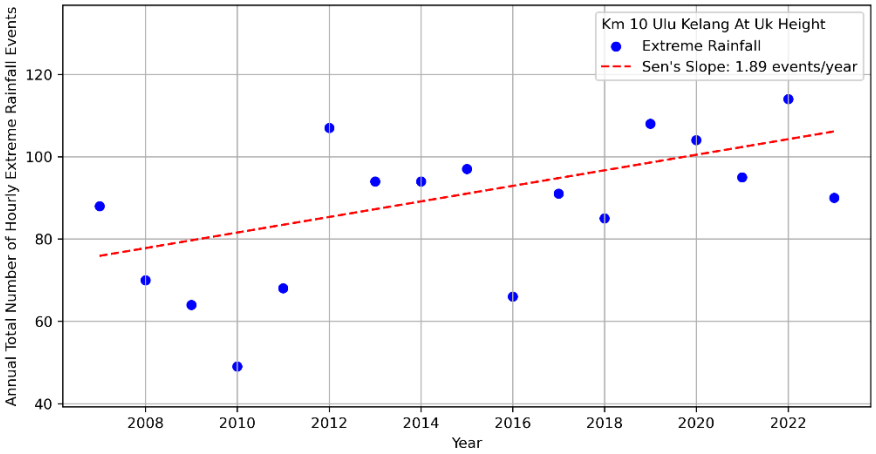


(b)

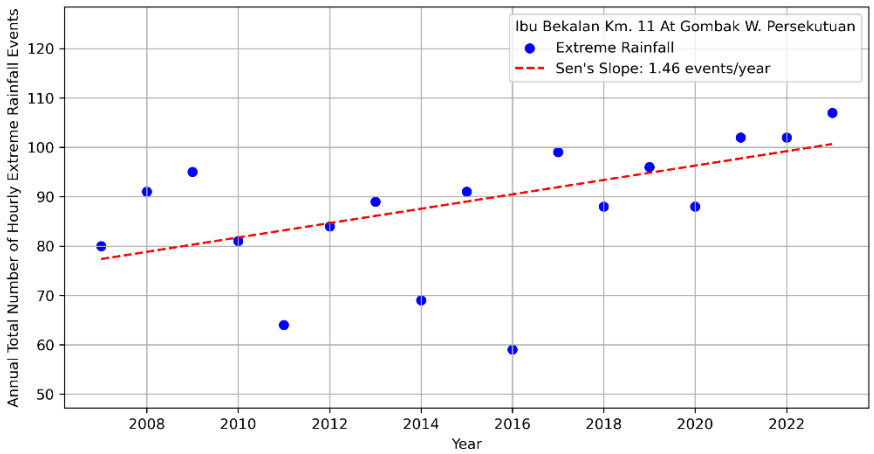


(c)

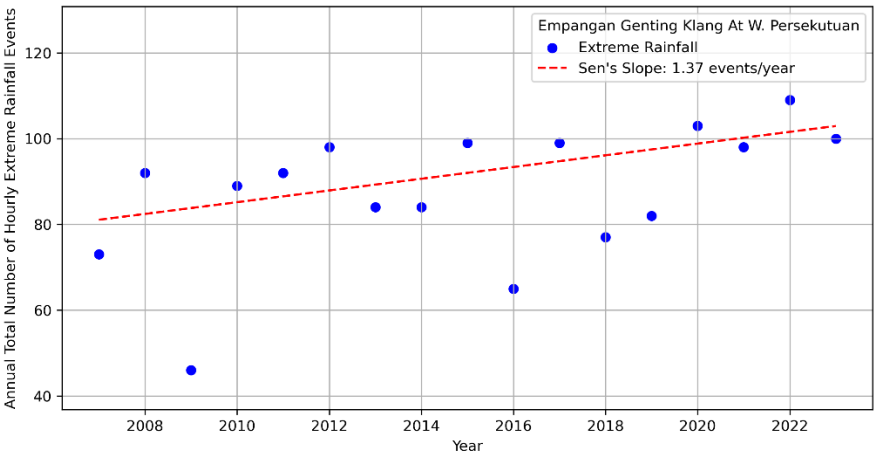
Figure 4.13: Sen's Slope for the Annual Total Number of Hourly Extreme Rainfall Events (events/year) at (a) Kolam Takungan Batu Station, (b) Taman Ehsan At Kepong W. Persekutuan Station and (c) Pusat Penyelidekan At Jps Ampang Selangor Station from 2007 to 2023.



(d)

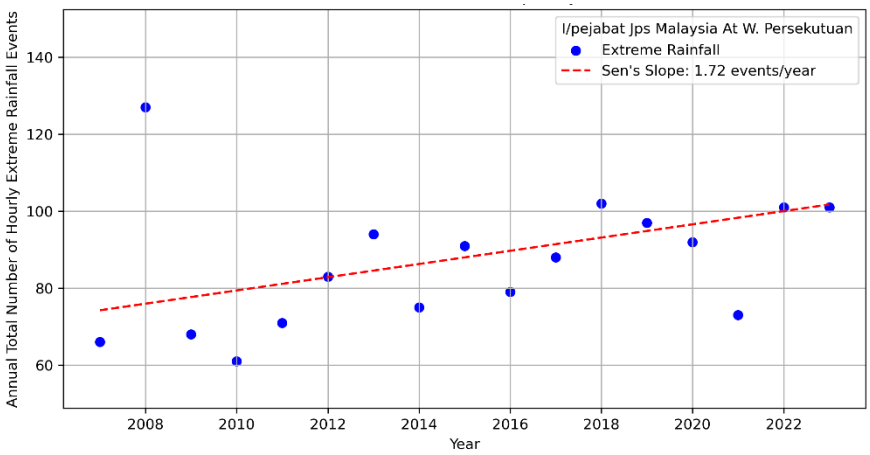


(e)

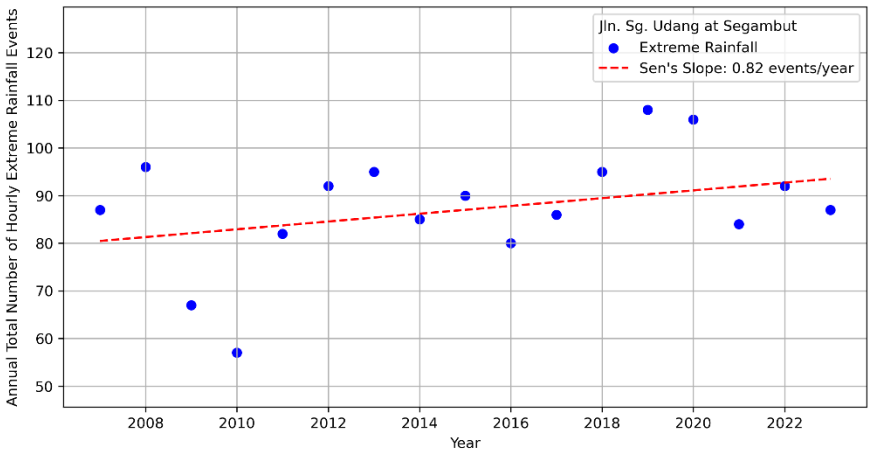


(f)

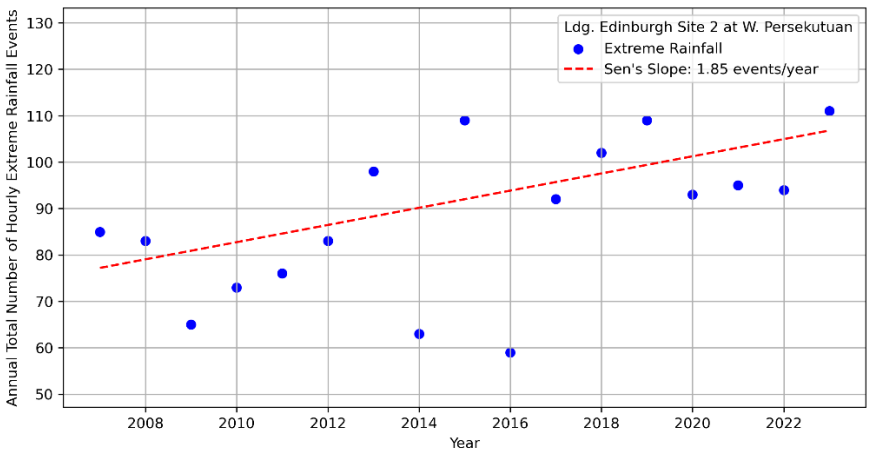
Figure 4.13: Sen's Slope for the Annual Total Number of Hourly Extreme Rainfall Events (events/year) at (d) Km 10 Ulu Kelang At Uk Height Station, (e) Ibu Bekalan Km. 11 At Gombak W. Persekutuan Station and (f) Empangan Genting Klang At W. Persekutuan Station from 2007 to 2023.



(g)



(h)



(i)

Figure 4.13: Sen's Slope for the Annual Total Number of Hourly Extreme Rainfall Events (events/year) at (g) I/pejabat Jps Malaysia At W. Persekutuan Station, (h) Jln. Sg. Udang At Segambut Station and (i) Ldg. Edinburgh Site 2 At W. Persekutuan Station from 2007 to 2023.

Similarly, the MMK test and Sen's slope estimator were applied to detect trends and rates of change in the annual total of hourly extreme rainfall at various stations from 2007 to 2023, as presented in Table 4.12 and Figure 4.14. All stations demonstrated positive trends in annual total hourly extreme rainfall, with increases ranging from 16.87 mm/year at the Empangan Genting Klang At W. Persekutuan Station to 50.48 mm/year at the Km 10 Ulu Kelang At Uk Height Station.

Five stations exhibited statistically significant increasing trends at the 95% confidence level, namely Kolam Takungan Batu (29.27 mm/year), Taman Ehsan At Kepong W. Persekutuan (36.19 mm/year), Km 10 Ulu Kelang At Uk Height (50.48 mm/year), Empangan Genting Klang At W. Persekutuan (16.87 mm/year), and I/pejabat Jps Malaysia At W. Persekutuan (45.10 mm/year).

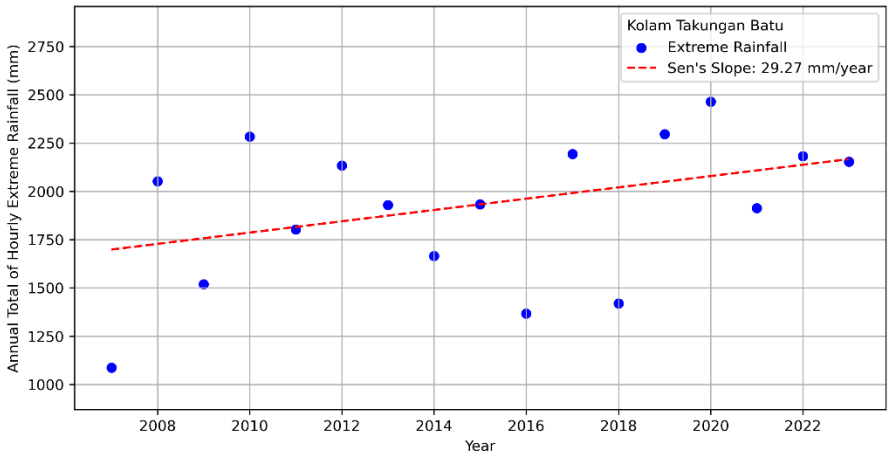
Although the Ldg. Edinburgh Site 2 At W. Persekutuan and Pusat Penyelidekan At Jps Ampang Selangor Stations recorded substantial increasing trends of 46.03 mm/year and 34.71 mm/year, respectively, these trends were not statistically significant, possibly due to high interannual variability.

Overall, the consistent presence of positive trends across all stations suggests a widespread intensification in the annual total of hourly extreme rainfall, primarily driven by more frequent extreme events, albeit with varying degrees of statistical confidence.

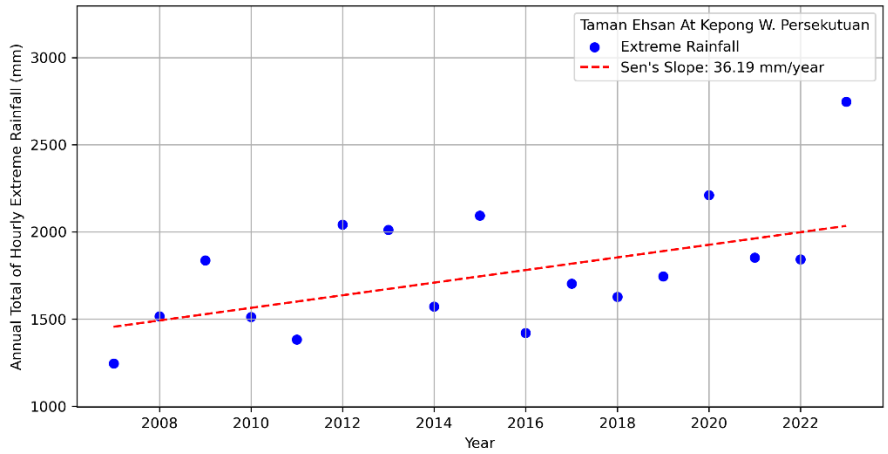
Table 4.12: Trends and Sen's Slope for the Annual Total of Hourly Extreme Rainfall (mm/year) at Selected Stations from 2007 to 2023.

No.	Station	Trend	Sen's Slope (mm/year)
1	**Kolam Takungan Batu (0231391RF)**	Positive*	29.27
2	**Taman Ehsan At Kepong W. Persekutuan (0231441RF)**	Positive*	36.19
3	**Pusat Penyelidekan At Jps Ampang Selangor (0231351RF)**	Positive	34.71
4	**Km 10 Ulu Kelang At Uk Height (0231401RF)**	Positive*	50.48
5	**Ibu Bekalan Km. 11 At Gombak W. Persekutuan (0230721RF)**	Positive	19.52
6	**Empangan Genting Klang At W. Persekutuan (0230631RF)**	Positive*	16.87
7	**I/pejabat Jps Malaysia At W. Persekutuan (0230641RF)**	Positive*	45.1
8	**Jln. Sg. Udang At Segambut (0231381RF)**	Positive	24.36
9	**Ldg. Edinburgh Site 2 At W. Persekutuan (0230651RF)**	Positive	46.03

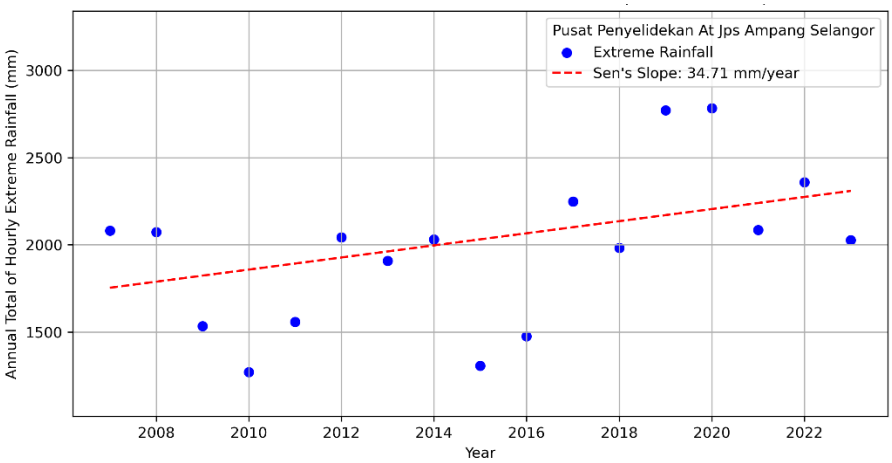
Note: An asterisk (*) indicates statistical significance at 95% confidence level.



(a)

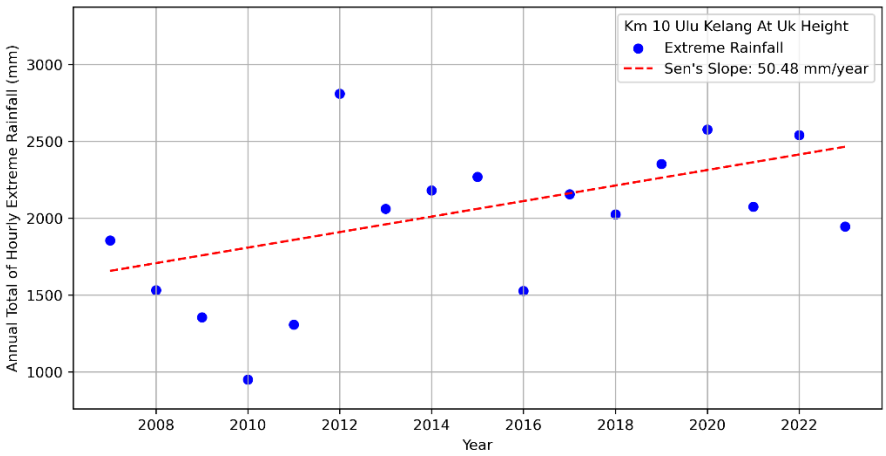


(b)

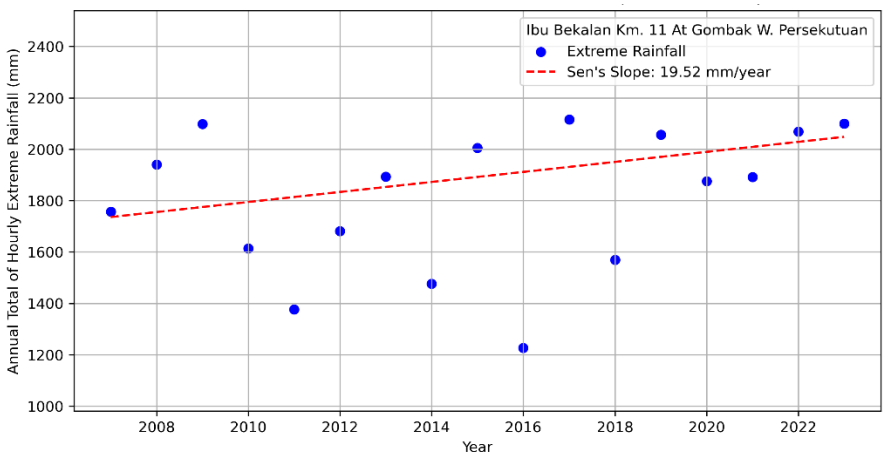


(c)

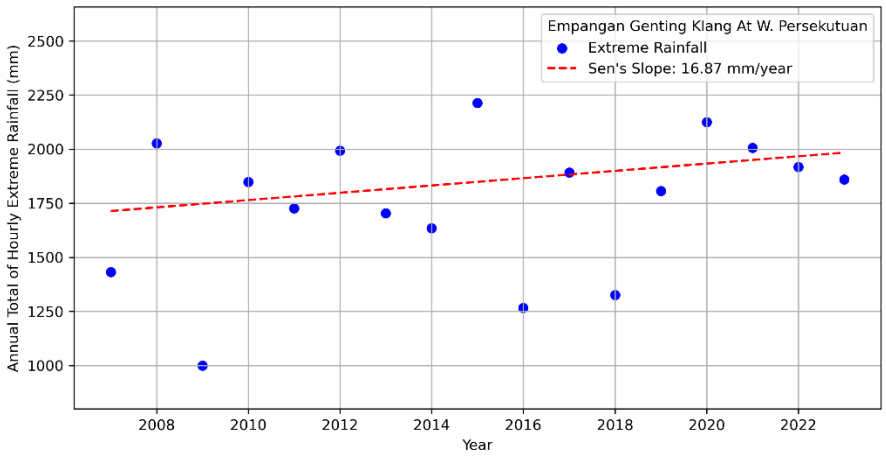
Figure 4.14: Sen's Slope for the Annual Total of Hourly Extreme Rainfall (mm/year) at (a) Kolam Takungan Batu Station, (b) Taman Ehsan At Kepong W. Persekutuan Station and (c) Pusat Penyelidekan At Jps Ampang Selangor Station from 2007 to 2023.



(d)

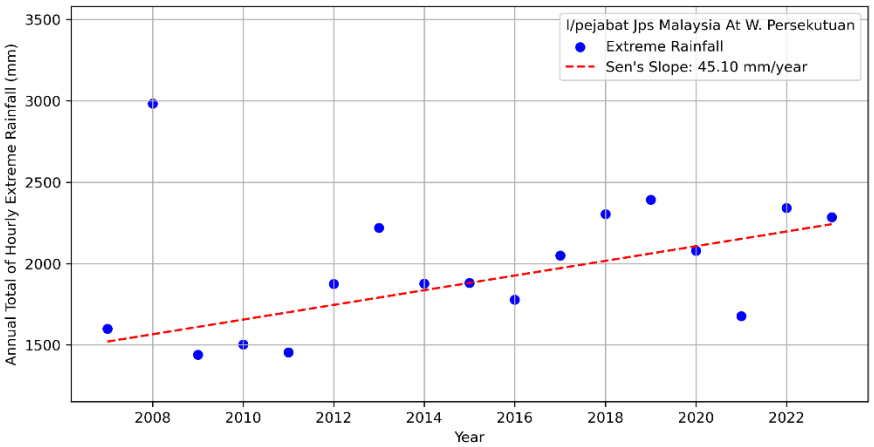


(e)

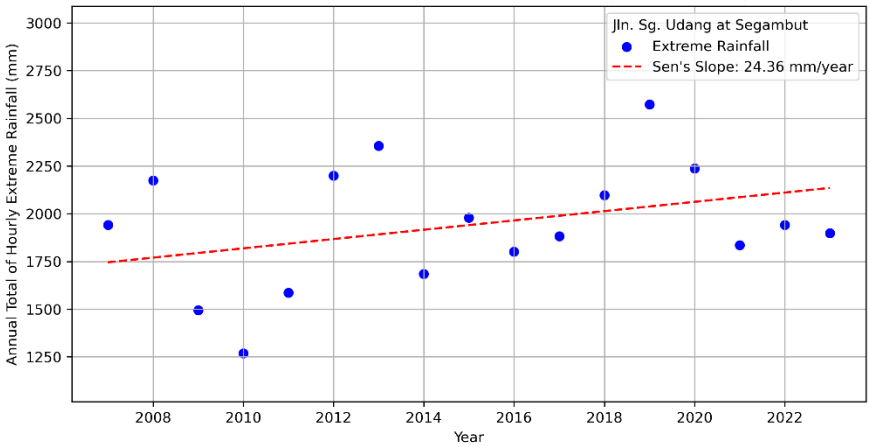


(f)

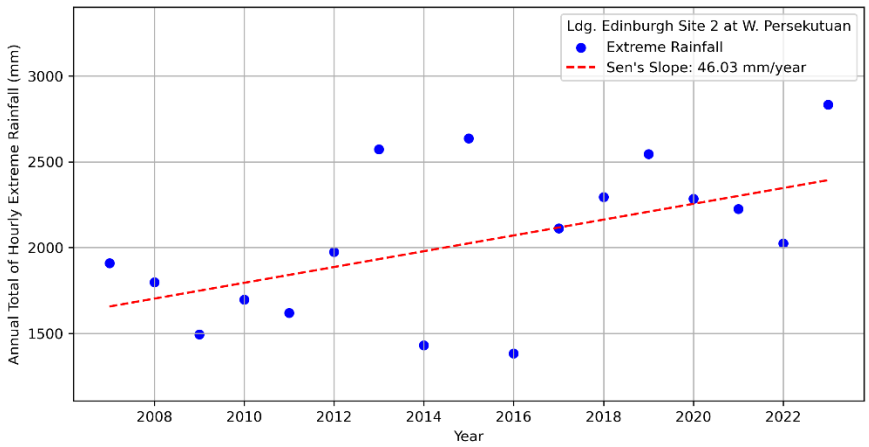
Figure 4.14: Sen's Slope for the Annual Total of Hourly Extreme Rainfall (mm/year) at (d) Km 10 Ulu Kelang At Uk Height Station, (e) Ibu Bekalan Km. 11 At Gombak W. Persekutuan Station and (f) Empangan Genting Klang At W. Persekutuan Station from 2007 to 2023.



(g)



(h)



(i)

Figure 4.14: Sen's Slope for the Annual Total of Hourly Extreme Rainfall (mm/year) at (g) I/pejabat Jps Malaysia At W. Persekutuan Station, (h) Jln. Sg. Udang At Segambut Station and (i) Ldg. Edinburgh Site 2 At W. Persekutuan Station from 2007 to 2023.

Table 4.13 and Figure 4.15 illustrate the trends and Sen's slope for the annual mean of hourly extreme rainfall at selected stations from 2007 to 2023. Similar to the comparative analysis, no statistically significant trends were observed at any of the stations.

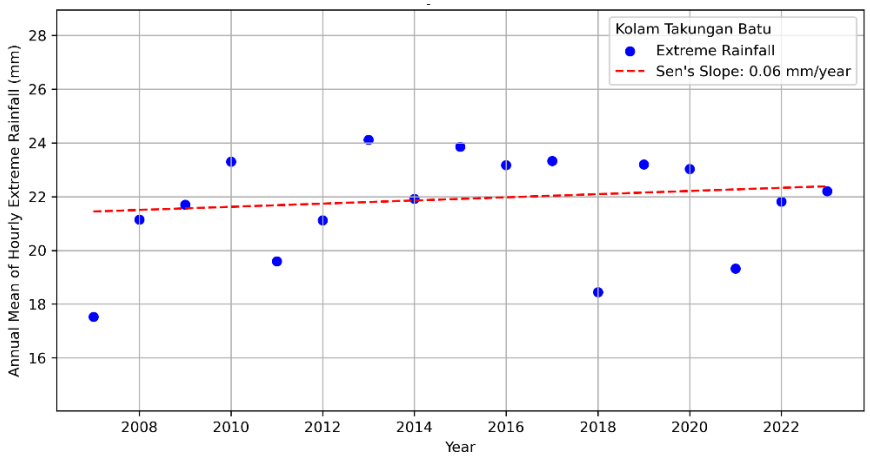
Four stations, Taman Ehsan At Kepong W. Persekutuan, Kolam Takungan Batu, Ldg. Edinburgh Site 2 At W. Persekutuan and Km 10 Ulu Kelang At Uk Height, exhibited extremely weak upward trends, with Sen's slope values ranging from 0.03 mm/year to 0.14 mm/year. Conversely, Jln. Sg. Udang At Segambut, Ibu Bekalan Km. 11 At Gombak W. Persekutuan and Empangan Genting Klang At W. Persekutuan Stations displayed extremely weak downward trends, ranging from -0.04 mm/year to -0.15 mm/year. Meanwhile, both the Pusat Penyelidekan At Jps Ampang Selangor and I/pejabat Jps Malaysia At W. Persekutuan Stations showed no observable trend.

In summary, the annual total number of hourly extreme rainfall events exhibited the most substantial changes across the study areas from 2007 to 2023, with 77.78% of the stations showing statistically significant trends at the 95% confidence level. This was followed by the annual total of hourly extreme rainfall, with 55.56% of stations displaying significant trends. In contrast, none of the stations showed statistically significant trends in the annual mean of hourly extreme rainfall. This finding suggests that the annual mean was relatively insensitive to changes in independent factors such as SUHII during the study period. It further indicates that no significant relationship exists between the annual SUHII and the annual mean of hourly extreme rainfall in the study areas. The relationships between the annual SUHII and the annual total number of hourly extreme rainfall events, as well as between the annual SUHII and the annual total of hourly extreme rainfall, were examined in the next section.

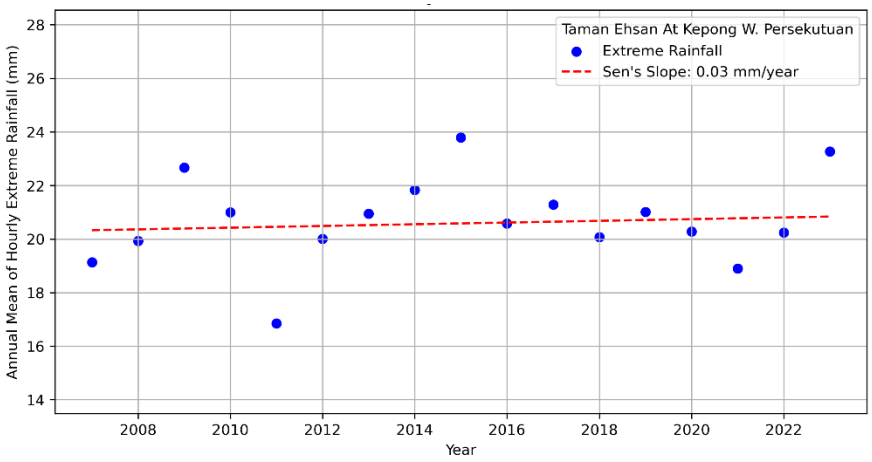
Table 4.13: Trends and Sen's Slope for the Annual Mean of Hourly Extreme Rainfall at Selected Stations from 2007 to 2023.

No.	Station	Trend	Sen's Slope (mm/year)
1	**Kolam Takungan Batu (0231391RF)**	Positive	0.06
2	**Taman Ehsan At Kepong W. Persekutuan (0231441RF)**	Positive	0.03
3	**Pusat Penyelidekan At Jps Ampang Selangor (0231351RF)**	No trend	0
4	**Km 10 Ulu Kelang At Uk Height (0231401RF)**	Positive	0.14
5	**Ibu Bekalan Km. 11 At Gombak W. Persekutuan (0230721RF)**	Negative	-0.11
6	**Empangan Genting Klang At W. Persekutuan (0230631RF)**	Negative	-0.15
7	**I/pejabat Jps Malaysia At W. Persekutuan (0230641RF)**	No trend	0
8	**Jln. Sg. Udang At Segambut (0231381RF)**	Negative	-0.04
9	**Ldg. Edinburgh Site 2 At W. Persekutuan (0230651RF)**	Positive	0.09

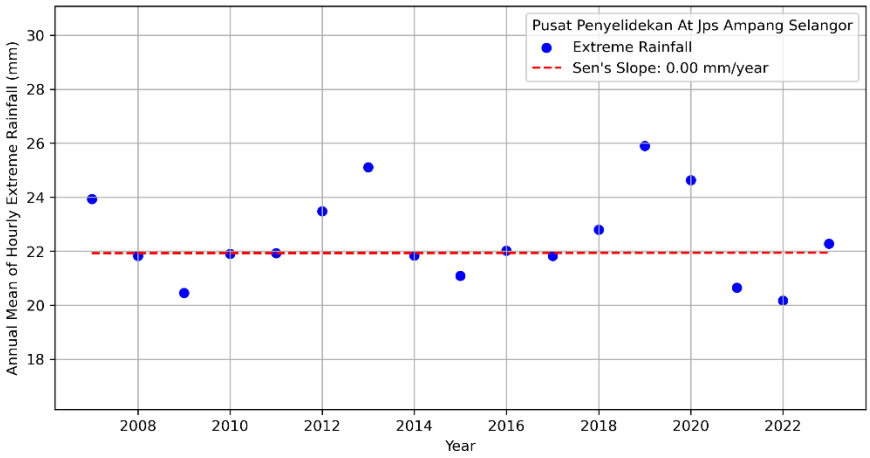
Note: An asterisk (*) indicates statistical significance at 95% confidence level.



(a)

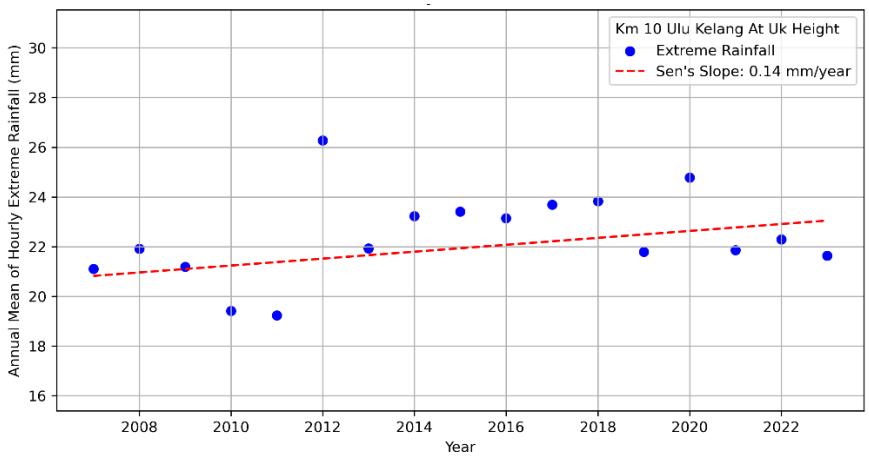


(b)

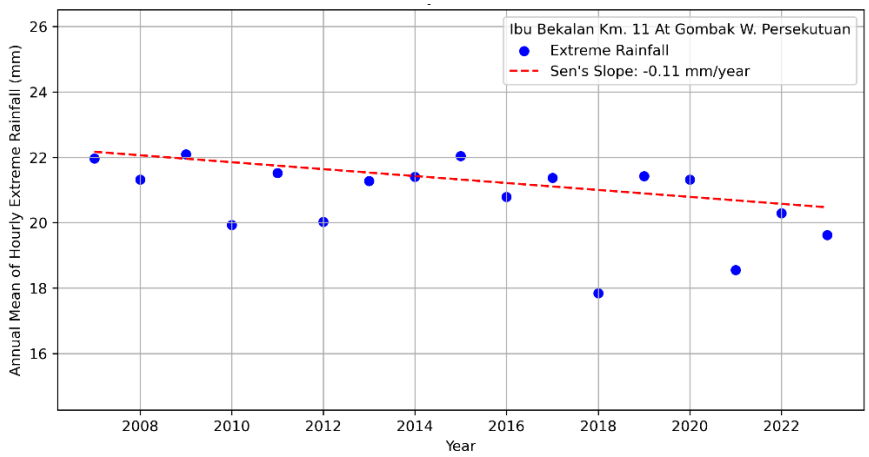


(c)

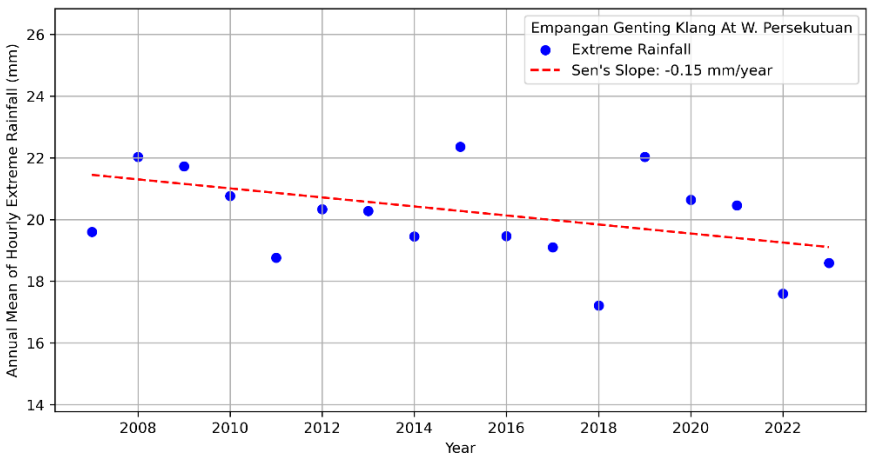
Figure 4.15: Sen's Slope for the Annual Mean of Hourly Extreme Rainfall (mm/year) at (a) Kolam Takungan Batu Station, (b) Taman Ehsan At Kepong W. Persekutuan Station and (c) Pusat Penyelidekan At Jps Ampang Selangor Station from 2007 to 2023.



(d)

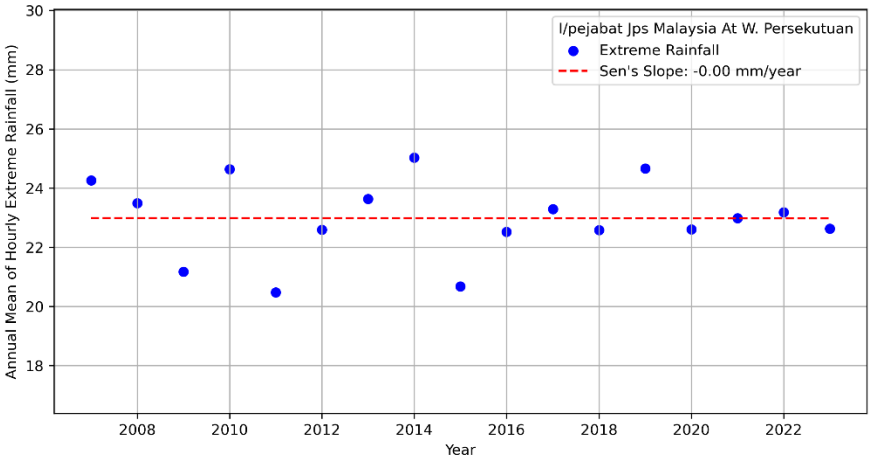


(e)

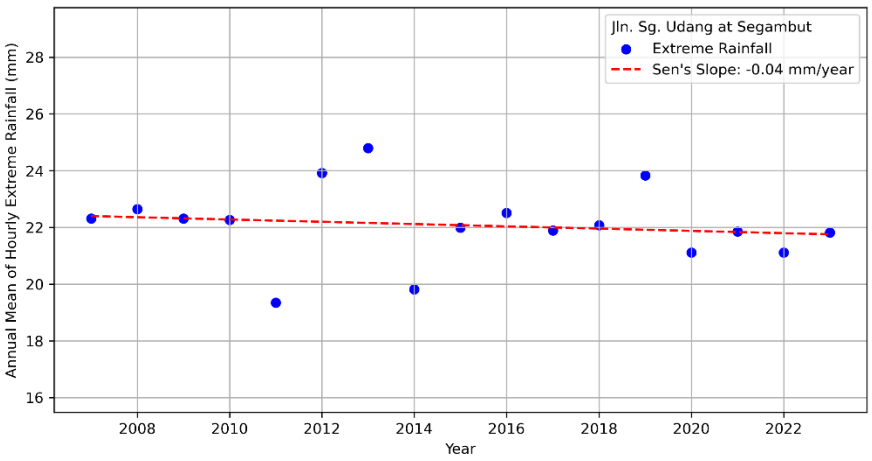


(f)

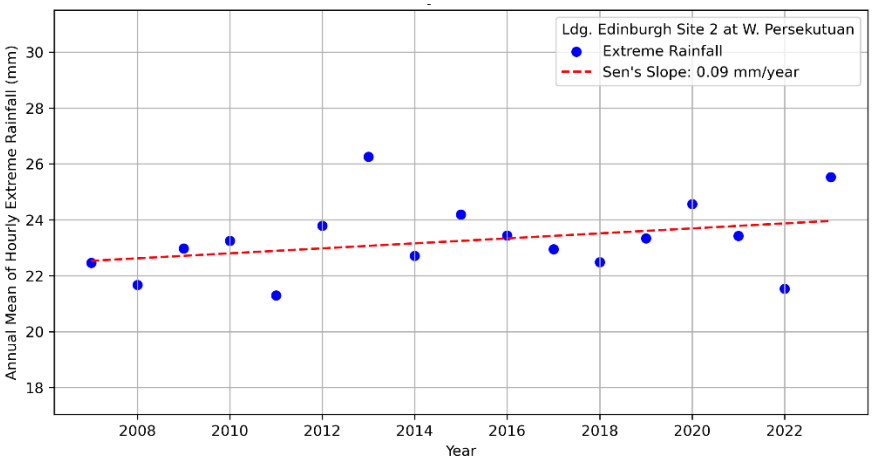
Figure 4.15: Sen's Slope for the Annual Mean of Hourly Extreme Rainfall (mm/year) at (d) Km 10 Ulu Kelang At Uk Height Station, (e) Ibu Bekalan Km. 11 At Gombak W. Persekutuan Station and (f) Empangan Genting Klang At W. Persekutuan Station from 2007 to 2023.



(g)



(h)



(i)

Figure 4.15: Sen's Slope for the Annual Mean of Hourly Extreme Rainfall (mm/year) at (g) I/pejabat Jps Malaysia At W. Persekutuan Station, (h) Jln. Sg. Udang At Segambut Station and (i) Ldg. Edinburgh Site 2 At W. Persekutuan Station from 2007 to 2023.

4.6 Relationships between UHI and Short-Duration Extreme Rainfall

To study the relationships between the annual SUHII and two key rainfall parameters, the annual total of hourly extreme rainfall and the annual total number of hourly extreme rainfall events, both the coefficient of determination (R^2) and Kendall's Tau correlation (τ) were employed. SUHII values from 2007, 2015 and 2023 were interpolated to estimate annual SUHII from 2007 to 2023, while hourly extreme rainfall was grouped by year. Appendix B1–B9 presents the annual SUHII and the annual rainfall for each station.

Table 4.14 and Figure 4.16 present the R^2 results, while Table 4.15 summarises the outcomes of the τ analysis. Generally, all stations demonstrated positive relationships between the annual SUHII and both rainfall parameters, though the strength of these relationships varied considerably.

4.6.1 Coefficient of Determination (R^2)

When analysing the annual total of hourly extreme rainfall as the dependent variable, the Taman Ehsan At Kepong W. Persekutuan and Km 10 Ulu Kelang At Uk Height Stations recorded the strongest statistically significant ($p < 0.05$) relationships, with R^2 values of 0.2629 and 0.2319, respectively. Other stations such as the Pusat Penyelidekan At Jps Ampang Selangor, Ldg. Edinburgh Site 2 At W. Persekutuan, Kolam Takungan Batu and Ibu Bekalan Km. 11 At Gombak W. Persekutuan, demonstrated weaker yet still positive relationships, with R^2 values ranging from 0.1237 to 0.2189. In contrast, the Jln. Sg. Udang At Segambut, Empangan Genting Klang At W. Persekutuan and I/Pejabat Jps Malaysia At W. Persekutuan Stations showed negligible relationships (R^2 = 0.0271 to 0.0870), indicating limited sensitivity of total rainfall volume to SUHII at these locations.

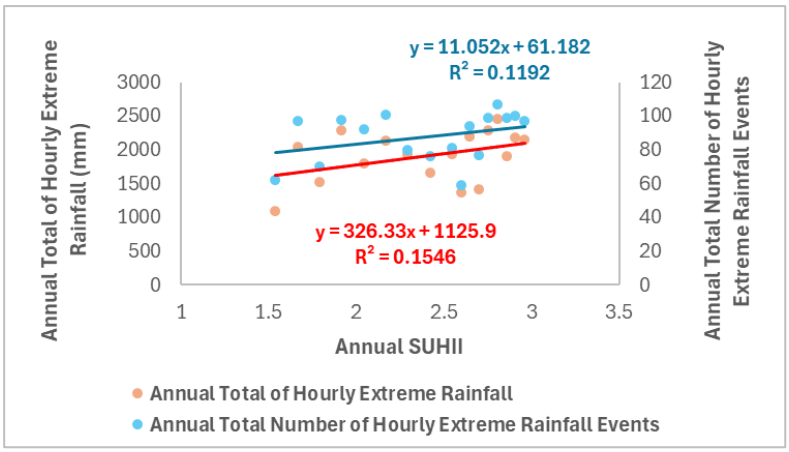
Stronger relationships were observed between the annual SUHII and the annual total number of hourly extreme rainfall events, suggesting that UHI effects may have a more direct influence on the frequency rather than the intensity of extreme rainfall events. Four stations exhibited statistically significant R^2 values at the 95% confidence level, with Ibu Bekalan Km. 11 At Gombak W. Persekutuan Station recording the highest (R^2 = 0.3088), followed by Km 10 Ulu Kelang At Uk Height Station (R^2 = 0.3035), Pusat

Penyelidekan At Jps Ampang Selangor Station (R^2 = 0.2880) and Taman Ehsan At Kepong W. Persekutuan Station (R^2 = 0.2530). The rest of the stations displayed weaker, non-significant relationships, ranging from 0.0407 at the I/Pejabat Jps Malaysia At W. Persekutuan Station to 0.1728 at the Empangan Genting Klang At W. Persekutuan Station.

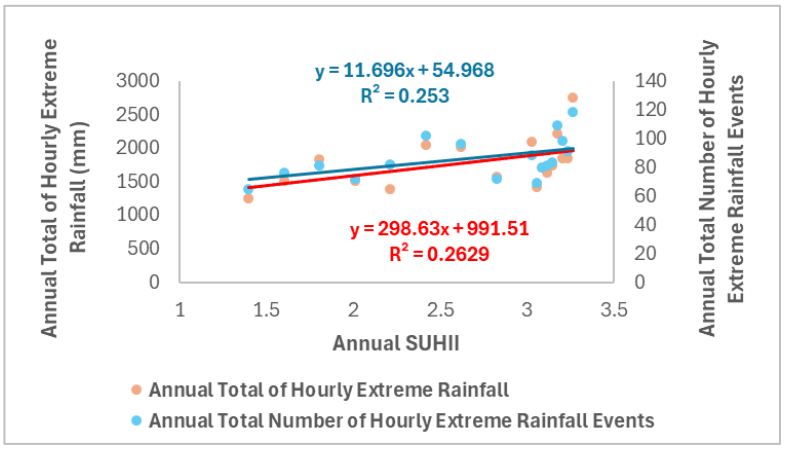
Table 4.14: Coefficient of Determination (R^2) between Annual SUHII and Annual Total of Hourly Extreme Rainfall, and between Annual SUHII and Annual Total Number of Hourly Extreme Rainfall Events at Selected Stations.

No.	Station	R^2 between Annual SUHII and Annual Total of Hourly Extreme Rainfall	R^2 between Annual SUHII and Annual Total Number of Extreme Rainfall Events
1	**Kolam Takungan Batu (0231391RF)**	0.1546	0.1192
2	**Taman Ehsan At Kepong W. Persekutuan (0231441RF)**	0.2629*	0.2530*
3	**Pusat Penyelidekan At Jps Ampang Selangor (0231351RF)**	0.2189	0.2880*
4	**Km 10 Ulu Kelang At Uk Height (0231401RF)**	0.2319*	0.3035*
5	**Ibu Bekalan Km. 11 At Gombak W. Persekutuan (0230721RF)**	0.1237	0.3088*
6	**Empangan Genting Klang At W. Persekutuan (0230631RF)**	0.0568	0.1728
7	**I/pejabat Jps Malaysia At W. Persekutuan (0230641RF)**	0.0271	0.0407
8	**Jln. Sg. Udang At Segambut (0231381RF)**	0.0870	0.1478
9	**Ldg. Edinburgh Site 2 At W. Persekutuan (0230651RF)**	0.1682	0.1381

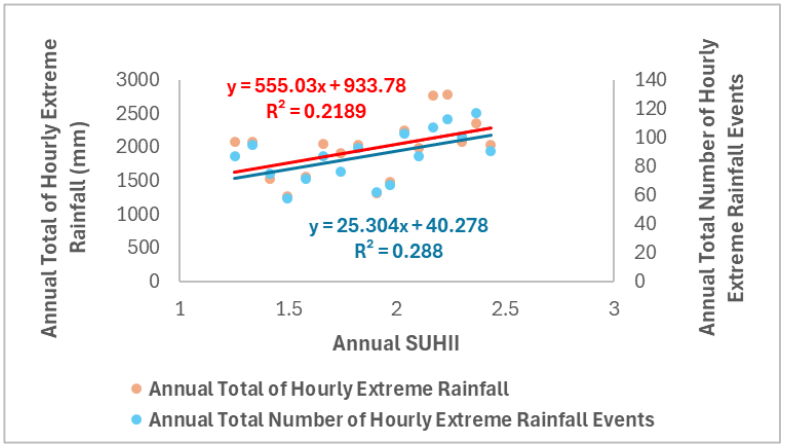
Note: An asterisk (*) indicates statistical significance at 95% confidence level.



(a)

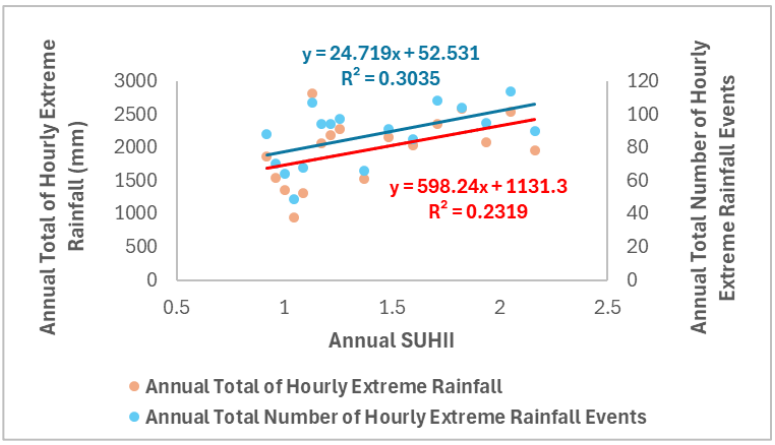


(b)

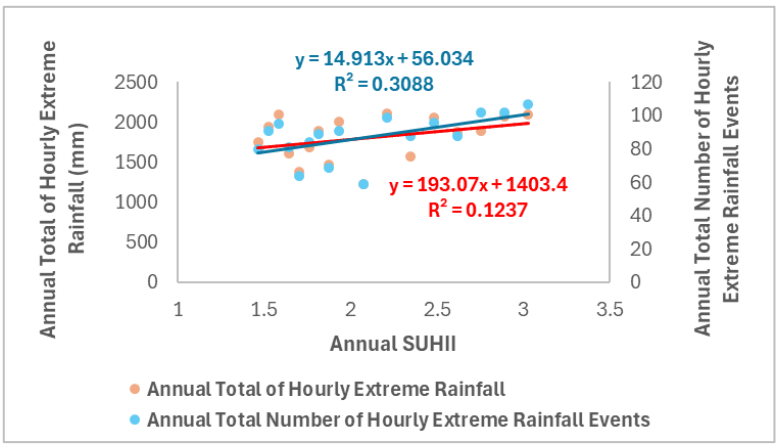


(c)

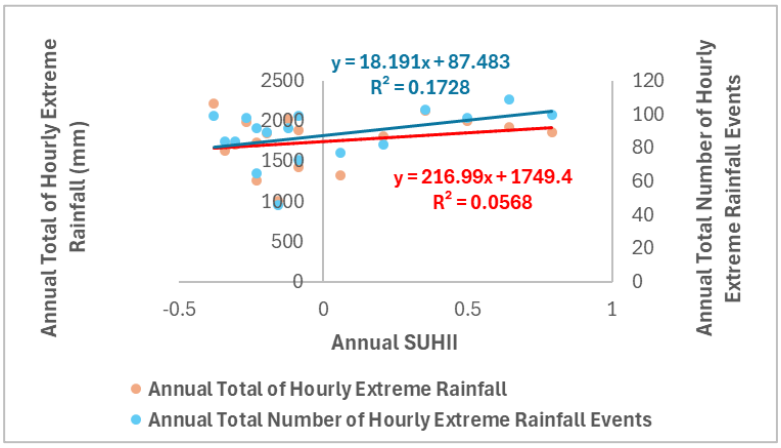
Figure 4.16: Coefficient of Determination (R^2) between Annual SUHII and Annual Total of Hourly Extreme Rainfall, and between Annual SUHII and Annual Total Number of Hourly Extreme Rainfall Events at (a) Kolam Takungan Batu Station, (b) Taman Ehsan At Kepong W. Persekutuan Station and (c) Pusat Penyelidekan At Jps Ampang Selangor Station.



(d)

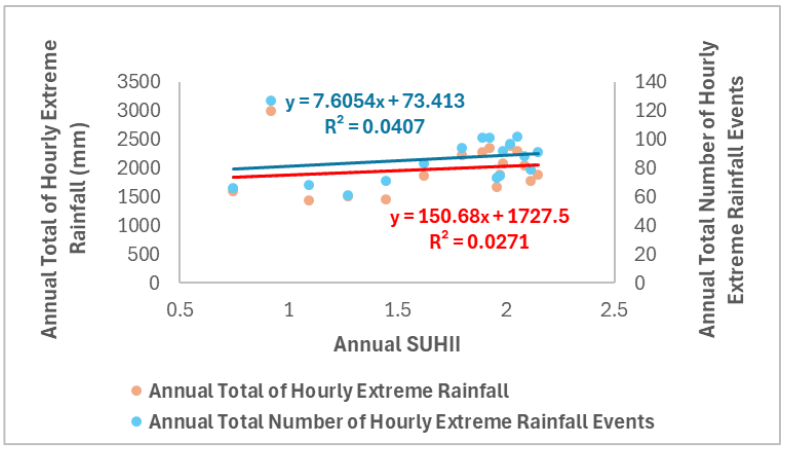


(e)

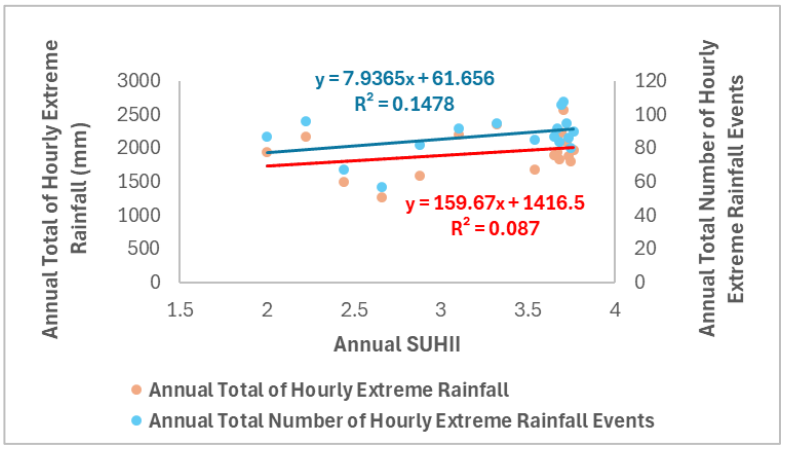


(f)

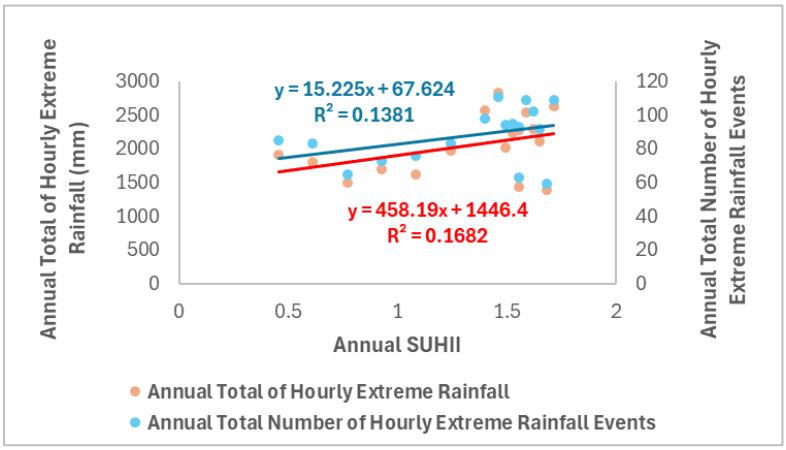
Figure 4.16: Coefficient of Determination (R^2) between Annual SUHII and Annual Total of Hourly Extreme Rainfall, and between Annual SUHII and Annual Total Number of Hourly Extreme Rainfall Events at (d) Km 10 Ulu Kelang At Uk Height Station, (e) Ibu Bekalan Km. 11 At Gombak W. Persekutuan Station and (f) Empangan Genting Klang At W. Persekutuan Station.



(g)



(h)



(i)

Figure 4.16: Coefficient of Determination (R^2) between Annual SUHII and Annual Total of Hourly Extreme Rainfall, and between Annual SUHII and Annual Total Number of Hourly Extreme Rainfall Events at (g) I/pejabat Jps Malaysia At W. Persekutuan Station, (h) Jln. Sg. Udang At Segambut Station and (i) Ldg. Edinburgh Site 2 At W. Persekutuan Station.

4.6.2 Kendall's Tau Correlation (τ)

For the Kendall's Tau correlation, all stations exhibited varying degrees of direct association between the annual SUHII and both the annual total of hourly extreme rainfall and the annual total number of hourly extreme rainfall events. When the dependent variable was the annual total of hourly extreme rainfall, the Taman Ehsan At Kepong W. Persekutuan Station showed the highest statistically significant correlation ($\tau = 0.4412$, $p < 0.05$) at the 95% confidence level. Although statistically insignificant, the remaining stations demonstrated weak to moderate correlations, ranging from 0.1176 at the Empangan Genting Klang At W. Persekutuan Station to 0.3235 at the Km 10 Ulu Kelang At Uk Height Station.

In contrast, stronger correlations were observed between the annual SUHII and the annual total number of hourly extreme rainfall events, with seven stations recording higher τ values. Notably, four of these stations showed statistically significant correlations at the 95% confidence level: Taman Ehsan At Kepong W. Persekutuan Station ($\tau = 0.4593$, $p < 0.05$), Ibu Bekalan Km. 11 At Gombak W. Persekutuan Station ($\tau = 0.4238$, $p < 0.05$), Pusat Penyelidekan At Jps Ampang Selangor Station ($\tau = 0.3941$, $p < 0.05$) and Km 10 Ulu Kelang At Uk Height Station ($\tau = 0.3616$, $p < 0.05$). Meanwhile, Jln. Sg. Udang At Segambut and Ldg. Edinburgh Site 2 At W. Persekutuan Stations exhibited lower correlations of 0.1413 and 0.2222, respectively, compared to those with the annual total of hourly extreme rainfall. The remaining stations displayed mild correlations between annual SUHII and the annual total number of extreme rainfall events, ranging from 0.2388 at the Empangan Genting Klang At W. Persekutuan Station to 0.3111 at the Kolam Takungan Batu Station.

The analysis using both the coefficient of determination and Kendall's Tau correlation revealed generally positive relationships between the annual SUHII and both the annual total of hourly extreme rainfall and the annual total number of hourly extreme rainfall events. Stronger and more statistically significant correlations were observed for rainfall frequency than rainfall volume, suggesting that UHI may have a greater influence on the frequency rather than the intensity of extreme rainfall events. Importantly, stations like Taman Ehsan At Kepong W. Persekutuan, Pusat Penyelidekan At

Jps Ampang Selangor, Km 10 Ulu Kelang At Uk Height and Ibu Bekalan Km. 11 At Gombak W. Persekutuan consistently showed stronger associations across both statistical methods.

Table 4.15: Kendall's Tau Correlation (τ) between Annual SUHII and Annual Total of Hourly Extreme Rainfall, and between Annual SUHII and Annual Total Number of Hourly Extreme Rainfall Events at Selected Stations.

No.	Station	Annual SUHII and Annual Total of Hourly Extreme Rainfall	Annual SUHII and Annual Total Number of Hourly Extreme Rainfall Events
1	**Kolam Takungan Batu (0231391RF)**	0.2647	0.3111
2	**Taman Ehsan At Kepong W. Persekutuan (0231441RF)**	0.4412*	0.4593*
3	**Pusat Penyelidekan At Jps Ampang Selangor (0231351RF)**	0.2794	0.3941*
4	**Km 10 Ulu Kelang At Uk Height (0231401RF)**	0.3235	0.3616*
5	**Ibu Bekalan Km. 11 At Gombak W. Persekutuan (0230721RF)**	0.2206	0.4238*
6	**Empangan Genting Klang At W. Persekutuan (0230631RF)**	0.1176	0.2388
7	**I/pejabat Jps Malaysia At W. Persekutuan (0230641RF)**	0.2206	0.2583
8	**Jln. Sg. Udang At Segambut (0231381RF)**	0.1765	0.1413
9	**Ldg. Edinburgh Site 2 At W. Persekutuan (0230651RF)**	0.2500	0.2222

Note: An asterisk (*) indicates statistical significance at 95% confidence level.

4.7 Summary

Urbanisation and LULC changes played a significant role in shaping rainfall patterns. Rapid development in the northern region of Kuala Lumpur intensified the UHI effect, which in turn enhanced localised convection and led to an increase in the total number of hourly extreme rainfall events. The UHI effect, driven by expanding urbanisation, altered the local climate and contributed to more frequent intense rainfall events (Li et al., 2020; Whitford et al., 2023; Siswato, Schrier, & Hurk, 2022; Jiang, Zhang, & Luo, 2023).

Additionally, the spatial variability in rainfall patterns supported the presence of localised influences. Some stations recorded greater increases in the total number of hourly extreme rainfall events than others, suggesting that beyond SUHII, factors such as local topography, wind patterns and LULC changes were also influential (Huang et al., 2025; Mwanthi et al., 2024; Pen et al., 2024). These variations highlight the complexity of rainfall dynamics, where different areas respond uniquely to broader climatic and environmental changes. Notably, stations such as Taman Ehsan At Kepong W. Persekutuan, Pusat Penyelidekan At Jps Ampang Selangor, Km 10 Ulu Kelang At Uk Height and Ibu Bekalan Km. 11 At Gombak W. Persekutuan demonstrated greater sensitivity to SUHII variation over the study period.

CHAPTER 5

CONCLUSIONS AND RECOMMENDATIONS

5.1 Conclusions

This study demonstrates that the UHI effect had a measurable influence on short-duration extreme rainfall in the northern region of Kuala Lumpur from 2007 to 2023. This conclusion is supported by the successful completion of the four study objectives.

Firstly, the spatiotemporal variations in LULC were effectively analysed using Landsat imagery and classified through SVM and RF algorithms. The results revealed substantial urban expansion between 2007 and 2015, characterised by an increase in built-up areas and a corresponding decline in vegetation cover and water bodies. Although a slight reduction in built-up areas was observed in 2023, the overall trend still indicated a net increase in urban development throughout the study period.

Secondly, the SUHII was estimated through the quantification of LST derived from thermal bands of Landsat imagery. The findings showed a rising trend in SUHII, particularly in the northern parts of Kuala Lumpur, reflecting the impact of dense urbanisation and increasing surface heating over time.

Thirdly, the spatiotemporal variation, diurnal distribution and trend of hourly extreme rainfall were examined, revealing an increase in the number of hourly extreme rainfall events, especially during late afternoon and early evening hours. Trend analysis further confirmed the presence of increasing short-duration extreme rainfall over the years.

Finally, statistical analyses using the coefficient of determination (R^2) and Kendall's Tau correlation (τ) investigated the relationship between the SUHII and hourly extreme rainfall. Four out of nine stations exhibited statistically significant moderate relationships between the annual SUHII and the annual total number of hourly extreme rainfall events ($R^2 = 0.2530 - 0.3088$; $\tau = 0.3616 - 0.4593$; $p < 0.05$). These results suggest that intensified urban heating may contribute to the enhancement of localised convective rainfall in Kuala Lumpur.

In conclusion, the integration of remote sensing, machine learning classification, statistical analysis and geospatial techniques provides meaningful insight into how the urbanisation-driven heat intensification can influence the occurrence and distribution of extreme rainfall events. These findings emphasise the importance of incorporating UHI considerations into urban climate resilience planning and flood management strategies.

5.2 Recommendations

Building on the findings of this research, the following recommendations are put forth to promote sustainable urban development and guide future research in urban climate dynamics:

1. Urban planners and policymakers should prioritise the implementation of green infrastructure, such as green roofs, urban parks and tree-lined streets, especially in areas exhibiting high SUHII. These measures can help reduce surface temperatures and potentially mitigate localised extreme rainfall events driven by urban heat.
2. Given the observed intensification of short-duration extreme rainfall, local authorities are encouraged to enhance drainage and flood mitigation infrastructure in rapidly urbanising zones to accommodate increased surface runoff and minimise flood risk.
3. To overcome the limitations caused by cloud cover in satellite imagery, future research should utilise data from satellites with higher temporal resolution, such as MODIS, or apply cloud-masking and gap-filling techniques to increase the availability and continuity of LULC, LST, and SUHII analyses across more time points.
4. Considering the sensitivity of machine learning models to data quality and parameter selection, future studies should explore ensemble methods or deep learning approaches that are more resilient to noise and variability. Integrating multi-source data, such as LiDAR or high-resolution aerial imagery, may further improve the reliability of LULC classification.
5. To address the limitations posed by sparse rainfall stations and the uncertainties associated with the IDW interpolation method, future

research should incorporate radar-based and satellite-derived rainfall products, such as TRMM. Validation using ground-based observations is recommended to assess and improve interpolation accuracy.

6. While this study focuses on SUHII and rainfall parameters, future investigations should include additional atmospheric variables, such as humidity, wind speed, atmospheric pressure and vertical motion, to develop a more comprehensive understanding of the physical mechanisms influencing rainfall under UHI conditions.

REFERENCES

Agrawal, Y., Pandey, H. and Tiwari, P.S., 2023. Analytical study of relation between Land surface temperature and Land Use/Land Cover using spectral indices: A case study of Chandigarh. *Journal of Geomatics*, [e-journal] 17(2). pp. 67-80. https://doi.org/10.58825/jog.2023.17.2.65.

Ahmed, I.A., Salam, R., Naikoo, M.W., Rahman, A., Praveen, B., Hoai, P.N., Pham, Q.B., Anh, D.T., Tri, D.Q. and Elkhrachy, I., 2022. Evaluating the variability in long-term rainfall over India with advanced statistical techniques. *Hydrology*, [e-journal] 70(1). pp. 801-818. https://doi.org/10.1007/s11600-022-00735-5.

Ahorloo, S., Jamarani, A., Kashfi, M., Kashani, M.H. and Najafizadeh, A., 2024. A systematic review of machine learning methods in software testing. *Applied Soft Computing*, [e-journal] 162(1). https://doi.org/10.1016/j.asoc.2024.111805.

Al Kafy, A., Al-Faisal, A., Rahman, M.S., Islam, M., Al Rakib, A., Islam, M.A., Khan, M.H.H., Skidar, M.S., Sarker, M.H.S., Mawa, J. and Sattar, G.S., 2021. Prediction of seasonal urban thermal field variance index using machine learning algorithms in Cumilla, Bangladesh. *Sustainable Cities and Society*, [e-journal] 64(1). https://doi.org/10.1016/j.scs.2020.102542.

Al-Taei, A.I., Alesheikh, A.A. and Boloorani, A.D., 2023. Land Use/Land Cover Change Analysis Using Multi-Temporal Remote Sensing Data: A Case Study of Tigris and Euphrates Rivers Basin. *Land*, [e-journal] 12(5). https://doi.org/10.3390/land12051101.

Athukorala, D. and Murayama, Y., 2021. Urban Heat Island Formation in Greater Cairo: Spatio-Temporal Analysis of Daytime and Nighttime Land Surface Temperatures along the Urban–Rural Gradient. *Remote Sensing*, [e-journal] 13(1). https://doi.org/10.3390/rs13071396.

Aucahuasi-Almidon, A., Cabrera-Carranza, C. and Garate-Quispe, J., 2024. Trend analysis and change-point detection of temperature and rainfall in southern Peruvian Amazon and its relation to deforestation. *Journal of Agrometeorology*, [e-journal] 26(4). pp. 425-430. https://doi.org/10.54386/jam.v26i4.2687.

Awad, M. and Khanna, R., 2015. *Efficient Learning Machines*. Berkeley, CA: Apress.

Benmoshe, N., 2025. A Simple Solution for the Inverse Distance Weighting Interpolation (IDW) Clustering Problem. *Sci*, [e-journal] 7(1). https://doi.org/10.3390/sci7010030.

Berman, J.J., 2016. *Data Simplification.* Burlington : Morgan Kaufmann.

Cheng, Y., Sang, Y., Wang, Z., Guo, Y. and Tang Y., 2021. Effects of Rainfall and Underlying Surface on Flood Recession—The Upper Huaihe River Basin Case. *International Journal of Disaster Risk Science*, [e-journal] 12(1). pp. 111-120. https://doi.org/10.1007/s13753-020-00310-w.

Cristóbal, J., Jiménez-Muñoz, J.C., Prakash, A., Mattar, C., Skoković, D. and Sobrino, J.A., 2018. An Improved Single-Channel Method to Retrieve Land Surface Temperature from the Landsat-8 Thermal Band. *Remote Sensing*, [e-journal] 10(3). https://doi.org/10.3390/rs10030431.

Croux, C. and Dehon, C., 2010. Influence functions of the Spearman and Kendall correlation measures. *Statistical Methods & Application*, [e-journal] 19(1). pp. 497-515. https://doi.org/10.1007/s10260-010-0142-z.

Dammayatri, R., Susantoro, T.M. and Wikantika, K., 2023. Green open space and barren land mapping for flood mitigation in Jakarta, the capital of Indonesia. Indonesian Journal of Geography, [e-journal] 55(2). pp.197-205. https://doi.org/10.22146/ijg.76452.

Darabi, H., Mehr, A.D., Kum, G., Sönmez, M.E., Dumitrache, C.A., Diani, K., Celebi, A. and Haghighi, T., 2023. Hydroclimatic Trends and Drought Risk Assessment in the Ceyhan River Basin: Insights from SPI and STI Indices. *Hydrology*, [e-journal] 10(8). https://doi.org/10.3390/hydrology10080157.

Dash, P., Sanders, S.L., Parajuli, P. and Ouyang, Y., 2023. Improving the Accuracy of Land Use and Land Cover Classification of Landsat Data in an Agricultural Watershed. *Remote Sensing*, [e-journal] 15(1). https://doi.org/10.3390/rs15164020.

Deopa, R., Thakur, D.A., Kumar, S., Mohanty, M.P. and Asha, P., 2024. Discerning the dynamics of urbanization-climate change-flood risk nexus in densely populated urban mega cities: An appraisal of efficient flood management through spatiotemporal and geostatistical rainfall analysis and hydrodynamic modelling. *Science of The Total Environment*, [e-journal] 952(1). https://doi.org/10.1016/j.scitotenv.2024.175882.

Doiphode, S. and Swami, V., 2024. Rainfall analysis and deriving design storm for hydrological modelling of Koyna dam catchment area in Maharashtra state of India. *Journal of Environmental Management*, [e-journal] 359(1). https://doi.org/10.1016/j.jenvman.2024.121082.

Doost, Z.H. and Yaseen, Z.M., 2023. The impact of land use and land cover on groundwater fluctuations using remote sensing and geographical information system: Representative case study in Afghanistan. *Environment, Development and Sustainability,* [e-journal]. https://doi.org/10.1007/s10668-023-04253-2.

Du, C., Ren, H., Qin, Q., Meng, J. and Zhao, S., 2015. A Practical Split-Window Algorithm for Estimating Land Surface Temperature from Landsat 8 Data. *Remote Sensing,* [e-journal] 7(1), pp. 647-665. https:// doi:10.3390/rs70100647.

Edan, M.H., Maarouf, R.M. and Hasson, J., 2021. Predicting the impacts of land use/land cover change on land surface temperature using remote sensing approach in Al Kut, Iraq. *Physics and Chemistry of the Earth,* [e-journal] 123(1). https://doi.org/10.1016/j.pce.2021.103012.

Fawagreh, K., Gaber, M.M. and Elyan, E., 2014. Random forests: from early developments to recent advancements. *Systems Science & Control Engineering,* [e-journal] 2(1). http://dx.doi.org/10.1080/21642583.2014.956265.

Fetene, A., 2025. Remote sensing analysis of urban heat island dynamics in Bahir Dar and Hawassa, Ethiopia: The role of vegetation, urbanization, and climate. *Environmental Challenges*, [e-journal] 19(1). https://doi.org/10.1016/j.envc.2025.101139.

Foody, G.M., 2020. Explaining the unsuitability of the kappa coefficient in the assessment and comparison of the accuracy of thematic maps obtained by image classification. *Remote Sensing of Environment,* [e-journal] 239(1). https://doi.org/10.1016/j.rse.2019.111630.

Fung, K.F., Chew, K.S., Huang, Y.F., Ahmed, A.N., Teo, F.Y., Ng, J.L. and Elshafie, A., 2022. Evaluation of spatial interpolation methods and spatiotemporal modeling of rainfall distribution in Peninsular Malaysia. *Ain Shams Engineering Journal*, [e-journal] 13(1). https://doi.org/10.1016/j.asej.2021.09.001.

García, D.H. and Díaz, J.A., 2021. Spatial and Multi-Temporal Analysis of Land Surface Temperature through Landsat 8 Images: Comparison of Algorithms in a Highly Polluted City (Granada). *Remote Sensing*, [e-journal] 13(5). https://doi.org/10.3390/rs13051012.

Gascon, M., Cirach, M., Martínez, D., Dadvand, P., Valentín, A., Plasència, A. and Nieuwenhuijsen,M.J., 2016. Normalized difference vegetation index (NDVI) as a marker of surrounding greenness in epidemiological studies: The case of Barcelona city. *Urban Forestry & Urban Greening*, [e-journal] 19(1). pp.88-94. https://doi.org/10.1016/j.ufug.2016.07.001.

Gyimah, R.R., 2023. The hot zones are cities: Methodological outcomes and synthesis of surface urban heat island effect in Africa. *Cogent Social Sciences*, [e-journal] 9(1). https://doi.org/10.1080/23311886.2023.2165651.

Hamed, K.H. and Ramachandra Rao, A., 1998. A modified Mann-Kendall trend test for autocorrelated data. *Journal of Hydrology*, [e-journal] 204(1-4). pp. 182-196. https://doi.org/10.1016/S0022-1694(97)00125-X.

Huang, W.R., Bui-Manh, H., Chiang, T.Y. and Koralegedara, S.B., 2025. Seasonal variation of diurnal rainfall characteristics over Vietnam. *Science of the Total Environment*, [e-journal] 971(1). https://doi.org/10.1016/j.scitotenv.2025.179058.

Huang, Y.F., Kasniza Jumari, N.A.S, Ahmed, A.N., Ng, J.L., Koo, C.H., Chong, K.L., Sherif, M. and Elshafie, A., 2023. Analysis of urban heat islands with landsat satellite images and GIS in Kuala Lumpur Metropolitan City. *Heliyon*, [e-journal] 9(8). https://doi.org/10.1016/j.heliyon.2023.e18424.

Hussien, K., Kebede, A., Mekuriaw, A., Beza, S.A. and Erena, S.H., 2023. Spatiotemporal trends of NDVI and its response to climate variability in the Abbay River Basin, Ethiopia. *Heliyon*, [e-journal] 9(3). https://doi.org/10.1016/j.heliyon.2023.e14113.

Jan, H. and Tomasz, K., 2011. Comparison of Values of Pearson's and Spearman's Correlation Coefficients on the Same Sets of Data. *Quaestiones Geographicae*, [e-journal] 30(2). pp. 87-93. https://doi.org/10.2478/v10117-011-0021-1.

Jedox, 2025. Error Metrics: How to Evaluate Your Forecasts [Online]. Available at: < https://www.jedox.com/en/blog/error-metrics-how-to-evaluate-forecasts/>. [Accessed on 28 February 2025].

Jia, S., Wang, Y., Liang, T.C., Weng, Q., Too, C., Chen, W. and Ding, X., 2024. Multiscale estimation of the cooling effect of urban greenspace in subtropical and tropical cities. *Urban Forestry & Urban Greening*, [e-journal] 98(1). https://doi.org/10.1016/j.ufug.2024.128390.

Jiang, X., Zhang, D.L. and Luo, Y., 2023. Influences of Urbanization on an Afternoon Heavy Rainfall Event over the Yangtze River Delta Region. *Monthly Weather Review*, [e-journal] 151(3). pp. 815-832. https://doi.org/10.1175/MWR-D-22-0165.1.

Kenabatho, P.K., 2025. Innovative trend analysis of long-term spatial-temporal rainfall patterns over Botswana: Implications for water resources management. *Journal of Hydrology: Regional Studies*, [e-journal] 58(1). https://doi.org/10.1016/j.ejrh.2025.102217.

Kendall, M.G., 1938. A New Measure of Rank Correlation. *Biometrika*, [e-journal] 30(1/2). pp. 81-93. https://doi.org/10.2307/2332226.

Lee, M., An, H., Lee, J., Um, M.J., Jun, Y., Jung, K., Kim, S. and Park, D., 2023. Spatiotemporal Variability of Regional Rainfall Frequencies in South Korea for Different Periods. *Sustianability*, [e-journal] 15(24). https://doi.org/10.3390/su152416646.

Li Y., Fowler, H.J., Argüeso, D., Blenkinsop, S., Evans, J.P., Lenderink, G., Yan, X., Guerreiro, S.B., Lewis, E. and Li, X.F., 2020. Strong Intensification of Hourly Rainfall Extremes by Urbanization. *Geophysical Research Letters*, [e-journal] 47(1). https://doi.org/10.1029/2020GL088758

Li, L., Zha Y. and Wang, R., 2020. Relationship of surface urban heat island with air temperature and precipitation in global large cities. *Ecological Indicators*, [e-journal] 117(1). https://doi.org/10.1016/j.ecolind.2020.106683.

Li, X.X., Koh, T.Y., Panda, J. and Norford, L.K., 2016. Impact of urbanization patterns on the local climate of a tropical city, Singapore: An ensemble study. Journal of Geophysical Research: *Atmospheres*, [e-journal] 121(9). pp. 4386-4403. https://doi.org/10.1002/2015JD024452.

Liu, Y., An, Z. and Ming, Y., 2024. Simulating influences of land use/land cover composition and configuration on urban heat island using machine learning. *Sustainable Cities and Society*, [e-journal] 108(1). https://doi.org/10.1016/j.scs.2024.105482.

Lu, D. and Weng, Q., 2007. A survey of image classification methods and techniques for improving classification performance. *International Journal of Remote Sensing*, [e-journal] 28(5), pp. 823-870. https://doi.org/10.1080/01431160600746456.

Madariya, G., Pandey, S.K. and Sharma, S.K., 2022. A comparative study on supervised and unsupervised techniques of land use and land cover classification. *The Pharma Innovation Journal*, [e-journal] 11(3), pp. 517-521. https://doi.org/10.1080/01431160600746456.

Mahata, B., Sahu, S.S., Sardar, A., Laxmikanta, R. and Maity, M., 2024. Spatiotemporal dynamics of land use/land cover (LULC) changes and its impact on land surface temperature: A case study in New Town Kolkata, eastern India. *Regional Sustainability*, [e-journal] 5(2). https://doi.org/10.1016/j.regsus.2024.100138.

Maithani, S., Nautiyal, G., Sharma, A. and Sharma, S.K., 2022. Simulation of Land Surface Temperature Patterns Over Future Urban Areas—A Machine Learning Approach. *Journal of the Indian Society of Remote Sensing*, [e-journal] 50(11), pp. 2145-2162. https://doi.org/10.1007/s12524-022-01590-z.

Mamun, A.A., Salleh, M.N. and Noor, H.M., 2018. Estimation of short-duration rainfall intensity from daily rainfall values in Klang Valley, *Malaysia. Applied Water Science*, [e-journal] 8(203). https://doi.org/10.1007/s13201-018-0854-z

Mandal, B., 2024. Leveraging machine learning for analyzing the nexus between land use and land cover change, land surface temperature and biophysical indices in an eco-sensitive region of Brahmani-Dwarka interfluve. *Results in Engineering*, [e-journal] 24(1). https://doi.org/10.1016/j.rineng.2024.102854.

McFeeters, S.K., 1996. The use of the Normalized Difference Water Index (NDWI) in the delineation of open water features. International Journal of Remote Sensing, [e-journal] 17(7). pp.1425-1432. https://doi.org/10.1080/01431169608948714.

Mekuria, E.T., Demissie, T.A. and Feyessa, F.F., 2025. Investigation of the spatial and temporal long-term hydro-climatic trends in Upper Omo Gibe Basin, Ethiopia. *Heliyon*, [e-journal] 11(3). https://doi.org/10.1016/j.heliyon.2025.e42265.

Miniandi, N.D., Muhammad, M.K.I., Jamal, M.H. and Shahid, S., 2024. Urbanization signature on hourly rainfall extremes of Kuala Lumpur. Sustainable Cities and Society, [e-journal] 112(1). https://doi.org/10.1016/j.scs.2024.105610.

Mirabi, E. and Davies, P.J., 2022. A systematic review investigating linear infrastructure effects on Urban Heat Island (UHI_{ULI}) and its interaction with UHI typologies. Urban Climate, [e-journal] 45(1). https://doi.org/10.1016/j.uclim.2022.101261

Mizutori, M. and Guha-Sapir, D., 2020. *Human cost of disasters: an overview of the last 20 years (2000-2019)*. [Online] Available at: < https://www.undrr.org/publication/human-cost-disasters-overview-last-20-years-2000-2019> [Accessed on 16 September 2024].

Moazzam, M.F.U., Kim, S. and Lee, B.G., (2024). Cities in the heat: Unveiling the urbanized impacted surface urban heat island of South Korea's metropolises. *Remote Sensing Applications: Society and Environment*, [e-journal] 36(1). https://doi.org/10.1016/j.rsase.2024.101271.

Mohammad, P., Goswami, A., Chauhan, S. and Nayak, S., 2022. Machine learning algorithm based prediction of land use land cover and land surface temperature changes to characterize the surface urban heat island phenomena over Ahmedabad city, India. *Urban Climate*, [e-journal] 42(1). https://doi.org/10.1016/j.uclim.2022.101116.

Mohd Hasmadi, I., Pakhriazad, H.Z. and Shahrin, M.F., 2009. Evaluating supervised and unsupervised techniques for land cover mapping using remote sensing data. *Geografia,* [e-journal] 5(1), pp. 1-10.

Motegaonkar, S. and Kashid, S., 2024. Remote sensing analysis of urban heat island dynamics in Bahir Dar and Hawassa, Ethiopia: The role of vegetation, urbanization, and climate. *Journal of Environmental Research, Engineering and Management,* [e-journal] 80(3). pp. 32-45. https://doi.org/ 10.5755/j01.erem.80.3.35146.

Mremi, A., Kimwaga, R., Mulungu, D.M.M. and Izdori, F.J., 2025. Spatial and temporal rainfall variability impacts on faecal sludge management services in Sinza River Catchment at Tandale Ward in Dar es Salaam, Tanzania. *Geomatics, Natural Hazards and Risk,* [e-journal] 16(1). https://doi.org/10.1080/19475705.2025.2462175.

Mwanthi, A.M., Mutemi, J.N., Opijah, F.J., Mutua, F.M., Atheru, Z. and Artan, G., 2024. Implications of WRF model resolutions on resolving rainfall variability with topography over East Africa. *Frontiers in Climate*, [e-journal] 6(1). https://doi.org/10.3389/fclim.2024.1311088.

Nakagawa, S. and Schielzeth, H., 2012. A general and simple method for obtaining R2 from generalized linear mixed-effects models. *Methods in Ecology and Evolution*, [e-journal] 4(2). pp. 133-142. https://doi.org/10.1111/j.2041-210x.2012.00261.x.

Nath, S., Mathew, A., Khandelwal, S. and Shekar, P.R., 2023. Rainfall and temperature dynamics in four Indian states: A comprehensive spatial and temporal trend analysis. *HydroResearch*, [e-journal] 6(1). pp. 247-254. https://doi.org/10.1016/j.hydres.2023.09.001.

Ng. C.Y., Wan Jaafar, W.Z., Othman, F., Lai, S.H., Mei, Y. and Juneng, L., 2024. Assessment of Evaporative Demand Drought Index for drought analysis in Peninsular Malaysia. *Science of the Total Environment,* [e-journal] 917(1). https://doi.org/10.1016/j.scitotenv.2024.170249.

Oh, S.G., Han, J.Y., Min, S.K. and Son, S.W., 2022. Impact of urban heat island on daily and sub-daily monsoon rainfall variabilities in East Asian megacities. *Climate Dynamics,* [e-journal] 61(1). https://doi.org/10.1007/s00382-022-06556-y.

Omer, A., Elagib, N.A., Ma, Z., Saleem, F. and Mohammed, A., 2020. Water scarcity in the Yellow River Basin under future climate change and human activities. *Science of The Total Environment*, [e-journal] 749(1). https://doi.org/10.1016/j.scitotenv.2020.141446.

Pande, C.B., Egbueri, J.C., Costache, R., Sidek, L.M., Wang, Q., Alshehri, F., Din, N.M., Gauam, V.K. and Pal, S.C., 2024. Predictive modeling of land surface temperature (LST) based on Landsat-8 satellite data and machine learning models for sustainable development. *Journal of Cleaner Production*, [e-journal] 444(1). https://doi.org/10.1016/j.jclepro.2024.141035.

Patel, S., Indraganti, M. and Jawarneh, R.N., 2024. A comprehensive systematic review: Impact of Land Use/ Land Cover (LULC) on Land Surface Temperatures (LST) and outdoor thermal comfort. *Building and Environment*, [e-journal] 249(1). https://doi.org/10.1016/j.buildenv.2023.111130.

Pen, S., Rad, S., Ban, L., Brang, S., Nuth, P and Liao, L., 2024. An Analysis of Extreme Rainfall Events in Cambodia. *Atmosphere*, [e-journal] 15(8). https://doi.org/10.3390/atmos15081017.

Phan, K.D., Can, T.N., Nguyen, K.D., Nguyen, T.H.D, Pham, T.B.T, Tran, G.H. and Thanh, N.P, 2024. Remote sensing for urban heat island research: Progress, current issues, and perspectives. *Remote Sensing Applications: Society and Environment*, [e-journal] 33(1). https://doi.org/10.1016/j.rsase.2023.101081.

Polovina, S., Radić, B., Ristić, R. and Milčanović, V., 2024. Application of Remote Sensing for Identifying Soil Erosion Processes on a Regional Scale: An Innovative Approach to Enhance the Erosion Potential Model. *Remote Sensing*, [e-journal] 16(13). https://doi.org/10.3390/rs16132390.

Puth, M.T., Neuhäuser, M. and Ruxton, G.D., 2015. Effective use of Spearman's and Kendall's correlation coefficients for association between two measured traits. *Animal Behaviour*, [e-journal] 102(1). pp. 77-84. https://doi.org/10.1016/j.anbehav.2015.01.010.

Qin, Y., Ghalambaz, S., Sheremet, M., Baro, M. and Ghalambaz, M., 2024. Deciphering Urban Heat Island Mitigation: A Comprehensive Analysis of Application Categories and Research Trends. *Sustainable Cities and Society*, [e-journal] 101(1). https://doi.org/10.1016/j.scs.2023.105081.

Rahaman, Z.A., Al Kafy, A., Saha, M., Rahim, A.A., Almulhim, A.I., Rahaman, S.N., Fattah, M.A., Rahman, M.T., Kalaivani, Al Faisal, A. and Al Rakib, A., 2022. Assessing the impacts of vegetation cover loss on surface temperature, urban heat island and carbon emission in Penang city, Malaysia. *Building and Environment*, [e-journal] 222(1). https://doi.org/10.1016/j.buildenv.2022.109335

Ran, L., Wang, Z., Yang, B., Margavi, A.A. and Alshahrani, N., 2024. Development of novel computational models based on artificial intelligence technique to predict the viscosity of ionic liquids-water mixtures. *Case Studies in Thermal Engineering*, [e-journal] 54(1). https://doi.org/10.1016/j.csite.2024.104076.

Rao, P., Tassinari, P. and Torreggiani, D., 2023. Exploring the land-use urban heat island nexus under climate change conditions using machine learning approach: A spatio-temporal analysis of remotely sensed data. *Heliyon*, [e-journal] 9(8). https://doi.org/10.1016/j.heliyon.2023.e18423

Rees, G., Baidy, L.H. and Belenok, V., 2024. Temporal Variations in Land Surface Temperature within an Urban Ecosystem: A Comprehensive Assessment of Land Use and Land Cover Change in Kharkiv, Ukraine. *Remote Sensing*, [e-journal] 16(9). https://doi.org/10.3390/rs16091637.

Reiners, P., Sobrino, J. and Kuenzer, C., 2023. Satellite-Derived Land Surface Temperature Dynamics in the Context of Global Change—A Review. *Remote Sensing*, [e-journal] 15(7). https://doi.org/10.3390/rs15071857.

Rendana, M., Idris, W.M.R., Rahim, S.A., Abdo, H.G., Almohamad, H., Dughairi, A.A.A. and Al-Mutiry, M., 2023. Relationships between land use types and urban heat island intensity in Hulu Langat district, Selangor, Malaysia. Ecological Processes, [e-journal] 12(1). https://doi.org/10.1186/s13717-023-00446-9.

Robertson, J.C., 1967. The Symap Programme for Computer Mapping. *The Cartographic Journal*, [e-journal] 4(2). pp. 108-113. https://doi.org/10.1179/caj.1967.4.2.108.

Rwanga, S.S. and Ndambuki, J.M., 2017. Accuracy Assessment of Land Use/Land Cover Classification Using Remote Sensing and GIS. *International Journal of Geosciences*, [e-journal] 8(1), pp. 611-622. https://doi.org/10.4236/ijg.2017.84033.

Sapawi, D.K.A., Chew, J.M., Shahidah, D.N., Rahman, M.N.A., Zulkifli, I., Hew, T.Y.A., Ismail, M.S., Wakid, S.A., Syazwan, W.M., Okamura, H., Horie, Y., Ong, M.C., Setyawan, A.D., Kumar, K., Cheng, W.H. and Yap, C.K., 2023. Public perception of flood risks in Klang Valley, Malaysia: A case study. *Sustainable Social Development*, [e-journal] 1(1). https://doi.org/10.54517/ssd.v1i1.2197.

Schober, P., Boer, C. and Schwarte, L., 2018. Correlation Coefficients: Appropriate Use and Interpretation. *Anesthesia & Analgesia*, [e-journal] 126(5). pp. 1763-1768. https://doi.org/10.1213/ANE.0000000000002864.

Sen, P.K., 1968. Estimates of the Regression Coefficient Based on Kendall's Tau. *Journal of the American Statistical Association,* [e-journal] 63(324). pp. 1379-1389. https://doi.org/10.1080/01621459.1968.10480934.

Sen, Z. (2012). Innovative trend analysis methodology. *Journal of Hydrologic Engineering,* [e-journal] 17(9). pp. 1042-1046. https://doi.org/10.1061/(ASCE)HE.1943-5584.0000556.

Sharma, A., Sajjad, H., Saha, T.K., Masroor, M., Sharma, Y. and Kumari G., 2024. Analyzing trend and forecasting of temperature and rainfall in Shimla district of Himachal Pradesh, India using non-parametric and bagging REPTree machine learning approaches. *Journal of Atmospheric and Solar–Terrestrial Physics*, [e-journal] 265(1). https://doi.org/10.1016/j.jastp.2024.106352.

Sharma, C.S., Panda, S.N., Pradhan, R.P., Singh, A. and Kawamura, A., 2016. Precipitation and temperature changes in eastern India by multiple trend detection methods. *Atmosphere Research*, [e-journal] 180(1). pp. 211-225. https://doi.org/10.1016/j.atmosres.2016.04.019.

Shi, H., Xian, G., Auch, R., Gallo, K. and Zhou, Q., 2021. Urban Heat Island and Its Regional Impacts Using Remotely Sensed Thermal Data—A Review of Recent Developments and Methodology. *Land*, [e-journal] 10(8). https://doi.org/10.3390/land10080867.

Siswanto, Schrier, G.V.D., Hurk, B.V.D., 2022. Observed Increase of Urban Extreme Rainfall as Surface Temperature Rise: The Jakarta Case. *Journal of the Meteorological Society of Japan,* [e-journal] 100(3). pp. 475-492. https://doi.org/10.2151/jmsj.2022-023

Steensen, B.M., Marelle, L., Hodnebrog, Ø. and Myhre, G., 2022. Future urban heat island influence on precipitation. *Climate Dynamics,* [e-journal] 58(1). https://doi.org/10.1007/s00382-021-06105-z.

Su, Q., Meng, X. and Sun, L., 2024. Investigation and Validation of Split-Window Algorithms for Estimating Land Surface Temperature from Landsat 9 TIRS-2 Data. *Remote Sensing,* [e-journal] 16(1). https://doi.org/10.3390/rs16193633.

Sun, Y., Gao, C., Li, J., Li, W. and Ma, R., 2018. Examining urban thermal environment dynamics and relations to biophysical composition and configuration and socio-economic factors: A case study of the Shanghai metropolitan region. *Sustainable Cities and Society,* [e-journal] 40(1). https://doi.org/10.1016/j.scs.2017.12.004.

Tahooni, A., Kakroodi, A.A. and Kiavarz, M., 2023. Monitoring of land surface albedo and its impact on land surface temperature (LST) using time series of remote sensing data. *Ecological Informatics*, [e-journal] 75(1). https://doi.org/10.1016/j.ecoinf.2023.102118.

Tahooni, A., Kakroodi, A.A. and Kiavarz, M., 2023. Monitoring of land surface albedo and its impact on land surface temperature (LST) using time series of remote sensing data. *Ecological Informatics*, [e-journal] 75(1). https://doi.org/10.1016/j.ecoinf.2023.102118.

Talekar, B. and Agrawal, S., 2020. A Detailed Review on Decision Tree and Random Forest. *Bioscience Biotechnology Research Communication,* [e-journal] 13(14). http://dx.doi.org/10.21786/bbrc/13.14/57.

Tan, J., Huffman, G.J., Bolvin, D.T. and Nelkin, E.J., 2019. IMERG V06: Changes to the Morphing Algorithm. *Journal of Atmospheric and Oceanic Technology,* [e-journal] 36(12). pp. 2471-2482. https://doi.org/10.1175/JTECH-D-19-0114.1.

Tang, J., Lan, X., Lian, Y., Zhao, F. and Li, T., 2022. Estimation of Urban–Rural Land Surface Temperature Difference at Different Elevations in the Qinling–Daba Mountains Using MODIS and the Random Forest Model. *International Journal of Environmental Research and Public Health*, [e-journal] 19(18). https://doi.org/10.3390/ijerph191811442

Tanoori, G., Soltani, A. and Modiri, A., 2024. Machine Learning for Urban Heat Island (UHI) Analysis: Predicting Land Surface Temperature (LST) in Urban Environments. Urban Climate, [e-journal] 55(1). https://doi.org/10.1016/j.uclim.2024.101962.

Tiwari, S., Jha, S.K. and Sivakumar, B., 2019. Reconstruction of daily rainfall data using the concepts of networks: Accounting for spatial connections in neighborhood selection. *Journal of Hydrology*, [e-journal] 579(1). https://doi.org/10.1016/j.jhydrol.2019.124185.

Verma, G., Kumar, B., Kumar, C., Ray, A. and Khandelwal, M., 2023. Application of KRR, K-NN and GPR Algorithms for Predicting the Soaked CBR of Fine-Grained Plastic Soils. *Arabian Journal for Science and Engineering*, [e-journal] 48(1). https://doi.org/10.1007/s13369-023-07962-y.

Vilcins, D., Sly, P.D., Scarth, P. and Mavoa, S., 2022. Green space in health research: an overview of common indicators of greenness. *Reviews on Environmental Health*, [e-journal] 39(2). pp.221-231. https://doi.org/10.1515/reveh-2022-0083

Wang, M., He, G., Zhang, Z., Wang, G., Wang, Z., Yin, R., Cui, S, Wu, Z. and Cao, X., 2019. A radiance-based split-window algorithm for land surface temperature retrieval: Theory and application to MODIS data. *International Journal of Applied Earth Observation and Geoinformation,* [e-journal] 76(1), pp. 204-217. https://doi.org/10.1016/j.jag.2018.11.015.

Wang, Y., Xu, Y., Tabari, H., Wang, J., Wang, Q., Song, S. and Hu, Z., 2020. Innovative trend analysis of annual and seasonal rainfall in the Yangtze River Delta, eastern China. Atmospheric Research, [e-journal] 231(1). https://doi.org/10.1016/j.atmosres.2019.104673.

Watson, J.T., Gayer, M. and Connolly, M.A., 2007. Epidemics after Natural Disasters. *Emerging Infectious Diseases*, [e-journal] 13(1). https://doi.org/10.3201%2Feid1301.060779.

Whitford, A.C., Blenkinsop, S., Pritchard, D. and Fowler, H.J., 2023. A gauge-based sub-daily extreme rainfall climatology for western Europe. *Weather and Climate Extremes*, [e-journal] 41(1). https://doi.org/10.1016/j.wace.2023.100585.

Wimala, S., Jirakajohnkool, S., Konisranukul, W. and Nakhapakorn, K., 2025. Comparison of Spatial Rainfall Interpolation by Using Statistical Methods at Thailand's Eastern Coast Basin. *Suranaree Journal of Social Science*, [e-journal] 19(1). https://doi.org/10.55766/sjss-1-2025-253620.

WorldData.info, 2024. The Climate in Malaysia [Online]. Available at: < https://www.worlddata.info/asia/malaysia/climate.php>. [Accessed on 21 August 2024].

Wu, Y., Xi, Y., Feng, M. and Peng, S., 2021. Wetlands Cool Land Surface Temperature in Tropical Regions but Warm in Boreal Regions. *Remote Sensing,* [e-journal] 13(8). https://doi.org/10.3390/rs13081439.

Xu, J., Jin, Y., Ling, Y., Sun, Y. and Wang, Y., 2025. Exploring the seasonal impacts of morphological spatial pattern of green spaces on the urban heat island. *Sustainable Cities and Society*, [e-journal] 125(1). https://doi.org/10.1016/j.scs.2025.106352.

Yan, H., Gao, Y., Wilby, R., Yu, D., Wright, N., Yin, J., Chen, X., Chen, J. and Guan, M., 2024. Urbanization Further Intensifies Short-Duration Rainfall Extremes in a Warmer Climate. *Geophysical Research Letters,* [e-journal] 51(1). https://doi.org/10.1029/2024GL108565

Zhou, D., Zhao, S., Liu, S., Zhang, L. and Zhu, C., 2014. Surface urban heat island in China's 32 major cities: Spatial patterns and drivers. *Remote Sensing of Environment,* [e-journal] 152(1), pp. 51-61. https://doi.org/10.1016/j.rse.2014.05.017.

APPENDICES

Appendix A: Threshold of 99th Percentile Hourly Extreme Rainfall.

No.	Station	Threshold (mm)
1	**Kolam Takungan Batu (0231391RF)**	9.34
2	**Taman Ehsan At Kepong W. Persekutuan (0231441RF)**	9.38
3	**Pusat Penyelidekan At Jps Ampang Selangor (0231351RF)**	10.09
4	**Km 10 Ulu Kelang At Uk Height (0231401RF)**	9.4
5	**Ibu Bekalan Km. 11 At Gombak W. Persekutuan (0230721RF)**	8.8
6	**Empangan Genting Klang At W. Persekutuan (0230631RF)**	8.6
7	**I/pejabat Jps Malaysia At W. Persekutuan (0230641RF)**	10.5
8	**Jln. Sg. Udang At Segambut (0231381RF)**	9.53
9	**Ldg. Edinburgh Site 2 At W. Persekutuan (0230651RF)**	10.5

Appendix B-1: Annual SUHII and Annual Rainfall at Kolam Takungan Batu Station.

Year	Annual SUHII	Annual Total of Hourly Extreme Rainfall (mm)	Annual Total Number of Hourly Extreme Rainfall Events
2007	1.5387702	1087	62
2008	1.664701	2051.5	97
2009	1.790633	1519.104	70
2010	1.916564	2284	98
2011	2.042495	1802.5	92
2012	2.168426	2133.279	101
2013	2.294358	1929.452	80
2014	2.420289	1665.395	76
2015	2.5462202	1932.356	81
2016	2.598028	1367.43	59
2017	2.649835	2192.603	94
2018	2.701643	1420.007	77
2019	2.75345	2296.613	99
2020	2.805258	2464.485	107
2021	2.857065	1913.592	99
2022	2.908873	2181.928	100
2023	2.9606803	2153.788	97

Appendix B-2: Annual SUHII and Annual Rainfall at Taman Ehsan At Kepong W. Persekutuan Station.

Year	Annual SUHII	Annual Total of Hourly Extreme Rainfall (mm)	Annual Total Number of Hourly Extreme Rainfall Events
2007	1.399024444	1244	65
2008	1.602457926	1514.5	76
2009	1.805891408	1836	81
2010	2.00932489	1511.723223	72
2011	2.212758372	1381.955653	82
2012	2.416191854	2040.996742	102
2013	2.619625336	2011.200153	96
2014	2.823058818	1572.049518	72
2015	3.0264923	2093.668941	88
2016	3.056046409	1420.239455	69
2017	3.085600518	1702.621839	80
2018	3.115154627	1626.00713	81
2019	3.144708736	1744.237113	83
2020	3.174262845	2210.800228	109
2021	3.203816955	1852.646877	98
2022	3.233371064	1842.308501	91
2023	3.262925173	2746.186276	118

Appendix B-3: Annual SUHII and Annual Rainfall at Pusat Penyelidekan At Jps Ampang Selangor Station.

Year	Annual SUHII	Annual Total of Hourly Extreme Rainfall (mm)	Annual Total Number of Hourly Extreme Rainfall Events
2007	1.254371701	2081.8	87
2008	1.335658487	2073.3	95
2009	1.416945273	1534	75
2010	1.498232059	1270.397434	58
2011	1.579518845	1557.112432	71
2012	1.660805631	2042.375695	87
2013	1.742092417	1908.411179	76
2014	1.823379203	2031.220355	93
2015	1.904665988	1307.293063	62
2016	1.970535783	1475.42824	67
2017	2.036405578	2247.202797	103
2018	2.102275373	1983.25733	87
2019	2.168145168	2770.501551	107
2020	2.234014963	2783.538226	113
2021	2.299884758	2085.539596	101
2022	2.365754552	2358.94464	117
2023	2.431624347	2027.398054	91

Appendix B-4: Annual SUHII and Annual Rainfall at Km 10 Ulu Kelang At Uk Height Station.

Year	Annual SUHII	Annual Total of Hourly Extreme Rainfall (mm)	Annual Total Number of Hourly Extreme Rainfall Events
2007	0.919739591	1857	88
2008	0.961661847	1533.5	70
2009	1.003584103	1355.5	64
2010	1.045506359	951	49
2011	1.087428615	1307.5	68
2012	1.129350871	2811.1	107
2013	1.171273127	2061.5	94
2014	1.213195383	2182.7	94
2015	1.255117638	2271	97
2016	1.368438935	1527	66
2017	1.481760232	2155.8	91
2018	1.595081529	2025.7	85
2019	1.708402826	2353	108
2020	1.821724122	2576.7	104
2021	1.935045419	2076.3	95
2022	2.048366716	2541.2	114
2023	2.161688013	1947.2	90

Appendix B-5: Annual SUHII and Annual Rainfall at Ibu Bekalan Km. 11 At Gombak W. Persekutuan Station.

Year	Annual SUHII	Annual Total of Hourly Extreme Rainfall (mm)	Annual Total Number of Hourly Extreme Rainfall Events
2007	1.465395149	1757	80
2008	1.524280038	1940	91
2009	1.583164926	2098.3	95
2010	1.642049815	1614	81
2011	1.700934703	1377	64
2012	1.759819592	1682.1	84
2013	1.81870448	1893.9	89
2014	1.877589369	1477	69
2015	1.936474257	2005.6	91
2016	2.072637018	1226.3	59
2017	2.208799779	2116.1	99
2018	2.34496254	1570	88
2019	2.481125301	2056.3	96
2020	2.617288062	1876	88
2021	2.753450823	1892.1	102
2022	2.889613584	2069.1	102
2023	3.025776345	2099.3	107

Appendix B-6: Annual SUHII and Annual Rainfall at Empangan Genting Klang At W. Persekutuan Station.

Year	Annual SUHII	Annual Total of Hourly Extreme Rainfall (mm)	Annual Total Number of Hourly Extreme Rainfall Events
2007	-0.084530554	1431	73
2008	-0.121236422	2026.5	92
2009	-0.157942289	999.5	46
2010	-0.194648156	1848.5	89
2011	-0.231354024	1726	92
2012	-0.268059891	1993.2	98
2013	-0.304765758	1703.3	84
2014	-0.341471626	1634.2	84
2015	-0.378177493	2214	99
2016	-0.232028469	1265.5	65
2017	-0.085879444	1891.3	99
2018	0.06026958	1325.6	77
2019	0.206418604	1806.5	82
2020	0.352567628	2125.5	103
2021	0.498716652	2005.4	98
2022	0.644865676	1918	109
2023	0.791014701	1859.1	100

Appendix B-7: Annual SUHII and Annual Rainfall at I/pejabat Jps Malaysia At W. Persekutuan Station.

Year	Annual SUHII	Annual Total of Hourly Extreme Rainfall (mm)	Annual Total Number of Hourly Extreme Rainfall Events
2007	0.745248782	1600.8	66
2008	0.920639947	2983.8	127
2009	1.096031111	1439.5	68
2010	1.271422276	1502.5	61
2011	1.44681344	1454	71
2012	1.622204605	1875.397573	83
2013	1.79759577	2221.176367	94
2014	1.972986934	1877.335745	75
2015	2.148378099	1881.553695	91
2016	2.116373114	1779.16392	79
2017	2.084368129	2049.466887	88
2018	2.052363144	2303.6317	102
2019	2.020358159	2391.921028	97
2020	1.988353174	2079.796643	92
2021	1.956348188	1677.591336	73
2022	1.924343203	2341.97777	101
2023	1.892338218	2285.379069	101

Appendix B-8: Annual SUHII and Annual Rainfall at Jln. Sg. Udang At Segambut Station.

Year	Annual SUHII	Annual Total of Hourly Extreme Rainfall (mm)	Annual Total Number of Hourly Extreme Rainfall Events
2007	2.002210688	1940.5	87
2008	2.222129709	2174	96
2009	2.442048731	1494.696529	67
2010	2.661967753	1269	57
2011	2.881886775	1586	82
2012	3.101805796	2200.44505	92
2013	3.321724818	2355.103075	95
2014	3.54164384	1684.264984	85
2015	3.761562861	1979.37709	90
2016	3.748155641	1800.682812	80
2017	3.734748421	1882.539994	86
2018	3.721341202	2097.066423	95
2019	3.707933982	2573.031511	108
2020	3.694526762	2237.565753	106
2021	3.681119542	1835.403479	84
2022	3.667712322	1941.727278	92
2023	3.654305102	1898.360627	87

Appendix B-9: Annual SUHII and Annual Rainfall at Ldg. Edinburgh Site 2 At W. Persekutuan Station.

Year	Annual SUHII	Annual Total of Hourly Extreme Rainfall (mm)	Annual Total Number of Hourly Extreme Rainfall Events
2007	0.45620937	1909	85
2008	0.613527711	1798.564663	83
2009	0.770846051	1493.5	65
2010	0.928164392	1697	73
2011	1.085482732	1618.5	76
2012	1.242801073	1974.428332	83
2013	1.400119413	2572.618284	98
2014	1.557437754	1430.805937	63
2015	1.714756094	2637.15881	109
2016	1.683358609	1383.220905	59
2017	1.651961123	2111.27459	92
2018	1.620563637	2293.883498	102
2019	1.589166151	2544.476238	109
2020	1.557768665	2284.056933	93
2021	1.52637118	2225.98593	95
2022	1.494973694	2024.644796	94
2023	1.463576208	2833.752027	111