BAREKENG: Journal of Mathematics and Its Applications
December 2026 Volume 20 Issue 4 Page 3035-3052
P-ISSN: 1978-7227 E-ISSN: 2615-3017
https://doi.org/10.30598/barekengvol20iss4pp3035-3052

CLIMATE COMFORT INDEX ANALYSIS USING SPATIO-TEMPORAL PCA-FASTMCD METHOD

Agus Yarcana[1*], **Henny Pramoedyo**[2], **Suci Astutik**[3]

[1,2,3]*Department of Statistics, Faculty of Mathematics and Natural Sciences, Universitas Brawijaya*
Jln. Veteran, Malang, 65145, Indonesia

[1]*Indonesian Agency for Meteorology, Climatology, and Geophysics (BMKG)*
Jln. Angkasa I No 2, Kemayoran Jakarta Pusat, 10720, Indonesia

*Corresponding author's e-mail: * **agus.yarcana@bmkg.go.id.com***

Article Info

Article History:

Received:26th September 2025
Revised: 8th May 2026
Accepted: 7th June 2026
Available online: 24th August 2026

Keywords:

Climate Comfort Index;
FASTMCD;
Functional Data Analysis;
Spatial Weight Matrix; Spatio-Temporal PCA (STPCA).

ABSTRACT

This study addresses the challenge of modeling spatio-temporal climate data that are often affected by outliers, which significantly bias conventional principal component analysis. The main contribution of this research is not merely the application of Spatio-Temporal Principal Component Analysis (STPCA), but its novel integration with the Fast Minimum Covariance Determinant (FASTMCD) method to obtain robust spatio-temporal components that are resilient to outliers in Bali's climate data. The core methodology involves transforming four climate variables (thermal comfort, cloud cover, rainfall, and wind speed) from 24 stations in Bali (2010–2019) using Fourier basis expansion, applying spatial weighting, and utilizing robust covariance estimation via FASTMCD. The results indicate that the Inverse Power Distance (IPD) weighting scheme optimally captures the spatial structure. Furthermore, the first robust principal component (STPC1) reveals the dominant climate variability, which is driven primarily by thermal comfort and wind speed. This component successfully highlights a clear spatial differentiation between coastal lowland and highland regions despite the presence of extreme observations. These findings imply that the robust STPCA-FASTMCD framework provides a highly stable representation of regional climate patterns, offering a reliable analytical tool for developing climate comfort indices and supporting climate-informed tourism planning in tropical regions.

How to cite this article:

A.Yarcana, H.Pramoedyo, and S.Astutik, "CLIMATE COMFORT INDEX ANALYSIS USING SPATIO-TEMPORAL PCA-FASTMCD METHOD", *BAREKENG: J. Math. & App.,* vol. 20, no. 4, pp. 3035-3052, Dec. 2026.

Journal homepage: *https://ojs3.unpatti.ac.id/index.php/barekeng/*
Journal e-mail: *barekeng.math@yahoo.com; barekeng.journal@mail.unpatti.ac.id*

Research Article · Access

1. INTRODUCTION

Studies on climate comfort and climate variability have been widely conducted using multivariate statistical approaches, including Principal Component Analysis (PCA). However, a significant scientific gap remains: while studies on climate comfort already exist, none have used a spatio-temporal robust approach based on STPCA- Fast Minimum Covariance Determinant (FASTMCD) with an evaluation of several spatial weight matrices. The novelty of this research lies in integrating STPCA with FASTMCD to produce principal components that are robust to outliers, testing different spatial weight matrices to identify the most suitable structure for climate data, and reconstructing a climate comfort index tailored to Bali's tropical conditions, which has not been addressed in previous studies.

Most existing studies rely on standard PCA or its extensions without explicitly accounting for the spatio-temporal structure of the data and the presence of outliers, which are common in climate observations. Data dimension reduction is one of the main functions of the multivariate method PCA, which reduces data by forming linear combinations of variables. These linear combinations assign optimized weights to each variable in the dataset, thereby resulting in lower-dimensional data [1]. The STPCA method is an extension of the PCA method, which is applied to data with spatial and temporal components. STPCA is capable of reducing variables and shows hidden patterns in various spatio-temporal datasets [2]. Given these functional characteristics, climate data are exceptionally well-suited for analysis via STPCA, as climate variables evolve continuously over time and exhibit spatial dependence across locations. However, several methodological considerations may influence the analytical outcomes.

Several factors can influence the outcomes of an STPCA. The first is the data processing stage, as simultaneous processing steps may alter the final results. This is because STPCA methods have two key stages: the transformation of time series data into multivariate functional data and the subsequent construction of the STPCA components. The second factor is the scale of the variables, as data measured on a ratio scale is known to yield more reliable reduction results. The third factor is the determination of an appropriate spatial weight matrix that is suitable for the characteristics of the research data [2]. Another factor is the choice of rotation, either orthogonal or oblique [3]. Despite these considerations, most STPCA applications assume spatio-temporal structures unaffected by outliers, which is often unrealistic in climate data.

The determination of the weight type strongly depends on its relevance to the research discipline [4]. For climate research, especially when it involves rainfall data variables, the most suitable spatial weight matrix is the K-Nearest Neighbors (KNN) [5]. Meanwhile, when it involves wind speed data variables, the best weight matrix is the inverse power distance (IPD) matrix [6]. Accordingly, climate data reduction using STPCA requires testing with several spatial weight matrices, such as KNN and inverse power distance. Spatial weighting schemes are widely used, but their systematic evaluation in climate studies with robust spatio-temporal methods remains limited.

Climate data often contain outliers, which can substantially affect the decomposition of the variance–covariance matrix in multivariate methods such as PCA. Because PCA relies on covariance structure, extreme observations may exert disproportionate influence, leading to inflated eigenvalues and distorted loading patterns, so that the resulting principal components reflect isolated extreme events rather than the dominant variability of the dataset [7]. In climate-related studies, such outliers frequently correspond to extreme meteorological events, including heatwaves, prolonged droughts, or intense rainfall episodes, which are critical phenomena but may bias the representation of general climate patterns [8]. Hence, it is necessary to reconstruct a robust variance–covariance matrix. FASTMCD was chosen over other robust methods because it is an efficient modification of the MCD estimator that can be combined with the Projection Pursuit (PP) method, and when applied within PCA, it produces principal components that are resistant to outliers and better represent the underlying structure of the data [9].

The practical benefit of these robust results is significant for tropical environments. This study will be applied to climate data used in constructing a climate comfort index, which will measure how atmospheric or climatic physical conditions affect human beings [10]. This index approach is used because weather and climate variables are highly diverse, and also to minimize differences in perceptions of climate comfort. There are several types of indices commonly used to measure climate comfort, which are: Heat Index (Humidex), Temperature Humidity Index (THI), Tourism Climate Index (TCI), and Holiday Climate Index (HCI). TCI and HCI climate comfort indices, in particular, were developed to support tourism activities. These two indices are commonly known to have a weight system and an objective rating system based on surveys that were conducted among tourists in subtropical regions. Therefore, when these indices were going to be applied in regions with different climate types, they required further adjustment and analysis [11].

One of the leading tourist destinations in Indonesia is Bali, and to maintain tourist arrivals in Bali, one of the efforts that was done is to provide information on climate comfort indices for tourism activities. Climate comfort indices are affected by the frequency of extreme weather events. The data shows that from 2010 until 2019, the frequency of extreme temperatures increased by 10% compared to the normal average conditions that were recorded from 1961 to 1990 [12]. In addition, from 2010-2019 represented a normal period of tourist arrivals prior to the COVID-19 global pandemic. After COVID-19 was declared in Indonesia in March 2020, the number of tourist arrivals in Bali decreased by 50% [13]. This declining amount of tourist arrivals was even more pronounced in subsequent months, continuing until 2022. Therefore, this study uses data from 2010-2019.

Drawing from the above background, this study is conducted to determine the optimal spatial weight matrix (KNN and IPD) for reducing climate data by using the STPCA-FASTMCD method and to group regions in Bali based on the derived STPCA-FASTMCD climate comfort index. The novelty of this research lies in integrating STPCA with FASTMCD to produce principal components that are robust to outliers, testing different spatial weight matrices to identify the most suitable structure for climate data, and reconstructing a climate comfort index tailored to Bali's tropical conditions, which has not been addressed in previous studies.

2. RESEARCH METHODS

2.1 Research Data

The data used in this research were obtained from two resources, which were obtained from two sources: secondary in situ weather observation data from BMKG stations in Bali and ERA5 reanalysis data. European Centre for Medium-Range Weather Forecasts (ECMWF). Reanalysis data are the result of data assimilation from global observation data, including synoptic observations, upper-air measurements, and remote sensing data, which are then compiled into three-dimensional datasets based on the latest global climate models [14]. This data serves as a proxy to be used in research conducted at locations without in situ observations. The weather observation was conducted at 24 locations, which are: 3 BMKG offices (I Gusti Ngurah Rai Meteorological Station, Denpasar Geophysics Station, and Bali Climatological Station), 1 synoptic observation post (Kahang Kahang Post), and 20 rainfall measurement posts. The complete location details of the climate observation sites are presented in Fig. 1.

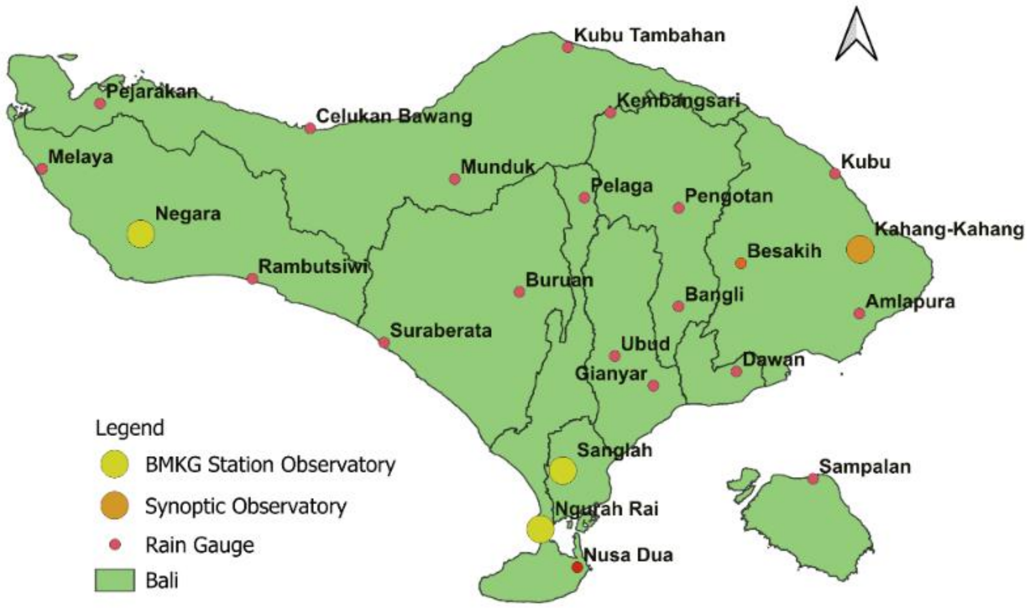


Figure 1. Climate Observation Sites in Bali

The variables that were used in this study consist of four parameters ($p = 4$), which were: thermal comfort (TC), cloud cover (A), rain rate (R), and wind speed (W). All variables were based on a monthly average from January 2010 until December 2019. To clarify the analytical framework, the initial data is structured on a monthly basis across a temporal dimension of $T = 120$ months for each of the $n = 24$ spatial stations. This forms an initial three-dimensional data structure of $n \times p \times T$. The data period corresponds to a 10–12-year (decadal) dataset, which has the same analysis result and prediction as the 30-year dataset (standard normal data). Both decadal and standard normal datasets are updated every 10 years [15]. The selection of these four variables is fundamentally driven by the HCI framework, a globally recognized standard for evaluating climate comfort in tourist destinations. Within this framework, air temperature and relative humidity are integrated into a single TC variable to measure heat stress and prevent multicollinearity, while A, R, and W represent the essential aesthetic and physical facets of outdoor tourism, making this configuration highly representative for classifying climate comfort in Bali. The complete list of variables is presented in Table 1.

Table 1. Research Variables Table

Variable	Symbol	Unit	Scale
Monthly Average Thermal Comfort (TC)	x_1	Celcius	Interval
Monthly Average Cloud Cover (A)	x_2	Percent	Ratio
Monthly Average Rain Rate (R)	x_3	Millimeter per day	Ratio
Monthly Average Wind Speed (W)	x_4	Kilometer per hour	Ratio

2.2 Research Stages

1. Establish the analytical framework and mathematical notation to ensure consistency throughout the analysis. The sequence of transformations is defined by the following matrix: the original observed data matrix ($\boldsymbol{X}$); the standardized data matrix ($\boldsymbol{X'}$); the number of significant bases (B); the resulting basis coefficient matrix ($\boldsymbol{A}$); which serves as the input for robust analysis; the spatial weight matrix ($\boldsymbol{W}$); and the final spatio-temporal matrix ($\boldsymbol{C}$). The core STPCA is executed on the basis of the coefficient space rather than the raw observations.
2. Obtain the necessary data for the analysis, which consists of maximum temperature, relative humidity, cloud cover, rainfall, and wind speed data from BMKG and ECMWF.
3. Perform data preprocessing, which consists of missing data imputation using the K-Nearest Neighbors (KNN) method with $k = 12$, outlier identification with Mahalanobis distance calculation, and data standardization.
4. Transform temporal data into Fourier basis functional data and determine the optimal number of Fourier bases (B_k) using the Bayesian Information Criterion (BIC).
5. Construct the basis coefficient matrix ($\boldsymbol{A}$) with dimensions $n \times (p \times B_{k+1})$, the spatial weight matrix KNN ($\boldsymbol{W}_{KNN}$) with dimensions $n \times n$, Inverse Power Distance (IPD) matrix with $r = 3$ ($\boldsymbol{W}_{IPD\ r3}$), and IPD with $r = 4$ ($\boldsymbol{W}_{IPD\ r4}$).
6. Calculate Moran's I spatial autocorrelation values based on the three spatial weights.
7. Determine the optimal subset size (h) using FASTMCD, and calculate outlyingness with the Stahel-Donoho index on matrix $\boldsymbol{A}$, to identify outlier locations. Operationally, the h-subset is determined by retaining the h spatial locations that have the smallest robust distances. The mechanism for identifying location outliers involves projecting the data onto multiple directional vectors
8. Construct the robust basis matrix $\boldsymbol{A}_{robust}$ with dimensions $h \times (p \times B_{k+1})$, three types of robust spatial weight matrix $\boldsymbol{W}_{robust}$ with dimensions $h \times h$**,** and the robust spatio-temporal matrix $\boldsymbol{C}_{robust}$ with dimensions $(p \times B_{k+1}) \times (p \times B_{k+1})$. The robust spatial weight matrix ($\boldsymbol{W}_{robust}$) is formed operationally by deleting the rows and columns corresponding to the identified outlier locations from the original spatial weight matrix ($\boldsymbol{W}$). This ensures that extreme locations are completely removed and do not distort the spatial dependency measurements.

9. Decompose the robust spatio-temporal matrix ($\boldsymbol{C}_{robust}$) and compute the loadings and spatio-temporal principal components ($STPC^{+}$and $STPC^{-}$).
10. Compare the eigenvalues of STPCA–FASTMCD across the three types of spatial weight matrix and determine the best spatial weighting scheme.
11. Interpret the results and classify the climate comfort index in Bali based on STPCA–FASTMCD.

2.3 Data Preprocessing

The data preprocessing stage consists of missing data imputation using KNN, outlier identification through Mahalanobis distance calculation, and data standardization. The KNN imputation procedure begins with determining the number of nearest neighbors (k), followed by computing the Euclidean distance (d_{ij}) using Eq. (1).

$$d_{ij} = \left(\sum_{\mathrm{k=1}}^{p}\left(y_{ik} - y_{jk}\right)^2\right)^{\frac{1}{2}}, \tag{1}$$

with:
y_{ik} : value of the k-th at location i;
$\widehat{\alpha}_{kb}$: value of the k-th at location j;
p : number of variables.

While referring to the Euclidean distance values, the next step is the imputation of KNN $\underline{x}_j$ calculation using Eq. (2), where $w_k = \frac{1}{d_{ij}}$ [16].

$$\hat{x}_j = \frac{\sum_{k=1}^{K} w_k\, y_{jk}}{\sum_{k=1}^{K} w_k}\ . \tag{2}$$

After obtaining the imputed missing data, the mean value (μ_2) is calculated. This mean is then compared with the mean prior to imputation (μ_1) using the Z-test.

$H_0 : \mu_1 = \mu_2$ (No difference in means); while

$H_1 : \mu_1 \neq \mu_2$ (There is a difference in means),

with a significance level of α (0.05).

The null hypothesis (H_0) is accepted at the significance level α if $|Z| \leq Z_{1-\frac{\alpha}{2}}$ which indicates that there is no difference between the mean values before and after imputation. Meanwhile, the null hypothesis (H_0) is rejected if $|Z| > Z_{1-\frac{\alpha}{2}}$ which indicates there is a difference between the mean values before and after imputation.

One of the commonly used methods to detect multivariate outliers is the Mahalanobis distance [17]. The Mahalanobis distance calculation d_i^2 is given in Eq. (3). Outliers are identified if the Mahalanobis distance calculation result exceeds the threshold of the Chi-Square distribution [18]. If the Mahalanobis follows the Chi-Square distribution, the data are not outliers.

$$d_i^2 = (\boldsymbol{x}_i - \boldsymbol{\mu})^{\mathrm{T}}\Sigma^{-1}(\boldsymbol{x}_i - \boldsymbol{\mu})\,, \tag{3}$$

With:
$\boldsymbol{x}_i$: random vector at the i-th observation;
$\boldsymbol{\mu}$: mean vector;
Σ : covariance matrix.

The scaling process is one of the preprocessing steps in modeling, one of which is data standardization. Data standardization is the process of rescaling with a mean value of 0 and a standard deviation of 1 [19]. Since climate variables have different units (for example, temperature in degrees Celsius, rainfall in mm/day, and wind speed in km/h), standardization is required according to Eq. (4).

$$x_{ijt}{}' = \left(x_{ijt} - \mu_j\right)/\sigma_j \tag{4}$$

With:
$x_{ijt}{}'$: the standardized value in locations i-th, variable j-th, times t-th;
μ_j : mean population in variable j-th;
σ_j : standard deviation of the population in variable j-th.

2.4 Fourier Basis Functional Transformation and Basis Function Coefficients

The transformation of temporal data into functional data aims to project the data into a functional space that is defined by an orthonormal basis. This process also needs the selection of basic function systems, such as Fourier, Spline, Wavelet, and others. The final result of this transformation is multivariate functional data (MFD), in which each variable is measured as a continuous function over time according to Eq. (5).

$$\boldsymbol{y}_k(t) = \sum_{b=0}^{B} \widehat{\boldsymbol{\alpha}}_{kb}\, \boldsymbol{\phi}_b(t). \tag{5}$$

With:
$\boldsymbol{y}_i(t)$: variable i-th as time function t;
$\widehat{\boldsymbol{\alpha}}_{kb}$: basic function coefficient estimation in variable k-th and b-th basis;
$\boldsymbol{\phi}_b(t)$: basic function vector Fourier in b-th basis;
B : number of significant bases.

This research uses a Fourier basis, where there are three criteria for basis functions according to Eq. (6), with b is basis and T is the total time period.

$$\phi_b(t) = \begin{cases} \phi_0(t) = \dfrac{1}{\sqrt{T}}, \\ \phi_{2b-1}(t) = \sqrt{\dfrac{2}{T}} \sin \dfrac{2\pi bt}{T}, \\ \phi_{2b}(t) = \sqrt{\dfrac{2}{T}} \cos \dfrac{2\pi bt}{T}. \end{cases} \tag{6}$$

Significant basis value B affect the smoothness level of the functional data. The smaller the B value, the greater the degree of curve smoothing. The optimal value of B (B_k) would be determined by comparing the Bayesian Information Criterion (BIC), where $BIC = \ln\left(\sum_{k=0}^{K}\left(y_k - \sum_{b=0}^{B} \hat{\alpha}_{kb}\, \phi_{kb}(t)\right)^2\right) + (B + 1)\left(\frac{\ln k}{k}\right)$ [20].

As a result of this transformation, the basis function coefficient vectors (α) is derived. These coefficient vectors are combined to form matrix $\boldsymbol{A} = (\alpha_1, \alpha_2, \alpha_3 \ldots \alpha_p)$. Matrix $\boldsymbol{A}$ as defined in Eq. (7), serves as input to construct the spatio-temporal matrix. In matrix $\boldsymbol{A}$, n represents the number of observations/locations, p represents the number of variables, and B_k represents the number of optimal basis functions.

$$\boldsymbol{A} = \begin{bmatrix} \alpha_{1(1.0)} & \alpha_{1(1.1)} & \cdots & \alpha_{1(1.B_{k+1})} & \alpha_{2(1.0)} & \cdots & \alpha_{p(1.B_{k+1})} \\ \alpha_{1(2.0)} & \alpha_{1(2.1)} & \cdots & \alpha_{1(2.B_{k+1})} & \alpha_{2(2.0)} & \cdots & \alpha_{p(2.B_{k+1})} \\ \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ \alpha_{1(n.0)} & \alpha_{1(n.1)} & \cdots & \alpha_{1(n.B_{k+1})} & \alpha_{2(n.0)} & \vdots & \alpha_{p(n.B_{k+1})} \end{bmatrix}_{n\times(p\times B_{k+1})}. \tag{7}$$

2.5 Spatial Weight and Autocorrelation Moran's I Test

Spatial weight matrix ($\boldsymbol{W}$) in Eq. (8), will determine the number of primary components in STPCA. The order of matrix $\boldsymbol{W}$ is $n \times n$, where n represents the number of observations/locations. Some of the spatial weight matrix methods that are commonly used in spatial analysis are KNN and IPD. Spatial weight in the KNN method is strongly influenced by the choice of k as described in Eq. (9). It is known that the number of the nearest neighbors k could be determined by calculating the square root values from the overall number of observation data [21].

$$\boldsymbol{W}_{ij} = \begin{bmatrix} 0 & w_{12} & \cdots & w_{1n} \\ w_{21} & 0 & \cdots & w_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ w_{n1} & w_{n2} & \cdots & 0 \end{bmatrix}_{n\times n}, \tag{8}$$

$$w_{ij} = \begin{cases} 1, & if\ j \in k \\ 0, & other \end{cases}. \tag{9}$$

The IPD spatial weight is a variation of power distance, where this weight will give a larger spatial effect to the nearest point. Meanwhile, for points that are farther apart, the spatial effect is smaller. This is based on the inverse power relationship with distance [22]. The optimal distance decay parameter for a small dataset (≤19 data) is 2, whereas for a larger dataset (≥25 data) the optimal distance decay parameters are 3 and 4 [23]. This distance calculation is performed using Euclidean distance formula, as fully presented in Eq. (10).

$$w_{ij} = \frac{1}{d_{ij}^{r}}. \tag{10}$$

Spatial autocorrelation could be calculated using various methods that evaluate to which extent an attribute value tends to be correlated or clustered geographically. These measurements are divided into two types: global and local. Global autocorrelation provides an overview of spatial association patterns across the same area, meanwhile local autocorrelation focuses on the specific relationships between a unit and its neighboring units [24]. One of the statistical methods commonly used to measure spatial autocorrelation and suitable for STPCA application is Moran's I. The equation for Moran's I index with standardized weight in Eq. (11) with n represents number of locations, $\overline{\boldsymbol{x}'}$ represents mean variable value across all locations, x'_{ijt} and x'_{kjt} are the variable values at locations i and k respectively. The Moran's I index ranges from -1 (perfect negative autocorrelation) to 1 (perfect positive autocorrelation). Values $I > 0$ indicate positive spatial autocorrelation, meaning the data exhibit clustering, while values of $I < 0$ indicate negative spatial autocorrelation, meaning the data are dispersed [25].

$$I\left(\overline{\boldsymbol{x}}'\right) = \frac{n}{\sum_{i=1}^{n}\sum_{k=1}^{n} w_{ik}} \frac{\sum_{i=1}^{n}\sum_{j=1}^{n} w_{ij}\left(x'_{ijt} - \overline{x}'_{jt}\right)\left(x'_{kjt} - \overline{x}'_{jt}\right)}{\sum_{i=1}^{n}\left(x'_{ijt} - \overline{x}'_{jt}\right)^2}. \tag{11}$$

2.6 STPCA-FASTMCD Application

The first step to the STPCA-FASTMCD application is to construct a spatio-temporal matrix ($\boldsymbol{C}$) with an order $p \times p$, as shown in Eq. (12).

$$\boldsymbol{C} = \left(\sum_{ij=1}^{n} w_{ij}\right)^{-1} \boldsymbol{A}^T \boldsymbol{W} \boldsymbol{A}. \tag{12}$$

However, since matrix $\boldsymbol{C}$ is sensitive to outliers; an approach is needed to ensure the FASTMCD method is robust. The FASTMCD method combines two key concepts: the measurements of outlyingness and the FASTMCD algorithm to produce robust principal components. The process begins with selecting a data subset (h) through the parameter α which ranges between 0.5 and 1 as shown at Eq. (13). The next step involves calculating outlyingness according to Eq. (14) [26].

$$h = max(\alpha_n, \lfloor (n + p + 1)/2 \rfloor), \tag{13}$$

$$O(\boldsymbol{x}_k) = max_{V\in B}\left(\left|x_k^T \boldsymbol{v} - \hat{\mu}_{MCD} x_i^{\mathrm{T}} \boldsymbol{v}\right| / \hat{S}_{MCD}\left(x_i^T \boldsymbol{v}\right)\right). \tag{14}$$

With:

n	: number of observations;
p	: number of locations;
$O(\boldsymbol{x}_k)$	: outlyingness value of vector $\boldsymbol{x}_k$;
$V \in B$	: set of directional vectors;
x_k	: the k-th observation being evaluated for outlyingness;
x_i	: the i-th observation ($i = 1,2,3 \dots n$);
$\boldsymbol{v}$	: directional vector considered significant;
$\hat{\mu}_{MCD}$	: robust mean estimate using MCD;

$\hat{S}_{MCD}$: robust standard deviation estimate using MCD.

The application of FASTMCD results in the identification of outliers according to the h-subset criterion in matrix $\boldsymbol{A}$, producing the robust matrix $\boldsymbol{A}_{robust}$. The same process also applies to the matrix $\boldsymbol{W}$, resulting in the formation of a robust spatial weight matrix $\boldsymbol{W}_{robust}$. Consequently, the robust spatio-temporal matrix $\boldsymbol{C}_{robust}$ as shown in Eq. (15).

$$\boldsymbol{C}_{robust} = \left(\sum_{ij=1}^{N} w_{ij}\right)^{-1} \boldsymbol{A}_{robust}{}^{T} \boldsymbol{W}_{robust}\, \boldsymbol{A}_{robust} \tag{15}$$

Decomposition matrix $\boldsymbol{C}_{robust}$ produces the eigenvalues (γ) and eigenvectors ($\boldsymbol{\omega}$) of STPCA-FASTMCD using the relationship $\boldsymbol{C}_{robust}\boldsymbol{\omega} - \gamma\boldsymbol{\omega} = 0$. The decomposition results in two types of spatio-temporal loading vectors: the positive eigenvector of order q ($\boldsymbol{\omega}_q^+$) and the negative eigenvector of order ($\boldsymbol{\omega}_q^-$). These eigenvectors are derived from the corresponding positive and negative eigenvalues. They are then used to calculate principal spatio-temporal component values at each location. The spatio-temporal principal component values for the positive and negative eigenvectors of order q with dimension $k \times 1$ and the basis function coefficients at location p with also dimension $k \times 1$ and both are formulated in Eqs. (16) and (17).

$$\begin{aligned} STPC_{pq}^{+} &= \left(\boldsymbol{\omega}_{qb}^{+}\right)^{\mathrm{T}} \boldsymbol{\alpha}_{pb} \\ STPC_{pq}^{+} &= \omega_{q1}^{+}\alpha_{p1} + \omega_{q2}^{+}\alpha_{p2} + \cdots \omega_{qb}^{+}\alpha_{pb} \end{aligned} \tag{16}$$

$$\begin{aligned} STPC_{pq}^{-} &= \left(\boldsymbol{\omega}_{qb}^{-}\right)^{\mathrm{T}} \boldsymbol{\alpha}_{pb} \\ STPC_{pq}^{-} &= \omega_{q1}^{-}\alpha_{p1} + \omega_{q2}^{-}\alpha_{p2} + \cdots \omega_{qb}^{-}\alpha_{pb} \end{aligned} \tag{17}$$

With:

$\boldsymbol{STPC}_{pq}^{+}$: positive spatio-temporal principal component score of order q at location p;
$\boldsymbol{STPC}_{pq}^{-}$: negative spatio-temporal principal component score of order q at location p;
$\boldsymbol{\omega}_{\mathrm{qb}}^{+}$: positive spatio-temporal eigenvector of order q for the b-th basis;
$\boldsymbol{\omega}_{\mathrm{qb}}^{-}$: negative spatio-temporal eigenvector of order q for the b-th basis;
$\boldsymbol{\alpha}_{p\mathrm{b}}$: basis function coefficient at location p for the b-th basis.

2.7 Quality Measure of STPCA–FASTMCD Eigenvalues

The total amount of eigenvalues from the STPCA-FASTMCD method could not be interpreted in the same way as in classic PCA. Therefore, a specific quality measure is used, which is based on the comparison of the two largest eigenvalues to the total of all positive and negative eigenvalues [2]. This measure is referred to as the eigenvalue quality measure, defined by Eqs. (18) and (19).

$$M^{+} = \frac{(\gamma_1^{+} + \gamma_2^{+})}{\sum_{i=1}^{r} \gamma_i^{+}} \times 100\%, \tag{18}$$

$$M^{-} = \frac{(|\gamma_1^{-}| + |\gamma_2^{-}|)}{\sum_{i=1}^{r} \gamma_i^{-}} \times 100\%. \tag{19}$$

With:

$\gamma_1^{+} + \gamma_2^{+}$: sum of the first and second positive eigenvalues;
$|\gamma_1^{-}| + |\gamma_2^{-}|$: sum of the first and second absolute negative eigenvalues;
γ_i^{+} : the i-th positive eigenvalue;
γ_i^{-} : the i-th negative eigenvalue;
r : total number of eigenvalues.

2.8 Climate Comfort Index

The level of climate comfort varies across regions and time periods. Differences in the climate comfort by region are influenced by the duration of each season. These differences can affect the attractiveness of tourist destinations and their development planning [27]. In general, there are two climate comfort indices for tourism activities: Tourism Comfort Index (TCI) and Holiday Comfort Index (HCI). Based on a study in

Europe over 10 years, an equation was derived for urban area and beach, each of them, namely Holiday Comfort Index: Urban (HCIU) and Holiday Comfort Index: Beach (HCIB), as can be seen in Eqs. (20) and (21)[10]. These equations become a basis for weighting comparison in the climate comfort index constructed using the STPCA-FASTMCD method.

$$HCIU = 4(TC) + 2(A) + 3(R) + (W), \tag{20}$$

$$HCIB = 2(TC) + 4(A) + 3(R) + (W). \tag{21}$$

Whereas TC is Thermal Comfort resulting from the calculation between maximum temperature (T Max) and relative humidity (RH), according to Eq. (22).

$$TC = 0.8(T\ Max) + (RH(T\ Max)/500). \tag{22}$$

Meanwhile, A is Aesthetic, meaning cloud cover is defined as a portion of the sky obscured by clouds, usually measured in tenths or eighths of the sky. Rainfall, often symbolized by R, represents precipitation water droplets with diameters of 0.5 mm or larger; where the diameters of the droplets are smaller than 0.5 mm are classified as drizzle. Wind speed (W) is defined as the comparison between the distance traveled by air and the time required to travel that distance.

3. RESULTS AND DISCUSSION

3.1 Results of Preprocessing Data

Missing data were imputed using the KNN method with k = 12. It was known that the R variable contained 61 missing values (NA). The mean value of before imputation (μ_1) is 5.487, while the mean value after imputation (μ_2) is 4.828. After the Z-test was conducted, it is known that p-value was 0.2214, which is greater than 0.05; thus, the null hypothesis (H_0) was accepted. This indicated that there is no significant difference between the mean value before and after imputation.

Mahalanobis distance calculations for the Bali climate data, which exceeded the Chi-Square distribution threshold, were used as the basis for outlier identification. The calculated value ranged from 59,8972 to 0.09076, with the highest number recorded at Kembangsari; meanwhile, the lowest value was recorded at Gianyar. Bali's climate data that were identified as outliers reached a total of 175 data points, while the other 2,705 were identified as non-outliers. The areas with the most outlier data were Pengotan, Kembangsari, and Nusa Dua. The calculation result showed that $p-\text{value} = 0$ and $p-\text{value} < 0.05$ Showed that the Mahalanobis distance distribution exceeds the Chi-Square threshold. Therefore, it could be concluded that there were outliers in Bali's climate data.

The last preprocessing step involves standardization of Bali's climate data to ensure that the differences between units and data distribution anomalies do not affect the analysis. Results from this step were shown in Table 2, with the highest maximum value at data standardization located at the R variable, while the lowest maximum value was recorded at the A variable. The highest minimum value is also located at the R variable, and the lowest one is located at the TC variable.

Table 2. Summary of Data Standardization Result

Variable	Monthly Average Thermal Comfort (TC)	Monthly Average Cloud Cover (A)	Monthly Average Rain Rate (R)	Monthly Average Wind Speed (W)
Minimum	-3.1588	-2.541969	-0.9777	-1.6564
Mean	0	0	0	0
Median	0.2357	0.0085	-0.2739	-0.1917
Maximum	2.3967	1.705248	6.1099	4.0580
Standard Deviation	1	1	1	1

3.2 Fourier Basis Functional Transformation

To mathematically clarify the functional data transformation, the discrete time domain of the observations ($t = 1, 2, \dots, 120$ months) was normalized to a periodic interval [0, B]. The Fourier basis expansion was explicitly chosen as it is highly suitable for capturing recurring seasonal patterns in the monthly climate data. Based on the BIC evaluated across 3 to 7 basis functions, the lowest values were mostly observed around 3 and 4, as shown in Fig. 2. However, because a Fourier series naturally consists of pairs of sine and cosine functions, the number of trigonometric basis functions must formally be an even number $(2k)$. Therefore, the optimal number of trigonometric functions was determined to be 4. Furthermore, to account for the overall mean of the data, a constant term (the 0-th basis function, $\phi_0(t) = 1$) is always included in the equation. Adding this baseline constant to the 4 optimal trigonometric functions yields a final total of 5 basis functions for each variable.

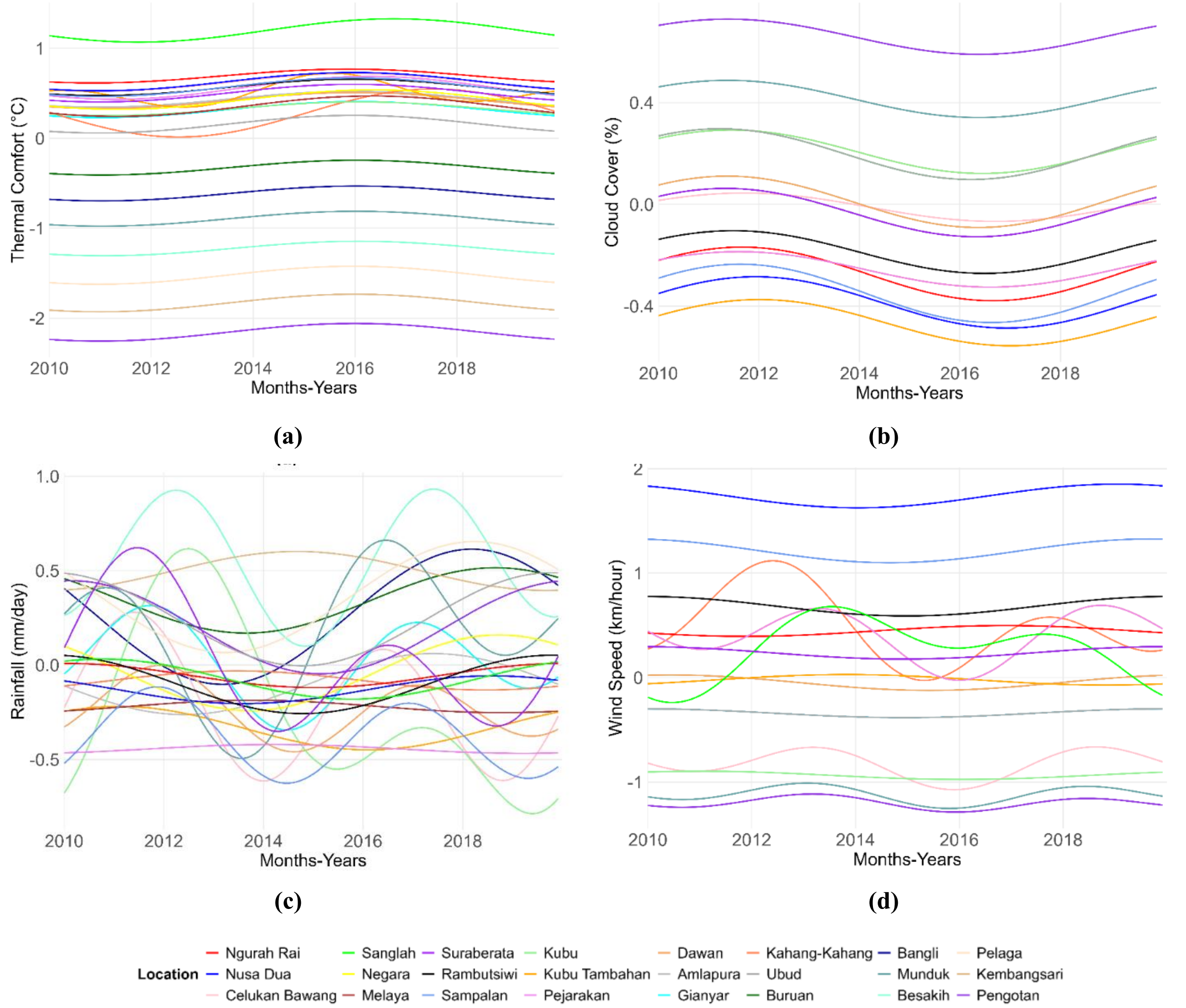


Figure 2. Fourier Basis Functions Transformations Graphs at 24 Locations in Bali from 2010 until 2019 (a) Monthly Average Thermal Comfort (TC), (b) Monthly Average Cloud Cover (A), (c) Monthly Average Rain Rate (R), and (d) Monthly Average Wind Speed (W)

3.3 Basis Function Coefficient Matrix and Spatial Weight Matrix

After acquiring the result of the five basis function transformations for each variable, the basis function coefficients were compiled. The basis function coefficients for each variable were then combined into a basis function coefficient matrix ($\boldsymbol{A}$) as shown in Eq. 23. Matrix $\boldsymbol{A}$ has dimensions of 24×20, representing 24 observation locations and 20 basis coefficients at each location. The values of matrix $\boldsymbol{A}$ are presented in Eq. (23).

$$\boldsymbol{A} = \begin{bmatrix} 7.5774 & -0.3214 & \dots & 0.6797 & -2.9936 & \cdots & -0.5452 \\ 6.8909 & -0.4697 & \dots & 0.7899 & -4.2201 & \cdots & -0.8942 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ -23.6177 & -0.4622 & \cdots & 0.4435 & 7.2180 & \vdots & -0.2369 \end{bmatrix}_{(24\times 20)} . \quad (23)$$

The spatial weight matrix ($\boldsymbol{W}$) was constructed using two distance-based approaches, KNN and Inverse Power Distance (IPD). The KNN spatial weight requires determining the value of k, which can be calculated using the square root of the number of observations in Eq. (24). For $k = 5$, each location is influenced by its five nearest neighbors.

$$k = \sqrt{n} = \sqrt{24} \approx 5. \tag{24}$$

$\boldsymbol{W}_{KNN}$ is as presented in Eq. (25).

$$\boldsymbol{W}_{KNN} = \begin{bmatrix} 0 & 1 & 0 & 1 & 0 & \cdots & 0 \\ 1 & 0 & 0 & 1 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & 0 & 0 & \vdots & 0 \end{bmatrix}_{(24\times24)}. \tag{25}$$

For the IPD spatial weights, this study used $r = 3$ and $r = 4$, which are more suitable for larger datasets ($\geq$ 25 observations). $\boldsymbol{W}_{\text{IPD}\,r3}$ matrix and $\boldsymbol{W}_{\text{IPD}\,r4}$ matrix can be seen in Eqs. (26) and (27).

$$\boldsymbol{W}_{IPD\,r3} = \begin{bmatrix} 0 & 1.9078 \times 10^{-12} & \cdots & 6.5823 \times 10^{-15} \\ 1.9078 \times 10^{-12} & 0 & \cdots & 5.4072 \times 10^{-15} \\ \vdots & \vdots & \ddots & \vdots \\ 6.5823 \times 10^{-15} & 5.4072 \times 10^{-15} & \cdots & 0 \end{bmatrix}_{(24\times24)}, \tag{26}$$

$$\boldsymbol{W}_{IPD\,r4} = \begin{bmatrix} 0 & 2.3663 \times 10^{-16} & \cdots & 1.2336 \times 10^{-19} \\ 2.3663 \times 10^{-16} & 0 & \cdots & 9.4908 \times 10^{-20} \\ \vdots & \vdots & \ddots & \vdots \\ 1.2336 \times 10^{-19} & 9.4908 \times 10^{-20} & \cdots & 0 \end{bmatrix}_{(24\times24)}. \tag{27}$$

3.4 Moran's I Autocorrelation Test

The results of Moran's I autocorrelation in Table 3 show that TC variables and W variables exhibit a significant positive spatial autocorrelation with $\boldsymbol{W}_{KNN}$. This means TC values and W values tend to form a similar pattern in neighbouring locations, which can be interpreted as both TC and W having relatively strong spatial structure within the study area. In contrast, A variables and R variables do not show significant spatial autocorrelation, meaning their distribution tends to be random and is not influenced by spatial proximity.

Table 3. Moran's I Autocorrelation Test Using Spatial Weights $\boldsymbol{W}_{KNN}$, $\boldsymbol{W}_{\text{IPD}\,r3}$, and $\boldsymbol{W}_{\text{IPD}\,r4}$

Spatial Weight	Variable	Moran's I Autocorrelation	p-value	Description ($\alpha = 0.05$)
$\boldsymbol{W}_{KNN}$	TC	0.3466	3.772×10^{-5}	Significantly Positive
	A	-0.0092	0.3661	Not Significant
	R	0.1028	0.07336	Not Significant
	W	0.1251	0.04418	Significantly Positive
$\boldsymbol{W}_{\text{IPD}\,r3}$	TC	0.4060	0.00054	Significantly Positive
	A	0.2472	0.01875	Significantly Positive
	R	0.2115	0.03498	Significantly Positive
	W	0.2287	0.02429	Significantly Positive
$\boldsymbol{W}_{\text{IPD}\,r4}$	TC	0.4734	0.0008938	Significantly Positive
	A	0.3149	0.01652	Significantly Positive
	R	0.2546	0.03912	Significantly Positive
	W	0.2567	0.03534	Significantly Positive

Moran's I autocorrelation test using $\boldsymbol{W}_{\text{IPD}\,r3}$ and $\boldsymbol{W}_{\text{IPD}\,r4}$, produced significant positive spatial autocorrelation across all variables. This indicates that all climate variables exhibit a clear spatial dependence, meaning that the value of a variable at one location is strongly influenced by its neighboring locations. Moreover, Moran's I value for TC variables was higher compared to other spatial weight matrices, further reinforcing the indication that the thermal comfort variable is strongly affected by spatial proximity among locations.

3.5 STPCA-FASTMCD Application

To clarify the unit of analysis and prevent the masking effect caused by classical covariance estimators, outlier detection was formally conducted in two consecutive phases. First, at the observation level, outliers were evaluated per combined location-month observation on the standardized data ($\boldsymbol{X}'$). By using a Mahalanobis Distance, we identified 175 transient outlier data points, which were treated as missing values and imputed via the KNN method. Subsequently, at the location level, after functional transformation, we evaluated outliers per location based on matrix $\boldsymbol{A}$ using the FASTMCD algorithm.

The h subset is measured based on the number of rows and columns in the matrix. It is known that matrix $\boldsymbol{A}$ has 24 rows and 20 columns. If it is determined that $\alpha = 0.75$, the subset size was determined to be $h = 22$. Next, the outlyingness calculation was performed using the Stahel-Donoho index at each location. The highest Stahel-Donoho index values were found at Pengotan and Kembangsari, which were therefore identified as outlier locations. By knowing that the h subset reaches 22, and the 2 locations with the highest SDO values are Pengotan and Kembangsari, thus $\boldsymbol{A}_{robust}$ matrix could be constructed.

$\boldsymbol{A}_{robust}$ matrix dimension is 22×20, which indicates there are 22 locations and 20 basis function coefficients at each location. This shows the dimension of $\boldsymbol{A}_{robust}$ matrix is consistent with h subset requirements at FASTMCD and the outlier locations have been successfully removed. By excluding spatial outliers from the $\boldsymbol{W}$ matrix, thus $\boldsymbol{W}_{KNN\ robust}$, $\boldsymbol{W}_{IPD\ r3\ robust}$, and $\boldsymbol{W}_{IPD\ r4\ robust}$ matrices were also constructed, with each matrix consisting of 22×22 dimensions. As a result, the relationships among the remaining locations became more stable. The spatial patterns obtained are more representative of the actual conditions because they are no longer influenced by locations with extreme characteristics.

After $\boldsymbol{A}_{robust}$ and $\boldsymbol{W}_{robust}$ matrices were obtained, a robust spatio-temporal matrix was constructed. There are 3 types of $\boldsymbol{C}_{robust}$ matrix where each of them has 20×20 dimensions which are shown in Eqs. (28)-(30).

$$\boldsymbol{C}_{KNN\ robust} = \begin{bmatrix} 7{,}1053 & -0{,}05149 & -0{,}03653 & \cdots & -0{,}2659 \\ -0{,}05149 & -0{,}0085 & 0{,}0010 & \cdots & -0{,}0120 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ -0{,}2659 & -0{,}01199 & 0{,}0025 & \cdots & -0{,}0165 \end{bmatrix}_{(20\times 20)}, \tag{28}$$

$$\boldsymbol{C}_{IPD\ r3\ robust} = \begin{bmatrix} 16{,}6568 & -0{,}1276 & 0{,}0082 & \cdots & -0{,}7654 \\ -0{,}1276 & -0{,}0237 & 0{,}0034 & \cdots & -0{,}0360 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ -0{,}7654 & -0{,}0360 & 0{,}0042 & \cdots & -0{,}0212 \end{bmatrix}_{(20\times 20)}, \tag{29}$$

$$\boldsymbol{C}_{IPD\ r4\ robust} = \begin{bmatrix} 18{,}5304 & -0{,}1105 & 0{,}0263 & \cdots & -0{,}8853 \\ -0{,}1105 & -0{,}0251 & 0{,}0028 & \cdots & -0{,}0399 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ -0{,}8853 & -0{,}0399 & 0{,}0030 & \cdots & -0{,}0152 \end{bmatrix}_{(20\times 20)}. \tag{30}$$

3.6 Best Robust Spatial Weight Matrix

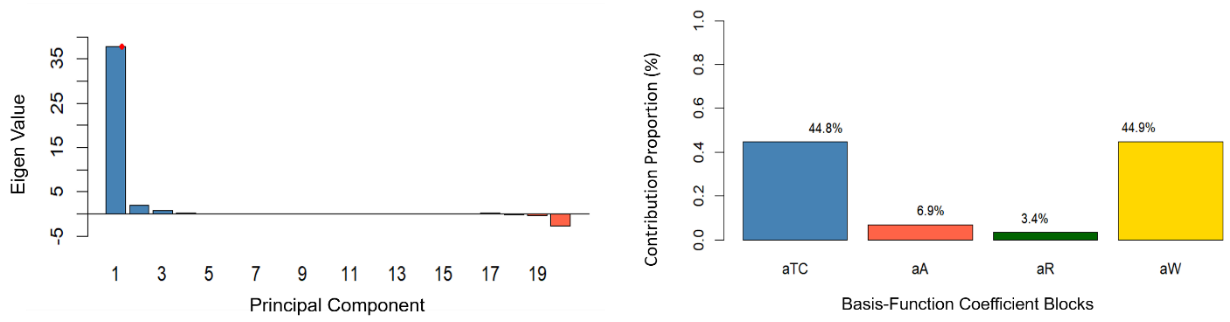
According to $\boldsymbol{C}_{robust}$ matrix, an eigenvalue decomposition was performed to determine the best $\boldsymbol{W}_{robust}$ matrix, which $M^+$and M^- calculation results could be seen at Table 4. It was found that all $\boldsymbol{W}_{robust}$ matrix show very high percentages. The highest M^+value was recorded at $\boldsymbol{W}_{IPD\ r3\ robust}$ with value of 97.3627%, meanwhile the highest M^- value was recorded $\boldsymbol{W}_{KNN\ robust}$ with a value of 91.8384%. However, since the first Spatio-Temporal Principal Component (STPC1) is positive and has the highest eigenvalue and also able to explain the greatest proportion of variance in the data, the best $\boldsymbol{W}_{robust}$ matrix could be seen from the highest M^+ value, which is $\boldsymbol{W}_{IPD\ r3\ robust}$. This finding is consistent with recommendations in the literature, which suggest that IPD is more suitable when the spatial model involves high uncertainty, such as when the number of data pairs is limited [23].

Tabel 4. Comparison of Eigen Value Quality Measures

Matrix $\boldsymbol{W}_{robust}$	M^+ value (%)	M^- value (%)
$\boldsymbol{W}_{KNN\ robust}$	96.885	91.838
$\boldsymbol{W}_{IPD\ r3\ robust}$	97.362	91.151
$\boldsymbol{W}_{IPD\ r4\ robust}$	96.897	90.894

3.7 STPCA-FASTMCD Principal Component

Fig. 3 presents the main outcomes of the eigenvalue decomposition obtained from the selected robust spatial weight matrix. Panel (a) displays the scree plot of STPCA-FASTMCD eigenvalues, while panel (b) shows the relative contributions of the basis-function coefficient blocks (Thermal Comfort, Cloud Cover, Rain Rate, Wind Speed) to the $STPC_1^+$.



(a) (b)

Figure 3. Screen Plot (a) and Contribution of the Basis Function Coefficient Blocks in Positive STPC1 (b) STPCA-FASTMCD

The results of the eigenvalue decomposition of the best $\boldsymbol{W}_{robust}$ matrix were then used to construct the principal components of STPCA-FASTMCD. As can be seen in Fig. 3 (a), the $STPC_1^+$ eigenvalue is 37.787, which is much larger than the other components across all plots. This indicates that a single principal component is sufficient to explain the dominant spatio-temporal variance in the data. The advantage of using the STPCA-FASTMCD method is also evident here, as it is able to capture negative eigenvalues in the final components. Unlike classical PCA, STPCA decomposes $\boldsymbol{C}_{robust}$ containing $\boldsymbol{W}_{robust}$ matrix that is not positive semi-definite, inherently allowing for the mathematical emergence of negative eigenvalues. Consequently, the calculated eigenvalue for $STPC_1^-$ reached -2.822, which could be interpreted as an indicator of the presence of opposite patterns between neighboring locations.

The contribution of the basis function coefficient blocks for $(a)\, STPC_1^+$ in Fig. 3 (b) shows that the aTC and aW blocks contribute almost equally, with percentages of 44.8% and 44.9%, respectively. Meanwhile, the contributions of the aA (6.9%) and aR (3.4%) are relatively small. These findings suggest that the spatio-temporal variability of Bali's climate is primarily influenced by thermal comfort and wind speed. Furthermore, based on the coefficient magnitudes in the equations of $STPC_1^+$ and $STPC_1^-$ it is evident that the TC and W variables are the most decisive in shaping the variability pattern. The complete forms of $STPC_1^+$ and $STPC_1^-$ for location p are shown in Eqs. (31) and (32).

$$\begin{aligned} STPC_{1p}^+ = {} & 0.669\alpha_{p1} + 0.003\alpha_{p2} - 0.001\alpha_{p3} + 0.004\alpha_{p4} + 0.006\alpha_{p5} - 0.263\alpha_{p6} + 0.007\alpha_{p7} - 0.007\alpha_{p8} \\ & + 0.003\alpha_{p9} - 0.006\alpha_{p10} - 0.177\alpha_{p11} + 0.037\alpha_{p12} - 0.032\alpha_{p13} + 0.019\alpha_{p14} \\ & + 0.012\alpha_{p15} + 0.668\alpha_{p16} - 0.027\alpha_{p17} - 0.035\alpha_{p18} - 0.01\alpha_{p19} - 0.022\alpha_{p20}. \end{aligned} \tag{31}$$

$$\begin{aligned} STPC_{1p}^- = {} & 0.494\alpha_{p1} + 0.134\alpha_{p2} - 0.001\alpha_{p3} - 0.024\alpha_{p4} + 0.043\alpha_{p5} - 0.028\alpha_{p6} - 0.357\alpha_{p7} + 0.02\alpha_{p8} \\ & + 0.005\alpha_{p9} - 0.006\alpha_{p10} - 0.007\alpha_{p11} - 0.019\alpha_{p12} - 0.165\alpha_{p13} + 0.019\alpha_{p14} \\ & + 0.036\alpha_{p15} - 0.104\alpha_{p16} + 0.001\alpha_{p17} - 0.028\alpha_{p18} - 0.698\alpha_{p19} - 0.176\alpha_{p20}. \end{aligned} \tag{32}$$

When compared with the HCIU and HCIB equations developed for subtropical regions, significant differences can be observed in the weights of A, R, and W. This demonstrates that the formulas do not fully represent conditions in tropical regions. Therefore, modifications to the HCI equations—particularly adjustments to the weights, ratings, and comfort thresholds—are necessary to make the indices more consistent with tourists' actual perceptions in the field [11].

3.8 Regional Grouping of the STPCA-FASTMCD Climate Comfort Index

The regional grouping of the climate comfort index can be carried out by projecting the scores of the two positive STPCs ($STPC_1^+$ and $STPC_2^+$) and the two negative STPCs ($STPC_1^-$ and $STPC_2^-$). The projection results of $STPC_1^+$ and $STPC_2^+$ are shown in Fig. 4. A clear spatial contrast can be observed between lowland and highland areas. Positive scores on $STPC_1^+$ generally appear in coastal and lowland regions (e.g., Ngurah Rai, Rambutsiwi, Celukan Bawang), characterized by a co-movement between αTC and αW. Conversely, negative scores were mostly found in highland areas (Bangli, Besakih, Ubud), reflecting an opposite direction between the two variables. Meanwhile, for positive $STPC_2^+$, the pattern is more reflective of the dominance of αR and αA, whereas the positive scores indicate that rainfall increases along with cloud cover, while negative scores suggest an inverse relationship between these variables.

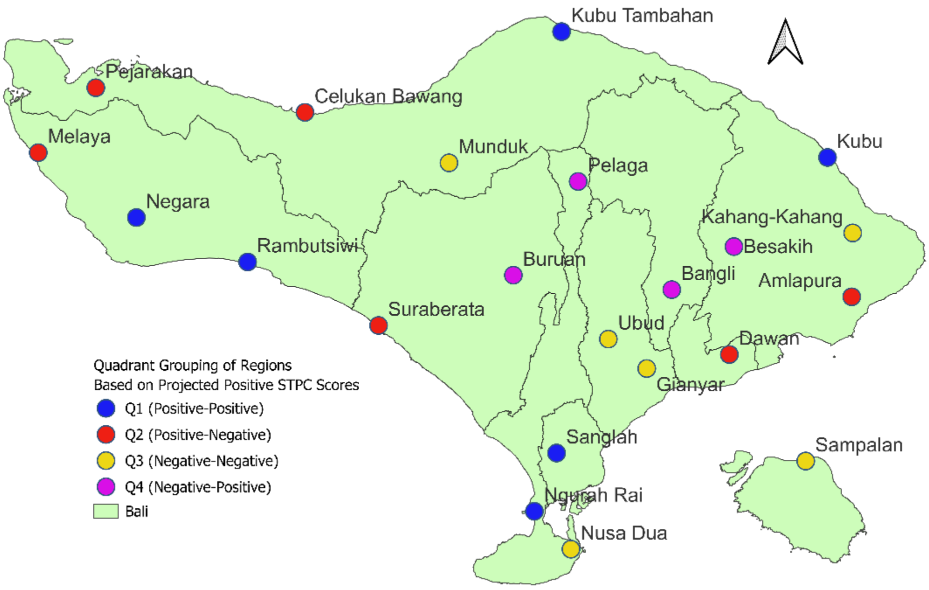


Figure 4. Regions Grouping Based on $STPC_1^+$ and $STPC_2^+$ Scores

The projection results of $STPC_1^-$ and $STPC_2^-$ is shown in Fig. 5, where the presence of microclimate areas can be observed. Positive scores appear in the west and southeast areas of Bali (Melaya, Negara), while combinations of positive and negative were found along the northern and southern coasts. Meanwhile, combinations of negative and positive scores were observed in the central–eastern lowlands (Suraberata, Besakih). In contrast, areas with negative scores were distributed across several locations (Ubud, Munduk, Pelaga, Amlapura), emphasizing that the negative components better represent local climate variations that complement the general patterns in the positive scores.

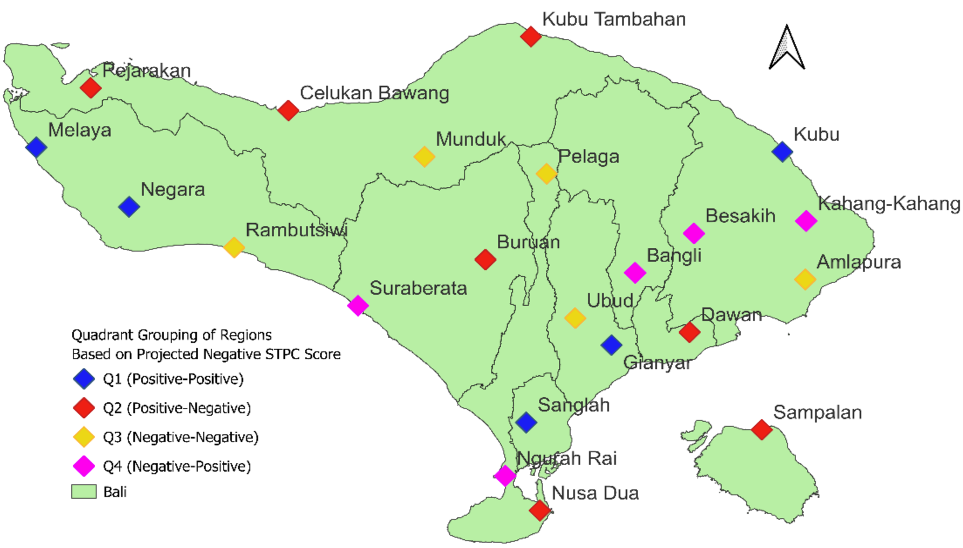


Figure 5. Regions Grouping Based on $STPC_1^-$ and $STPC_2^-$ Scores

Compared to prior climate comfort assessments in Bali, the spatial grouping obtained from the STPCA-FASTMCD method offers a more refined understanding of the spatial structure of climatic comfort. Previous investigations, [28] and [29], consistently demonstrated that comfort levels are strongly influenced by topography and regional orientation, with higher comfort observed in elevated and south–western areas of the island. The present STPCA-FASTMCD analysis refines them by revealing coherent high-comfort grouping across the southern and central highlands and detecting microclimate areas that explain localized variations. Substantively, these findings reveal that rather than broad regional groupings, Bali's climate is characterized by localized microclimate pockets. While macroclimates cover wide areas, these microclimates reflect site-specific variations driven by local topography, vegetation, and surface properties that often shift independently of general regional patterns [30]. These patterns highlight that the robust spatio-temporal approach not only reproduces established comfort gradients but also effectively captures these localized physical dynamics across regions.

These patterns show that the robust spatio-temporal approach not only confirms the general comfort gradients previously identified in Bali but also uncovers more localized climatic differences between regions. Such detailed spatial information is particularly valuable for tourism-oriented planning. Beyond methodological improvements, the findings have clear practical relevance for the tourism sector. Variations in the climate comfort index affect the suitability of outdoor activities, seasonal travel preferences, and overall destination attractiveness, especially in tourism-dependent regions. Changes in comfort levels may influence visitor flows, length of stay, and the timing of peak seasons, ultimately impacting local businesses and tourism revenue [31], [32]. Therefore, combining robust spatio-temporal analysis with climate comfort assessment provides meaningful insights to support tourism management, destination planning, and climate adaptation strategies within the hospitality industry.

Although these findings offer valuable insights for the tourism sector, the accuracy of our regional grouping still depends on a few practical mathematical assumptions. First, we assume that $\boldsymbol{X}'$ is enough to make the different climate variables truly comparable, and the Fourier method effectively captures their natural monthly cycles. Second, we assume that the spatial relationships between locations do not change over the ten-year study period. Finally, we assume that even after removing the extreme outlier locations, the remaining data accurately represent Bali's overall climate dynamics.

Beyond these methodological assumptions, this study also has several practical limitations that open opportunities for further research. The climate comfort index was constructed using the standard variables defined in the holiday climate framework, meaning that the assessment focuses mainly on measurable meteorological conditions. Future research could enrich this approach by incorporating tourist perception data or satisfaction surveys to better capture how visitors actually experience climatic comfort. In addition, the STPCA-FASTMCD method was applied specifically to Bali's tropical climate. Testing this framework in regions with different climatic characteristics would enable meaningful comparisons and help evaluate its broader applicability. Such efforts would further strengthen the relevance and generalizability of robust spatio-temporal climate comfort analysis in tourism studies.

4. CONCLUSION

This study confirms that STPCA-FASTMCD is effective for analyzing climate data that contains outliers. The method works by applying FASTMCD to the basis coefficient matrix ($\boldsymbol{A}$) to identify outlier locations, which are then excluded to form a robust spatio-temporal matrix ($\boldsymbol{C}_{robust}$). The $\boldsymbol{W}_{IPD\ r3\ robust}$ matrix is proven to be the most suitable for Bali's climate data. $STPC_1^+$ explains the largest spatio-temporal variance, with the main influences coming from thermal comfort and wind speed. Regional grouping patterns based on $STPC_1^+$ and $STPC_2^+$ scores highlight contrasts between coastal/lowland and highland areas, while grouping based on $STPC_1^-$ and $STPC_2^-$ reveals microclimates that are significant for the tourism sector. Furthermore, the findings confirm that climate comfort indices developed for subtropical regions are not fully applicable to Bali Province. Therefore, adjustments to the weights and comfort thresholds are necessary to make the index more representative of the local conditions.

Practically, these results allow tourism planners to move beyond general weather forecasts and focus on localized microclimate pockets. This information helps policymakers, tourism planners, and local businesses in Bali in identifying seasonally favorable areas, optimizing tourism calendars, and developing climate-adaptive strategies for sustainable long-term tourism planning. Overall, the STPCA-FASTMCD method provides a solid evidence-based foundation for constructing a locally adaptive climate comfort index to support strategic planning in Bali's tourism sector.

Author Contributions

Agus Yarcana: Formal Analysis, Funding Acquisition, Software, Data Curation, Writing—Review and Editing. Henny Pramoedyo: Methodology, Supervision, Validation, Writing—Review and Editing. Suci Astutik: Methodology, Supervision, Validation, Writing—Review and Editing. All authors discussed the results and contributed to the final manuscript.

Funding Statement

We would like to extend our sincere appreciation to the Indonesian Agency for Meteorology, Climatology, and Geophysics (BMKG) for the financial support provided in this study.

Acknowledgment

Thank you to the Indonesian Agency for Meteorology, Climatology, and Geophysics (Badan Meteorologi, Klimatologi, dan Geofisika or simply BMKG) for providing the climate data that made this research possible.

Declarations

The authors declare no competing interests.

Declaration of Generative AI and AI-assisted Technologies

Generative AI tools (e.g., ChatGPT) were used solely for language refinement, including grammar, spelling, and clarity. The scientific content, analysis, interpretation, and conclusions were developed entirely by the authors. All final text was reviewed and approved by the authors.

REFERENCES

[1] W. K. Härdle and L. Simar, "PRINCIPAL COMPONENTS ANALYSIS," in *Applied Multivariate Statistical Analysis*, Cham: Springer International Publishing, 2019, pp. 299–336. doi: https://doi.org/10.1007/978-3-030-26006-4_11

[2] M. Krzyśko, P. Nijkamp, W. Ratajczak, W. Wołyński, and B. Wenerska, "SPATIO-TEMPORAL PRINCIPAL COMPONENT ANALYSIS," *Spat. Econ. Anal.*, vol. 19, no. 1, pp. 8–29, 2024. doi: https://doi.org/10.1080/17421772.2023.2237532.

[3] S. Stahlschmidt, W. K. Härdle, and H. Thome, "AN APPLICATION OF PRINCIPAL COMPONENT ANALYSIS ON MULTIVARIATE TIME-STATIONARY SPATIO-TEMPORAL DATA," *Spat. Econ. Anal.*, vol. 10, no. 2, pp. 160–180, 2015. doi: https://doi.org/10.1080/17421772.2015.1023339.

[4] F. Janatabadi and A. Ermagun, "ACCESS WEIGHT MATRIX: A PLACE AND MOBILITY INFUSED SPATIAL WEIGHT MATRIX," *Geogr. Anal.*, pp. 746–767, 2024. doi: https://doi.org/10.1111/gean.12395

[5] N. Yu and T. Haskins, "KNN, AN UNDERESTIMATED MODEL FOR REGIONAL RAINFALL FORECASTING," 2021, [Online]. Available: http://arxiv.org/abs/2103.15235

[6] E. Ozelkan, G. Chen, and B. B. Ustundag, "SPATIAL ESTIMATION OF WIND SPEED: A NEW INTEGRATIVE MODEL USING INVERSE DISTANCE WEIGHTING AND POWER LAW," *Int. J. Digit. Earth*, vol. 9, no. 8, pp. 733–747, 2016. doi: https://doi.org/10.1080/17538947.2015.1127437

[7] P. J. Rousseeuw and M. Hubert, "ANOMALY DETECTION BY ROBUST STATISTICS," *Wiley Interdiscip. Rev. Data Min. Knowl. Discov.*, vol. 8, no. 2, pp. 1–14, 2018. doi: https://doi.org/10.1002/widm.1236.

[8] J. T. Abatzoglou, S. Z. Dobrowski, and S. A. Parks, "MULTIVARIATE CLIMATE DEPARTURES HAVE OUTPACED UNIVARIATE CHANGES ACROSS GLOBAL LANDS," *Sci. Rep.*, vol. 10, no. 1, pp. 1–9, 2020. doi: https://doi.org/10.1038/s41598-020-60270-5.

[9] D. I. Purnama and P. R. Sihombing, "PERBANDINGAN ANALISIS KOMPONEN UTAMA DAN ROBUST PCA (ROBPCA)," *J. Bayesian J. Ilm. Stat. dan Ekon.*, vol. 1, no. 1, pp. 67–76, 2021. doi: https://doi.org/10.46306/bay.v1i1.7.

[10] D. Scott, M. Rutty, B. Amelung, and M. Tang, "AN INTER-COMPARISON OF THE HOLIDAY CLIMATE INDEX (HCI) AND THE TOURISM CLIMATE INDEX (TCI) IN EUROPE," *Atmosphere (Basel).*, vol. 7, no. 6, 2016. doi: https://doi.org/10.3390/atmos7060080.

[11] N. M. Hidayat, R. Hidayati, A. Turyanti, and S. F. Al Maula, "MODIFICATION OF THE THERMAL COMFORT INDEX BASED ON PERCEPTIONS FOR URBAN TOURISM AROUND JAKARTA," *J. Meteorol. dan Geofis.*, vol. 25, no. 1, pp. 1–15, 2024. doi: https://doi.org/10.31172/jmg.v25i1.1051.

[12] *The Global Climate 2011-2020*, no. 1338. 2020.

[13] I. G. A. A. Wulandari, "THE COVID-19 PANDEMIC IMPACT ON TOURISM BUSINESS IN KUTA BEACH BALI: A

NATURALISTIC QUALITATIVE STUDY," *Int. J. Tour. Hosp. Asia Pasific*, vol. 6, no. 1, pp. 80–96, 2023. doi: https://doi.org/10.32535/ijthap.v6i1.2192.
[14] M. R. Abdillah *et al.*, "EXTREME WIND VARIABILITY AND WIND MAP DEVELOPMENT IN WESTERN JAVA, INDONESIA," *Int. J. Disaster Risk Sci.*, vol. 13, no. 3, pp. 465–480, 2022. doi: https://doi.org/10.1007/s13753-022-00420-7.
[15] WMO, "WMO GUIDELINES ON THE CALCULATION OF CLIMATE NORMALS," *WMO-No. 1203*, no. 1203, p. 29, 2017, [Online]. Available: https://library.wmo.int/doc_num.php?explnum_id=4166
[16] U. Mawarsari, "IMPUTASI MISSING DATA DENGAN K-NEAREST NEIGHBOR DAN ALGORITMA GENETIKA," *AdMathEdu J. Ilm. Pendidik. Mat. Ilmu Mat. dan Mat. Terap.*, vol. 6, no. 1, 2016. doi: https://doi.org/10.12928/admathedu.v6i1.4764.
[17] H. Ghorbani, "MAHALANOBIS DISTANCE AND ITS APPLICATION FOR DETECTING MULTIVARIATE OUTLIERS," *Facta Univ. Ser. Math. Informatics*, p. 583, 2019. doi: https://doi.org/10.22190/FUMI1903583G.
[18] F. M. Barus and Sutarman, "MENDETEKSI OUTLIER PADA DATA MULTIVARIAT DENGAN METODE JARAK MAHALANOBIS-MINIMUM COVARIANCE DETERMINANT (MMCD)," *IJM Indones. J. Multidiscip.*, vol. 1, no. 3, pp. 1164–1172, 2023.
[19] A. Ambarwari, Q. J. Adrian, and Y. Herdiyeni, "ANALISIS PENGARUH DATA SCALING TERHADAP PERFORMA ALGORITME MACHINE LEARNING UNTUK IDENTIFIKASI TANAMAN," *J. RESTI (Rekayasa Sist. dan Teknol. Informasi)*, vol. 4, no. 1, pp. 117–122, 2020. doi: https://doi.org/10.29207/resti.v4i1.1517
[20] T. Górecki, M. Krzyśko, Ł. Waszak, and W. Wołyński, "SELECTED STATISTICAL METHODS OF DATA ANALYSIS FOR MULTIVARIATE FUNCTIONAL DATA," *Stat. Pap.*, vol. 59, no. 1, pp. 153–182, 2018. doi: https://doi.org/10.1007/s00362-016-0757-8.
[21] F. Martínez, M. P. Frías, M. D. Pérez, and A. J. Rivera, "A METHODOLOGY FOR APPLYING K-NEAREST NEIGHBOR TO TIME SERIES FORECASTING," *Artif. Intell. Rev.*, vol. 52, no. 3, pp. 2019–2037, 2019. doi: https://doi.org/10.1007/s10462-017-9593-z.
[22] M. Masoudi, "ESTIMATION OF THE SPATIAL CLIMATE COMFORT DISTRIBUTION USING TOURISM CLIMATE INDEX (TCI) AND INVERSE DISTANCE WEIGHTING (IDW) (CASE STUDY: FARS PROVINCE, IRAN)," *Arab. J. Geosci.*, vol. 14, no. 5, 2021. doi: https://doi.org/10.1007/s12517-021-06605-6.
[23] U. Barudžija, J. Ivšinović, and T. Malvić, "SELECTION OF THE VALUE OF THE POWER DISTANCE EXPONENT FOR MAPPING WITH THE INVERSE DISTANCE WEIGHTING METHOD—APPLICATION IN SUBSURFACE POROSITY MAPPING, NORTHERN CROATIA NEOGENE," *Geosci.*, vol. 14, no. 6, 2024. doi: https://doi.org/10.3390/geosciences14060155.
[24] J. Fan, W. Wang, and Y. Zhong, "AN ∞ EIGENVECTOR PERTURBATION BOUND AND ITS APPLICATION TO ROBUST COVARIANCE ESTIMATION," *J. Mach. Learn. Res.*, vol. 18, pp. 1–42, 2018, [Online]. Available: https://www.jmlr.org/papers/volume18/16-140/16-140.pdf
[25] N. Fat'Ha and H. T. Sutanto, "IDENTIFIKASI AUTOKORELASI SPASIAL PADA PENGANGGURAN DI JAWA TIMUR MENGGUNAKAN INDEKS MORAN," *MATHunesa J. Ilm. Mat.*, vol. 8, no. 2, pp. 89–92, 2020. doi: https://doi.org/10.26740/mathunesa.v8n2.p89-92.
[26] B. E. Susilowati and P. R. Sihombing, "METODE ROBPCA (ROBUST PRINCIPAL COMPONENT ANALYSIS) DAN CLARA (CLUSTERING LARGE AREA) PADA DATA DENGAN OUTLIER," *J. Ilmu Komput.*, vol. 13, no. 2, p. 11, 2020. doi: https://doi.org/10.24843/JIK.2020.v13.i02.p04.
[27] X. Yao, M. Zhang, Y. Zhang, H. Xiao, and J. Wang, "RESEARCH ON EVALUATION OF CLIMATE COMFORT IN NORTHWEST CHINA UNDER CLIMATE CHANGE," *Sustain.*, vol. 13, no. 18, pp. 1–19, 2021. doi: https://doi.org/10.3390/su131810111.
[28] K. Sumaja, I. K. M. Satriyabawa, S. Maharani, and W. A. Mustika, "THE CLIMATE COMFORT AND RISK ASSESSMENT FOR TOURISM IN BALI, INDONESIA," *Springer Proc. Phys.*, vol. 290, no. July, pp. 545–553, 2023. doi: https://doi.org/10.1007/978-981-19-9768-6_50.
[29] L. R. Z. Dini and Sobirin, "TINGKAT KENYAMANAN IKLIM DI PULAU BALI BERDASARKAN TOURISM CLIMATE INDEX," *Ind. Res. Work. Natl. Semin.*, pp. 678–684, 2017, [Online]. Available: https://jurnal.polban.ac.id/index.php/proceeding/article/view/602/457
[30] J. Kemppinen *et al.*, "MICROCLIMATE, AN IMPORTANT PART OF ECOLOGY AND BIOGEOGRAPHY," no. March, 2024. doi: https://doi.org/10.1111/geb.13834
[31] A. S. Prinsloo and J. M. Fitchett, "QUANTIFYING CLIMATIC SUITABILITY FOR TOURISM IN SOUTHWEST INDIAN OCEAN TROPICAL ISLANDS : APPLYING THE HOLIDAY CLIMATE INDEX TO RÉUNION ISLAND," *Int. J. Biometeorol.*, vol. 68, no. 9, pp. 1717–1728, 2024. doi: https://doi.org/10.1007/s00484-024-02700-x.
[32] X. Huang, Y. Hui, J. Chen, Z. Huang, X. Li, and X. Yang, "Research Progress on the Evaluation of Tourism Climate Comfort and Its Application in China : A Bibliometrics-Based Review," pp. 1–23, 2025. doi: https://doi.org/10.3390/atmos16060714