International Journal of GEOMATE, Sept., 2016, Vol. 11, Issue 25, pp. 2422-2428
Geotec., Const. Mat. & Env., ISSN: 2186-2982(Print), 2186-2990(Online), Japan

OPTIMAL DESIGN OF RAIN GAUGE NETWORK IN JOHOR BY USING GEOSTATISTICS AND PARTICLE SWARM OPTIMIZATION

Mohd Khairul Bazli Mohd Aziz[1], Fadhilah Yusof[2], Zalina Mohd Daud[3], Zulkifli Yusop[4], Mohammad Afif Kasno[5]
[1]Centre of Preparatory and General Studies, TATI University College, Malaysia, [1,2]Department of Mathematical Sciences, Faculty of Science, Universiti Teknologi Malaysia, Malaysia, [3]UTM Razak School of Enggineering and Advanced Technology, Universiti Teknologi Malaysia, Malaysia, [4]Institute of Environmental and Water Resource Management (IPASA), Faculty of Civil Enggineering, Universiti Teknologi Malaysia, Malaysia, [5]Malaysia – Japan International Institute of Technology (MJIIT),Universiti Teknologi Malaysia, Malaysia.

ABSTRACT: This study proposes particle swarm optimization (PSO) approach to determine the optimal number and locations for the optimal rain gauge network in Johor state. The existing network of 84 rain gauges in Johor is also restructured into new locations by using daily rainfall, humidity, solar radiation, temperature and wind speed data collected during the monsoon season (November – February) of 1975 until 2008. This study used the combination of geostatistics method (variance-reduction method) and particle swarm optimization as the algorithm of optimization during the restructured proses. The numerical result shows that the new rain gauge location provides minimum value of estimated variance. This shows that the proposed method can serve as an analysis tool for a decision making to assist hydrologist in the selection of prime sites for the installation of rain gauge stations.

Keywords: Rainfall Network, Geostatistics, Particle Swarm Optimization

1. INTRODUCTION

Collectively, rainfall data deliver important input for effective planning, designing, operating and managing of water resources projects. Rainfall data are employed in numerous water resources management tasks such as water budget analysis and assessment, flood analysis and forecasting, streamflow estimation, and design of hydraulic structures. A reliable and optimal rain gauge network can provide accurate and precise rainfall data that is crucial for effective and economic design of hydraulic structures for flood control. This will help the researchers to minimize the hydrological and economic risk and errors involved in different water resources projects. Rain gauge network is usually installed to facilitate the direct measurement of rainfall data that characterize the spatial and temporal variations of local rainfall patterns in a catchment [14]. A rain gauge network should be denser than networks used to measure other meteorological elements (e.g. temperature), because the highly variable rainfall patterns and its spatial distribution cannot be represented effectively without having a network of enough spatial density [2]. A well-designed rain gauge network thus should contain a sufficient number of rain gauges, which reflect the spatial and temporal variability of rainfall in a catchment [3].

In hydrological studies, the rain gauge network used is often sparsed and thus incapable of providing adequate rainfall estimation necessary for effective hydrological analysis and design of water resources projects. Significant design errors in the water resources projects will eventually cause in the immense loss of lives and property damages resulted from use of inaccurate rainfall data.

Thus, identification and selection of the optimal rain gauge network configuration with optimal number and locations of rain gauge stations is the main objective of the network design. Hence, the optimal rain gauge network should contain the number and locations of rain gauge stations in such a way that it can produce optimum rainfall information and data with minimum uncertainty and cost [2,4]. One can approach the problem either by removing redundant stations from the network to minimize the cost or by expanding the network with installation of additional stations to reduce the estimation uncertainty [5]. Some studies applied the kriging technique in combination with other techniques such as entropy [3] and multivariate factor analysis [7] for the network design. A few studies also combined optimization method based on simulation tools (e.g. simulated annealing) with the kriging technique [2, 8, 9] to obtain the optimal rain gauge network.

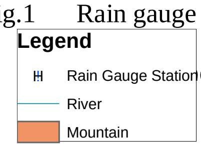
A network design methodology was developed in this study to determine optimal number and locations of the existing stations in the current rain gauge network located in the Johor state, Malaysia. The procedure involves a methodical search for the optimal number and locations of rain gauge stations in the network that minimize estimated variance.

The methodology presented in this study is in line with that of [2] who used the kriging-based geostatistical and simulated annealing as an optimization method to determine the optimal number and location of rain gauge stations in the existing The major contribution here is that unlike the work of [2], the developed methodology considered the particle swarm optimization technique as an optimization method to provide fast convergence optimization procedure. The rest of the paper has been organized as follows. First, details of the study area and datasets used are presented, which is followed by the methodology. The results are summarized next and finally, the conclusions are drawn.

2. STUDY AREA AND DATA DESCRIPTION

Johor is the second largest state in the Malaysia Peninsular, with an area of 18,941 km^2. The rivers in Johor and its streams are important sources of water supply for the people of Johor. The catchment area contains a dense rain gauge network, 84 rain gauges covering 19,210 km^2 in Johor (see Figure 1). The current overall network density for the whole of Johor state is about 194 km^2 per gauge, which surpasses even the ideal WMO recommendation of one gauge per 600 – 900 km^2 for flat areas. In fact, it also fulfills the ideal density for mountainous region of 100 – 200 km^2 per gauge. The data used to perform the analysis was from the daily rainfall, elevation, humidity, temperature, solar radiation and wind speed measurement from November until February of 1975 through 2008 from 84 rain gauge stations that are located all over Johor. Figure 4 shows the mapping of the humidity, temperature, solar radiation, wind speed and elevation in Johor. From Figure 4, the highest humidity area is the centre of Johor, west side of Johor receives more heat based on the solar radiation and the temperature, and the wind speed is highest along the coast. The data was obtained from Department of Irrigation and Drainage (DID) Malaysia and Malaysia Meteorological Department (MMD).

Fig.1 Rain gauge locations



3. GY

3.1 Geostatistical Method

The network design problem consists in obtaining the number N and the location of rain gauges stations that give the best estimate areal mean rainfall. The estimation variance σ^2 is a basic tool of variance reduction techniques for optimal selection of sampling locations. For the application of the variance reduction method to optimal location of sampling sites, a semivariogram must be modelled.

A semivariogram, $\gamma(h)$ is one of the significant functions to indicate spatial correlation in observations measured at sample locations. Semivariogram is represented as a graph that shows the difference in measure with distance between all pairs of sampled locations. The estimated variance depends on the semivariogram model, the number N of rain gauges and its spatial location. Therefore, choosing an appropriate semivariogram model is vital in determined the optimal estimation variance.

Let h be the lag or distance, and Z be an intrinsic random function and let $Z(x_i)$, for $i = 1,2,\ldots,N$ be a sampling of size N. Then the following expression is an unbiased estimator for the semivariogram of the random function:

$$\gamma(h) = \frac{1}{2n(h)}\sum_{i=1}^{n(h)}\left[Z(x_i + h) - Z(x_i)\right]^2 \quad (1)$$

Equation 1 is used to compute experimental semivariogram from the data under study. By changing h, both in distance and direction, a set of the sample (or experimental) semivariograms for the data is obtained [10].

The exponential semivariogram models are selected to fit the data. The exponential semivariogram model equation is as follow:

$$\gamma(h) = C\left(1 - e^{-\frac{3h}{a}}\right) \quad (2)$$

where C is the sill and a is the range.

Once the model of the semivariogram is fixed, the estimation variance only depends on the number N and the location of the rain gauges. To calculate the estimation variance using ordinary kriging,

$$\sigma^2(x_0) = 2\sum_{i=1}^{k}\lambda_i\gamma(x_i, x_0) - \sum_{i=1}^{k}\sum_{j=1}^{k}\lambda_i\lambda_j\gamma(x_i, x_j) \quad (3)$$

Where

$$\hat{Z}(x_0) = \sum_{i=1}^{k} \lambda_i Z(x_i) \quad (4)$$

Subject to $\sum_{i=1}^{k} \lambda_i = 1$.

This is an algorithm for the ordinary kriging estimation to calculate the estimation variance (Olea, 2003):

1. Calculate each term in matrix *G*.
   Let x_i's be the sampling sites of a sample subset of size k, $i = 1,2,\ldots,k$ and let $\gamma(x_i, x_j)$'s be the experimental variogram. Then the *G* is the matrix

$$G = \begin{bmatrix} \gamma(x_1,x_1) & \gamma(x_2,x_1) & \cdots & \gamma(x_k,x_1) & 1 \\ \gamma(x_1,x_2) & \gamma(x_2,x_2) & \cdots & \gamma(x_k,x_2) & 1 \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ \gamma(x_1,x_k) & \gamma(x_2,x_k) & \cdots & \gamma(x_k,x_k) & 1 \\ 1 & 1 & \cdots & 1 & 0 \end{bmatrix} \quad (5)$$

2. Calculate each term in matrix g.
   Let x_0 be the estimation location, then the *g* is the matrix

$$g = [\gamma(x_0,x_1) \quad \gamma(x_0,x_2) \quad \cdots \quad \gamma(x_0,x_k) \quad 1]' \quad (6)$$

3. Solve the system of equations

$$GW = g,$$
$$W = gG^{-1},$$

   Where $W = [\lambda_1 \quad \lambda_2 \quad \cdots \quad \lambda_k \quad -\mu]'$.

4. Calculate the ordinary kriging estimation variance

$$\sigma^2(x_0) = g'W = g'G^{-1}g. \quad (7)$$

3.2 Partical Swarm Optimization

PSO was developed by Kennedy and Eberhart, [12] in 1995, inspired by the social behavior of bird flocking. The movement of each swarming particle is determined by a combination of a stochastic element and a deterministic element, [1]. A population (swarm) of particles is initialized in an *n*-dimensional search space in which each particle $x_i = (x_{i1}, x_{i2}, \ldots, x_{in})$ represents a possible solution. Each particle is aware of its current position, its own personal best position, its current velocity and the single global (or local) best position. The global best position is represented as $g_i = (g_{i1}, g_{i2}, \ldots, g_{in})$ and symbolizes the best position of all particles in the population (swarm). The personal best position represents the best position found by a particle so far and is denoted as $p_i = (p_{i1}, p_{i2}, \ldots, p_{in})$.

The velocity $v_i = (v_{i1}, v_{i2}, \ldots, v_{in})$ gives the position change of a particle. In the original proposed PSO of Kennedy and Eberhart, the new velocity for each particle is calculated according to Eq. (1). To update the new position of each particle, Eq. (2) is used:

$$v_i = v_i + c_1.r_1.(p_i - x_i) + c_2.r_2.(g - x_i) \quad (8)$$

$$x_i = x_i + v_i \quad (9)$$

where $i = 1,2,\ldots,N$, with *N* as population size. c_1 and c_2 are two positive constants; r_1 and r_2 are two random numbers with range (0,1).

In PSO, Eq. (8) is used to calculate the new velocity according to its previous velocity and to the distance of its current position from both its own best historical position and the best position of the entire population or its neighbourhood. Generally, the value of each component of *v* can be constricted by setting the maximum velocity $v_{\max}$ to the upper and lower bounds of the decision variable range. Then the particle flies toward a new position according to Eq. (9). This process is repeated until a user-defined stopping criterion is reached.

The minimisation of objective function given in Eq. (3) is a problem of combinatorial optimization. A typical procedure of rain gauge network design has to look for a combination among all rain gauge stations in such a way that minimizes the estimation variance and/or maximizes the information content for the observed data. This can be achieved either by optimal positioning of additional and redundant stations or simply removing redundant stations that forms the scope of this paper.

The steps of PSO algorithms applied for determined optimal rain gauge network as follow:

1. Randomly initialize the *x* particles each representing different combinations of *N* rain gauge placements.
2. Evaluate each particle in the population using the fitness function, Eq. (3) and minimum fitness function is set as *pbest*.
3. Compare particle's fitness evaluation with particle's *pbest*. If current value is < than *pbest*, then set *pbest* as *gbest*.
4. Generate a new set of *x* particles by Eq. (8).

5. Stop the algorithm if the stopping criterion is satisfied or maximum number of iterations is reached.

4. RESULTS AND DISCUSSION

The variance reduction technique requires an appropriate semivariogram model that fitted the observed data. Initially, an experimental semivariogram from the experimental data is derived. Then, a functional semivariogram model is fitted to the experimental variogram. The obtained semivariogram model has essential information to be used in kriging interpolation of observed data. Fitting and selection of suitable semivariogram model can be accomplished through the variogram modelling technique. Once a proper semivariogram model is selected for the observed dataset, kriging is applied for the generation of interpolated surfaces and the estimation of the corresponding kriging error. The semivariogram model fitted to the experimental data is shown as in table below:

Table 1 Exponential semivariogram model

Nugget	Sill	Range
0.93917	1.5568	1.05449

When the semivariogram is successfully fitted to the empirical rainfall data, the particle swarm optimization method is applied to find the minimum objective function (equation 3) in order to get the optimum number and location of rain gauge stations. The optimization technique done based on the steps mentioned earlier. The results of the optimization process were shown in Figure 1 and Table 2.

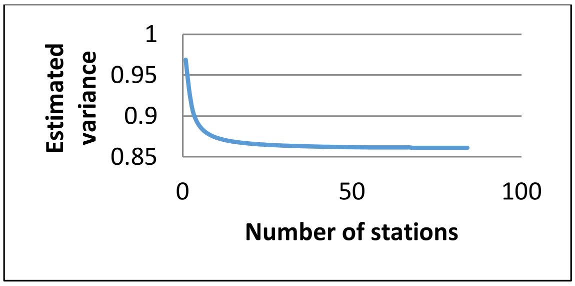


Fig. 2 Estimation variance versus number of stations

Table 2 Estimated variance value with optimal number of stations

Estimated variance	Stations selected	Stations removed
0.8618	67	17

This new optimal rain gauge network demonstrates that the high density of existing rain gauge in Johor has been reduced by removing several stations. Fig. 2 depicts that the estimated variance value decreased as the number of rain gauges increased. This value approach optimal value when the optimization process gives the number of optimal rain gauge stations. In Table 2, it is shown that the optimal number of rain gauge stations is 67 stations with estimated variance value 0.8618. The locations of the 67 selected locations are as shown in Fig. 3 below. From Fig. 3, most of the removed stations are redundant stations that are unnecessary needed.

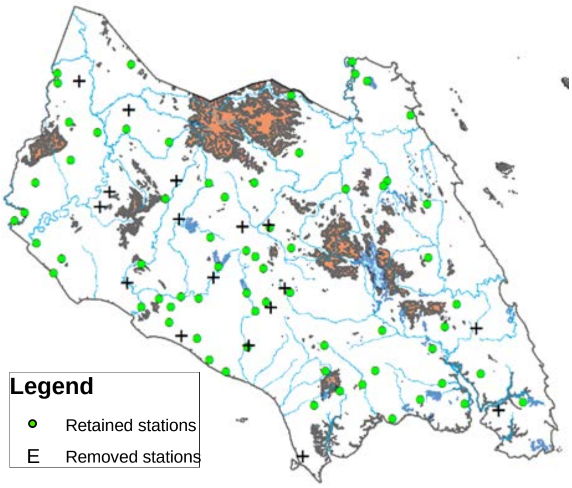


Fig. 3 Optimal locations for 67 selected rain gauge stations.

Table 3 shows the comparison of observed and estimated areal average rainfall for the existing and new optimal rain gauge stations. It shows the new optimal rain gauge stations manage to increase the mean areal rainfall despite the increase in the error due to the lack number of stations.

Table 3 Comparison between observed and estimated areal average rainfall

Estimated Mean Areal Rainfall (mm) for the rain gauge network		**Error (%) obtained for the rain gauge network**	
Existing	Optimal	Existing network	Optimal Network
5.562404	5.5721	2.83	3.35

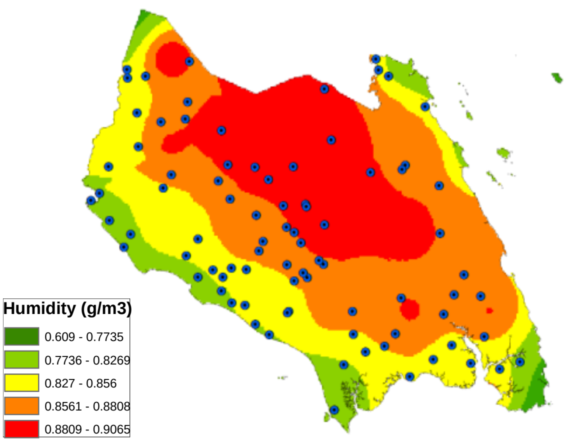
After the execution of the optimization process, two scenarios were considered for the 17 removed stations in Table 2.

- Scenario-1: Optimal positioning of 17 removed stations fitted into Figure 3 to find their optimal locations.
- Scenario-2: All 84 existing stations were reset into new optimal locations

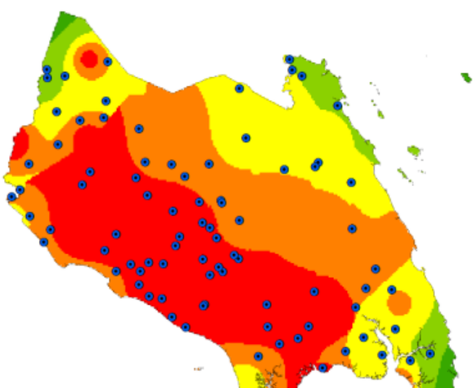
For the redesignation purpose of the network, the whole study area is discretized into 250 square grids with each unit equivalent to 100km² (Fig. 4f). This is in line with the criterion set by the World Meteorological Centre [11] which stated that every rain gauge in the mountain region of temperate, Mediterranean and tropical zone need to be in the area of 100-250km² each. In order to redesign the network, the following criteria were also considered and the design is based on these criteria:

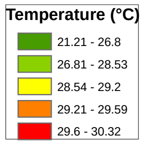
1. Each selected grid cannot have more than one station.
2. The new rain gauge must be put in an area where it is not more than 100 meters above sea level. This is because most of the existing rain gauge stations in Johor are located not more than 100 meters above the sea level as can be seen in Fig. 4(e).
3. The new rain gauge must be put in an area with high humidity reading value as this shows that the area receives more rainfall.
4. The new rain gauge must be put in an area with low temperature as lower temperature shows that the area is humid and receives more rainfall.
5. The new rain gauge must be put in an area where the value solar radiation is low as this shows that the area is humid and has low temperature.
6. The new rain gauge must be put in an area that is protected from the wind in all direction. Area with low wind speed ensures that the collected data is not influenced by the wind speed.

The PSO algorithm is applied once again to determine the optimal rain gauge location for the existing 84 stations from the 250 possible sites. The humidity, temperature, solar radiation, wind speed and elevation data in Fig. 4 will help the algorithm locates suitable locations for the rain gauges.

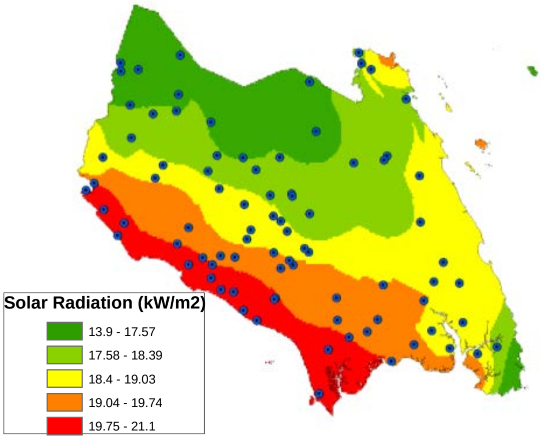


a. humidity

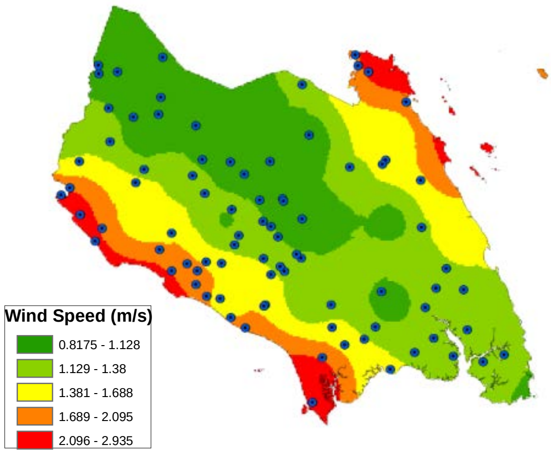




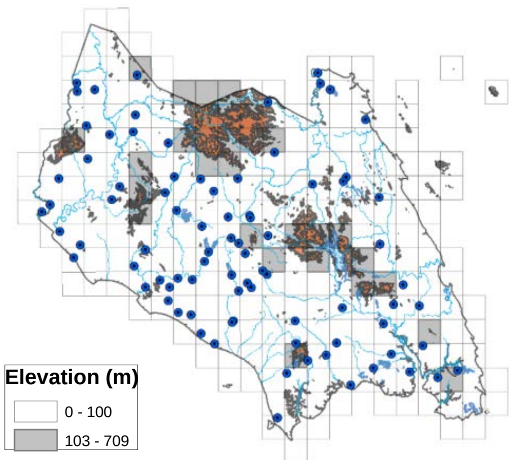
b. temperature



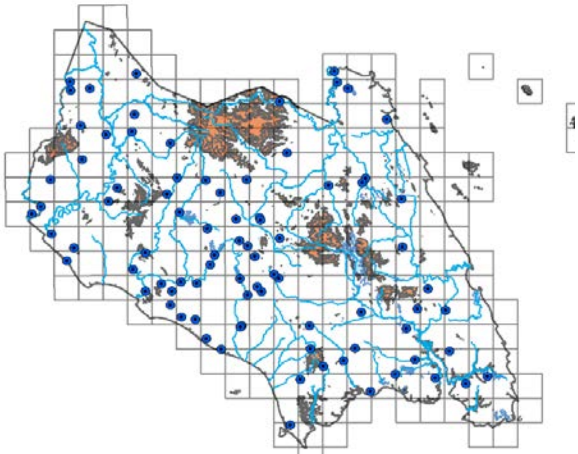
c. solar radiation



d. wind speed



e. elevation



f. Johor with grids

Fig. 4 Humidity, Temperature, Solar Radiation, Wind Speed, Elevation and Grids

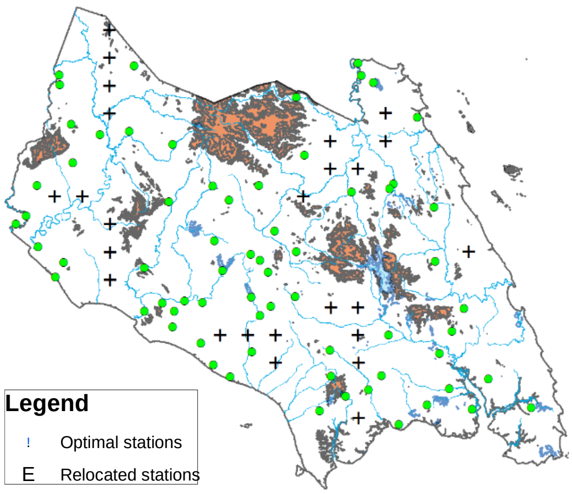
Table 4 shows the resulted estimated variance for both scenarios. For scenario 1, the estimated variance value reduces when the removed stations were located into new optimal locations. This shows that the removed stations were able to improve the accuracy of the network by placing it in an optimum location. For scenario 2, the estimated variance decreases lower than scenario 1 and previous estimated variance. This shows that in order to achieve an optimal network with the existing stations, all stations must be reorganized and relocated into new locations. Elevation, temperature, wind speed, solar radiation and humidity data has help in identifying the optimal locations in the discretized Johor.

The new locations of the restructured rain gauge networks for both scenarios are as shown in Fig.5. In Fig. 5a, the 17 removed stations were relocated into new optimal locations. The placements of the 17 removed stations allow the stations to fill the gap in empty potential grid in Fig. 4f. These 17 stations will act like an additional stations to already optimal 67 stations obtained in the first process. The placements of these additional stations will improve estimation accuracy that leads to an optimal rain gauge network.

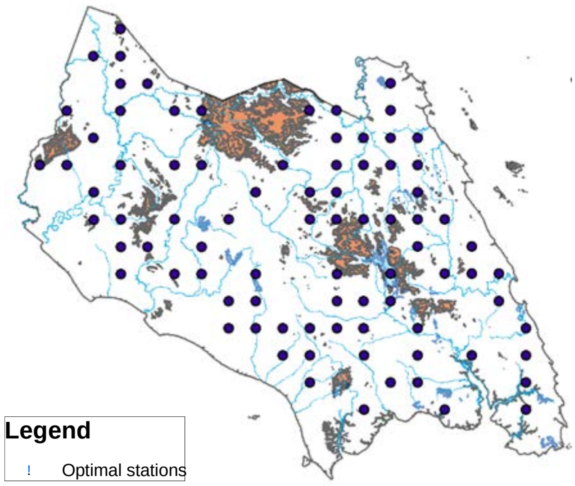
In Fig. 5b, the 84 existing rain gauge stations were restructured into a whole new optimal rain gauge network. This process considered the number of existing stations but need to be located into a new possible optimal location. Figure 5b also depicts locations of all rain gauge stations in Johor that are no longer with redundant stations located into new locations which are complied with the criterion to not have stations located in one same grid. The PSO algorithm managed to determine a new site for each rain gauge stations and it is well placed into a flat and terrain area with elevation not more than 100m above sea level. It is because, when determining the theoretical placement of any rain gauge, the manufacturer often assumes the absence of any physical obstacles that could reduce or obstruct the rainfall. Placing a rain gauge close or on a hilly area will severely alter the rainfall data.

Table 4 Estimated variance

	Scenario 1	Scenario 2
Estimated Variance	0.8473	0.7163



(a) Scenario 1



(b) Scenario 2

Fig. 5 New optimal locations for 84 rain gauge stations

5. CONCLUSIONS

Combination of geostatistics methods and particle swarm optimization as an algorithm of optimization can be used as a framework for rain gauge network design models as it improves the existing rainfall network by minimizing the variance of estimation value. Through this study, it was found that the optimal number of stations for Johor is 67 stations with the estimated variance of 0.8618. Particle swarm optimization as an algorithm of numerical optimisation also improves the optimal network of rain gauge stations in Johor by variance reduction method with the help of rainfall, elevation, humidity, solar radiation, temperature and wind speed data. From the data analysis, it was also found that the 17 stations were either removed or relocated into new locations. This study considered two scenarios. Scenario 1 consideration was to optimally relocate the 17 removed stations to new locations and scenario 2 considerations was to redesign all the existing stations. Scenario 1 was most likely to be applied when the rain gauge network need an additional stations while scenario 2 was more practical in the early planning and designing of an optimal rain gauge network. Redesignation of all 84 stations depicts the minimum optimum estimated variance compare to only 17 relocated stations in scenario 1. This approach is more expensive because it requires a completely new network to be established but the accuracy and the reliability of the rain gauge network is surely improved.

Overall, this study has illustrated that the geostatistics method with particle swarm optimization can be used as the optimization method to provide the solution in designing an optimal rain gauges network system. This optimal network also is essential in providing better rainfall data. This model and methodology can provide information that will help decision makers to understand the relationship between numbers and locations of rain gauge stations in order to provide a better and more accurate rainfall data. Further researches on the optimization technique are required for better results.

6. ACKNOWLEDGEMENTS

The authors would like to thank the Malaysian Ministry of Higher Education, TATI University College and Universiti Teknologi Malaysia for their financial funding and support.

7. REFERENCES

[1] Yang, X. S., Enggineering optimization: An introduction with metaheuristics applications. Wiley Publications. 2010.

[2] Pardo-Igúzquiza, E. Optimal selection of number and location of rainfall gauges for areal rainfall estimation using geostatistics and simulated annealing. Journal of Hydrology 210: 206–220. DOI: 10.1016/S0022-1694(98)00188-7. 1998.

[3] Yeh, H. C., Chen, Y.C., Wei, C., Chen, R. H., Entropy and kriging approach to rainfall network design. Paddy and Water Environment 9: 343–355. DOI: 10.1007/s10333-010-0247-x. 2011.

[4] Kassim, A. H. M., Kottegoda, N.T. Rainfall network design through comparative kriging methods. Hydrological Sciences Journal 36: 223–240. DOI: 10.1080/02626669109492505. 1991.

[5] Mishra, A. K., Coulibaly, P. Developments in hydrometric network design: a review. Reviews of Geophysics 47: RG2001. DOI: 10.1029/2007RG000243. 2009.

[6] Chen, Y. C., Wei, C., Yeh, H. C. Rainfall network design using kriging and entropy. Hydrological Processes 22: 340–346. DOI: 10.1002/hyp.6292. 2008.

[7] Shaghaghian, M. R., Abedini, M. J. Rain gauge network design using coupled geostatistical and multivariate techniques. Scientia Iranica 20: 259–269. DOI: 10.1016/j.scient.2012.11.014. 2013.

[8] Barca, E., Passarella, G., Uricchio, V. Optimal extension of the rain gauge monitoring network of the Apulian regional consortium for crop protection. Environmental Monitoring and Assessment 145: 375–386. DOI: 10.1007/s10661-007-0046-z. 2008.

[9] Chebbi, A., Bargaoui, Z. K., Cunha, M. D. C., Optimal extension of rain gauge monitoring network for rainfall intensity and erosivity index interpolation. Journal of Hydrologic Engineering 16: 665–676. DOI: 10.1061/(ASCE)HE.1943-5584.0000353. 2011.

[10] Burrough, P. A. Principles of Geographical Information Systems for Land Resources Assessment. Clarendon Press, Oxford, pg. 147-165. 1987.

[11] Guide to Meteorological Instruments and Methods of Observation, (WMO-No.8, Seventh edition), WMO, Geneva, 2008.

[12] Eberhart, R. and Kennedy, J. A new optimizer using particle swarm theory. In Proceedings of the Sixth International Symposium on Micro Machine and Human Science (MHS ’95), pages 39–43. IEEE Press, 1995.

[13] Olea, R. A. *Geostatistics for Engineers and Earth Scientists*, Kluwer Academic Publishers, ISBN 0-7923-8523-3, Massachusetts, 2003.

[14] Sajal, K. A., Abdullah, G. Y. and Nitin, M. Optimal design of rain gauge network in the Middle Yarra River catchment, Australia. Hydrol.Process.2014.

International Journal of GEOMATE, Sept, 2016, Vol. 11, Issue 25, pp.2422-2428.

MS No. 5137 received on July 19, 2015 and reviewed under GEOMATE publication policies.

Corresponding Author: M. K. Bazli Mohd Aziz