HYDROLOGICAL PROCESSES
Hydrol. Process. (2014)
Published online in Wiley Online Library
(wileyonlinelibrary.com) DOI: 10.1002/hyp.10389

Optimal design of rain gauge network in the Middle Yarra River catchment, Australia

Sajal Kumar Adhikary,[1]* Abdullah Gokhan Yilmaz[1] and Nitin Muttil[1,2]

[1] *College of Engineering and Science, Victoria University, PO Box 14428, Melbourne, Victoria, 8001, Australia*
[2] *Institute for Sustainability and Innovation, Victoria University, PO Box 14428, Melbourne, Victoria, 8001, Australia*

Abstract:

Rainfall data are a fundamental input for effective planning, designing and operating of water resources projects. A well-designed rain gauge network is capable of providing accurate estimates of necessary areal average and/or point rainfall estimates at any desired ungauged location in a catchment. Increasing network density with additional rain gauge stations has been the main underlying criterion in the past to reduce error and uncertainty in rainfall estimates. However, installing and operation of additional stations in a network involves large cost and manpower. Hence, the objective of this study is to design an optimal rain gauge network in the Middle Yarra River catchment in Victoria, Australia. The optimal positioning of additional stations as well as optimally relocating of existing redundant stations using the kriging-based geostatistical approach was undertaken in this study. Reduction of kriging error was considered as an indicator for optimal spatial positioning of the stations. Daily rainfall records of 1997 (an El Niño year) and 2010 (a La Niña year) were used for the analysis. Ordinary kriging was applied for rainfall data interpolation to estimate the kriging error for the network. The results indicate that significant reduction in the kriging error can be achieved by the optimal spatial positioning of the additional as well as redundant stations. Thus, the obtained optimal rain gauge network is expected to be appropriate for providing high quality rainfall estimates over the catchment. The concept proposed in this study for optimal rain gauge network design through combined use of additional and redundant stations together is equally applicable to any other catchment. © 2014 The Authors. *Hydrological Processes* published by John Wiley & Sons Ltd.

KEY WORDS rain gauge network; geostatistical analysis; ordinary kriging; kriging error; variogram modelling; Middle Yarra River catchment

Received 7 July 2014; Accepted 28 October 2014

INTRODUCTION

Rainfall data provide essential input for effective planning, designing, operating and managing of water resources projects. Rainfall data are employed in various water resources management tasks such as water budget analysis and assessment, flood frequency analysis and forecasting, streamflow estimation, and design of hydraulic structures. A reliable rain gauge network can provide immediate and precise rainfall data that are crucial for effective and economic design of hydraulic structures for flood control. This helps to minimize the hydrological and economic risk involved in different water resources projects. Rain gauge networks are usually installed to facilitate the direct measurement of rainfall data that characterize the spatial and temporal variations of local rainfall patterns in a catchment. 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 (Pardo-Igúzquiza, 1998). 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 (Yeh *et al.*, 2011).

Hydrologists are often required to estimate areal average rainfall over the catchment and/or point rainfall at unsampled locations from observed sample measurements at neighbouring locations. This task can be accomplished accurately with an optimally designed rain gauge network and is, therefore, regarded as an indispensable component of any hydrological study. However, the rain gauge network used in most of the hydrological studies are often sparse and thus incapable of providing adequate rainfall estimates necessary for effective hydrological analysis and design of water resources projects. Use of inaccurate rainfall data may result in significant design errors in the water resources projects, which may eventually result in the immeasurable loss of lives and property damages.

*Correspondence to: Sajal Kumar Adhikary, College of Engineering and Science, Victoria University, PO Box 14428, Melbourne, Victoria 8001, Australia.
E-mail: sajal.adhikary@live.vu.edu.au

Thus, identification and selection of the best network configuration having optimal number and locations of rain gauge stations is the sole 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 yield optimum rainfall information with minimum uncertainty and cost (Kassim and Kottegoda, 1991; Basalirwa *et al.,* 1993; Pardo-Igúzquiza, 1998). One can approach the problem either by eliminating redundant stations from the network to minimize the cost or by expanding the network with installation of additional stations to reduce the estimation uncertainty (Mishra and Coulibaly, 2009).

The design of hydrometric networks is a well-identified problem in hydrometeorology (Mishra and Coulibaly, 2009), which has received considerable attention from the researchers for many years. Because hydrometric network design is associated with myriad concerns, many approaches have been developed for optimal network design. Among others, one type of approach is a kriging-based geostatistical approach that finds wide applications in the rain gauge network design. An important feature of this approach is the provision of kriging error that forms the basis for the rain gauge network design. Optimal network configuration can be achieved by minimizing the kriging error that involves a process of methodical search to find an optimal combination of the appropriate number and locations of stations producing the minimum kriging error. More detailed information about kriging-based geostatistics can be found in the literature (Isaaks and Srivastava, 1989; Webster and Oliver, 2007). In many studies, the kriging technique alone was employed for the rain gauge network design (Shamsi *et al.,* 1988; Kassim and Kottegoda, 1991; Loof *et al.,* 1994; Papamichail and Metaxa, 1996; Tsintikidis *et al.,* 2002; Cheng *et al.,* 2008). However, some studies applied the kriging technique in combination with other techniques such as entropy (Yeh *et al.,* 2011; Chen *et al.,* 2008) and multivariate factor analysis (Shaghaghian and Abedini, 2013) for the network design. In those studies, the sole function of the kriging was to generate rainfall data by interpolation in locations where prospective stations might be installed, whereas entropy was used to measure the information content of each station, and the factor analysis along with clustering technique was used to prioritize stations in terms of information content, respectively. Although in most of the past studies, trial-and-error procedure was used to minimize the kriging error, a few studies combined optimization method based on simulation tools (e.g. simulated annealing) with the kriging technique (Pardo-Igúzquiza, 1998; Barca *et al.,* 2008; Chebbi *et al.,* 2011) to obtain the optimal rain gauge network.

Four different objectives are usually considered with regard to optimal rain gauge network design and assessment by the kriging-based geostatistical approach:

- Expanding the existing rain gauge network with additional stations to achieve appropriate network density for the reduction of estimation uncertainty (Loof *et al.,* 1994; Papamichail and Metaxa, 1996; Tsintikidis *et al.,* 2002; Barca *et al.,* 2008; Chebbi *et al.,* 2011).
- Identifying and establishing the optimal location of additional rain gauge stations in the network to improve the estimation accuracy (Pardo-Igúzquiza, 1998; Chen *et al.,* 2008).
- Prioritizing the rain gauges with respect to their contribution in error reduction in the network (Kassim and Kottegoda, 1991; Cheng *et al.,* 2008; Yeh *et al.,* 2011).
- Choosing an optimal subset of stations from an existing dense rain gauge network to achieve optimum rainfall information (Shaghaghian and Abedini, 2013).

Expansion of the existing network by adding supplementary stations has been the main underlying criterion to achieve the optimal network in most of the past studies. However, the placement and adjustment of stations significantly influence the quality of the obtained hydrological variable in a network (Yeh *et al.,* 2011). Furthermore, an existing network may consist of redundant stations (Mishra and Coulibaly, 2009) that may make little or no contribution to the network performance for providing quality data. Therefore, the optimal positioning of both additional and redundant stations linked to the existing rain gauge network constitutes the main scope of this paper. Hence, the objective of this study is to design an optimal rain gauge network through optimal positioning of additional stations as well as optimally relocating of existing redundant stations using the kriging-based geostatistical approach.

A network design methodology was developed in this study to determine optimal locations of the additional stations and existing redundant stations in the current rain gauge network located in the Middle Yarra River catchment in Victoria, Australia. The procedure involves a methodical search for the optimal number and locations of rain gauge stations in the network that minimize the kriging error of areal and/or point rainfall estimates over the catchment. The methodology presented in this paper is in line with that of Loof *et al.* (1994) who used the kriging-based geostatistical approach to determine the optimal location of additional rain gauge stations in the existing network by using only a selected variogram (e.g. exponential) model. The major contribution here is that unlike the work of Loof *et al.* (1994), the developed methodology considered the likely presence of the redundant stations within the existing network along with the additional stations to obtain the optimal rain gauge network. Furthermore, the use of a selected variogram model in kriging may not give appropriate results for all types of catchments depending on the rainfall and catchment characteristics. Therefore, instead of using a selected variogram model in the kriging, the best fitted variogram model from a set of commonly used variogram models in

DOI: 10.1002/hyp

hydrology was used to compute the kriging error in the network design. The best variogram model was chosen for the kriging applications on the basis of different goodness-of-fit criteria and cross-validation statistics.

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.

STUDY AREA AND DATA DESCRIPTION

The study area

The middle segment of the Yarra River catchment located in Victoria, Australia, is selected as the case study catchment. Approximate location of the catchment is shown in Figure 1. The water resources management is an important and complex issue in the Yarra River catchment because of its wide range of water uses as well as its downstream user requirements and environmental flow provisions (Barua *et al.,* 2012). The catchment is home to more than one-third of Victoria's population (approximately 1.5 million) and native plant and animal species, where the Yarra River acts as the only lifeline. Although the Yarra River catchment is not large with respect to other Australian catchments, it produces the fourth highest water yield per hectare of the catchment in Victoria, making it a very productive catchment (Melbourne Water, 2013). The Yarra River has thus played a key role in the way Melbourne has developed and grown.

The catchment lies north and east of Melbourne covering an area of 4044 km^2. The Yarra River travels about 245 km from its source, on the southern slopes of the Great Dividing Range in the forested Yarra Ranges National Park, and runs through the catchment into the end of its estuary, at Port Phillip Bay. There are seven storage reservoirs located within the catchment (Figure 1) that support water supply to Melbourne. Because of the diversity of water use activities and significant changes in the rainfall patterns, pressure upon the water resources management has become more intense in the catchment (Barua *et al.,* 2012).

The Yarra River catchment is divided into three distinctive subcatchments, namely, Upper Yarra, Middle Yarra and Lower Yarra segments (Barua *et al.,* 2012) based on the different land use patterns (Figure 1). Forest, agricultural and urban areas are the major land use patterns of the catchment (Sokolov and Black, 1996). The Upper Yarra segment of the catchment, beginning from the Yarra Ranges National Park to the Warburton Gorge at Millgrove, consists of mainly forested and mountainous areas with minimum human settlement. This segment is used as a closed water supply catchment for Melbourne, and about 70% of Melbourne's drinking water supply comes from this pristine upper segment. Thus, it has been reserved for more than 100 years for water supply purposes (Barua *et al.,* 2012; Melbourne Water, 2013). The Middle Yarra segment, from the Warburton Gorge to

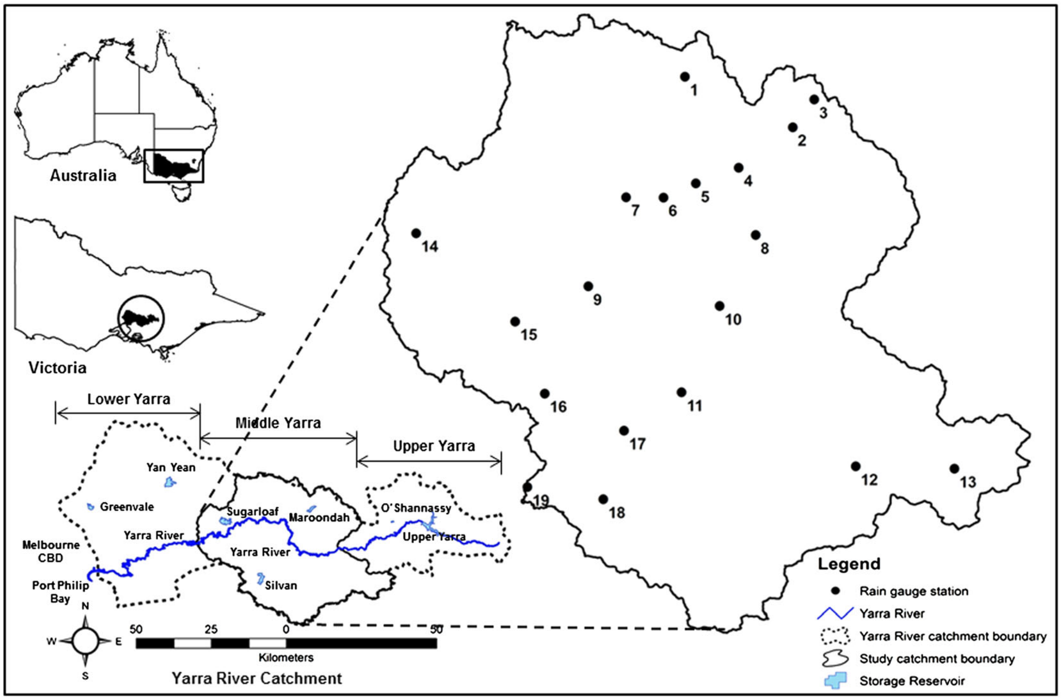

Figure 1. Yarra River catchment showing the study area with rain gauge stations

Warrandyte Gorge, is notable as the only part of the catchment with an extensive flood plain, which is mainly used for agricultural activities. The Lower Yarra segment of the catchment, located downstream of Warrandyte, is mainly characterized by the urbanized floodplain areas of Melbourne city. Most of the land along rivers and creeks in the Middle and Lower segments of the catchment has been cleared for the agricultural or urban development (Melbourne Water, 2013).

The management of water resources in the Yarra River catchment is of great importance considering the greater variation of the rainfall patterns through its different segments. The mean annual rainfall varies across the catchment from about 1100 mm in the Upper Yarra segment to 600 mm in the Lower Yarra segment (Daly *et al.,* 2013). The Middle Yarra segment (case study area) covers an area of 1511 km^2 (Figure 1). The area consists of three reservoirs, namely, Maroondah, Silvan and Sugarloaf reservoirs, that supports water supply for a range of activities including urban and agricultural activities. The main intention of improving reservoir operation in Australia is to store as much water as possible for satisfying the water demand during shortage of streamflows while keeping provision for flood control during excess streamflows (Melbourne Water, 2013). Decreasing rainfall patterns will reduce the streamflows, which in turn will lead to the reduction in reservoir inflows and hence impact the overall water availability. Moreover, the reduced streamflows may cause increased risk of bushfires in the catchment. However, increasing rainfall patterns and the occurrence of extreme rainfall events will result in excess amount of streamflows that may cause flash floods in the urbanized lower segment of the catchment and make it vulnerable and risk prone. The urbanized Lower segment of the catchment is also dependent on the water supply from the storage reservoirs mainly located in the middle and upper reaches of the catchment. Accurate rainfall information in the Middle and Upper segments of the Yarra River catchment is essential to determine the future streamflows accurately for optimal reservoir operation and effective flood control in the lower segment. Therefore, the design of an optimal rain gauge network has great importance for the Middle Yarra River catchment.

Dataset used

There are 19 (number 1 to 19 in Figure 1) rain gauge stations in the study area (Middle Yarra River catchment), which are currently operated and maintained by the Bureau of Meteorology (BoM), Australia. Among those, two stations, namely, Monbulk (Spring Road) (station 18) and Ferny Creek (station 19), were installed by BoM in 2011 (Figure 1 and Table I). The rain gauge network in this study thus consisted of 17 stations before 2011 (will be called as the network before 2011 in this paper) and consists of 19 stations after 2011 (will be called as the network after 2011 in this paper). The objective of this paper is to determine the optimal location of the two new additional stations (stations 18 and 19) in the network after 2011 as well as the existing redundant stations in the network before 2011.

Daily rainfall data for all 19 stations in the rain gauge network of the Middle Yarra River catchment were collected from the SILO climate database for the period of 1980–2012. The SILO database (http://www.longpaddock.qld.gov.au/silo/) has been selected for this study because SILO data are free from missing records. This database allows filling up the missing records based on its own interpolation algorithm using available records in the surrounding stations (Jeffrey *et al.,* 2001).

Several studies reported that the rainfall variability in the eastern parts of Australia (including the study area) is strongly influenced by the El Niño Southern Oscillation (ENSO) phenomenon (Allan, 1988; Nicholls and Kariko, 1993; Murphy and Ribbe, 2004; Dutta *et al.,* 2006; Chowdhury and Beecham, 2010; Mekanik *et al.,* 2013). The ENSO is quantitatively defined by the Southern Oscillation Index (SOI) that divides ENSO into El Niño and La Niña phenomena. The El Niño phenomenon is a persistent negative value of the SOI, which usually corresponds to a decrease in rainfall. The La Niña phenomenon is the reverse process of El Niño and is responsible for causing more rainfall than normal (Chowdhury and Beecham, 2010). The spatial variability of rainfall caused by ENSO effect was considered for the design of the rain gauge network in this study by considering daily rainfall records of one El Niño and one La Niña year.

The SOI data were obtained from BoM for the period of 1980–2012. It was found that maximum negative SOI occurred in the year 1997 for the El Niño period, whereas maximum positive SOI occurred in the year 2010 for the La Niña period. Therefore, daily rainfall data of 1997 (El Niño) and 2010 (La Niña) were selected for the rain gauge network design in this study. Use of rainfall records from the selected years for El Niño and La Niña periods in the network design will be a better representation of the high rainfall variability experienced in the Middle Yarra River catchment. The statistics of daily rainfall data at different rain gauge stations for the selected El Niño and La Niña years are given in Table I.

METHODOLOGY

A rain gauge network design methodology was developed in this study using the kriging-based geostatistical approach. The framework of the developed methodology is composed of the following three steps:

1. Data preparation and transformation – Performing exploratory data analysis and normality test for the observed data.

DOI: 10.1002/hyp

Table I. Summary of rain gauge stations and rainfall data used in the case study

Station no.[a]	BoM ID	BoM rain gauge station name	Elevation (m)	Daily rainfall (mm)			
				El Niño year		La Niña year	
				Mean	SD[b]	Mean	SD[b]
1	86142	Toolangi (Mount St Leonard DPI)	595	2.418	5.395	4.343	12.601
2	86366	Fernshaw	210	2.057	4.887	3.759	8.499
3	86009	Black Spur	567	2.504	6.228	4.461	11.069
4	86070	Maroondah Weir	174	1.787	4.125	3.508	7.385
5	86385	Healesville (Mount Yule)	100	1.477	3.488	3.759	8.499
6	86363	Tarrawarra	124	1.414	3.827	3.119	7.343
7	86364	Tarrawarra Monastery	100	1.392	3.540	2.656	6.516
8	86219	Coranderrk Badger Weir	360	2.255	5.167	3.708	8.048
9	86383	Coldstream	83	1.425	3.444	2.813	6.685
10	86229	Healesville (Valley View Farm)	156	1.725	3.807	3.134	6.716
11	86367	Seville	181	1.685	3.818	3.029	6.392
12	86358	Gladysdale (Little Feet Farm)	295	2.140	4.681	4.056	8.949
13	86094	Powelltown DNRE	189	2.481	5.373	4.048	9.685
14	86059	Kangaroo Ground	183	1.449	3.466	2.684	6.704
15	86066	Lilydale	130	1.492	3.662	2.922	6.915
16	86076	Montrose	170	2.529	5.726	3.202	6.917
17	86106	Silvan	259	2.193	5.063	3.299	7.261
18	86072	Monbulk (Spring Road)	—	2.270	4.929	4.140	8.934
19	86266	Ferny Creek	513	2.285	4.895	4.640	9.581

BoM, Bureau of Meteorology.
[a] Station numbers are same as in Figure 1.
[b] SD – standard deviation of daily rainfall records.

2. Variogram modelling and kriging interpolation – This includes fitting and selection of appropriate variogram models, performing kriging interpolations and estimation of the kriging error.
3. Rain gauge network design – This step involves exploring the existing rain gauge network for finding appropriate locations for additional stations and removing and/or relocating redundant stations to reduce kriging error in the network to achieve the best network output.

Each of the three steps is discussed in detail in the following subsections.

Data preparation and transformation

The normal distribution of data is a basic requirement of the kriging-based geostatistical approach (Barca *et al.,* 2008). Kriging assumes that the data come from a stationary stochastic process. Kriging leads to an optimum estimator and yields best results when the data are normally distributed. Thus, the inconsistency present in observed datasets should be identified and fixed in the beginning before going for the model development and analysis. To accomplish this, the exploratory data analysis (i.e. detection and removal of trends and outliers, performing the normality test for the observed data and applying the data transformation for non-normal datasets) was undertaken. Transformation of the inconsistent data is very useful to make it symmetrical, linear and constant in variance. The log transformation is often used for hydrological data that have skewed or non-normal distributions. After employing the log transformation on the observed data, the skewness of the transformed data becomes close to zero, and the data follow the normal distribution (Johnston *et al.,* 2001). In this study, the log transformation was applied for data transformation when datasets could not satisfy the normal distribution hypothesis.

Once the data transformation is completed, the transformed data must be tested to check whether the evidence is sufficient to accept the normal distribution hypothesis. A statistical test known as Kolmogorov–Smirnov (K–S) test was applied for this purpose, because it is a simple and straightforward test to check the normality. Details of the K–S test can be found in McCuen (2003). Further investigation can be performed for the normal distribution through the visual examination of quantile–quantile (Q–Q) plot and skewness coefficients obtained from the observed data (Johnston *et al.,* 2001). Data were accepted as normally distributed if most of the transformed data in the Q–Q plot were laid on or very close to a straight line, and skewness coefficients of the transformed data were reduced or close to zero. After transforming and testing of all observed datasets for the normal distribution, resulting datasets were used for the variogram modelling and kriging interpolation.

DOI: 10.1002/hyp

Variogram modelling and kriging interpolation

The kriging technique requires an appropriate variogram model that defines the spatial structure of the observed data in the computation process. Initially, an experimental variogram from the observed data is derived. A functional variogram model is then fitted to the experimental variogram. The obtained variogram model contains necessary information to be used in kriging interpolation of observed data. Fitting and selection of appropriate variogram model can be accomplished through the variogram modelling technique. Once a proper variogram model is chosen for the observed dataset, kriging is employed for the generation of interpolated surfaces and the estimation of the corresponding kriging error.

Variogram modelling. The degree of spatial dependence is generally expressed by a variogram model in kriging. A variogram is a mathematical function of the distance and direction separating two locations used to quantify the spatial autocorrelation in regionalized variables (RVs). An RV is a variable that can take values according to its spatial location. Variogram modelling is a process of developing relationship among sampling locations to quantify the variability associated with RV. Variogram function is a key tool for the kriging method and frequently employed to exercises that involve estimating desired values at new unsampled locations based on observed values at neighbouring locations.

The kriging method requires a theoretical variogram function that is to be fitted with an experimental variogram of the observed data. The experimental variogram, $\gamma(h)$, is calculated from the observed data as a function of the distance of separation, h, and is given by

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

where $N(h)$ is the number of sample data points separated by a distance h; x_i and $(x_i + h)$ represent sampling locations separated by a distance h; $Z(x_i)$ and $Z(x_i + h)$ indicate values of the observed variable Z, measured at the corresponding locations x_i and $(x_i + h)$, respectively. The theoretical variogram function, $\gamma^*(h)$, allows the analytical estimation of variogram values for any distance and provides the unique solution for weights required for kriging interpolation. Several variogram models are possible depending on the shape of the variogram function that include exponential, gaussian, spherical, circular, linear, K-bessel, J-bessel, rational quadratic, stable and hole effect models (Johnston *et al.,* 2001; Webster and Oliver, 2007). However, exponential, gaussian and spherical variogram models are mostly used in hydrology and are expressed by Equations (2)-(4).

$$\gamma^*(h) = C_0 + C_1\left[1 - \exp\left(-\frac{3h}{a}\right)\right]_{\text{Exponential}} \quad (2)$$

$$\gamma^*(h) = C_0 + C_1\left[1 - \exp\left(-\frac{3h^2}{a^2}\right)\right]_{\text{Gaussian}} \quad (3)$$

$$\gamma^*(h) = C_0 + C_1\left[1.5\left(\frac{h}{a}\right) - 0.5\left(\frac{h^3}{a^3}\right)\right]_{\text{Spherical}} \quad (4)$$

where C_0, a and $(C_0 + C_1)$ represent nugget, range and sill, respectively, commonly called as variogram parameters. These parameters describe a variogram model and hence affect the kriging computation. Nugget represents measurement error and/or microscale variation at spatial scales that are too fine to detect and is seen as a discontinuity at the origin of the variogram model. Range is a distance beyond which there is little or no autocorrelation among variables. Sill is the constant semivariance of the RV beyond the range.

For variogram modelling, three isotropic theoretical variogram functions (i.e. exponential, gaussian and spherical models) were fitted to the experimental variogram ignoring directional influences and assuming isotropic condition. Isotropy is a property in which direction is unimportant and the spatial dependence or autocorrelation changes only with the distance between two locations. The corresponding variogram parameters of the theoretical models were inferred on the basis of the experimental variogram. Manual (visual) and automatic fitting methods (ESRI, 2009) were applied to obtain the best fitted parameters for variogram models. The variogram parameters (nugget, sill and range coefficients) were iteratively changed to obtain the best fitted model. The best model was selected on the basis of the coefficient of determination (R), residual sum of square (RSS), root mean square error (RMSE) and mean absolute error (MAE) values. The variogram model that gave the highest R with the lowest RSS, RMSE and MAE values was chosen for kriging interpolation.

Kriging interpolation. Kriging is an optimal surface interpolation technique based on spatially dependent variance. Kriging refers to a family of generalized least-square regression methods in geostatistics. It is the best linear unbiased estimator of unknown variable values at unsampled locations in space, where no measurements are available based on the known sampling values from surrounding area (Isaaks and Srivastava, 1989; Webster and Oliver, 2007). Ordinary kriging (OK) technique from the family of the classical geostatistical methods was used in this study for interpolation of the rainfall data and estimation of the kriging error. The kriging estimator is expressed as

$$Z^*(x_0) = \sum_{i=1}^{n} w_i Z(x_i) \quad (5)$$

DOI: 10.1002/hyp

where $Z^*(x_0)$ refers to the estimated value of Z at desired location x_0; w_i represents weights associated with the observation at the location x_i with respect to x_0; and n indicates the number of observations within the domain of search neighbourhood of x_0 for performing the estimation of $Z^*(x_0)$.

The kriging variance, $\sigma_z^2(x_0)$, in the OK can be computed by Equation (6) as

$$\sigma_z^2(x_0) = \mu_z + \sum_{i=1}^{n} w_i \gamma(h_{0i}) \quad \text{for} \sum_{i=1}^{n} w_i = 1 \tag{6}$$

where $\gamma(h)$ is the variogram value for the distance h; h_{0i} is the distance between observed data points x_i and x_j; μ_z is the Lagrangian multiplier in the Z scale; h_{0j} is the distance between the unsampled location x_0 (where estimation is desired) and sample locations x_i; and n is the number of sample locations.

When a log transformation is applied to data, OK is converted to log-normal kriging (LNK). The log-transformed predicted values obtained in the LNK are then back-transformed to its original states. However, the back-transformed values are biased predictor (Johnston *et al.,* 2001). Therefore, the kriging variance, $\sigma_z^2(x_0)$, in the LNK is obtained by Equation (7), which is derived from the unbiased expression of the LNK estimator given by Cressie (1993):

$$\sigma_z^2(x_0) = \{Z^*(x_0)\}^2 \exp\left[\left\{\sigma_y^2(x_0)\right\} - 1\right] \tag{7}$$

where $Z^*(x_0)$ refers to the estimated value of Z at desired location x_0 and $\sigma_y^2(x_0)$ is the LNK error in Y scale. The square root of the kriging variance is termed as the kriging standard error (KSE) that forms the basis for the rain gauge network design and evaluation.

Performance assessment of variogram models. The final form of the theoretical variogram model for kriging applications was selected on the basis of the results of a validation scheme known as the cross-validation procedure. Cross-validation is a simple leave-one out validation procedure (Haddad *et al.,* 2013) that involves eliminating the data values individually one by one from the observed data and then predicting each data value by using the remaining data values. This validation scheme helps to evaluate the prediction performance of kriging by comparing observed and estimated values. Cross-validation statistics serves as diagnostics to demonstrate whether the performance of the adopted model is acceptable. The statistics are used to check whether the prediction is unbiased, as close as possible to the measured value, and the variability of the prediction is correctly assessed. Model performances were evaluated on the basis of the following cross-validation parameters (Johnston *et al.,* 2001):

- The mean standardized prediction error (MSS) was used to check if the model is unbiased and should be close to zero for unbiased estimates (the closer the MSS values to zero, the better the performance of the model).
- The root mean square prediction error (RMSE) was used to check whether the prediction is close to the measured values (the smaller the RMSE value, the closer the prediction is to the measured value).
- The variability of the predicted data was assessed in two ways; first, by comparison of the RMSE and average KSE values. If RMSE and KSE values are closer, this indicates that the variability in the prediction is correctly assessed. Second, the variability was assessed by the root mean square standardized (RMSS) prediction error. If the root mean square standardized value is close to one, then the estimation variances are consistent and the variability of the prediction is correctly assessed. If it is greater than one, then it is underestimated, and otherwise, it is overestimated.

Rain gauge network design

An optimal rain gauge network should neither suffer from lack of rain gauge stations nor be oversaturated with redundant rain gauge stations. 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 variance reduction approach under the kriging-based geostatistical approach was used for the rain gauge network design in this study. Reduction of the kriging error was considered as an indicator to achieve the optimal network. The underlying principle is that optimal positioning of additional as well as redundant stations in high variance zones will reduce the kriging error in the network and hence improve the network performance. Applying this principle repeatedly, a certain stage will come when the optimal combination of existing and additional stations can be obtained that yield high network performance to form the optimal rain gauge network. It is important to note that topographic effects (elevation) can be used as a secondary variable in kriging process in the form of co-kriging (Goovaerts, 2000; Mair and Fares, 2011; Feki *et al.,* 2012). However, this variable was not considered in the current study for the kriging interpolation and network design because it is beyond the scope of this study. The following sequential steps were used for the rain gauge network design:

1. Perform variogram modelling and kriging analysis for the network before 2011 (BoM's base network) consisting of 17 stations and computation of the kriging error (KSE_{Old}).

DOI: 10.1002/hyp

Table II. Summary of skewness values and normality test for mean daily rainfall data

Year/period	Rain gauge network used(no. of rain gauge stations)	Skewness		K–S*
		Without transformation	With log transformation	
El Niño year	Network before 2011 ($n = 17$)	0.1740	0.0408	0.1855
	Network after 2011 ($n = 19$)	−0.0442	—	0.1728
La Niña year	Network before 2011 ($n = 17$)	0.3075	0.1102	0.1145
	Network after 2011 ($n = 19$)	0.2045	0.0067	0.1169

*K–S: Kolmogorov–Smirnov statistic value.
$KS_{17,\ 0.05} = 0.3180$.
$KS_{19,\ 0.05} = 0.3010$.

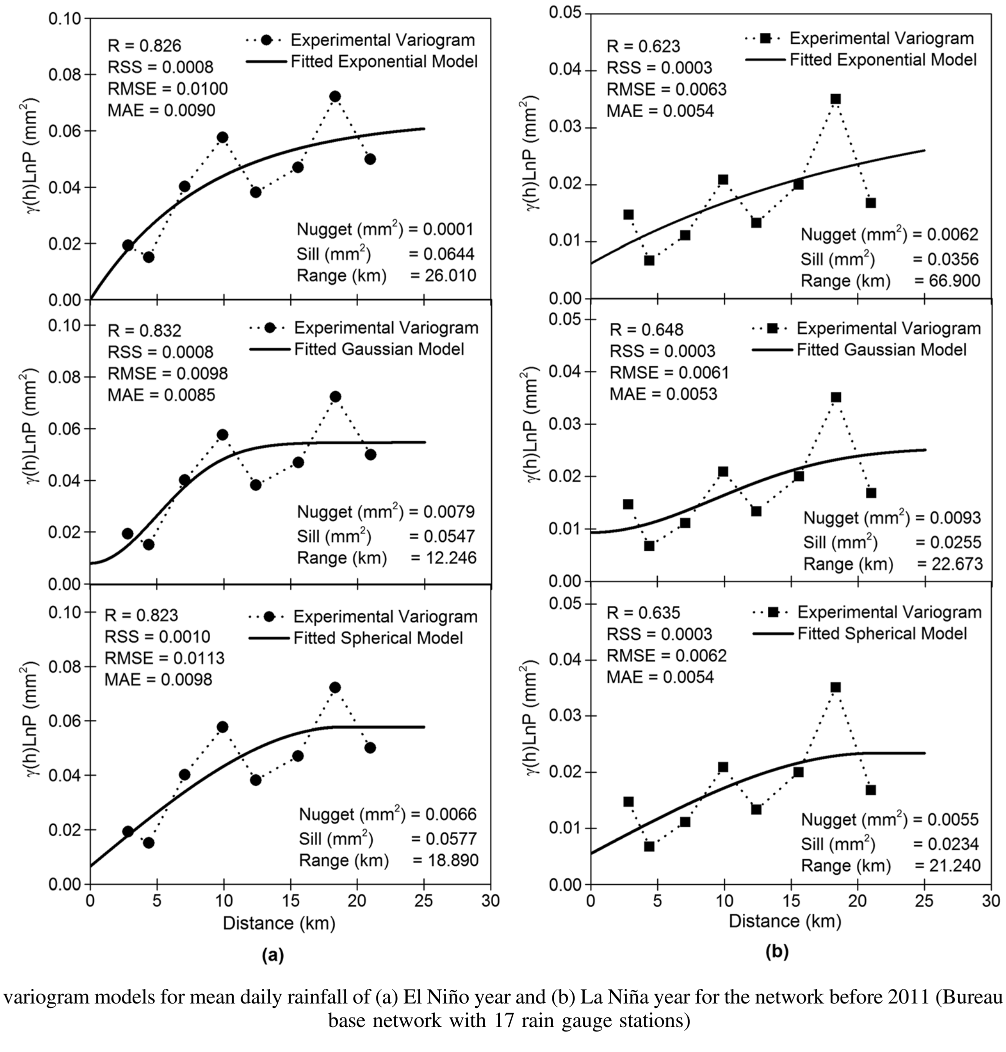

Figure 2. Fitted variogram models for mean daily rainfall of (a) El Niño year and (b) La Niña year for the network before 2011 (Bureau of Meteorology's base network with 17 rain gauge stations)

2. Consider the network after 2011 (BoM's augmented network) consisting of 19 stations and again perform variogram modelling and kriging analysis to estimate the corresponding kriging error (KSE_{New}).
3. Compute the relative error (RE) reduction to check the network performance for error reduction as

$$RE(\%) = \frac{KSE_{Old} - KSE_{New}}{KSE_{Old}} \times 100 \qquad (8)$$

4. Generate KSE map for the network after 2011 (BoM's augmented network) and identify high variance zones in the KSE map. Locate additional stations directly in

DOI: 10.1002/hyp

high variance zones, whereas remove redundant stations are from low variance zones and locate them in high variance zones in the KSE map.

5. Locate either a single rain gauge or a set of rain gauges in high variance zones of the KSE map from the group of additional and redundant stations. Calculate the corresponding RE each time (repeat steps two to four). It is worth noting that redundant stations were identified from the network before 2011, and additional stations were considered from the network after 2011.
6. Plot RE values against various combinations of existing and additional rain gauge stations.
7. Select the combination that gives the maximum RE value indicating the optimal location of rain gauge stations and thus yield the optimal rain gauge network.

RESULTS AND DISCUSSION

Data preparation and transformation

Kriging-based geostatistical interpolation methods lead to optimum estimators when data values are normally distributed. Thus, the goodness of fit of the normal distribution was first investigated for rainfall datasets. Exploratory data analysis and visual inspection of Q–Q plots for rainfall datasets were performed to explore the normal distribution hypothesis. Skewness values of the histograms obtained in the exploratory data analysis were used initially to check whether the rainfall data could approach the normal distribution. If the skewness values are close to zero, this means that data are free from skewness and thus fit the normal distribution. The skewness values of observed mean daily rainfall data for El Niño and La Niña years are shown in Table II.

The results indicate that data are positively skewed and corresponding skewness values are not close to zero in all cases except the network after 2011 during El Niño year. This means that they do not follow a normal distribution and appropriate transformation is necessary to make them normally distributed. Log transformation was applied to those positively skewed datasets, and histograms were formed again. Skewness values for obtained histograms of log-transformed datasets are also listed in Table II. It is

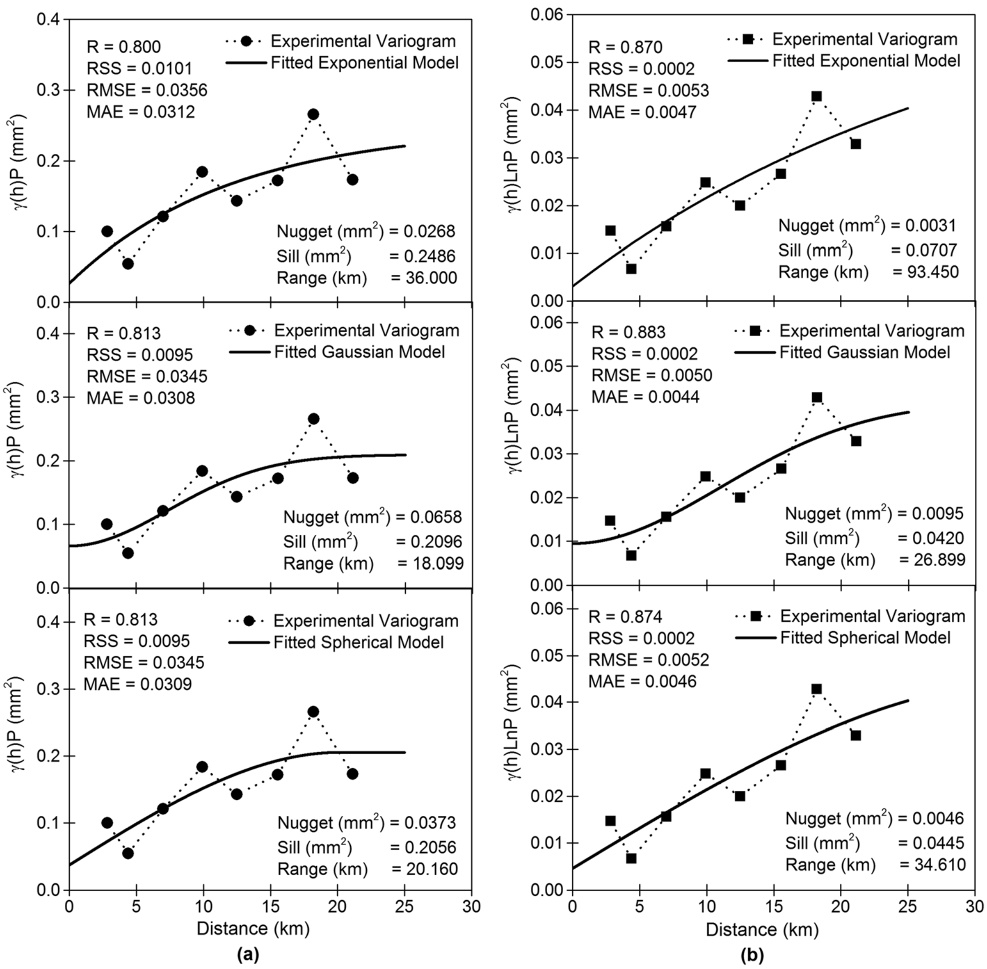

Figure 3. Fitted variogram models for mean daily rainfall of (a) El Niño year and (b) La Niña year for the network after 2011 (Bureau of Meteorology's augmented network with 19 rain gauge stations)

DOI: 10.1002/hyp

seen that the log transformation has greatly reduced the skewness values close to zero, and hence, transformed datasets can be treated as normally distributed.

Fitting of normal distribution for all rainfall datasets was confirmed by the K–S test with a 5% significance level. Datasets were accepted as normally distributed (i.e. null hypothesis was accepted) if the K–S test statistic value was less than the corresponding critical value ($KS_{17} = 0.3180$ and $KS_{19} = 0.3010$) for the 5% level of significance. In the current study, the null hypothesis is defined as the condition that indicates datasets are normally distributed. It is evident from Table II that the null hypothesis of both non-transformed and transformed datasets normality cannot be rejected at the 0.05 level of significance, and thus, K–S test has accepted the normal distribution in the 95% confidence level. By considering the exploratory data analysis and K–S test results, it can be concluded that all positively skewed rainfall datasets approach a normal distribution after applying the log transformation.

Variogram models for rainfall datasets

In the variogram modelling process, an experimental variogram was first computed by using Equation (1) based on normally distributed rainfall datasets. The binning process that defines average values of variance in several distance lags was followed in computing the experimental variogram. A lag represents a line vector that separates any two sample locations and thus has length (distance) and direction (orientation). In the experimental variogram computation, a lag size of 2.835 km equals to the minimum interstation distance, and a total of 8 lag intervals that covers half of the maximum interstation distance were used (Johnston *et al.,* 2001; Webster and Oliver, 2007). Three variogram functions, namely, exponential, gaussian and spherical models described by Equations (2)-(4), were fitted to the experimental variogram. For simplicity of modelling, effect of anisotropy on variogram parameters was ignored, and isotropy was assumed. Because of the less number of stations in the network, formation of directional variograms gave only few data pairs that were too chaotic to form directional variogram. Therefore, directional influence was ignored, and isotropic variogram model was fitted to the experimental variogram in all cases.

Figures 2 and 3 show the computed experimental and fitted variogram models with corresponding variogram parameters and goodness-of-fit measures for rainfall

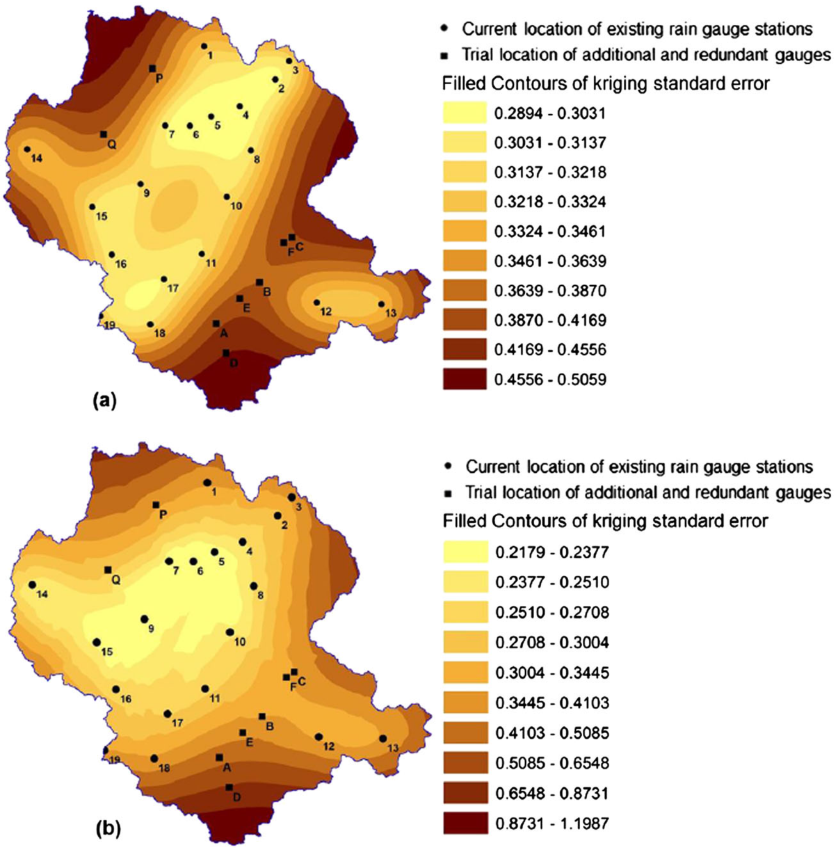

Figure 4. Kriging standard error (KSE) map for (a) El Niño year and (b) La Niña year with trial locations of additional stations (stations 18 and 19) and redundant stations (stations 5 and 6) in the high variance zones of the KSE map

DOI: 10.1002/hyp

datasets of El Niño and La Niña years for the network before 2011 and network after 2011, respectively. As seen in Figures 2 and 3, the gaussian model gives the highest *R* value and lowest RSS, RMSE and MAE values for both El Niño and La Niña years and was, therefore, selected as the best variogram model. The computed cross-validation statistics for all models indicate that the gaussian variogram model satisfies the unbiased and consistent estimates of variances for all four rainfall datasets in El Niño and La Niña years for the network before 2011 and network after 2011, respectively, and thus applicable for the kriging analysis. Therefore, it can be concluded that the gaussian variogram model and the corresponding parameters are adequate to describe the spatial structure of the observed rainfall data.

Kriging of rainfall datasets

In this study, OK was implemented through ArcGISv9.3.1 software (Redlands, CA, USA) (ESRI, 2009) and its geostatistical analyst extension (Johnston *et al.,* 2001). OK was performed for non-transformed rainfall datasets using the modelled variograms to estimate the kriging error. However, in case of log-transformed rainfall datasets, LNK was performed, and back-transformed values were used to estimate the kriging error. For log-transformed datasets, predicted values computed by kriging were automatically back-transformed to the original values before a map was produced by the ArcGIS software (Johnston *et al.,* 2001). Figure 4 shows the kriging error (KSE) map produced by kriging interpolation for the network after 2011 during El Niño and La Niña years, respectively.

Figure 4 demonstrates that less gauge density exists in the eastern and south-eastern part of the Middle Yarra River catchment where high variance zones are observed. It is seen that locations near existing stations have lower KSE values, whereas higher KSE values can be found in areas having less or no rain gauge stations. For example, areas having stations 4, 5, 6 and 7 exhibit lower KSE values because of the high network density. It reveals the likely presence of redundant stations in that region. However, in the north-western, eastern and south-eastern part of the catchment, it is observed that the network density is comparatively less and therefore requires placement of the additional stations to reduce the kriging error. Thus, the rain gauge density in the

Table III. Proposed positions of Bureau of Meteorology stations in high variance areas of kriging error map

Trial no.	Rain gauge station used[a]	New location in high variance zone in KSE map[b]	Location coordinates (m) Easting	Northing
Case-1: Optimal positioning of new additional stations (stations 18 and 19) in the high variance areas				
1	Station 18	18A	369 163	5 806 421
2	Station 18	18B	374 495	5 811 531
3	Station 18	18C	378 494	5 817 084
4	Station 19	19D	370 385	5 802 755
5	Station 19	19E	372 051	5 809 531
6	Station 19	19 F	377 494	5 816 418
7	Station 18	18A	369 163	5 806 421
	Station 19	19D	370 385	5 802 755
8	Station 18	18A	369 163	5 806 421
	Station 19	19E	372 051	5 809 531
9	Station 18	18A	369 163	5 806 421
	Station 19	19 F	377 494	5 816 418
10	Station 18	18B	374 495	5 806 421
	Station 19	19D	370 385	5 802 755
11	Station 18	18B	374 495	5 806 421
	Station 19	19F	377 494	5 816 418
12	Station 18	18C	378 494	5 806 421
	Station 19	19E	372 051	5 809 531
Case-2: Relocating and optimal positioning of redundant stations (stations 5 and 6) in the high variance areas				
13	Station 5	5P	361 276	5 837 856
14	Station 6	6Q	355 278	5 829 747
15	Station 5	5P	361 276	5 806 421
	Station 6	6Q	355 278	5 802 755

KSE, kriging standard error.
[a] Station numbers are the same as Figure 1.
[b] Station locations are shown in Figure 4.

DOI: 10.1002/hyp

Table IV. Skewness and kurtosis values of mean daily rainfall datasets of different combinations of rain gauge stations in the network after 2011

Trial no.	Rain gauge combination in network after 2011 ($n = 19$)[a]	Rainfall (mm) in El Niño year					Rainfall (mm) in La Niña year				
		Without transformation		With log transformation			Without transformation		With log transformation		
		Skew	Kurt	Skew	Kurt	K–S*	Skew	Kurt	Skew	Kurt	K–S*
Case-1: Optimal positioning of new additional stations (stations 18 and 19) in the high variance areas											
1	17 + (station 18 and 19)	−0.031	1.433	—	—	0.1723	0.211	1.853	0.009	1.793	0.1198
2	17 + (station 18 and 19)	0.054	1.488	—	—	0.1676	0.318	2.029	0.090	1.925	0.0980
3	17 + (station 18 and 19)	0.041	1.485	—	—	0.1685	0.355	2.046	0.127	1.938	0.0939
4	17 + (station 18 and 19)	−0.040	1.383	—	—	0.1723	0.287	1.980	0.062	1.851	0.1104
5	17 + (station 18 and 19)	0.039	1.489	—	—	0.1687	0.125	1.838	−0.068	1.818	0.1329
6	17 + (station 18 and 19)	0.058	1.494	—	—	0.1674	0.199	1.877	0.001	1.849	0.1090
7	17 + (station 18 and 19)	−0.025	1.405	—	—	0.1717	0.296	2.016	0.066	1.875	0.1133
8	17 + (station 18 and 19)	0.052	1.518	—	—	0.1683	0.122	1.861	−0.074	1.833	0.1363
9	17 + (station 18 and 19)	0.071	1.523	—	—	0.1670	0.197	1.902	−0.004	1.865	0.1123
10	17 + (station 18 and 19)	0.059	1.455	—	—	0.1669	0.408	2.212	0.152	2.018	0.0985
11	17 + (station 18 and 19)	0.157	1.605	−0.019	1.547	0.1669	0.289	2.124	0.056	2.026	0.0771
12	17 + (station 18 and 19)	0.125	1.590	−0.049	1.534	0.1687	0.252	2.081	0.025	1.993	0.1087
Case-2: Relocating and optimal positioning of redundant stations (stations 5 and 6) in the high variance areas											
13	18 + (station 5)	−0.073	1.488	—	—	0.1421	0.301	1.871	0.100	1.811	0.1267
14	18+ (station 6)	−0.076	1.442	—	—	0.1657	0.179	1.755	−0.007	1.684	0.1185
15	17 + (station 5 and 6)	−0.112	1.531	—	—	0.1437	0.273	1.805	0.081	1.724	0.1277

[a] Station numbers are the same as in Figure 1.
*K–S: Kolmogorov–Smirnov statistic value.
$KS_{19,\ 0.05} = 0.3010$.

Table V. Summary of variogram parameters for selected best fitted variogram models for different rain gauges combinations used in the network after 2011

Trial no.	Rain gauge combination in network after 2011[a] ($n = 19$)	Variogram parameters				Goodness-of-fit measures			
		Model	Nugget, C_0 (mm^2)	Sill, $C_0 + C_1$ (mm^2)	Range, a (km)	RMSE	MAE	RSS	R
(a) El Niño year									
1	17 + (stations 18 and 19)	Exponential	0.0285	0.2303	33.630	0.0408	0.0359	0.0133	0.732
2	17 + (stations 18 and 19)	Exponential	0.0264	0.2118	28.050	0.0371	0.0324	0.0110	0.742
3	17 + (stations 18 and 19)	Exponential	0.0357	0.2254	34.830	0.0268	0.0253	0.0058	0.831
4	17 + (stations 18 and 19)	Exponential	0.0001	0.1563	20.400	0.0488	0.0387	0.0191	0.716
5	17 + (stations 18 and 19)	Exponential	0.0263	0.2046	27.240	0.0339	0.0299	0.0092	0.749
6	17 + (stations 18 and 19)	Exponential	0.0173	0.1856	21.210	0.0310	0.0275	0.0077	0.748
7	17 + (stations 18 and 19)	Gaussian	0.0144	0.1355	12.454	0.0578	0.0426	0.0267	0.681
8	17 + (stations 18 and 19)	Exponential	0.0122	0.1858	27.150	0.0395	0.0335	0.0125	0.712
9	17 + (stations 18 and 19)	Spherical	0.0020	0.1247	12.020	0.0497	0.0399	0.0198	0.668
10	17 + (stations 18 and 19)	Gaussian	0.0334	0.1451	13.233	0.0456	0.0356	0.0167	0.702
11	17 + (stations 18 and 19)	Gaussian	0.0110	0.0476	13.614	0.0080	0.0070	0.0005	0.828
12	17 + (stations 18 and 19)	Exponential	0.0015	0.0553	25.710	0.0062	0.0058	0.0003	0.883
13	18 + (station 5)	Exponential	0.0010	0.1400	24.600	0.0584	0.0447	0.0273	0.632
14	18+ (station 6)	Exponential	0.0350	0.4290	89.580	0.0262	0.0248	0.0055	0.901
15	17 + (stations 5 and 6)	Exponential	0.0168	0.2907	75.930	0.0438	0.0367	0.0153	0.817
(b) La Niña year									
1	17 + (stations 18 and 19)	Exponential	0.0011	0.0383	115.710	0.0140	0.0121	0.0016	0.851
2	17 + (stations 18 and 19)	Exponential	0.0006	0.0396	104.280	0.0112	0.0098	0.0010	0.850
3	17 + (stations 18 and 19)	Exponential	0.0010	0.0508	153.300	0.0108	0.0099	0.0009	0.915
4	17 + (stations 18 and 19)	Exponential	0.0005	0.0106	32.460	0.0162	0.0134	0.0021	0.614
5	17 + (stations 18 and 19)	Exponential	0.0004	0.0156	44.670	0.0115	0.0096	0.0011	0.607
6	17 + (stations 18 and 19)	Exponential	0.0008	0.0148	43.860	0.0108	0.0091	0.0009	0.610
7	17 + (stations 18 and 19)	Exponential	0.0004	0.0139	43.380	0.0152	0.0126	0.0018	0.639
8	17 + (stations 18 and 19)	Exponential	0.0005	0.0225	71.220	0.0106	0.0088	0.0009	0.653
9	17 + (stations 18 and 19)	Exponential	0.0004	0.0244	76.590	0.0097	0.0082	0.0008	0.647
10	17 + (stations 18 and 19)	Exponential	0.0001	0.0130	37.530	0.0136	0.0113	0.0015	0.610
11	17 + (stations 18 and 19)	Exponential	0.0012	0.0266	87.000	0.0079	0.0064	0.0005	0.629
12	17 + (stations 18 and 19)	Exponential	0.0010	0.0345	121.140	0.0076	0.0067	0.0005	0.735
13	18 + (station 5)	Exponential	0.0001	0.0354	153.300	0.0177	0.0156	0.0025	0.864
14	18+ (station 6)	Exponential	0.0001	0.0611	153.300	0.0152	0.0134	0.0019	0.929
15	17 + (stations 5 and 6)	Exponential	0.0001	0.0371	153.300	0.0207	0.0173	0.0034	0.885

[a] Station numbers are the same as in Figure 1.
RMSE, root mean square error; MAE, mean absolute error; RSS, residual sum of square; R, coefficient of determination.

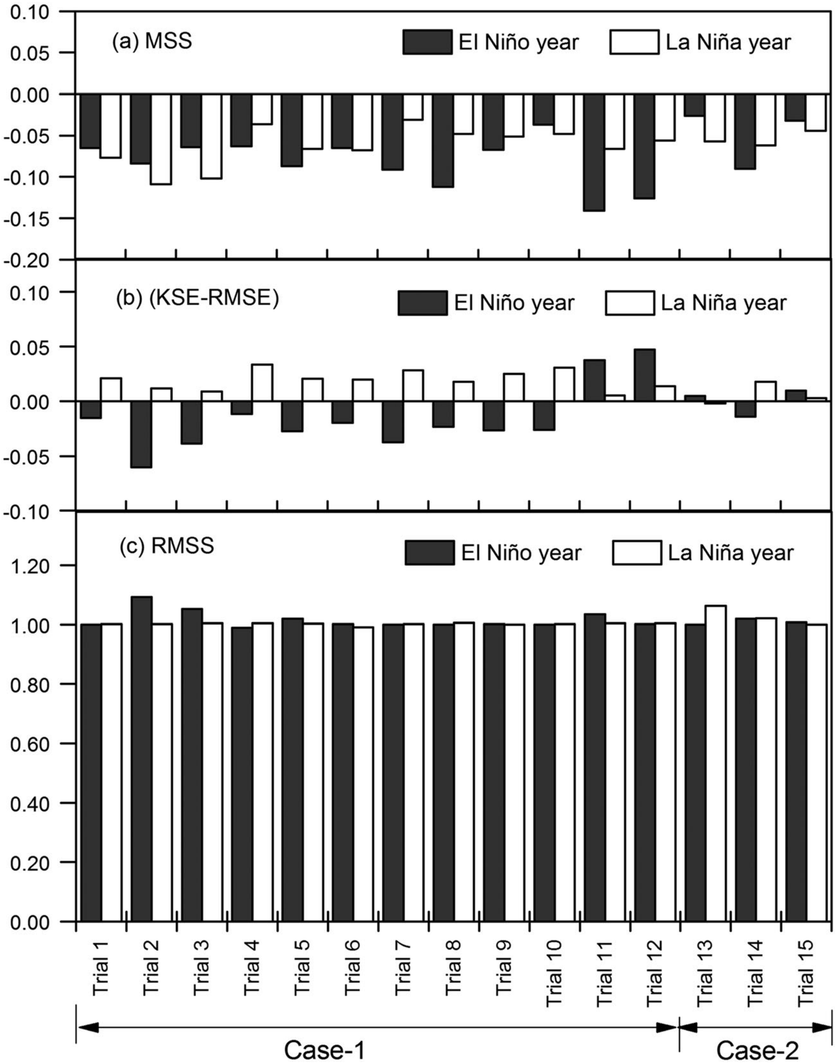

Figure 5. Cross-validation statistics (a) mean standardized prediction error (MSS), (b) difference of kriging standard error (KSE) and root mean square error (RMSE) and (c) root mean square standardized (RMSS) for El Niño and La Niña years. MSS and (KSE-RMSE) values close to 0 indicate an accurate model, whereas RMSS value close to 1 indicates an accurate model

network affects the kriging interpolated values and the corresponding KSE map. The reason is that in areas of sparse stations, the experimental variogram is more chaotic in nature and simulated surface produced by kriging carries large uncertainties (Journel and Huijbregts, 1978). Therefore, the placement of additional rain gauge stations in that area will help to minimize the kriging error and improve estimation accuracy that leads to an optimal rain gauge network.

According to the aforementioned considerations, two cases were considered, and further analysis was performed using the kriging method:

- Case-1: Optimal positioning of additional stations in the high variance areas. In this study, stations 18 and 19 installed by BoM in the network after 2011 were considered as additional stations, and determining their optimal locations was explored.
- Case-2: Redundant stations are either removed or are optimally relocated from low to high variance zones in the network. In this study, stations 5 and 6 in the network before 2011 (i.e. BoM's base network) were identified as redundant stations from Figure 4, and determining their optimal locations was explored.

Based on the aforementioned two cases, additional (stations 18 and 19) and redundant (stations 5 and 6) stations were placed in the high variance zones of the network after 2011, and their corresponding locations are shown in Figure 4. This results in a number of possible combinations for locating rain gauge stations that are given in Table III. This will allow one to check the optimality of the BoM's augmented network (i.e. network after 2011). This enables a decision maker to select the best combination of the number and location of stations in the network after 2011 that yield

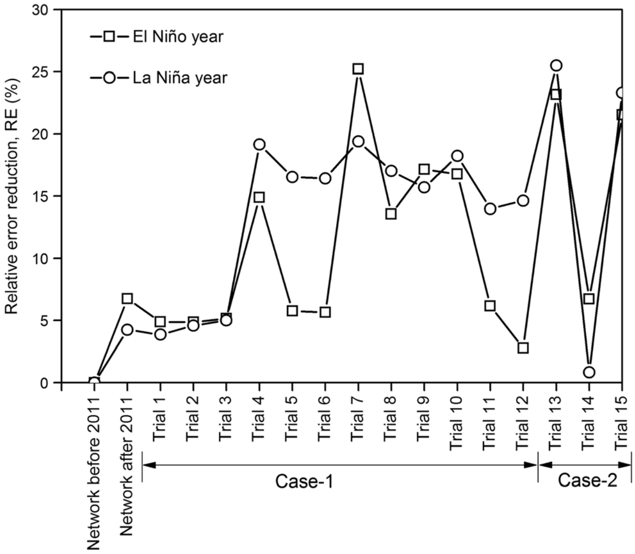

Figure 6. Relative error (RE) reduction for different combinations of stations in the network after 2011

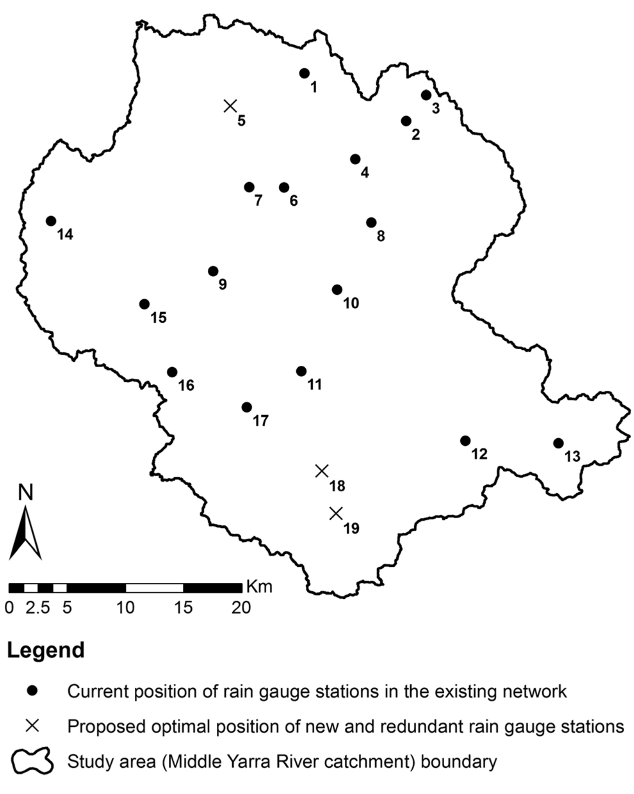

Figure 7. Devised optimal rain gauge network in the Middle Yarra River catchment

the highest reduction in KSE and hence the best network output and estimation accuracy. However, the remaining 17 rain gauge stations in the BoM's base network (i.e. network before 2011) were not relocated in the subsequent network augmentation because it is not practically feasible to relocate all existing rain gauge stations.

After selecting potential rain gauge sites in the high variance zones, rainfall values were estimated in those locations by kriging prediction. Predicted rainfall values of additional and redundant stations for scenarios 1 and 2 were combined with observed rainfall datasets of the remaining stations. Therefore, each combination of rain gauge stations got individual rainfall datasets. For each combination, exploratory data analysis, checking of data normality and variogram modelling were performed again for El Niño and La Niña years. Summary of the exploratory data analysis and data normality conditions for all rainfall datasets is given in Table IV. It can be seen from Table IV that in most cases of the El Niño year, the skewness values of rainfall datasets are close to zero, and hence, no transformation was necessary. However, log transformation was applied to the remaining datasets that were positively skewed, and it was found that skewness values approached to zero after log transformation. Furthermore, the K–S test confirms that the spatial datasets for all combinations have accepted the normal distribution in the 95% confidence level for both El Niño and La Niña years. Therefore, rainfall datasets for all trial combinations exhibit a normal distribution and are thus ready for further analysis.

Summary of the variogram modelling and cross-validation statistics for all rainfall datasets is provided in Table V and Figure 5, respectively. As seen from Table V, the exponential variogram model gives satisfactory result for all trial combinations in the La Niña year as well as for most cases in the El Niño year and thus can be applied further for the kriging analysis. The results also indicate that gaussian and spherical models produce satisfactory results for few combinations. The cross-validation statistics show that the unbiased and consistent estimates of variances are achieved for all rainfall datasets in the El Niño and La Niña years.

Optimal design of rain gauge network

For rain gauge network design, KSE values were computed for the network before 2011 and network after 2011. Although the estimation error and variability were reduced in some parts of the network, a number of regions exhibiting high KSE were still present. Rain gauge stations (additional and redundant) were placed at areas having higher values of KSE, and the process was repeated until the KSE values could not be reduced further. In this way, the process was repeated for a number of trials where the network was optimized with different combination of stations having individual rainfall datasets. The sites that resulted in the most significant reduction in KSE values were identified as the locations for placing additional as well as redundant stations. In the methodical search procedure, to obtain the optimal network for case-1, attempt was made first by one additional station (either station 18 or 19) and then with both

DOI: 10.1002/hyp

Table VI. Comparison between observed and estimated areal average rainfall

Option	Observed mean rainfall (mm)		Estimated mean rainfall (mm) for the rain gauge network				Error (%) obtained for the rain gauge network			
			El Niño year		La Niña year		El Niño year		La Niña year	
	El Niño year	La Niña year	BoM	Optimal	BoM	Optimal	BoM	Optimal	BoM	Optimal
Case-1	1.946	3.541	1.905	1.904	3.509	3.526	−2.2	−2.2	−0.9	−0.4
Case-2	1.946	3.541	1.905	1.949	3.509	3.520	−2.1	0.2	−0.9	−0.6

Case-1: Optimal positioning of new additional stations (stations 18 and 19) in the high variance areas
Case-2: Relocating and optimal positioning of redundant stations (stations 5 and 6) to the higher variance areas
BoM, Bureau of Meteorology.

additional stations (stations 18 and 19 together). Similarly, the optimal network search procedure for case-2 followed the relocation of one redundant station (either station 5 or 6) and then with both redundant stations (stations 5 and 6 together). The computation was performed for both the El Niño and La Niña years, and corresponding KSE values were estimated to compute the KSE reduction (Figure 6). As expected, the estimated KSE for different combination of stations was reduced as the number of stations was optimally combined that led to the optimal rain gauge network. Figure 6 shows that KSE values have decreased as the network is optimized with different combination of additional and redundant stations.

The result for case-1 demonstrates that high variance regions in the KSE map was defined by station combination of trial 7, which had the maximum reduction in KSE values. As seen in Figure 6, the maximum reduction in the KSE was obtained for this particular combination for both El Niño and La Niña years. Therefore, the optimal locations of the additional two stations in the network after 2011 (BoM's augmented network) can be represented by the station combination of trial 7. However, the result obtained for case-2 indicates that relocation of only one redundant station (station 5) defined by station combination of trial 13 in the high variance zone results in the maximum reduction in KSE values. It is interesting to note that relocation of both redundant stations (stations 5 and 6 together) defined by station combination of trial 15 in high variance zones gives a similar reduction in the KSE value. However, this alternative is less preferred because relocation and re-installing of a rain gauge station is associated with cost and manpower. Thus, the accepted feasible solution is the station combination defined by trial 13 for case-2 is selected as the optimal positioning of redundant stations. In this way, the optimal rain gauge network is achieved for the Middle Yarra River catchment, which is shown in Figure 7. The optimal rain gauge network consists of 19 stations, which includes the original 16 stations (stations 1–4, 6, 7–17), the redundant (station 5) station from the network before 2011 and two additional stations (stations 18 and 19) from the network after 2011 with their corresponding new locations.

The BoM's existing rain gauge network (Figure 1) and the designed optimal rain gauge network in this study (Figure 7) demonstrate that the optimal rainfall network provides more accurate estimates of areal average and point rainfalls in the Middle Yarra River catchment (Table VI). Although the improvement is insignificant in terms of mean daily rainfall scale as compared with actual datasets, the devised optimal network improves the areal average as well as point rainfall estimates. Therefore, it can be used for relevant hydrological applications.

CONCLUSIONS

On the basis of the results obtained in this study, the following conclusions can be drawn:

- The spatial structure and continuity of the rainfall data were modelled by using thee different variogram models (exponential, gaussian and spherical) for the selected El Niño and La Niña years. Cross-validation statistics were applied to test the validity of different variogram models and adequacy of estimated model parameters to be used for kriging applications. The results show that an isotropic gaussian model had the best fit with the experimental variogram generated from the mean daily rainfall datasets for both the network before 2011 and network after 2011.
- Kriging error map produced by OK shows that locations adjacent to the rain gauge stations exhibit lower error, whereas higher error is found in regions having less or no stations. It can be concluded that additional stations are necessary in regions that lack rain gauge stations and should be located accordingly to reduce the kriging error.
- It was found that if the additional stations (stations 18 and 19 together) installed in the network by BoM after 2011 are optimally located (as indicated in Figure 7), then the network yields improved estimates of areal average and

DOI: 10.1002/hyp

point rainfall for the La Niña year only, whereas no improvement could be achieved for the El Niño year.
- It was also found that if BoM's installed additional stations (stations 18 and 19 together) were considered to be in their original positions in the network after 2011, and only one redundant station (station 5) was optimally located without relocating other existing stations, the network yielded significant improvement in areal average and point rainfall estimates for both El Niño and La Niña years.
- Thus, this study has developed an optimal rainfall network for the Middle Yarra River catchment that consists of an optimal combination of rain gauge stations with the capability of providing more accurate areal average as well as point rainfall estimates.

The main recommendation arising from the results obtained in this study is to instal and maintain additional and redundant rain gauges in the Middle Yarra River catchment at locations indicated in Figure 7. It is not necessary to relocate the other existing rain gauge stations in the current network because of the associated costs. The concept proposed in this study for optimal design of rain gauge network through combined use of additional and redundant stations together is equally applicable to any other catchment.

ACKNOWLEDGEMENTS

This research was undertaken as a part of the first author's PhD programme at Victoria University, Australia. The authors express their appreciation to the University and the Commonwealth Government of Australia for providing the financial support in the form of IPRS scholarship. The authors are also thankful to the BoM for providing the necessary data.

REFERENCES

Allan RJ. 1988. El Niño Southern Oscillation influence in the Australasian region. *Progress in Physical Geography* **12**: 313–348. DOI: 10.1177/030913338801200301.

Barca E, Passarella G, Uricchio V. 2008. 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.

Barua S, Muttil N, Ng AWM, Perera BJC. 2012. Rainfall trend and its implications for water resource management within the Yarra River catchment, Australia. *Hydrological Processes* **27**: 1727–1738. DOI: 10.1002/hyp.9311.

Basalirwa CPK, Ogallo LJ, Mutua FM. 1993. The design of regional minimum rain gauge network. *International Journal of Water Resources Development* **9**: 411–424. DOI: 10.1080/07900629308722598.

Chebbi A, Bargaoui ZK, Cunha MDC. 2011. 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.

Chen YC, Wei C, Yeh, HC. 2008. Rainfall network design using kriging and entropy. *Hydrological Processes* **22**: 340–346. DOI: 10.1002/hyp.6292.

Cheng KS, Lin YC, Liou JJ. 2008. Rain-gauge network evaluation and augmentation using geostatistics. *Hydrological Processes* **22**: 2554–2564. DOI: 10.1002/hyp.6851.

Chowdhury RK, Beecham S. 2010. Australian rainfall trends and their relation to the southern oscillation index. *Hydrological Processes* **24**: 504–514. DOI: 10.1002/hyp.7504.

Cressie N. 1993. *Statistics for Spatial Data* (Revised Edition). John Wiley & Sons: New York.

Daly E, Kolotelo P, Schang C, Osborne CA, Coleman R, Deletic A, McCarthy DT. 2013. Escherichia coli concentrations and loads in an urbanised catchment: the Yarra River, Australia. *Journal of Hydrology* **497**: 51–61. DOI: 10.1016/j.jhydrol.2013.05.024.

Dutta SC, Ritchie JW, Freebairn DM, Abawi GY. 2006. Rainfall and streamflow response to El Nino Southern Oscillation: a case study in a semiarid catchment, Australia. *Hydrological Sciences Journal* **51**: 1006–1020. DOI: 10.1623/hysj.51.6.1006.

ESRI. 2009. *ArcGIS 9.3.1 (Computer Software Product of ESRI)*. Environmental Systems Research Institute (ESRI): Redlands, CA, USA.

Feki H, Slimani M, Cudennec C. 2012. Incorporating elevation in rainfall interpolation in Tunisia using geostatistical methods. *Hydrological Sciences Journal* **57**: 1294–1314. DOI: 10.1080/02626667.2012.710334.

Goovaerts P. 2000. Geostatistical approaches for incorporating elevation into the spatial interpolation of rainfall. *Journal of Hydrology* **228**: 113–129. DOI: 10.1016/S0022-1694(00)00144-X.

Haddad K, Rahman A, Zaman MA, Shrestha S. 2013. Applicability of Monte Carlo cross validation technique for model development and validation using generalised least squares regression. *Journal of Hydrology* **482**: 119–128. DOI: 10.1016/j.jhydrol.2012.12.041.

Isaaks EH, Srivastava RM. 1989. *Applied Geostatistics*. Oxford University Press: New York.

Jeffrey SJ, Carter JO, Moodie KB, Beswick AR. 2001. Using spatial interpolation to construct a comprehensive archive of Australian climate data. *Environmental Modelling and Software* **16**: 309–330. DOI: 10.1016/S1364-8152(01)00008-1.

Johnston K, VerHoef JM, Krivoruchko K, Lucas N. 2001. *Using ArcGIS Geostatistical Analyst*. ArcGIS Manual by ESRI: Redlands, CA, USA.

Journel AG, Huijbregts CJ. 1978. *Mining Geostatistics*. Academic Press: New York.

Kassim AHM, Kottegoda NT. 1991. Rainfall network design through comparative kriging methods. *Hydrological Sciences Journal* **36**: 223–240. DOI: 10.1080/02626669109492505.

Loof R, Hjorth P, Raut OB. 1994. Rainfall network design using the kriging technique: a case study of Karnali river basin, Nepal. *International Journal of Water Resources Development* **10**: 497–513. DOI: 10.1080/07900629408722649.

Mair A, Fares A. 2011. Comparison of rainfall interpolation methods in a mountainous region of a tropical island. *Journal of Hydrologic Engineering* **16**: 371–383. DOI: 10.1061/(ASCE)HE.1943-5584.0000330.

McCuen RH. 2003. *Modeling Hydrologic Change: Statistical Methods*. Lewis Publishers: USA.

Mekanik F, Imteaz MA, Gato-Trinidad S, Elmahdi A. 2013. Multiple regression and artificial neural network for long-term rainfall forecasting using large scale climate models. *Journal of Hydrology* **503**: 11–21. DOI: 10.1016/j.jhydrol.2013.08.035.

Melbourne Water. 2013. Port Phillip and Westernport Regional River Health Strategy Yarra catchment. Available at: http://melbournewater.com.au/aboutus/reportsandpublications/key-strategies/Documents/Port%20Phillip%20and%20Westernport%20Regional%20River%20Health%20Strategy%20-%20Yarra%20catchment.pdf (Accessed on: 26 December, 2013).

Mishra AK, Coulibaly P. 2009. Developments in hydrometric network design: a review. *Reviews of Geophysics* **47**: RG2001. DOI: 10.1029/2007RG000243.

Murphy BF, Ribbe J. 2004. Variability of southern Queensland rainfall and climate indices. *International Journal of Climatology* **24**: 703–721. DOI: 10.1002/joc.1018.

Nicholls N, Kariko A. 1993. East Australian rainfall events: interannual variations, trends, and relationships with southern oscillation. *Journal of Climate* **6**: 1141–1152. DOI: 10.1175/1520-0442(1993)006<1141:EAREIV>2.0.CO;2.

Papamichail DM, Metaxa IG. 1996. Geostatistical analysis of spatial variability of rainfall and optimal design of rain gauge network. *Water Resources Management* **10**: 107–127. DOI: 10.1007/BF00429682.

Pardo-Igúzquiza E. 1998. 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.

Shaghaghian MR, Abedini MJ. 2013. Rain gauge network design using coupled geostatistical and multivariate techniques. *Scientia Iranica* **20**: 259–269. DOI: 10.1016/j.scient.2012.11.014.

Shamsi UM, Quimpo RG, Yoganarasimhan GN. 1988. An application of kriging to rainfall network design. *Nordic Hydrology* **19**: 137–152. DOI:10.2166/nh.1988.010.

Sokolov S, Black KP. 1996. Modelling the time evolution of water-quality parameters in a river: Yarra River, Australia. *Journal of Hydrology* **178**: 311–335. DOI: 10.1016/0022-1694(95)02797-1.

Tsintikidis D, Georgakakos KP, Sperfslage JA, Smith DE, Carpenter TM. 2002. Precipitation uncertainty and rain gauge network design within Folsom Lake watershed. *Journal of Hydrologic Engineering* **7**: 175–184. DOI: 10.1061/(ASCE)1084-0699(2002)7:2(175).

Webster R, Oliver MA. 2007. *Geostatistics for Environmental Scientists* (2nd Ed). John Wiley & Sons: Chichester, United Kingdom.

Yeh HC, Chen YC, Wei C, Chen RH. 2011. Entropy and kriging approach to rainfall network design. *Paddy and Water Environment* **9**: 343–355. DOI: 10.1007/s10333-010-0247-x.