Received: August 1, 2022. Revised: September 13, 2022.

International Journal of
Intelligent Engineering & Systems

http://www.inass.org/

Quality Evaluation of Wind Energy Data with Complete Linkage Clustering

Sakon Klongboonjit[1] **Tossapol Kiatcharoenpol**[1]*

[1]*Industrial Engineering Department, School of Engineering, King Mongkut's Institute of Technology Ladkrabang, Bangkok, Thailand*
* Corresponding author's Email: tossapol_k@yahoo.com

Abstract: Although wind is an important free energy and most investors or farmers would like to invest in wind energy projects, they sometimes lack of the wind quality data in alternative areas for making decision. It should be definitely good to have some simple methods to classify the quality of wind energy for alternative areas. In this study, Complete Linkage method combining with the Euclidean distance calculation, which is really a simple method for users, is introduced to cluster wind energy quality of alternative areas. In a case of 13 alternative areas in the south of Thailand, the data of average wind velocity along 12 months from the secondary data source can be used to generate the initial distance matrix before continuously improving with Complete Linkage method. Finally, these 13 alternative areas are suitable clustered at C.D. = 5.11 into 3 groups of the low wind quality area with I.D. = 1.21, the medium wind quality area with I.D. = 1.41 and the high wind quality area with I.D. = 1.45.

Keywords: Complete linkage method, Wind energy potential, Wind quality evaluation, Wind quality cluster.

1. Introduction

Energy is one of basic human needs and is really important in our daily lives. In Thailand, the final total energy consumption in one month of January 2022 was really high about 7,174 ktoe of US$ 3,400 million and almost 80% of the final total energy consumption is from fossil fuel [1]. To decrease the final energy consumption by fossil, renewable energy from natural sources should be increasingly invested. Currently, the electricity of Thailand is cumulative generated form many kinds of renewable energy: solar energy 500 MW, wind energy 800 MW, hydroelectric energy 324 MW, biogas 120 MW and waste energy 160 MW [2]. It is clearly that cumulative electric energy form renewable energy is mostly from solar energy and wind energy since Thailand is located near the equator. That is why researchers in Thailand are so interested in studying these two renewable energy sources. However, some researchers would have liked to study more wind energy since the wind blows all day and night to generate the electric energy all 24 hours. Data of wind energy in Thailand from Department of Alternative Energy Development and Efficiency (DEDE) [2] show that the average of wind speed in Thailand is around 3-5 m/s or in low to moderate level and the range of these wind speed is high enough to continuously generate energy all 24 hours in a day. There are many researchers studying on the wind potential of Thailand. Thitipong, Adun and Arne [3] used hourly wind speed data from year 2008 to 2010 at the height of 10 m, 30 m and 40 m of Ubonratchathani province in Thailand to determine the potential of wind power generation. Sakkarin and Somchai [4] reviewed the status of wind energy in Thailand and concluded that wind energy was agreed to be one of good renewable energy sources to generate electric power. Pham and Thananchai [5] statistically analyzed one year wind measurement data of three sites to investigate wind energy potential at high level at the height of 65 m, 90 m and 120 m. Their results introduced that the wind turbines with low cut-in wind speed were strongly recommended to install at these three sites. From [3-5], these researchers had to collect and measure all related wind data by themselves with long data collection time and high budget cost. These are so

Received: August 1, 2022. Revised: September 13, 2022.

important obstacles for Thai investors who would like to invest in the wind energy projects but have not any wind measuring instruments and enough budgets.

According to above researches, it would be good if some methods are introduced to evaluate wind quality in areas with secondary wind data for Thai investors to reduce data collecting time and budgets. Sakon and Tossapol [6] applied Analytical Hierarchy Process (AHP) to secondary wind data [7]. The calculation results showed that AHP could prioritize wind energy of alternative areas with the data of average wind velocity generally receiving from the secondary data sources. Although the investors can indicate the best alternative area among many areas with AHP, it is still not simple for them to cluster alternative areas into a few groups with the data of average wind velocity. To achieve clustering alternative areas with their secondary wind data, the agglomerative hierarchical cluster methods should be applied to these secondary wind data because the agglomerative hierarchical cluster methods are multivariate method in statistics which are uncomplicated to understand, simply to interpreting the results and easy to simultaneously analyze multi-variables. Hamid [8] formed 11 machines into machine cells base on their components. Two clustering methods of Single Linkage and Average Linkage were applied to this problem. These 11 machines were formed into 2 machine cells with Single Linkage method but some bottleneck machines were not properly grouped while Average Linkage method formed these 11 machines into 3 machine cells and all bottleneck machines were properly put into the suitable machine cell. Odilia and Kylee [9] applied three agglomerative clustering methods of Single Linkage, Complete Linkage and Average Linkage to group 77 bilingual people who use Cantonese and English languages. This study showed the way to group bilingual people base on different variable data. According to dataset of 3 variables: daily use of both languages, Cantonese proficiency and English proficiency, all people were grouped into 4 groups: group A of frequent language users with high Cantonese proficiency, group B of frequent language users with intermediate Cantonese proficiency, group C of moderate language users with high Cantonese proficiency and group D of moderate language users with high Cantonese proficiency. Noor, Sabri and Safiek [10] compared performances of Single Linkage method and Complete Linkage method by clustering 200 tourists in Kapas Island into subgroups. With these two methods, the results showed that Complete Linkage method was more proper method to analyze and group these 200 tourists than Single Linkage method since the chain effect grouped the 191 tourists into one group. Vijava, Shweta and Neha [11] comparatively studied three methods of Single Linkage method, Complete Linkage method and Ward method. In their studying, these three methods were applied to cluster the shopping customers by using their total income and their expenditure on shopping. From customers' dataset, Single Linkage method could not clearly cluster shopping customers while Complete Linkage method and Ward method similarly clustered all customers into 5 groups. Yanuwar, Nurissaidah, Tony and Moh [12] applied three methods of Single Linkage method, Complete Linkage method and Average Linkage method to cluster people's community welfare in East Java. The results showed that methods of Complete Linkage and Average Linkage were really suitable methods to cluster people into 3 groups with data of their community welfare while Single Linkage again clustered most of people into one big group. Hassan, Reda and Araby [13] introduced a method of Modified Single Linkage to the application of cellular manufacturing. With this method, all similar coefficients were generated with non-Jaccard's measure technique before forming the initial similarity matrix. This modified method gave really suitable machine cellular formation. Abdullah, Robert and Sanguthevar [14] studied dataset of patients' record of health information. The problem was these data from different health agencies and one patient could have many records in different forms from the health agencies. This was not easy to identify all records belonging to the same patient. With Complete Linkage method, the similar records from the same patient were grouped together nearly 100% accuracy. Lenin, Liliana and Enrique [15] compared traffic conflicts occurring at two-lane roundabout and turbo-roundabout with Complete Linkage method. The results showed that the number of traffic conflicts at turbo-roundabout were lower than them at two-lane roundabout 72%. Moreover, there were similar traffic conflicts at turbo-roundabout so that these similar traffic conflicts were grouped together with Complete Linkage method while there were many kinds of non-similar traffic conflicts at two-lane roundabout. Tanongsak and Sakon [16] introduced the method of connection correlation to generate the initial similarity matrix before analyzing with Complete Linkage method. This technique was applied to group 105 machine components in an industrial printer powertrain set. All 105 machine components were suitable clustered into 7 groups with different numbers of surface interconnection.

Researches [8-16] really show that the agglomerative hierarchical cluster methods can be applied to different types of datasets to cluster data into groups with multi-variables. However, Single Linkage method may not good to cluster dataset with closely values. Meanwhile, methods of Complete Linkage, Average Linkage or Ward method are general method that can be applied to general dataset. Among of these cluster methods, Complete Linkage method is quite of simple and easy to understand for most users. Moreover, non-Jaccard's measure technique can be also employed together with Complete Linkage method to suitable cluster similar data into same group.

With this study, the quality of wind energy in alternative areas would be analyzed with simply method and low budget. Since the secondary wind dataset [7] already informs the values of wind power in each alternative areas so that users can conveniently rank from the lowest wind energy area to the highest wind energy area, however, these datasets should be more useful if the data in this wind dataset can be clustered into groups with same wind quality in each group. Among many agglomerative hierarchical cluster methods [8-16], the simplest method without the effect of the closely data values and using together with other measure techniques should be Complete Linkage Method.

2. Purpose

The main purpose of this study is to introduce Complete Linkage method with non-Jaccard's measure technique of simply Euclidean distance method to cluster the secondary wind dataset [7] into the area groups of same wind quality in each group. This method would be expected to be a suitable method of evaluating wind quality and site selections for Thai investors or farmers in rural areas who have not enough budgets and time during site selecting process before invest in their green energy projects.

3. Research methodology

To cluster quality of wind energy data, there are 4 main steps as following: 1) Select alternative areas for wind energy investment and collect wind energy data, 2) Calculate a pairwise Euclidean distance value for wind energy data of alternative areas and generate an initial distance matrix, 3) Improve initial distance matrix with Complete Linkage method and 4) Develop the dendrogram for all alternative areas and select the suitable value of the cluster distance.

3.1 Select alternative areas for wind energy investment and collect wind energy data

In the first step, the secondary data of wind energy [7] for N alternative areas are searched. These data consist of the average wind velocity at the height level of 40 m along 12 months of a year from 41 wind data centers in Thailand. There are 16 wind datasets for north, 3 wind dataset for middle and west, 9 north-east, 6 wind dataset for east and 13 wind datasets for south. The average wind velocity in Thailand at the height level of 40 m does not exceed 4 m/s.

3.2 Calculate a pairwise euclidean distance for wind energy data of alternative areas and generate an initial distance matrix

In this step, the wind energy data of N alternative areas in the first stage are pairwise calculated with the Euclidean distance in Eq. (1) to receive their pairwise distance value of wind energy for alternative area i and j (D_{ij} or D_{ji}). This pairwise distance value (D_{ij}) depends on the average wind velocity of alternative area i of the k^{th} month (x_{ik}) and the average wind velocity of alternative area j of the k^{th} month (y_{ik})

$$D_{ij} = D_{ji} = \sqrt{\sum_{k=1}^{12}(y_{ik} - x_{ik})^2}, i \neq j$$

and

$$D_{ii} = 0 \qquad (1)$$

Finally, an N x N initial distance matrix as shown in Fig. 1 can be generated from all D_{ij} and D_{ji} values. Next, this initial distance matrix will be improved with Complete Linkage method to group alternative areas with same wind quality level together.

3.3 Improve initial distance matrix with complete linkage method

To group alternative areas with same wind quality level together, the N x N initial distance matrix in Fig. 2 should be improved with 2 steps of determine cluster distance values and develop new

	A_1	A_2	A_3	…	A_N
A_1	0	D_{12}	D_{13}	…	D_{1N}
A_2	D_{21}	0	D_{23}	…	D_{2N}
A_3	D_{31}	D_{32}	0	…	D_{3N}
⋮	⋮	⋮	⋮	⋱	⋮
A_N	D_{N1}	D_{N2}	D_{N3}	…	0

Figure. 1 The initial distance matrix (N x N) for N alternative areas

	A_2, A_3	A_1	A_4	…	A_N
A_2, A_3	0	$E_{(2,3),1}$	$E_{(2,3),4}$	…	$E_{(2,3),N}$
A_1	$E_{1,(2,3)}$	0	D_{14}	…	D_{1N}
A_4	$E_{4,(2,3)}$	D_{41}	0	…	D_{4N}
⋮	⋮	⋮	⋮	⋱	⋮
A_N	$E_{N,(2,3)}$	D_{N1}	D_{N4}	…	0

Figure. 2 New distance matrix (N-1 x N-1) for N alternative areas

Table 1. The relationship of all C.D. values and the matrix size of their distance matrix for N alternative areas

The ith improving round	Cluster Distance (C.D.)	The matrix size of the distance matrix in each improving round	The number of group
0	-	N x N	N
1	C.D.$_1$	(N-1) x (N-1)	N-1
2	C.D.$_2$	(N-2) x (N-2)	N-2
3	C.D.$_3$	(N-3) x (N-3)	N-3
⋮	⋮	⋮	⋮
N-1	C.D.$_{N-1}$	1 x 1	1

distance matrix. Then, these 2 steps should be repeated for N-1 rounds until all alternative areas are group together into one group.

Step1: Determine Cluster Distance Value (C.D.)

In the distance matrix, identify the lowest value of all matrix cells. In Fig. 1, if the lowest value of D_{ij} is D_{23} or D_{32} ($D_{23} = D_{32}$), the cluster distance value for the nth improving round is D_{23} or D_{32}. If there are 2 or more the lowest values of D_{ij}, the cluster distance can be any equaling D_{ij} values.

Step 2: Develop New Distance Matrix

After the cluster distance value (D_{ij}) is selected, the alternative areas of A_i and A_j can be combined into the same group. If the cluster distance value (C.D.) is D_{23} or D_{32}, area 2 and 3 are grouped together and new distance matrix is as shown in Fig. 2. All values of $E_{i,(j,k)} = E_{(j,k),i}$ are calculated with Eq. (2) base on the maximum distance of nearby groups.

$$E_{i,(j,k)} = E_{(j,k),i} = \max \{D_{ij}, D_{ik}\} \qquad (2)$$

Repeat Step 1 and Step 2 for N-1 rounds, then, all alternative areas should be grouped together into one group. With the N alternative areas, the relationship of all C.D. values, the matrix size of their distance matrix and the number of group can be shown in Table 1.

3.4 Develop the dendrogram for all alternative areas and select the suitable value of the cluster distance

In this step, all cluster distance values (C.D.'s) and their distance matrix are transformed into the dendrogram as shown in Fig 3. Finally, the suitable value of the cluster distance (C.D.) should be selected by analysts to cluster al N alternative areas into the proper numbers of group. With the good selected C.D. value, each area group should have the closely value of internal group average distance (I.D.) that is the average distance of group members and their group centroid point and the two-group centroid distance (G.D.) should be longer than all I.D.'s.

4. Results and discussion

For this study, 13 areas in the south of Thailand are selected to be alternative areas since on the left side and right side of these 13 alternative areas is Andaman Sea and Gulf of Thailand, respectively. That means they are the possible potential areas for wind green energy projects. The Locations of these alternative areas are shown in Table 2 in the latitude and longitude form [17]. The average wind velocity data at the height level of 40 m for these 13 alternative areas [17] and their wind power values are shown in Table 3. Generally, the area with higher wind power value should clearly have more wind energy potential.

According to the data of wind power in Table 3, the wind energy potential of these 13 alternative areas can be clearly ranked from the highest wind

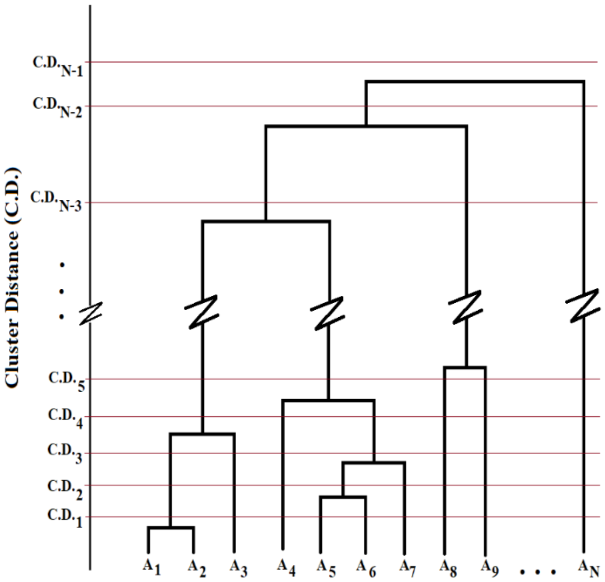

Figure. 3 The example of dendrogram for N alternative areas

potential area to the lowest wind potential area as follows: A_1, A_2, A_3, A_4, A_5, A_6, A_7, A_8, A_9, A_{10}, A_{11}, A_{12} *and* A_{13}, respectively. The monthly data of average wind velocity at the height of 40 m for these 13 alternative areas are not quite clearly different so that it is not easy to cluster these 13 alternative areas into group base on their wind quality. To evaluate and cluster these 13 alternative areas into groups, Complete Linkage method is applied to these data in Table 3. Before applying Complete Linkage method, the initial distance matrix in Table 4 should be firstly generated with calculating the values of a Pairwise Euclidean Distance from the monthly data of average wind velocity at the height of 40 m with Eq. (1).

For the D_{ij} and D_{ji} calculation example, if area A_1 is compared with area A_2,

$$
\begin{aligned}
D_{12} &= D_{21} \\
&= \sqrt{\begin{matrix}(2.47-3.95)^2 + (2.92-3.73)^2 \\ +(2.41-2.44)^2+\ldots+(2.06-3.03)^2\end{matrix}} \\
&= 2.42 \text{ m/s}
\end{aligned}
$$

Note that D_{ii}'s are always equaling to 0.

Next, the cluster distance value (C.D.) is determined from the lowest value in Table 4 with C.D.$_1$ = 0.69. That means area A_6 and A_7 are grouped together since the wind quality of these 2 areas are

Table 2. The location of all 13 alternative areas [17]

Alternative areas	The location of alternative areas	
	North Latitude	East Longitude
A_1	7-32-58	99-35-46
A_2	8-21-27.7	99-24-58
A_3	7-08-57.1	100-33-13
A_4	9-36-07	98-28-39
A_5	7-46-01	100-18-00
A_6	9-22-27	99-57-04
A_7	7-58-28	98-58-05
A_8	8-19-41	98-15-50
A_9	6-57-16	100-49-11
A_{10}	7-06-00.4	100-40-12
A_{11}	7-21-26	100-11-16
A_{12}	6-56-54	101-16-41
A_{13}	7-53-41	100-20-46

Table 3. The average wind velocity data at the height level of 40 m for all 13 alternative areas [17]

Alternative Areas	Average Wind Velocity (m/s) at the height of 40 m												Wind Power (W/m^2)
	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec	
A_1	3.95	3.73	2.44	1.69	1.24	1.37	1.42	1.69	1.37	1.46	1.89	3.03	5.75
A_2	2.47	2.92	2.41	2.01	1.96	1.92	2.03	2.09	2.11	1.61	1.72	2.06	5.91
A_3	4.31	3.93	3.38	2.76	2.41	2.67	2.83	3.05	2.85	2.52	2.46	3.88	17.88
A_4	3.69	3.06	2.64	2.38	2.85	3.60	3.71	3.76	2.86	2.36	2.88	3.68	18.59
A_5	3.77	3.58	3.21	2.64	3.27	3.65	3.71	4.07	4.03	3.25	2.77	3.45	23.84
A_6	4.11	3.71	3.70	2.90	3.00	3.10	3.26	3.39	3.44	3.15	3.42	4.29	24.27
A_7	4.32	3.99	3.46	2.93	3.05	3.26	3.40	3.63	3.34	2.83	3.24	4.07	25.35
A_8	3.25	3.39	3.54	3.46	4.21	5.15	5.47	5.50	3.39	2.46	2.12	2.70	30.50
A_9	5.22	5.09	4.41	3.68	2.93	3.08	3.05	3.29	3.01	2.94	3.15	4.20	30.75
A_{10}	5.85	5.53	4.61	3.60	2.73	2.78	2.89	3.22	2.73	2.83	3.51	5.01	32.79
A_{11}	6.30	5.72	4.84	3.71	2.73	2.87	3.17	3.58	2.93	3.22	4.06	5.30	43.43
A_{12}	5.02	5.19	4.52	3.98	3.61	3.67	3.89	4.36	4.07	3.87	3.67	5.05	45.02
A_{13}	6.17	5.32	4.76	3.91	3.44	4.00	3.75	4.29	4.15	3.82	4.37	5.89	55.77

Table 4. The initial distance matrix for all 13 alternative areas

	A_1	A_2	A_3	A_4	A_5	A_6	A_7	A_8	A_9	A_{10}	A_{11}	A_{12}	A_{13}
A_1	0.00	2.42	3.67	4.75	5.75	5.15	5.10	8.01	5.49	5.85	6.77	7.57	8.33
A_2	2.42	0.00	3.67	4.00	4.90	4.84	4.81	6.79	5.75	6.54	7.45	7.49	8.58
A_3	3.67	3.67	0.00	2.10	2.49	1.70	1.56	5.11	2.33	3.13	3.98	4.04	4.99
A_4	4.75	4.00	2.10	0.00	1.78	2.06	1.80	3.81	3.59	4.49	5.11	4.43	5.46
A_5	5.75	4.90	2.49	1.78	0.00	1.72	1.54	3.37	3.20	4.28	4.74	3.41	4.68
A_6	5.15	4.84	1.70	2.06	1.72	0.00	0.69	4.57	2.15	3.02	3.56	2.91	3.93
A_7	5.10	4.81	1.56	1.80	1.54	0.69	0.00	4.21	1.98	2.93	3.50	2.86	3.91
A_8	8.01	6.79	5.11	3.81	3.37	4.57	4.21	0.00	5.28	6.33	6.65	4.93	6.16
A_9	5.49	5.75	2.33	3.59	3.20	2.15	1.98	5.28	0.00	1.29	2.01	2.41	3.19
A_{10}	5.85	6.54	3.13	4.49	4.28	3.02	2.93	6.33	1.29	0.00	1.05	2.78	2.94
A_{11}	6.77	7.45	3.98	5.11	4.74	3.56	3.50	6.65	2.01	1.05	0.00	2.56	2.27
A_{12}	7.57	7.49	4.04	4.43	3.41	2.91	2.86	4.93	2.41	2.78	2.56	0.00	1.66
A_{13}	8.33	8.58	4.99	5.46	4.68	3.93	3.91	6.16	3.19	2.94	2.27	1.66	0.00

at same quality level. With applying Eq. (2), all values of $E_{i,(6,7)} = E_{(6,7),i}$ are as follows:

$E_{1,(6,7)} = E_{(6,7),1} = \max\{5.15, 5.10\} = 5.15$,
$E_{2,(6,7)} = E_{(6,7),2} = \max\{4.84, 4.81\} = 4.84$,
$E_{3,(6,7)} = E_{(6,7),3} = \max\{1.70, 1.56\} = 1.70$,
$E_{4,(6,7)} = E_{(6,7),4} = \max\{2.06, 1.80\} = 2.06$,
$E_{5,(6,7)} = E_{(6,7),5} = \max\{1.72, 1.54\} = 1.72$,
$E_{8,(6,7)} = E_{(6,7),8} = \max\{4.57, 4.21\} = 4.57$,
$E_{9,(6,7)} = E_{(6,7),9} = \max\{2.15, 1.98\} = 2.15$,
$E_{10,(6,7)} = E_{(6,7),10} = \max\{3.02, 2.93\} = 3.02$,
$E_{11,(6,7)} = E_{(6,7),11} = \max\{3.56, 3.50\} = 3.56$,
$E_{12,(6,7)} = E_{(6,7),12} = \max\{2.91, 2.86\} = 2.91$,
$E_{13,(6,7)} = E_{(6,7),13} = \max\{3.93, 3.91\} = 3.93$,
and $E_{(6,7),(6,7)} = 0$.

Then, the initial distance matrix is improved to be the 1st distance matrix with C.D.$_1$ = 0.69 as seen in Table 5.

Again, the cluster distance for the 2nd distance matrix is determined from the lowest value in Table 5 with C.D.$_2$ = 1.05. For this 2nd improvement, area A_{10} is grouped with area A_{11} and all values of $E_{i,(10,11)} = E_{(10,11),i}$ are as follows:

$E_{1,(10,11)} = E_{(10,11),1} = \max\{5.85, 6.77\} = 6.77$,
$E_{2,(10,11)} = E_{(10,11),2} = \max\{6.54, 7.45\} = 7.45$,
$E_{3,(10,11)} = E_{(10,11),3} = \max\{3.13, 3.98\} = 3.98$,
$E_{4,(10,11)} = E_{(10,11),4} = \max\{4.49, 5.11\} = 5.11$,
$E_{5,(10,11)} = E_{(10,11),5} = \max\{4.28, 4.74\} = 4.74$,
$E_{8,(10,11)} = E_{(10,11),8} = \max\{6.33, 6.65\} = 6.65$,
$E_{9,(10,11)} = E_{(10,11),9} = \max\{1.29, 2.01\} = 2.01$,
$E_{12,(10,11)} = E_{(10,11),12} = \max\{2.78, 2.56\} = 2.78$,
$E_{13,(10,11)} = E_{(10,11),13} = \max\{2.94, 2.27\} = 2.94$,
$E_{(6,7),(10,11)} = E_{(10,11),(6,7)} = \max\{3.02, 3.56\} = 3.56$,
and $E_{(10,11),(10,11)} = 0$.

Finally, the 2nd distance matrix with C.D.$_2$ = 1.05 is generated as seen in Table 6.

Repeat the same steps, the nth distance matrix with its C.D. is gradually improved and all alternative areas are also gradually grouped together until all 13 alternative areas are grouped into one group after 12 repeat rounds. Table 7 shows the values of C.D., the size of distance matrix, the number of groups and the members of each group for each nth round. From the 1st round to the 9th round, there are only 12 from 13 alternative areas which are grouped. Finally, all 13 alternative areas are firstly grouped with 3 groups at the 10th round. After that, all 13 alternative areas are grouped into 2 groups at the11th round and all 13 alternative areas are finally grouped into one group at the 12th round.

Table 5. The 1st distance matrix for all 13 alternative areas

C.D.$_1$ = 0.69	A_6, A_7	A_1	A_2	A_3	A_4	A_5	A_8	A_9	A_{10}	A_{11}	A_{12}	A_{13}
A_6, A_7	0.00	5.15	4.84	1.70	2.06	1.72	4.57	2.15	3.02	3.56	2.91	3.93
A_1	5.15	0.00	2.42	3.67	4.75	5.75	8.01	5.49	5.85	6.77	7.57	8.33
A_2	4.84	2.42	0.00	3.67	4.00	4.90	6.79	5.75	6.54	7.45	7.49	8.58
A_3	1.70	3.67	3.67	0.00	2.10	2.49	5.11	2.33	3.13	3.98	4.04	4.99
A_4	2.06	4.75	4.00	2.10	0.00	1.78	3.81	3.59	4.49	5.11	4.43	5.46
A_5	1.72	5.75	4.90	2.49	1.78	0.00	3.37	3.20	4.28	4.74	3.41	4.68
A_8	4.57	8.01	6.79	5.11	3.81	3.37	0.00	5.28	6.33	6.65	4.93	6.16
A_9	2.15	5.49	5.75	2.33	3.59	3.20	5.28	0.00	1.29	2.01	2.41	3.19
A_{10}	3.02	5.85	6.54	3.13	4.49	4.28	6.33	1.29	0.00	1.05	2.78	2.94
A_{11}	3.56	6.77	7.45	3.98	5.11	4.74	6.65	2.01	1.05	0.00	2.56	2.27
A_{12}	2.91	7.57	7.49	4.04	4.43	3.41	4.93	2.41	2.78	2.56	0.00	1.66
A_{13}	3.93	8.33	8.58	4.99	5.46	4.68	6.16	3.19	2.94	2.27	1.66	0.00

Table 6. The 2nd distance matrix for all 13 alternative areas

C.D.$_2$ = 1.05	A_6, A_7	A_{10}, A_{11}	A_1	A_2	A_3	A_4	A_5	A_8	A_9	A_{12}	A_{13}
A_6, A_7	0.00	3.56	5.15	4.84	1.70	2.06	1.72	4.57	2.15	2.91	3.93
A_{10}, A_{11}	3.56	0.00	6.77	7.45	3.98	5.11	4.74	6.65	2.01	2.78	2.94
A_1	5.15	6.77	0.00	2.42	3.67	4.75	5.75	8.01	5.49	7.57	8.33
A_2	4.84	7.45	2.42	0.00	3.67	4.00	4.90	6.79	5.75	7.49	8.58
A_3	1.70	3.98	3.67	3.67	0.00	2.10	2.49	5.11	2.33	4.04	4.99
A_4	2.06	5.11	4.75	4.00	2.10	0.00	1.78	3.81	3.59	4.43	5.46
A_5	1.72	4.74	5.75	4.90	2.49	1.78	0.00	3.37	3.20	3.41	4.68
A_8	4.57	6.65	8.01	6.79	5.11	3.81	3.37	0.00	5.28	4.93	6.16
A_9	2.15	2.01	5.49	5.75	2.33	3.59	3.20	5.28	0.00	2.41	3.19
A_{12}	2.91	2.78	7.57	7.49	4.04	4.43	3.41	4.93	2.41	0.00	1.66
A_{13}	3.93	2.94	8.33	8.58	4.99	5.46	4.68	6.16	3.19	1.66	0.00

Table 7. Data of the C.D. values, the distance matrix size, the group numbers and the group members at each nth improving round for all 13 alternative areas

The nth round	The C.D. values	The distance matrix size	The group numbers	The group members				
				Group 1	Group 2	Group 3	Group 4	Group 5
0	-	13x13	-	-	-	-	-	-
1	0.69	12x12	1	A_6, A_7	-	-	-	-
2	1.05	11x11	2	A_6, A_7	A_{10}, A_{11}	-	-	-
3	1.66	10x10	3	A_6, A_7	A_{10}, A_{11}	A_{12}, A_{13}	-	-
4	1.70	9x9	3	A_3, A_6, A_7	A_{10}, A_{11}	A_{12}, A_{13}	-	-
5	1.78	8x8	4	A_4, A_5	A_3, A_6, A_7	A_{10}, A_{11}	A_{12}, A_{13}	-
6	2.01	7x7	4	A_4, A_5	A_3, A_6, A_7	A_9, A_{10}, A_{11}	A_{12}, A_{13}	-
7	2.42	6x6	5	A_1, A_2	A_4, A_5	A_3, A_6, A_7	A_9, A_{10}, A_{11}	A_{12}, A_{13}
8	2.49	5x5	4	A_1, A_2	A_3, A_4, A_5, A_6, A_7	A_9, A_{10}, A_{11}	A_{12}, A_{13}	-
9	3.19	4x4	3	A_1, A_2	A_3, A_4, A_5, A_6, A_7	A_9, A_{10}, A_{11}, A_{12}, A_{13}	-	-
10	5.11	3x3	3	A_1, A_2	A_3, A_4, A_5, A_6, A_7, A_8	A_9, A_{10}, A_{11}, A_{12}, A_{13}	-	-
11	6.65	2x2	2	A_1, A_2	A_3, A_4, A_5, A_6, A_7, A_8, A_9, A_{10}, A_{11}, A_{12}, A_{13}	-	-	-
12	8.58	1x1	2	A_1, A_2, A_3, A_4, A_5, A_6, A_7, A_8, A_9, A_{10}, A_{11}, A_{12}, A_{13}	-	-	-	-

Table 8. The summary results of grouping all 13 alternative areas with complete linkage method

	The grouping results at C.D. = 5.11 with Complete Linkage method												
	The low wind quality area		The medium wind quality area						The high wind quality area				
	A_1	A_2	A_3	A_4	A_5	A_6	A_7	A_8	A_9	A_{10}	A_{11}	A_{12}	A_{13}
Wind power (W/m^2)	5.75	5.91	17.88	18.59	23.84	24.27	25.35	30.50	30.75	32.79	43.43	45.02	55.77
	I.D. = 1.21		I.D. = 1.41						I.D. = 1.45				
	G.D. of Low and Medium wind quality group = 4.77												
	G.D. of Low and High wind quality group = 6.80												
	G.D. of Medium and High wind quality group = 3.52												

Finally, the dedrogram for these 13 alternative areas is developed with the Complete Linkage method as seen in Fig. 4. If the data of wind power in Table 3 are considered together with data in Table 7 and Fig. 4, the summary results for these 13 alternative areas are concluded as shown in Table 8. In this study, these 13 alternative areas can be suitable grouped into 3 groups at the value of C.D. = 5.11: Group 1 for the low wind quality area consists of 2 members of A_1 and A_2 with I.D. =1.21, Group 2 for the medium wind quality area consists of 6 members of A_3, A_4, A_5, A_6, A_7 and A_8 with I.D. = 1.41 and Group 3 for the high wind quality area consists of 5 members of A_9, A_{10}, A_{11}, A_{12} and A_{13} with I.D. = 1.45. Also, all values of G.D. of any two groups are 4.77, 6.80 and 3.52 surely longer than all values of I.D. that means these 13 alternative areas are well grouped at this C.D. = 5.11.

Comparing the members in each group in Table 8 with the wind power of these 13 alternative areas in Table 3, members of Group 1 are from the first two areas (The wind power values of 5.75 W/m^2 and 5.91 W/m^2), members of Group 2 are from the next six areas (The wind power values of 17.88

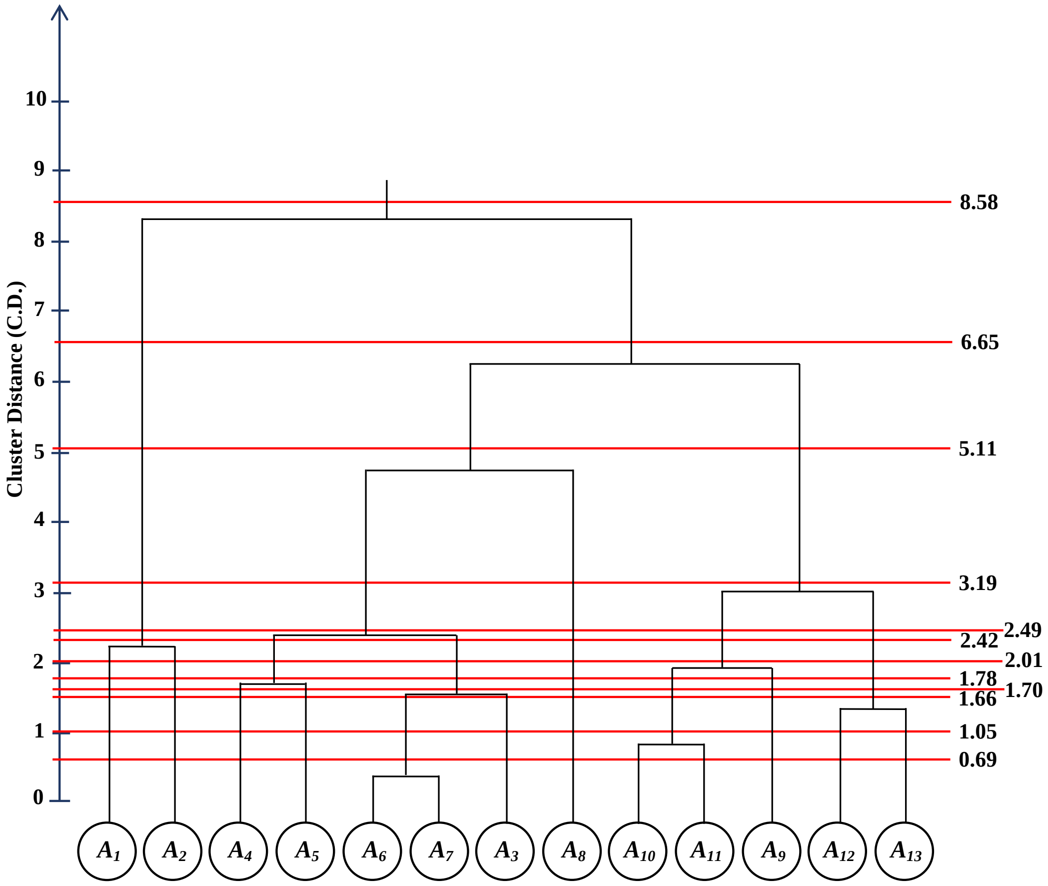

Figure. 4 The dendrogram for all 13 alternative areas with complete linkage method

W/m^2, 18.59 W/m^2, 23.84 W/m^2, 24.27 W/m^2, 25.35 W/m^2 and 30.50 W/m^2) and members of Group 3 are from the last five areas (The wind power values of 30.75 W/m^2, 32.79 W/m^2, 43.43 W/m^2, 45.02 W/m^2 and 55.77 W/m^2). That means the calculation results of Complete Linkage method with the Euclidean distance just group areas of same wind quality together without having any conflict with the area ranking order by the wind power data. If the wind secondary dataset of alternative areas are available, this method should be applied to these data since the steps of calculation are not too complex like other agglomerative hierarchical cluster methods and it can also evaluate the appropriate of the cluster results with the values of 1) the internal group average distance (I.D.) and 2) the two-group centroid distance (G.D.).

5. Conclusion

This study is to introduce the non-complex way to evaluate and cluster wind energy quality of alternative areas. With Complete Linkage method combining with the Euclidean distance calculation, it can be simply applied to the secondary data of average wind velocity along 12 months of a year in 13 places in Thailand Southern alternative areas that is quite different from the conventional method with high budget of data collecting and analyzing. The results shows that all 13 alternative areas can be suitable clustered into 3 groups of the low wind quality area, the medium wind quality area and the high wind quality areas. If this method is applied to all big secondary dataset of average wind velocity of Thailand, the map of wind quality cluster should be developed and this map should be really one of useful tools for Thai investors or farmers who would like to invest in wind energy projects in the future.

Conflicts of Interest

The authors declare no conflict of interest.

Author Contributions

The paper conceptualization, methodology, software, validation, formal analysis, investigation, resources, data curation, writing-original draft preparation, writing-review and editing and visualization have been done by the 1st author. The supervision and project administration have been done by the 2nd author.

References

[1] Department of Alternative Energy Development and Efficiency, Ministry of Energy dataset: *https://www.dede.go.th/ewtadmin/ewt/dede_web/download/state_65/1_SIT_Jan%20 65.pdf*, [Online; accessed 1 July 2022].

[2] Department of Alternative Energy Development and Efficiency, Ministry of Energy dataset: *https://webkc.dede.go.th/testmax/landing_kc2?category=11*, [Online; accessed 1 July 2022].

[3] T. Unchai, A. Janyalertadum, and A. E. Hold, "Wind Energy Potential Assessment as Power Generation Source in Ubonratchathani Province, Thailand", *Wind Engineering*, Vol. 36, No. 2, pp. 131-144, 2012.

[4] S. Chingulpitak and S. Wongwises, "Critical Review of the Current Status of Wind Energy in Thailand", *Renewable and Sustainable Energy Reviews*, Vol. 31, pp. 312-318, 2014.

[5] P. Quan and T. Leephakpreeda, "Assessment of Wind Energy Potential for Selecting Wind Turbines: An Application to Thailand", *Sustainable Energy Technologies and Assessments*, Vol. 11, pp. 17-26, 2015.

[6] S. Klongboonjit and T. Kaitcharoenpol, "Prioritization of Wind Energy Data with Analytical Hierarchy Process (AHP)", *International Journal of Intelligent Engineering and Systems*, Vol. 14, No. 6, pp. 369-376, 2021, doi: 10.22266/ijies2021.1231.33.

[7] Department of Alternative Energy Development and Efficiency, Ministry of Energy dataset: *http://www2.dede.go.th/km_it/windmap40m/windmap40m.html*, [Online; accessed 1 July 2022].

[8] H. Seifoddini, "Machine Grouping-Expert Systems: Comparison between Single Linkage and Average Linkage Clustering Techniques in formin Machince Cells", *Computers and Industrial Engineering*, Vol. 11, No. 1, pp. 210-216, 1988.

[9] O. Yim and K. T. Ramdeen, "Hierarchical Cluster Analysis: Comparison of Three Linkage Measures and Application to Psychological Data", *The Quantitative Methods for Psychology*, Vol. 11, No. 1, pp. 8-21, 2015.

[10] N. R. Rashidah, A. Sabri, and M. Safiek, "A Comparison between Single Linkage and Complete Linkage in Agglomerative Hierarchical Cluster Analysis for Identifying Tourists Segments", *IIUM Engineering Journal*, Vol. 12, No. 6, pp. 105-116, 2011.

[11] S. Vijaya, S. Sharma, and N. Batra, "Comparative Study of Single Linkage, Complete Linkage, and Ward Method of Agglomerative Clustering", In: *Proc. of International Conference on Machine Learning, Big Data, Cloud and Parallel Computing (Com-IT-Con)*, pp. 568-573, 2019.

[12] Y. Reinaldi, N. Ulinnuha, T. Hartono, and M. Hafiyusholeh, "Comparison of Single Linkage, Complete Linkage, and Average Linkage Methods on Community Welfare Analysis in Cities and Regencies in East Java", *Jurnal Matematika, Statistika & Komputasi*, Vol. 18, No. 1, pp. 130-140, 2021.

[13] H. M. Selim, R. M. S. A. Aal, and A. I. Majdi, "Formation of Machine Groups and Part Families: A Modified SLC method and Comparative Study", *Integrated Manufacturing Systems*, Vol. 14, No. 2, pp. 123-137, 2003.

[14] A. A. Mamun, R. Aseltine, and S. Rajasekaran, "Efficient Record Linkage Algorithms using Complete Linkage Clustering", *PLoS ONE*, Vol. 11, No. 4, pp. 1-21, 2016.

[15] L. B. Cruz, L. Lyons, and E. Darghan, "Complet-Linkage Clustering Analysis of Surrogate Measures for Road Safety Assessment in Roundabouts", *Revista Colombiana de Estadistica-Applied Statistics*, Vol. 44, No. 1, pp. 91-121, 2021.

[16] T. Kongsin and S. Klongboonjit, "Machine Component Clustering with Connection Correlation Method", *International Journal of Intelligent Engineering and Systems*, Vol. 14, No. 3, pp. 537-544, 2021, doi: 10.22266/ijies2021.0630.45.

[17] Department of Alternative Energy Development and Efficiency, Ministry of Energy dataset: *http://www2.dede.go.th/km_it/windmap40m/windmap40m.html*, [Online; accessed 1 July 2022].