pagepress

Geospatial Health 2023; volume 18:1202

# Province clustering based on the percentage of communicable disease using the BCBimax biclustering algorithm

**Muhammad Nur Aidi,[1] Cynthia Wulandari,[1] Sachnaz Desta Oktarina,[1] Taufiqur Rakhim Aditra,[2] Fitrah Ernawati,[3] Efriwati,[3] Nunung Nurjanah,[3] Rika Rachmawati,[3] Elisa Diana Julianti,[3] Dian Sundari,[3] Fifi Retiaty,[3] Aya Yuriestia Arifin,[3] Rita Marleta Dewi,[3] Nazarina Nazaruddin,[3] Salimar,[3] Noviati Fuada,[3] Yekti Widodo,[3] Budi Setyawati,[3] Nuzuliyati Nurhidayati,[3] Sudikno,[3] Irlina Raswanti Irawan,[3] Widoretno[3]**

*[1]IPB Bogor University, Bogor, Indonesia; [2]Airlangga University, Surabaya, Indonesia; [3]National Research and Innovation Agency, Jakarta, Indonesia*

Correspondence: Muhammad Nur Aidi, Department of Statistics, IPB University, Bogor, West Java, Indonesia, Kampus IPB, Jalan Meranti Wing 22 Level 4, Dramaga, Babakan, Kec. Dramaga, Kabupaten Bogor, Jawa Barat 16680, Indonesia.
Tel.: +62 251 8625481 - Fax: +62 251 8625708
E-mail: muhammadai@apps.ipb.ac.id

Key words: biclustering; BCBimax; communicable disease; diarrhoea; Papua Province; Indonesia.

Contributions: all authors made substantial contributions to this research and approved the final manuscript. MNA, CW, SDO contributed to the data analysis and interpretation, writing and review. TRA, FE, E, NN contributed to every step (research concept, design, investigation, data interpretation, writing, editing, and review). RR, EDJ, DS, FR, AYA, RMD contributed to writing, data interpretation and review. NaN, Sa, NF, YW, BS, NazN, Su, IR and W contributed to investigation, data curation, writing and editing.

Conflict of interest: the authors declare no potential conflict of interest, and all authors confirm accuracy.

Ethics approval and consent to participate: the study was conducted in accordance with the Declaration of Helsinki and approved by Ethical Committee of National Health Institute of Research and Development, Ministry of Health, Republic of Indonesia (No: LB.02.01/2KE./267/2017). All authors have read and agreed to the published version of the manuscript.

Availability of data and materials: the data presented in this study are from the public domain (the 2018 National Report of the Basic Helth Research (Riskedas).

Funding: the study was based on freely provided secondary data from the 2018 National Report of the Basic Helth Research (Riskedas).

Acknowledgements: the authors are grateful to the National Institute of Health Research and Development, Indonesian Ministry of Health.

Received: 28 March 2023.
Accepted: 9 August 2023.


*Geospatial Health 2023; 18:1202*
*doi:10.4081/gh.2023.1202*



## Abstract

Indonesia needs to lower its high infectious disease rate. This requires reliable data and following their temporal changes across provinces. We investigated the benefits of surveying the epidemiological situation with the imax biclustering algorithm using secondary data from a recent national scale survey of main infectious diseases from the National Basic Health Research (Riskesdas) covering 34 provinces in Indonesia. Hierarchical and k-means clustering can only handle one data source, but BCBimax biclustering can cluster rows and columns in a data matrix. Several experiments determined the best row and column threshold values, which is crucial for a useful result. The percentages of Indonesia's seven most common infectious diseases (ARI, pneumonia, diarrhoea, tuberculosis (TB), hepatitis, malaria, and filariasis) were ordered by province to form groups without considering proximity because clusters are usually far apart. ARI, pneumonia, and diarrhoea were divided into toddler and adult infections, making 10 target diseases instead of seven. The set of biclusters formed based on the presence and level of these diseases included 7 diseases with moderate to high disease levels, 5 diseases (formed by 2 clusters), 3 diseases, 2 diseases, and a final order that only included adult diarrhoea. In 6 of 8 clusters, diarrhea was the most prevalent infectious disease in Indonesia, making its eradication a priority. Direct person-to-person infections like ARI, pneumonia, TB, and diarrhoea were found in 4-6 of 8 clusters. These diseases are more common and spread faster than vector-borne diseases like malaria and filariasis, making them more important.

## Introduction

The population growth in Indonesia has become one of the highest in the world during the last few decades. According to Statistics Indonesia the Indonesian population reached 275,773,800 in the mid 2022. To offset the risk of a decrease of the health indices, the Government seeks to improve the health sector by reducing the spread of communicable diseases through increased study and control activities in all provinces; these diseases are known to be responsible for the increased mortality rate in the Indonesian population (Germas, 2020). The impact of communicable diseases in a country is inseparable from the prevailing socioeconomic, and Pradana *et al*. (2021) have stated that the spread and development of infectious diseases occurs when there is an imbalance between humans, the natural environment and disease-carrying agents. This situation cannot easily be changed but

better knowledge of the distribution of communicable diseases in the country should improve the situation.

The cluster approach is arelatively new approach to the study of the spatial distribution of communicable diseases that has been used in different parts of the world, *e.g*., work on tuberculosis (TB) has been done in Mexico by Bastida *et al.* (2012), in Nigeria by James (2016), in Zimbabwe by Chirenda *et al*. (2020) and in Bandung, Indonesia by Puspita *et al*. (2021). Similar work regarding malaria has been carried out in Ghana by Magna *et al.* (2019) and the presence of clustered areas of people affected by diarrhoeal disease in Karkala Karnataka, India was recently reported by Dmello *et al*. (2022). At the same time, Almasi *et al.* (2022) investigated the global occurrence from 2000 to 2017 of clustered areas of diarrhoea-related mortality (DRM) in children under five and found that Asian and African countries had the highest incidence of DRM in this period. Their findings also revealed that the mortality due to diarrhoea was most common in Asian countries before 2010, but that this has shifted to Africa in the following decade. Likewise, regional grouping of people affected by pneumonia in Malawi has been studied by Uwemedimo *et al*. (2018) and in Bogota, Colombia by Payares-Garcia *et al*. (2023), while Santos *et al*. (2017) and Yamada *et al*. (2021) investigated the distribution of hepatitis in northern Brazil.

The dominant infectious diseases common throughout most of the provinces in Indonesia include acute respiratory infection (ARI), pneumonia, diarrhoea, TB, hepatitis, malaria and filariasis. As in other developed and developing countries, Indonesia has a number of adults with infectious diseases as well as cases invulnerable groups, such as toddlers and other children under five, in particular with with ARI, pneumonia and diarrhoea. Saputra (2021) has stated that ARI is the highest cause of death and morbidity in young children amounting to 4.25 million cases annually. According to the Ministry of Health of the Republic of Indonesia (MoH), the provinces of Banten, Bengkulu, East Nusa Tenggara, Papua and West Papua are the five provinces with the highest percentage of ARI incidents in adults (MoH, 2021).

Infectious diseases are one of the main targets of the United Nations' sustainable development goals (SDGs) (United Nation, 2017; Prasetyo *et al.* 2017). In Indonesia, the strategy for prevention and control of infectious diseases includes expanding the scope of access to health services by early detection of through surveillance, improved competence of health workers and ensurance of the availability of drugs and vaccines, including rapid diagnostic tools for control (MoH, 2020). Information on the potential spread of infectious diseases in several provinces can help implement prevention and control of infectious disease, particularly those that show clustered distributions.

Biclustering, a bidirectional technique based on grouping rows and columns with similar characteristics, can be a solution to identify and group together infectious diseases that are predominantly prevalent in some provinces. Biclustering is a data-mining approach that finds subsets of rows that have similar characteristics along a subset of columns. It finds submatrices by partitioning the data matrix based on the algorithm. To date, a number of different biclustering algorithms have been developed and compared with respect to finding the optimal biclusters (Liu & Wang, 2007; Jamail & Moussa, 2020) and Padilha and Campello (2017) mention 17 biclustering algorithms that have been compared with respect to finding optimal biclusters through various experimental scenarios. However, Wang *et al.* (2016) argue that biclustering algorithms are non-specific so there is no rule how to choose the right one for certain criteria or datasets. The selection of a biclustering algorithm should be based on several considerations, namely ease of implementation, absence of disturbance by noisy datasets and speed in finding the most suitable structure in the data matrix (Castanho *et al.,* 2022; Chu *et al.,* 2022).

We took an interest in applying the BCBimax algorithm for the study of clustering with respect to percentage of infectious diseases in Indonesia as this is a biclustering algorithm classified as very fast in finding a simple structure (Castanho *et al*., 2022). The resulting bicluster can provide information, e.g., on the magnitude of infectious disease problems in each cluster throughout a province based on the disease distribution. The specific aim of this study was to describe the presence of high clustering of percentages of infectious disease cases since this would be an indication of the potential for high transmission zones in Indonesia.

## Materials and Methods

### Sampling

The data analyzed in this study are secondary data from 300,000 households derived from the 2018 National Report of the Basic Health Research (Riskesdas), a national-scale survey with cross-sectional and non-interventional design visited. The number of individuals interviewed was 1,017,290 across 34 provinces in Indonesia (MoH, 2019).

### Variable selection

The variables used were i) respondents under the age of five years suffering from one or more of ARI, pneumonia or diarrhoea and ii) adult respondents suffering from one or more of these three afflictions plus TB, hepatitis, malaria and filariasis. Criteria for suffering from these diseases in the report were symptoms diagnosed by health workers. The percentage of people affected with each of the 10 types of infection per province was determined by the formula:

$$\frac{Number\ of\ infected}{Total\ number\ of\ respondents} x\ 100$$

### Procedure and selection of optimal biclustering

To find the optimal bicluster, we used a stepwise process involving three stages: i) pre-processing of the data collected; ii) running the biclustering algorithm incrementally; and iii) evaluation as described below and shown in Figure 1.

#### *Stage 1. Data perparation*

The pre-processing of the data in this study included scaling and searching for outliers by boxplots and heatmaps (Hutson, 2018; Qian, 2016; Tomy *et al.*, 2021) based on a initial data matrix using normal standardization by which the percentages of each target infectious disease in the different provinces were converted into a Z-score according to the following formula:

$$Z = \frac{x-\bar{x}}{s}, \qquad \text{Eq. 1}$$

where $\bar{x}$ is the average; ans $S$ the standard deviation (SD). Z=0 indicates that the percentage of the infectious disease under study is equal to the average percentage of that infectious disease in a specific province. When Z>0 this percentage is higher than the average and when Z<0, it is lower than the average.

After obtaining the scale, the data in the matrix were explored by preparing and displaying a heatmap and a boxplot to provide an initial picture of the data in the matrix. The former visualizes the percentage distribution of each infection in the provinces where the highest percentages stand out, while the latter provides information about the distribution of observations including extreme data, *i.e.* outliers.

### *Stage 2. Incremental, stepwise run of the algorithm*

The BCBimax biclustering algorithm was run in incremental steps using a search process based on several criteria to findan optimal submatrix (Dolnicar *et al.* 2012). These steps included: i) a change of the initial data matrix into a binary data matrix, where those elements of the initial data matrix that had valueshigher than the specified threshold were given the value = 1, and those with lower values than this threshold were given the value= 0; ii) this created binary data matrix was divided into two columns, **CU** and **CV**, where all rows containing the value=1 would become sub-sections of the **CU** column and those with the value=0, subsections of the **CV**column; iii) the **CU** and **CV** columnswerethen divided into three rows, **GU**, **GW** and**GV**, which produced new submatrices (note the redish and blue boxes in Fig 5. Two sub-matrices would then be processed iteratively until finding a sub-matrix having all elements of the value=1 based on the row and column thresholds set; and iv) the thresholds in the BCBimax algorithm depend on the size of the matrix formed (optimal bicluster).

The four steps given above are schematiclly displayed in Figure 2, with Figure 3 showing how the BCBimax algorithm finds a submatrix containing element one. The algorithm itself has several criteria in finding the optimal bicluster result. It can be applied after the digitalization, *i.e.* transformation of the data matrix into a binary data matrix using a predetermined threshold value, is an essential initial stage. When searching for the optimal bicluster, the algorithm also pays attention to combined bicluster sizes. Several experiments on the minimum size combination of rows and columns to be used in finding a biclustermust also be considered since the size of the bicluster dimensions affect the number and evaluation value when selecting the optimal bicluster. The range of minimum bicluster sizes that can be used is in the range of 1 to 10 for each minimum row and column set.

### *Stage 3. Evaluation*

This research used two stages of evaluation of the BCBimax algorithm, namely the average residue (ASR) and the Liu and Wang index.

Using the ASR value as an evaluation of the intra-bicluster function, Yang *et al.* (2002) and Lee *et al.* (2009) evaluated the overall total number of biclusters (n) from the experimental results of a number of bicluster results based on consideration of the minimum combination of rows and columns using the ASR by calculating the mean square residue (MSR), which can be written as follows:

$$E_{MSR}(I',J') = \frac{\sum_{i\in I'}\sum_{j\in J'}(m_{ij}-m_{ij'}-m_{i'j}+m_{i'j'})^2}{|I'|\times|J'|} \quad \text{Eq. 2}$$

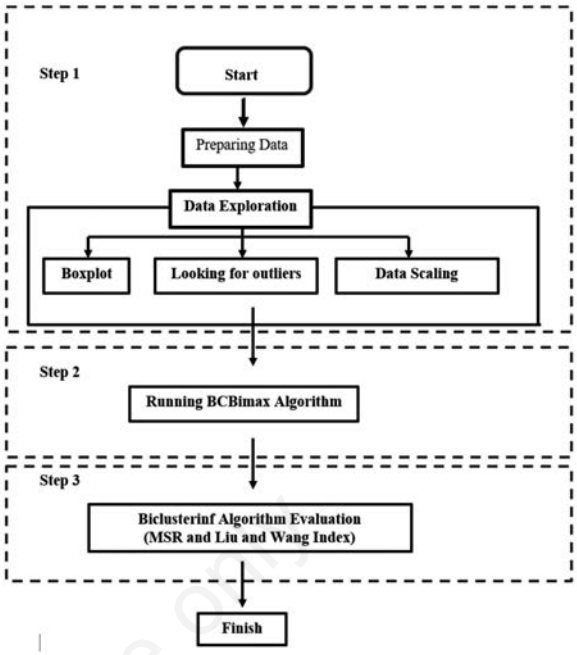



**Figure 1. The stepwise biclustering algorithm procedure.**

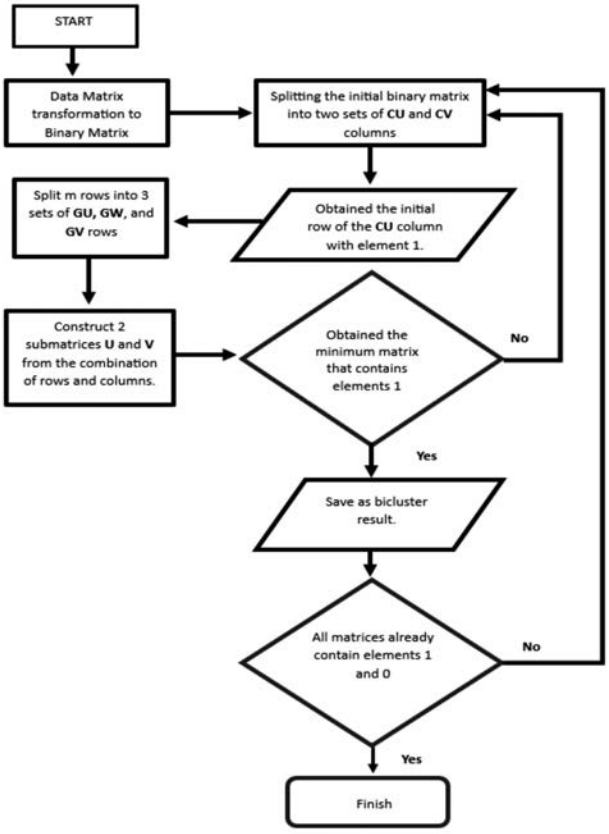



**Figure 2. The incremental, stepwise way to finding the optimal submatrix.**

where $m_{IJ}$ defines the average of the whole bicluster; $m_{Ij}$ the average in column $j$; $m_{iJ}$ the average in row $i$, $|I'|$ the row dimension of the bicluster, $|J'|$ the column dimension of the bicluster. In this way, $E_{MSR}(I', J')$ shows the variation of the interaction between rows and columns in the bicluster. Meanwhile, the ASR equation can be written as follows:

$$ASR = \frac{1}{n}\left(\sum_{i=1}^{n} E_{MSR_i}(I', J')\right) \quad \text{Eq. 3}$$

The quality of the bicluster results can be seen from the evaluation function used. The evaluation value improves when the ASR or average MSR value gets smaller and closer to 0. In addition, the initial research objectives are also taken into consideration in choosing the optimal bicluster.

This research also used the Liu and Wang index as an inter-bicluster evaluation function to see the biclustering algorithm's performance in measuring the similarity between biclusters. The Liu and Wang index can see the likeness of each bicluster group from each minimum combination of rows and columns formed (Liu & Wang 2007) and the equation is written as follows:

$$I_{Liu\&Wang}(M_{opt}, M) = \frac{1}{K_{opt}} \sum_{i=1}^{K_{opt}} \max\left(\frac{[G_i \cap G_j] + [C_i \cap C_j]}{[G_i \cup G_j] + [C_i \cup C_j]}\right) \quad \text{Eq. 4}$$

where $M$ denotes all the resulting bicluster groups; $M_{opt}$ the optimal bicluster of $M$ selected based on the average value of the residual divided by the smallest volume; $K_{opt}$ the number of bicluster in $M_{opt}$; $[G_i \cap G_i]$ the number of rows ($G$) of the optimal bicluster ($M_{opt}$) that intersect with the rows in $M$; $[C_i \cap C_i]$ the number of columns of the optimal bicluster ($M_{opt}$) that intersects with the columns in $M$; $[G_i \cup G_i]$ the number of combined rows in $M_{opt}$ and $M$; and $[C_i \cup C_i]$ the number of combined columns ($C$) in $M_{opt}$ and $M$. The higher the Liu and Wang index, the more similar the bicluster with the same membership characteristics formed will become (Lee *et al.*, 2011; Al-Akwaa, 2012; Peng *et al.*, 2014; Pandove & Malhi, 2021).

This study used the median threshold of each disease when performing the 'binarization' process. The threshold values were chosen based on the consideration that each disease had a different level of vulnerability. In addition, the result of binary matrix transformation with matrix element no. 1 showed that each region or province concerned was found to have a relatively wide distribution and thresholds using each variable's median would help answer the research objectives more effectively. One hundred trials of the combination of rows and columns were tried in this study with a value range of 1 to 10 as the minimum limit for the algorithm to find the optimal biclusters. Values outside the range of these numbers would result in the complete absence of biclusters, a fact that would defy the purpose of the research. Experimenting with combinations of minimum row and column size constraints, resulted in the formation of various biclusters; however, those that did not provide useful information were not selected for further evaluation as they were not the optimal bicluster candidates.

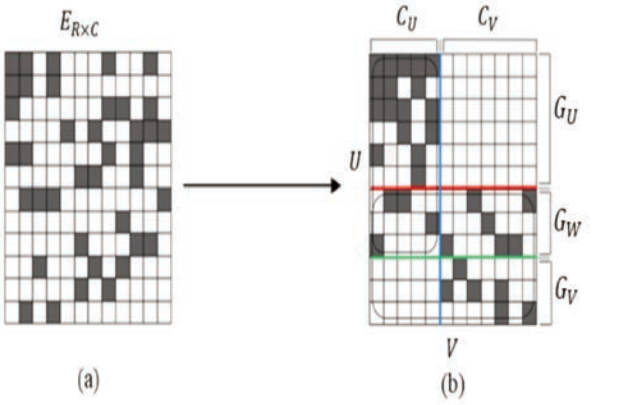



**Figure 3. Ilustration biclustering algorithm. The following figure shows an illustration of how the BCBimax algorithm works in dividing the binary data matrix into three submatrices. G denotes the set of rows in the binary data matrix; and C the set of columns. The black boxes show the matrix elements that have a value of 1; and the white boxes those that have a value value of 0. The binary data matrix that has been divided into three submatrices will then be partitioned into two submatrices (U and V). These two submatrices will be processed recursively until all the elements of the submatrix have the value of 1.**

| No | Provinces | % Toddler ARI | % Adult ARI | % Adult Pneumonia | % Toddler Pneumonia | % TBC | % Hepatitis | % Adult Diarrhoea | % Todddler Diarrhoea | % Malaria | % Filariasis |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | Aceh | -0,169 | 0,271 | -0,388 | -1,150 | 0,723 | 0,103 | 0,902 | 1,222 | -0,319 | -1,357 |
| 2 | Noth Sumatera | -0,654 | -0,860 | -0,212 | -0,488 | -0,536 | -0,286 | 0,902 | 1,573 | -0,319 | -1,048 |
| 3 | West Sumatera | 0,515 | 0,315 | -0,916 | -0,753 | -0,470 | -0,383 | 1,018 | 0,949 | -0,344 | -1,048 |
| 4 | Riau | -1,053 | -0,729 | -1,092 | -1,415 | -1,067 | -0,091 | -0,027 | -0,298 | -0,352 | -1,048 |
| 5 | Jambi | -1,053 | -1,425 | -1,444 | -1,348 | -0,735 | -0,091 | -1,770 | -1,077 | -0,295 | -1,048 |
| 6 | South Sumatera | -0,682 | -0,816 | -0,476 | -0,224 | 0,989 | -0,674 | -1,189 | -0,259 | -0,303 | -0,738 |
| 7 | Bengkulu | 1,541 | 1,315 | 0,931 | 0,305 | 0,193 | 0,297 | 1,076 | 1,144 | 0,225 | -0,738 |
| 8 | Lampung | 0,087 | -0,599 | -1,004 | -0,885 | -0,337 | -0,966 | -1,537 | -0,883 | -0,299 | -0,738 |
| 9 | Bangka Belitung | -0,768 | -0,816 | -0,828 | -0,158 | -1,929 | -2,132 | -1,828 | -1,506 | 0,034 | -0,738 |
| 10 | Riau Islands | -0,625 | -0,990 | -1,444 | -1,481 | -0,603 | -1,452 | -1,886 | -2,091 | -0,271 | -0,738 |
| 11 | Jakarta | 0,629 | -0,120 | -0,388 | -0,488 | 0,856 | 0,783 | -0,202 | -0,142 | -0,384 | -0,428 |
| 12 | West Java | 1,056 | 1,054 | 0,404 | 0,372 | 1,652 | 0,497 | 0,611 | 1,183 | -0,384 | -0,428 |
| 13 | Central Java | 0,800 | -0,120 | -0,740 | -0,687 | -0,138 | -1,160 | 0,495 | 0,403 | -0,388 | -0,428 |
| 14 | Yogyakarta | -0,112 | -0,816 | -0,476 | 2,224 | -1,465 | -0,480 | 0,553 | -1,038 | -0,368 | -0,428 |
| 15 | East Java | 1,769 | 0,315 | -0,740 | -0,356 | -0,603 | 0,006 | 0,031 | -0,259 | -0,392 | -0,428 |
| 16 | Banten | 1,912 | 1,358 | 0,580 | 0,372 | 2,514 | 0,200 | 0,960 | 1,105 | -0,364 | -0,428 |
| 17 | Bali | 0,771 | 0,402 | -0,828 | 0,107 | -1,663 | -0,383 | 0,437 | -0,688 | -0,384 | -0,428 |
| 18 | West Nusa Tenggara | 0,087 | 1,271 | 0,316 | 0,239 | -0,404 | 1,560 | 1,541 | 1,456 | -0,137 | -0,118 |
| 19 | East Nusa Tenggara | 2,168 | 2,881 | 2,339 | 2,555 | -0,735 | -0,480 | -0,550 | -0,142 | 0,408 | -0,118 |
| 20 | West Kalimantan | -0,369 | -0,164 | 0,140 | -0,091 | -0,138 | -1,063 | 0,670 | 1,222 | -0,246 | -0,118 |
| 21 | Central Kalimantan | 1,170 | 0,054 | -0,476 | -0,422 | 0,060 | 0,006 | -1,189 | -0,922 | -0,327 | -0,118 |
| 22 | South Kalimantan | -1,082 | -0,729 | -0,652 | -1,415 | 0,193 | -0,674 | -0,608 | -0,532 | -0,352 | -0,118 |
| 23 | East Kalimantan | -0,055 | -0,294 | -0,740 | -0,819 | -0,337 | -0,091 | -0,840 | -0,610 | -0,319 | 0,191 |
| 24 | Noth Kalimantan | -0,882 | -0,860 | -0,212 | 0,041 | 0,922 | -0,674 | 0,089 | 0,170 | -0,336 | 0,191 |
| 25 | Noth Sulawesi | -1,367 | -1,121 | -0,212 | -0,555 | 0,060 | 0,686 | -0,492 | -0,960 | -0,210 | 0,191 |
| 26 | Central Sulawesi | 0,087 | 0,271 | 1,459 | 1,562 | 0,060 | 2,143 | 1,599 | 1,183 | -0,100 | 0,191 |
| 27 | South Sulawesi | -0,654 | -0,207 | 0,756 | 0,239 | -0,138 | 0,297 | 0,960 | 0,170 | -0,348 | 0,501 |
| 28 | Southeast Sulawesi | -0,255 | -0,294 | 0,052 | 0,372 | 0,193 | -0,091 | -0,376 | -0,649 | -0,311 | 0,501 |
| 29 | Gorontalo | -1,253 | 0,315 | 1,635 | 1,099 | 0,259 | 1,463 | 0,728 | 0,481 | -0,352 | 1,430 |
| 30 | West Sulawesi | -1,053 | -0,816 | 0,580 | 0,702 | -0,470 | 1,657 | 0,495 | 0,481 | -0,315 | 1,430 |
| 31 | Maluku | -0,568 | -0,120 | -0,036 | 0,041 | 0,060 | -0,577 | -0,550 | -0,493 | 0,091 | 1,740 |
| 32 | Noth Maluku | -1,424 | -1,338 | 0,052 | 0,041 | -0,536 | -0,674 | -1,247 | -1,272 | 0,152 | 1,740 |
| 33 | West Papua | 0,629 | 1,532 | 1,635 | 0,636 | 0,989 | 0,200 | 0,147 | -0,649 | 3,109 | 1,740 |
| 34 | Papua | 0,857 | 1,880 | 2,427 | 1,827 | 2,580 | 2,532 | 1,076 | 1,728 | 4,502 | 2,979 |
| | | High | | | | | | | | | |
| | | Low | | | | | | | | | |

ARI, acute respiratory tract infection; TB, tuberculosis.

**Figure 4. Heatmap of scaled data matrix. Blue cells indicate areas where the percentage of infectious diseases is higher than the average, which means the percentage of the disease is relatively high; yellow cells indicate areas where the percentage of infectious diseases is lower than the average.**

## Results

### Data exploration

After converting the initial data matrix into a scaling data matrix, the initial description of the relationship between the infectious disease cases by province was presented in a heatmap diagram (Figure 4), which shows that several provinces are vulnerable with respect to these diseases. This figure also shows that as many as 10 infectious diseases (toddler ARI, adult ARI, toddler pneumonia, adult pneumonia, toddler diarrhoea, adult diarrhoea, TB, hepatitis, malaria and filariasis) were above the national percentage in Papua and Bengkulu provinces. Meanwhile, the percentages of the same diseases in Riau and East Kalimantan provinces were found to be below the national percentage.

With regard to the percentage of toodler ARI, adult pneumonia, adult diarrhoea and toddler diarrhoea there were no provinces that are considered outliers in the percentage value (Figure 5). In the percentage of adult ARI, however, one province was considered an outlier, namely the East Nusa Tenggara Province, while in the percentage of toddler pneumonia there were two such provinces, namely the provinces Yogyakarta and East Nusa Tenggara. The percentage of TB had four provinces as outliers namely Banten, West Java, Bangka Belitung and Papua. With respect to percentage of hepatitis, there were three such provinces (Central Sulawesi, Bangka Belitung and Papua). The provinces Papua, West Papua, Bengkulu and East Nusa Tenggara were considered outliers with regard to the percentage of malaria. Finally, for the percentage of filariasis, the provinces Maluku, Riau Island, Papua and West Papua were found to be outliers.

### Provincial distribution of the diseases under study

#### *ARI in toddlers*

The national average percentage of ARI in toddlers was 11.0% (median = 10.5%). There were 15 provinces with a higher percentage of ARI in toddlers (Lampung, West Nusa Tenggara, Central Sulawesi, West Sumatra, Jakarta, West Papua, Bali, Central Java, Papua, West Java, Central Kalimantan, Bengkulu, East Java, Banten and East Nusa Tenggara). The remaining 18 provinces had a lower percentage of this type of ARI.

#### *ARI in adults*

The national average percentage of ARI in adults was 8.8% (median = 8.5%). There were 14 provinces with a higher percentage of adult ARI (Central Kalimantan, Aceh, Central Sulawesi, West Sumatra, East Java, Gorontalo, Bali, West Java, West Nusa Tenggara, Bengkulu, Banten, West Papua, Papua and East Nusa Tenggara. The remaining 19 provinces had a lower percentage of this type of ARI.

#### *Pneumonia in toddlers*

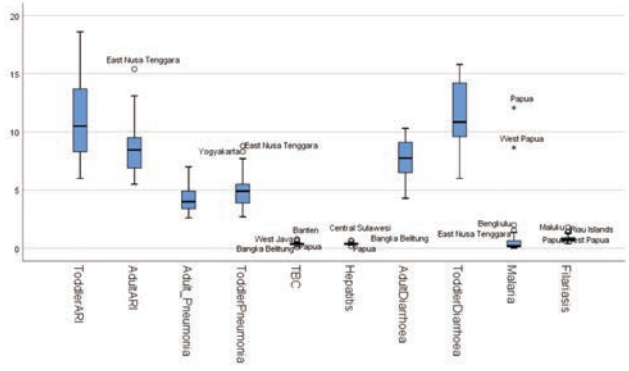
The national average percentage of pneumonia in toddlers was 4.9% (median = 4.9%). There were 17 provinces with a higher percentage of pneumonia in toddlers (North Kalimantan, Maluku, North Maluku, Bali, West Nusa Tenggara, South Sulawesi, Bengkulu, West Java, Banten, Southeast Sulawesi, West Papua, West Sulawesi, Gorontalo, Central Sulawesi, Papua, Yogyakarta and East Nusa Tenggara). The remaining 16 provinces had a lower percentage of this type of pneumonia.

#### *Pneumonia in adults*

The national average percentage of pneumonia was 4.2% (median = 4.0%). There were 14 provinces with a higher percentage of adult pneumonia (Southeast Sulawesi, North Maluku, West Kalimantan, West Nusa Tenggara, West Java, Banten, West Sulawesi, South Sulawesi, Bengkulu, Central Sulawesi, Gorontalo, West Papua, East Nusa Tenggara and Papua). The remaining 19 provinces had a lower percentage of this type of pneumonia.

#### *Diarrhoea in toddlers*

The national average percentage of diarrhoea in toddlers was 11.4% (median = 10.9%). There were 15 provinces with a higher percentage of diarrhoea in toddlers (North Kalimantan, South Sulawesi, Central Java, West Sulawesi, Gorontalo, West Sumatra, Banten, Bengkulu, West Java, Central Sulawesi, West Kalimantan, Aceh, West Nusa Tenggara, North Sumatera and Papua). The remaining 18 provinces had a lower percentage of this type of diarrhoea.




ARI, acute respiratory tract infection; TB, tuberculosis.

**Figure 5. Boxplot of data matrix according to type of disease.**

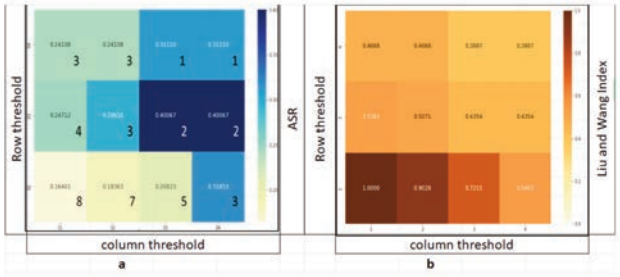



**Figure 6. Heatmap of ASR value evaluation; a) the average residue (ASR); b) the Liu and Wang Index based on candidate of bicluser optimal selected from BCBimax algorithm minimum threshold combination experiment.**

### *Diarrhoea in adults*

The national average percentage was 7.6% (median = 7.8%). There were 19 provinces with a higher percentage of adult diarrhoea (East Java, North Kalimantan, West Papua, Bali, Central Java, West Sulawesi, Yogyakarta, West Java, West Kalimantan, Gorontalo, North Sumatera, Aceh, Banten, South Sulawesi, West Sumatera, Bengkulu, Papua, West Nusa Tenggara and Central Sulawesi. The remaining 14 provinces had a lower percentage of ofthis type of diarrhoea.

### *Tuberculosis*

The national average percentage of TB was 0.4% (median = 0.4%). There were 16 provinces with a high percentage of TB (North Sulawesi, Central Kalimantan, Maluku, Central Sulawesi, South Kalimantan, Bengkulu, Southeast Sulawesi, Gorontalo, Aceh, Jakarta, North Kalimantan, South Sumatera, West Papua, West Java, Banten and Papua). The remaining 17 provinces had a lower percentage of TB.

### *Hepatitis*

The national average percentage of hepatitis was 0.4 % (median = 0.4%). There were 15 provinces with a higher percentage of hepatitis (East Java, Central Kalimantan, Aceh, West Papua, Banten, South Sulawesi, Bengkulu, West Java, North Sulawesi, Jakarta, Gorontalo, West Nusa Tenggara, West Sulawesi, Central Sulawesi and Papua. The remaining 18 provinces had a lower percentage of hepatitis.

### *Malaria*

The national average percentage of malaria was1.0 % (median = 0.2%). There were 7 provinces with a higher percentage of malaria (Bangka Belitung, Maluku, North Maluku, Bengkulu, East Nusa Tenggara, West Papua and Papua). The remaining 26 provinces had a lower percentage of malaria.

### *Filariasis*

The national average percentage of filariasis was 0.8 % (median = 0.8%). There were 12 provinces have a higher percentage of filariasis (Central Kalimantan, South Sumatera, North Maluku, Bengkulu, South Kalimantan, North Kalimantan, Jambi, West Nusa Tenggara, West Sulawesi, East Nusa Tenggara, Papua and Banten). The remaining 21 provinces had a lower percentage of filariasis.

As given by the text above and displayed in Table 1, it can be concluded that the mean percentage of disease was greater than the median disease percentage (in same cases the same) for all the diseases studied except for adult diarrhoea. Thus, the percentage distribution was right-skewed, which means that the mean was greater than the median. The mean overestimates the most common values in a positively skewed distribution, while it underestimates them in a negatively skewed distribution, *i.e.* the mean has a lower value than the median in a left-skewed distribution. The explanation in our case was that there were outliers in some provinces. In addition, the average percentages of TB, hepatitis, malaria and filariasis had values below 1%, while it was above 1% for ARI, pneumonia, and diarrhoea, both in toddlers and adults. Taking the SD into account, only TB, hepatitis and filariasis had values below 1%.

### *Outliers*

Although the average percentage of malaria was just below 1%, the SD was more than double that (2.5). This indicates that several provinces must have a much higher percentage, e.g., West



**Table 1. Percentage distribution of the ten most common communicable diseases in Indonesia.**

| | Toddler ARI (%) | Adult ARI (%) | Adult pneumonia (%) | Toddler pneumonia (%) | TB (%) |
|---|---|---|---|---|---|
| Study provinces | 34 | 34 | 34 | 34 | 34 |
| Provincial data missed | 0 | 0 | 0 | 0 | 0 |
| Mean | 10.99 | 8.78 | 4.24 | 4.94 | 0.38 |
| Median | 10.50 | 8.45 | 4.00 | 4.90 | 0.36 |
| Standard deviation | 3.51 | 2.30 | 1.14 | 1.51 | 0.15 |
| Variance | 12.31 | 5.29 | 1.29 | 2.29 | 0.02 |
| Range | 12.60 | 9.90 | 4.40 | 6.10 | 0.68 |
| Minimum | 6.00 | 5.50 | 2.60 | 2.70 | 0.09 |
| Maximum | 18.60 | 15.40 | 7.00 | 8.80 | 0.77 |
| | Hepatitis (%) | Toddler diarrhoea (%) | Adult diarrhoea (%) | Malaria (%) | Filariasis (%) |
| Study provinces | 34 | 34 | 34 | 34 | 34 |
| Provincial data missed | 0 | 0 | 0 | 0 | 0 |
| Mean | 0.40 | 11.36 | 7.55 | 0.99 | 0.84 |
| Median | 0.39 | 10.85 | 7.75 | 0.21 | 0.75 |
| Standard deviation | 0.10 | 2.57 | 1.72 | 2.46 | 0.32 |
| Variance | 0.011 | 6.585 | 2.97 | 6.06 | 0.10 |
| Range | 0.48 | 9.80 | 6.00 | 12.05 | 1.40 |
| Minimum | 0.18 | 6.00 | 4.30 | 0.02 | 0.40 |
| Maximum | 0.66 | 15.80 | 10.30 | 12.07 | 1.80 |

ARI, acute respiratory tract infection; TB, tuberculosis.

Papua Province and Papua Province as they displayed malaria percentages of 8.6%, and 12.1%, respectively. Those values are far from the average percentage of the other provinces (0.4%), a fact that tells us that these two provinces are outliers and that they need to pay special attention to the malaria situation.

The largest percentage of adult ARI amounted to 15.4% and was found in East Nusa Tenggara Province, a percentage that makes this province a pronuanced outlier with regard to this infection. East Nusa Tenggara and Yogyakarta Provinces were the provinces with the largest and the second largest percentages of toddler pneumonia with 8.8% and 8.3%, respectively, which make them outliers with respect to this disease. The provinces of Banten and West Java had TB percentages of 0.8% and 0.6%, respectively, while Papua Province and Central Sulawesi Province showed hepatitis percentages of 0.7% and 0.6%, respectively, all levels making them outliers, the former with respect to TB and the latter regarding hepatitis.

## Bicluster selection

The selected optimal bicluster candidates from all the threshold combination experiments and the results of the evaluation stage are presented in the heatmap (Figure 6). This study set the minimum size of column 1 to 4 as the optimal bicluster result and the minimum size of row 2 to 4 as the selected optimal bicluster candidate. The heatmap in Figure 6a displays the results of the ASR value evaluation carried out, while Figure 6b displays the evaluation results using the inter-bicluster function using the calculation based on the Liu and Wang index. Both evaluation stages show that the minimum size combination of row 2 and column 1 had the smallest ASR (0.16401). Evaluation using the Liu and Wang index showed that the BCBimax algorithm produces different bicluster result memberships for each candidate trial combination. The algorithm set the minimum size of row 2 and column 1 as the optimal bicluster result with eight biclusters formed, *i.e.* the number of biclusters classified as informative in helping to answer the research objectives. Figure 7 shows the distribution map of optimal bicluster results throughout Indonesia using the BCBimax algorithm.

Based on the 34 provinces and 10 diseases classed using BCBimax biclustering were only able to classify 17 provinces (Bengkulu, Papua, West Java, Central Sulawesi, Banten, West Papua, Aceh, West Nusa Tenggara, South Sulawesi, West Sulawesi, Jakarta, Central Kalimantan, East Nusa Tenggara, Gorontalo, North Sumatera, Central Java and North Borneo) and seven of the diseases (adult ARI, adult pneumonia, TB, hepatitis, adult diarrhoea, malaria and filariasis (Table 2). The results formed eight groups (Table 2, Figure 7). Toddler ARI, toddler pneumonia and toddler diarrhoea could not be used to classify provinces based on the BCBimax method because there was no similarity in their percentages anywhere.

Bicluster 1 is characterized as 2 x 7, which means that this group consists of two provinces (Bengkulu and Papua) and seven diseases (adult ARI, adult pneumonia, adult diarrhoea, TB, hepati-

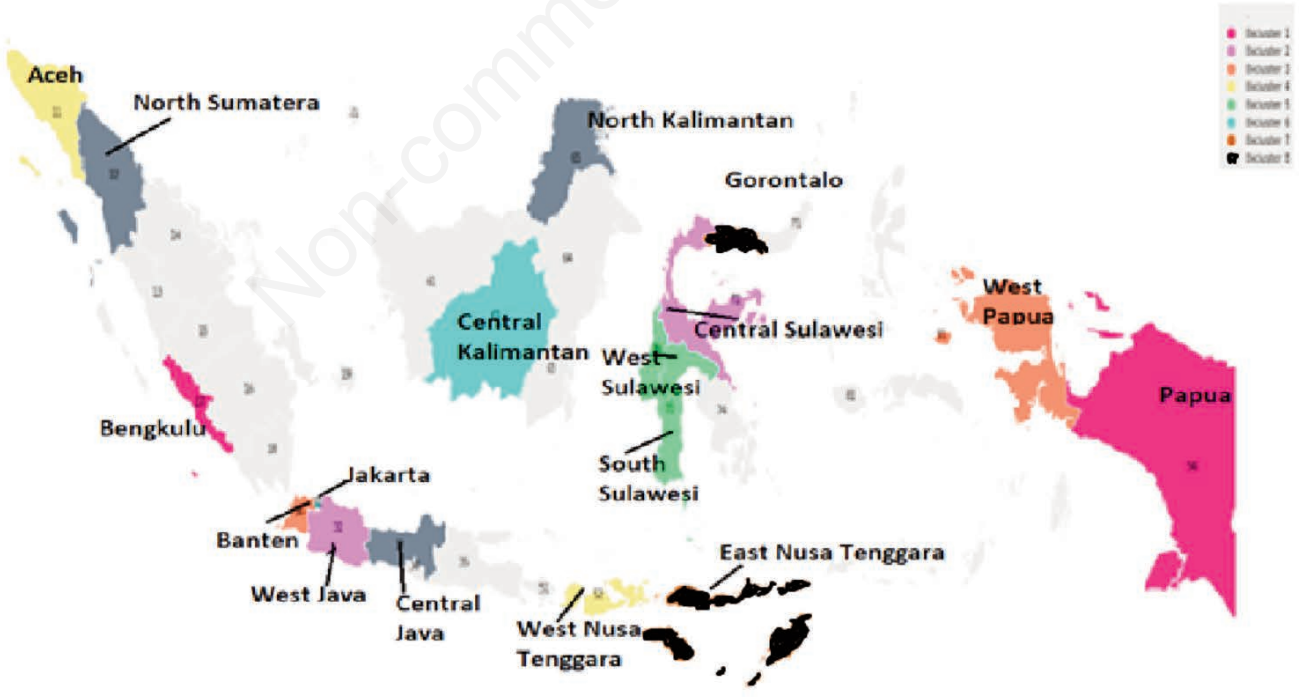



**Figure 7. Distribution map of optimal bicluster results of the BCBimax algorithm. Bicluster 1: Bengkulu and Papua (crimson); Bicluster 2: West Java and Central Sulawesi (violet); Bicluster 3: Banten and West Papua (brown); Bicluster 4: Aceh and West Nusa Tenggara (yellow); Bicluster 5: South Sulawesi and West Sulawesi (green); Bicluster 6: Jakarta and Central Kalimantan (blue); Bicluster 7: East Nusa Tenggara and Gorontalo (black); and Bicluster 8: North Sumatera, Central Java and North Kalimantan (grey).**

tis, malaria and filariasis). These two provinces were selected because the 7 diseases were found to be present at similar percentage levels (Figure 7). However, Papua Province had a high percentage of all seven diseases, while Bengkulu Province only showed high percentages of adult ARI, adult pneumonia, and adult diarrhoea, with a moderate presence of TB, hepatitis, malaria and filariasis (Table 2, Figure 7). The two provinces are spatially very far apart, with a distance between them of 3,962 km.

Bicluster 2 is characterized as 2 x 5 and consists of two provinces (West Java and Central Sulawesi) and five diseases (adult ARI, adult pneumonia, TB, hepatitis and adult diarrhoea). The two provinces are in the same group because of similar percentages of the five diseases. However, while West Java Province had a high percentage of all of them, Central Java Province had high percentages of adult pneumonia, hepatitis and adult diarrhoea, with moderate percentages of adult ARI and TB (Table 2, Figure 7). The two provinces are spatially separated by 1,651 km.

Bicluster 3 contains Banten and West Papua provinces, with five diseases (adult ARI, adult pneumonia, TB, hepatitis and adult diarrhoea). They belong to the same bicluster due to their similar percentages of these 5 diseases. Banten Province had high percentages of adult ARI, adult pneumonia, adult diarrhoeaand TB but only a moderate percentage of hepatitis, while West Papua Province had high percentages of adult ARI, adult pneumonia and TB, but moderate presence of both adult diarrhoea and hepatitis (Table 2, Figure 7). The two provinces are spatially very far apart (3,057 km).

Bicluster 4 consists of Aceh and West Nusa Tenggara provinces with only three diseases (adult ARI, adult diarrhoea and hepatitis). These provinces are presented together because of similar percentages with respect to these three diseases. While West Nusa Tenggara Province had high percentages of all of them, Aceh Province had only a high percentage of adult diarrhoea, with a moderate presence of adult ARI and hepatitis (Table 2, Figure 7). The two provinces are spatially far apart (3,834 km).

Bicluster 5 contains South Sulawesi and West Sulawesi provinces. They are in the same group because both had high percentages of adult pneumonia and adult diarrhoea (Table 2, Figure 7). The two provinces have a joint border.

Bicluster 6 consists of Jakarta and Central Kalimantan provinces. Both have similar percentages of TB and hepatitis, but their presence was high in Jakarta but only moderate in Central

**Table 2. Membership characteristics of the optimal bicluster matrix.**

| Bicluster | Size of Bicluster | Province | Toddler ARI | Adult ARI | Adult Pneumonia | Toddler Pneumonia | TBC | Hepatitis | Adult Diarrhoea | Toddler Diarrhoea | Malaria | Filariasis |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 2x7 | Bengkulu | | H | H | | M | M | H | | M | M |
| | | Papua | | H | H | | H | H | H | | H | H |
| 2 | 2x5 | West Java | | H | H | | H | H | H | | | |
| | | Central Sulawesi | | M | H | | M | H | H | | | |
| 3 | 2x5 | Banten | | H | H | | H | M | H | | | |
| | | West Papua | | H | H | | H | M | M | | | |
| 4 | 2x3 | Aceh | | M | | | | M | H | | | |
| | | West Nusa Tenggara | | H | | | | H | H | | | |
| 5 | 2x2 | South Sulawesi | | | H | | | | H | | | |
| | | West Sulawesi | | | H | | | | H | | | |
| 6 | 2x2 | Jakarta | | | | | H | H | | | | |
| | | Central Kalimantan | | | | | M | M | | | | |
| 7 | 2x2 | East Nusa Tenggara | | H | H | | | | | | | |
| | | Gorontalo | | M | H | | | | | | | |
| 8 | 3x1 | North Sumatera | | | | | | | H | | | |
| | | Central Java | | | | | | | H | | | |
| | | North Kalimantan | | | | | | | M | | | |

Low Percentage
Middle Percentage
High Percentage

First Group
Second Group
Third Group
Fourth Group
Fifth group

Kalimantan Province (Table 2, Figure 7). The distance between these provinces is 1,352 km. Bicluster 7 has two members (East Nusa Tenggara and Gorontalo) and two diseases (adult ARI and adult pneumonia). While East Nusa Tenggara Province had a high percentages of both diseases, Gorontalo Province had only a high percentage of adult pneumonia, with a moderate percentage of adult ARI (Table 2, Figure 7). The distance between the two provinces is 2,098 km. Bicluster 8 consists of three provinces (North Sumatera, Central Java and North Kalimantan) but is only concerned with one disease (adult diarrhoea) that exists as high in the two former provinces and only moderate in North Kalimantan (Table 2, Figure 7).

## Discussion

Based on the information on the bicluster approach, the level of infectious disease problem can be summarized as follows: with the presence of seven different infections, bicluster 1 has the most serious problems with respect to infectious diseases, while biclusters 2 and 3 each 'only' have five such diseases. Bicluster 4 comes third with 3 diseases (adult ARI and diarrhoea; TB; and hepatitis) followed by biclusters 5, 6 and 7 with only 2 diseases each: adult ARI and adult pneumonia, where only one of them, adult ARI, exists in two different biclusters. Bicluster 8 stands out as the group with the least problems with infectious diseases as it is only characterized by diarrhoea.

Papua Province is a province with a high percentage of seven diseases (adult ARI, adult diarrhoea, adult pneumonia, TB, hepatitis, malaria and filariasis). With the exception of malaria and filariasis, West Java Province also requires special attention because of the risk of spreading these diseases. South Sulawesi and West Sulawesi provinces, on the other hand, will need to focus more on adult pneumonia and adult diarrhoea because of its high percentagea of these two diseases, while Jakarta Province presents a strong risk of the spread of TB and hepatitis.

Based on the bicluster results with the BCBimax algorithm, it can be concluded that diarrhoea disease has the widest range by covering six bicluster groups out of the eight formed. While ARI, hepatitis and adult pneumonia had the second largest range (five groups out of the eight formed), TB malaria and filariasis, in that order, showed increasingly lower spread.

With the exception of the research on TB in Bandung, Indonesia (Puspita *et al*., 2021), which supports the use of the cluster approach presented here, cannot easily be compared to other projects mentioned in the Introduction section since the situations described are very different with regard to climate, socio-economic situation and, not the least, the geographical situation, which is not commonly met in other countries. Importantly, however, all published work we considered used clustering methodology that functioned well for the study of communicable diseases.

The use of the BCBimax method in grouping provinces along the percentages of at set of target diseases as the key variable resulted in eight bicluster groups. This grouping used the percentage of of ARI, pneumonia, TB, hepatitis, diarrhoea, malaria and filariasis and produced pairs of provinces (in one case a triplet) with similar problems with regard to certain communicable diseases in spite of the pair/triplet members often being geographically far from each other.

## Conclusions

The finding that provinces situated far from each other often show similarities indicates that other variables than close contact, e.g., socioeconomic variables, also can play an important role. Thus, application of the BCBimax algorithm points towards new, potentially successful approaches to public health problems. Collaboration with respect to these diseases in the pairs/triplet found should contribute to control and prevention of the disease in question. Another valuable finding was the visualization of diarrhoea as having the most extensive range as it appeared in six of the groups out of 8 groups formed. This confirms the strong need for diarrhoeal disease eradication in Indonesia.

## References

Al-Akwaa FM, 2012. Analysis of Gene Expression Data Using Biclustering Algorithms. In Functional Genomics. Edited by Germana Meroni and Francesca Petrera. Published 12 September 2012. doi:10.5772/3117. isbn978-953-51-0727-9. ebook (pdf) isbn: 978-953-51-5316-0. IntechOpen 5, Princes Gate Count. London, SW 7 20J, UK.

Almasi A, Zangeneh A, Ziapour A, Saeidi A, Teimouri R, Ahmadi T, Khezeli M, Moradi G, Soofi M, Salimi Y, Rajabi-GilanN, Ghasemi SR, Heydarpour F, Moghadam S,Yigitcanlar T, 2022. Investigating Global Spatial Patterns of Diarrhoea-Related Mortality in Children Under Five. Front Public Health 10:861629.

Bastida AZ, Tellez MHN, Montes LPB, Torres IM, Paniagua JNSJ, Tes Maejb, Barrera, Rez-Duran NR, 2017. Spatial and temporal distribution of tuberculosis in the State of Mexico, Mexico. Vet Ital 53:39-46.

Castanho EN, Aidos H, Madeira SC, 2022. Biclustering fMRI time series: a comparative study. BMC Bioinformatics 23:1–30.

Chirenda J, Gwitira I, Warren RM, Sampson SL, Murwira A, Masimirembwa C, Mateveke KM, Duri C, Chonzi P, Rusakaniko S, Streicher EM, 2020. Spatial distribution of Mycobacterium tuberculosis in metropolitan Harare, Zimbabwe. PLoS One 15:116:e0231637.

Chu HM, Liu JX, Zhang K, Zheng CH, Wang J, Kong XZ, 2022. A binary biclustering algorithm based on the adjacency diference matrix for gene expression data analysis. BMC Bioinformatics 23:381.

Dmello MK, Badiger S, Kumar S, Kumar N, Dsousa N, 2022. Spatial and space-time clustering of diarrhoeal cases among under-five children in Karkala, Karnataka: a geospatial analysis. J Clin Diagn Res 16:1-5

Dolnicar S, Kaiser S,Lazarevski K, Leisch F, 2012. Biclustering: Overcoming data dimensionality problems in market segmentation. J Travel Res 51:41-

Germas, 2020. Rencana Aksi Kegiatan. Promosi Kesehatan dan Pemberdayaan Masyarakat Tahun 2020-2024 [Activity Action Plan. Health Promotion and Community Empowerment 2020-2024.] Directorate of Health Promotion and Community Empowerment. Ministry of Health of the Republic of Indonesia [Direktorat Promosi Kesehatan dan Pemberdayaan Masyarakat. Kementerian Kesehatan Republik Indonesia.]

Hutson AV, 2018. Statistics in the Health Sciences Theory, Applications, and Computing. 1st Edition. Published August

29, 2022 by Chapman & Hall Book. 2-6 Boundary Row, London, SE1 8HN, UK

James DO, 2016. Spatial distribution of tuberculosis in Nigeria and its socioeconomic correlates. Faculty of Health and Medicine, Lancaster University, Doctoral thesis.

Jamail I, Moussa A, 2020. Current State-of-the-Art of Clustering Methods for Gene Expression Data with RNA-Seq. In Applications of Pattern Recognition, edited by Carlos M. Travieso-Gonzalez. ISBN: 978-1-78985-561-6. IntechOpen 5 Princes Gate Count. London, SW 7 2QJ, UK.

Lee Y, Lee J-H, Jun C-H, 2009. Validation measures of bicluster solutions. Ind Eng Manag Syst 8:101–108.

Lee Y, Lee J-H, Jun C-H, 2011. Stability-based validation of bicluster solutions. Pattern Recognition 44:252-64.

Liu X, Wang L, 2007. Computing the maximum similarity biclusters of gene expression data. BMC Bioinformatics 23:50-6.

Magna EK, Dabi M, Tadri P, 2019. Spatial distribution of malaria in the semi-arid zone of Ghana: A case of upper west region using GIS approach. J Environ Health and Sustain Dev 4:670-7.

Ministry of Health (MoH), 2019. Laporan Nasional Riskesdas 2018 (Basic Health Research National Report 2018), Badan Penelitian dan Pengembangan Kesehatan. Jakarta: Indonesia. Available from: http://labdata.litbang.kemkes.go.id/images/download/laporan/RKD/2018/Laporan_Nasional_RKD2018_FINAL.pdf.

MoH, 2020. Peraturan Menteri Kesehatan Republik Indonesia Nomor 21 Tahun 2020 tentan Rencana Strategis Kementerian Kesehatan Tahun 2020-2024 [Regulation of the Minister of Health of the Republic of Indonesia Number 21 of 2020 concerning the Strategic Plan of the Ministry of Health for 2020-2024]. Indonesia.

MoH, 2021. Profil Kesehatan Indonesia Tahun 2020 (Indonesia Health Profile 2020). Jakarta: Indonesia. doi: 10.1524/itit.2006.48.1.6.

Padilha VA, Campello RJGB, 2017. A systematic comparative evaluation of biclustering techniques. BMC Bioinformatics 18:1–25.

Pandove D, Malhi A, 2021. A Correlation based recommendation system for large data sets. J Grid Computing 19:42.

Payares-Garcia D, Quintero-Alonso B, Carlos Melo-Martinez CE. 2023. Determinants of Pneumonia mortality in Bogota, Colombia: A spatial econometrics approach. Spatial and Spat Spatiotemporal Epidemiol 45:100581.

Peng X, Cai L, Liao B, Chen H, Zhu W, 2014. Detecting the Maximum Similarity Bi-Clusters of Gene Expression Data with Evolutionary Computation. J Comput Theor Nanosci 11:1585-91.

Pradana AA, Pramitaningrum IK, Aslam M, Anindita R, 2021. Epidemiologi Penyakit Menular Pengantar Bagi Mahasiswa Kesehatan (Epidemiology of Infectious Diseases Introduction to Health Students). Rajawali Press, Jakarta. ISBN 9786232318144

Prasetyo R, Siagian TH, 2017. Determinan Penyakit Berbasis Lingkungan pada Anak Balita di Indonesia [Determinants of Environmental Based Diseases in Toddlers in Indonesia]. Jurnal Kependudukan. Indonesia 12:93–104.

Puspita T, Suryatma A, Simarmata OS, Veridona G, Lestary H, Anwar A, Pambudi I, Sulistyo, Pakasi TT, 2021. Spatial variation of tuberculosis risk in Indonesia 2010-2019. Health Sci J Indonesia 12:104-10.

Qian SS, 2016. Environmental and Ecological Statistics with R. Second Edition. CRC Press. Taylor and Francis Group. New York, USA.

Santos MB, dos Santos AD, da Silva PP, Barreto AS, dos Santos EO, França AVC, Barbosa CS, de Araújo KCGM, 2017. Spatial analysis of viral hepatitis and schistosomiasis coinfection in an endemic area in Northeastern Brazil. Rev Soc Bras Med Trop 50:383-7.

Saputra HA, 2021. Faktor-Faktor yang Berhubungan Dengan Kejadian Infeksi Saluran Pernafasan Akut (Ispa) pada Balita [Factors Associated with the Incidence of Acute Respiratory Infection (ARI) in Toddlers).] J Public Health 8:16-27.

Tomy L, Chesneau C, Madhav AK, 2021. Statistical Techniques for Environmental Sciences: A Review. Math Comput Appl 26;74.

United Nation, 2017. The Sustainable Development Goals. Report. United Nations. New York, USA.

Uwemedimo OT, Lewis TP, Essien EA, Chan GJ, Nsona H, Kruk ME, Leslie HH, 2018. Distribution and determinants of pneumonia diagnosis using Integrated Management of Childhood Illness guidelines: a nationally representative study in Malawi. BMJ Glob Health 3:e000506..

Wang B, Miao Y, Zhao H, Jin J, Chen Y, 2016. A biclustering-based method for market segmentation using customer pain points. Eng Appl Artif Intell 47:101–109.

Yang J, Wang W, Wang H, Yu P, 2002. /spl delta/-clusters: capturing subspace correlation in a large data set, Proceedings 18th International Conference on Data Engineering, San Jose, CA, USA. pp. 517-528. doi:10.1109/icde.2002.994771.

Yamada ABF, de Freitas PL, da Silva RF, Souto FJD, 2021. Trends and spatial distribution of Hepatitis D in the North of Brazil, 2009-2018: an ecological study. Epidemiol Serv Saude Brasília 30:e2020867.