Journal of Geospatial Science and Analytics

Journal homepage:
https://ejournal.foundae.com/index.php/jgsa/index
e-ISSN: 3090-8086 | p-ISSN: 3090-532X

Spatial ecological determinants of infectious diseases using National Health Insurance data: A multivariate canonical correlation analysis

Rafli Akbar Ramadhan
Indonesia Defense University, Indonesia

Purnomo Husnul Khotimah*
National Research and Innovation Agency (BRIN), Indonesia

Rifani Bhakti Natari
National Research and Innovation Agency (BRIN), Indonesia

Dianadewi Riswantini
National Research and Innovation Agency (BRIN), Indonesia

Devi Munandar
National Research and Innovation Agency (BRIN), Indonesia

Muh. Hafizh Izzaturrahim
National Research and Innovation Agency (BRIN), Indonesia

Article Info

Article history:

Received: July 01, 2025
Revised: August 05, 2025
Accepted: September 01, 2025

Keywords:

Canonical Correlation
Climate
Environment
Infectious diseases
National Health Insurance (BPJS)

DOI:
10.58524/jgsa.v1i3.56

Abstract

Uncertain climate change and population growth, which increase every year, constitute a varied lifestyle that can escalate the incidence of some infectious diseases. In West Java, the incident of dengue hemorrhagic fever (DHF), malaria, and pneumonia, along with the number of claims under National Health Insurance (BPJS) for these diseases, increased in 2022. This research aims to explore the relationship between ecological factors and the incidence of infectious diseases using multivariate canonical correlation analysis. The response variables are the incidence of DHF, malaria, and pneumonia based on disease-related visits to healthcare facilities captured in the BPJS data sample. The ecological factors used as explanatory variables include population size, average humidity, average rainfall, average temperature, the amount of waste transported to landfills per ton, and the percentage of households with access to adequate sanitation. The results showed a high correlation between ecology and disease incidence. Based on canonical loading and cross loading, the ecological factors that significantly contribute to disease incidence are population, average rainfall, average temperature, and the amount of waste transported to landfills per ton. Statistically, reduction in waste sent to landfills can decrease the incidence of DHF, malaria, and pneumonia. Reduction in population can reduce disease incidence, and decrease in temperature can lower disease incidence. Meanwhile, increase in rainfall can also reduce disease incidence. Therefore, efforts to control population growth, improve access to proper sanitation, and implement effective waste management can have a positive impact on reducing the incidence of infectious diseases in West Java. Therefore, efforts to control the population, improve access to proper sanitation, and effective waste management can have a positive impact on reducing infectious disease incidence in West Java.

***Corresponding Author:**
Purnomo Husnul Khotimah, National Research and Innovation Agency (BRIN), INDONESIA,
Email: purn005@brin.go.id

To cite this article:
Ramadhan, R. A., Khotimah, P. H., Natari, R. B., Riswantini, D., Munandar, D., & Izzaturrahim, M. H. (2025). Spatial ecological determinants of infectious diseases using National Health Insurance data: A multivariate canonical correlation analysis. *Journal of Geospatial Science and Analytics,* 1(3), 203–214.

INTRODUCTION

Indonesia, as a tropical country, often faces complex public health challenges. West Java Province, one of Indonesia's economic and population centers, faces complex challenges related to public health and environmental sustainability. Ecological factors such as weather and environmental changes deeply affect the health conditions of the population (de Claro et al., 2024; Lebedeva & Shkuropadska, 2024). Climate change increases the risk of disease through increased temperatures, xtreme rainfall, and other ecological factors that can affect the health profile of present and future generations (Abdullah et al., 2022; de Souza Fernandes Duarte et al., 2024; Komara et al., 2024). In 2022, the use of BPJS in the West Java province increased significantly, especially in diseases caused by mosquitoes and the air. However, in 2020 BPJS services related to diseases caused by mosquito vectors and air tend to decrease, this may be due to the COVID-19 season, which allows people to be more aware of the surrounding environment and climates.

Tabel 1. Summary Data

Variable	Min	Max	Mean	Stdv
Population 2022	2.06	55.66	18.09	12.05
Population 2021	2.03	54.80	18.00	11.88
Population 2020	1.83	60.80	18.40	13.26
Population 2019	1.81	59.60	18.20	12.96
Humidity 2022	78.46	88.71	84.12	2.76
Humidity 2021	77.09	88.24	83.11	2.79
Humidity 2020	76.95	89.47	83.70	3.19
Humidity 2019	72.86	86.97	79.71	3.75
Rainfall 2022	6.86	11.33	9.09	1.16
Rainfall 2021	7.00	9.32	8.12	0.65
Rainfall 2020	6.58	11.62	8.94	1.49
Rainfall 2019	4.70	9.14	6.44	1.10
Temperature 2022	25.79	27.30	26.63	0.40
Temperature 2021	25.75	27.29	26.63	0.41
Temperature 2020	25.68	27.49	26.71	0.50
Temperature 2019	24.98	27.56	26.57	0.76
Amount of Waste Transported to Landfills 2022	48.37	1444.34	653.86	389.82
Amount of Waste Transported to Landfills 2021	29.70	1430.00	653.86	355.51
Amount of Waste Transported to Landfills 2020	57.48	2173.05	414.40	492.86
Amount of Waste Transported to Landfills 2019	57.48	2173.05	402.56	492.86
Sanitation 2022	45.80	96.21	71.48	14.96
Sanitation 2021	39.64	97.54	70.57	16.20
Sanitation 2020	34.93	97.48	70.81	17.30
Sanitation 2019	39.79	96.30	68.64	17.66

According to the West Java Central Bureau of Statistics (BPS), Open Data Jabar of West Java and Copernicus Climate Data Store in Table 1 show that population, humidity, rainfall, temperature, amount of waste, and sanitation conditions experience significant variations from year to year. The high standard deviation indicates that these data vary significantly, while the mean value gives an idea of the general trend of each variable over the observed period. Significant variations in these data indicate that there are large fluctuations in these variables from year to year. For example, the population variable shows considerable changes in the range of minimum and maximum values from 2019 to 2022. This indicates the dynamics of population growth or decline within this timeframe.

Population density plays a prominent roles in the increase of public health diseases (Oliveira et al., 2023). Population growth that increases every year is sometimes not supported by adequate infrastructure. such as proper sanitation. Poor sanitation can lead to environmental and water contamination, which in turn increases the risk of spreading infectious diseases (Ahmed et al., 2019). Not only that, erratic and extreme rainfall conditions create stagnant water for a long time, making it a perfect place for mosquitoes lay eggs and breed (Al-Osaimi et al., 2024). After periods of heavy rainfall, there is often an increase in the number of dengue cases due to the increased mosquito population. Varies rainfall also creates an environment with varying levels of air humidity and temperature, indirectly affecting public health. High average air humidity can create suitable conditions for the proliferation of disease vectors such as mosquitoes. Mosquitoes are a major vector in the transmission of diseases such as dengue hemorrhagic fever (DHF) and malaria (Kumar, 2024). When air humidity is high, mosquitoes tend to be more active and breed faster, increasing the likelihood of disease transmission. High humidity can worsen the conditions of people with asthma and other respiratory diseases.

In addition, air temperature is also one of the main ecological factors, playing a significant role in human health. One disease that is strongly affected by the temperature is dengue hemorrhagic fever (DHF). High temperatures accelerate the life cycle of the Aedes aegypti mosquito, the main vector of DHF, from egg to adult stage (Oyarzún G et al., 2021). Mosquitoes breed faster in hot conditions, and high temperatures also accelerate dengue virus replication in the mosquito body. At temperatures above the optimum temperature of 32-35 °C according to Ministry of Health, (2012), the life cycle of the Aedes mosquitos become shorter, on average only 7 days. This makes mosquitoes breed more often. In addition, the size of the mosquito is different from usual, making its movements more aggressive. Not only vector-borne diseases, but temperature also affects respiratory diseases such as pneumonia. Extreme temperature fluctuations, especially sudden transitions from hot to cold temperatures or vice versa, can lower the immune system and make individuals more susceptible to respiratory infections as well as influenza-associated pneumonia at low ambient temperatures (Amorosi et al., 2024; He et al., 2023). Climate change could increase the prevalence of certain diseases, such as dengue hemorrhagic fever (DHF), malaria, skin diseases, respiratory diseases due to poor air quality, and other infectious diseases. As a consequence, the burden of health services and claims on BPJS will increase as people need better access to health services to manage these impacts. BPJS health data shows that the prevalence of infectious diseases such as dengue hemorrhagic fever (DHF), malaria, and pneumonia vary from year to year. Taking four years of data from the latest BPJS utilization data, the highest level of DHF service in the most recent BPJS health records were in 2022, which was 1356. In terms of the prevalence of malaria, 464 people sought treatment for malaria. The prevalence of pneumonia in BPJS health data also varies in the number of services, and the highest with the most recent BPJS health service records in 2022 was 2432.

To determine the factors that contribute to the spread of diseases caused by climate and environment, canonical correlation analysis was used (Osborne et al., 2024; Siti Rahmawati Hindo et al., 2018). Using canonical correlation analysis, according to research conducted by Hindo et al in 2018 said that the environment is highly correlated with the increase of pneumonia in both under-fives and severe pneumonia. However, of the many factors in environmental prevalence, there is one factor that has a high correlation with clean drinking water sources (Viyana, 2019).

Meanwhile, according to research conducted by Ria in 2019 using canonical correlation analysis, climate is also highly correlated with the incidence of non-respiratory ARI and dengue hemorrhagic fever (DHF), however, from the correlation of the two prevalences, it is the wind speed factor that has a tight correlation with non-pneumonia ARI (Lydia H V Franklinos, MSc et al., 2019). Research done by Frima in 2022 using the canonical correlation method said that climate has a high correlation with the prevalence of malaria and dengue hemorrhagic fever (DHF), especially in the factors of high air temperature and rainfall, which have a high impact on the incidence of malaria (Winkler et al., 2020). According to the background above, weather uncertainty is one of the factors that can cause the emergence of unexpected diseases (Asbah & Safitri, 2019). Therefore, this research will conduct a canonical correlation analysis to determine the correlation of ecological prevalence, which includes climate and environmental factors, to public health diseases including dengue hemorrhagic fever (DHF), malaria, and pneumonia with the help of R software. Ecological prevalence in this research is aided by bringing up climatic factors, namely average air temperature, average humidity, and average rainfall. Meanwhile, environmental factors are described by population densities, the number of houses that have proper sanitation, and the amount of waste in each housing unit that is transported to the landfill in tons.

This study uses canonical multivariate analysis because this analysis involves more than one variable, both dependent and independent variables (Osborne et al., 2024; Siti Rahmawati Hindo et al., 2018). This research also examines how population density and community waste disposal dynamics affect the spread of disease, how variations in climate change and the effectiveness of proper sanitation programs affects the spread of disease. With some assumptions about the data that will be fulfilled, the canonical correlation analysis can be done. The final results that will be achieved are canonical weight, canonical loading, and canonical cross-loading. Bringing out the three interpretations of the canonical correlation analysis will unravel the correlation relationship between ecologies and public health diseases.

METHOD

This research is an application of the canonical correlation analysis method for measuring the proximity of the correlation among a group of related variables (diseases recorded by BPJS) and the collection of independent variables (ecological and demographic factors) in West Java in 2022. The dependent variables in this research were sourced from *Badan Penyelenggara Jaminan Sosial Kesehatan* (BPJS Kesehatan, or BPJS), a governmental agency that administers healthcare services insurance https://data.bpjs-kesehatan.go.id/bpjs-portal/action/landingPage.cbi. The dependent variables in 2022 include dengue hemorrhagic fever (DHF)(Y1), malaria (Y2), and pneumonia (Y3). Meanwhile, the independent variables include population size (X1), average humidity (X2), average rainfall (X3), average temperature (X4), amount of waste transported to landfills per ton (X5), and percentage of households with access to proper sanitation (X6). Data on the independent variables used were taken with a record of 2022 sourced from the West Java Central Bureau of Statistics (BPS) (https://jabar.bps.go.id), Open Data Jabar (https://opendata.jabarprov. go.id/id), and Copernicus Climate Data Store (https://climate.copernicus. eu/).

1. Canonical Correlation Analysis

Canonical correlation analysis serves as a statistical method utilized to understand the correlation between two sets of multivariate variables (many variables). This technique identifies and measures the association among the optimal linear combinations of two sets of variables, with the main objective of finding the pair of variables that has the highest correlation (Osborne et al., 2024; Sherry & Henson, 2005; Wickramasinghe, 2019). Suppose that the set of dependent variables comprises q dependent variables. and the set of independent variables consists of p independent variables. Then the basic form of canonical correlation analysis is:

$$y_1 + y_2 + y_3 + \dots + y_q = x_1 + x_2 + \dots + x_p. \quad (1)$$

Before doing a correlation analysis, some assumptions must be made, such as: Multivariate normality of data, Multicollinearity test, and Homoscedasticity test (Purwadi et al., 2024; Wei et al., 2021).

2. Determination of Canonical Function and Canonical Estimator

Canonical correlation analysis investigates the association between the composite linear expressions of the dependent variables $Y' = (y_1, y_2, \ldots, y_p)$. The combination of pairs of linear functions is called canonical functions, and the relationships between them are known as canonical correlations. The letter y represents dependent variables, while the letter x represents independent variables. Suppose that the number of dependent variables is denoted by q and the number of independent variables is denoted by p where $p \leq q$. Suppose Z is a matrix partitioned by $(p \times p)$ variance-covariance matrices Σ_{XX}, $(p \times p)$ variance-covariance matrices Σ_{YY}, and $(p \times q)$ and $(q \times p)$ matrix Σ_{XY} and Σ_{YX}. Therefore, its covariance matrix is

$$\Sigma_Z = \begin{bmatrix} \Sigma_{XX} & \Sigma_{XY} \\ \Sigma_{YX} & \Sigma_{YY} \end{bmatrix} \tag{2}$$

Σ_{XX} is the matrix of variance and covariance between X and X of size $(p\times p)$, Σ_{YY} is the matrix of variance and covariance between X and Y of size $(q \times q)$, Σ_{XY} is the matrix of variance and covariance between X and Y of size $(p \times q)$, and Σ_{YX} is the matrix of variance and covariance between Y and X of size $(q\times p)$. If there are p independent variables $x_1, x_2, \ldots, x_p$ and q dependent $y_1, y_2, \ldots, y_p$ then there are a minimum number of canonical variable pairs q and p. Suppose $a' = [a_1, a_2, \ldots, a_k]$ and $b' = [b_1, b_2, \ldots, b_k]$ the linear combination formed is

$$U = a_1x_1 + a_2x_1 + \cdots + a_qx_q = a_k' X, \tag{3}$$

$$V = b_1y_1 + b_2y_1 + \cdots + a_py_p = b_k'Y. \tag{4}$$

Then the canonical function k:

$$\begin{array}{ll} U_1 = a_1X & V_1 = a_1X \\ U_2 = a_2X & V_1 = a_2X \\ \vdots & \vdots \\ U_k = a_kX & V_k = a_kX \end{array} \tag{5}$$

$$\text{where } X = \begin{bmatrix} x_1 \\ | \\ x_q \end{bmatrix} \text{ dan } Y = \begin{bmatrix} y_1 \\ | \\ y_p \end{bmatrix}. \tag{6}$$

According to the matrix $\Sigma_{YY}^{-1}\, \Sigma_{YX}\, \Sigma_{XX}^{-1}\, \Sigma_{XY}$ using the eigenvectors $f_1, f_2, \ldots, f_k$ and $\Sigma_{XX}^{-1}\, \Sigma_{XY} \Sigma_{YY}^{-1}\, \Sigma_{YX}$ using the eigenvectors $e_1, e_2, \ldots, e_k$, we will obtain the coefficient vectors a and b by finding $\lambda_1^2 > \lambda_2^2 > \ldots > \lambda_k^2$. The canonical coefficient vectors a and b are obtained as follows:

$$a_k = \left(\frac{1}{\sqrt{(e_k'\Sigma_{XX}e_k)}}\right) e_k \tag{7}$$

$$b_k = \left(\frac{1}{\sqrt{(f_k'\Sigma_{YY}f_k)}}\right) f_k \tag{8}$$

with k = 1,2,3, ..., r being the minimum number of p and q. Canonical correlation U and V:

$$\rho_{(U,V)} = \frac{\text{Kov}(U,V)}{\sqrt{\text{Var(U)Var(V)}}} = \frac{a'\, \Sigma_{XY}}{\sqrt{a'\Sigma_{XX}\, a}\sqrt{b'\Sigma_{YY}\, b}} \tag{9}$$

eigenvalue:

$$\Sigma_{XX}^{-1}\, \Sigma_{XY}\, \Sigma_{YY}^{-1}\, \Sigma_{YX} - \lambda I = 0 \tag{10}$$

$\Sigma_{YY}^{-1}\, \Sigma_{YX} \Sigma_{XX}^{-1}\, \Sigma_{XY} - \lambda I = 0$. The selection of canonical functions is also considered with large to small eigenvalues. To determine the number of canonical variables to be interpreted, several criteria can be used. Several criteria should be considered and linked to each other to select the right canonical variables, including the magnitude of canonical correlation and redundancy measures for the percentage of diversity that can be explained by canonical variables and the significance level of each function.

3. Canonical Weight

Canonical weights are normalized canonical coefficients and can be interpreted as a measure of the contribution of the original variable to the canonical variable. If the sign of a variable's weight is opposite to the sign of its canonical variable, it indicates an inverse correlation with the other variable. However, canonical weights have some drawbacks the fact that it only shows the amount of contribution of the original variable to the canonical variable, making them inaccurate in describing the correlation between variables (Weinberger, 2016).

4. Canonical Loading

The original variable's correlation strength and direction with the canonical variable are indicated. A high canonical load value (>0.5) for an original variable suggests its importance in the group of variables, with the load's sign showing the correlation direction. The larger the load value, the more crucial the original variable's role in the group. The formula obtains the independent canonical loading

$$\Sigma_{XU} = \Sigma_{XX}\, ak, \tag{11}$$

the simple correlation between X and a is the canonical coefficient vector of the variate U. While the formula obtains the independent canonical loading.

$$\Sigma_{YV} = \Sigma_{YY}\, bk, \tag{12}$$

the simple correlation between Y and b is the vector of canonical coefficients of variate V. Canonical weights are more effective in interpreting the correlation between variables than canonical weights, as canonical weights have some disadvantages.

5. Canonical Cross Loading

Shows the correlation between the original variables from one set and the canonical variables from the other set, identifying interactions between sets of variables. The formula obtains the independent canonical cross-loading

$$\Sigma_{XV} = \Sigma_{XU}\, \rho k, \tag{13}$$

Σ_{XV} is the canonical charge of the independent variable and ρk is the correlation value of the k-th canonical function. The formula is used to obtain the dependent canonical cross-charge.

$$\Sigma_{YU} = \Sigma_{YV}\, \rho k, \tag{14}$$

Σ_{YU} is the canonical charge of the dependent variable and ρk is the correlation value of the k-th canonical function.

RESULTS AND DISCUSSIONS

1. Assumption Check

Before processing the data, data testing is needed to ensure the fulfillment of several assumptions, namely data normality, multicollinearity, and homoscedasticity. The QQ-Plot results in Figure 1 show that, descriptively, the data spreads double normal because the data plot or the black dots are still around the straight line. Next, multicollinearity testing is conducted by examining the correlation between variables or by checking the values of the variance inflation factor (VIF) (Frima Aji Umargani, 2022).

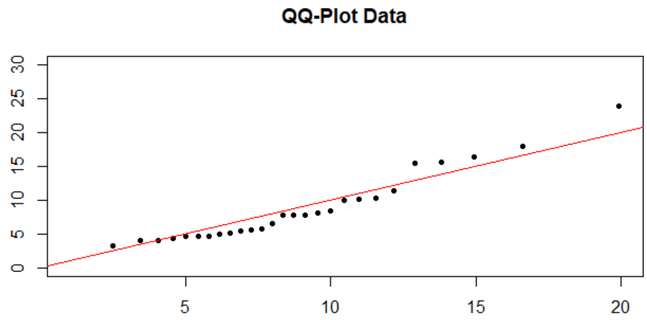

Figure 1. QQ-Plot of data distribution.

Table 2. Variance Inflation Factor (VIF)

Dimension	Variable	VIF
	Y1	3.8578
Public Health Diseases	Y2	1.7649
	Y3	1.3173
	X1	2.0069
	X2	1.4221
Ecology	X3	2.9934
	X4	2.2167
	X5	2.2074

Based on Table 2, all Varian Inflation Factor (VIF) value is lower than 10, and it is concluded that there is no multicollinearity or that the assumption of non-multicollinearity is met. In the next stage, with the help of RStudio software tests, the assumption of homoscedasticity of the data is tested by the results of the standardized Breusch-Pagan test. The p-value is 0.2266, which means there is not enough evidence to reject the assumption of homoscedasticity at the 0.05 level of significance. Therefore, we cannot conclude the presence of heteroscedasticity in the data.

2. Canonical Correlation Analysis Results

With the assumptions of the canonical correlation test fulfilled, the analysis can proceed. Data processing in the canonical correlation analysis uses the RStudio version 4.3.1 program. The minimum number of variables used determines the number of canonical functions formed. In this research, the independent variable group consists of 6 variables, while the dependent variable group only consists of 3 variables, so the canonical function forms as many as 3 canonical functions. Determination of canonical functions that can be used must consider statistical significance tests, diversity contributions, and correlations within the canonical functions.

Table 3. Canonical Function Analysis and Its Significance

CanR	CanRSQ	Eigen	Percent	Pr (> F)
0.86	0.75	30.52	78.5	0.00
0.60	0.36	0.58	15.0	0.16
0.44	0.19	0.24	6.4	0.32

The results of the program according generated a total of three functions which explain that function number 1 accommodated 78.5% of the canonical correlation, while the rest were very small and possibly meaningless. Using the minimum limitation criterion of diversity contribution above 70%, it is sufficient to take the first canonical function only. The table above shows that the canonical correlation number for the first function is 0.86 and has a redundancy coefficient of 0.75, while for functions number 2 and number 3, there is a very weak correlation with a limit of 0.5 for the strength of the correlation of two variables. The significance of function number 1 is far below 0.05, while functions number 2 and 3 are far above 0.05, so they are considered unable to be processed further into further analysis

3. Canonical Weight

The table below displays the relative contribution of the independent variable and the dependent variable to the variate variable as indicated by the value of the coefficient or canonical weight.

Table 4. Coefficient of relative contribution of independent variable group

Variable	Function 1
X1	0.001
X2	-0.036
X3	0.008
X4	-0.778
X5	-0.002
X6	-0.003

Table 5. Coefficient of the relative contribution of the dependent variable group

Variable	Function 1
Y1	-0.006
Y2	-0.040
Y3	-0.010

Based on Table 5 and 4, By ignoring the second and third functions based on the size of the coefficients, it can be concluded that the order of relative contribution from the largest to the smallest of the dependent variables is pneumonia (Y3), DHF (Y1), and malaria (Y2). In this case, the cluster of public health diseases, the pneumonia variable is the most concerning in increasing other health diseases. The order of relative contribution from largest to smallest in the dependent variable is Temperature (X4), Humidity (X2), Rainfall (X3), Proper Sanitation (X6), Total Garbage Disposal (X5), Population (X1). In this case in the ecological variable cluster, the temperature variable is most concerned about the increase in public health diseases. Then the equation Based on the canonical correlation coefficient analysis that has been carried out, a linear combination of the two variable groups is obtained as follows

$$U_1 = -0.001\, X_1 - 0.036\, X_2 + 0.008\, X_3 - 0.778 X_4 - 0.002\, X_5 - 0.003 X_6 \tag{15}$$

$$V_1 = -0.063\, Y_1 - 0.040\, Y_2 - 0.010 Y_3 \tag{16}$$

4. Loading

The canonical load reflects the relationship between the initial variable and the canonical variables or variates. According to the program that has been run, the results of the canonical load on the dependent and independent variables are as follows Shown in the table 6, the canonical charge

Table 6. Canonical charge in the canonical variable function

Dimension	Variable	Weight
Public Health Diseases	Y1	-0.911
	Y2	-0.712
	Y3	-0.977
Ecology	X1	-0.667
	X2	0.297
	X3	0.605
	X4	-0.605
	X5	-0.953
	X6	-0.003

of public health disease variables that have the closest correlation with the first canonical function are DHF (Y1), malaria (Y2), and pneumonia (Y3) with negative correlation values of -0.911, -0.712, and -0.977 respectively. Meanwhile, the canonical load of ecological independent variables that have the closest correlation with the first canonical function is the amount of waste transported to landfill (X5), population (X1), temperature (X4), and rainfall (X3) with correlation values of -0.953, -0.667, -0.605, and 0.605.

5. Canonical Cross Loading

According to the program that has been run, the canonical cross-loading contained in the first canonical function on the independent and dependent variables is as follows:

Table 7. Canonical cross-loading in variable functions

Dimension	Variable	Weight
Public Health Diseases	Y1	-0.790
	Y2	-0.618
	Y3	-0.848
Ecology	X1	-0.579
	X2	0.258
	X3	0.525
	X4	-0.525
	X5	-0.827
	X6	-0.002

As shown in the table 7 above the canonical cross-loading between the independent variables to the canonical function shows that the most closely related to the first canonical function is the amount of waste transported to the landfill (X5) with a negative correlation value of -0.827 the total population (X1) with a negative correlation value of -0.579, temperature (X4) with a negative correlation value of -0.525, and rainfall (X3) with a positive correlation value of 0.525. Meanwhile, the canonical cross-loading between the dependent variables and the canonical function shows that the most closely related to the first canonical function is the proportion of DHF (Y1) with a correlation value of -0.790, malaria (Y2) with a correlation value of -0.618, dan pneumonia (Y3) with a correlation value of -0.848.

DISCUSSION

From the output of the canonical correlation analysis, three canonical functions were found, but only the first function was significant with a correlation coefficient of 0.82. This result indicates a very close correlation between public health diseases and ecological factors such as climate, population, access to proper sanitation, and the amount of waste transported to landfills. This close correlation indicates that changes in these ecological factors significantly affect the incidence of diseases in the community.

In this research, the most dominant public health diseases were Dengue Hemorrhagic Fever (DHF), malaria, and pneumonia. The most dominant constituent ecological factors include the amount of waste transported to landfills, with a negative correlation value of -0.827, population with a negative correlation value of -0.579, temperature with a negative correlation value of -0.525, and have a significant effect on disease incidence. The combination of these variables shows how environmental conditions and demographics affect the level of public health.

When there is less waste going to landfills, less population, and lower temperatures, the incidence of dengue hemorrhagic fever (DHF), malaria, and pneumonia tends to decrease. A cleaner environment and reduced population pressure can decrease the spread of these diseases. Other findings showed that an increase in rainfall with a positive correlation value of 0.525 intensity correlated with a decrease in the incidence of dengue hemorrhagic fever (DHF), malaria, and pneumonia This may be due to the effects of excessive rainfall, which can reduce populations of disease vectors such as mosquitoes through washing and inundating their habitats, or changes in human habits during rainy periods. Although pneumonia is generally not directly linked to rainfall, high rainfall conditions can increase air humidity which may improve air quality and help reduce the risk of transmission of respiratory diseases.

CONCLUSION

Three canonical functions were found from the canonical correlation study of ecological factors on public health disease using data from 2019 to 2023. Only the first function, which had a high correlation coefficient of 0.86 and a determination coefficient of 0.75, was statistically significant, explaining 78.5% of the total variation. These findings demonstrate that ecological factors, including climate, population density, and waste management, play a crucial role in the spread of public health diseases. The study highlights that dengue hemorrhagic fever (DHF), malaria, and pneumonia are the dominant public health diseases, while key ecological factors influencing their spread include waste accumulation in landfills, population density, temperature, and rainfall. Statistically, every 1-ton reduction in waste sent to landfills can decrease the incidence of DHF, malaria, and pneumonia with a negative correlation of -0.827. Every 1 million decrease in population can reduce disease incidence with a negative correlation of -0.579. Every 1°C decrease in temperature can lower disease incidence with a negative correlation of -0.525. Meanwhile, every 1 mm/day increase in rainfall can reduce disease incidence with a positive correlation of 0.525.

Overall, this study confirms the importance of managing ecological factors to improve public health. Efforts to control population size, improve access to proper sanitation, and implement effective waste management can have a positive impact on reducing the incidence of communicable and non-communicable diseases in West Java. For future research, it is recommended to expand the scope of ecological variables, such as air quality, and other socio-economic factors that may affect public health.

AUTHOR CONTRIBUTIONS

RAR designed the research, analyzed data, and wrote the introduction, methods, and results sections; PHK and DR conceptualized the study and conducted statistical interpretation; RBN assisted in collecting disease-related health data from BPJS and provided additional interpretation; DM, PHK, and DR validated the mathematical equations; MHI collected climate and weather data for analysis using resources from Copernicus Climate Data Store. RAR = Rafli Akbar Ramadhan; PHK = Purnomo Husnul Khotimah; DR = Dianadewi Riswantini; DM = Devi Munandar; MHI = Muh. Hafizh Izzaturrahim; RBN = Rifani Bhakti Natari.

ACKNOWLEDGMENT

Gratitude is extended to the Faculty of Mathematics and Natural Sciences at Indonesian Defense University, the Research Center for Data and Information Sciences, National Research and Innovation Agency, Indonesia, and the Environmental and Public Health Research Program for their support in facilitating this research. Special appreciation is given to the West Java Health Office and the National Health Insurance (BPJS) for providing essential data sources. The authors also sincerely thank healthcare professionals, environmental specialists, academic mentors, and colleagues for their invaluable insights and constructive feedback, which significantly enriched the depth and quality of this study. This study is supported by Riset dan Inovasi untuk Indonesia Maju (RIIM) Batch 4 of National Research and Innovation Agency (BRIN) and Indonesia Endowment Fund for Education Agency (LPDP); Contract Number: B-3836/II.7.5/FR.06.00/11/2023.

REFERENCES

Abdullah, N. A. M. H., Dom, N. C., Salleh, S. A., Salim, H., & Precha, N. (2022). The association between dengue case and climate: A systematic review and meta-analysis. *One Health*, *15*, 100452. https://doi.org/10.1016/j.onehlt.2022.100452

Ahmed, T., Hyder, M. Z., Liaqat, I., & Scholz, M. (2019). Climatic conditions: conventional and nanotechnology-based methods for the control of mosquito vectors causing human health issues. *International Journal of Environmental Research and Public Health*, *16*(17), 3165. https://doi.org/10.3390/ijerph16173165

Al-Osaimi, H. M., Kanan, M., Marghlani, L., Al-Rowaili, B., Albalawi, R., Saad, A., Alasmari, S., Althobaiti, K., Alhulaili, Z., Alanzi, A., Alqarni, R., Alsofiyani, R., & Shrwani, R. (2024). A systematic review on malaria and dengue vaccines for the effective management of these mosquito borne diseases: Improving public health. *Human Vaccines & Immunotherapeutics*, *20*(1), 2337985. https://doi.org/10.1080/21645515.2024.2337985

Amorosi, L., Padellini, T., Puerto, J., & Valverde, C. (2024). A mathematical programming approach to sparse canonical correlation analysis. *Expert Systems with Applications*, *237*, 121293. https://doi.org/10.1016/j.eswa.2023.121293

Asbah, M. F., & Safitri, D. (2019). Penentuan koefisien korelasi kanonik dan interpretasi fungsi kanonik multivariat. *Jurnal Gaussian*, *2*(2), 119–128. https://doi.org/10.14710/j.gauss.2.2.119%20-%20128

de Claro, V., Lava, J. B., Bondoc, C., & Stan, L. (2024). The role of local health officers in advancing public health and primary care integration: Lessons from the ongoing universal health coverage reforms in the Philippines. *BMJ Global Health*, *9*(1), e014118. https://doi.org/10.1136/bmjgh-2023-014118

de Souza Fernandes Duarte, E., Lucio, P. S., Costa, M. J., Salgueiro, V., Salgado, R., Potes, M., Hoelzemann, J. J., & Bortoli, D. (2024). Pollutant-meteorological factors and cardio-respiratory mortality in Portugal: Seasonal variability and associations. *Environmental Research*, *240*(Pt 2), 117491. https://doi.org/10.1016/j.envres.2023.117491

Frima Aji Umargani, N. : 18106010014. (2022). *Perbandingan analisis korelasi kanonik antara metode klasik dan robust (estimasi m) (studi kasus: pengaruh hubungan perubahan iklim dengan penyakit malaria dan dbd di indonesia tahun 2020)* [Thesis, UIN Sunan Kalijaga Yogyakarta].

He, Q., Liu, Y., Yin, P., Gao, Y., Kan, H., Zhou, M., Chen, R., & Li, Y. (2023). Differentiating the impacts of ambient temperature on pneumonia mortality of various infectious causes: A nationwide, individual-level, case-crossover study. *EBioMedicine*, *98*, 104854. https://doi.org/10.1016/j.ebiom.2023.104854

Komara, E., Wahyuningsih, N. E., & Setiani, O. (2024). Hubungan cuaca dan kepadatan penduduk dengan kejadian DBD: Literature review: *Media Publikasi Promosi Kesehatan Indonesia (MPPKI)*, *7*(4), 864–870. https://doi.org/10.56338/mppki.v7i4.5172

Kumar, D. V. (2024). Climate change impacts on human health. *Current Research in Medical Sciences*, *3*(1), 1–9. https://www.pioneerpublisher.com/crms/article/view/633

Lebedeva, L., & Shkuropadska, D. (2024). Resilience of EU healthcare systems. *Foreign Trade: Economics, Finance, Law*, *133*(2), 120–133. https://doi.org/10.31617/3.2024(133)07

Lydia H V Franklinos, MSc, Prof Kate E Jones, PhD, David W Redding, PhD, & Prof Ibrahim Abubakar, PhD. (2019). The effect of global change on mosquito-borne disease. *The Lancet Infectious Diseases*, *19*(9), e302–e312. https://www.thelancet.com/journals/laninf/article/PIIS1473-3099(19)30161-6/abstract

Oliveira, J. B., Murari, T. B., Nascimento Filho, A. S., Saba, H., Moret, M. A., & Cardoso, C. A. L. (2023). Paradox between adequate sanitation and rainfall in dengue fever cases. *Science of The Total Environment*, *860*, 160491. https://doi.org/10.1016/j.scitotenv.2022.160491

Osborne, A. J., Bierzynska, A., Colby, E., Andag, U., Kalra, P. A., Radresa, O., Skroblin, P., Taal, M. W., Welsh, G. I., Saleem, M. A., & Campbell, C. (2024). Multivariate canonical correlation analysis identifies additional genetic variants for chronic kidney disease. *Npj Systems Biology and Applications*, *10*(1), 28. https://doi.org/10.1038/s41540-024-00350-8

Oyarzún G, M., Lanas Z, F., Wolff R, M., & Quezada L, A. (2021). The impact of climate change on health. *Revista Medica De Chile*, *149*(5), 738–746. https://doi.org/10.4067/s0034-98872021000500738

Purwadi, J., Gumelar, B., Widiantoro, T., & Ningsih, Z. N. (2024). Canonical correlation analysis of economic growth and unemployment rate. *Barekeng: Jurnal Ilmu Matematika Dan Terapan*, *18*(2), 1273–1282. https://doi.org/10.30598/barekengvol18iss2pp1273-1282

Sherry, A., & Henson, R. K. (2005). Conducting and interpreting canonical correlation analysis in personality research: A user-friendly primer. *Journal of Personality Assessment*, *84*(1), 37–48. https://doi.org/10.1207/s15327752jpa8401_09

Siti Rahmawati Hindo, Puridewi, A. E., Aulia Ananda Yuhana, Muhajir, M., S Si, Mahasiswa Jurusan Statistika, Dosen Program, Studi Statistika, Fakultas Matematika, Ilmu, D., & Pengetahuan Alam. (2018). A*nalisis hubungan tingkatan status pneumonia pada balita dan indikator lingkungan hidup sehat di indonesia menggunakan korelasi kanonikal*. Unpublished. https://doi.org/10.13140/RG.2.2.21075.89121

Viyana, A. S. (2019). *Penerapan Analisis Korelasi Kanonik Untuk Melihat Hubungan Iklim dengan Kejadian DBD dan ISPA Bukan Pneumonia di Kota Medan Tahun 2015-2017* [Thesis, Universitas Sumatera Utara]. https://repositori.usu.ac.id/handle/123456789/26614

Wei, K., Kong, W., & Wang, S. (2021). An improved multi-task sparse canonical correlation analysis of imaging genetics for detecting biomarkers of alzheimer's disease. *IEEE Access*, *9*, 30528–30538. https://doi.org/10.1109/ACCESS.2021.3059520

Weinberger, S. E. (2016). Curbing firearm violence: Identifying a specific target for physician action. *Annals of Internal Medicine*, *165*(3), 221–222. https://doi.org/10.7326/M16-0968

Wickramasinghe, N. D. (2019). Canonical correlation analysis: an introduction to a multivariate statistical analysis. *Journal of the College of Community Physicians of Sri Lanka*, *25*(1), 37. https://doi.org/10.4038/jccpsl.v25i1.8204

Winkler, A. M., Renaud, O., Smith, S. M., & Nichols, T. E. (2020). Permutation inference for canonical correlation analysis. *NeuroImage*, *220*, 117065. https://doi.org/10.1016/j.neuroimage.2020.117065