CAUCHY – Jurnal Matematika Murni dan Aplikasi
Volume 11 (1) (2026), Pages 135-151
p-ISSN: 2086-0382; e-ISSN: 2477-3344

Enhancing Spatio-Temporal PCA with FASTMCD for Climate Comfort Assessment

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

[1] *Department of Statistics, Faculty of Mathematics and Natural Sciences, Brawijaya University, Malang, Indonesia*
[2] *Gusti Ngurah Rai Meteorological Office, Indonesian Agency for Meteorology Climatology and Geophysics, Denpasar, Indonesia*

Abstract

This study presents a robust formulation of the Spatio-Temporal Principal Component Analysis (STPCA) by integrating the Fast Minimum Covariance Determinant (FASTMCD) estimator into the spatio-temporal decomposition framework. Unlike classical STPCA—which constructs the spatio-temporal matrix from sample-based means and is therefore highly sensitive to extreme observations—the proposed STPCA–FASTMCD replaces the classical mean and scatter structure with robust estimates derived from FASTMCD. The method incorporates functional Fourier-based temporal smoothing and an inverse power–distance spatial weight matrix to better capture the underlying spatio-temporal patterns. Monthly climate data (thermal comfort, cloud cover, rainfall, and wind speed) from 24 monitoring locations in Bali during 2010–2019 are analyzed. Performance is evaluated using mean-shift analysis, eigenvalue-stability assessment, and eigenvector perturbation diagnostics. The classical STPCA produces inflated and unstable leading components, with the first eigenvalue reaching 63.36, whereas STPCA–FASTMCD reduces this value to 37.79 and yields smoother, more coherent spatial loading patterns. The robust STPC1 reveals a clear thermal–wind variability mode, enhancing the interpretability of spatial gradients relevant to climate comfort. Overall, the proposed formulation substantially improves the stability and climatic relevance of dominant spatio-temporal modes, providing a more reliable foundation for climate comfort assessment in Bali.

Keywords: Bali Climate; Climate Comfort; Eigenvalue Stability; FASTMCD; Robust Estimation; Spatio-Temporal PCA.

1. Introduction

Global climate change has intensified temperature fluctuations, altered rainfall patterns, and increased atmospheric variability across many regions, including tropical areas such as Indonesia [1]. These changes directly influence climate comfort, a multidimensional concept shaped by various meteorological factors such as air temperature, humidity, rainfall, wind speed, and cloud cover [2]. Assessing climate comfort is increasingly important for climate-sensitive sectors—particularly tourism—because tourists' perception of comfort is closely linked to daily and seasonal weather conditions at a destination [3]. Given these considerations, a standardized climate comfort index

*Corresponding author. E-mail: agus.yarcana@bmkg.go.id

Submitted: October 29, 2025 Reviewed: December 4, 2025 Accepted: December 9, 2025
DOI: https://doi.org/10.18860/cauchy.v11i1.37866

is required to translate complex meteorological conditions into a form that reflects how tourists experience weather.

The Holiday Climate Index (HCI) is a tourism-focused climate comfort metric that integrates key meteorological variables using a weighting and rating system based on tourist preference surveys. It comprises two variants—HCI-Urban (HCIU) and HCI-Beach (HCIB)—which combine thermal comfort (TC), cloud cover (A), rainfall (R), and wind speed (W) with different weighting schemes [3]. Variable TC in this study is calculated from maximum air temperature and relative humidity, representing perceived temperature in degrees Celsius [4]. Because HCI is derived from climate variables that vary across both space and time, its interpretation requires methods capable of summarizing dominant spatio-temporal patterns to support climate comfort analysis.

Bali, as one of the world's major tourist destinations, possesses a humid tropical climate with complex spatial–temporal characteristics. Its climatic variability is strongly shaped by the island's diverse topography, ranging from coastal lowlands to mountainous highlands [5]. This heterogeneity results in climate data that exhibit a clear spatio-temporal structure, where conditions vary simultaneously across locations and over time. Consequently, conventional analytical approaches—which treat observations solely as time series or as independent spatial points—are unable to fully capture the dynamic interaction between space and time.

Principal Component Analysis (PCA) is among the most widely used dimensionality-reduction techniques [6]. However, classical PCA does not account for spatial–temporal dependencies and is highly sensitive to outliers, which frequently arise in tropical climate datasets due to extreme rainfall, strong winds, or measurement anomalies. Spatio-Temporal PCA (STPCA) extends PCA by decomposing variability into dominant spatial–temporal modes [7], but the standard STPCA formulation still relies the sample spatio-temporal matrix, making it unstable when outliers are present. This sensitivity is particularly problematic when HCI variables exhibit extreme values, as outlier-induced distortion may lead to misleading representations of climate comfort patterns. To address this limitation, robust covariance estimators—particularly the Fast Minimum Covariance Determinant (FASTMCD)—offer high breakdown-point performance and are effective for detecting structure in multivariate datasets affected by outliers [8, 9]. Despite its advantages, the integration of FASTMCD into the STPCA framework has not been comprehensively examined, especially in the context of climate data. A robust formulation is therefore essential to ensure that spatio-temporal modes used to interpret HCI are not dominated by extreme meteorological events or measurement anomalies.

The contributions of this study are threefold. First, it develops a robust formulation of Spatio-Temporal PCA (STPCA-FASTMCD) by integrating Fourier-based temporal smoothing and Inverse Power Distance (IPD) spatial weighting, and replacing the classical covariance matrix with the FASTMCD robust estimator within the STPCA operator. Second, it provides a systematic comparison between classical STPCA and the proposed robust STPCA-FASTMCD in terms of mean shifts, eigenvalue stability, and spatial–temporal loading structures. Third, the proposed formulation is applied to Bali's climate dataset (2010–2019) to demonstrate how robust estimation improves the extraction of dominant climate-variability modes that are relevant for interpreting climate comfort patterns. Consequently, the primary novelty of this work lies in the specific integration of the FASTMCD estimator into the STPCA framework, establishing a robust methodology designed to mitigate the impact of extreme outliers inherent in tropical climate data.

2. Methods

This section describes the methodological framework used in this study, beginning with the data structure and proceeding through outlier detection, spatial weighting, functional transformation, and robust spatio-temporal decomposition. The following subsections present each methodological component in detail, highlighting their roles within the proposed STPCA–FASTMCD framework.

2.1. Data

This study utilizes a combined dataset derived from observational records of the Indonesian Agency of Meteorology, Climatology, and Geophysics (BMKG) and reanalysis data from the ERA5 product of the European Centre for Medium-Range Weather Forecasts (ECMWF). The dataset covers 24 climate monitoring locations across Bali Province over a ten-year period from 2010 to 2019, with all variables compiled at monthly intervals.

For four BMKG meteorological stations—Ngurah Rai, Sanglah, Negara, and Kahang-Kahang—most variables were obtained directly from BMKG observations. However, the cloud cover variable (A) for these stations was supplemented using ERA5 ECMWF data due to incomplete local measurements. For the remaining 20 rainfall-monitoring sites, BMKG provides complete rainfall data (R), while ERA5 ECMWF supplies the additional variables, namely thermal comfort (TC), cloud cover (A), and wind speed (W), which are not recorded at these locations. Thermal comfort (TC) is calculated using the following formulation: $TC = 0.8T_{\max} + \frac{RH \times T_{\max}}{500}$ where $T_{\max}$ denotes the monthly maximum air temperature and RH represents relative humidity.

The spatial distribution of the monitoring sites is presented in Table 1. These locations represent Bali's diverse climatic conditions, ranging from coastal lowlands to mountainous regions exceeding 1,000 m above sea level. The use of Universal Transverse Mercator (UTM) coordinates and elevation information allows for the analysis of both horizontal and vertical climate gradients. With 24 spatial units observed over 120 months, the resulting dataset forms a balanced panel suitable for multivariate and spatio-temporal analyses.

Table 1: Climate Observation Sites in Bali

Location	UTM Latitude	UTM Longitude	Elevation (m)
Ngurah Rai	298595.33	9032573.83	3
Nusa Dua	304216.60	9026793.78	9
Celukan Bawang	263192.79	9093603.27	10
Sanglah	301945.73	9041530.96	16
Negara	237477.19	9077321.91	24
Melaya	222383.44	9087187.63	24
Suraberata	274629.03	9060967.36	25
Rambutsiwi	274629.03	9060967.36	28
Sampalan	340041.87	9040465.23	40
Kubu	343203.34	9087047.25	41
Kubu Tambahan	302415.29	9106177.58	56
Pejarakan	231141.33	9097200.90	57
Dawan	328298.18	9056762.29	63
Amlapura	346981.33	9065680.61	95
Gianyar	315677.26	9054558.88	119
Kahang-Kahang	347099.55	9075480.21	141
Ubud	309785.34	9059049.20	218
Buruan	295210.10	9068783.15	386
Bangli	319448.65	9066679.60	544
Munduk	285273.42	9085972.67	698
Besakih	328935.52	9073323.36	834
Pelaga	305057.84	9083207.72	907
Kembangsari	308979.85	9096220.96	1027
Pengotan	319416.65	9081702.48	1155

A summary of the climate and comfort-related variables employed in this study is provided in Table 2.

Table 2: Research Variables

Variable	Symbol	Unit
Monthly Average Thermal Comfort	x_1	Celsius
Monthly Average Cloud Cover	x_2	Percent
Monthly Average Rain Rate	x_3	Millimeter per day
Monthly Average Wind Speed	x_4	Kilometer per hour

2.2. Outlier Identification

One of the most common methods for detecting multivariate outliers is the Mahalanobis distance (MD) [10]. For the i-th observation, the squared Mahalanobis distance is computed as

$$d_i^2 = (\mathbf{x}_i - \boldsymbol{\mu})' \boldsymbol{\Sigma}^{-1} (\mathbf{x}_i - \boldsymbol{\mu}),$$

where $\mathbf{x}_i$ denotes the multivariate observation vector for index i, $\boldsymbol{\mu}$ is the mean vector, and $\boldsymbol{\Sigma}$ is the covariance matrix.

The hypothesis tested is H_0: d_i^2 follows a χ^2 distribution with p degrees of freedom (no multivariate outliers) versus H_1: d_i^2 does not follow a χ^2 distribution (presence of outliers), where p denotes the number of variables. An observation is classified as an outlier if d_i^2 exceeds the chi-square threshold $\chi^2_{(p,0.05)}$.

2.3. Spatial Weight Matrix Construction and Moran's I

The choice of the spatial weight matrix ($\mathbf{W}$) plays a crucial role in determining the number of principal components produced by STPCA. This matrix has an order of $n \times n$, where n denotes the number of observation locations. One commonly used spatial weighting scheme is the Inverse Power Distance (IPD). The IPD method assigns weights based on distance, giving stronger spatial influence to closer locations and weaker influence to those farther away. IPD weights follow an inverse-power relationship with distance, as shown in Eq. (1).

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

with d_{ij} denoting the distance between locations i and j, and r representing the distance-decay parameter. Previous studies recommend $r = 2$ for relatively small spatial samples (≤ 19 sites) and $r = 3$–4 for larger networks (≥ 25 sites) [11]. Because this study involves 24 locations—positioned between these two regimes—the parameter $r = 3$ is adopted as an intermediate choice that provides adequate spatial differentiation while avoiding excessive emphasis on localized effects.

Spatial autocorrelation can be assessed through both global and local measures, which indicate how strongly values of a variable are spatially clustered or dispersed. Global measures describe overall spatial patterns, while local measures examine the relationship between each unit and its neighbors [12]. A widely used measure suitable for STPCA is Moran's I, expressed in Eq. (2) with standardized spatial weights. Moran's I ranges from -1 (perfect negative autocorrelation) to 1 (perfect positive autocorrelation). Positive values indicate spatial clustering, whereas negative values suggest spatial dispersion [13].

$$I(X) = \frac{n}{\sum_{i=1}^n \sum_{j=1}^n w_{ij}} \frac{\sum_{i=1}^n \sum_{j=1}^n w_{ij}(X_i - \bar{X})(X_j - \bar{X})}{\sum_{i=1}^n (X_i - \bar{X})^2}, \tag{2}$$

where X_i is the observed value at location i, $\bar{X}$ denotes the mean of X, and w_{ij} represents the spatial weight between locations i and j.

2.4. Functional Data Transformation (Fourier Basis)

Functional data refer to smoothed realizations of random sample observations collected over space and time [14]. These data take the form of curves, requiring a specialized framework known

as functional data analysis. Its objectives include generating new functional representations, identifying variability in data characteristics, exploring temporal patterns, and estimating the variation in response variables based on predictor information. Transforming temporal data into functional form involves projecting observations onto a functional space defined by orthonormal bases, such as Fourier, spline, or wavelet functions, resulting in multivariate functional data (MFD) where each variable is expressed as a continuous function over time as shown in Eq. (3).

$$y_k(t) = \sum_{b=0}^{B} \hat{\alpha}_{kb}\, \phi_b(t), \tag{3}$$

where $y_k(t)$ denotes the k-th variable as a function of time t, $\hat{\alpha}_{kb}$ is the estimated coefficient of the k-th variable associated with the b-th basis function, $\phi_b(t)$ represents the b-th Fourier basis function, and B is the number of significant basis functions.

For climate-related data, the Fourier basis is particularly suitable because it decomposes discrete and potentially complex temporal patterns into simple sinusoidal components [15]. This transformation is effective for meteorological and climatological variables that exhibit strong periodic behavior, such as seasonal rainfall and dry-season patterns [16]. The Fourier basis consists of three orthonormal components as shown in Eq. (4):

$$\phi_b(t) = \begin{cases} \phi_0(t) = \dfrac{1}{\sqrt{T}}, \\ \phi_{2b-1}(t) = \sqrt{\dfrac{2}{T}} \sin\left(\dfrac{2\pi b t}{T}\right), \\ \phi_{2b}(t) = \sqrt{\dfrac{2}{T}} \cos\left(\dfrac{2\pi b t}{T}\right), \end{cases} \tag{4}$$

where b denotes the basis index and T represents the total time period.

The number of basis functions B directly influences the smoothness of the functional data, where smaller values of B produce smoother curves. The optimal number of basis functions B (B_k) can be selected by evaluating and comparing the Bayesian Information Criterion (BIC). The BIC is defined as shown in Eq. (5):

$$\text{BIC} = \ln\left(\sum_{k=0}^{K}\left(y_k - \sum_{b=0}^{B} \hat{\alpha}_{kb}\, \phi_b(t)\right)^2\right) + (B+1)\left(\frac{\ln K}{K}\right), \tag{5}$$

This transformation produces a set of basis-function coefficient vectors ($\boldsymbol{\alpha}$), which are then assembled into the matrix $\mathbf{A} = (\boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2, \boldsymbol{\alpha}_3, \ldots, \boldsymbol{\alpha}_p)$. The matrix $\mathbf{A}$ serves as the input for constructing the spatio-temporal matrix.

2.5. Spatio-Temporal PCA

The extension of PCA to more complex data structures, particularly those exhibiting spatio-temporal characteristics, underlies the development of the STPCA framework. This methodology is intended to more effectively characterize the relationship between spatial variation across geographic locations and the corresponding temporal dynamics of the data [17]. To accomplish this, STPCA constructs a matrix that encodes both types of dependencies and serves as the basis for the eigenvector decomposition. Based on this formulation, the spatio-temporal matrix $\mathbf{C}$ is obtained from the joint processing of the functional coefficient matrix $\mathbf{A}$ and the spatial weight matrix $\mathbf{W}$, and is expressed as shown in Eq. (6).

$$\mathbf{C} = \left(\sum_{i=1}^{n}\sum_{j=1}^{n} w_{ij}\right)^{-1} \mathbf{A}'\mathbf{W}\mathbf{A}, \tag{6}$$

The matrix $\mathbf{C}$ in Eq. (6) is symmetric ($\mathbf{C} = \mathbf{C}^{\mathsf{T}}$) but may be indefinite due to the influence of spatial weights and cross-location interactions. For this reason, $\mathbf{C}$ is more appropriately interpreted as a spatio-temporal matrix rather than a conventional covariance matrix, because true covariance matrices must be positive semi-definite with non-negative diagonal entries [18]. Despite this property, the eigendecomposition of $\mathbf{C}$ is well defined, and the corresponding eigenvalue problem is expressed as Eq. (7):

$$\mathbf{C}\boldsymbol{\omega}_q - \gamma_q \boldsymbol{\omega}_q = \mathbf{0}, \tag{7}$$

Positive eigenvalues ($\gamma_q > 0$) correspond to positive principal directions, whereas negative eigenvalues ($\gamma_q < 0$) generate negative principal directions. Together, positive and negative eigenvalues characterize opposing modes of spatio-temporal variation. A positive eigenvector corresponds to a principal direction in which locations with high loadings move in the same dominant pattern, whereas a negative eigenvector reflects an inverse or contrasting pattern relative to that direction. These opposing structures indicate how different regions may express the same spatio-temporal mode with either aligned or reversed intensities, making the decomposition useful for distinguishing complementary spatio-temporal behaviors in the data.

Within the STPCA framework, each eigenvector $\boldsymbol{\omega}_q$ yields a pair of loading vectors—the positive eigenvector $\boldsymbol{\omega}_q^+$ and the negative eigenvector $\boldsymbol{\omega}_q^-$—derived from the respective positive and negative eigenvalues. These loadings are used to compute the spatio-temporal principal component scores at each location. For the q-th component, the positive score at location p is computed as shown in Eq. (8):

$$\mathrm{STPC}_{pq}^{+} = (\boldsymbol{\omega}_q^{+})'\boldsymbol{\alpha}_p = \omega_{q1}^{+}\alpha_{p1} + \omega_{q2}^{+}\alpha_{p2} + \cdots + \omega_{qk}^{+}\alpha_{pk}, \tag{8}$$

and the negative component as shown in Eq. (9):

$$\mathrm{STPC}_{pq}^{-} = (\boldsymbol{\omega}_q^{-})'\boldsymbol{\alpha}_p = \omega_{q1}^{-}\alpha_{p1} + \omega_{q2}^{-}\alpha_{p2} + \cdots + \omega_{qk}^{-}\alpha_{pk}. \tag{9}$$

Here, $\boldsymbol{\alpha}_p$ denotes the basis-coefficient vector at location p, while $\boldsymbol{\omega}_q^+$ and $\boldsymbol{\omega}_q^-$ represent the positive and negative spatio-temporal eigenvectors that capture the underlying spatio-temporal patterns. The resulting inner products quantify the extent to which each pattern is manifested at a given location.

where:

STPC_{pq}^{+}	: positive spatio-temporal principal component score of order q at location p;
STPC_{pq}^{-}	: negative spatio-temporal principal component score of order q at location p;
ω_{qb}^{+}	: positive spatio-temporal eigenvector of order q for the b-th basis;
ω_{qb}^{-}	: negative spatio-temporal eigenvector of order q for the b-th basis;
α_{pb}	: basis-function coefficient at location p for the b-th basis.

2.6. Robust Covariance Estimation Using FAST-MCD

The Minimum Covariance Determinant (MCD) estimator provides robust estimates of the mean and covariance by selecting a subset of size h as shown in Eq. (10):

$$h = \frac{n + p + 1}{2}, \tag{10}$$

whose covariance determinant is minimal. The robust estimates are computed as shown in Eq. (11) and Eq. (12):

$$\bar{\mathbf{M}}_{\mathrm{MCD}} = \frac{1}{h}\sum_{i=1}^{h}\mathbf{X}_i, \tag{11}$$

$$\mathbf{S}_{\mathrm{MCD}} = \frac{1}{h-1}\sum_{i=1}^{h}\left(\mathbf{X}_i - \bar{\mathbf{M}}_{\mathrm{MCD}}\right)\left(\mathbf{X}_i - \bar{\mathbf{M}}_{\mathrm{MCD}}\right)', \tag{12}$$

Because MCD is resistant to outliers, the resulting eigenvalues and eigenvectors remain stable even when extreme observations are present. Prior studies [19, 20] confirm that PCA based on MCD yields more reliable component scores than classical PCA when data are contaminated. The main limitation of MCD is its computational burden, which led Rousseeuw and Van Driessen [21] to introduce the faster FASTMCD algorithm. FASTMCD determines a subset using α between 0.5 and 1, with

$$h = \max\left(\alpha n, \left\lfloor \frac{n+p+1}{2} \right\rfloor\right), \tag{13}$$

Observations with the smallest Stahel–Donoho outlyingness values are selected to form the FASTMCD subset h, from which robust estimates of the mean and covariance are obtained. These robust estimates are then incorporated into the formation of the robust spatio-temporal matrix, which replaces the classical spatio-temporal matrix $\mathbf{C}$ in the eigendecomposition step. This ensures that the resulting spatio-temporal principal components are less influenced by outliers and more accurately capture the underlying spatio-temporal structure of the data.

2.7. Eigenvector Perturbation

The eigenstructure of STPCA depends directly on the spatio-temporal matrix, making it highly sensitive to extreme observations. When FASTMCD is incorporated into the STPCA framework, the classical spatio-temporal matrix is replaced with its robust version, $\mathbf{C}_{\text{robust}}$, which is computed from a robust mean and a reduced subset of observations. This shift in spatio-temporal structure results in perturbed eigenvalues (γ) and eigenvectors ($\boldsymbol{\omega}$).

Based on eigenvector perturbation theory [22], the effect of a small change in the covariance matrix ($\boldsymbol{\Sigma}$) can be described through the following relation:

$$(\boldsymbol{\Sigma} + \delta\boldsymbol{\Sigma})(\boldsymbol{\omega} + \delta\boldsymbol{\omega}) = (\gamma + \delta\gamma)(\boldsymbol{\omega} + \delta\boldsymbol{\omega}), \tag{14}$$

Eq. (14) serves as the fundamental identity from which first-order perturbation expressions are derived. The leading approximation of the eigenvalue perturbation is obtained by projecting $\delta\boldsymbol{\Sigma}$ onto the direction of the eigenvector as shown in Eq. (15):

$$\delta\gamma \approx \boldsymbol{\omega}'(\delta\boldsymbol{\Sigma})\boldsymbol{\omega}, \tag{15}$$

This expression shows that the alignment between the perturbation matrix $\delta\boldsymbol{\Sigma}$ and the eigenvector $\boldsymbol{\omega}$ determines the magnitude of the perturbation in the eigenvalue. Because FASTMCD modifies the underlying matrix only through robustification of location and scatter, the resulting $\delta\boldsymbol{\Sigma}$ is comparatively small. As a result, the eigenvalues and eigenvectors obtained from the STPCA–FASTMCD procedure exhibit improved stability in the presence of outliers and provide a more reliable representation of the underlying spatio-temporal structure.

3. Results and Discussion

This section presents and interprets the empirical findings obtained from the proposed methodology. The following subsections report the results step by step, beginning with the robust spatio-temporal decomposition and continuing with supporting analyses that explain the observed spatial and temporal patterns.

3.1. Spatio-Temporal PCA-FASTMCD

The integration of FASTMCD begins by determining the subset size h using Eq. (13). According to [8], the subset must satisfy $p < h \leq n$, where n is the number of observations and p is the number of variables. Using this subset, the matrices $\mathbf{A}_{\text{robust}}$ and $\mathbf{W}_{\text{robust}}$ are constructed, leading to the formulation of the robust spatio-temporal matrix as shown in Eq. (16):

$$\mathbf{C}_{\text{robust}} = \left(\sum_{i=1}^{n}\sum_{j=1}^{n} w_{ij}\right)^{-1} \mathbf{A}'_{\text{robust}}\mathbf{W}_{\text{robust}}\mathbf{A}_{\text{robust}}. \tag{16}$$

The matrix $\mathbf{C}_{\text{robust}}$ is then decomposed through its eigenvalue problem as shown in Eq. (17):

$$\mathbf{C}_{\text{robust}}\,\boldsymbol{\omega}_q^{(\text{rob})} - \gamma_q^{(\text{rob})}\,\boldsymbol{\omega}_q^{(\text{rob})} = \mathbf{0}. \tag{17}$$

The resulting eigenvalues and eigenvectors form the basis for computing the robust spatio-temporal principal components. The use of FASTMCD ensures that the eigenstructure is not distorted by outliers, resulting in more stable and representative components.

To evaluate how the robust spatio-temporal matrix modifies the eigenstructure, the eigenvector perturbation framework is applied. The perturbation form of the eigenvector equation is written as Eq. (18):

$$(\mathbf{C} + \delta\mathbf{C})(\boldsymbol{\omega}_q + \delta\boldsymbol{\omega}_q) = (\gamma_q + \delta\gamma_q)(\boldsymbol{\omega}_q + \delta\boldsymbol{\omega}_q). \tag{18}$$

Multiplying both sides by $\boldsymbol{\omega}_q'$ gives:

$$\boldsymbol{\omega}_q'(\mathbf{C} + \delta\mathbf{C})(\boldsymbol{\omega}_q + \delta\boldsymbol{\omega}_q) = (\gamma_q + \delta\gamma_q)(\boldsymbol{\omega}_q'\boldsymbol{\omega}_q + \boldsymbol{\omega}_q'\delta\boldsymbol{\omega}_q).$$

Because $\boldsymbol{\omega}_q'\boldsymbol{\omega}_q = 1$ and $\boldsymbol{\omega}_q'\delta\boldsymbol{\omega}_q = 0$, the expression simplifies to:

$$\boldsymbol{\omega}_q'\mathbf{C}\boldsymbol{\omega}_q + \boldsymbol{\omega}_q'\mathbf{C}\delta\boldsymbol{\omega}_q + \boldsymbol{\omega}_q'\delta\mathbf{C}\boldsymbol{\omega}_q + \boldsymbol{\omega}_q'\delta\mathbf{C}\delta\boldsymbol{\omega}_q = \gamma_q + \delta\gamma.$$

Ignoring second-order terms $\boldsymbol{\omega}_q'\delta\mathbf{C}\delta\boldsymbol{\omega}_q$ and using $\boldsymbol{\omega}_q'\mathbf{C}\boldsymbol{\omega}_q = \gamma_q$, the first-order perturbation yields:

$$\delta\gamma = \boldsymbol{\omega}_q'\delta\mathbf{C}\boldsymbol{\omega}_q.$$

This becomes, in terms of the robust and classical means:

$$\delta\gamma = \frac{1}{h-1}\sum_{i=1}^{h}\left(\boldsymbol{\omega}_q'(\mathbf{X}_i - \bar{\mathbf{M}}_{\text{MCD}})\right)^2 - \frac{1}{n-1}\sum_{i=1}^{n}\left(\boldsymbol{\omega}_q'(\mathbf{X}_i - \bar{\mathbf{X}})\right)^2.$$

This expression shows that robust eigenvalues depend on the robust mean $\bar{\mathbf{M}}_{\text{MCD}}$, which is unaffected by outliers, while the classical mean $\bar{\mathbf{X}}$ may shift toward extreme observations. As a result, robust eigenvalues may increase when the classical mean suppresses core variability, or decrease when the classical mean artificially inflates variability due to outliers. In all cases, robust estimation provides a more accurate reflection of the central data structure.

For FASTMCD to be properly integrated in this study, the data must satisfy several basic requirements. Specifically, the number of observations should exceed the number of variables ($n > p$), the variables should be continuous and numerical without perfect multicollinearity, and the proportion of outliers should not be so large that it dominates the sample and undermines the stability of the robust subset selection.

3.2. Outlier Identification of Climate Data

The MD values for the Bali climate dataset range from 0.0908 to 59.8972. The highest distance was observed in Kembangsari, while the lowest occurred in Gianyar. Based on the chi-square criterion at a significance level of 0.05, a total of 175 observations were classified as outliers, and 2,705 observations were identified as non-outliers.

The locations with the largest number of flagged outliers were Pengotan, Kembangsari, and Nusa Dua. The MD calculation for the climate data is reported in Table 3. The resulting p-value of 2.1140×10^{-10}, which is well below the 0.05 threshold, indicates that the MD distribution exceeds the expected chi-square boundary. This confirms the presence of multivariate outliers in the Bali climate dataset.

Table 3: Result of Outlier Identification

Outlier Detection Result	**Value**
Maximum Mahalanobis distance	59.8972
Minimum Mahalanobis distance	0.0908
Number of detected outliers	175
Locations with most outliers	Pengotan, Kembangsari, and Nusa Dua
Significance level (α)	0.05
p-value	2.1140×10^{-10}

To further assess the robustness of this identification, the Mahalanobis-based outlier counts were compared with the Stahel–Donoho Outlyingness (SDO) index. The scatter plot (Fig. 1) shows that locations with high MD-based outlier frequency generally exhibit elevated SDO values, although several stations display disproportionately large SDO indices relative to their MD counts. This divergence suggests that the classical MD may underestimate outlyingness in the presence of heavy-tailed or skewed data, thereby reinforcing the need for robust estimation methods in subsequent analyses. Additionally, the same key outlying locations identified by the MD also appear among the highest SDO values, although the robust measure flags a slightly broader set of anomalous stations. For the purposes of the STPCA–FASTMCD analysis, the final classification of outlying locations is determined using the SDO criterion, as it provides a more stable and reliable assessment under heavy-tailed spatio-temporal variability.

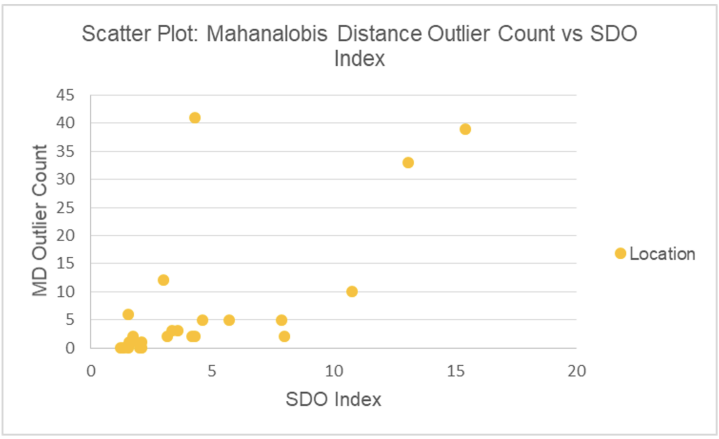


Fig. 1: Scatter Plot Mahanalobis Distance Outlier Count vs SDO Index

3.3. Result of Spatial Weight Matrix Construction and Moran's I

The spatial weights in this study were constructed using the IPD method with a distance-decay parameter of $r = 3$, resulting in the spatial weight matrix $\mathbf{W}_{\text{IPD}r3}$ as defined in Eq. (19):

$$\mathbf{W}_{\text{IPD}r3} = \begin{bmatrix} 0 & 5.0297 \times 10^{-2} & 7.5062 \times 10^{-5} & \cdots & 1.7353 \times 10^{-4} \\ 5.0297 \times 10^{-2} & 0 & 5.4708 \times 10^{-5} & \cdots & 1.4255 \times 10^{-4} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1.7353 \times 10^{-4} & 1.4255 \times 10^{-4} & 1.3889 \times 10^{-4} & \cdots & 0 \end{bmatrix}_{24\times 24}. \quad (19)$$

Spatial dependence was then evaluated using Moran's I for four climatic variables under the weight configuration $\mathbf{W}_{\text{IPD}r3}$. Moran's I assesses whether measurements at nearby locations tend to be similar (indicating spatial clustering) or whether they follow a random pattern. The results are summarized in Table 4.

Table 4: Result of Moran's I

Variable	**Moran's I**	**p value**	**Interpretation ($\alpha = 0.05$)**
TC	0.4060	0.00054	Positive Significant
A	0.2472	0.01875	Positive Significant
R	0.2115	0.03498	Positive Significant
W	0.2287	0.02429	Positive Significant

All variables show significant positive spatial autocorrelation when analyzed using $\mathbf{W}_{\text{IPD}r3}$. This indicates that observations at one location tend to resemble those of neighboring locations, confirming the presence of non-random spatial structure across all variables. Among the four variables, TC shows the highest Moran's I value, indicating the strongest spatial clustering

pattern. This suggests that the spatial process underlying TC varies more smoothly across the region, with clear grouping of high or low values. In contrast, variables with lower Moran's I values display weaker, yet still significant, spatial dependence, indicating more localized spatial variation.

3.4. Result of Functional Data Transformation (Fourier Basis)

The transformation of temporal data into functional form was conducted using several candidate numbers of basis functions, followed by the computation of the BIC for each option. The optimal number of basis functions was selected based on the minimum BIC. In this study, the number of basis functions evaluated ranged from 3 to 7, guided by the reference value of four basis functions suggested by [7].

Once the candidate basis sizes were specified, five functional transformations were generated for each variable. The BIC results indicate that the lowest values predominantly occurred for basis sizes of 3 and 4. A summary of the frequency of minimum BIC values across variables is presented in Table 5. Overall, a basis size of 3 most frequently produced the lowest BIC, although variables R and W include several locations where a basis size of 4 was optimal.

Table 5: Summary of Basis Function Selection by Minimum BIC

Variable	Basis Size (k)	Frequency
TC	3	24
A	3	24
R	3	16
R	4	8
W	3	15
W	4	9

After identifying 4 basis functions as the optimal number, the corresponding basis coefficients were constructed. Although $k = 3$ produced the lowest BIC most frequently, a uniform basis size of 4 was ultimately chosen for all variables to ensure complete Fourier harmonic pairs (sine–cosine) and to maintain consistency across variables within the STPCA framework. Under the Fourier system, selecting $k = 4$ yields five basis functions (a constant term plus two sine–cosine pairs), which explains why matrix $\mathbf{A}$ contains 20 basis coefficients per location (4 variables $\times$ 5 basis functions). The coefficients for each variable were then combined into the basis coefficient matrix $\mathbf{A}$ according to the prescribed formulation. Matrix $\mathbf{A}$ has dimensions 24×20, representing 24 observation locations and 20 basis coefficients at each location. A sample portion of matrix $\mathbf{A}$ is shown in Eq. (20):

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

3.5. Spatio-Temporal Matrix Classic and Robust

To evaluate how robust estimation affects the spatio–temporal structure, the classical spatio-temporal matrix $\mathbf{C}$, which serves as the core matrix in the construction of STPCA, is compared with its robust counterpart $\mathbf{C}_{\text{robust}}$, which is used in STPCA-FASTMCD. The classical version uses the sample mean, whereas the robust version applies the FASTMCD estimator to both spatial and temporal components. In this analysis, the number of Fourier basis functions was fixed at $k = 4$, following the rationale described in the Functional Data Transformation results, where a complete sine–cosine harmonic pairing was required for a balanced Fourier representation. The formulation of matrix $\mathbf{C}$ is presented in Eq. (21).

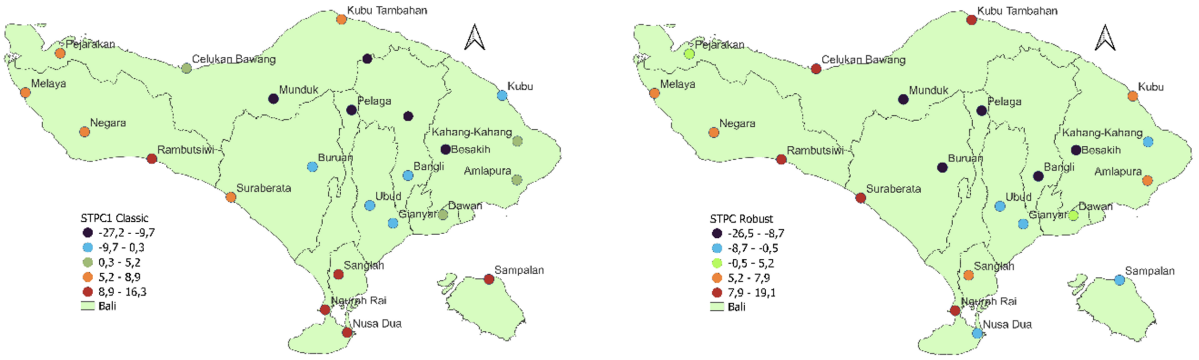
$$\mathbf{C} = \begin{bmatrix} 37.4865 & -2.2613 & -2.8612 & \cdots & -3.6638 \\ 2.7162 & 0.4433 & -0.1595 & \cdots & 0.4586 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ -3.6638 & 0.4586 & 0.4742 & \cdots & 0.5142 \end{bmatrix}_{20\times 20}. \tag{21}$$

The determination of subset h was carried out to identify the most representative portion of the observations as part of the STPCA-FASTMCD procedure. The subset size was based on the number of rows and columns of matrix $\mathbf{A}$, which contains 24 rows and 20 columns. A breakdown value of $\alpha = 0.75$ was used because it is a standard choice in FASTMCD applications, offering a balance between robustness to outliers and statistical efficiency. Using a constant $\alpha = 0.75$, the resulting subset size was $h = 22$. The Stahel–Donoho outlyingness index was then computed for matrix $\mathbf{A}$, with Pengotan and Kembangsari showing the highest values and therefore identified as outliers. Given $h = 22$ and the identification of these two locations as the most outlying, the construction of the robust matrices $\mathbf{A}_{\text{robust}}$ and $\mathbf{W}_{\text{robust}}$ can subsequently be performed within the STPCA-FASTMCD framework. The formulation of matrix $\mathbf{C}_{\text{robust}}$ is presented in Eq. (22).

$$\mathbf{C}_{\text{robust}} = \begin{bmatrix} 16.6568 & -0.1276 & 0.0082 & \cdots & -0.7654 \\ -0.1276 & -0.0237 & 0.0034 & \cdots & -0.0360 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ -0.7654 & -0.0360 & 0.0042 & \cdots & -0.0212 \end{bmatrix}_{20\times 20}. \tag{22}$$

The classical spatio-temporal matrix $\mathbf{C}$ contains substantially larger diagonal values (for example, 37.4865; 0.4433; 0.4742). These inflated diagonal contributions indicate that several extreme observations pull the classical mean toward outlying directions. In contrast, the diagonal elements of $\mathbf{C}_{\text{robust}}$ (for example, 16.6568; –0.0237; 0.0042) are much smaller because they represent variation around the FASTMCD mean, which is resistant to outliers. The off-diagonal elements also decrease in magnitude, showing that the robust procedure reduces spurious contributions to the spatial–temporal matrix.

Overall, the contrast between $\mathbf{C}$ and $\mathbf{C}_{\text{robust}}$ shows that outliers strongly distort the classical spatio-temporal structure, whereas FASTMCD stabilizes it and produces more reliable principal components. To further illustrate this difference, a spatial comparison of the STPC1 patterns provides clear evidence of how outliers influence the resulting spatial structure, as shown in Fig. 2.



(a) Distribution of STPC1 Classic **(b)** Distribution of STPC1 Robust

Fig. 2: Comparison of STPC1 distributions between classical and robust approaches.

In the distribution of STPC1, the absence of negative STPC1 values in Nusa Dua and Sampalan arises from the combined effects of the spatial weight structure and the influence of outliers. Under the $\mathbf{W}_{\text{IPD}r3}$ configuration, spatial weights decrease rapidly with distance, causing central and inland stations with dense spatial connectivity to exert greater influence on the component

structure than the more isolated coastal sites. At the same time, the classical spatio-temporal matrix **C** is affected by outlying highland stations such as Pengotan and Kembangsari, whose inflated diagonal contributions distort the eigenvector orientation toward patterns dominated by inland variability. As a result, moderate or contrasting spatial contributions from coastal locations become suppressed in the classical STPC1 representation.

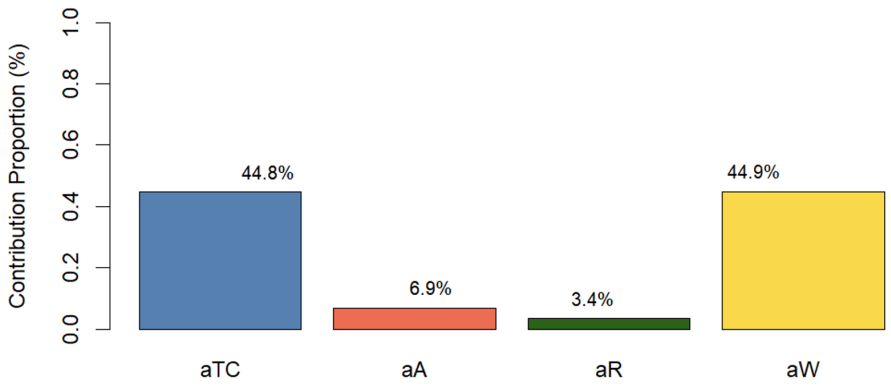


Fig. 3: Block Contribution of Climate Variables to STPC1 Robust

The robust formulation mitigates these effects by reducing the influence of extreme observations and stabilizing the underlying spatio-temporal structure, which allows coastal variability to re-emerge in the STPC1 pattern. To further interpret the physical meaning of this component, the contribution of each climate variable to STPC1 Robust was evaluated using block contribution aggregation of the matrix A. As shown in Figure 3, thermal comfort (44.8%) and wind speed (44.9%) dominate the structure of STPC1, whereas humidity (6.9%) and rainfall (3.4%) contribute only marginally. This indicates that STPC1 reflects a thermal–wind variability mode, meaning that the dominant spatio-temporal signal in Bali is driven primarily by temperature-based comfort and wind dynamics. The strong thermal–wind signature also explains why coastal locations regain negative STPC1 values under the robust formulation: once the influence of high-altitude outliers is reduced, the natural thermal–wind gradient between inland and coastal areas becomes more clearly represented. This block contribution interpretation provides a physically meaningful explanation for the spatial patterns observed in the robust STPC1.

3.6. Eigenvalue and Mean Shift Analysis

After obtaining matrix **A** and $\mathbf{A}_{\text{robust}}$, the mean coefficients of the basis functions from STPCA and STPCA-FASTMCD were compared. The mean basis coefficients for STPCA ($\bar{X}$) were derived from matrix **A**, whereas those for STPCA-FASTMCD ($\bar{M}_{\text{MCD}}$) were obtained from matrix $\mathbf{A}_{\text{robust}}$. Differences between these mean values affect the construction of the spatio-temporal matrix as well as the resulting eigenvectors and eigenvalues.

Table 6 shows notable discrepancies in several basis coefficients, particularly $\alpha_{1.0}$, $\alpha_{2.0}$, $\alpha_{3.0}$, and $\alpha_{4.0}$. These large differences arise because some STPCA mean coefficients are close to zero, while none of the STPCA-FASTMCD means approach zero. This finding indicates that outliers exert a substantial influence on pulling the data distribution.

Table 6: Comparison of Mean Basis Coefficients: STPCA vs. STPCA-FASTMCD

BFC	MBC $\mathbf{A}$ ($\bar{X}$)	MBC $\mathbf{A}_{\text{robust}}$ ($\tilde{M}_{\text{MCD}}$)	Difference ($\tilde{M}_{\text{MCD}} - \bar{X}$)
$\alpha_{1.0}$	5.18×10^{-15}	1.985287	1.985287
$\alpha_{1.1}$	-0.54892	-0.55688	-0.00795
$\alpha_{1.2}$	-0.5316	-0.52514	0.006461
$\alpha_{1.3}$	0.568796	0.567538	-0.00126
$\alpha_{1.4}$	0.625715	0.642817	0.017103
$\alpha_{2.0}$	4.22×10^{-15}	-0.09663	-0.09663
$\alpha_{2.1}$	0.557217	0.558117	0.0009
$\alpha_{2.2}$	0.387047	0.398068	0.011021
$\alpha_{2.3}$	0.285102	0.281156	-0.00395
$\alpha_{2.4}$	-1.07545	-1.08564	-0.01018
$\alpha_{3.0}$	1.33×10^{-15}	-0.3479	-0.3479
$\alpha_{3.1}$	-0.10035	-0.12682	-0.02646
$\alpha_{3.2}$	0.582395	0.585055	0.00266
$\alpha_{3.3}$	0.786708	0.867595	0.080887
$\alpha_{3.4}$	-1.03827	-0.95947	0.078798
$\alpha_{4.0}$	0	0.608484	0.608484
$\alpha_{4.1}$	0.224513	0.222799	-0.00171
$\alpha_{4.2}$	0.330626	0.37101	0.040384
$\alpha_{4.3}$	-0.83866	-0.88336	-0.04471
$\alpha_{4.4}$	-0.49068	-0.52075	-0.03006

The eigenvalue decomposition under STPCA and STPCA-FASTMCD shows clear differences, particularly in the leading components. The first four eigenvalues obtained from STPCA are substantially larger than those from STPCA-FASTMCD—for example, the first component decreases from 63.3570 to 37.7872. The detailed comparison can be seen in Table 7. This reduction indicates that classical STPCA is influenced by outlier-induced distortions, which inflate spatio-temporal relationships and produce large but unstable eigenvalues. With the application of FASTMCD, the impact of outliers is minimized, resulting in smaller yet more reliable eigenvalues that better capture the underlying structure of the data.

For the smaller eigenvalues (components 5 to 20), the differences between the two methods are relatively minor, and some components even show positive differences. This pattern suggests that outliers primarily affect the leading components, while their influence diminishes in later components. In line with the comparison of mean basis coefficients, several STPCA means close to zero contribute to instability in constructing the spatio-temporal matrix, whereas FASTMCD produces more stable means. As a result, the STPCA-FASTMCD decomposition yields a more consistent and robust eigenvalue structure.

In addition, the improved performance of STPCA-FASTMCD is also influenced by the more compact spatial structure produced after robust estimation. The reduction of effective spatial distances between observation locations leads to stronger and more coherent spatial relationships, which enhances the stability of the decomposition process. This aligns with the findings of [7], who showed that compact spatial structures improve the robustness of spatio-temporal PCA.

Table 7: Comparison of Eigenvalues in STPCA and STPCA-FASTMCD

Eigenvalue	STPCA	STPCA-FASTMCD	Difference
1	63.36	37.79	-25.57
2	7.0843	2.0119	-5.0725
3	1.9824	0.8297	-1.1526
4	0.6289	0.1967	-0.4322
5	0.1104	0.0508	-0.0596
6	0.0234	0.0007	-0.0227
7	0.0106	0.0001	-0.0105
8	0.0002	0.0000	-0.0002
9	0.0000	0.0000	0.0000
10	0.0000	0.0000	0.0000
11	0.0000	-0.0001	0.0000
12	-0.0001	-0.0001	0.0000
13	-0.0002	-0.0001	0.0000
14	-0.0014	-0.0014	0.0000
15	-0.0304	-0.0122	0.0182
16	-0.0501	-0.0450	0.0051
17	-0.0757	-0.1008	-0.0251
18	-0.1427	-0.1476	-0.0050
19	-0.7406	-0.3433	0.3972
20	-2.9580	-2.8222	0.1358

4. Conclusion

This study develops a robust formulation of Spatio-Temporal Principal Component Analysis (STPCA–FASTMCD) to address the sensitivity of classical STPCA to extreme observations commonly found in tropical climate datasets. The incorporation of FASTMCD successfully stabilizes the spatio-temporal structure by producing robust mean and covariance estimates, thereby reducing diagonal contributions inflation and mean shifts. This improvement is reflected in the decrease of the first eigenvalue from 63.36 under classical STPCA to 37.79 under STPCA–FASTMCD. The stabilizing effect is theoretically supported by the perturbation relationship

$$\delta\gamma = \omega_q' \, \delta C \, \omega_q,$$

where the reduced magnitude of δC under FASTMCD produces smaller perturbations in the eigenvalues. Empirically, this aligns with the smoother spatial and temporal loading patterns obtained in the robust analysis, confirming that the STPCA–FASTMCD framework yields more reliable and interpretable spatio-temporal modes of climate variability.

The comparison between STPCA and STPCA–FASTMCD shows clear differences in spatial patterns. Outliers originating from high-altitude sites distort the eigenvector orientation in the classical model, suppressing coastal contrasts and producing spatial patterns that do not align with Bali's known climate gradients. In contrast, the robust formulation restores these gradients by reducing the influence of extreme observations, resulting in more coherent and physically meaningful spatial representations. The analysis of block contributions further indicates that thermal comfort and wind speed dominate the structure of the first spatio-temporal component, confirming that the leading mode captures a thermal–wind pattern that is consistent with Bali's coastal–inland climatic dynamics.

This study still has limitations in handling outliers, which resulted in two locations being removed from the analysis. Future research should therefore consider integrating robust distance measures as adaptive spatial weights rather than exclusion criteria. Such an extension would allow all locations to remain within the model while proportionally reducing the influence of sites exhibiting extreme conditions, thereby producing a more flexible and fully inclusive robust spatio-temporal framework.

CRediT Authorship Contribution Statement

Agus Yarcana: Conceptualization, Methodology, Writing—-Original Draft. **Henny Pramoedyo:** Supervision, Validation, Editing. **Suci Astutik:** Supervision, Validation, Editing.

Declaration of Generative AI and AI-assisted technologies

This research was carried out entirely by the authors without the use of generative AI or any AI-assisted tools. All parts of the study, including data collection, data analysis, and manuscript writing, were completed independently by the authors.

Declaration of Competing Interest

The authors declare that there are no competing interests related to this study. The entire research process was conducted independently, without any external influence that could affect the results or conclusions.

Funding and Acknowledgments

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

Data and Code Availability

This study uses two data sources. BMKG sites observations (TC, A, R, and W for four full-observation stations, and rainfall R for the remaining 20 sites) are available upon reasonable request and subject to the official BMKG data-access policy. ERA5 reanalysis data (TC, A, and W) used to supplement incomplete BMKG measurements are freely accessible through the Copernicus Climate Data Store at the following link https://cds.climate.copernicus.eu/datasets/reanalysis-era5-single-levels-monthly-means?tab=overview.

Acknowledgments

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

References

[1] Intergovernmental Panel on Climate Change (IPCC). *Summary for Policymakers of the Intergovernmental Panel on Climate Change*. 2023. DOI: 10.59327/IPCC/AR6-9789291691647.001.

[2] M. Tang. *Comparing the 'Tourism Climate Index' and 'Holiday Climate Index' in Major European Urban Destinations*. [Online]. Available: https://uwspace.uwaterloo.ca/handle/10012/7638. 2013.

[3] D. Scott et al. "An inter-comparison of the Holiday Climate Index (HCI) and the Tourism Climate Index (TCI) in Europe". In: *Atmosphere (Basel)* 7.6 (2016), pp. 1–20. DOI: 10.3390/atmos7060080.

[4] K. Sumaja et al. "The Climate Comfort and Risk Assessment for Tourism in Bali, Indonesia". In: *Springer Proc. Phys.* Vol. 290. 2023, pp. 545–553. DOI: 10.1007/978-981-19-9768-6_50.

[5] L. R. Z. Dini and Sobirin. "Tingkat Kenyamanan Iklim di Pulau Bali Berdasarkan Tourism Climate Index". In: *Ind. Res. Work. Natl. Semin.* [Online]. Available: https://jurnal.polban.ac.id/index.php/proceeding/article/view/602/457. 2017, pp. 678–684.

[6] E. D. Lusiana et al. "Identifying key factors determining dynamics of environmental characteristics in Lesser Sunda island, Indonesia". In: *IOP Conf. Ser. Earth Environ. Sci.* 1191.1 (May 2023), p. 012004. DOI: 10.1088/1755-1315/1191/1/012004.

[7] M. Krzyśko et al. "Spatio-temporal principal component analysis". In: *Spat. Econ. Anal.* 19.1 (2024), pp. 8–29. DOI: 10.1080/17421772.2023.2237532.

[8] P. J. Rousseeuw and K. Van Driessen. "A fast algorithm for the minimum covariance determinant estimator". In: *Technometrics* 41.3 (1999), pp. 212–223. DOI: 10.1080/00401706.1999.10485670.

[9] D. I. Purnama and P. R. Sihombing. "Perbandingan Analisis Komponen Utama Dan Robust Pca (Robpca)". In: *J. Bayesian J. Ilm. Stat. dan Ekon.* 1.1 (2021), pp. 67–76. DOI: 10.46306/bay.v1i1.7.

[10] H. Ghorbani. "Mahalanobis Distance and Its Application for Detecting Multivariate Outliers". In: *Facta Univ. Ser. Math. Informatics* (2019), pp. 583–592. DOI: 10.22190/fumi1903583g.

[11] 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". In: *Geosci.* 14.6 (2024). DOI: 10.3390/geosciences14060155.

[12] D. W. S. Wong and F. Wang. "Spatial Analysis Methods". In: *Comprehensive Geographic Information Systems*. Ed. by B. Huang. Oxford: Elsevier, 2018, pp. 125–147. DOI: 10.1016/B978-0-12-409548-9.09598-1.

[13] N. Fat'Ha and H. T. Sutanto. "Identifikasi Autokorelasi Spasial Pada Pengangguran Di Jawa Timur Menggunakan Indeks Moran". In: *MATHunesa J. Ilm. Mat.* 8.2 (2020), pp. 89–92. DOI: 10.26740/mathunesa.v8n2.p89-92.

[14] M. Benko, W. Härdle, and A. Kneip. "Common functional principal components". In: *Ann. Stat.* 37.1 (2009), pp. 1–34. DOI: 10.1214/07-AOS516.

[15] L. F. Ichsari et al. "Studi Komparasi Hasil Pengolahan Pasang Surut Dengan 3 Metode (Admiralty, Least Square Dan Fast Fourier Transform) Di Pelabuhan Malahayati, Banda Aceh". In: *Indones. J. Oceanogr.* 2.2 (2020), pp. 121–128. DOI: 10.14710/ijoce.v2i2.7985.

[16] P. Septiawan and S. Nurdiati. "Analisis Empirical Orthogonal Function (EOF) Dan Transformasi Fourier Pada Sinyal Curah Hujan Indonesia". In: *Semin. Mat. Dan Pendidik. Mat.* 2017, pp. 179–186. DOI: 10.31227/osf.io/8e2f3.

[17] S. Stahlschmidt, W. K. Härdle, and H. Thome. "An Application of Principal Component Analysis on Multivariate Time-stationary Spatio-temporal Data". In: *Spat. Econ. Anal.* 10.2 (2015), pp. 160–180. DOI: 10.1080/17421772.2015.1023339.

[18] M. Krzyśko et al. "A novel Spatio-temporal principal component analysis based on Geary's contiguity ratio". In: *Comput. Environ. Urban Syst.* 103 (2023), p. 101980. DOI: 10.1016/j.compenvurbsys.2023.101980.

[19] S. D. A. Larasati, K. Nisa, and N. Herawati. "Robust Principal Component Trimmed Clustering of Indonesian Provinces Based on Human Development Index Indicators". In: *J. Phys. Conf. Ser.* 1751.1 (2021), p. 012021. DOI: 10.1088/1742-6596/1751/1/012021.

[20] R. N. M. Sanusi, D. R. S. Saputro, and R. Setiyowati. "Application of Robust Principal Component Analysis for Gross Regional Domestic Product of Provinces in Indonesia". In: *J. Phys. Conf. Ser.* 1776.1 (2021), p. 012056. DOI: 10.1088/1742-6596/1776/1/012056.

[21] M. Hubert, P. J. Rousseeuw, and K. Vanden Branden. "ROBPCA: A new approach to robust principal component analysis". In: *Technometrics* 47.1 (2005), pp. 64–79. DOI: 10.1198/004017004000000563.

[22] J. Fan, W. Wang, and Y. Zhong. *An ∞ eigenvector perturbation bound and its application to robust covariance estimation*. [Online]. Available: https://www.jmlr.org/papers/volume18/16-140/16-140.pdf. 2018.