Grouping Scheme Discriminant Canonical Correlation Analysis for Time-Dependent Multivariate Data Structure

Gyasi-Agyei, Kwame Asare[1*], Nkansah, Bismark Kwao[2], Eyiah-Bediako, Francis[2], Nkansah, Henrietta[3]

[1] Department of Applied Mathematics, FAST, KTU, P. O. Box KF 981, Koforidua, Ghana, W/A.

[2] Department of Statistics, School of Physical Sciences, UCC, Cape Coast, Ghana, W/A.

[3] Department of Mathematics, School of Physical Sciences, UCC, Cape Coast, Ghana, W/A.

Correspondence: gyasi-agyei.kwame@ktu.edu.gh

DOI: https://doi.org/10.30212/JITI.202503.001

Submitted: Nov. 20, 2024 Accepted: Feb. 01, 2025

ABSTRACT

One popular covariance analysis methodology, canonical correlation analysis, is not primarily intended for multivariate multiple time-dependent data structures that may be appropriately divided into two response and predictor variables groups. This means that conventional canonical correlation analysis does not yield practical results for such data problems. This study, therefore, designs and implements a grouping scheme discriminant canonical correlation analysis to handle this problem so that the time effect is adequately captured in the computation of the correlation coefficient between the two sets of variables. We also show how multiple discriminant analyses obtain the ideal value. This process is known as grouping scheme discriminant canonical correlation analysis in this study. Therefore, the grouping scheme discriminant canonical correlation analysis is a method designed to handle time-dependent multivariate data efficiently by integrating the time effect into the canonical correlation through discriminant analysis. Based on data on six weather conditions, the demonstrations show that the correlation coefficient between heating and cooling sets of weather conditions varies at different time points, and that the overall correlations are higher than that obtained from data assumed to be time-independent. The detection techniques for multiple discriminant analysis and the currently used canonical correlation analysis are compared in this paper. The results are validated through simulation and real performance review. According to the findings, determining the genuine correlation between the two sets of variables with time-dependent structure is significantly improved by seven-group discriminant analysis. Furthermore, seven-group discriminant analysis yielded the best results for the combination method based on multiple discriminant analysis and conventional canonical correlation analysis. It has been noted that the time impact is successfully incorporated into the calculation of the canonical correlation when the grouping scheme is used in discriminant canonical correlation analysis. The results finally show that incorporating the time effect into canonical correlation analysis achieves a more reasonable relationship between subset variables within the data.

Keywords: Canonical correlation, Canonical discriminant function, Grouping scheme, Time-dependent multivariate data, Weather conditions.

1. Introduction

Canonical correlation analysis (CCA) is one covariance analysis technique used in various disciplines to examine the relationship between a set of response variables, **Y** and a set of predictor variables, **X,** that make up the same dataset. A variety of fields have used CCA in several applications. The goal of CCA is to ascertain whether the predictor set of variables has an impact on the response sets of variables [1]. It has been noticed that CCA has drawn a lot of interest as a potent approach for fusing multimodal features [2]. Like that of regression, the aim of CCA is to measure the strength of the link between the two sets of data [3, 4]. It is comparable to factor analysis for constructing variable composites. It is also similar to discriminant analysis in that it can generate independent dimensions with the aim of obtaining the highest correlation between the dimensions for each set of variables [5].

Given two interrelated random vectors $\mathbf{Y} = (Y_1, Y_2, ..., Y_p)'$ and $\mathbf{X} = (X_1, X_2, ..., X_q)'$, assume for convenience, that $p \leq q$ [6]. CCA combines the two new variables provided by the linear combinations: $U_i = \alpha_i' \mathbf{Y}$ and $V_j = \beta_j' \mathbf{X}$, such that the correlation between the two linear functions U_i and V_j is maximized, where, $\alpha_i = (\alpha_{i1}, \alpha_{i2}, ..., \alpha_{ip})'$ and $\beta_j = (\beta_{j1}, \beta_{j2}, ..., \beta_{jq})'$ are the coefficient vectors of **Y** and **X**, respectively. This conventional objective of CCA is, however, independent of the possible time effect.

The basic structure of the multivariate multiple time-dependent data (MMTDD) is fairly represented in the literature $(n \times d \times T)$, where *n* is the total number of observations (subject count), *d* is the total number of time-dependent variables, and *T* is the total number of time events. CCA and such data may be considered time-independent if the effect of T is not appropriately captured in the procedure design. In order to incorporate the time effect, the stated basic structure of the MMTDD could further be partitioned as $(n_1 \times d \times t_1)$, $(n_2 \times d \times t_2)$,…, $(n_g \times d \times t_g)$ [3, 6] for suitably determined t_g, $g < T$, such that the average of the CCA could provide a reasonable reflection of the relationship between **Y** and **X** [7].

In this study, we demonstrate how the optimal value of *g* is determined via multiple discriminant analysis. This procedure is referred to in this study as grouping scheme discriminant canonical correlation analysis (GSDCCA). Thus, the GSDCCA is an approach intended to effectively incorporate the time effect into the CCA via discriminant analysis to handle the time-dependent multivariate data (TDMD). Attempts at capturing the time effect in CCA have been pursued in a few notable approaches by [8, 9, 10] from the perspective of optimization incorporated into the basic formulation of CCA.

To comprehend the coupling dynamics and temporal variations between the two time-varying sources, [8] created the Time-dependent CCA. Time-dependent canonical vectors can be extracted from multilevel time series data using the Time-dependent CCA technique. They suggest a convex formulation of the issue that uses the singular value decomposition (SVD) characteristics present in all answers to the CCA problem. They test the proposed approach using simulated datasets. They show that the Time-dependent CCA-based strategy outperforms previous feature extraction and temporal variation detection techniques. Discriminative multiple CCA is used by [11] for the analysis and synthesis of multimodal data. They discovered that it could extract more distinct features from multimodal information representations. To better use the multimodal data, they carefully selected projected trajectories that increased within and decreased between classes.

It is challenging to understand the concept of CCA because of how it is presented, which appears complicated. This might be because the method is mathematically intensive. As a result, it is necessary to conduct a study that provides a CCA in a rational and approachable manner. Therefore, the study’s objectives are to provide simple procedures for producing canonical variables using generated codes and explicitly state the logical justification for the results. It has facilitated the application of CCA to multivariate multiple time-dependent data structures.

The intellectual underpinnings and other writers’ perspectives are considered in Section 2, Review of Related Literature. Section 3 describes the methodology for the study and points out how discriminant analysis (DA) is incorporated into the CCA. Relevant hypothesis testing techniques have also been presented in this section. Section 4 demonstrates the proposed procedure on a suitable data set. Section 5 presents the discussion and conclusion, respectively.

2. Literature Review

Canonical Correlation Analysis (CCA) is being used as a key approach in increasing studies. A comprehensive review of relevant literature is conducted to ascertain the use of CCA. A quick survey of the literature on recommendation systems and other types of CCA follows.

From the work of [12], we use the discriminant analysis (DA) method to evaluate the available data when the response variable is categorical and the predictor variable is of an interval type. A response variable is divided into different categories when referred to as a categorical variable. For example, one of the three dummy variables, 1, 2, or 3, can be the category answer variable. A discriminant function is a linear collection of predictor variables that accurately differentiates between the response variable categories. Creating discriminant functions is the aim of the DA method. The number of categories the response variables have determines how DA is defined. Since statistics holds that all propositions are true until infinity, the type used in this case is Two-group DA when the dependent variable has two categories. When the dependent variable has three or more categories, Multiple-group DA is utilized [12].

[6] examined the connection between Indonesia's economic development and unemployment rates in each province in 2021 using CCA, one of the dependent methods used for multivariate analysis. Both variables are regarded as dependent factors, and their research uses five independent variables: the development of the human index, the minimum wage by region, the percentage of poor citizens, investment, and the farmer rate value in each province. This approach allowed them to identify the variable with the strongest correlation between the independent and dependent variables. According to their findings, the unemployment rate had the most significant influence on the connection between the dependent and independent variables, while the human index development had the most substantial relationship.

The work of [13] is also worthy of note. Their work investigates bias and fairness in CCA to look at the relationship between two sets of variables. Reducing the correlation disparity error linked to protected traits offers a framework that lessens unfairness. Their method guarantees that global projection matrices from all data points produced by CCA have similar correlation levels to group-specific projection matrices. The effectiveness of their approach in lowering correlation disparity error without sacrificing CCA accuracy is demonstrated by experimental evaluation on both synthetic and real-world datasets.

The literature makes it abundantly evident that little research has been done on using CCA in multivariate time-dependent data. Thus, this area remains grey for further exploration.

3. Research Design

We examine the methodology used in this paper and present a review of data notation for numerous multivariate multiple time-dependent (MMTD) variables. The review also covers key matrices and the fundamental prerequisites and presumptions for using the matrices are pertinent. The section provides some multivariate multiple DA methods for selecting a reasonable model and CCA methods for these models. The primary goal is to generate ideas for GSDCCA.

3.1 The Concept of Canonical Correlation Analysis

Given two interrelated random vectors $\mathbf{Y} = (Y_1, Y_2, \ldots, Y_p)'$ and $\mathbf{X} = (X_1, X_2, \ldots, X_q)'$, assume for convenience that $p \le q$. The number of variables in each set of variables is used to determine the random vectors [1, 14]. The combined covariance matrix is produced by using the enhanced random vectors.

Let $E(\mathbf{Y}) = \mu_y$ and $E(\mathbf{X}) = \mu_x$ be the respective expectations of **Y** and **X**. The resulting combined random vector, **Z**, and its mean vector, μ are, respectively, given as

$$\mathbf{Z} = \begin{bmatrix} Y \\ \cdots \\ X \end{bmatrix} \text{ and } \mu = \begin{bmatrix} E(Y) \\ \cdots \\ E(X) \end{bmatrix}$$

The combined covariance matrix ($\mathbf{\Sigma}$) for the enhanced random vector is calculated as follows:

$$\mathbf{\Sigma} = E(\mathbf{Z} - \mu)(\mathbf{Z} - \mu)^{/} = \left[\begin{array}{c:c} E(Y-\mu_y)(Y-\mu_y)^{/} & E(Y-\mu_y)(X-\mu_x)^{/} \\ \hdashline E(X-\mu_x)(Y-\mu_y)^{/} & E(X-\mu_x)(X-\mu_x)^{/} \end{array}\right] = \left[\begin{array}{c:c} \Sigma_{11} & \Sigma_{12} \\ \hdashline \Sigma_{21} & \Sigma_{22} \end{array}\right] \tag{1}$$

where, $D(\mathbf{Y}) = \Sigma_{11}$ is $(p \times p)$ and $D(\mathbf{X}) = \Sigma_{22}$ is $(q \times q)$ sample covariance matrices for **Y** and **X**, respectively, and $Cov(\mathbf{Y}, \mathbf{X}) = \Sigma_{12}$ [6, 14].

For a fixed number of i = 1, 2, ..., p and j = 1, 2, ..., q, the linear combinations $U_i = \alpha_i^{/}\mathbf{Y}$ and $V_j = \beta_j^{/}\mathbf{X}$

are such that the optimal correlation

$$\rho_i^{*} = \underbrace{\max}_{\alpha,\beta} Corr(\alpha^{/}\mathbf{Y},\ \beta^{/}\mathbf{X}) = \alpha^{/}\Sigma_{12}\beta \tag{2}$$

is subject to the constraints: $\alpha^{/}\Sigma_{11}\alpha = \beta^{/}\Sigma_{22}\beta = 1$ $Cov\left(\alpha_i^{/}\mathbf{Y},\ V_j\right) = Cov\left(U_i, \beta_j^{/}\mathbf{X}\right) = 0$, and $Cov\left(\alpha_i^{/}\mathbf{Y},\ U_i\right) = Cov\left(V_j, \beta_j^{/}\mathbf{X}\right) = 1$.

3.2 Fundamental Approaches to Canonical Correlation Analysis

Following the Cauchy-Schwarz Inequality approach, the key matrices for canonical correlation analysis may be presented in a compact form as

$$\mathbf{Q}_i = \Sigma_{ii}^{-\frac{1}{2}}\mathbf{P}_i\Sigma_{ii}^{\frac{1}{2}} = \Sigma_{ii}^{-1}\Sigma_{ik}\Sigma_{kk}^{-1}\Sigma_{ki} \tag{3}$$

where,

$$\mathbf{P}_i = \Sigma_{ii}^{-\frac{1}{2}}\Sigma_{ik}\Sigma_{kk}^{-1}\Sigma_{ki}\Sigma_{ii}^{-\frac{1}{2}}, \qquad k = \begin{cases} i+1; & i=1 \\ i-1; & i=2 \end{cases} \tag{4}$$

Thus, $\mathbf{Q}_i$ has the same eigenvalues as $\mathbf{P}_i$, with the corresponding eigenvectors given as

$$E_{N_i} = Eig(\mathbf{P}_i) \quad \text{and} \quad Eig(\mathbf{Q}_i) = \Sigma_{ii}^{-\frac{1}{2}}Eig(\mathbf{P}_i) \tag{5}$$

where

$$Eig(\mathbf{P}_i) = \begin{cases} E_{N_1}; & i=1 \\ \Sigma_{22}^{-\frac{1}{2}}\Sigma_{21}\Sigma_{11}^{-\frac{1}{2}}E_{N_1}; & i=2 \end{cases} \tag{6}$$

It is noted that by defining the matrix, $\mathbf{A} = \Sigma_{11}^{-\frac{1}{2}}\Sigma_{12}\Sigma_{22}^{-\frac{1}{2}}$, then $\mathbf{P}_1 = \mathbf{A}\mathbf{A}^{/}$ and $\mathbf{P}_2 = \mathbf{A}^{/}\mathbf{A}$. It follows from the definitions that the canonical variables are the ordered eigenvectors of the matrices $\mathbf{Q}_i$, i = 1, 2, and that the first five matrices **A,** $\mathbf{P}_1$, $\mathbf{P}_2$, $\mathbf{Q}_1$ and $\mathbf{Q}_2$ have the same non–zero eigenvalues.

Let $\mathbf{Q}_{it}$ be a sequence of input matrices for t = {1, 2, 3, ..., g}, $g < T$ where t_s (s = 1, 2, ..., g) are some g partitions of the original time period T and $t_1 = T$ under various schemes $1, 2, \ldots, T-1$. The optimal scheme is equivalent to the multiple-group discriminant analysis (m-group DA) for that scheme that yields more optimal values than the 2-group discriminant canonical correlation analysis.

For $\mathbf{Q}_{it}$, let the corresponding sequence of eigenvectors be $Eig(\mathbf{Q}_{it})$ and $\lambda_t(\mathbf{Q}_{it})$ be the corresponding values. Thus, for any t, the eigen Equation will be

$$[\mathbf{Q}_{it} - \mathbf{\Lambda}_t(\mathbf{Q}_{it})]\, Eig(\mathbf{Q}_{it}) = 0 \qquad (7)$$

where, $\mathbf{\Lambda}_t(.)$ is $f \times f$ diagonal matrix of eigenvalues of $(.)$, $f = min(p, q)$. The solution to Equation (7) gives the matrix, $\mathbf{\Lambda}_t$. The overall canonical correlation between **Y** and **X** is for the average matrix given as

$$\mathbf{\Lambda} = \frac{1}{g}\sum_{t=1}^{g} \times \mathbf{\Lambda}_t \qquad (8)$$

3.3 Link between Canonical Correlation Analysis and Discriminant Analysis

It is possible to construct a multiple-group discriminant analysis as a canonical correlation analysis problem with group membership as the dependent variable coded using dummy variables [15, 16]. Let the observations $\mathbf{y}_k$ be a collection of heating variables, $\mathbf{x}_k$ label for cooling variables, C_k is the collection of points, n is the total number of observations and n_k is the number of observations in class k, where k = 1, 2, ..., g then, an average of the observations for class k is

$$\mathbf{m}_k = \frac{1}{n_1}\sum_{k:x_{ki}\in C_k}(\mathbf{y}_k) \qquad (9)$$

The sum of squares total is defined by

$$\mathbf{S}_T = \sum_{k=1}^{g}\sum_{k:x_k\in C_k}^{n_i}(\mathbf{y}_k\mathbf{y}_k^{/}) = (n-1)\mathbf{S}_{yy} \qquad (10)$$

where $\mathbf{S}_{yy}$ is the sample covariance matrix of **Y**. The sum of squares total $\mathbf{S}_T$ can be divided into the sum of squares within class $\mathbf{S}_W$ and between class $\mathbf{S}_B$ such that

$$\mathbf{S}_W = \sum_{k=1}^{g}\sum_{k:x_{ki}\in C_k}^{n_i}(\mathbf{y}_k - \mathbf{m}_k)(\mathbf{y}_k - \mathbf{m}_k)^{/} \qquad (11)$$

$$\mathbf{S}_B = \sum_{k=1}^{g}(n_k\mathbf{m}_k\mathbf{m}_k^{/}) \qquad (12)$$

and hence $\mathbf{S}_T = \mathbf{S}_B + \mathbf{S}_W$ [15].

If we define $\mathbf{y}_k = \mathbf{I}_{(\mathbf{x}_k\in C_k)}$ as the label matrix with kth entry being described as the indicator function, $\mathbf{X}\in\Re^{n\times g}$ as a matrix whose rows are the observations, $\mathbf{x}_k$, the class labels matrix, $\mathbf{Y}\in\Re^{n\times g}$ and **I** as the indicator function [1, 17], we have

$$\mathbf{Y} = \begin{bmatrix} \mathrm{I}n_1 & 0 & \dots & 0 \\ 0 & \mathrm{I}n_2 & \dots & 0 \\ :: & : & . & : \\ 0 & 0 & \dots & \mathrm{I}n_g \end{bmatrix}, \text{ where } \quad \mathbf{I} = \begin{bmatrix} 1 & 0 & \dots & 0 \\ 0 & 1 & \dots & 0 \\ :: & : & . & : \\ 0 & 0 & \dots & 1 \end{bmatrix}$$

It is clear from this that if $\mathbf{m}_k$ is the average of the observations for class *k*, then

$$\mathbf{S}_{YX} = \mathbf{Y}^{/}\mathbf{X} = \begin{bmatrix} n_1 m_1^{/} \\ n_2 m_2^{/} \\ : \\ n_g m_g^{/} \end{bmatrix}. \text{ It follows that: } \qquad \mathbf{S}_{YY}^{-1} = (\mathbf{Y}^{/}\mathbf{Y})^{-1} = \begin{bmatrix} \frac{1_1}{n_1} & 0 & \dots & 0 \\ 0 & \frac{1}{n_2} & \dots & 0 \\ :: & : & . & : \\ 0 & 0 & \dots & \frac{1}{n_g} \end{bmatrix}$$

Thus, $$\mathbf{S}_{XY}\mathbf{S}_{YY}^{-1} = \sum_{k=1}^{g} n_k \mathbf{m}_k \mathbf{m}_k^{/} = \mathbf{S}_B \tag{13}$$

We begin with the sample canonical correlation matrix equation given in Equation (14), where ρ_H is the canonical correlation coefficient (CCC) and α is the coefficient vector for the response variable [1, 3, 18].

$$\mathbf{S}_{XY}\,\mathbf{S}_{YY}^{-1}\,\mathbf{S}_{YX}\,\alpha = \rho_H^2\,(n-1)\,\mathbf{S}_{yy}\alpha \tag{14}$$

Substituting Equations (10) and (13) into Equation (14) yields

$$\mathbf{S}_B\,\alpha = \rho_H^2\,\mathbf{S}_T\,\alpha \tag{15}$$

Since $\mathbf{S}_T = \mathbf{S}_B + \mathbf{S}_W$, then it follows that: $$\mathbf{S}_B\,\alpha = \left(\frac{\rho_H^2}{1-\rho_H^2}\right)\mathbf{S}_W\,\alpha \tag{16}$$

This link will be relevant by first determining the optimal groupings in the data for which discrimination is maximum.

3.4 Formulation of Grouping Schemes Discriminant Canonical Correlation Analysis

Grouping scheme (GS) is an approach that is intended in this study to incorporate the time effect into the discriminant canonical correlation analysis (DCCA) in order to handle the time-dependent multivariate data (TDMD) effectively. The number of possible group analyses that can be performed is denoted as

G = (2, 3, ..., *t* – 1) for data on *t* years. Thus, for multiple-group (*m*-group) DCCA), $m \in G$. The formulations of GS discriminant canonical correlations for 2-group DCCA (G_2) and 3-group DCCA (G_3) are given, respectively, as shown in the following matrices:

$$G_{2M} = \begin{bmatrix} 1_{y_1} & 1_{y_1} & 1_{y_1} & 1_{y_1} & \cdots & 1_{y_1} \\ 2_{y_2} & 1_{y_2} & 1_{y_2} & .1_{y_2} & \cdots & 1_{y_2} \\ 2_{y_3} & 2_{y_3} & 1_{y_3} & 1_{y_3} & \cdots & 1_{y_3} \\ 2_{y_4} & 2_{y_4} & 2_{y_4} & 1_{y_4} & \cdots & \vdots \\ \vdots & \vdots & \vdots\ . & \vdots & . & 1_{y_{t-1}} \\ 2_{y_t} & 2_{y_t} & 2_{y_t} & 2_{y_t} & \cdots & 2_{y_t} \end{bmatrix} \tag{17}$$

In this section, we assigned 1 to the first 12 months (1_{y_1}) and 2 to the remaining months for the first column, grouping scheme one (GS_1). For the second column, grouping scheme two (GS_2), we assigned 1 to the first 24 months (1_{y_1} to 1_{y2}) and 2 to the remaining months. For the third column, grouping scheme three (GS_3), we assigned 1 to the first 36 months (1_{y_1} to 1_{y_3}) and 2 to the remaining months, as shown in the matrix G_{2M} above. We repeated this procedure for all the grouping schemes and used SPSS to run the resulting data, which G_{2M} denotes a 2-group matrix.

$$G_{3M} = \begin{bmatrix} 1_{y_1} & 1_{y_1} & 1_{y_1} & 1_{y_1} & \cdots & 1_{y_1} \\ 2_{y_2} & 1_{y_2} & 1_{y_2} & .1_{y_2} & \cdots & 1_{y_2} \\ 3_{y_3} & 2_{y_3} & 1_{y_3} & 1_{y_3} & \cdots & 1_{y_3} \\ 3_{y_4} & 3_{y_4} & 2_{y_4} & 1_{y_4} & \cdots & \vdots \\ 3_{y_5} & 3_{y_5} & 3_{y_5} & .2_{y_5} & . & 1_{y_{t-2}} \\ \vdots & \vdots & \vdots & \vdots & \cdots & 2_{y_{t-1}} \\ 3_{y_t} & 3_{y_t} & 3_{y_t} & 3_{y_t} & \cdots & 3_{y_t} \end{bmatrix} \tag{18}$$

The general formulation of GS discriminant canonical correlation analysis for *m*-group DCCA (G_m) is given as

$$G_{mM} = \begin{bmatrix} 1_{y_1} & 1_{y_1} & \cdots 1_{y_1} & \cdots & 1_{y_1} & \cdots & 1_{y_1} & 1_{y_1} \\ 2_{y_2} & 1_{y_2} & \cdots 1_{y_2} & \cdots & 1_{y_2} & \cdots\cdots & 1_{y_2} & 1_{y_2} \\ m_{y_3} & 2_{y_3} & .\ 1_{y_3} & \cdots & \vdots & \cdots & \vdots & \vdots \\ \vdots & \vdots & \cdots 1_{y_4} & \cdots & 1_{y_j} & \cdots & 1_{y_{t-m}} & 1_{y_{t-m+1}} \\ m_{y_m} & b_{y_m} & \cdots 2_{y_5} & . & 2_{y_{j+1}} & \cdots & \vdots & \vdots \\ m_{y_d} & m_{y_d} & \cdots \vdots & \cdots & \vdots & \cdots & b_{y_{t-2}} & a_{y_{t-2}} \\ \vdots & \vdots & \cdots 3 & \cdots & b_{y_{t-1}} & \cdots & m_{y_{t-1}} & b_{y_{t-1}} \\ m_{y_t} & m_{y_t} & \cdots m_{y_t} & \cdots & m_{y_t} & \cdots & m_{y_t} & m_{y_t} \end{bmatrix} \tag{19}$$

where $a = (m - 2)$, $b = (m - 1)$, $d = (m + 1)$, $m = 3, 4, ..., t{-}1$; $i = 3, 4, ..., t{-}1$, and C_{y_t} means that all observations in the year y_t are coded as *C*. Suppose $G_m(:, j)$ is the best grouping, and $\mathbf{Z} = [\mathbf{Y} \vdots \mathbf{X}]$

is the combined variables of Set 1 and Set 2, then the augmented data that incorporates the scheme is given by

$$\mathbf{Z}_F = [\mathbf{Y} \mid \mathbf{X} \mid G_m(:, j)] \tag{20}$$

where $\mathbf{Z}_F \in \Re^{p+q+m}$ and $G_m(:, j) \in \Re^m$.

3.5 Hypotheses Testing of Discriminant Canonical Correlation Analysis

We need to analyze some related hypotheses in CCA and DA to arrive at the optimal GS. Let ρ_r be the r^{th} canonical correlation, r = 1, 2,..., f. The null and alternative hypotheses for computing the statistical significance of CCA [14], are given as follows:

$$H_0 : \rho_1 = \rho_2 = \ldots = \rho_f = 0$$
$$H_1 : \rho_1 \neq \rho_2 \neq \ldots \neq \rho_f \neq 0$$

where $f = min(p, q)$. We can test the hypotheses using a variety of test statistics. The test statistic based on Wilks' Lambda (**Λ**) is given by Equation 21.

$$\mathbf{\Lambda} = \Pi_{r=1}^{f}(1-\rho_r^2) \tag{21}$$

The closer Lambda is to zero, the more likely canonical correlation will be statistically significant. The statistical significance of **Λ** or the likelihood ratio is tested by using the test statistic [19] given as

$$\mathbf{B} = -[n-1-\tfrac{1}{2}(p+q+1)]\ln \mathbf{\Lambda} \tag{22}$$

The Hotelling's *T*-square test statistic is given by

$$T^2 = \tfrac{n_1 n_2}{n_1+n_2}(\mu_1-\mu_2)^{\prime}\mathbf{S}^{-1}(\mu_1-\mu_2) \tag{23}$$

where $\mathbf{S} = \frac{(n_1-1)S_1+(n_2-1)S_2}{n_1+n_2-2}$ is the pooled sample covariance matrix of **Y** and **X**. $\mathbf{S}_1$, and $\mathbf{S}_2$ are the sample covariance matrices of **Y** and **X**, respectively. Large n T^2 is approximately distributed as Chi-square (χ^2) with p degrees of freedom [19, 20, 21].

If λ_r is the r^{th} eigenvalue of the matrix $\mathbf{P}_1$ as defined above, for r = 1, 2, ..., f, then the other relevant statistic measures in Table 1 are helpful [22] about the stated hypotheses.

Table 1. Multivariate Tests of Significance

Test Statistic	**Method**	**Remarks**
Pillai's trace	$\sum_{r=1}^{f}\frac{\lambda_r}{1+\lambda_r}$	Should be large and between 0 and 1
Wilks	$\sum_{r=1}^{f}\frac{1}{1+\lambda_r}$	Should be small

Roys Largest Root	$\lambda_r(1+\lambda_r)$	Should be large

4. Results and Discussions

The data used for the implementation covers six weather conditions variables obtained from Ghana Metrological Agency (GMet) from January 2002 to December 2023 and involves 264 observations. The data is partitioned into response variables **Y** = (maximum temperature, minimum temperature, solar radiation) and predictor variables **X** = (precipitation, wind, relative humidity). The time series plots of the monthly weather conditions are shown in Figure 1.

4.1 Some Explorations

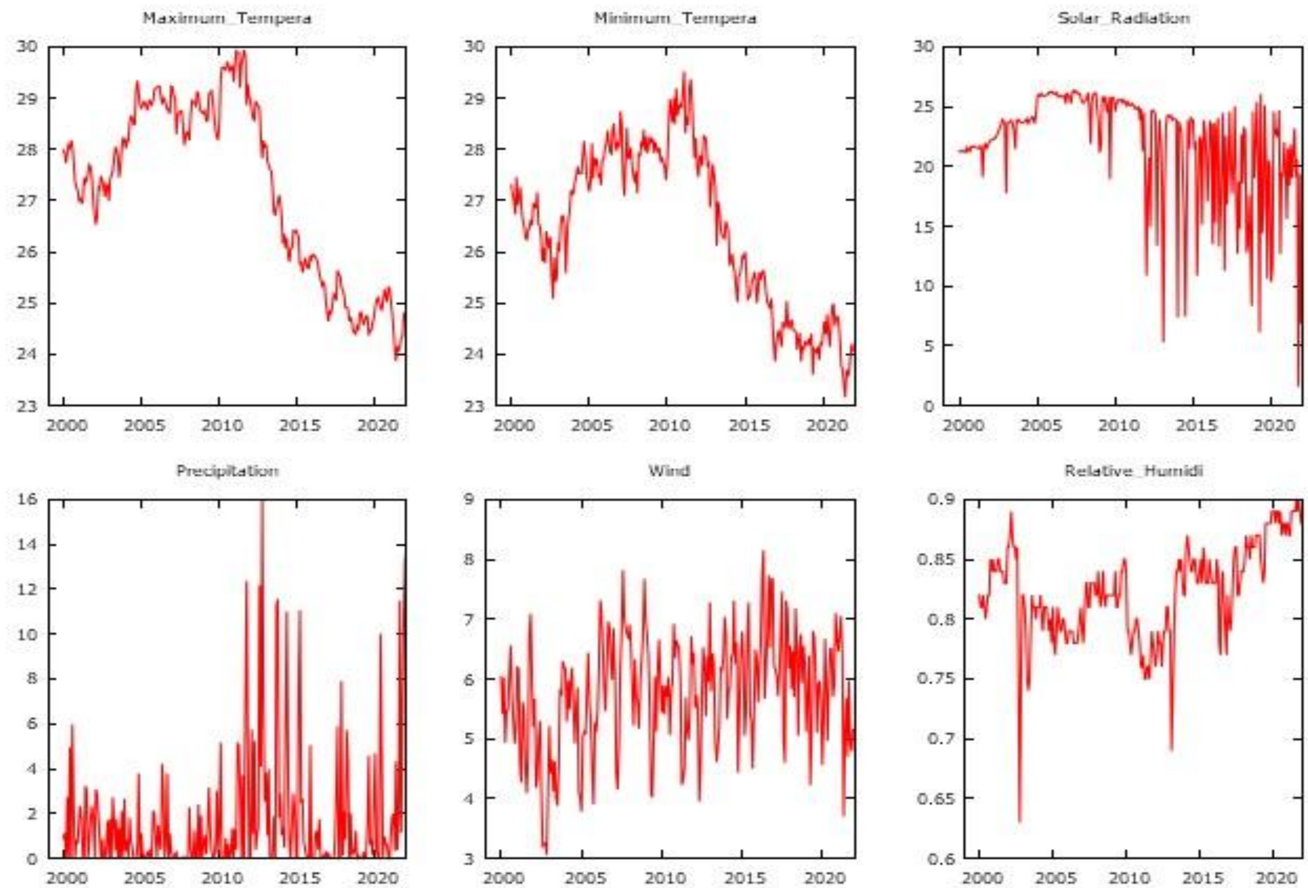

Figure 1. Series plots of data on six monthly weather conditions

The two temperature variables have similar characteristics with slightly higher variation in the second. In both cases, there is a gently increasing linear trend with considerable variability up to about the year 2013, after which the pattern assumes a sharp negative linear trend till the end of the series, with variation in the last few portions being quite high. However, the general trend in all six series suggests evidence of non-stationarity in the mean. Generally, there appears to be a noticeable change in the behaviour of all the series around 2013.

Table 2 gives the relevant statistical measures on CCA of the illustrative data without reference to the time component.

Table 2. Relevant Canonical Correlation Analysis Statistic Measures from Original Data

Root No.	CCC	Eigenvalue	Var. Exp. (%)	Cum. Var. Exp. (%)	W. Lambda	Significant
1	0.731	1.145	87.270	87.270	0.398	0.000
2	0.345	0.119	10.316	97.586	0.854	0.000
3	0.175	0.031	2.414	100.000	0.969	0.004

Source: Researchers' computations (2024)

From Table 2, the column CCC shows the canonical correlation coefficient for the data without considering the time effect. All of these coefficients are significant. The results demonstrate that the relationship between the heating and cooling climate variables is generally substantial, even without considering the time component. Table 3 depicts the overall tests of significance for statistical measures of the original data. The results show that all statistical measures are significant, indicating that the canonical correlations differ significantly.

Table 3. Overall Tests of Significance for Statistic Measures of the Original Data

Test Statistic	Hotelling's	Pillai's	Wilks	Roy's	Eigenvalue	CCC
Value	0.813	4.339	0.187	0.813	4.339	0.901
Sign	0.000	0.000	0.000	0.000	0.000	0.000

Source: Researchers' computations (2024)

4.2 Grouping Scheme Discriminant Canonical Correlation Functions

The explorations above are based on the data without referencing the time component. Now, consider the effect of the year by first assuming that the year introduces only 2-group discrimination in the data. Since *m* should be two or greater than two, higher statistical values from 2-group discriminant canonical correlation analysis (DCCA) would serve as the foundation for further research on *m*-group DCCA.

4.2.1 Two-group grouping scheme discriminant canonical correlation function

The 2-group DCCA design is presented in the grouping scheme (GS) pictorial map given in Figure 2, where the colours denote dummy variables. The grouping schemes are used to run several G_2 from grouping scheme one (GS_1) through to grouping scheme twenty (GS_{20}) to determine the one that has the optimal statistic measures, as shown in Table 4.

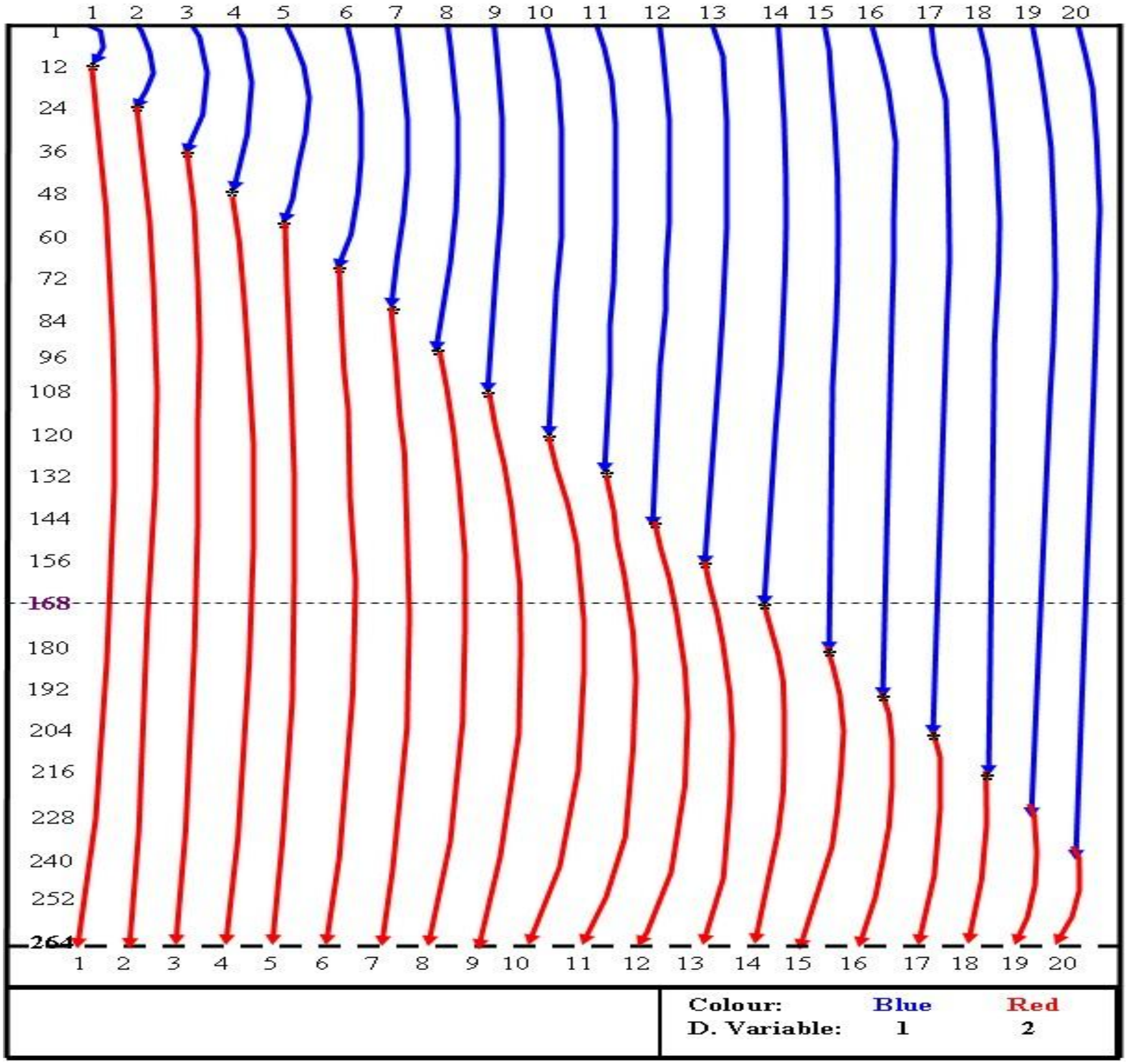

Figure 2. Grouping Scheme pictorial map for the 2-group discriminant canonical correlation function

Table 4 reports the first discriminant CCC (FDCCC) values and other relevant statistics. Grouping scheme fourteen (GS_{14}) has the optimal statistic measures among all the schemes, which falls on the 168th month, the end of 2013 as indicated in Figure 2. That is, assuming that the data may be suitably segmented into two, then the partition that is provided by month 168 is the one that offers the best correlation between the two subsets of weather conditions. It is worth noting that the highest FDCCC obtained in this case is 0.907, slightly higher than that in Table 3.

Table 4. Classification Measures of 2-group Discriminant Canonical Correlation Function

GS	FDCCC	Eigenvalue	W. Lambda	Chi-Square

1	0.141	0.020	0.980	5.221
2	0.219	0.050	0.952	12.697
3	0.341	0.131	0.884	31.965
4	0.425	0.220	0.819	51.598
5	0.466	0.278	0.783	63.502
6	0.580	0.507	0.663	106.263
7	0.556	0.448	0.690	95.964
8	0.578	0.501	0.666	105.117
9	0.628	0.652	0.605	129.981
10	0.700	0.962	0.510	174.606
11	0.781	1.564	0.390	243.834
12	0.851	2.619	0.276	333.130
13	0.898	4.143	0.194	424.140
14	**0.907**	**4.627**	**0.178**	**447.452**
15	0.878	3.377	0.228	382.372
16	0.835	2.311	0.302	310.115
17	0.720	1.474	0.404	234.660
18	0.720	1.074	0.482	188.894
19	0.632	0.666	0.600	132.276
20	0.541	0.251	0.707	89.849

Source: Researchers' computations (2024)

4.2.2 *Multiple-group grouping scheme discriminant canonical correlation functions*

We run many DCCA groups, ranging from 3-group DCCA (G_3) to 20-group DCCA (G_{20}). The primary goal is to determine whether all statistic measures in a given *m*-group (G_m) are more optimal than those of the GS_{14} 2-group DCCA (G_2) in Table 4. The procedure identifies 7-group DCCA (G_7) to produce the most optimal results. Figure 3 depicts the plots of the GS pictorial map for 7-group DCCA, confirming the most optimal values for all four discrimination measures.

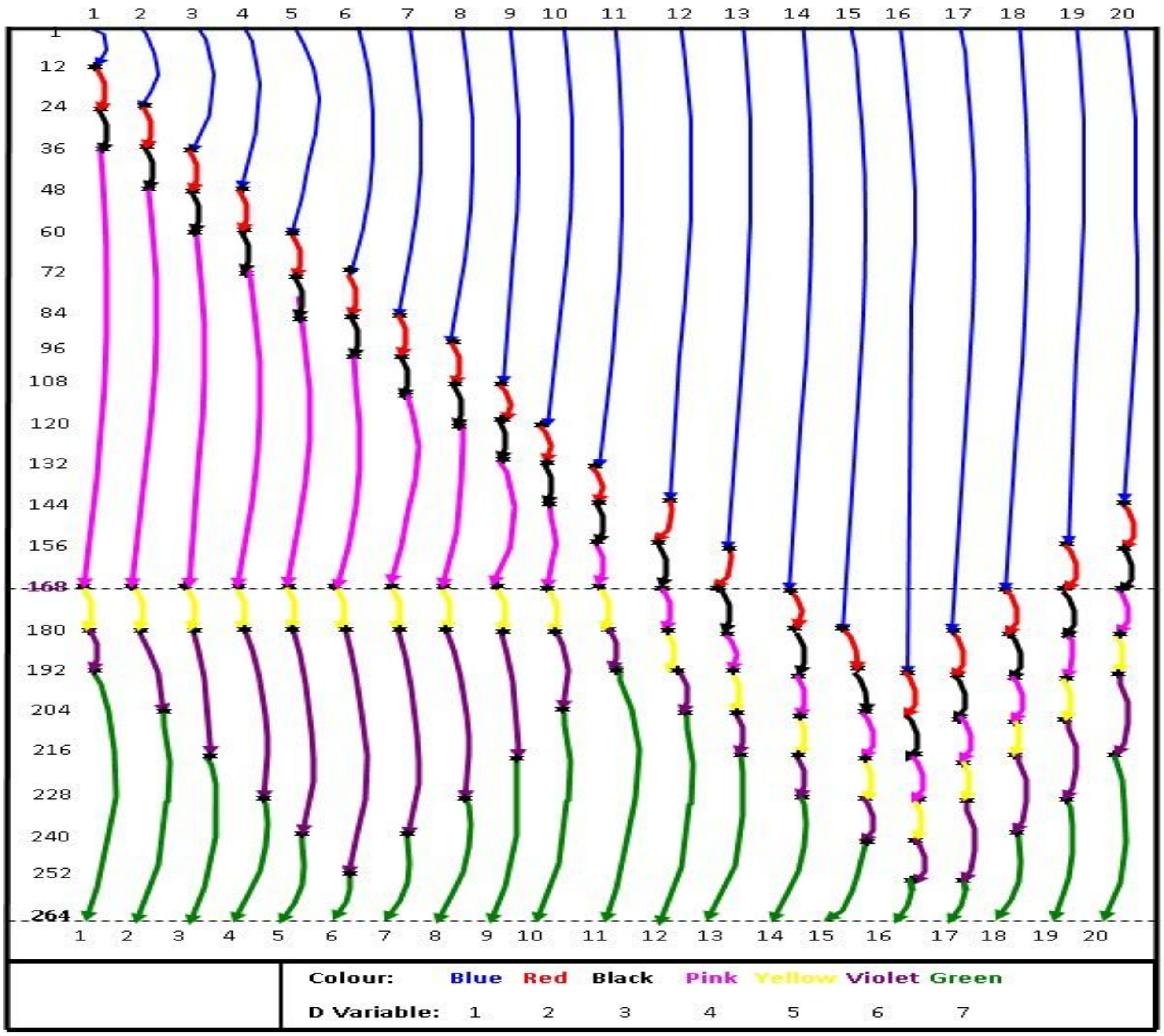

Figure 3. Grouping Scheme pictorial map for the 7-group discriminant canonical correlation function

Table 5 depicts the statistical measures of 7-group DCCA. The table shows that the measures of each grouping scheme are more optimal than the corresponding values of the 2-group DCCA in Table 4. The grouping scheme fourteen (GS_{14}) reports the most optimal values for all measures, which also fall in the 168th month. See Appendix for Table 5.

Table 6 summarizes grouping schemes with optimal values for all *m*-group DCCA for *m* = 2, 3, 4, ..., 16 groups. Since the best GS may not report the highest correct classification, the correct classification is not provided in the table.

The table confirms that 7-group DCCA outperforms all the other grouping schemes.

Table 6. Summary Statistic Measures for all Best Multiple-Group Discriminant Canonical Correlation Functions

Group	FDCCC	Eigenvalue	Wilks Lambda	Chi-Square	GS	Month	Year

2	0.907	4.627	0.178	447.452	14	168th	2013
3	0.934	6.854	0.105	582.402	4	48th	2003
4	0.948	8.884	0.074	672.814	4	48th	2003
5	0.967	14.375	0.045	799.903	14	168th	2013
6	0.961	12.123	0.048	777.800	14	168th	2013
7	**0.968**	**14.928**	**0.032**	**881.004**	**14**	**168th**	**2013**
8	0.956	10.601	0.046	789.584	14	168th	2013
9	0.964	13.190	0.047	784.399	14	168th	2013
10	0.965	13.628	0.047	791.777	5	60th	2004
11	0.965	13.715	0.040	866.505	4	48th	2003
12	0.962	12.294	0.038	808.021	14	168th	2013
13	0.963	12.924	0.036	828.047	14	168th	2013
14	0.965	13.669	0.034	876.168	14	168th	2013
15	0.964	12.996	0.034	841.741	4	48th	2003
16	0.962	12.385	0.043	752.701	4	48th	2003

Source: Researchers' computations (2024)

Out of all the *m*-groups, the grouping scheme fourteen (GS_{14}) of 7-group DCCA gives the most optimal statistic for all four measures. Again 60% $\left(\frac{9}{15}\right)$ of the possible *m*-group DCCA yields GS_{14} as the best grouping scheme. The results further indicate that the discriminant correlation between the subset of heating variables and the subset of cooling variables improves from 0.901 to 0.968 when the time impact is considered.

4.3 Simulation Studies

This section conducts simulation studies to assess the proposed procedure's effectiveness in reproducing the original series' features in Figure 1. We first examine the case of the time-independent data followed by that of the time-dependent data structure.

4.3.1 Simulation studies of time-independent data structure

The parameters for simulation in the case of the time-independent data are presented in Table 7, where n represents the total number of observations, d represents the mean of the variable, and T represents the standard deviation of data over the entire time.

Table 7. Parameters for Time-Independent Simulation

Summary Statistic	n	d	T	**Min. Value**	**Max. Value**
Maximum Temperature	264	27.281	1.761	22.250	32.285
Minimum Temperature	264	26.500	1.612	21.897	31.079
Solar Radiation	264	22.071	4.555	9.061	35.014
Precipitation	264	1.779	2.786	-6.168	9.704
Wind	264	5.793	0.978	3.000	8.571
Relative Humidity	264	0.824	0.040	0.708	0.938

Source: Researchers' computations (2024)

The plots of the monthly time-independent simulated weather conditions data are shown in Figure 4.

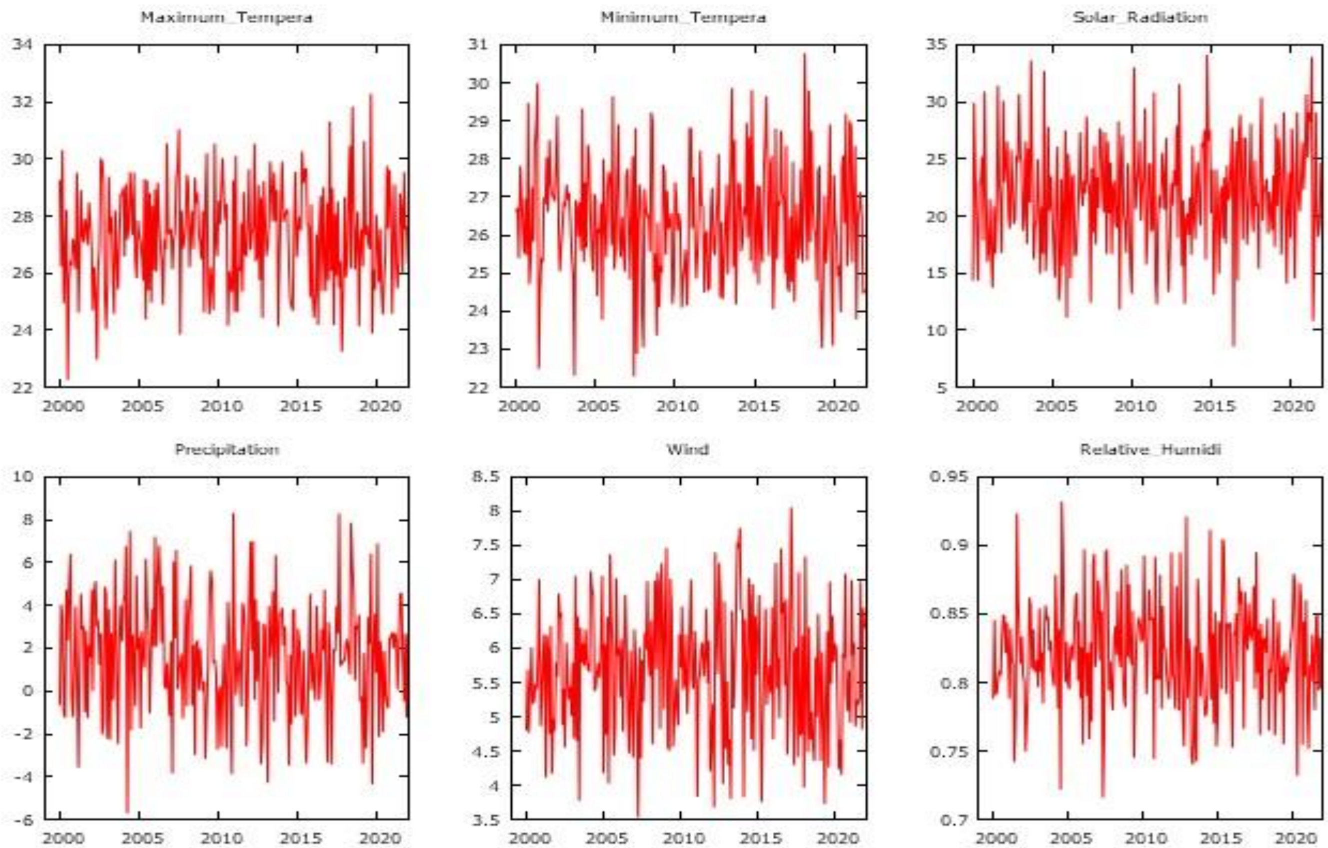

Figure 4. Series plots of data on six monthly time-independent simulated weather conditions

It can be seen that the data exhibit random changes with no noticeable patterns for all variables. The pattern appears the same for all variables, with generally wide variation from the mean. Thus, the pattern deviates from the original series in Figure 1 for most variables. This observation shows that the canonical correlation coefficients of the time-independent data may not reflect reality. Table 8 shows the overall time-independent simulated values of all statistic measures.

Table 8. Overall Tests of Significance for Statistic Measures of the Time-Independent Data

Test Statistic	**Hotelling's**	**Pillai's**	**Wilks Lambda**	**Roy's**	**Eigenvalue**	**CCC**
Value	0.510	1.039	0.491	0.510	1.039	0.714
Sign	0.000	0.000	0.000	0.000	0.000	0.000

Source: Researchers' computations (2024)

The simulated data is used to generate 2-group to 16-group DCCA to check if the results are better than the overall results in Table 8. Results beyond the 16-group DCCA were not found to be interesting; hence, they were omitted. Table 9 summarizes grouping schemes with optimal values for all *m*-group DCCA for m = 2, 3, 4, ..., 16 grouping schemes based on time-independent simulation. The table confirms that the grouping scheme fourteen (GS_{14}) of 7-group DCCA has the optimal values. Here, 40% $\left(\frac{6}{15}\right)$ of the possible *m*-group DCCA yields GS_{14} as the best scheme as shown in Table 9 compared with 60% $\left(\frac{9}{15}\right)$ in the original data in Table 6. See Appendix for Table 9.

4.3.2 *Simulation studies of time-dependent data structure*

This part of the simulation assumes a time-dependent data structure that incorporates the time effect identified in the 7-group DCCA in Table 5. The series plots of the simulated time-dependent monthly weather conditions data are shown in Figure 5. Table 10 presents the parameters for the time-dependent simulation data. See Appendix for Table 10.

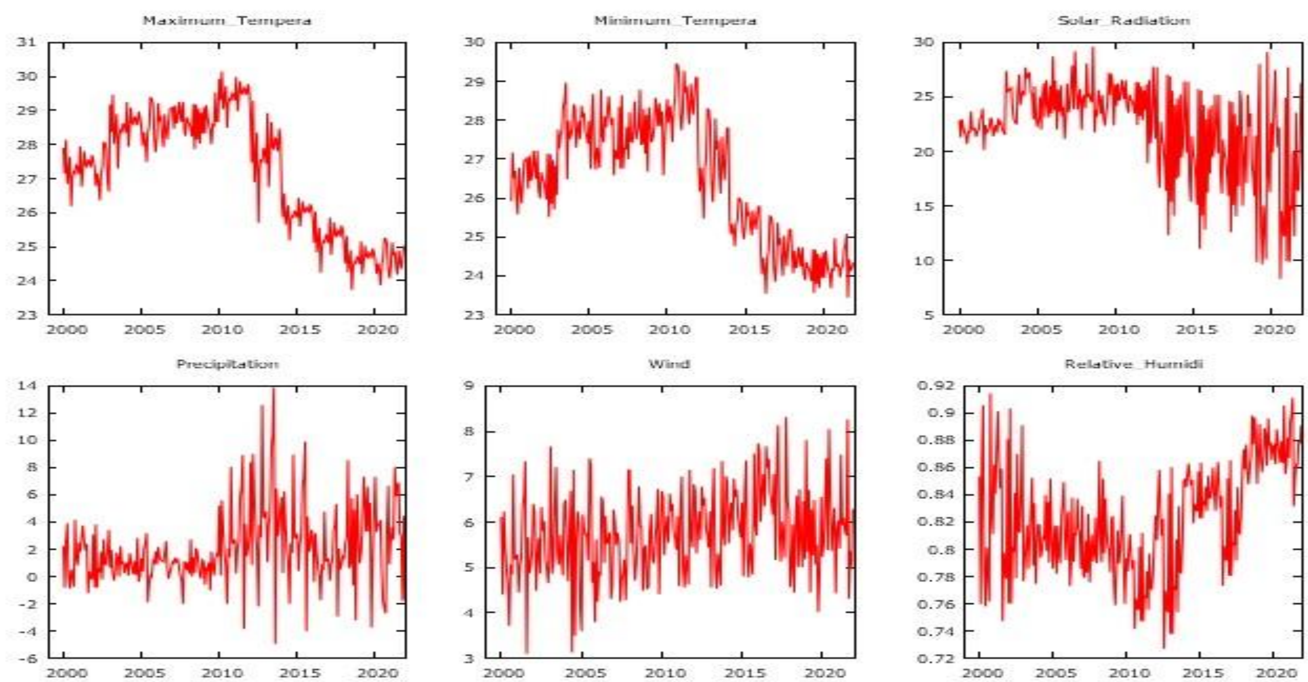

Figure 5. Series plots of data on six monthly time-dependent simulated weather conditions

All six variables, especially the two temperature variables, are seen to follow the pattern in the original data. This implies that the canonical correlation coefficients based on the time-dependent data reflect reality.

Table 11 shows the overall discriminant canonical correlation analysis statistic measures for time-dependent simulated data. The simulated data is used to generate 2 to 16-group discriminant canonical correlation analysis and to determine the results that are better than the overall results in Table 11.

Table 11. Overall Tests of Significance for Statistic Measures of the Time-Dependent Data

Test Statistic	**Hotelling's**	**Pillai's**	**Wilks Lambda**	**Roy's**	**Eigenvalue**	**CCC**

Value	0.785	3.655	0.215	0.785	3.655	0.886
Sign	0.000	0.000	0.000	0.000	0.000	0.000

Source: Researchers' computations (2024)

Table 12 confirms that 7-group DCCA with grouping scheme fourteen (GS_{14}) produces the most optimal results. It is also found that 67% $\left(\frac{10}{15}\right)$ of the possible *m*-group DCCA yields GS_{14} as the best scheme. It can also be observed that the values of all the statistic measures in Table 12 are consistent with the corresponding values in Table 6 for the original time-dependent data. The results support the observation that incorporating the time component improves the correlation between the two subset weather variables.

Table 12. Summary Statistic Measures for All Best DCC Functions Based on Time-Dependent Simulation

Group	FDCCC	Eigenvalue	Wilks Lambda	Chi-Square	GS	Month	Year
2	0.903	4.413	0.185	437.42	14	168th	2013
3	0.937	7.181	0.117	554.65	3	36th	2002
4	0.950	9.317	0.089	625.48	3	36th	2002
5	0.968	14.959	0.050	773.70	14	168th	2013
6	0.958	11.092	0.064	707.63	12	144th	2011
7	**0.980**	**24.350**	**0.022**	**975.57**	**14**	**168th**	**2013**
8	0.966	14.101	0.046	787.56	14	168th	2013
9	0.971	16.291	0.039	826.67	14	168th	2013
10	0.960	11.638	0.050	765.34	5	60th	2004
11	0.961	12.154	0.049	767.57	14	168th	2013
12	0.954	10.027	0.058	723.09	14	168th	2013
13	0.953	9.903	0.057	725.12	14	168th	2013
14	0.967	14.363	0.036	837.06	3	36th	2002
15	0.956	10.577	0.048	769.37	14	168th	2013
16	0.962	12.530	0.039	812.25	14	168th	2013

Source: Researchers' computations (2024)

Table 13 summarizes the results of the optimal values of 7-group DCCA for the original data, the time-independent simulation data and the time-dependent simulation based on GS_{14}.

Table 13. Summary Statistic Measures of Optimal 7-group Discriminant CCF for all Data

Statistic Measures	Original data	Time-independent data	Time-dependent data
FDCCC	0.968	0.252	0.980
Eigenvalue	14.928	0.065	24.350
Wilks Lambda	0.032	0.875	0.022
Chi-square	881.004	34.106	975.570

Source: Researchers' computations (2024)

Table 14 presents the canonical correlation coefficients for each of the seven time periods of the 7-group discriminant canonical correlation analysis of the time-dependent data based on the GS_{14} along with their mean values. It is clear from the table that the correlation between heating and cooling climate variables could be as high as 0.917. The results, therefore, show that incorporating the time effect into canonical correlation analysis achieves a more reasonable relationship between subset variables within the data.

Table 14. Time-dependent CCC using Group fourteen of 7-group DCCF

CCC	ρ_1	ρ_2	ρ_3
Λ_1	0.660	0.503	0.135
Λ_2	0.696	0.265	0.139
Λ_3	0.832	0.302	0.157
Λ_4	0.835	0.328	0.218
Λ_5	**0.917**	0.347	0.141
Λ_6	0.640	0.246	0.169
Λ_7	0.837	0.482	0.308
Average	**0.774**	**0.353**	**0.181**

Source: Researchers' computations (2024)

5. Conclusions

To arrive at a specific grouping scheme for a given multiple-group discriminant analysis, the method described in this paper has shown that the inclusion of the time effect in canonical correlation analysis aids in the determination of more realistic correlation coefficients between subsets of variables. It has been noted that significant changes resulting from extreme observations made at different times may impact the outcomes of the suggested technique. Identifying the key matrices that cause the necessary modification of the original variables is essential to understand the process of extracting canonical variables. Additionally, generalized correlations between these important matrices were found in the study. Six main techniques that usually cover the independence of the new variables and unit variance have been used to analyze and explain the theoretical properties of the new canonical variables. It has been demonstrated that the inverse matrix must combine the matrices that generate the required canonical variables. The study identifies the proper division, dataset structure, and pertinent matrices that allow us to arrive at the desired theoretical outcome. The ability to extract canonical variables from unprocessed, centred, or normalized data has been demonstrated. Multivariate multiple time-dependent and CCA were the main subjects of the study. To characterize the three separate data structures, the generic form of multivariate multiple time-dependent has been described in three different ways. The literature doesn't consider the data format for multivariate multiple time-dependent applications. Even though the benefits of expanded

CCA methods have been shown in individual research, a thorough comparison of CCA and DA methods is still lacking. From the illustrative dataset, higher overall correlation coefficients are obtained for the two sets of variables when the time-dependent structure is considered than when the data is assumed to be time-independent. In particular, correlations could be much higher between the two sets of variables for some years than others. The results, therefore, reflect reality and justify the technique adopted. Further research could improve the procedure by controlling the extremes in the data.

5. 1 Recommendation for Future Research

The study unequivocally shows that the proper protocols for multivariate multiple time dependencies are being followed. The outcomes of this study have shown how beneficial the suggested Grouping Scheme Discriminant Canonical Correlation Analysis is. This will lay the groundwork for experimental validation and verification and provide additional insight into future grouping scheme methods. It is demonstrated that multivariate multiple time-dependent data structures may not be appropriate for the conventional Canonical Correlation Analysis. For now, there is considerable time for interaction with the data using the proposed methodology. Therefore, the Grouping Scheme Discriminant Canonical Correlation Analysis procedure requires some enhancement to reduce implementation time. The approach could also serve as a fundamental step for obtaining what may be known as a "Fundamental Multivariate Canonical Time Series Modelling (FMCTSM)". By this, it should be possible to determine canonical correlations between the two variables over time.

Consent: Not applicable.

Ethical Approval: Not applicable.

Computing Interest

The authors have declared that no potential computing interests exist concerning this work's authorship, research, or publication.

References

[1] Samarov, D.V. The analysis and advanced extensions of canonical correlation analysis. Chapel Hill, 2009. Available at: https://core.ac.uk.

[2] Yang, X., Liu, W., Wei, L. and Tao, D. A survey on canonical correlation analysis. Information Sciences, 2021, 33(6), 2349–2368. DOI: 10.1109/TKDE.2019.2958342.

[3] Xin, L., Lei, Z. and Zhang, Y. Multi-view discriminative fusion on canonical correlation analysis in event classification. IEEE Explore, 2023. DOI: 10.1109/ICCRD56364.2023.10080508.

[4] Kabir, A., Merrill, R.D., Abu, A.S. and Rolf, D.W. Canonical correlation analysis of infant's size at birth and maternal factors: A study in rural Northwest Bangladesh. PLoS ONE, 2014.

[5] Roungu, A.M., Matair, R., Sanwar, H. and Azizur, R. Canonical correlation analysis - an application to bank performance and consumers satisfaction. International Journal of Advanced Scientific and Technical Research, 2013, 3(1). Available at: http://www.rspublication.com/ijst/index.html.

[6] Purwadi, J., Gumelar, B., Widiantoro, T. and Ningsih, Z.N. Canonical correlation analysis of economic growth and unemployment rate. BAREKENG: Journal of Mathematics and Applications, 2024, 18(2), 1273-1282.

[7] Casa, A., Bouveyron, C., Erosheva, E. and Menardi, G. Co-clustering of time-dependent data via the shape invariant model. Journal of Classification, 2021, 38, 626–649. DOI: 10.1007/s00357-021-09402-8.

[8] Xuefei, C., Jun, K., Sandstede, B. and Luo, C. Time-dependent canonical correlation analysis for multilevel time series, 2019. DOI: 10.1101/650101.

[9] Chu, D., Liao, L.Z., Ng, M.K. and Zhang, X. Sparse canonical correlation analysis: new formulation and algorithm. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2013, 35(12), 3050–3065.

[10] Abdeldjalil, A.E.B. and Abd-Krim, S. Sparse canonical correlation analysis based on rank-1 matrix approximation and its application for fMRI signals. 41st IEEE International Conference on Acoustics, Speech and Signal Processing, Shanghai, China, 2016, 4678-4682. DOI: 10.1109/ICASSP.2016.7472564.

[11] Gao, L., Qi, L. and Guan, L. Discriminative multiple canonical correlation analysis for information fusion, 2021.

[12] Shelley, M. Multivariate techniques for dichotomous dependent variables: An application to public policy. Public Administration and Public Policy, 2007.

[13] Zhou, Z., Davoud, A.T., Bojian, H., Tong, B., Xu, J., Yanbo, F., Long, Q. and Shen, L. Fair canonical correlation analysis. 37th Conference on Neural Information Processing Systems (NeurIPS), 2023.

[14] Ashish, S. and Dale, E.S. Pattern matching in multivariate time series databases using a moving-window approach. Industrial & Engineering Chemistry Research, 2022.

[15] Akour, I., Al Rahamneh, A., Al Kurdi, B., Alhamad, A., Al-Makhariz, I., Alshurideh, M. and Al-Hawary, S. Using the canonical correlation analysis method to study students' levels in face-to-face and online education in Jordan. Information Sciences Letters, 2023. DOI: 10.18576/isl/120229.

[16] Zuobin, W., Kezhi, M. and Ng, G.W. Effective feature fusion for pattern classification based on intra-class and extra-class discriminative canonical correlation analysis. In Proceedings, 2017, 1–8. DOI: 10.23919/ICIF.2017.8009795.

[17] Shu, H., Qu, Z. and Zhu, H. D-GCCA: Decomposition-based generalized canonical correlation analysis for multi-view high-dimensional data. License: CC-BY 4.0, 2022. Available at: http://jmlr.org/papers/v23/20-021.html.

[18] Johnson, R.A. and Wichern, D.W. Applied multivariate statistical analysis. 6th ed., Pearson Prentice Hall, Upper Saddle River, New Jersey, 2007.

[19] Kim, J., Kim, Y., Lee, J. and Lee, S. Post Hotelling's T-Squared procedure to identify fault variables. Journal of Statistical Computation and Simulation, 2023, 94(1), 1-28. DOI: 10.1080/00949655.2023.2228958.

[20] Rathbun, S., Andrew, W., Srabashi, B., Roths, S. and Romer, M. Applied multivariate statistical analysis. STAT 505, Elberly College of Science, Pennsylvania State University, 2023.

[21] Zongliang, H., Tiejun, T. and Marc, G.G. Diagonal likelihood ratio test for equality of mean vectors in high-dimensional data. Biometrics, 2019.

[22] Cankaya, S., Balkaya, A. and Karaagac, O. Canonical correlation analysis for the determination of relationships between plant characters and yield components in red pepper. Bulgarian Journal of Agricultural Science, 2011.

APPENDIX

Table 5. Classification Measures of 7-group Discriminant Canonical Correlation Function

GS	FDCCC	Eigenvalue	W. Lambda	Chi-Square
1	0.950	9.347	0.059	727.005
2	0.951	9.509	0.050	766.905
3	0.946	8.534	0.047	785.478
4	0.943	7.986	0.052	757.882
5	0.943	8.090	0.058	731.603
6	0.925	5.952	0.075	663.451
7	0.931	6.524	0.071	677.124
8	0.939	7.434	0.064	704.327
9	0.939	7.399	0.057	735.468
10	0.935	6.929	0.050	767.748
11	0.937	7.148	0.053	754.499
12	0.952	9.574	0.048	778.181
13	0.956	10.717	0.055	743.672
14	**0.968**	**14.928**	**0.032**	**881.004**
15	0.947	8.614	0.050	770.788
16	0.931	6.471	0.059	724.858
17	0.935	6.963	0.066	697.764
18	0.942	7.815	0.068	689.833
19	0.945	8.386	0.058	728.640
20	0.938	7.264	0.058	728.935

Source: Researchers’ computations (2024)

Table 9. Summary Statistic Measures for all best DCC Functions Based on Time-Independent Simulation

Group	FDCCC	Eigenvalue	Wilks Lambda	Chi-Square	GS	Month	Year
2	0.163	0.027	0.973	7.014	3	36th	2002
3	0.214	0.048	0.945	14.532	4	48th	2003
4	0.215	0.048	0.924	20.895	4	48th	2003
5	0.240	0.061	0.897	28.097	14	168th	2013
6	0.241	0.061	0.885	31.320	13	156th	2012
7	**0.252**	**0.065**	**0.875**	**34.106**	**14**	**168th**	**2013**
8	0.224	0.053	0.877	33.592	14	168th	2013
9	0.203	0.043	0.895	28.283	11	132nd	2010
10	0.166	0.028	0.928	19.019	5	60th	2004
11	0.216	0.049	0.904	25.617	4	48th	2003
12	0.202	0.042	0.901	26.575	14	168th	2013
13	0.211	0.046	0.886	30.823	14	168th	2013
14	0.197	0.040	0.879	32.521	14	168th	2013

15	0.192	0.038	0.914	22.631	4	48th	2003
16	0.208	0.045	0.901	26.234	4	48th	2003

Source: Researchers' computations (2024)

Table 10. Parameters for Time-Dependent Simulation

Dummy Variable	**Summary Statistic**	**Maximum Temp.**	**Minimum Temp.**	**Solar Radiation**	**Precipitation**	**Wind**	**Relative Humidity**
1	n_1 d T_1	168 28.3936 0.8371	168 27.4733 0.8919	168 23.6454 2.8411	168 1.5621 2.5036	168 5.5853 0.9600	168 0.8052 0.0340
2	n_2 d T_2	12 26.2167 0.2435	12 25.6392 0.2788	12 19.0975 6.2086	12 2.5600 3.1881	12 6.0675 0.7672	12 0.8450 0.0144
3	n_3 d T_3	12 25.8606 0.2349	12 25.3883 0.2862	12 20.0675 4.0123	12 0.5092 3.6183	12 5.9283 0.8360	12 0.8392 0.0108
4	n_4 d T_4	12 25.5267 0.3678	12 25.0075 0.5557	12 20.2108 4.3460	12 0.4183 0.6053	12 6.8758 0.8802	12 0.8133 0.0287
5	n_5 d T_5	12 25.1575 0.3160	12 24.5017 0.2335	12 18.9542 4.1757	12 1.4875 2.6869	12 6.2342 0.8720	12 0.8283 0.0233
6	n_6 d T_6	12 24.6850 0.2452	12 24.833 0.2203	12 17.6292 5.1353	12 1.4158 1.8420	12 6.2825 0.5686	12 0.8600 0.0095
7	n_7 d T_7	36 24.7079 0.3680	36 24.1875 0.4357	36 18.5814 5.8268	36 2.4558 3.6634	36 5.7525 0.8422	36 0.8783 0.0161

Source: Researchers' computations (2024)