Optimal Design of a Rain Gauge Network to Improve Streamflow Forecasting

by

Sajal Kumar Adhikary

B.Sc. Engg. (Civil), M. Engg. (Water)

VICTORIA UNIVERSITY
MELBOURNE AUSTRALIA

Thesis submitted in fulfillment of the requirements for the degree of

Doctor of Philosophy

College of Engineering and Science, Victoria University, Australia

March 2017

Abstract

Enhanced streamflow forecasting has always been an important task for researchers and water resources managers. However, streamflow forecasting is often challenging owing to the complexity of hydrologic systems. The accuracy of streamflow forecasting mainly depends on the input data, especially rainfall as it constitutes the key input in transforming rainfall into runoff. This emphasizes the need for incorporating accurate rainfall input in streamflow forecasting models in order to achieve enhanced streamflow forecasting. Based on past research, it is well-known that an optimal rain gauge network is necessary to provide high quality rainfall estimates. Therefore, this study focused on the optimal design of a rain gauge network and integration of the optimal network-based rainfall input in artificial neural network (ANN) models to enhance the accuracy of streamflow forecasting. The Middle Yarra River catchment in Victoria, Australia was selected as the case study catchment, since the management of water resources in the catchment is of great importance to the majority of Victorians.

The study had three components. First, an evaluation of existing kriging methods and universal function approximation techniques such as genetic programming (GP) and ANN were performed in terms of their potentials and suitability for the enhanced spatial estimation of rainfall. The evaluation confirmed that the fusion of GP and ordinary kriging is highly effective for the improved estimation of rainfall and the ordinary cokriging using elevation can enhance the spatial estimation of rainfall.

Second, the design of an optimal rain gauge network was undertaken for the case study catchment using the kriging-based geostatistical approach based on the variance reduction framework. It is likely that an existing rain gauge network may consist of redundant stations, which have no contribution to the network performance for providing quality rainfall estimates. Therefore, the optimal network was achieved through optimal placement of additional stations (network augmentation) as well as eliminating or optimally relocating of redundant stations (network rationalization). In order to take the rainfall variability caused by climatic factors like El Niño Southern Oscillation into account, the network was designed using rainfall records for both El

Niño and La Niña periods. The rain gauge network that gives the improved estimates of areal average and point rainfalls for both the El Niño and La Niña periods was selected as the optimal network. It was found that the optimal network outperformed the existing one in estimating the spatiotemporal estimates of areal average and point rainfalls. Additionally, optimal positioning of redundant stations was found to be highly effective to achieve the optimal rain gauge network.

Third, an ANN-based enhanced streamflow forecasting approach was demonstrated, which incorporated the optimal rain gauge network-based input instead of using input from an existing non-optimal network to achieve the enhanced streamflow forecasting. The approach was found to be highly effective in improving the accuracy of streamflow forecasting, particularly when the current operational rain gauge network is not an optimal one. For example, it was found that use of the optimal rain gauge network-based input results in the improvement of streamflow forecasting accuracy by 7.1% in terms of normalized root mean square error (NRMSE) compared to the current rain gauge network based-input. Further improvement in streamflow forecasting was achieved through augmentation of the optimal network by incorporating additional fictitious rain gauge stations. The fictitious stations were added in sub-catchments that were delineated based on the digital elevation model. It was evident from the results that 18.3% improvement in streamflow forecasting accuracy was achieved in terms of NRMSE using the augmented optimal rain gauge network-based input compared to the current rain gauge network-based input. The ANN-based input selection technique that was employed in this study for streamflow forecasting offers a viable technique for significant input variables selection as this technique is capable of learning problems involving very non-linear and complex data.

Declaration

"I, Sajal Kumar Adhikary, declare that the PhD thesis by Publication entitled 'Optimal Design of a Rain Gauge Network to Improve Streamflow Forecasting' is no more than 100,000 words in length including quotes and exclusive of tables, figures, appendices, bibliography, references and footnotes. This thesis contains no material that has been submitted previously, in whole or in part, for the award of any other academic degree or diploma. Except where otherwise indicated, this thesis is my own work".

Signature: [redacted] Date: 20/03/2017

Dedication

This thesis and all the works behind it are dedicated to the memory of my late father

Acknowledgements

I would like to express my utmost appreciation and acknowledgement to my Principal Supervisor, Dr. Nitin Muttil for providing me with the invaluable guidance, necessary supports and constant inspirations throughout my PhD research at Victoria University. I am grateful to him for keeping his faith on my ability and accepting me as one of his research students, advising me in all of my academic pursuits, sharing his in-depth knowledge and expertise with me, and for the continuous support throughout my research. I have been especially fortunate to work with such a great supervisor. His insightful comments and suggestions in reviewing my writing were immensely helpful to keep me focused and successfully write the journal and conference papers on which this thesis is based. I believe that this thesis would not have been shaped to its present form without his wonderful supervision, proper guidance and motivation.

I owe my gratitude and sincere appreciation to my associate supervisor, Dr. Abdullah Gokhan Yilmaz for his discussions, helpful suggestions, friendly attitude in supervision, valuable comments and feedbacks on my research works and writing, which were very supportive for successful completion of my PhD research at Victoria University.

I would like to thank the Government of Australia and Victoria University for providing me with the full financial support through the International Postgraduate Research Scholarship (IPRS) scheme for this research project. I gratefully acknowledge the financial supports provided by the Secomb Conference and Travel Fund and the College of Engineering and Science Conference Funding Scheme in Victoria University to attend two conferences at Gold Coast and Hobart in Australia. I am highly indebted to Australian Bureau of Meteorology (BoM), SILO climate database, Geoscience Australia and Melbourne Water Corporation for supplying the necessary data for this research.

I wish to thank the respected faculty members of Water Resources Research Group, the officers and staff members of Graduate Research Center and Scholarship Unit in

Victoria University who helped me out directly or indirectly throughout my PhD research. I would also like to thank my fellow PhD colleagues for their enduring support and encouragement. I would like to especially thank Sushil Kumar Das, Stephanie C. Ashbolt and Prasad M. Inamdar for their cordial supports in data collection.

On a personal note, I would like to express my deepest love and gratitude to my mother, elder brother, and sister-in-law for their unconditional love, endless mental support, benevolent prayers and sacrifice during my long stay away from home, which has always been a constant source of my motivation and inspiration. Last but not least, I would like to thank my newly wedded wife, Sanjukta Das for her patience, unconditional love and sacrifice, continuous support and understanding, which were always there for me during this long journey.

List of Publications

Journal Articles

1. **Adhikary SK,** Yilmaz AG, Muttil N. 2015a. Optimal design of rain gauge network in the Middle Yarra River catchment, Australia. *Hydrological Processes* **29(11):** 2582-2599. DOI: 10.1002/hyp.10389.

2. **Adhikary SK,** Muttil N, Yilmaz AG. 2016a. Genetic programming-based ordinary kriging for spatial interpolation of rainfall. *Journal of Hydrologic Engineering* **21(2):** 04015062. DOI: 10.1061/(ASCE)HE.1943-5584.0001300.

3. **Adhikary SK,** Muttil N, Yilmaz AG. 2016b. Ordinary kriging and genetic programming for spatial estimation of rainfall in the Middle Yarra River catchment, Australia. *Hydrology Research* **47(6):** 1182-1197. DOI: 10.2166/nh.2016.196.

4. **Adhikary SK,** Muttil N, Yilmaz AG. 2017a. Cokriging for enhanced spatial interpolation of rainfall in two Australian catchments. *Hydrological Processes* **31(12):** 2143-2161. DOI: 10.1002/hyp.11163.

5. **Adhikary SK,** Muttil N, Yilmaz AG. 2017b. Improving streamflow forecasts using optimal rain gauge network-based input to artificial neural network models. *Hydrology Research* (Revised based on the reviewers' comments and resubmitted to the journal).

Refereed Conference Papers

6. **Adhikary SK,** Muttil N, Yilmaz AG. 2015b. Improved spatial interpolation of rainfall using genetic programming. *Proceedings of the 21st International Congress on Modelling and Simulation 2015*, *MODSIM 2015*, Gold Coast, Queensland, Australia, 29 November-6 December, 2015, (CD-ROM) pp 2214-2220.

   Accessible through the congress website on the following link:
   http://www.mssanz.org.au/modsim2015/L7/adhikary.pdf

7. **Adhikary SK,** Yilmaz AG, Muttil N. 2015c. Kriging-based geostatistical approach for optimal design of rain gauge network - A case study in the Middle Yarra River catchment, Australia. *Proceedings of the 36th Hydrology and Water Resources Symposium 2015, HWRS 2015*, Hobart, Tasmania, Australia, 7-10 December, 2015 (CD-ROM). pp 791-798.

   Accessible through the symposium website on the following link:
   https://search.informit.com.au/documentSummary;dn=824513108125420;res=IELENG

VICTORIA UNIVERSITY
MELBOURNE AUSTRALIA

PART A:

DETAILS OF INCLUDED PAPERS: THESIS BY PUBLICATION

Please list details of each Paper included in the thesis submission. Copies of published Papers and submitted and/or final draft Paper manuscripts should also be included in the thesis submission

Item/ Chapter No.	**Paper Title**	**Publication Status** (e.g. published, accepted for publication, to be revised and resubmitted, currently under review, unsubmitted but proposed to be submitted)	**Publication Title and Details** (e.g. date published, impact factor etc.)
3	Genetic Programming-Based Ordinary Kriging for Spatial Interpolation of Rainfall	Published	Journal of Hydrologic Engineering, 2016 SCImago Journal Rank: Q1 Impact Factor: 1.694
4	Ordinary Kriging and Genetic Programming for Spatial Estimation of Rainfall in the Middle Yarra River Catchment, Australia	Published	Hydrology Research, 2016 SCImago Journal Rank: Q1 Impact Factor: 1.779
	Cokriging for Enhanced Spatial Interpolation of Rainfall in Two Australian Catchments	Published	Hydrological Processes, 2017 SCImago Journal Rank: Q1 Impact Factor: 3.014
5	Optimal Design of Rain Gauge Network in the Middle Yarra River Catchment, Australia	Published	Hydrological Processes, 2015 SCImago Journal Rank: Q1 Impact Factor: 3.014
6	Improving Streamflow Forecast Using Optimal Rain Gauge Network-Based Input to Artificial Neural Network Models	Revised and resubmitted to the journal	Hydrology Research, 2017 Submitted in February 2017 SCImago Journal Rank: Q1 Impact Factor: 1.779

Declaration by [candidate name]:	**Signature:**	**Date:**
Sajal Kumar Adhikary	[redacted]	20/03/2017

Table of Contents

Chapter 2

Chapter 3

Chapter 4

Chapter 5

Chapter 6

Chapter 7

Chapter 1

Introduction

1.1 Background

Accurate forecasting of streamflow is essential for many of the activities associated with the effective planning and operation of the components of risk-based water resources systems. Particularly, streamflow forecasting is of vital importance to flood control and mitigation, and water resources planning and management systems. The analysis and design of hydraulic structures such as dams and bridges, management of extreme events including floods and droughts, optimal operation of reservoir for activities including irrigation water requirement, hydropower generation, domestic and industrial water supply objectives are a few examples wherein streamflow forecasting provides crucial information (Jain and Kumar, 2007; Londhe and Charhate, 2010). The purpose of forecasting is to minimize the risk in a decision taken, at any given point of interest, in various water engineering applications. Hence, there is a growing need for both short-term and long-term forecasting of future streamflow in order to optimize the water resources systems in an efficient way (Akhtar et al., 2009). According to Ruiz et al. (2007), a reliable and improved forecasting of streamflow ensures rational regulation of river runoff, which ultimately results in the enhanced flood control and protection. Improving the accuracy and performance of streamflow forecasting has, therefore, always been an absolute necessity for researchers and hydrologists.

While floods contribute to a greater loss of life and property damage leading to huge economic loss, droughts induced water scarcity limits water uses across the world (Makkeasorn et al., 2008). In Australia, floods and droughts have also affected many regions of several states (van den Honert and McAneney, 2011; National Climate Centre, 2011; van Dijk et al., 2013; Low et al., 2015). Australia is often referred to as the land of 'drought and flooding rains' (Pittock and Connell, 2010). This contrast was more evident than ever in the last two decades, when an extensive period of drought (the 1997-2009 Millennium Drought) was followed by a series of large-scale floods occurring in 2010-2011, which caused havoc in many parts of southeastern and eastern Australia (Bureau of Meteorology, 2011; Victorian Government, 2011). For example, floods that affected eastern Australia in January 2011 caused extensive damage worth approximately $A126 Million (AIDR, 2015). Floods are usually Australia's most expensive natural hazard with the average annual cost of flooding being about $A377 Million (BITRE, 2008). Likewise, severe drought in 2002–2003 caused a $A7.4 Billion drop in Australia's agricultural production and for that year, economic losses were equivalent to 1.6 percent of country's gross domestic product (Lu and Hedley, 2004). Floods and droughts induced water management problems have thus become the important concerns, especially in southeastern Australia, which require immediate action. Australia's population and agricultural production are highly concentrated in the southeastern part of the country (Murphy and Timbal, 2008). Water resources play a key role in the economic development of the region. The region's growing population and resulting new demands on limited water resources require efficient management of existing water resources as well as identifying alternative sources of water. In order to address these water management challenges, it is widely recognized that maximizing water management efficiency based on streamflow forecasting is vital.

Due to the importance of hydrologic forecasting, a large number of forecasting models and methodologies have been developed and applied in streamflow forecasting. These streamflow forecasting models can be categorized as process-driven models and data-driven models (Wang, 2006). The process-driven models depend on the underlying physical processes, which require a large amount of data and huge computational efforts. Various forms of lumped, semi-distributed and distributed rainfall-runoff models are included in this category. Data-driven models, on the other hand, are based on a limited knowledge of the internal physical mechanism of the catchment and rely on data describing input and output characteristics. They are essentially black-box models that characterize the relationships between input and outputs without explicit simulation of the underlying physical processes. They may include regression and time series models as well as machine learning-based models like artificial neural networks (ANN) and genetic programming (GP) models. Recently, data-driven models have gained widespread popularity in streamflow forecasting due to the data availability from monitoring stations, real-time data retrieval, and availability of advanced computing facilities (Wang, 2006). These models usually perform better than others in situations where the underlying interactions and dependencies of physical processes are unknown or only partially understood (Maier and Dandy, 2000). These models are easy to set up and are able to produce acceptable results with minimum input data (rainfall and discharge) (e.g., Talei et al., 2010; Londhe and Charhate, 2010; Yilmaz and Muttil, 2014). Due to this simplicity, data-driven models, particularly ANN have been widely used for streamflow forecasting across the word (e.g., Zealand et al., 1999; Dibike and Solomatine, 2001; Birikundavyi et al., 2002; Huang et al., 2004; Wu et al. 2005; Srinivasulu and Jain, 2009; Sivapragasam et al., 2014; Taormina et al., 2015).

Generally, streamflow forecasting is challenging due to the complexity of hydrologic systems. However, the key challenge of achieving enhanced accuracy of streamflow forecasting remains. This is by no means an easy task because there is no single streamflow forecasting method that provides optimum forecast results for all types of catchments and under all circumstances (Shamseldin, 2004). Furthermore, there is always uncertainty when it comes to forecasting because it is unlikely to forecast the exact future conditions. As a result of these difficulties, it is necessary to devise a viable streamflow forecasting approach for improved estimation of future streamflows with ease and better accuracy. The accuracy of streamflow forecasting primarily depends on the input data, especially rainfall data as it constitutes the key input in transforming rainfall into runoff (Maskey et al., 2004; Jones et al., 2006; Ekström and Jones, 2009). Hence, rainfall is one of the most important inputs to develop streamflow forecasting models. Since streamflow is a consequence of rainfall, uncertainty associated with rainfall causes uncertainty in estimated streamflow and adversely affects the accuracy of streamflow forecasting (Faurès et al., 1995; Tsintikidis et al., 2002; Moulin et al., 2009). This emphasizes the importance of using accurate rainfall data in streamflow forecasting models to meet the challenge of enhanced streamflow forecasting. Thus, it can be expected that the enhanced accuracy in streamflow forecasting can be achieved if catchment rainfall is estimated accurately and fed into streamflow forecasting models.

Rain gauge networks are usually installed to get direct measurements of rainfall. However, many of the water resources systems are large in spatial extent and often consist of a rain gauge network that is very sparse due to financial, logistics and geological factors. This results in considerable uncertainty in the rainfall data that are available (Zealand et al., 1999). Rainfall often shows significant spatial variations within a catchment or region (Faurès et al., 1995; Krajewski et al., 2003). As an example, rainfall variability in eastern Australia is mainly caused by the dominant influence of El Niño Southern Oscillation (ENSO) effect (Murphy and Ribbe, 2004; Pittock et al., 2006), which also cause uncertainty in rainfall data. As a result, existing standard rain gauge networks established in the past in Australian catchments often fail to adequately address the rainfall variability. Therefore, the design and establishment of an optimal rain gauge network is an important task in this study to obtain high quality rainfall estimates. An essential advantage of having such an optimal network is to achieve improved streamflow forecasting. The optimal network improves the accuracy in streamflow forecasting by providing accurate estimates of rainfall data with little or no uncertainty (Tsintikidis et al., 2002) that are used as the input to streamflow forecasting models.

Rainfall is often considered independent of streamflow simulation and forecasting such as estimation of areal average rainfall over a catchment (e.g., Bastin et al., 1984; Seed and Austin, 1990; Mishra, 2013) or the design of rain gauge networks (e.g., Papamichail and Metaxa, 1996; Pardo-Igúzquiza, 1998; Tsintikidis et al., 2002; Cheng et al., 2008; Aziz et al., 2016; Feki et al., 2017). However, this does not allow one to focus on the strength and weakness of an established optimal rain gauge network that really matter when rainfall data from the optimal network are fed into streamflow forecasting models. Therefore, it is logical to design an optimal rain gauge network for providing a satisfactory solution to the specific requirement of enhanced streamflow forecasting for which the network is being established. In the past, rainfall data were used in streamflow forecasting models directly from the existing rain gauge network (which may not be an optimal network) rather than using improved rainfall data from an optimally designed rain gauge network (e.g., St-Hilaire et al., 2003; Dong et al., 2005; Anctil et al., 2006; Xu et al., 2006; Bárdossy and Das, 2008; Ekström and Jones, 2009; Moulin et al., 2009; Xu et al., 2013; Tsai et al., 2014). This ultimately results in the less accurate forecasting of streamflow. Thus, it can be hypothesized that the integration of rainfall input from an optimal rain gauge network with streamflow forecasting models is expected to improve the accuracy of streamflow forecasting. Therefore, this thesis focused on the optimal design of a rain gauge network in order to improve the accuracy of streamflow forecasting. The results of this thesis would contribute to the development of appropriate water management plans and optimal reservoir operation strategies for effective flood control and drought management specifically in southeastern Australia.

1.2 Aims of the Study

The main aim of this study was to improve the performance of streamflow forecasting at the outlet of a catchment by incorporating the optimal rain gauge network-based input (instead of incorporating input from the existing non-optimal rain gauge network) to artificial neural network (ANN)-based streamflow forecasting models. The research framework of this study consists of two integral parts, which have been linked together to establish an enhanced data-driven modelling framework in order to improve the accuracy of streamflow forecasting. In the first part, the aim was to analyze the spatial variability of rainfall and re-design the existing non-optimal rain gauge network using the kriging-based geostatistical approach to achieve the optimal rain gauge network. In the second part, the aim was to use rainfall data from an optimally designed rain gauge network (obtained from the first part) in addition to streamflow records as the input to the ANN-based streamflow forecasting models to forecast the streamflow values at the catchment outlet. The following specific aims were considered under the above mentioned main aim of the study:

- Assessment of the potential of incorporating universal function approximation-based variogram models under the kriging-based geostatistical framework for improved estimation of rainfall.
- Evaluation of the performance of different conventional and universal function approximation-based kriging methods for spatial interpolation of rainfall.
- Analysis of the potential of using supplementary information (elevation) along with primary variable (rainfall) in the kriging-based geostatistical framework for enhanced estimation of rainfall.
- Development of a simple and effective rain gauge network design methodology under the variance reduction framework using the kriging-based geostatistical approach for optimal design of a rain gauge network through optimal positioning of additional rain gauge stations (network augmentation) as well as removing and/or optimally relocating of existing redundant rain gauge stations (network rationalization).
- Assessment of the effectiveness of incorporating rainfall input from an optimal rain gauge network (instead of using input from an existing non-optimal rain gauge network) in the ANN-based streamflow forecasting models to achieve the enhanced accuracy in streamflow forecasting.

1.3 Research Methodology in Brief

In order to achieve the aforementioned aims as stated in Section 1.2, the following tasks were undertaken in this research project.

1. Review of rain gauge network design and its impact on the accuracy of streamflow simulation.
2. Development of a universal function approximation-based kriging for improved estimation of rainfall.
3. Evaluation of different univariate and multivariate kriging methods for enhanced spatial interpolation of rainfall.

4. Optimal design of a rain gauge network using the kriging-based geostatistical approach.
5. Improvement of streamflow forecasts using optimal rain gauge network-based input.

The aforementioned tasks under the methodological framework of this research project were demonstrated through a case study area, which comprises the middle segment of the Yarra River catchment in Victoria, Australia. The study area is referred to as the 'Middle Yarra River catchment' in this thesis. A brief description of the study area along with its selection and importance for the effective management of water resources is given in Section 1.4. It is important to note that the methodological framework that is proposed and demonstrated in this thesis can be used for any other study area or catchment. Brief description of each of the aforementioned tasks is given in the following.

1.3.1 Review of rain gauge network design and its impact on the accuracy of streamflow simulation

There are several methods that have been used in the past for the design and evaluation of a rain gauge network (Mishra and Coulibaly, 2009). Among those, two basic methods namely kriging-based geostatistical approach (e.g., Shamsi et al., 1988; Kassim and Kottegoda, 1991; Loof et al., 1994; Papamichail and Metaxa, 1996; Pardo-Igúzquiza, 1998; Ghahraman and Sepaskhah, 2001; Tsintikidis et al., 2002; Nour et al., 2006; Barca et al., 2008; Chen et al., 2008; Cheng et al., 2008; Chebbi et al., 2011; Yeh et al., 2011; Putthividhya and Tanaka, 2013; Shaghaghian and Abedini, 2013; Shafiei et al., 2014; Adib and Moslemzadeh, 2016; Aziz et al., 2016; Haggag et al., 2016; Chang et al., 2017; Feki et al., 2017) and information theory-based entropy approach (e.g., Husain, 1989; Krstanovic and Singh, 1992a,b; Al-Zahrani and Husain, 1998; Yoo et al., 2008; Karimi-Hosseini et al., 2011; Ridolfi et al., 2011; Vivekanandan and Jagtap, 2012; Li et al., 2012; Xu et al., 2015; Werstuck and Coulibaly, 2016; Xu et al., 2017) are mostly used for optimal design of a rain gauge network by researchers and professionals around the world.

A number of issues was considered for selecting the rain gauge network design method, including understanding of the concept and simplicity of use (i.e., time and effort required to arrive at a conclusion), strength and weakness of the approach (i.e., quantity and quality of necessary data, data availability, and ability to handle uncertainty), and availability of user-friendly software packages to accomplish the analysis. Based on the above-mentioned issues and factors, the kriging-based geostatistical approach was selected for optimal design of rain gauge network in this research project. Thus, a review of the kriging-based geostatistical approach for analyzing the spatial variability of rainfall and optimal design of a rain gauge network was undertaken, at first, to understand the concepts, applicability, existing gaps of the approach and likely modifications for further improvement. In addition, there are several studies that have been explored in the past to study the spatial rainfall variability and impact of rain gauge network density on streamflow estimation and forecasting (e.g., Faurès et al., 1995; Arnaud et al., 2002; St-Hilaire et al., 2003; Dong et al., 2005; Anctil et al., 2006; Xu et al., 2006; Bárdossy and Das, 2008; Ekström and Jones, 2009; Moulin et al., 2009; Volkmann et al., 2010; Xu et al., 2013; Tsai et al., 2014; Kar et al., 2015; Zeinivand, 2015; Chen et al., 2017). Hence, a detailed review of those studies was also carried out to identify the existing gaps and likely modifications in order to formulate an effective approach to achieve the enhanced accuracy in streamflow forecasting.

1.3.2 Development of a universal function approximation-based kriging for improved estimation of rainfall

Based on the literature survey of the different kriging-based geostatistical approaches for optimal design of a rain gauge network (Section 1.3.1), it was found that traditional kriging has a major weakness because it requires a priori definition of the mathematical function for a variogram model. The variogram model represents spatial correlations among data points, which plays a vital role in the kriging process and thus significantly impacts the performance of the kriging method (Wackernagel, 2003; Webster and Oliver, 2007). Thus, the robustness of kriging method heavily depends on the fitting and estimation of a correct variogram model. Furthermore, selection of an appropriate variogram model, finding the optimal variogram parameters (i.e., nugget, sill and range) and the computational burdens involved are some of the difficulties associated with the traditional kriging (Teegavarapu, 2007; Oliver and Webster, 2014).

As a solution to those issues, a new universal function approximation-based kriging was developed in this research project where genetic programming (GP) was used as a universal function approximator to derive the variogram model. This new variant of kriging is referred to as the genetic programming-based ordinary kriging (GPOK) in which the standard parametric variogram models (i.e., exponential, gaussian, spherical models) in traditional ordinary kriging were replaced by the GP-derived variogram model. The GPOK was developed in this research project in order to overcome the limitations associated with the traditional ordinary kriging. The performance of the GPOK was then tested against the performance of traditional ordinary kriging as well as artificial neural network (ANN)-based ordinary kriging for improved estimation of rainfall at desired unsampled locations in the case study catchment.

1.3.3 Evaluation of different univariate and multivariate kriging methods for enhanced spatial interpolation of rainfall

The literature survey undertaken in Section 1.3.1 indicated that a wide variety of deterministic and stochastic (kriging-based geostatistics) interpolation methods have been used in the past for spatial interpolation of spatially distributed point rainfall values and production of high quality continuous rainfall datasets in a range of different regions and climates around the world (e.g., Hevesi et al., 1992; Dirks et al., 1998; Goovaerts, 2000; Lloyd, 2005; Hsieh et al., 2006; Moral, 2010; Di Piazza et al., 2011; Mair and Fares, 2011; Chen and Liu, 2012; Feki et al., 2012; Delbari et al., 2013). It is often challenging to select the suitable interpolation method arbitrarily, which performs best for estimating the spatial distribution of rainfall for a particular study area. The reason is that the performance of an interpolation method depends on a number of factors including catchment size and characteristics, sampling density and its spatial distribution, surface type, data variance, interpolation grid resolution, quality of available auxiliary information (Li and Heap, 2011).

There had been also a number of spatial rainfall interpolation studies undertaken at a regional or national scale in Australia (e.g., Hutchinson, 1995; Hancock and Hutchinson, 2006; Jones et al., 2009; Li and Shao, 2010; Woldemeskel et al., 2013; Yang et al., 2015). However, none of these studies was conducted at a local or catchment scale. Furthermore, elevation and rainfall relationship locally have been relatively little studied in Australia. As a solution to the aforementioned issues, comparative evaluation of a range of selected univariate and multivariate interpolation methods was carried out to identify the most appropriate interpolator for spatial interpolation of rainfall and generation of continuous rainfall maps for the case study catchment. The GPOK approach was also evaluated for spatial interpolation of rainfall and generation of continuous rainfall maps for the case study catchment, which was found to be the best interpolation method for rainfall estimation (Section 1.3.2). Based on the rainfall-topography relationship of the case study catchment, the elevation was used as an auxiliary variable in addition to rainfall (i.e., primary variable) for the adopted multivariate or cokriging methods in this research project. A number of cross-validation criteria were used for the evaluation of the adopted univariate and multivariate interpolation methods in order to select the best interpolation methods for the case study catchment.

1.3.4 Optimal design of a rain gauge network using kriging-based geostatistical approach

As was mentioned in Section 1.3.1, several rain gauge network design methods were developed in the past for rain gauge network design and evaluation around the world (Mishra and Coulibaly, 2009). The kriging-based geostatistical approach was found to be the most suitable rain gauge network design technique across the world (Section 1.3.1) (e.g., Shamsi et al., 1988; Kassim and Kottegoda, 1991; Loof et al., 1994; Papamichail and Metaxa, 1996; Pardo-Iguzquiza, 1998; Ghahraman and Sepaskhah, 2001; Tsintikidis et al., 2002; Cheng et al., 2008; Shaghaghian and Abedini, 2013; Adib and Moslemzadeh, 2016; Aziz et al., 2016; Haggag et al., 2016; Feki et al., 2017), which was also used in this research project for optimal design of a rain gauge network in the case study catchment. Network expansion with additional stations (network augmentation) for variance reduction has been the underlying criterion to achieve the optimal network in most of the past studies. However, an existing rain gauge network may consist of redundant stations, which have little or no contribution to the network performance for providing the high quality rainfall estimates (St-Hilaire et al., 2003; Mishra and Coulibaly, 2009).

As a solution to the aforementioned issue, both additional and redundant stations were considered to achieve the optimal rain gauge network design in this research. The optimal rain gauge network for the case study catchment was thus achieved through optimal positioning of additional rain gauge stations (network augmentation process) in addition to removing and/or optimally relocating of existing redundant rain gauge stations (network rationalization process). The spatial variability of rainfall in southeastern Australia (where the case study catchment is located) is usually caused by the El Niño and La Niña processes of the ENSO effect (Murphy and Ribbe, 2004; Dutta et al., 2006; Pittock et al., 2006). As a solution to this issue, the rain gauge network was designed independently based on rainfall records obtained for both El Niño and La Niña periods. The rain gauge network that gave the enhanced estimates of areal average and point rainfalls for both El Niño and La Niña periods was then selected as the optimal rain gauge network for the case study catchment.

1.3.5 Improvement of streamflow forecasting using optimal rain gauge network-based input

As was stated in Section 1.1, streamflow forecasting models in most of the past studies generally used the rainfall data directly obtained from the existing rain gauge network (e.g., Kumar et al., 2004; Kişi, 2007; Londhe and Charhate, 2010), which might exhibit high variance and hence not be the optimal rain gauge network. Since streamflow is a consequence of rainfall, uncertainty associated with the rainfall input propagates through the streamflow simulation (Faurès et al., 1995; Arnaud et al., 2002; St-Hilaire et al., 2003; Dong et al., 2005; Ekström and Jones, 2009; Moulin et al., 2009), which may ultimately lead to the less accurate forecasting of streamflows. As a solution to this issue, an enhanced streamflow forecasting approach was developed in the current research using the rainfall input from an optimal rain gauge network (achieved through Section 1.3.4) instead of using the input from an existing rain gauge network (which may not be an optimal network) for enhanced streamflow forecasting. The GPOK approach, which was found to be the best interpolation method for rainfall estimation at ungauged locations (Section 1.3.2) was used to estimate rainfall data at the identified optimal locations (ungauged points) for the corresponding additional and redundant rain gauge stations in the optimal rain gauge network (Section 1.3.4).

Streamflow forecasting models were developed using the rainfall input from both current and optimal rain gauge networks in addition to streamflow records. The forecasting performance of those models was then evaluated and compared to test the robustness of the proposed streamflow forecasting approach. The artificial neural network (ANN), which was found to be the best suitable technique for streamflow forecasting in many past studies (e.g., Zealand et al., 1999; Dibike and Solomatine, 2001; Birikundavyi et al., 2002; Huang et al., 2004; Wu et al. 2005; Akhtar et al., 2009; Srinivasulu and Jain, 2009; Sivapragasam et al., 2014; Yilmaz and Muttil, 2014; Linares-Rodriquez et al., 2015; Taormina et al., 2015), was also used in the current study to formulate streamflow forecasting models. Since the streamflow-rainfall relationship is highly non-linear and ANN is able to capture such highly non-linear relationship between the input and the output of a system, the ANN-based input selection approach was used to select significant input variables for developing the ANN-based streamflow forecasting models. Furthermore, based on the ANN-based input variable selection approach, an indirect way of identifying the optimal locations of rain gauge stations in the final operational optimal rain gauge network was carried out in order to achieve the best streamflow forecasting performance of the network. Finally, the robustness and efficacy of the proposed approach was investigated through a several quantitative standard statistical performance evaluation measures.

1.4 The Study Area

1.4.1 Background and selection of the study area

As was stated in Section 1.3, the Middle Yarra River catchment (the middle segment of the Yarra River catchment) located in Australian state of Victoria was selected as the case study area in the current research study. The location of the study area is shown in

Figure 1-1. The Middle Yarra River catchment covers an area of 1511 km^2, which is mainly characterized by the rural floodplains and valleys with limited urban developments. The study area is notable as the only part of the Yarra River catchment with an extensive flood plain, wherein most of the lands are used for agricultural activities (Gardner, 1994; Carty and Pierotti, 2010).

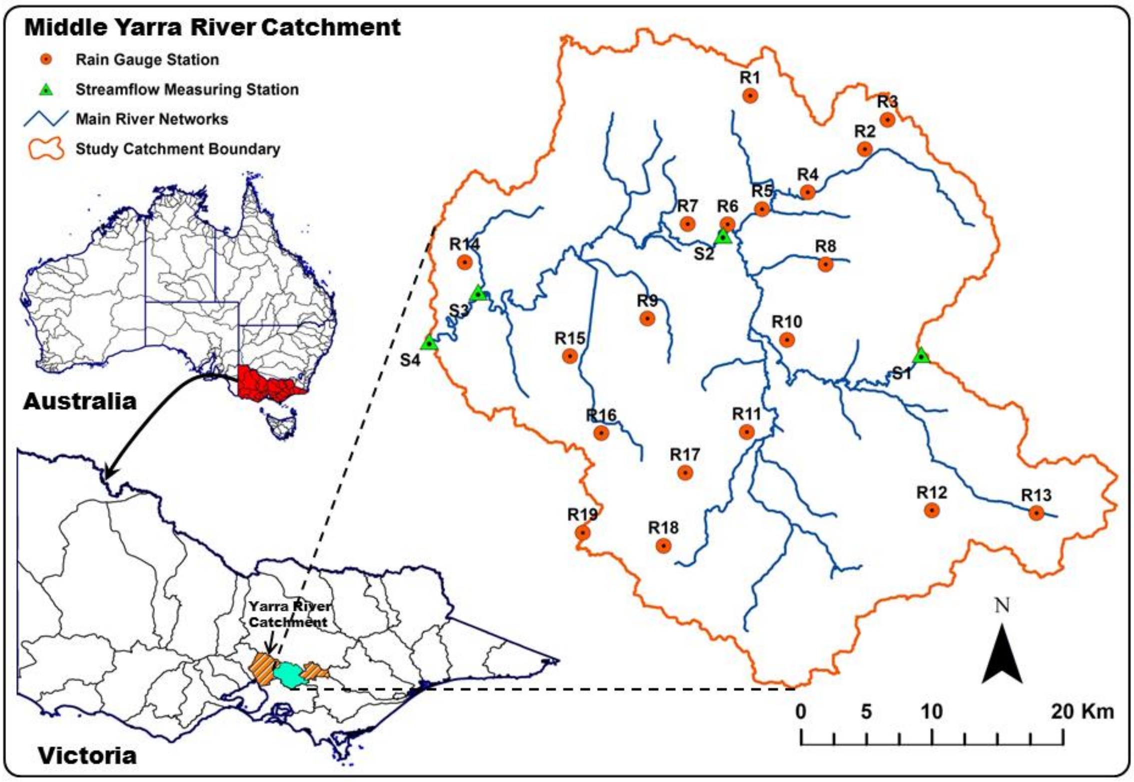

Figure 1-1. Location map of the Middle Yarra River catchment (the study area in this study) with rain gauge and streamflow measuring stations

The Yarra River catchment is an important water resources catchment of Victoria because majority of Victorians depends on the water resources of this catchment. The catchment is home to more than one-third of Victoria's population (approximately 1.8 million) and native plant and animal species, where the Yarra River acts as the only lifeline. The water resources management in the Yarra River catchment is of great importance considering the diverse water use activities and high variability of rainfall (Muttil et al., 2009; Barua et al., 2012). The catchment water resources support a wide range of water uses valued by the Melbourne's community, including urban water supply, agricultural, horticultural industries and downstream user requirements as well as flow requirements for maintaining environmental flows. The Yarra River catchment covers an area of 4044 km^2, which is shown in Figure 1-2. The Yarra River flows from east to west and 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 (EPA Victoria, 1999).

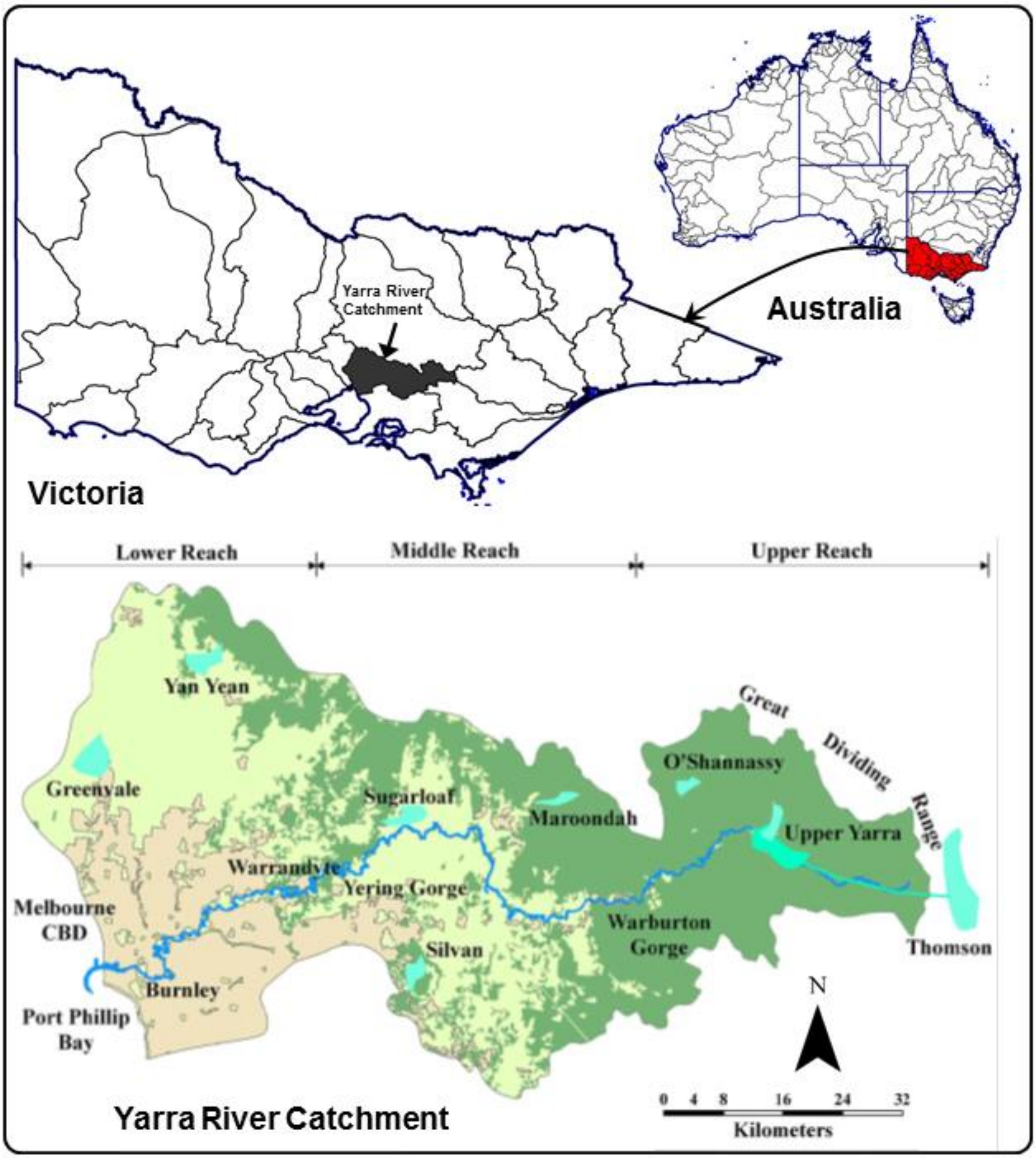

Figure 1-2. Details of the Yarra River catchment in Victoria, Australia

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, 2015). Melbourne city is located in the lower part of the Yarra River catchment. There are seven major storage reservoirs (Upper Yarra, O'Shannassy, Maroondah, Silvan, Sugarloaf, Yan Yean and Greenvale) located within the catchment as shown in Figure 1-2. These reservoirs are mainly used for urban water supply to Melbourne and storage purposes. In addition, flows from Thomson reservoir, which is located outside the catchment, is transferred to the Upper Yarra reservoir mainly for urban water supply purposes. There are many farm dams and licensed water extraction points in the Yarra River and its tributaries within the catchment (Melbourne Water, 2013). A range of recreational activities, metropolitan parks and biodiversity conservation is also located around the catchment waterways. The Yarra River, being the main river of the catchment, has thus played a key role in the way Melbourne has developed and grown.

Rainfall is the main source of water resources in the Yarra River catchment like many other catchments in Australia. The average 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), contributing to higher flows in the Yarra River during winter and spring (Melbourne Water, 2015). However, the annual average rainfall has declined during the last decade compared to the long-term historical average (Muttil et al, 2009). Figure 1-3 shows the annual average rainfall based on the 22 rain gauge stations in the Yarra River catchment for the period of 1960 to 2008 as analyzed by Barua (2010). As can be seen from the figure, the average annual rainfall was 831.1 mm after 1997 whereas it was 1031.9 mm before 1997. As a result of such high spatial and temporal variability of rainfall and diverse water use activities across the catchment, the water resources management in the Yarra River catchment remains an important and challenging task for the water managers.

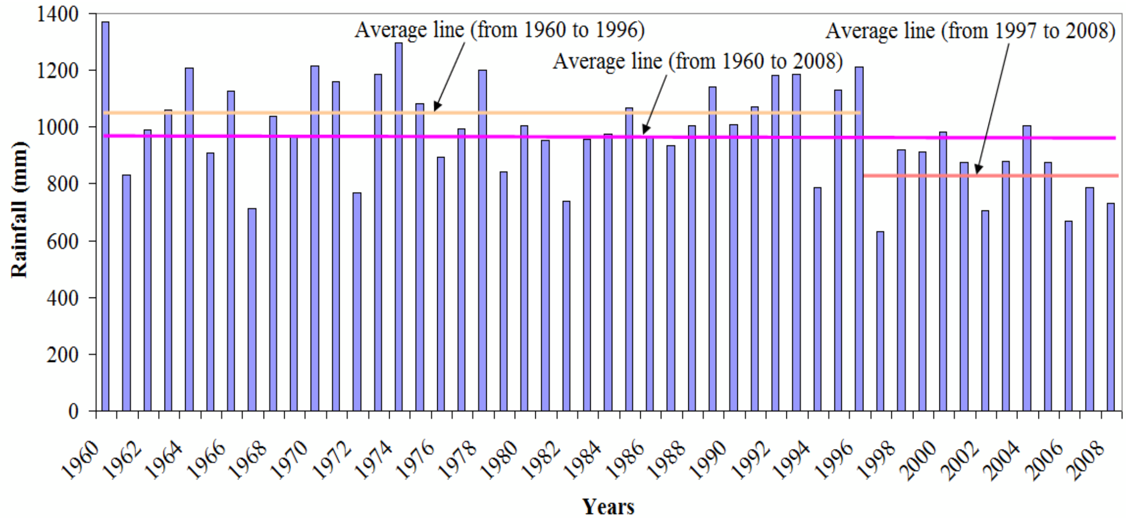

Figure 1-3. Annual average rainfalls in the Yarra River catchment (Barua, 2010)

The Yarra River catchment is characterized by three major land use types: forest, agricultural, and urban areas (Sokolov and Black, 1996). Approximately 21 percent of the catchment retains its natural vegetation, 57 percent is agricultural and 22 percent is urbanized (Victorian Government Department of Sustainability and Environment, 2006). Based on different land use patterns and natural sub-divisions, the Yarra River catchment is divided into three distinctive sub-catchments, namely Upper Yarra, Middle Yarra, and Lower Yarra segments (EPA Victoria, 1999; Barua et al., 2012). Each segment has its own distinct characteristics. The Upper Yarra segment of the catchment, beginning from the Great Dividing Range to the Warburton Gorge at Millgrove, consists of mainly dense and extensive forested area with minimum human settlement. Major tributaries of the upper segment flow through these forested and mountainous areas, which have been reserved for water supply purposes for more than 100 years (Barua et al., 2012). This segment is used as a closed water supply catchment for Melbourne and about 70 percent of Melbourne's drinking water supply comes from this pristine upper segment (Barua et al., 2012; Melbourne Water, 2013).

The Middle Yarra segment of the catchment (the study area as shown in Figure 1-1), from the Warburton Gorge to Warrandyte Gorge, consists of mainly rural floodplains and valleys with limited urban development. The river gradient decreases and valley widens as the river approaches downstream. There are several gorges in this area, particularly the Yering Gorge, which restrict flows in the river. The Middle Yarra segment is notable as the only part of the catchment with an extensive flood plain area. Most of the land in this segment is mainly used for agricultural activities (Gardner, 1994; Carty and Pierotti, 2010). The Lower Yarra segment of the catchment, located downstream of Warrandyte, mainly consists of the urbanized floodplain areas of Melbourne city. Large areas of this segment comprises hard surfaces such as paved roads, roofs, car parks, concrete channels, etc. (EPA Victoria, 1999). Most of the land along the rivers and creeks in the middle segment (where the study area is located) and the urbanized lower segment (where the Melbourne city is located) of the Yarra River catchment has been cleared for agricultural or urban development (Melbourne Water, 2013).

The extensive clearing of lands in the middle and lower segment often results in high flows in rivers during intense storms, causing frequent flash flooding within the study area (Middle Yarra River catchment) and the urbanized lower segment located immediate downstream of the study area (Melbourne Water, 2013). Furthermore, increasing rainfall patterns and the occurrence of extreme rainfall events due to the impact of potential climate change (as reported in Whetton et al., 1993; Yilmaz and Perera, 2014; Yilmaz et al., 2014) will result in excess amount of streamflows that may cause intense flash floods in the middle segment (the study area) and the urbanized lower segment of the catchment. This has a direct impact on the urbanized lower segment of the Yarra River catchment, making it vulnerable and risk-prone due to potential flash flood problems. For example, Australian rainfall in 2009–2010 was 13 percent greater than the long-term (1911-2010) average (Bureau of Meteorology, 2011). This eventually caused the 2010-2011 devastating floods that affect three eastern states of Australia including Victoria (where the study area is located), Queensland and New South Wales (van den Honert and McAneney, 2011; Victorian Government, 2011). The urbanized Lower segment of the catchment also depends on the water supply from the storage reservoirs (shown in Figure 2-1) mainly located in the middle and upper reaches of the catchment (Barua et al., 2012). The main intention of reservoir operation in Australia is to store as much water as possible for satisfying the water demand during shortage of streamflows caused by droughts while keeping provision for flood control during excess streamflows caused by floods (Melbourne Water, 2013).

The study area has three storage reservoirs, namely Maroondah, Silvan and Sugarloaf reservoirs (shown in Figure 1-2), which support urban water supply systems of Melbourne. However, several major droughts occurred over the Australian state of Victoria in the past (Keating, 1992; Tan and Rhodes, 2008), particularly the 1997-2009 Millennium Drought (Grant et al., 2013; Chiew et al., 2014) have caused the substantial reduction in rainfall, inflows and storage volumes of major water harvesting reservoirs of Victoria (Low et al., 2015). For example, in the period 1998-2007, the annual average rainfall in Victoria declined by approximately 13 percent from the long-term average (Victorian Government Department of Sustainability and Environment, 2008). As a result of this rainfall declination, the inflows to Melbourne's main water supply reservoirs located within the Yarra River catchment declined by about 38 percent from the long-term average of the period 1913-1996 (Wallis et al., 2009). This substantial reduction in rainfalls and inflows has affected the overall water availability within the catchment (Melbourne Water, 2013). Hence, accurate rainfall information and improved estimation of future streamflows particularly in the middle and upper segments of the catchment could be beneficial for optimal operation of reservoirs mainly in the drought periods, and effective flood control particularly in the urbanized lower segment of the Yarra River catchment. Therefore, the Middle Yarra River catchment (as shown in Figure 1-1) was selected as the case study area for this study.

1.4.2 Topography

A digital elevation model (DEM) of the study area was collected from Geoscience Australia to produce the topographic map of the study area. Figure 1-4 shows the topographic map of the study area. As can be seen from the figure, the study area is characterized by a relatively flat to mountainous topography. The elevation of the study area varies from 25 to 1243 m with a mean elevation of 621 m (above mean seal level) from west to east. The lower regions are found mainly across in the central, north-western and western part while the higher regions are located mainly in the northern, north-eastern and eastern part of the study area. The river gradient decreases and valley widens as the river approaches downstream. Surface relief of the catchment converges from the east, north and south towards the central portion of the catchment, which finally meets the outlet of the study area at west.

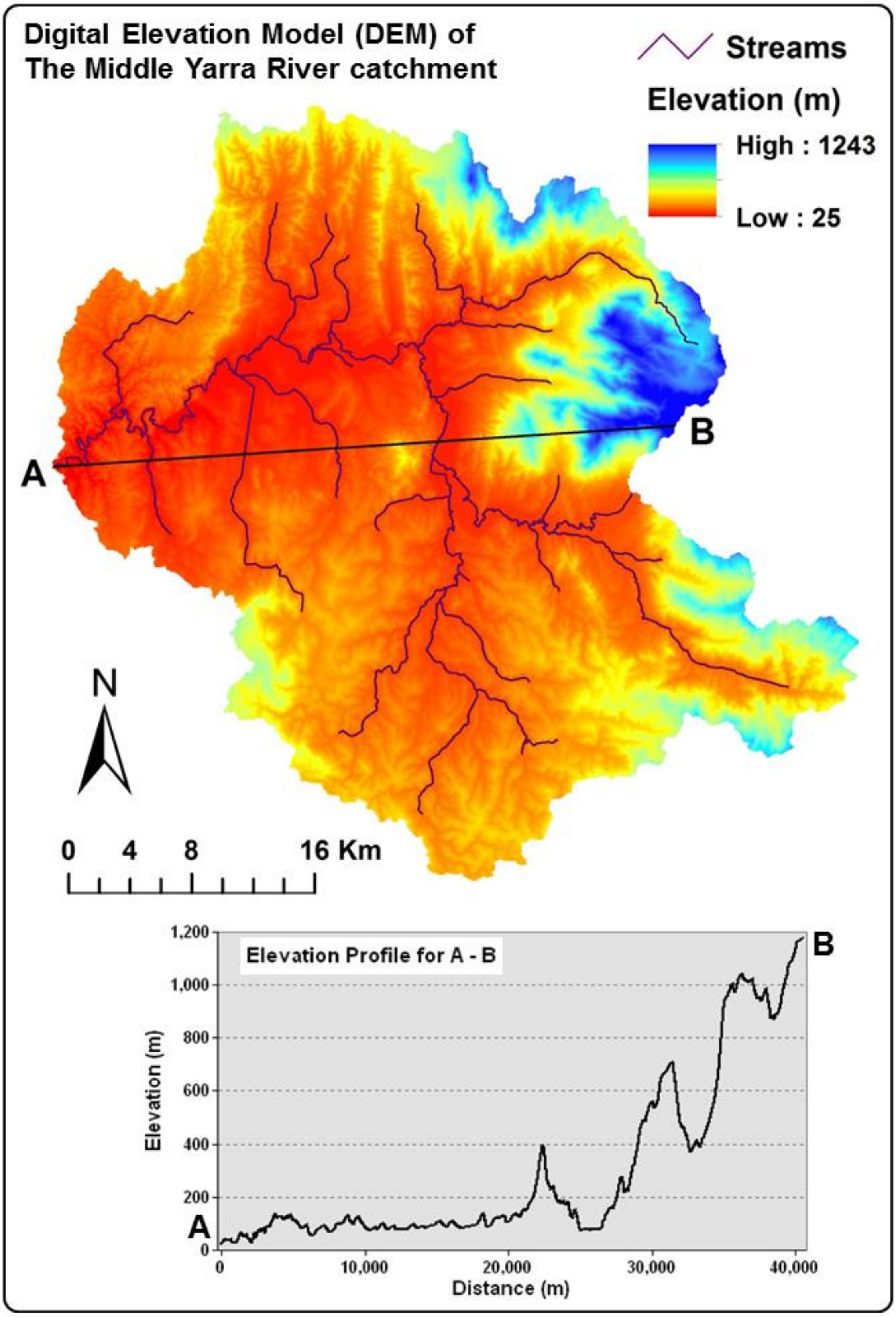

Figure 1-4. DEM of the study area with a cross-sectional elevation profile A-B

The figure also shows that the elevation difference is approximately 1165 m along the cross-sectional profile A-B extending about 40 km from the eastern side of the catchment to its outlet. Several studies have revealed that rainfall usually exhibits a high correlation with elevation and hence tends to increase with higher elevations due to the orographic effects of mountainous topography (e.g., Hevesi et al., 1992; Dirks et al., 1998; Goovaerts, 2000; Lloyd, 2005; Moral, 2010; Mair and Fares, 2011; Feki et al., 2012). Therefore, the elevation data extracted from the DEM of the study area was used as an auxiliary variable in the cokriging analysis for enhanced spatial interpolation of rainfall.

1.4.3 Land use and agriculture

The Australian Land Use and Management (ALUM) Classification system is used for the national (1 : 2,500,000) and catchment scale (1 : 25,000 to 1 : 1,000,000) land use mapping in Australia (ABARES, 2012). The ALUM Classification system provides a nationally consistent method to collect and present land use information for a wide range of users across Australia. There are six major classes of land use (each further being divided into two extra tiers) in the ALUM classification system. Percent distribution of the six major land use categories in the study area is shown in Figure 1-5. As can be seen from the figure, the dominant land use category in the study area is "Production from dryland agriculture and plantations" in which the main land use type is pasture productions for livestock.

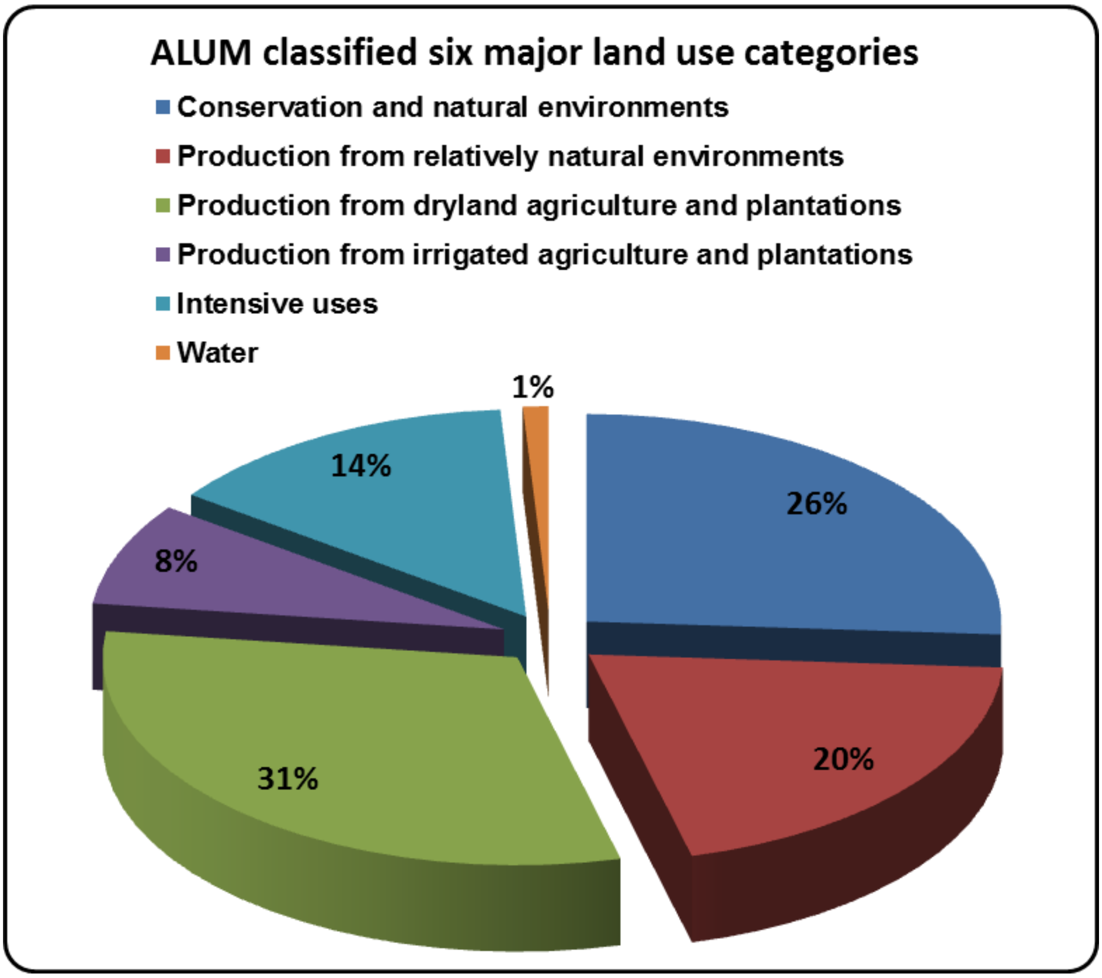

Figure 1-5. Distribution of ALUM classified six major land use types in the study area

The land use map for the study area was prepared from the catchment scale 50m grid raster data collected from the Australian Bureau of Agricultural and Resource Economics and Sciences (ABARES, 2012). Figure 1-6 shows details of different land use types in the study area. The map covers data for the period of 1997 to May 2011. The ALUM classified major land use categories for the study area (Figure 1-5) were re-classified to produce the final land use map of the study area showing each land use types with their coverage (Figure 1-6). As can be seen from Figure 1-6, the upper part of the study area is mainly mountainous forest, and the most downstream part is covered with developed rural and urban areas. The majority of the middle portion of the study area is covered with agricultural and cropping lands, dominated by pasture and covering about 32% of the total area. Cultivations of cereals and hays, potatoes and different fruits, pasture productions for livestock are considered as the major agricultural activities.

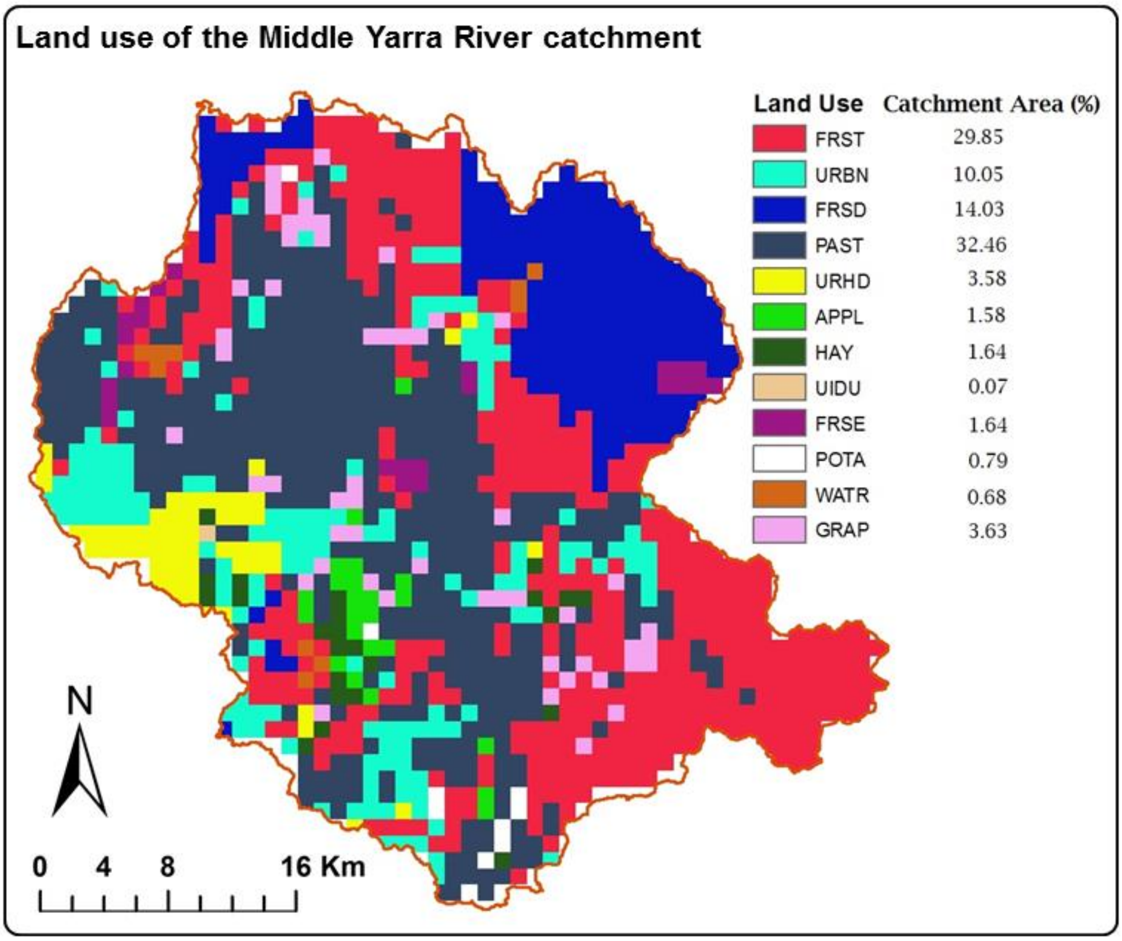

Figure 1-6. Land use map of the study area

1.4.4 Soil

In this research project, the soil type data were collected from the Australian Soil Resource Information System (ASRIS) at http://www.asris.csiro.au to produce a digital soil map of the study area. ASRIS is a product of the Australian Collaborative Land Evaluation Program (ACLEP), which is developed by Commonwealth Scientific and Industrial Research Organization (CSIRO) and Department of Agriculture, Fisheries and Forestry (DAFF) in collaboration with state and territory agencies. The soil map of the study area is shown in Figure 1-7. As can be seen from the figure, the main types of soils in the study area include: strong texture contrast soils (Sodosols, Chromosols), soils that lack strong texture contrast (Dermosols, Ferrosols), sandy soils of weakly developed texture (Tenosols) and clay-rich soils of uniform texture (Vertosols).

However, the dominant soil types among them are the Sodosols (covers approximately 54% of the study area) and the Dermosols (covers approximately 35% of the study area). It is important to note that the soil classification is shown in accordance with the Australian Soil Classification defined in Isbell (2002).

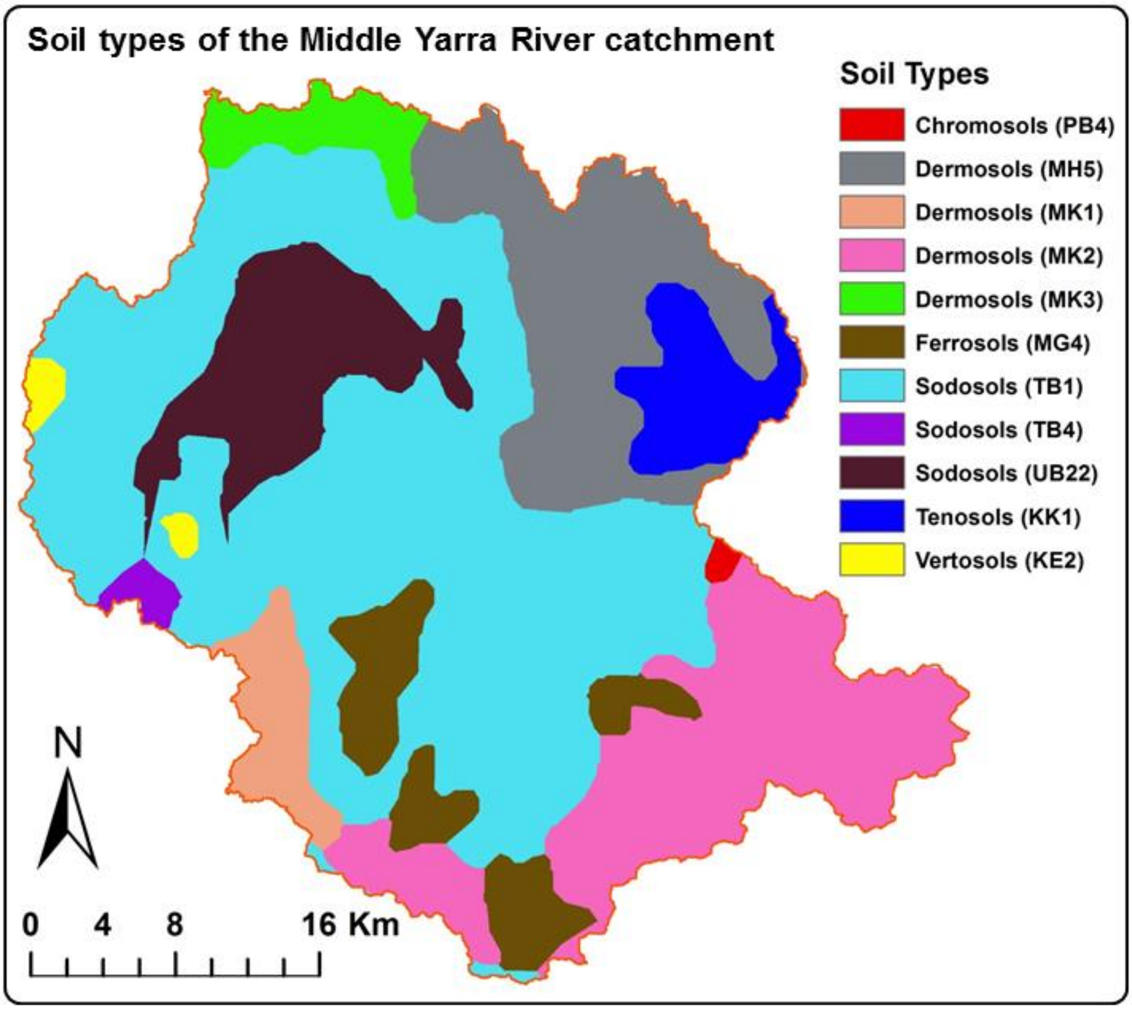

Figure 1-7. Soil map of the study area

1.4.5 Climate and streamflow

The climate over south-eastern Australia (where the case study area is located) is influenced by a number of factors including Northwest Cloudbands, Sub-tropical Ridge, Cut-off Lows, El Niño and La Niña events of El Niño Southern Oscillation (ENSO), and the Southern Annular Mode (SAM) (Bureau of Meteorology, 2016). Hence, the gradient of rainfall exists in the south-north direction over the study area.

In this research project, nineteen rain gauge stations located within the study area (indicated with R1 to R19 in Figure 1-1) were considered. Daily rainfall data for all nineteen stations were collected from the Scientific Information for Land Owners (SILO) climate database (http://www.longpaddock.qld.gov.au/silo/) and the Bureau of Meteorology (http://www.bom.gov.au/climate/data/) for the period of thirty-three years from 1980to 2012. The major advantage of the SILO climate database is that the SILO data are highly quality-controlled and totally free from missing records (Jeffrey et al., 2001). Therefore, the SILO rainfall data were used in this research project. The station numbers, names and corresponding BoM identification numbers and geographical coordinates of each of the rain gauges are given in Table 1-1 and their spatial locations are shown in Figure 1-1.

Table 1-1. Details of rain gauge stations considered in the study

Station Number	Station ID	Name of Station	Latitude (Degree)	Longitude (Degree)
R1	86142	Toolangi (Mount St Leonard DPI)	-37.57	145.50
R2	86366	Fernshaw	-37.61	145.60
R3	86009	Black Spur	-37.59	145.62
R4	86070	Maroondah Weir	-37.64	145.55
R5	86385	Healesville (Mount Yule)	-37.65	145.51
R6	86363	Tarrawarra	-37.66	145.48
R7	86364	Tarrawarra Monastery	-37.66	145.45
R8	86219	Coranderrk Badger Weir	-37.69	145.56
R9	86383	Coldstream	-37.72	145.41
R10	86229	Healesville (Valley View Farm)	-37.74	145.53
R11	86367	Seville	-37.80	145.49
R12	86358	Gladysdale (Little Feet Farm)	-37.86	145.65
R13	86094	Powelltown DNRE	-37.86	145.74
R14	86059	Kangaroo Ground	-37.68	145.25
R15	86066	Lilydale	-37.76	145.36
R16	86076	Montrose	-37.80	145.37
R17	86106	Silvan	-37.83	145.44
R18	86072	Monbulk (Spring Road)	-37.88	145.42
R19	86266	Ferny Creek	-37.87	145.35

Station numbers are same as in Figure 1-1

Station ID is as defined by the Bureau of Meteorology (BoM), Australia at http://www.bom.gov.au/climate/data/stations/

Figure 1-8 shows the annual catchment average rainfall for the Middle Yarra River catchment (the study area) based on the nineteen rain gauge stations for the period from 1980 to 2012. As can be seen from the figure, the annual catchment average rainfall was around 1081.6 mm when considering all years from 1980 to 2012. The annual catchment average rainfall in the aforementioned period varies from 710.4 mm to 1422.2 mm. The figure also shows that the annual catchment average rainfall from 1980 to 1996 was around 1164.2 mm, whereas it was only around 931.4 mm for the period from 1997 to 2009. It can be also seen that there was an abrupt drop in the annual catchment average rainfall by more than 200 mm within the case study area during the period from 1997 to 2009 when the Millennium Drought (Grant et al., 2013; Chiew et al., 2014) was occurred. This sudden change is found to be similar to the findings shown in Figure 1-3 for the entire Yarra River catchment. Furthermore, the figure shows that from 2010 onwards, the annual catchment average rainfall within the case study catchment started to grow up when there was a flood in Victoria (van den Honert and McAneney, 2011; Victorian Government, 2011). This change in rainfall has great implications on the water resources management of the study area.

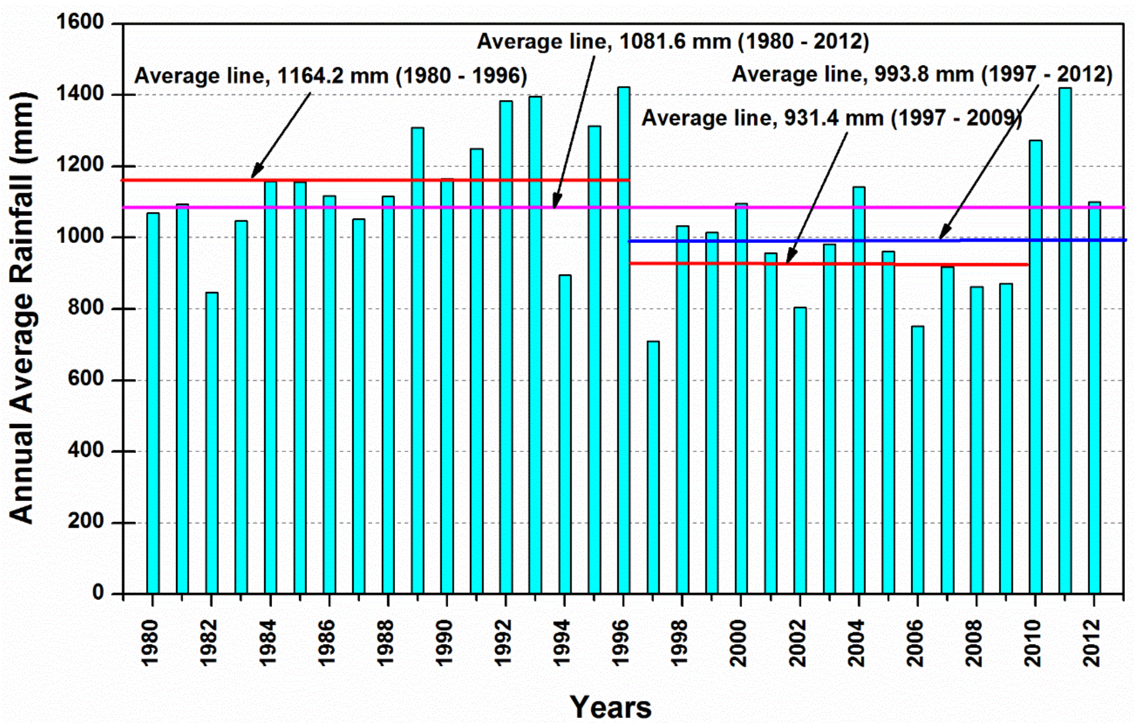

Figure 1-8. Annual average rainfall variations over the study area

The monthly average rainfall and temperature (max and min) in the study area is shown Figure 1-9. As can be seen from the figure, the September is the wettest month (rainfall amount equals to 112.5 mm) and have the highest rainfall variation. The figure also shows that the driest month is February (rainfall amount equals to 56.4 mm), which have the second highest rainfall variability. It was also found that the southern and south-eastern part of the study area experiences the highest rainfall whereas the lowest rainfall occurs in the north-western part. Approximately 60% of the rainfall occurs in the winter (June-August) and spring (September-November) seasons, which contributes mostly to streamflows. In general, the summer (December-February) is very dry compared to the winter (June-August) and spring (September-November). The average monthly maximum temperature varies from 11.4 ^{0}C (in July) to 25.3 ^{0}C (in February), and minimum temperature varies from 4.4 ^{0}C (in July) to 12.3 ^{0}C (in February).

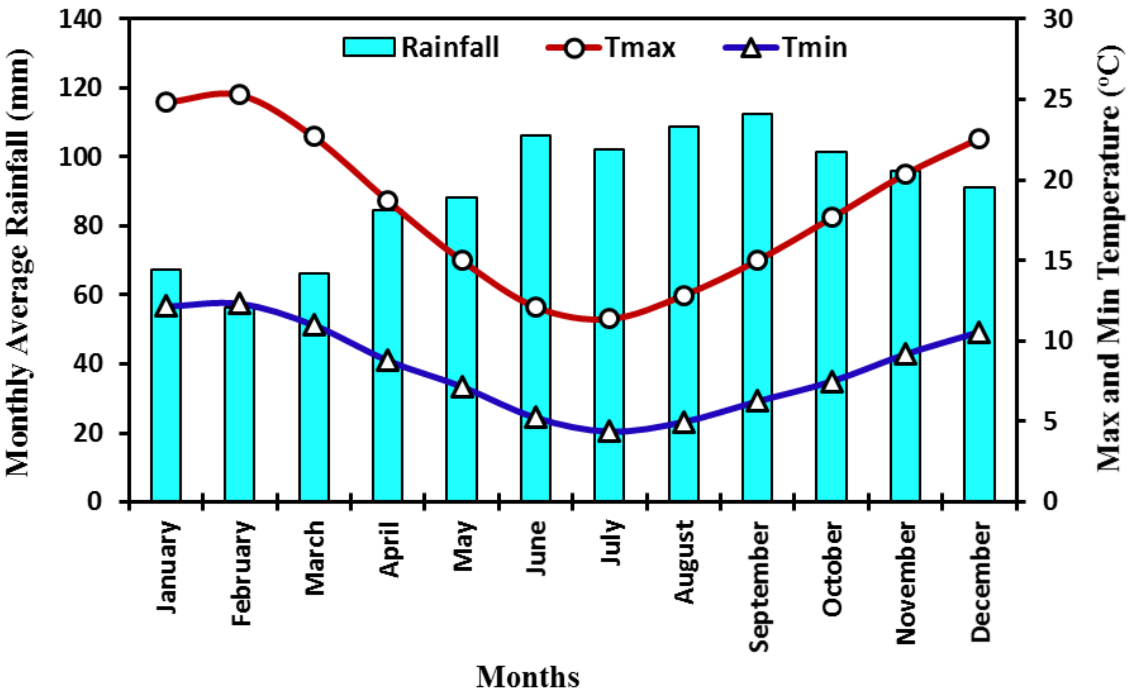

Figure 1-9. Monthly average rainfall and temperature variations in the study area

There are six streamflow measuring stations along the main course of the Yarra River within the study area. Four streamflow measuring stations among them (indicated with S1 to S4 in Figure 1-1) were used in this research project. In the Australian state of Victoria, Melbourne Water Corporation (MWC) provides the access the continuous streamflow data. Daily streamflow records for these four stations were thus collected from MWC database for a period of thirty years from 1980 to 2009. The choice of this study period is based on the availability of high quality of data with no missing records for an extended period. The station numbers, names and corresponding MWC identification numbers and geographical coordinates of each of the streamflow gauges are given in Table 1-2 and their spatial locations are also shown in Figure 1-1.

Table 1-2. Description of streamflow measuring stations considered in the study

Station Number	**Station ID**	**Name of Station**	**Latitude (Degree)**	**Longitude (Degree)**
S1	229212	Yarra River at Millgrove	-37.75	145.66
S2	229653	Yarra River at Yarra Grange	-37.67	145.48
S3	229608	Watsons Creek at Kangaroo Ground South	-37.70	145.26
S4	229200	Yarra River at Warrandyte	-37.74	145.22

Station numbers are same as in Figure 1-1

Station ID is as defined by the Melbourne Water Corporation, Australia

1.5 Research Significance

This research project has produced several significant contributions in the field of water resources management, especially for the effective management of water resources in Australian conditions. These contributions are outlined below.

In the optimal rain gauge network design using the kriging-based geostatistical approach, kriging is often used to estimate rainfall values along with the kriging variance in ungauged locations based on the observed rainfall values at neighboring gauged locations. In traditional kriging, accurate kriging results highly depend on the use of an appropriate variogram model, which represents spatial correlations among data points and plays a vital role in the kriging process. Therefore, the robustness of kriging interpolation heavily depends on the fitting and estimation of a correct variogram model. Furthermore, selection of an appropriate variogram model, finding the optimal variogram parameters (i.e., nugget, sill and range) and the computational burdens involved are some difficulties associated with the traditional kriging. As a potential solution to these issues, a new universal function approximation-based kriging was explored in this study where genetic programming (GP) was used as a universal function approximator to derive the variogram model. This new variant of kriging is referred to as the genetic programming-based ordinary kriging (GPOK) in which the standard parametric variogram models (i.e., exponential, gaussian, spherical models) in traditional ordinary kriging were replaced by the GP-derived variogram model. The GPOK demonstrated in this study overcomes the limitations associated with the traditional ordinary kriging and gives improved estimation of rainfalls at ungauged locations. According to the best knowledge of the author, there is no record in the published literature on the variogram modelling and developing a kriging model using GP prior to this study. Variogram modelling using GP offers several advantages. For example, GP does not require a pre-defined mathematical form or architecture unlike ANN to generate the functional variogram models. In addition, GP can generate variogram models, which consists of similar mathematical structure as having with the standard variogram models. Furthermore, GP-derived variogram model does not require identifying the variogram parameters in advance, unlike the standard parametric variogram models in traditional kriging. Therefore, the GPOK is completely free from the trial and error process of estimating the variogram parameters. The function approximation capability of GP produces the best fitted GP-derived variogram models compared to the standard models, which were found to give the best functional variogram models. Thus, the GP-derived variogram models offer a viable alternative to the existing standard variogram models from which to choose the best variogram model for kriging interpolation of rainfall.

Furthermore, a key advantage of kriging over deterministic and other conventional interpolation methods is that while providing the kriging variance for rain gauge network design, it is capable of complementing the sparsely sampled primary variable, such as rainfall by the correlated densely sampled secondary variable, such as elevation to improve the estimation accuracy of primary variable. This multivariate extension of kriging is referred to as the cokriging method, which was investigated in this study for enhanced spatial interpolation of rainfall. Based on the rainfall-topography relationship of the case study catchment, the elevation was used as an auxiliary variable in addition to rainfall in the cokriging methods. However, it is often challenging to select the best interpolation method arbitrarily from a wide variety of available kriging and deterministic interpolation methods for estimating the spatial distribution of rainfall for a particular area because the performance of an interpolation method depends on many factors. As a solution to this issue, a comparative evaluation of a range of univariate and multivariate kriging and deterministic interpolation methods was performed in this study to identify the most appropriate interpolator for enhanced spatial interpolation of rainfall and generation of continuous rainfall maps. Ordinary kriging yields the best results among all univariate interpolation methods. Ordinary kriging combined with genetic programming overcomes some of the limitations of ordinary kriging and thus offers several advantages in spatial mapping of rainfall, which produces similar results as given by ordinary kriging. The ordinary cokriging using elevation information outperforms all other methods when comparing all univariate and multivariate methods. The ordinary cokriging was thus found to be the most suitable interpolator for enhanced spatial distribution of rainfall in the Middle Yarra River catchment. The major advantage of this method is that it allows the use of correlated auxiliary information to improve the accuracy of kriging interpolation of rainfall in a catchment with the mountainous and/or complex terrain.

Conventionally, increasing network density through network augmentation (using additional rain gauge stations) to reduce the kriging variance of the network is used for the design of a rain gauge network. However, it is not very uncommon for a rain gauge network that it may consist of redundant stations, which have little or no contribution to the network performance for providing quality rainfall estimates. Since operation and maintenance of a rain gauge station involves large costs, removal of redundant stations can contribute to substantial cost reductions. Therefore, optimal rain gauge network design based on the variance reduction principle using the kriging-based geostatistical approach, it is important to consider both additional and redundant stations simultaneously in order to achieve the variance reduction and cost-effectiveness objectives of the optimal network. As a solution to these issues, a simple and effective rain gauge network design technique considering both additional and redundant stations under the variance reduction framework was investigated in this study. The technique involves a methodical search for the optimal number and locations of rain gauge stations in the network that minimize the kriging variance of areal and/or point rainfall estimates over the catchment. The technique allows achieving the optimal rain gauge network through optimal positioning of additional stations (network augmentation) as well as removing and/or optimally relocating of existing redundant stations (network rationalization). Furthermore, the spatial variability of rainfall in the case study catchment is usually caused by the El Niño and La Niña processes of ENSO phenomenon. As a solution to this issue, the rain gauge network was designed separately based on rainfall records for both the El Niño and La Niña periods and the network that gave the enhanced estimates of areal average and point rainfalls was selected as the optimal network. It was found that the optimal network achieved in this way gives the improved spatiotemporal estimates of areal average and point rainfalls when comparing with the existing rain gauge network for the case study catchment.

Rainfall is often considered independent of streamflow simulation and forecasting in many hydrological studies such as the design of a rain gauge network for estimation of areal average rainfall over a catchment or region. But this does not allow one to focus on the strength and weakness of an established optimal rain gauge network that really matter when rainfall data from the optimal network are fed into the streamflow forecasting models. Since streamflow is a consequence of rainfall, correct rainfall input is essential for accurate and enhanced streamflow forecasting. However, the majority of the streamflow forecasting techniques in the past studies used rainfall data directly from the existing rain gauge network which might exhibit high variance and hence not be the optimal network. This ultimately adversely affects the performance of streamflow forecasting models and results in less accurate streamflow forecasts. As a solution to this issue, an ANN-based enhanced streamflow forecasting approach was explored in this study, which incorporated the optimal rain gauge network-based input instead of using existing rain gauge network-based input in order to achieve the enhanced streamflow forecasting. According to the best knowledge of the author, there is no single research study in the published literature at present, which has employed the optimal rain gauge network-based input to improve the accuracy in streamflow forecasts. The approach was found to be highly effective to achieve better accuracy in streamflow forecasts and thus could be a viable option for enhanced streamflow forecasting. The ANN-based input selection technique demonstrated in this thesis offers a high potential for significant input variables selection in data-driven modelling in hydrology. Furthermore, the ANN-based input selection technique offers an indirect way of identifying the optimal locations of rain gauge stations in the final optimal rain gauge network through the input significance of each rain gauge station, which produces the best streamflow forecasting performance of the network. Also, this study is the first of this kind in Australia to incorporate the input from an optimal rain gauge network within the ANN-based data-driven framework for improved forecasting of catchment streamflows, and is perhaps one of the very few studies across the world.

As was stated in Section 1.4.1, the occurrence of extreme rainfalls in the Yarra River catchment of Victoria has caused the urban areas (mostly located in the lower segment of the catchment) highly vulnerable to floods. On the other hand, the occurrence of droughts (e.g., the 1997-2009 Millennium Drought) has led to a substantial reduction in rainfall and inflows of major water harvesting reservoirs of Victoria including Melbourne's main water supply reservoirs located within the Yarra River catchment. The ANN-based enhanced streamflow forecasting approach demonstrated in this study can assist water managers for efficient planning and operation of the risk-based water resources systems to mitigate the impacts and risks caused by the occurrence of floods and droughts. Accurate rainfall information and improved estimation of future streamflows achieved through the approach especially in the middle segment (the case study area) of the Yarra River catchment can be supportive in the optimal operation of the water supply reservoirs for drought management as well as in the effective flood control planning in the urbanized lower segment of the catchment.

1.6 Outline of the Thesis

This thesis is prepared based on journal articles and conference papers published, submitted or in under review (refer to the list of publications). The outline of this thesis is presented through an interconnected flow chart as shown in Figure 1-10. As can be seen from the figure, this thesis consists of seven chapters. Each chapter of the thesis from Chapter 3 to Chapter 6 includes one or more publications as described in the following.

The first chapter addresses the subject of this thesis including the detailed background of the research project, the research aims and a brief methodology to achieve the stipulated aims. The significance of the research project (i.e., integrating the optimal rain gauge network-based input for enhanced streamflow forecasting) and the description of the study area are also presented in this chapter.

The second chapter includes the literature review conducted at the initiation of this research project and provides a general overview of the rain gauge network design methods and impact of spatial rainfall variability and rain gauge network density on streamflow forecasting. Thus, the first two chapters cover the research design for this project. First, a detailed literature survey on the optimal rain gauge network design and evaluation using the kriging-based geostatistical approach is presented in this chapter. Then, the chapter also includes a review on the impact of spatial rainfall variability and rain gauge network density on the streamflow simulation and/or forecasting. It is worth mentioning that although Chapter 2 covers the majority of literature review of this research project, parts of the literature review are also distributed through Chapters 3 to 6 in which it was particularly required to demonstrate the proposed methodology and discuss the application through a case study.

The third chapter details the development of a new universal function approximation-based kriging method using GP and its potential application for the improved estimation of rainfall. The proposed variant of kriging was referred to as GPOK in this thesis where the standard parametric variogram models (i.e., exponential, gaussian, spherical models) in traditional ordinary kriging were replaced by the GP-derived variogram model. GP was used as a universal function approximator in this study in order to overcome some of the limitations associated with the existing standard variogram as well as other universal function approximator such as ANN-derived variogram models. This chapter also shows that GP-based variogram modelling offers several advantages. For example, GP does not require a pre-defined mathematical form or architecture unlike ANN to generate the functional variogram model. In addition, GP-derived variogram model does not require identifying the variogram parameters in advance, unlike the standard variogram models. As a result, the proposed GPOK method was completely free from the trial and error process of estimating the variogram parameters. Finally, the robustness of the proposed GPOK method over the traditional and ANN-based ordinary kriging methods for improved estimation of rainfall is detailed in this chapter.

The fourth chapter presents the performance evaluation of different univariate and multivariate kriging interpolation methods to identify the best interpolator for enhanced spatial interpolation of rainfall and subsequent production of high quality continuous rainfall datasets in the form of rainfall maps. Based on the rainfall-topography relationship of the case study catchment, the elevation was used as an auxiliary variable in addition to rainfall (i.e., primary variable) for the multivariate or cokriging methods (ordinary cokriging and kriging with an external drift) in this research project. This chapter is divided into two parts. The first part of this chapter addresses the performance comparison of different univariate kriging interpolation methods including four traditional and a universal function approximation-based method using GP for spatial interpolation of rainfall. The second part of this chapter details the performance comparison of five traditional univariate kriging interpolation methods and two cokriging interpolation methods for enhanced spatial interpolation of rainfall. The second part of this chapter also includes the analysis for an additional case study area (Ovens River catchment in Victoria). Findings of the comparative performance of different univariate and multivariate kriging methods are also explained in this chapter.

In chapter five, development of a simple and effective rain gauge network design methodology in order to achieve an optimal rain gauge network under the variance reduction framework using the kriging-based geostatistical approach is presented. Network expansion with additional stations (network augmentation) for variance reduction has been the underlying criterion to achieve the optimal network in most of the past studies. However, an existing rain gauge network may consist of redundant stations, which have little or no contribution to the network performance for providing quality rainfall estimates. As a solution to this issue, both additional and redundant stations were considered to achieve the optimal rain gauge network design in this research project. The optimal rain gauge network for the case study catchment was thus achieved through optimal positioning of additional rain gauge stations (network augmentation) in addition to removing and/or optimally relocating of existing redundant rain gauge stations (network rationalization), which is detailed in this chapter. In this study, the spatial variability of rainfall in the case study catchment caused by the El Niño and La Niña processes of ENSO phenomenon was taken into consideration and thus the rain gauge network was designed independently based on rainfall records obtained for both El Niño and La Niña periods. The rain gauge network that gave the best estimates of areal average and point rainfalls was selected as the optimal rain gauge network for the case study catchment. Findings related to the optimal rain gauge network design and comparative estimates of rainfall from the existing and optimal network are finally explained in this chapter.

In chapter six, an enhanced streamflow forecasting approach to achieve the improved accuracy in streamflow forecasting using the optimal rain gauge network-based input to ANN modelling framework is presented. In this chapter, the effectiveness of incorporating input from an optimally designed rain gauge network (instead of using input from an existing non-optimal rain gauge network) in the ANN-based streamflow forecasting models to improve streamflow forecasts is explored. The robustness of the proposed approach was tested using different quantitative performance evaluation measures and the corresponding findings, which are explained in this chapter. The potential of using the ANN-based input selection approach in significant input variables selection for developing the ANN-based streamflow forecasting models is also detailed in this chapter. Based on the ANN-based input variable selection approach, the chapter finally presents an indirect way of identifying the optimal locations of rain gauge stations in the final optimal rain gauge network in order to achieve the best streamflow forecasting performance of the network.

Finally, a summary of the thesis, the conclusions drawn from the study and some recommendations for the future work are presented in chapter seven. Figure 1-10 shown in the following depicts the interconnection between the works described in the journal papers (refer to the list of publications indicated earlier) included in this thesis.

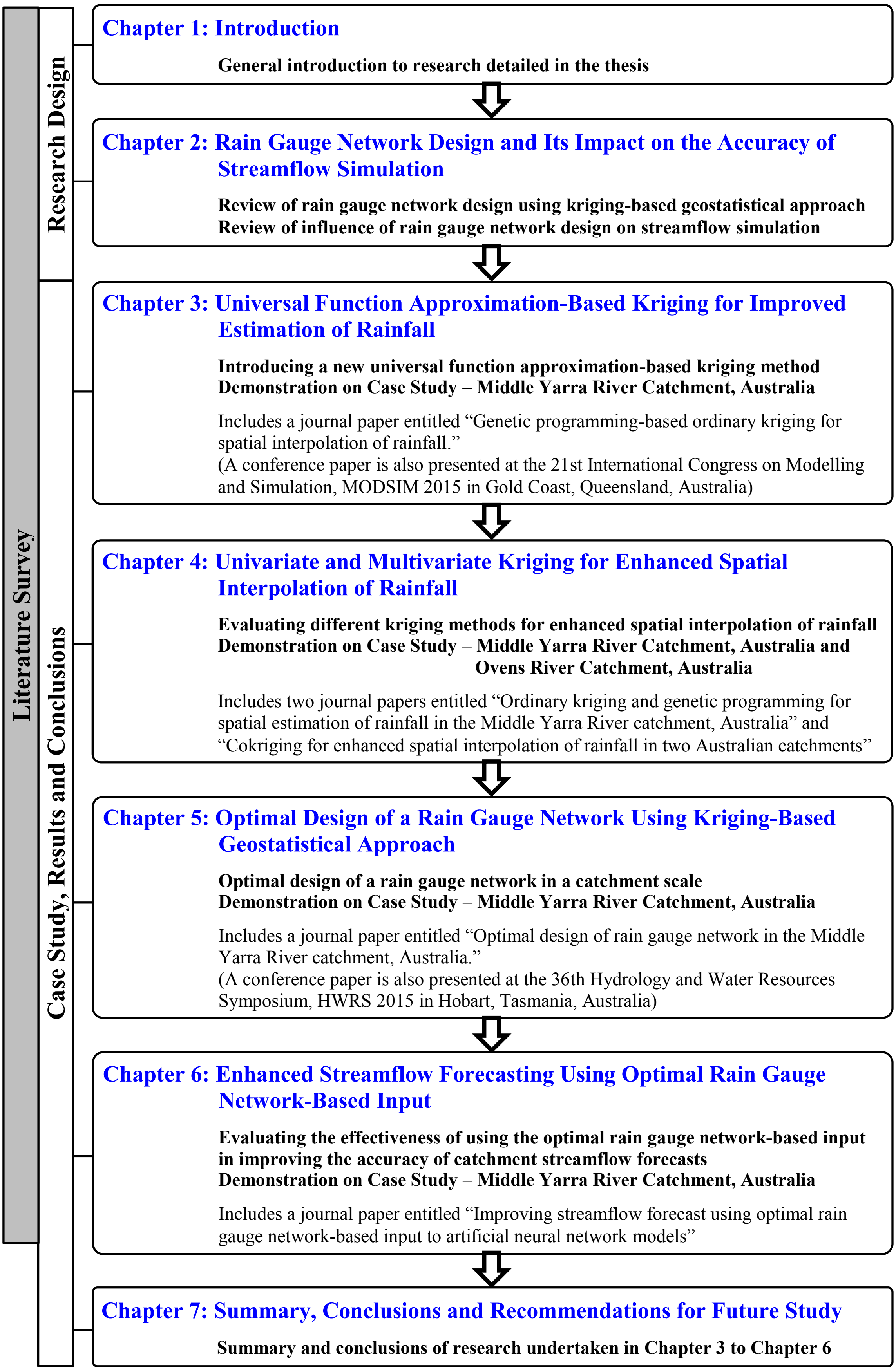

Figure 1-10. Interconnection between works described in the thesis

Chapter 2

Rain Gauge Network Design and Its Impact on the Accuracy of Streamflow Simulation

2.1 Introduction

One of the most important elements of the effective planning and management of water resource is the assessment of this invaluable resource, which takes into account the identification of the sources and the evaluation of their capacity, reliability and quality, implying the measurement and collection of data of interest (Mishra and Coulibaly, 2009). For this purpose, monitoring networks are designed and sometimes optimized for decision making according to the water management objectives (Loucks et al., 2005). The World Meteorological Organization (WMO) stated in 1981 that "the objective of a monitoring network is to ensure a density and distribution of stations in a region such that, by spatial interpolation between datasets at different stations, it will be possible to determine the characteristics of the basic hydrological and meteorological elements anywhere in the region with sufficient accuracy" (van der Made 1988, p.20).

A vital part of modern water management covers the measurement of different processes of the hydrologic cycle. Monitoring networks are established to acquire measured data of different hydrologic variables with useful information content. A monitoring network can be defined as a set of strategically located measurement devices that collect data of interest about a water system at a given temporal scale. The measured data of hydrologic variables from the monitoring network are used by water managers and decision-makers to manage water resources systems in an efficient manner or to maintain a good performance of water resource systems. Therefore, monitoring networks are considered as one of the important components of any hydrological study because they collect data that, after being interpreted, provide insights for decision-making. This thesis concentrates on rainfall monitoring network (or rain gauge network), which is commonly used for measurement and monitoring of rainfall data over a catchment or watershed.

Rainfall is one of the most sought-after variables of the hydrological processes, which is used as a fundamental input for most hydrological modelling and analyses of water resources systems. However, many of the water resources systems are large in spatial extent and often consist of a rain gauge network that is very sparse due to large cost involvement, logistics, and geological factors. This results in considerable uncertainty in the collected rainfall data from these sparse rain gauge networks (Zealand et al., 1999; Goovaerts, 2000). These problems have created the need for establishing an optimal rain gauge network that can provide high quality rainfall data for effective hydrological analysis and design of water projects. An optimal rain gauge network can be defined as a balanced network that neither suffers from lack of rain gauge stations nor is over-saturated with redundant rain gauge stations (Mishra and Coulibaly, 2009; Shaghaghian and Abedini, 2013). Hence, an optimal rain gauge network is often considered an indispensable component of any hydrological study. Therefore, a review of the existing optimal rain gauge network design approaches is vital to come up with a suitable approach for optimal design of rain gauge network in this study.

As was mentioned in Section 1.1, the accuracy of streamflow simulation and forecasting primarily depends on the input data, wherein the largest impact on estimated streamflow is caused by the catchment rainfall (Faurès et al., 1995; Maskey et al., 2004; Jones et al., 2006; Ekström and Jones, 2009). Since streamflow is a consequence of rainfall, the uncertainty associated with rainfall leads to the uncertainty in estimated streamflow and can adversely affect the accuracy of streamflow simulation and forecasting (Faurès et al., 1995; Tsintikidis et al., 2002; Moulin et al., 2009). There are several methods that have been used in the past to assess the impact of rain gauge network density on streamflow simulation and forecasting. Understanding these techniques will be beneficial in developing a suitable streamflow forecasting approach for enhanced streamflow forecasting.

There are two aims of this chapter: (i) to review the existing optimal rain gauge network design approaches that have been used for optimal design of rain gauge network, and (ii) to review the existing methods that have been used to evaluate the impact of rain gauge network design on streamflow simulation and forecasting. The outcome of this chapter will be the identification of the most suitable optimal rain gauge network design technique and development of an enhanced streamflow forecasting approach incorporating the optimal network for use in this research project.

The chapter first starts with the review of the existing optimal rain gauge network design approaches, followed by the review of the existing methods that explores different aspects of the impact of rain gauge network design on streamflow simulation and forecasting.

2.2 Overview of Rain Gauge Network Design Methods

There is a long history of developing and establishing rain gauge networks around the world. Due to imbalanced regional development and less effective design during the early stages, the installation of new rain gauge stations is often required to satisfy the growing demands for rainfall monitoring in certain areas of a catchment or watershed (Wang et al., 2015). However, deploying infinite number of additional rain gauge stations in a network is unrealistic in real-world applications considering the complex environmental constraints and limited resources. As a result of these difficulties, optimal design of a rain gauge network is essential in order to achieve the highest rainfall monitoring efficiency for improving the resource utilization. The optimal configuration problem of rain gauge stations can be considered a facility location problem (Wang et al., 2015). In general, the aim of the rain gauge network design is to deploy a certain number of rain gauge stations at specific locations in the network to fulfill the water management objectives of a catchment or watershed.

The design and evaluation of rain gauge networks is associated with a number of key challenges, which range from establishing proper temporal and spatial scales to defining their scope at minimum costs. In theory, hydrologists usually take into account these challenges when developing new approaches for rain gauge network design. However, in practice, it has been reported that the rainfall data collected by the existing rain gauge networks remains, in general, inadequate for understanding and explaining the dynamics of natural water resources systems (Canadian Water Resources et al., 1994; IUCN, 1980). This may be because the criteria to establish the final rain gauge network are driven in practice by non-scientific aspects, such as political and social viewpoints.

Existing rain gauge networks are often evaluated to verify that the objectives for which the network was designed are fulfilled. The outcome of this evaluation comprises the reconfiguration of the existing rain gauge network, which may include a redefinition of the size and scope of the network. This can lead either to the inclusion of additional rain gauge stations through network augmentation process in places where rainfall data cannot be adequately inferred from the existing network or the elimination (due to redundancy or uselessness of the collected data) of the redundant rain gauge stations through network rationalization process (St-Hilaire et al., 2003; Mishra and Coulibaly, 2009). In general, the same methods used for the design of monitoring networks are used for their evaluation.

The design of hydrometric networks is a classical problem in hydrometeorology (Mishra and Coulibaly, 2009), which has received significant attention from the researchers for many years. Since myriad concerns are associated with hydrometric network design (Li et al, 2012), many approaches have been developed for that purpose around the world. A number of available network design and evaluation methods can be found in Mishra and Coulibaly (2009). These approaches can be broadly classified as statistical methods, entropy methods, optimization methods, basin physiographic characteristics and sampling strategies (as schematized in Figure 2-1), a comprehensive review of which is presented in the work by Mishra and Coulibaly (2009).

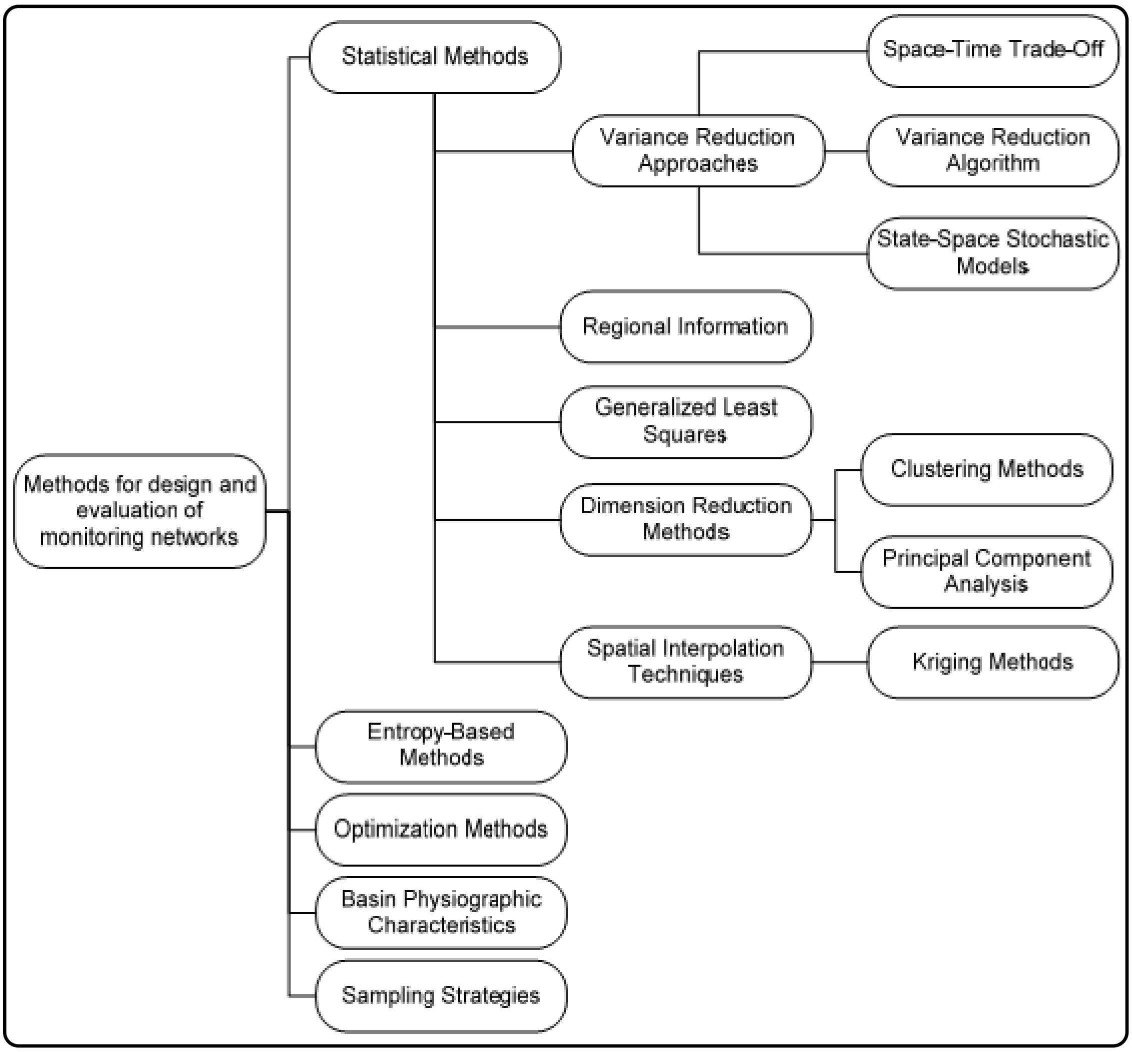

Figure 2-1. Classification of methods for hydrometric networks design and evaluation

Mishra and Coulibaly (2009) indicated that statistical methods are the most developed. Furthermore, it is found based on the available literature that among all the statistical methods, the kriging-based geostatistical method is the most commonly used method for the design and evaluation of rain gauge networks. An important advantage of this method is that it can be implemented with the combination of other methods such as entropy method, multivariate statistical techniques, and different multi-objective optimization techniques using genetic algorithms, simulated annealing, and particle swarm optimization in order to achieve a balanced or optimal rain gauge network. It is also found that the entropy method is successfully implemented for rain gauge network design and optimization in several studies. Therefore, a detailed review of these two basic methods for the design of a rain gauge network is presented in this chapter.

2.3 Kriging-Based Geostatistical Approach

2.3.1 Fundamental aspects of geostatistics

The term 'geostatistics' designates the statistical study of natural phenomena, which can be generally characterized by a distribution of one or more variables in space. These variables are referred to as regionalized variables (Journel and Huijbregts, 1978; Webster and Oliver, 2007). The unique feature of a regionalized variable is that it can take values according to its spatial location (Chebbi et al., 2011). Geostatistics is thus based on the theory of regionalized variables, which allows modelling of the spatial variability of the variable based on the spatial dependence between neighboring observations. The degree of spatial dependence is generally expressed by variogram (also called semivariogram) in geostatistics, which has the structural (spatial variability) information required on a regionalized variable. A variogram is a mathematical function of the distance and direction separating two locations used to quantify the spatial autocorrelation in regionalized variables (Webster and Oliver, 2007). The variogram has a key role in geostatistics, which was first applied and developed through kriging.

Kriging refers to a family of generalized least square regression methods in geostatistics (Isaaks and Srivastava, 1989; Goovaerts, 1997) that estimate values at unsampled locations using the sampled observations in a specified search neighborhood. As it will be seen, kriging-based geostatistical method for the design of rain gauge networks depends on the correct estimation of variogram models in which the design criteria are often related to the accuracy of the spatial estimation (i.e., the kriging standard error or the kriging variance) (e.g., Isaaks and Srivastava, 1989; ASCE Task Committee, 1990a,b). Therefore, prior to any kriging-based geostatistical assessment, the variogram must be computed from the regionalized variable data.

Initially, an experimental variogram $\hat{\gamma}(d)$ is computed based on the regionalized variable data using the Equation (2-1) as given in the following:

$$\hat{\gamma}(d) = \frac{1}{2N(d)} \sum_{i=1}^{N(d)} [Z(s_i + d) - Z(s_i)]^2 \qquad (2-1)$$

where $Z(s_i)$ and $Z(s_i + d)$ are the variable values at corresponding sampling locations s_i and $(s_i + d)$, respectively located at d distance apart and $N(d)$ is the number of data pairs (Isaaks and Srivastava, 1989; Webster and Oliver, 2007).

A variogram cloud is generated first using Equation (2-1) for observations at any two data points, in which all semivariance values are plotted against their separation distance. The experimental variogram is computed from the variogram cloud by sub-dividing it into a number of lags and taking an average of each lag interval (Johnston et al., 2001; Robertson, 2008). The average distances in each class interval are plotted against the values obtained from Equation (2-1), which is called the experimental variogram model (Oliver and Webster, 2014). The experimental variogram is thus composed of variogram $\hat{\gamma}(d)$ values for a finite set of discrete lags (Wackernagel, 2003). A lag is a line vector that separates any two data points (Johnston et al., 2001). A typical variogram cloud based on Equation (2-1) and a typical experimental variogram with a typical fitted model are shown in Figure 2-2.

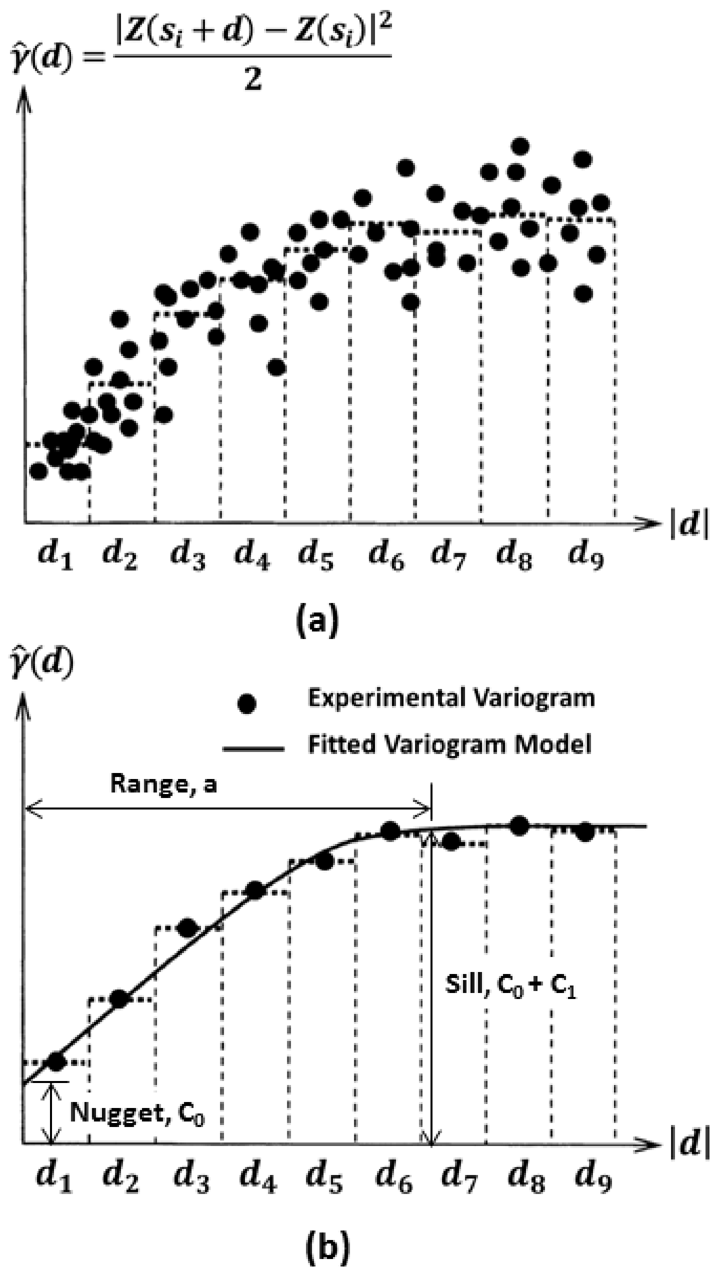

Figure 2-2. (a) A typical variogram cloud for a finite set of discrete lags, and (b) a typical experimental variogram based on the variogram cloud fitted by a typical variogram model with its parameters.

The experimental model data can be fitted to several analytical variogram $\gamma(d)$ models, which is a basic requirement in kriging-based geostatistics. Several standard variogram models are used for this purpose depending on the shape of the experimental variogram, which include exponential, gaussian, spherical, circular, linear, K-Bessel, J-Bessel, rational quadratic, stable and hole effect models. The shape and corresponding functional form of these standard variogram models can be found in Davis (1973), Goovaerts, 1997; Johnston et al. (2001), Wackernagel (2003) and Webster and Oliver (2007). However, exponential, gaussian and spherical variogram models are mostly used in hydrology (Teegavarapu, 2007; Abo-Monasar and Al-Zahrani, 2014). The functional forms of these commonly used variogram models are given in Table 2-1.

Table 2-1. Commonly used standard variogram models

Model name	**Functional form of the variogram model**
Exponential	$\gamma(d) = C_0 + C_1\left[1 - \exp\left(-\frac{3d}{a}\right)\right]$
Gaussian	$\gamma(d) = C_0 + C_1\left[1 - \exp\left(-\frac{3d^2}{a^2}\right)\right]$
Spherical	$\gamma(d) = C_0 + C_1\left[\frac{3}{2}\left(\frac{d}{a}\right) - \frac{1}{2}\left(\frac{d^3}{a^3}\right)\right], \quad d < a$
	$= C_0 + C_1, \quad d \geq a$

Note: C_0 = nugget coefficient, $C_0 + C_1$ = Sill, a = range, d = average lag distance

The nugget coefficient (C_0), sill ($C_0 + C_1$) and range (a) of the variogram models (Table 2-1 and Figure 2-2) are commonly referred to the variogram parameters. A variogram model is essentially described by these parameters that affect computation in the kriging process. 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 regionalized variables beyond the range (Johnston et al., 2001).

Spatial dependence, given the fact that the neighboring points tend to have similar characteristics, is expressed mathematically by variogram models. As an example,

Figure 2-3 shows schematization of the exponential and spherical variogram models. As can be seen from the figure, for both spherical and exponential curves, the variogram model steadily increases at a maximum value and then approaches a constant value, which is called the sill (C_0 + C_1) parameter. Furthermore, the distance corresponding to the variogram sill can be distinguished from the figure, which is called the range (*a*) of the variogram model. In the estimation of regionalized variables by variogram analyses, range divides the samples into two categories: (i) all the samples whose distances to the point to be estimated are less than or equal to range provide information about the point (dependent), and (ii) all the samples outside the neighborhood defined by the range are independent observations with respect to the point to be estimated and may be disregarded, because they do not provide any information about the point (Isaaks and Srivastava, 1989). Therefore, it can be shown from the spherical variogram model that $\gamma(d) = (C_0 + C_1)$ when $d = a$, meaning that the sill is reached at a distance equal to the variogram range. In contrast, the exponential semivariogram has a sill of (C_0 + C_1) that is approached asymptotically. In practice, a finite range is taken during variogram modelling for which $\gamma(a) = 0.95*(C_0 + C_1)$ (ASCE Task Committee, 1990a).

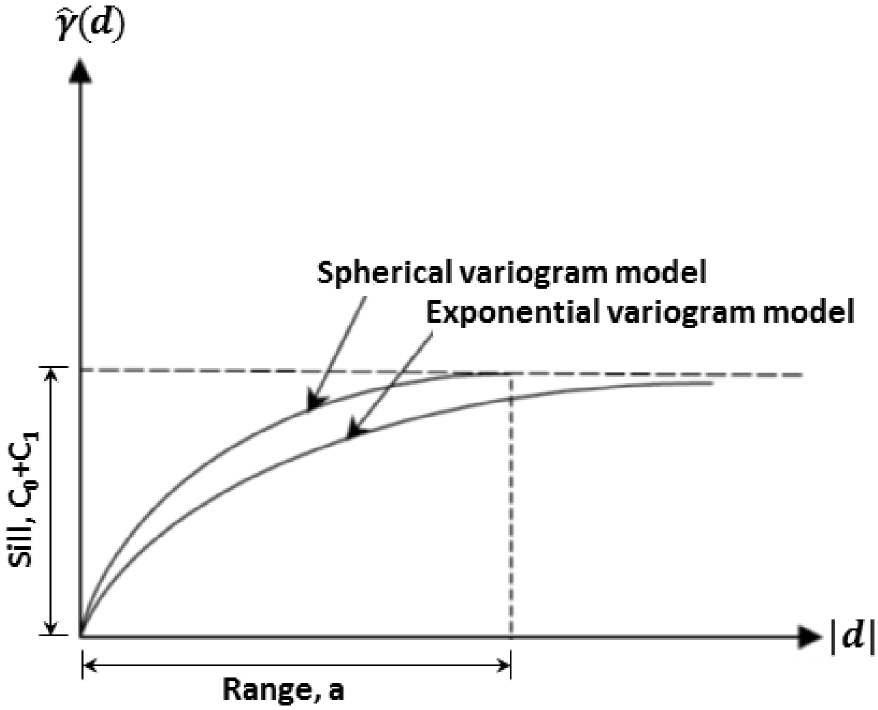

Figure 2-3. Schematization of the exponential and spherical variogram models

Furthermore, an important observation can be found from Figure 2-2, which indicates that the experimental variogram values is a function of the average lag distance between stations. This indicates that the experimental vs lag distance relationship can be modelled using suitable techniques (if any). Isaaks and Srivastava (1989) demonstrated that the experimental variogram (using Equation 2-1) can be modelled by any new mathematical function other than the available standard variogram models (exponential, gaussian, spherical models in Table 2-1). It is important to note that different universal function approximation-based techniques including artificial neural networks (ANN), genetic programming (GP), support vector machine (SVM), fuzzy theory and genetic algorithm (GA) could be a viable option for variogram modelling in such cases, which can be used to fit the experimental variogram (e.g., Chang et al., 2005; Teegavarapu, 2007; Sitharam et al., 2008; Kholghi and Hosseini, 2009; Teegavarapu et al., 2009; Huang et al., 2012). The most important advantage of these techniques is that they do not require any pre-defined functional form to fit the experimental variogram unlike the standard variogram models. For this reason, they are able to fit the experimental variogram of any shape based on their high approximation capability (e.g., Teegavarapu, 2007).

It is worth mentioning that developing new variogram models by fitting of the experimental variogram through the aforementioned universal function approximation-based techniques may not result in a unique and stable solution for the kriging weights, if the ‘positive definiteness condition’ is not satisfied (Wackernagel, 2003; Teegavarapu, 2007). Therefore, the universal function approximation-based variogram models should satisfy the positive definiteness condition before using them in kriging computation. The positive definiteness condition for a variogram model demonstrates that the kriging variance obtained based on the adopted variogram model remains always positive (Webster and Oliver, 2007). It is important to note that the standard variogram models given in Table 2-1 consist of the positive definite functions and hence always satisfy the positive definiteness condition. Therefore, the standard variogram models always generate unique and stable solutions for the kriging weights (Wackernagel, 2003) and hence, they can be directly used in kriging analysis.

2.3.2 Concept of kriging-based geostatistical approach in rain gauge network design

Kriging-based geostatistical method has been demonstrated to have many potential applications in hydrological research (Delhomme, 1978; ASCE Task Committee, 1990b). In particular, the optimal design of a rain gauge network over a catchment using the concept of variance reduction under the kriging-based geostatistical framework is quite popular among the hydrologists. In most of the applications, rain gauge network design and evaluation focuses on minimizing the kriging variance of the areal average and/or point rainfall estimates across a particular study area or catchment (Cheng et al., 2008). The kriging variance, $\sigma_z^2(s_0)$, of a regionalized variable, Z can be estimated using the Equation (2-2) as given in the following:

$$\sigma_z^2(s_0) = \sum_{i=1}^{n} \omega_i \gamma(d_{0i}) + \mu_z \qquad \text{for} \sum_{i=1}^{n} \omega_i = 1 \qquad (2-2)$$

where, $\gamma(d)$ is the variogram value for the distance d; d_{0i} is distance between sampled data points s_i and s_j; μ_z is the Lagrangian multiplier in the Z scale; and d_{0i} is distance between the unsampled location s_0 (where estimation is desired) and sampled locations s_i; n is number of sampled locations (Webster and Oliver, 2007). The square root of the kriging variance is termed as the kriging standard error that forms the basis for the rain gauge network design and evaluation.

In the kriging-based geostatistical method, the main objective is to minimize the kriging variance or kriging standard error of the network to obtain the optimal rain gauge network. The conceptual framework of the variance reduction technique using the kriging-based geostatistical method for rain gauge network design (e.g., Loof et al., 1994) is shown in Figure 2-4. As can be seen from the figure, the optimization of a rain gauge network is achieved through minimizing the kriging variance or kriging standard error (KSE) of the network. The underlying principle is that optimal positioning of rain gauge stations in the high variance zones will reduce KSE of the network and hence will improve the network performance. Applying this principle repeatedly, at a certain stage when the optimal combination of all rain gauge stations are achieved that yield high network performance, the optimal rain gauge network can be obtained. Thus, the optimal rain gauge network can be achieved by minimizing the kriging error, which involves a process of methodical search to find an optimal combination of the appropriate number and locations of stations producing the minimum KSE.

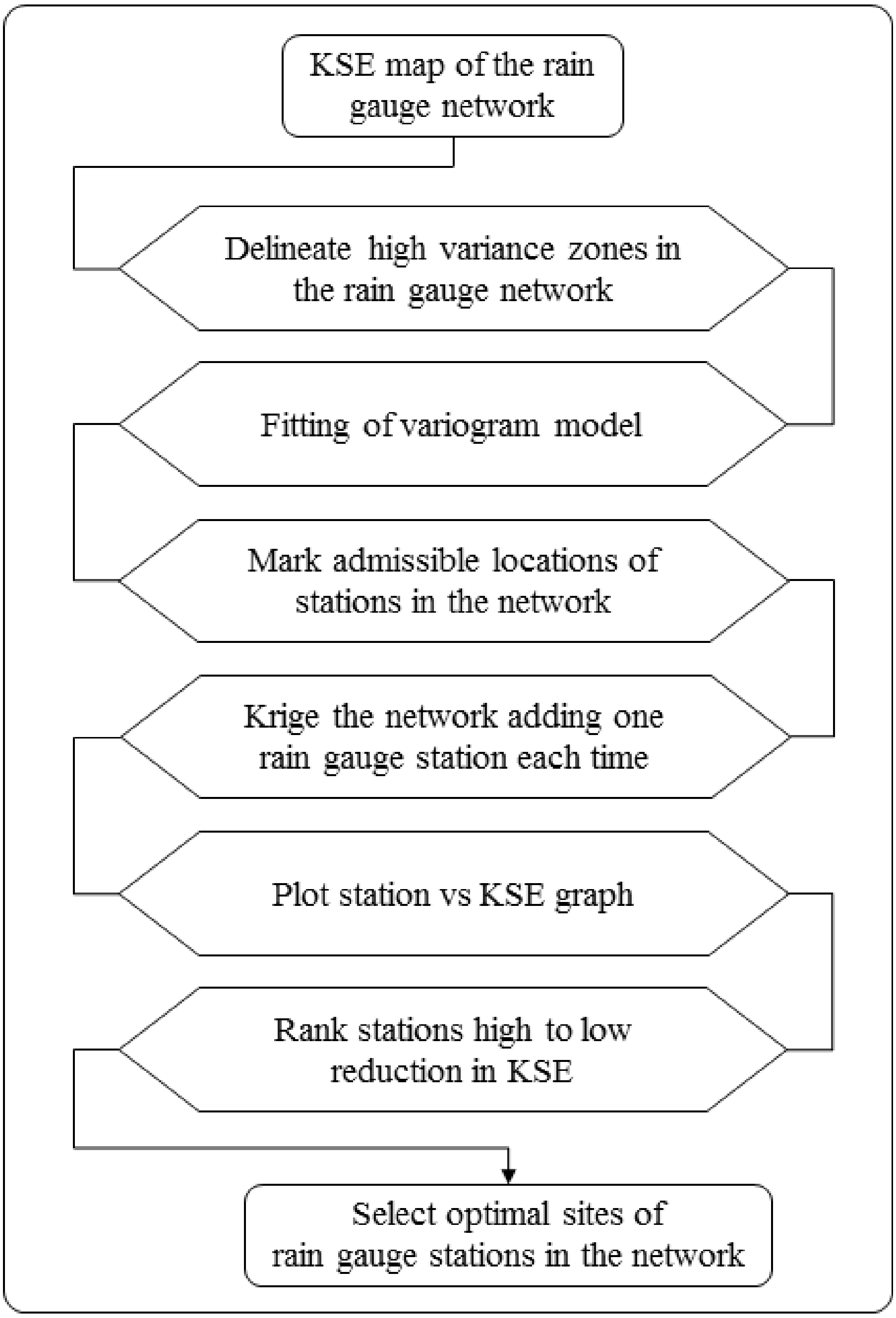

Figure 2-4. Flowchart for the design of a rain gauge network using the kriging-based geostatistical approach (Adapted from Loof et al., 1994)

2.3.3 Application of kriging-based geostatistical approach in rain gauge network design

Kriging-based geostatistical approach has been extensively used in network design and evaluation around the world including the design of rain gauge network (e.g., Cheng et al., 2008; Yeh et al., 2011; Haggag et al., 2016; Chang et al., 2017; Feki et al., 2017), drought monitoring network (e.g, Bonaccorso et al., 2003), evaporation monitoring network (e.g., Ashraf et al., 1997), groundwater head monitoring network (e.g, Cameron and Hunter, 2002; Kumar et al., 2005; Yang et al., 2008; Triki et al., 2013), groundwater quality monitoring network (e.g., Narany et al., 2015; Júnez-Ferreira et al., 2016) and air temperature monitoring network (e.g., Ahmed, 2004). In particular, the approach finds wide applications in rain gauge network design and evaluation due to its simplicity of application and easy to use (e.g., Shamsi et al., 1988; Kassim and Kottegoda, 1991; Loof et al., 1994; Papamichail and Metaxa, 1996; Pardo-Igúzquiza, 1998; Ghahraman and Sepaskhah, 2001; Tsintikidis et al., 2002; Nour et al., 2006; Barca et al., 2008; Chen et al., 2008; Cheng et al., 2008; Chebbi et al., 2011; Yeh et al., 2011; Shaghaghian and Abedini, 2013; Shafiei et al., 2014; Adib and Moslemzadeh, 2016; Aziz et al., 2016; Haggag et al., 2016; Chang et al., 2017; Feki et al., 2017).

In most of the aforementioned studies for rain gauge network design, kriging was implemented alone where only rainfall was considered as a single regionalized variable in kriging analysis. However, it is also found from the literature that the kriging-based geostatistical approach offers an advantage over other network design methods because kriging is able to complement the sparsely sampled primary variable, such as rainfall by the correlated densely sampled secondary variable, such as elevation, radar-based rainfall data etc. This is often referred to as the cokriging method (Goovaerts, 2000; Feki et al., 2012). Therefore, some studies also implemented cokriging for rain gauge network design applications (e.g., Moore et al., 2000; Bradley et al., 2002; Volkmann et al., 2010; Putthividhya and Tanaka, 2013). In most of these studies, radar-based rainfall was used as the secondary information. However, high quality radar data is not frequently available in many cases, which could be one of the major constraints that limit the application of the cokriging method in the design and evaluation of a rain gauge network.

In regard to the rain gauge network design and evaluation based on the kriging based geostatistical approach, three different attributes have been identified as given in the following:

- Augmentation of the network with additional rain gauge stations to achieve the desired network density and then identifying the optimal location of the additional stations in the network to improve the rainfall estimation accuracy as much as possible (e.g., Loof et al., 1994; Papamichail and Metaxa, 1996; Pardo-Igúzquiza, 1998; Tsintikidis et al., 2002; Barca et al., 2008; Chebbi et al., 2011; Haggag et al., 2016)
- Prioritizing the rain gauges in the network with respect to their contribution in estimation error reduction and accuracy improvement (e.g., Kassim and Kottegoda, 1991; Chen et al., 2008; Cheng et al., 2008; Yeh et al., 2011; Chang et al., 2017; Feki et al., 2017)
- Choosing an optimal representative subset of rain gauges from an existing dense network to achieve as much information as possible maintaining the desired level of accuracy (e.g., Shamsi et al., 1988; Shaghaghian and Abedini, 2013; Adib et al., 2016; Aziz et al., 2016).

It can be also found in most of the past studies that expansion of the existing network by adding supplementary stations has been the main underlying criterion to achieve the optimal rain gauge network (detailed in Figure 2-4). 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 (St-Hilaire et al., 2003; Mishra and Coulibaly, 2009) that may make little or no contribution to the network performance for providing quality data. Mishra and Coulibaly (2009) suggested that 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 rainfall estimation uncertainty. Therefore, optimal placement of redundant stations as well as additional stations must be ensured in order to achieve the optimal rain gauge network.

2.3.4 Limitations of kriging-based geostatistical approach in rain gauge network design

In spite of the overwhelming advantages offered by the kriging-based geostatistical approach, the technique is defected with some limitations. The approach cannot be employed to initiate a network; that is, it cannot be used for design purposes unless a priori monitoring points with collected data are available. This is also true for any other statistical technique such as entropy method (discussed in Section 2.4) that is used to design and evaluate a monitoring network. In particular, availability of data from a number of monitoring points is one of the important requirements to implement kriging-based geostatistical technique in order to compute the kriging variance or kriging standard error, which is finally used for rain gauge network design and evaluation. Therefore, presence of an existing network with a number of rain gauge stations is vital for successful implementation of this approach. In regard to this, a region with unavailable data, specifically for the completely ungauged catchments with no data will restrict the application of the simplest form of this approach. In such case, remotely sensed data could be a viable option (e.g., Dai et al., 2017) and the kriging-based geostatistical approach may be applied with modification using the remotely sensed data for rain gauge network design and evaluation.

2.4 Entropy Method

2.4.1 Overview of entropy theory

The term ‘entropy’ as a scientific concept was first used in thermodynamics as early as the 1850s by Clasius. Later in 1877, Boltzmann provided a probabilistic explanation of the concept within the context of statistical mechanics. The explicit relationship between entropy and probability was developed in the early 1900s by Planck (e.g., Harmancioglu and Singh, 1998; Singh, 2013). Shannon (1948a, b) used the concept to provide an economic interpretation of the properties of long sequences of symbols, and applied the results to a number of basic problems in coding theory and data transmission. Shannon finally developed, with his remarkable contributions in this area, the basis of modern information theory. Later, Jaynes (1957a, b) re-evaluated the method of maximum entropy and applied it to a variety of problems.

2.4.2 Concept and application of entropy method in rain gauge network design

Application of engineering principles to the problem of data collection calls for a minimum number of signals to be received to obtain the maximum amount of information. Redundant information does not help, which reduce the uncertainty further. Hence, this only increases the costs of obtaining data. In the case of redundant information, for example, an existing monitoring network should be reduced or rationalized and in the case of shortage of information, the existing network should be expanded (Mogheir and Singh, 2002). These considerations represent the essence of the field of communications and therefore hold equally true for hydrologic data sampling, which is essentially communicating with the natural system. Since the reduction of the uncertainty by means of making observations is equal to the amount of information gained, the entropy criterion indirectly measures the information content of a given series of data (Harmancioglu and Yevjevich, 1987).

There have been a wide range of applications of entropy theory in the field of hydrology and water resources (e.g., Singh 1997; Singh, 2013). In particular, the entropy principle can be effectively used to develop suitable network design criteria on the basis of quantitatively expressed information expectations and information availability in the rain gauge network. The entropy theory may be used to set up an information-based design strategy to design the rain gauge network. This is justified in the sense that a rain gauge network is basically an information system (Krstanovic and Singh, 1992a, b). In fact, past studies on application of the entropy principle in rain gauge network design have revealed promising results particularly in the selection of technical design features, such as optimal locations of rain gauge stations and identification of redundant stations (e.g., Husain, 1989; Krstanovic and Singh, 1992a, b; Al-Zahrani and Husain, 1998; Yoo et al., 2008; Karimi-Hosseini et al., 2011; Ridolfi et al., 2011; Vivekanandan and Jagtap, 2012; Li et al., 2012; Wei et al., 2014; Xu et al., 2015; Werstuck and Coulibaly, 2016; Xu et al., 2017). There are also a number of studies that used the entropy theory in design of monitoring networks in different fields including water quality monitoring network design (e.g., Alpaslan et al., 1992; Harmancioglu and Alpaslan, 1992; Ozkul et al., 2000), and groundwater quality monitoring network design (e.g., Mogheir and Singh, 2002; Mogheir et al., 2009).

2.4.3 Limitations of entropy method in rain gauge network design

Past studies have demonstrated that entropy method is also a promising method in rain gauge network design problems because it allows the design of a network through quantitative assessment of information. However, there are some limitations of the entropy method, which must be noted (Harmancioglu et al., 1994). A sound evaluation of rain gauge network features by the entropy method requires the availability of reliable and sufficient data. Applications with inadequate data often cause numerical difficulties and hence unreliable results. For example, when considering spatial and temporal frequencies in the multivariate case, the major numerical difficulty is related to the proprieties of the covariance matrix. Furthermore, when the determinant of the matrix is very small, entropy measures cannot be determined reliably since the matrix becomes ill-conditioned. This often occurs when the available sample size is very small (e.g., Harmancioglu and Alpaslan, 1992). On the other hand, the question in regard to the data availability is how many data points would be considered sufficient. Particularly, it is quite challenging to determine when a data record can be considered sufficient (Harmancioglu et al., 1994). The presence of gaps in data series puts limitations on entropy estimates particularly in the time domain such that temporal design cannot be realized after certain lags (Harmancioglu et al., 1999). The same difficulty extends to space/time design of network, which leads to unreliable results.

Another problem of the application of the entropy theory is the mathematical definition of entropy concepts for continuous variables. Shannon's basic definition of entropy is developed for a discrete random variable, and the extension of this definition to the continuous case entails the problem of selecting the discretizing class intervals to approximate probabilities with class frequencies. Different measures of entropy vary with class intervals such that each selected value of class intervals constitutes a different base level or scale for measuring uncertainty. Consequently, the same variable investigated assumes different values of entropy for each selected value of class interval. It may even take on negative values which contradict the positivity property of the entropy function in theory. A possible procedure to overcome the above deficiency may be to define confidence levels for entropy measures, particularly for transinformation. This is still an issue that needs to be investigated. This can be resolved by an a priori probability distribution function, but there are then controversies over the choice of an a priori distribution so that the problem still remains unsolved.

2.5 Hybrid or Mixed Method

In many cases, more than one statistical technique of rain gauge network design is merged together in order to achieve the more accurate network design methodology. Combination of two statistical methods offers some advantages compared to when each of them is used individually for network design. These methods are referred to as the hybrid or mixed methods. In general, kriging technique alone was used for the design of rain gauge networks in most of the past studies (e.g., Shamsi et al., 1988; Kassim and Kottegoda, 1991; Loof et al., 1994; Papamichail and Metaxa, 1996; Tsintikidis et al., 2002; Cheng et al., 2008; Haggag et al., 2016; Feki et al., 2017). However, in many studies, the kriging technique was used in combination with other techniques such as entropy (e.g., Chen et al., 2008; Yeh et al., 2011; Awadallah, 2012; Mahmoudi-Meimand et al., 2015) and multivariate factor analysis (e.g., Shaghaghian and Abedini, 2013) for the design of rain gauge networks. In those studies, the kriging technique was used to estimate rainfall data by interpolation in locations where prospective rain gauge stations are required to be installed, while 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. An important advantage of these hybrid methods is that the combination of kriging and entropy or kriging and factor analysis can effectively determine the optimum number and spatial distribution of rain gauge stations in a catchment or watershed.

2.6 Optimization Method

Generally, optimization refers to the study of problems in which one seeks to minimize or maximize a real function by systematically choosing the values of real or integer variables from within an allowed set. The basic application of optimization in network design is to maximize information with respect to minimizing cost (Mishra and Coulibaly, 2009). According to Langbein (1979), the design of rain gauge networks need not essentially be based on formal schemes of optimization, such as the minimum cost of attaining data accuracy. A design can be based upon judgmental analyses to accommodate a mix of design criteria. Available literature indicate that most of the past studies used a sequential trial and error procedure to minimize the kriging variance of rain gauge networks to optimize the network (e.g., Shamsi et al., 1988; Loof et al., 1994; Papamichail and Metaxa, 1996; Haggag et al., 2016). However, a few studies used optimization techniques such as simulated annealing, genetic algorithm, particle swarm optimization for minimizing the kriging variance in order to optimize rain gauge networks (e.g., Pardo-Igúzquiza, 1998; Barca et al., 2008; Chebbi et al., 2011; Adib et al., 2016; Aziz et al., 2014; Aziz et al., 2016). Several optimization techniques have been proposed in the literature since the early work of Bras and Rodríguez-Iturbe (1976) and Delhomme (1978), who demonstrated a methodology of rain gauge network design based on the minimization of the mean areal kriging variance. The adoption of optimization techniques in combination with the kriging-based geostatistical method for rainfall network sizing and augmentation was also performed by Pardo-Igúzquiza (1998), Barca et al. (2011), Chebbi et al. (2011), Aziz et al. (2014); Adib et al. (2016) and Aziz et al. (2016).

In Delhomme (1978), the optimal location of rain gauges was identified using a technique called the fictitious point method while an automatic optimization technique, namely simulated annealing was adopted in Pardo-Igúzquiza (1998) and Aziz et al. (2014). Barca et al. (2008) provided a methodology for assessing the optimal location of new rain gauge stations within an existing rain gauge network. The methodology used kriging and probabilistic techniques (simulated annealing) combined with geographic information system (GIS). Chebbi et al. (2011) have considered mono-objective criteria in simulated annealing technique assuming 1-hour rainfall intensity interpolation and erosivity factor interpolation and using one single extreme rainfall event to perform the analysis. Rainfall quantities retained in previous studies were mainly taken in a deterministic way. Effectively, a single rainfall pattern was selected for which the average kriging variance was minimized to achieve the best new rain gauge locations (Delhomme, 1978; Pardo-Igúzquiza, 1998; Chebbi et al., 2011). In the recent past, genetic algorithm and particle swarm optimization techniques in combination with geostatistics were also used for rain gauge network design to minimize the kriging variance for network optimization in two recent studies (e.g., Adib et al., 2016; Aziz et al., 2016). The major concern of the optimization method is that this method often results in a network configuration, which is often not practically suitable to be implemented as the optimal network. Because it is practically unrealistic to re-locate and re-install all existing rain gauge stations in their new identified locations (which may be frequently occurred as a part of the solution) based on the optimization method, which involves large cost and logistics supports. As a result of these difficulties, the sequential trial and error procedure (e.g., Shamsi et al., 1988; Loof et al., 1994) to minimize the kriging variance to optimize rain gauge networks is usually preferred, in particular to identify the optimal locations of the additional and likely redundant stations in the network.

2.7. Impact of Rain Gauge Network Design on the Accuracy of Streamflow Simulation

Rainfall-runoff models are widely used to forecast the streamflow for a catchment of known rainfall that comes from an established rain gauge network with a number of rain gauge stations. In places where streamflow statistics are scarce, the models are used for design purposes to infer flows of a particular frequency by applying a design storm of the same frequency to the model (Retnam and Williams, 1988). Hydraulic structures at the catchment outlet are then designed to cope with this streamflow. Another important application of the models is in the area of flood forecasting, where observed rainfall is used by the model to predict streamflows. Depending on the complexity of the model and the precision required, the rainfall is specified as spatially averaged and varying in time. But the rainfall measurements in most cases are, in fact, taken by continuous (in time) rain gauges at points (in space) from an established network, which gives only a limited spatial picture of rainfall (Hamlin, 1983; Retnam and Williams, 1988).

The models may be simple regression equations applied to a particular catchment or they may be a more general conceptual form requiring a range of data describing the catchment (Hamlin, 1983; Retnam and Williams, 1988). Furthermore, the validity of a model depends on the accuracy and reliability of input parameters and initial and boundary conditions (Zhu et al., 2013). Of these parameters and data, rainfall input is proved to be the most important (Golding, 2009). Additionally, accurate rainfall is required in the calibration of hydrological models to yield parameter sets, which represent catchment characteristics. Extensive use of hydrological models has demonstrated the need for accurate rainfall fields in order to produce runoff and streamflow forecasting with a high degree of confidence (Beven and Hornberger, 1982; Cole and Moore, 2008; Xu et al., 2013). According to McMillan et al. (2011, page 84), "no model, however, well-founded in physical theory or empirically justified by past performance, can produce accurate runoff predictions if forced with inaccurate rainfall data." Inaccurate rainfall data directly compromise the integrity of the model and the associated critical decisions made using model output (Golding, 2009; McMillan et al., 2011). In particular, for small catchments, the timing and location of rainfall is critical in reproducing streamflow hydrographs. There is thus an urgent need to acquire reliable rainfall estimates at high spatial and temporal resolutions, which can be achieve through establishment of an optimal rain gauge network.

Rainfall is usually measured through the ground-based rain gauge stations which are often installed in a network for a catchment or watershed. However, this rainfall measurement technique often results in errors due to lack of sufficient number of stations in the network due to financial, logistics and political factors (Goovaerts, 2000). This obviously can limit their capability of producing accurate rainfall input for hydrological models. Furthermore, hydrologists intending to forecast floods do not usually have the opportunity to design or redesign the rain gauge network (Retnam and Williams, 1988). As a result of these difficulties, spatially variability of rainfall is often inferred from existing point measurements obtained from the rain gauge stations. Hence, the spatial variation of rainfall presents two particular problems when hydrologic processes at catchment or basin scale are analysed via rainfall-runoff models (Hamlin, 1983; Retnam and Williams, 1988). The first is the extent to which point rainfall measurements adequately reflect the areal distribution of rainfall. This can be achieved through spatial variability analysis of rainfall and network design, which is discussed in the rain gauge network design part of this chapter. The second is the extent to which uncertainty in the true catchment rainfall and network configuration affect the catchment model parameters and response and hence the decision outcome.

High variability in rainfall has a demonstrated impact on runoff modelling (Schilling and Fuchs, 1986). The impact of rainfall variability was investigated by Faurés et al. (1995) who studied the effect of varying density and placement of rain gauge stations in a network on hydrological modelling outcomes for a semi-arid watershed in southeastern Arizona of United States of America. By varying the gauges used to generate the rainfall input for the model, they found that the peak runoff and the runoff volume varied significantly with a coefficient of variation ranging from 9 to 76% and 2 to 65%, respectively. This study demonstrated that in an environment dominated by high-intensity rainfall with significant spatial variability, rain gauge density and placement can strongly influence predicted streamflows from hydrological modelling, leading to increased uncertainty in model results. The errors within rain gauge measurements due to systematic and calibration issues also often result in significant error in subsequent modelling efforts.

Several researches have demonstrated the impact on network density on streamflow using conceptual and/or physically distributed models. Krajewski et al. (1991) described a Monte Carlo study of a physically based distributed parameter hydrologic model for the simulation of overland flow and streamflow. In this study, sensitivity of the model response with respect to the spatial and temporal sampling density of rainfall input was investigated. The input data were generated by a space-time stochastic model of rainfall. The generated rainfall fields were sampled by the varied density synthetic rain gauge networks. The basin response, based on 5-min increment input data from a network of high density with about one gauge per 0.1 km^2, was assumed to be the 'ground truth', and other results were compared against it. Results were interpreted in terms of hydrograph characteristics such as peak magnitude, time to peak, and total runoff volume. The results indicated that higher sensitivity of basin response with respect to the temporal resolution than to the spatial resolution of the rainfall data. Georgakakos et al. (1995) investigated the effect of the number of rain gauges (from 1 to 11) on the simulation performance (cross-correlation coefficient between observed and simulated flow) in two American river basins with an area of 2000 km^2. The results revealed that the cross-correlation coefficient increased considerably until 5 rain gauges were reached. Therefore they concluded that 11 rain gauges are more than adequate to represent mean areal precipitation over the catchment for their research purpose (the linkage of catchment climatology and hydrology to time scale).

St-Hilaire et al. (2003) investigated the impact of network density at two temporal scales, i.e., for the estimation of total annual rainfall and for the estimation of daily rainfall during specific rain events. In this investigation, kriging was used as an interpolator to estimate the spatial distribution and variance of rainfall. Kriged rainfall from two network scenarios (sparse and dense) was used as input into the HSAMI hydrological model, and simulations were compared on five drainage basins in the Mauricie watershed of Quebec, Canada. A comparison of the distribution of total annual rainfall interpolated from the two network scenarios showed that addition of rain gauge stations changed the distribution and magnitude of rainfall in the study area. High rainfall cells were better defined with the denser network, and decreases in the relative spatial variance were observed. Finally, simulations performed with the HSAMI model were generally improved when the rainfall input were estimated using a denser station network for most drainage basins studied, as expressed by increased Nash coefficients and a decreased root-mean-square error. Peak flows during important summer flood events were generally better simulated when a denser network was used to calculate the mean daily rainfall used as input. Total cumulated volume estimations during the rain events were also generally improved with a denser rain gauge network.

Dong et al. (2005) used statistical analyses and hydrological modelling to find the appropriate number and location of rain gauges for a river basin for streamflow simulation. First, a statistical method is used to identify the appropriate number of rain gauges, where the effect of the number of rain gauges on the cross correlation coefficient between areal averaged rainfall and discharge is investigated. Second, a lumped HBV model (a conceptual hydrological model for continuous prediction of runoff, developed at the Swedish Meteorological and Hydrological Institute) is used to investigate the effect of the number of rain gauges on hydrological modelling performance. The Qingjiang River basin with 26 rain gauges in China is used for this case study. The results show that both cross correlation coefficient and modelling performance increase hyperbolically and level off after five rain gauges (therefore identified to be the appropriate number of rain gauges) for this basin. The study concludes with the identification of geographical locations of rain gauges which give the best and worst hydrological modelling performance, which shows that there is a strong dependence on the local geographical and climatic patterns.

The influence of the spatial resolution of the rainfall input on the model calibration and application was investigated by Bárdossy and Das (2008). The analysis was carried out by varying the distribution of the rain gauge network. The semi-distributed HBV model was calibrated with the rainfall interpolated from the available observed rainfall of the different rain gauge networks. The meteorological input was interpolated using the external drift kriging method from the point measurements of the selected rain gauge stations. A mesoscale catchment located in the southwest of Germany was selected for this case study. They highlighted that the number and spatial distribution of rain gauges significantly affect the streamflow simulation results. It was found that the overall performance of the model worsens radically with an excessive reduction of rain gauges, and there is no significant improvement of the model by increasing the number of rain gauges more than a certain threshold number.

An approach to improve the runoff forecasting based on the optimization of the mean daily areal rainfall series was investigated by Anctil et al. (2006). The experimental protocol in this investigation was structured in two phases. First, the rain gauge network was randomly sampled to produce subsets of a specific number of rain gauges, in order to assess the impact of reduced rainfall knowledge on streamflow forecasting performance. Then, a genetic algorithm was used to orient the rain gauge combinatorial problem toward improved forecasting performance. Random sampling revealed that median performance diminishes rapidly when 10 rain gauge stations or fewer (out of 23) are used to compute the mean areal rainfall time series. Results also indicated that some rain gauge combinations lead to better forecasts than when all available rain gauges are used to estimate the mean areal rainfall. These findings justify the genetic search performed in the second phase of the study. The best performance improvement was achieved when the mean areal rainfall was computed from a specific 12–rain gauges combination. Many other combinations also lead to noticeable improvements in streamflow forecasting, revealing the complexity of the identification of an optimal sub-network. From an optimization point of view, and through the filter of a lumped neural network rainfall-runoff model, these results demonstrate that it may be beneficial to reduce the size of the total rain gauge network.

In recent years, the availability of spatially continuous radar rainfall data has led to its widespread utilization by hydrologists and inclusion in particular sorts of model. Tsai et al. (2014) developed a novel semi-distributed, data-driven, rainfall-runoff model for the Shihmen catchment, Taiwan to forecast reservoir inflow for a range of different forecasting horizons. The aim was to demonstrate how different levels of semi-distribution, applied to continuous radar rainfall data inputs, operating in a data-driven rainfall-runoff modelling framework, affect the performance of multi-step-ahead reservoir inflow forecasts in Taiwan. To perform simulations, a set of models based on different levels of radar rainfall spatial disaggregation was formulated using the adaptive network-based fuzzy inference system (ANFIS), from which a model with the preferred level of input distribution is identified. Different levels of spatially aggregated radar-derived rainfall data are used to generate 4, 8 and 12 sub-catchment input drivers. They found that continuous rainfall data appears to provide better performance over discrete, point-based spatial data for reservoir inflow forecasting in Taiwan. It was also found that further performance improvement can be achieved by using a semi-distributed modelling framework based on the spatially disaggregated radar rainfall input.

The influence of the rain gauge density and their distribution in network on the modelling results is still a challenging topic in hydrology. Rain gauge network configuration is not only depends on the station density, station location also plays an important role in determining whether information is gained properly, which is to be used for hydrological modelling. Xu et al. (2013) investigated the influence of different rain gauge density, network distribution and location of the rain gauge stations on the performance of the model in simulating the streamflow using the Xinanjiang model applied in Xiangjiang River basin, China. The basin consists of a dense rain gauge network with long and high quality data. Hence, the mean areal rainfall estimated by different rain gauge densities achieved through the stochastically-based rain gauge network design technique was analyzed at first. Secondly, the influence of different rain gauge density and distribution on the model performance was comprehensively evaluated. The results show that the error range of the indices in analyzing mean areal rainfall and simulated runoff narrowed gradually with increasing number of rain gauges up to some threshold, and beyond which the model performance did not show considerable improvements. The study found that the networks with 10 rain gauge stations yield the lowest model performance, whereas the highest model performance can be obtained using the 128 rain gauges network. Furthermore, when the number of rain gauges is higher than 93, there is no noticeable improvement in the model performance. Based on the stochastically generated rain gauge network, this study quantitatively demonstrated how the rain gauge stations' geographic locations impact the streamflow simulation results.

During the times of floods and natural calamities, it becomes difficult to collect information from all rain gauges in a catchment. Furthermore, it is reported that the impact of climatic change affects rainfall amounts, rainfall patterns, runoff amounts, and runoff coefficients (Ponce et al., 1997). Therefore, establishing a key rain gauge network is vital, which is capable of effective forecasting of floods with the desired accuracy. Kar et al. (2015) presented a methodology for the design of the key rain gauge network using multi-criteria decision analysis and clustering techniques for flood forecasting and demonstrated through a case study in the Kantamal sub-catchment, Mahanadi River basin, India. The objective was to achieve the key rain gauge network (instead of taking information from all rain gauge stations) that can be used for making reasonably accurate flood forecasts particularly during the time of distress (when the rainfall data of not all the stations are available due to various reasons). At first, different possible key rain gauge networks were designed from the existing rain gauge network in the catchment using Hall's method, analytical hierarchical process (AHP), self-organization map (SOM) and hierarchical clustering (HC) techniques. Efficiency of the key networks is tested by artificial neural network (ANN), fuzzy logic and NAM rainfall-runoff models, and the best network was used for flood forecasting. Further, flood forecasting has been carried out with the key rain gauge networks. Although, the best rain gauge network has shown the highest efficiency, simultaneously other networks were also tested with certain designated efficiency in order to use them at the time of failure of the best rain gauge network. Furthermore, flood forecasting has been carried out using the three most effective rain gauge networks which uses only 7 rain gauges instead of 14 rain gauges established in the catchment. The fuzzy logic applied on the key rain gauge network derived using AHP yielded the best result for flood forecasting with efficiency of 83% for 1-day ahead forecast. This study conclusively proves the need of the key rain gauge network designed for effective flood forecasting, particularly when there is difficulty in gathering the information from all rain gauges over a catchment.

The review of the analysis and results of all the aforementioned studies clearly highlight the importance and necessity of the optimal rain gauge network design for the enhanced rainfall estimations and incorporating the optimal rain gauge network-based input for accurate and improved streamflow forecasting.

Chapter 3

Universal Function Approximation-Based Kriging for Improved Estimation of Rainfall

3.1 Introduction

Many hydrological analyses and investigations need the estimates of hydrologic variables such as rainfall at ungauged locations in a catchment where no direct observations are readily available. In this chapter, development and application of a variance dependent universal function approximation-based kriging for the improved estimation of rainfall at an ungauged location in the case study area is detailed. In general, variance dependent stochastic interpolation methods such as kriging are the most frequently used methods for estimating the point rainfall values at any ungauged (or target) location in a catchment based on the observed rainfall values available at all other surrounding rain gauge stations (Isaaks and Srivastava, 1989; Goovaerts, 1997; Vieux, 2001; Webster and Oliver, 2007). However, the major issue associated with traditional kriging is that it requires correct estimation of a variogram model that represents spatial correlations among data points. The variogram model plays a vital role in the kriging process to estimate the kriging weights and thus significantly impacts the performance of the methods. The robustness of kriging thus heavily depends on how the variogram model is obtained (Oliver and Webster, 2014). In traditional kriging, it is the common practice to use a number of pre-defined standard variogram functions (e.g., exponential, gaussian, spherical models) to fit the experimental variogram of observed rainfall data. Furthermore, selection of an appropriate variogram model from the available standard models, finding the optimal variogram parameters (i.e., nugget, range, sill coefficients) and the computational burden involved are some of the difficulties associated with the traditional kriging (Teegavarapu, 2007; Oliver and Webster, 2014). In other words, it is generally not clear which particular variogram model from a set of available standard variogram models can perform well with the traditional kriging for a particular rainfall dataset.

As a potential solution to the aforementioned issues, a new universal function approximation-based kriging was developed using genetic programming (GP) where GP (Koza, 1992; Banzhaf et al., 1997) was used as a universal function approximator to derive the variogram model. The new variant of kriging is referred to as the genetic programming-based ordinary kriging (GPOK) in which the standard parametric variogram models (i.e., exponential, gaussian, spherical models) in traditional ordinary kriging were replaced by the GP-derived variogram model. Variogram modelling using GP offers a number of advantages. For example, GP does not require a pre-defined mathematical formula to generate the functional variogram models unlike the standard variogram model functions (as discussed in Section 2.3.1 and shown in Table 2-1). In addition, GP is able to generate variogram models of simple form, which consist of similar mathematical structure as in the standard variogram models unlike other universal function approximators such as ANN (e.g., Teegavarapu, 2007). Furthermore, GP-derived variogram model does not require identifying the variogram parameters in advance, unlike the standard parametric variogram models in traditional kriging. Therefore, the GPOK method is found to be completely free from the trial and error process of estimating the variogram parameters. The GPOK developed in this way can overcome the aforementioned limitations associated with the traditional ordinary kriging and give the enhanced estimation of rainfalls at ungauged locations.

In kriging, it is a common practice to select the best variogram model through the cross-validation and positive definiteness tests before using it for final interpolation of variable of interest. It is worth mentioning that the standard variogram models always satisfy the positive definite condition and thus results in a unique solution for the kriging weights (Wackernagel, 2003; Teegavarapu, 2007). Therefore, the cross-validation test alone is adequate to detect the best standard variogram model. However, developing and selecting a new variogram model (for example, the GP-derived variogram model detailed in this chapter) other than the standard ones cannot result in a unique and stable solution for the kriging weights, if the positive definiteness condition is not satisfied (Teegavarapu, 2007). The positive definiteness condition for a variogram model demonstrates that the variance obtained through kriging interpolation based on the adopted variogram model remains always positive (Wackernagel, 2003; Webster and Oliver, 2007). Therefore, the GP-derived model need to satisfy the cross-validation and positive definiteness tests simultaneously before it can be finally used in the GPOK method for spatial interpolation of rainfall. In order to test the positive definiteness of the GP-derived variogram model, the kriging variance for all rain gauge locations was calculated through the GPOK method based on the adopted GP-derived variogram model. The positive value of the kriging variance can confirm that the GP-derived model and its corresponding covariance function is capable of providing a unique and stable solution for the kriging weights in GPOK method. This indicates the suitability of the GP-derived model for kriging applications.

Furthermore, ordinary kriging does not ensure getting positive kriging weights and thus, negative weights can be obtained as a part of the solution to satisfy the requirement of unbiased constraints of kriging. The reason of getting negative weights in ordinary kriging is due to the fact that there is no constraint in ordinary kriging algorithm that can force the kriging process to take positive values for the kriging weights (Szidarovszky et al., 1987). Negative kriging weights when assigned to high rainfall values at a rain gauge station can lead to the negative estimation of rainfall values at the target ungauged location, which does not make practical sense. In case of this problem, Szidarovszky et al. (1987) and Deutsch (1996) have suggested that negative kriging weights, if it is obtained as a part of the solution, should be corrected and converted to positive weights through the positive kriging technique (Barnes and Johnson, 1984). As a solution to this problem, a variant of positive kriging using the mathematical programming approach was explored in this chapter, in which an additional non-negativity constraint of the kriging weights is assigned. This approach can limit the kriging weights to non-negative values. Finally in this chapter, the performance and robustness of the GPOK technique was tested against three traditional ordinary kriging as well as ANN-based ordinary kriging methods numerically and graphically for the enhanced estimation of rainfall at an ungauged location in the case study catchment.

This chapter consists of the following journal paper.

1. **Adhikary SK**, Muttil N, Yilmaz AG. 2016a. Genetic programming-based ordinary kriging for spatial interpolation of rainfall. *Journal of Hydrologic Engineering* 21(2): 04015062. DOI: 10.1061/(ASCE)HE.1943-5584.0001300. (SCImago Journal Rank indicator: Q1; Impact Factor: 1.694).

VICTORIA UNIVERSITY
MELBOURNE AUSTRALIA

GRADUATE RESEARCH CENTRE

3.2 DECLARATION OF CO-AUTHORSHIP AND CO-CONTRIBUTION: PAPERS INCORPORATED IN THESIS BY PUBLICATION

This declaration is to be completed for each conjointly authored publication and placed at the beginning of the thesis chapter in which the publication appears.

1. PUBLICATION DETAILS (to be completed by the candidate)

Title of Paper/Journal/Book: Genetic Programming-Based Ordinary Kriging for Spatial Interpolation of Rainfall

Surname: Adhikary First name: Sajal Kumar

College: College of Engineering & Science Candidate's Contribution (%): 80

Status:

Accepted and in press: ☑ Date: 06/08/2015

Published: ☑ Date: 11/09/2015

2. CANDIDATE DECLARATION

I declare that the publication above meets the requirements to be included in the thesis as outlined in the HDR Policy and related Procedures – policy.vu.edu.au.

[redacted]	20/02/2017
Signature	**Date**

3. CO-AUTHOR(S) DECLARATION

In the case of the above publication, the following authors contributed to the work as follows:

The undersigned certify that:

1. They meet criteria for authorship in that they have participated in the conception, execution or interpretation of at least that part of the publication in their field of expertise;

2. They take public responsibility for their part of the publication, except for the responsible author who accepts overall responsibility for the publication;

3. There are no other authors of the publication according to these criteria;

4. Potential conflicts of interest have been disclosed to a) granting bodies, b) the editor or publisher of journals or other publications, and c) the head of the responsible academic unit; and

5. The original data will be held for at least five years from the date indicated below and is stored at the following **location**(**s**):

College of Engineering and Science, Victoria University, Melbourne, Australia

Name(s) of Co-Author(s)	**Contribution (%)**	**Nature of Contribution**	**Signature**	**Date**
Sajal Kumar Adhikary	80	Conceptual ideas, data analysis, model simulations, paper writing	[redacted]	20/02/2017
Nitin Muttil	10	Critical comments, discussion on conceptual ideas	[redacted]	20/02/2017
Abdullah Gokhan Yilmaz	10	Critical comments, discussion on conceptual ideas	[redacted]	20/02/2017

Updated: June 2015

Genetic Programming-Based Ordinary Kriging for Spatial Interpolation of Rainfall title by *S.K. Adhikary, N. Muttil, and A.G. Yilmaz* was published in the peer review journal, *Journal of Hydrologic Engineering, 21/2, 2016.*

The article is available from: https://doi.org/10.1061/(ASCE)HE.1943-5584.0001300

Chapter 4

Univariate and Multivariate Kriging for Enhanced Spatial Interpolation of Rainfall

4.1 Introduction

Estimates of spatial rainfall distribution are vital for many hydrological analyses and modelling applications such as water budget analysis, flood modelling, climate change studies, drought management, irrigation scheduling and water management (Moral, 2010; Delbari et al., 2013). The rain gauge network acts as the primary data source in this rainfall estimation. However, the rain-gauge network is often sparse in the field because the number of stations in a network is often limited by economic, logistics and geological factors (Goovaerts, 2000). For example, in Australia, the spatial density of rain gauge stations varies greatly across the country. Furthermore, most of the rain gauges are located along the coastal area, while the vast inland area has less than twenty per cent of the total number of rain gauges (Sun et al., 2013). As a result, point rainfall values are commonly accessible from a limited number of rain gauges. These limitations increase the need for using suitable spatial estimation methods to obtain the spatial distribution of rainfall and generate rainfall maps from the point rainfall values. Moreover, the network is often not deployed on a regular grid and rainfall data may not be available in the target location where it is most required (Lloyd, 2005). In such cases, spatial interpolation method plays a vital role to simulate rainfall in areas having no stations based on the observed rainfall values in the surrounding areas.

A number of interpolation methods ranging from simple conventional and deterministic methods to complex stochastic or geostatistical methods have been employed to interpolate point rainfall data from rain-gauges and generate spatial distribution of rainfall map over a catchment. Kriging-based geostatistical methods have been shown superior to the conventional and deterministic methods for spatial interpolation of rainfall (Isaaks and Srivastava, 1989; Goovaerts, 1997). Several studies have reported that rainfall is generally characterized by a significant spatial variation (e.g., Goovaerts, 2000; Lloyd, 2005; Delbari et al., 2013). This suggests that interpolation methods which are explicitly able to incorporate the spatial variability of rainfall into the estimation process should be employed. In kriging, the spatial variability of rainfall is quantified by the variogram model that defines the degree of spatial correlation between the data points (Webster and Oliver, 2007). Therefore, kriging has become the most widely used geostatistical method for spatial interpolation of rainfall.

The ability of kriging to produce spatial distribution of rainfall has been demonstrated in many studies across the world (e.g., Goovaerts, 2000; Jeffrey et al., 2001; Lloyd, 2005; Moral, 2010; Yang et al., 2015). The major advantage of kriging is that it can take into account the spatial correlation between data points through the variogram model and provides unbiased estimates with a minimum variance. Another key advantage of kriging over the conventional and deterministic interpolation methods is that while providing a measure of prediction standard error (also called the kriging variance), it is capable of complementing the sparsely sampled primary variable, such as rainfall by the correlated densely sampled secondary variable, such as elevation to improve the estimation of primary variable, rainfall (Hevesi et al., 1992; Goovaerts, 2000). This multivariate extension of kriging is referred to as the cokriging method.

Kriging in geostatistics thus can be broadly categorized into two groups: univariate and multivariate kriging techniques. The univariate kriging employs a single primary variable (rainfall) for interpolations, which include ordinary kriging, simple kriging, universal kriging, log-normal kriging and indicator kriging. More recently, alternative univariate kriging methods such as universal function approximation-based kriging using ANN (e.g., Teegavarapu, 2007) and GP (as demonstrated in Chapter 3) are also used for the enhanced estimation of rainfall. On the other hand, the multivariate kriging allows the inclusion of secondary variables such as elevation, slope, weather radar-rainfall data that are more densely sampled to complement the primary variable to improve rainfall estimation, which include ordinary co-kriging, kriging with an external drift, and regression kriging. Multivariate kriging usually reduces the prediction error variance and specifically outperforms the univariate kriging method if the secondary variable is highly correlated (correlation coefficient higher than 0.75) with the primary variable (i.e., rainfall) and is known at many more points (Goovaerts, 2000). Furthermore, it can incorporate the exhaustive secondary variable (i.e., elevation) to give an enhanced estimation of primary variable (i.e., rainfall) when dealing with a low density rain-gauge network.

In this chapter, the performance evaluation of different univariate and multivariate kriging interpolation methods to identify the best interpolator for enhanced spatial interpolation of rainfall and production of high quality continuous rainfall datasets in the form of rainfall maps is presented. This chapter is divided into two parts. The first part of this chapter addresses the performance evaluation of different univariate kriging interpolation methods including four traditional and a universal function approximation-based kriging method using GP for spatial interpolation of rainfall. The second part of this chapter details the performance evaluation of five traditional univariate interpolation methods including kriging and two cokriging interpolation methods for enhanced spatial interpolation of rainfall. The second part of this chapter also includes the analysis for an additional case study area, the Ovens River catchment of Victoria, Australia.

In spatial interpolation of rainfall, it is often quite challenging to distinguish the best interpolation method to estimate the spatial distribution of rainfall for a particular catchment or study area. The reason is due to the fact that the performance of an interpolation method depends on a number of factors such as catchment size and characteristics, sampling density, sample spatial distribution, sample clustering, surface type, data variance, grid size or resolution, quality of auxiliary information to be used as well as the interactions among these factors. Moreover, it is unclear how the aforementioned factors affect the performance of the spatial interpolation methods

(Dirks et al., 1998, Li and Heap, 2011). As a solution this issue, the best interpolation method for a particular catchment or study area is usually established through the comparative assessment of different interpolation methods (e.g., Goovaerts, 2000; Hsieh et al., 2006; Moral, 2010; Di Piazza et al., 2011; Mair and Fares, 2011; Feki et al., 2012; Delbari et al., 2013), which was also implemented in this thesis. The interpolation method that gives the most realistic results with the lowest bias and the highest accuracy in rainfall estimation was chosen as the best interpolator to generate the continuous rainfall datasets for the case study catchment.

Rainfall data distribution highly depends on local conditions, especially in regions having the complex terrain (e.g., Phillips et al., 1992; Lloyd, 2005; Subyani and Al-Dakheel, 2009; Moral, 2010; Feki et al., 2012). The distribution can be rather different from coast to inland, high altitude to low altitude, and upper slope to down slope etc. Sharma and Shakya (2006) highlighted that any analysis of hydroclimatic variables should be carried out at the local scale rather than at a large or global scale because of the variations of hydroclimatic situations from one region to another. Hence, it is vital to consider the local environment that affects the distribution of rainfall data first, especially topographic conditions including elevation, slope, and distance to coast in order to improve the accuracy of the gridded rainfall data. Furthermore, rainfall-topography relationships have been relatively little studied at a local or catchment scale in Australia and no such studies have been undertaken specifically within the Middle Yarra River catchment (the case study area in this research). Spatial distribution of monthly precipitation is also largely influenced by the topography in the case study catchment. Therefore, the elevation was used as an auxiliary variable in addition to rainfall (i.e., primary variable) in order to improve the estimation of rainfall using the multivariate kriging methods based on the rainfall-topography relationship of the case study catchment.

This chapter consists of the following two journal papers. The first paper focuses on different univariate kriging interpolation methods, and the second paper details different multivariate as well as univariate kriging interpolation methods for enhanced spatial interpolation of rainfall and generation of continuous rainfall datasets in the form of rainfall maps.

1. **Adhikary SK**, Muttil N, Yilmaz AG. 2016b. Ordinary kriging and genetic programming for spatial estimation of rainfall in the Middle Yarra River catchment, Australia. *Hydrology Research* 47(6): 1182-1197. DOI: 10.2166/nh.2016.196. (SCImago Journal Rank indicator: Q1; Impact Factor: 1.779).

2. **Adhikary SK**, Muttil N, Yilmaz AG. 2017a. Cokriging for enhanced spatial interpolation of rainfall in two Australian catchments. *Hydrological Processes* 31(12): 2143-2161. DOI: 10.1002/hyp.11163 (SCImago Journal Rank indicator: Q1; Impact Factor: 3.014).

VICTORIA UNIVERSITY
MELBOURNE AUSTRALIA

GRADUATE RESEARCH CENTRE

4.2 DECLARATION OF CO-AUTHORSHIP AND CO-CONTRIBUTION: PAPERS INCORPORATED IN THESIS BY PUBLICATION

This declaration is to be completed for each conjointly authored publication and placed at the beginning of the thesis chapter in which the publication appears.

1. PUBLICATION DETAILS (to be completed by the candidate)

Title of Paper/Journal/Book: Ordinary Kriging and Genetic Programming for Spatial Estimation of Rainfall in the Middle Yarra River Catchment, Australia

Surname: Adhikary First name: Sajal Kumar

College: College of Engineering & Science Candidate's Contribution (%): 80

Status:

Accepted and in press: ☑ Date: 01/12/2015

Published: ☑ Date: 29/01/2016

2. CANDIDATE DECLARATION

I declare that the publication above meets the requirements to be included in the thesis as outlined in the HDR Policy and related Procedures – policy.vu.edu.au.

Signature	Date
[redacted]	20/02/2017

3. CO-AUTHOR(S) DECLARATION

In the case of the above publication, the following authors contributed to the work as follows:

The undersigned certify that:

1. They meet criteria for authorship in that they have participated in the conception, execution or interpretation of at least that part of the publication in their field of expertise;

2. They take public responsibility for their part of the publication, except for the responsible author who accepts overall responsibility for the publication;

3. There are no other authors of the publication according to these criteria;

4. Potential conflicts of interest have been disclosed to a) granting bodies, b) the editor or publisher of journals or other publications, and c) the head of the responsible academic unit; and

5. The original data will be held for at least five years from the date indicated below and is stored at the following **location**(**s**):

College of Engineering and Science, Victoria University, Melbourne, Australia

Name(s) of Co-Author(s)	**Contribution (%)**	**Nature of Contribution**	**Signature**	**Date**
Sajal Kumar Adhikary	80	Conceptual ideas, data analysis, model simulations, organize results, paper writing	[redacted]	20/02/2017
Nitin Muttil	10	Critical comments, discussion on conceptual ideas		20/02/2017
Abdullah Gokhan Yilmaz	10	Critical comments, discussion on conceptual ideas		20/02/2017

Updated: June 2015

Ordinary Kriging and Genetic Programming for Spatial Estimation of Rainfall in the Middle Yarra River Catchment, Australia by *S.K. Adhikary, N. Muttil, and A.G. Yilmaz* was published in the peer review journal, *Hydrology Research, 47/6, 1182-1197, 2016.*

The article is available from: https://doi.org/10.2166/nh.2016.196

GRADUATE RESEARCH CENTRE

DECLARATION OF CO-AUTHORSHIP AND CO-CONTRIBUTION: PAPERS INCORPORATED IN THESIS BY PUBLICATION

This declaration is to be completed for each conjointly authored publication and placed at the beginning of the thesis chapter in which the publication appears.

1. PUBLICATION DETAILS (to be completed by the candidate)

Title of Paper/Journal/Book: Cokriging for Enhanced Spatial Interpolation of Rainfall in Two Australian Catchments

Surname: Adhikary First name: Sajal Kumar

College: College of Engineering & Science Candidate's Contribution (%): 80

Status:

Accepted and in press: ☑ Date: 01/03/2017

Published: ☑ Date: 15/06/2017

2. CANDIDATE DECLARATION

I declare that the publication above meets the requirements to be included in the thesis as outlined in the HDR Policy and related Procedures – policy.vu.edu.au.

[redacted]	15/06/2017
Signature	**Date**

3. CO-AUTHOR(S) DECLARATION

In the case of the above publication, the following authors contributed to the work as follows:

The undersigned certify that:

1. They meet criteria for authorship in that they have participated in the conception, execution or interpretation of at least that part of the publication in their field of expertise;

2. They take public responsibility for their part of the publication, except for the responsible author who accepts overall responsibility for the publication;

3. There are no other authors of the publication according to these criteria;

4. Potential conflicts of interest have been disclosed to a) granting bodies, b) the editor or publisher of journals or other publications, and c) the head of the responsible academic unit; and

5. The original data will be held for at least five years from the date indicated below and is stored at the following **location**(**s**):

College of Engineering and Science, Victoria University, Melbourne, Australia

Name(s) of Co-Author(s)	**Contribution (%)**	**Nature of Contribution**	**Signature**	**Date**
Sajal Kumar Adhikary	80	Conceptual ideas, data analysis, model simulations, organize results, paper writing	[redacted]	10/03/2017
Nitin Muttil	10	Critical comments, discussion on conceptual ideas		10/03/2017
Abdullah Gokhan Yilmaz	10	Critical comments, discussion on conceptual ideas		10/03/2017

Received: 13 January 2016 | Accepted: 1 March 2017
DOI: 10.1002/hyp.11163

WILEY

RESEARCH ARTICLE

4.4

Cokriging for enhanced spatial interpolation of rainfall in two Australian catchments

Sajal Kumar Adhikary[1,2] | **Nitin Muttil**[1,2] | **Abdullah Gokhan Yilmaz**[3]

[1] College of Engineering and Science, Victoria University, P.O. Box 14428, Melbourne, Victoria 8001, Australia

[2] Institute for Sustainability and Innovation, Victoria University, P.O. Box 14428, Melbourne, Victoria 8001, Australia

[3] College of Engineering, University of Sharjah, P.O. Box 27272, Sharjah, United Arab Emirates

Correspondence
Sajal Kumar Adhikary, College of Engineering and Science, Victoria University, Footscray Park Campus, P.O. Box 14428, Melbourne, Victoria 8001, Australia.
Email: sajal.adhikary@live.vu.edu.au

Abstract

Rainfall data in continuous space provide an essential input for most hydrological and water resources planning studies. Spatial distribution of rainfall is usually estimated using ground-based point rainfall data from sparsely positioned rain-gauge stations in a rain-gauge network. Kriging has become a widely used interpolation method to estimate the spatial distribution of climate variables including rainfall. The objective of this study is to evaluate three geostatistical (ordinary kriging [OK], ordinary cokriging [OCK], kriging with an external drift [KED]), and two deterministic (inverse distance weighting, radial basis function) interpolation methods for enhanced spatial interpolation of monthly rainfall in the Middle Yarra River catchment and the Ovens River catchment in Victoria, Australia. Historical rainfall records from existing rain-gauge stations of the catchments during 1980–2012 period are used for the analysis. A digital elevation model of each catchment is used as the supplementary information in addition to rainfall for the OCK and kriging with an external drift methods. The prediction performance of the adopted interpolation methods is assessed through cross-validation. Results indicate that the geostatistical methods outperform the deterministic methods for spatial interpolation of rainfall. Results also indicate that among the geostatistical methods, the OCK method is found to be the best interpolator for estimating spatial rainfall distribution in both the catchments with the lowest prediction error between the observed and estimated monthly rainfall. Thus, this study demonstrates that the use of elevation as an auxiliary variable in addition to rainfall data in the geostatistical framework can significantly enhance the estimation of rainfall over a catchment.

KEYWORDS

cross-variogram model, digital elevation model, kriging with an external drift, ordinary cokriging, positive-definite condition, variogram model

1 | INTRODUCTION

Rainfall data provide an essential input for many hydrological investigations and modelling tasks. Accuracy of various hydrological analyses such as water budget analysis, flood modelling, climate change studies, drought management, irrigation scheduling, and water management greatly depends on the correct estimation of the spatial distribution of rainfall (Delbari, Afrasiab, & Jahani, 2013; Moral, 2010). This usually requires a dense rain-gauge network with a large number of stations (Adhikary, Muttil, & Yilmaz, 2016a). However, the rain-gauge network is often sparse in the field because the number of stations in a network is often restricted by economic, logistics, and geological factors (Goovaerts, 2000). As a result, point rainfall data are generally accessible from a limited number of stations. These limitations increase the need for using suitable spatial estimation methods to obtain the spatial distribution of rainfall and generate rainfall map from the point rainfall values. Moreover, the network is often not deployed on a regular grid and rainfall data may not be available in the target location where it is most required (Adhikary et al., 2016a). In such cases, spatial interpolation plays a vital role to simulate rainfall in areas with no stations based on the observed rainfall values in the surrounding areas.

Several interpolation methods have been frequently employed to interpolate rainfall data from rain-gauge stations and produce the spatial distribution of rainfall over a catchment. Examples of such methods range from simple conventional (e.g., Thiessen polygons [Thiessen, 1911], isohyet mapping [ASCE, 1996], simple trend surface interpolation [Gittins, 1968]), and deterministic methods (i.e., inverse distance weighting (IDW; ASCE, 1996), radial basis function (RBF; Di Piazza, Conti, Noto, Viola, & Loggia, 2011) to complex stochastic or geostatistical methods (i.e., ordinary kriging [OK], ordinary cokriging [OCK] and kriging with an external drift [KED; Goovaerts, 2000]). Although the conventional and deterministic methods have been improved over time, their limitations continue to exist. These limitations have been described elaborately in Goovaerts (2000) and Teegavarapu and Chandramouli (2005).

Geostatistical methods have been shown superior to the conventional and deterministic methods for spatial interpolation of rainfall (Goovaerts, 1997; Isaaks & Srivastava, 1989). Several studies have reported that rainfall is generally characterised by a significant spatial variation (e.g., Delbari et al., 2013; Lloyd, 2005). This suggests that interpolation methods, which are explicitly able to incorporate the spatial variability of rainfall into the estimation process should be employed. In view of that, kriging has become the most widely used geostatistical method for spatial interpolation of rainfall. The ability of kriging to produce spatial predictions of rainfall has been distinguished in many studies (e.g., Adhikary et al., 2016a; Goovaerts, 2000; Jeffrey, Carter, Moodie, & Beswick, 2001; Lloyd, 2005; Moral, 2010; Yang, Xie, Liu, Ji, & Wang, 2015). The major advantage of kriging is that it takes into account the spatial correlation between data points and provides unbiased estimates with a minimum variance. The spatial variability in kriging is quantified by the variogram model that defines the degree of spatial correlation between the data points (Webster & Oliver, 2007).

Another key advantage of kriging over the conventional and deterministic methods is that while providing a measure of prediction standard error (also called kriging variance), it is capable of complementing the sparsely sampled primary variable, such as rainfall by the correlated densely sampled secondary variable, such as elevation to improve the estimation accuracy of primary variable (Goovaerts, 2000; Hevesi, Istok, & Flint, 1992). This multivariate extension of kriging is referred to as the cokriging method. The standard form of cokriging is the OCK method, which usually reduces the prediction error variance and specifically outperforms kriging method if the secondary variable (i.e., elevation) is highly correlated (correlation coefficient higher than .75) with the primary variable (i.e., rainfall) and is known at many more points (Goovaerts, 2000). The KED is another commonly applied cokriging method, which can incorporate the exhaustive secondary variable (i.e., elevation) to give an enhanced estimation of rainfall when dealing with a low-density rain-gauge network. Thus, cokriging including the OCK and KED methods has been the increasing preferred geostatistical methods all over the word. As highlighted by Goovaerts (2000) and Feki, Slimani, and Cudennec (2012), rainfall and elevation tend to be related because of the orographic influence of mountainous topography. Therefore, topographic information such as the digital elevation model (DEM) can be used as a convenient and valuable source of secondary data for the OCK and KED methods. The efficacy of incorporating elevation into the interpolation procedure for enhanced estimation of rainfall has been highlighted in many studies across the world (e.g., Di Piazza et al., 2011; Feki et al., 2012; Hevesi et al., 1992; Lloyd, 2005; Martínez-cob, 1996; Moral, 2010; Phillips, Dolph, & Marks, 1992; Subyani & Al-Dakheel, 2009).

A wide variety of spatial interpolation methods have been developed for the interpolation of spatially distributed point rainfall values. However, it is often challenging to distinguish the best interpolation method to estimate the spatial distribution of rainfall for a particular catchment or study area. The reason is that the performance of an interpolation method depends on a number of factors such as catchment size and characteristics, sampling density, sample spatial distribution, sample clustering, surface type, data variance, grid size or resolution, quality of auxiliary information to be used as well as the interactions among these factors. Moreover, it is unclear how the aforementioned factors affect the performance of the spatial interpolation methods (Dirks, Hay, Stow, & Harris, 1998; Li & Heap, 2011). Hence, the best interpolation method for a particular study area is usually established through the comparative assessment of different interpolation methods (e.g., Delbari et al., 2013; Dirks et al., 1998; Goovaerts, 2000; Hsieh, Cheng, Liou, Chou, & Siao, 2006; Mair & Fares, 2011; Moral, 2010). The comparison among different interpolation methods is made through a validation procedure. The interpolation method that provides better results with lower bias and higher accuracy in rainfall estimation is identified as the best interpolation method.

In the past, many studies have been devoted to the comparison and evaluation of different deterministic and geostatistical interpolation methods in a range of different regions and climates around the world. Dirks et al. (1998) compared four spatial interpolation methods using rainfall data from a network of 13 rain-gauges in Norfolk Island concluding that kriging provided no substantial improvement over any of the simpler Thiessen polygon (TP), IDW, or areal-mean methods. Goovaerts (2000) employed three multivariate geostatistical methods (OCK, KED, simple kriging with varying local means [SKVM]), which incorporate a DEM as secondary variable and three univariate methods (OK, TP, and IDW) that do not take into account the elevation for spatial prediction of monthly and annual rainfall data available at 36 rain-gauge stations. The comparison among these methods indicated that the three multivariate geostatistical methods gave the lowest errors in rainfall estimation. Martínez-cob (1996) compared OK, OCK, and modified residual kriging to interpolate annual rainfall and evapotranspiration in Aragón, Spain. The results indicated that OCK was superior for rainfall estimation, reducing estimation error by 18.7% and 24.3% compared with OK and modified residual kriging, respectively. Hsieh et al. (2006) evaluated OK and IDW methods using daily rainfall records from 20 rain-gauges to estimate the spatial distribution of rainfall in the Shih-Men Watershed in Taiwan. The results demonstrated that IDW produced more reasonable representations than OK. Moral (2010) compared three univariate kriging (simple kriging [SK], universal kriging, and OK) with three multivariate kriging methods (OCK, SKVM, and regression kriging) to interpolate monthly and annual rainfall data from 136 rain-gauges in Extremadura region of Spain. The results showed that multivariate kriging outperformed univariate kriging and among multivariate kriging, SKVM and regression kriging performed better than OCK.

Ly, Charles, and Degré (2011) used IDW, TP, and several kriging methods to interpolate daily rainfall at a catchment scale in Belgium.

The results indicated that integrating elevation into KED and OCK did not provide improvement in the interpolation accuracy for daily rainfall estimation. OK and IDW were considered to be the suitable methods as they gave the smallest error for almost all cases. Mair and Fares (2011) compared TP, IDW, OK, linear regression, SKVM to estimate seasonal rainfall in a mountainous watershed concluding that OK provided the lowest error for nearly all cases. They also found that incorporating elevation did not improve the prediction accuracy over OK for the correlation between rainfall and elevation lower than 0.82. Delbari et al. (2013) used two univariate methods (IDW and OK), and four multivariate methods (OCK, KED, SKVM, and linear regression) for mapping monthly and annual rainfall over the Golestan Province in Iran. They reported that KED and OK outperformed all other methods in terms of root mean square error (RMSE). Jeffrey et al. (2001) derived a comprehensive archive of Australian rainfall and climate data using a thin plate smoothing spline to interpolate daily climate variables and OK to interpolate daily and monthly rainfall. The aforementioned studies on spatial interpolation of rainfall indicate that each method has its advantages and disadvantages and thus performs in a dissimilar way for different regions. There is no single interpolation method that can work well everywhere (Daly, 2006). Therefore, the best interpolator for a particular study area or catchment should essentially be achieved through the comparative assessment of different interpolation methods.

To date, many studies have been conducted on spatial interpolation of rainfall at a regional and national scale in Australia (Gyasi-Agyei, 2016; Hancock & Hutchinson, 2006; Hutchinson, 1995; Jeffrey et al., 2001; Johnson et al., 2016; Jones, Wang, & Fawcett, 2009; Li & Shao, 2010; Woldemeskel, Sivakumar, & Sharma, 2013; Yang et al., 2015). However, none of these studies was conducted at a local or catchment scale. Likewise, elevation and rainfall relations locally have been relatively little studied in Australia and no such studies have been undertaken specifically within the Yarra River catchment and the Ovens River catchment in Victoria, Australia. Sharma and Shakya (2006) highlighted that any analysis of hydroclimatic variables should be carried out at the local scale rather than at a large or global scale because of the variations of hydroclimatic situations from one region to another. Therefore, the main aim of this study is to assess if relatively more complex geostatistical interpolation methods that take into account the elevation and rainfall relation provide any benefits over simpler methods for enhanced estimation of rainfall within the Yarra River catchment and the Ovens River catchment in Victoria, Australia. The specific focus is to evaluate the effectiveness of the cokriging methods including OCK and KED that make use of elevation as a secondary variable over those methods including OK, IDW, and RBF that do not make use of such information to estimate the spatial distribution of rainfall within the catchment. This study is expected to provide an important contribution towards the enhanced estimation of rainfall in the aforementioned two Australian catchments using the cokriging methods by incorporating elevation as an auxiliary variable in addition to rainfall data. One specific contribution of this paper is in explaining how rainfall varies with elevation from catchment to catchment.

The Yarra River catchment and the Ovens River catchment in Victoria are selected for this study because they are two important water resources catchments in Australia in terms of water supply and agricultural production (Adhikary, Yilmaz, & Muttil, 2015; EPA Victoria, 2003; Schreider, Jakeman, Pittock, & Whetton, 1996; Yu, Cartwright, Braden, & de Bree, 2013). The Yarra River catchment is a major source of water for more than one third of Victoria's population in Australia (Barua, Muttil, Ng, & Perera, 2012). Although the catchment is not large with respect to other Australian catchments, it produces the fourth highest water yield per hectare of the catchment in Victoria (Adhikary et al., 2015). There are seven storage reservoirs in the catchment that supports about 70% of drinking water supply in Melbourne city (Barua et al., 2012). The Ovens River catchment is another important source of water in northeast Victoria, which forms a part of the Murray-Darling basin (Yu et al., 2013). The Ovens River is considered one of the most important tributaries of the Murray-Darling Basin due to the availability of sufficient volume of water with acceptable quality and its good ecological condition. The average annual flow of the river constitutes approximately 7.3% of the total flow of the state of Victoria (Schreider et al., 1996). The catchment contributes approximately 14% of Murray-Darling basin flows in spite of its relatively small catchment area of less than 1% of the total Murray-Darling basin area (EPA Victoria, 2003; Yu et al., 2013). Thus, both the catchments have significant contribution towards the sustainable development of Victoria's economy. However, high rainfall variation and diverse water use activities in these catchments has complicated the water management tasks, which has further created strong burden on the water managers and policy makers for effective water resources management. Therefore, enhanced estimation of rainfall and its spatial distribution is important, which could be beneficial for effective water supply and demand management, and sustainable agricultural planning in both the catchments.

The rest of the paper is arranged as follows. First, a brief description of the study area and data used are presented, which is followed by the detailed description of the methodology adopted in this study. The results are summarized next, and finally, the conclusions drawn from this study are presented.

2 | STUDY AREA AND DATA USED

2.1 | The study area

This study considers two catchments in Victoria as the case study area, which includes the Middle Yarra River catchment and Ovens River catchment in south-eastern Australia. Figure 1 shows the approximate location of the case study area. The Yarra River catchment is located in northeast of Melbourne covering an area of 4,044 km^2. The water resources management is an important and multifaceted 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 significantly contributes to the water supply in Melbourne and has been playing an important role in the way Melbourne has been developed and grown (Adhikary et al., 2015). The Yarra River catchment consists of three distinctive subcatchments including Upper, Middle, and Lower Yarra segments based on different land use activities (Barua et al., 2012). This study concentrates on the Middle segment of the Yarra River catchment, which forms part of the case study area in Figure 1. The Upper Yarra segment includes mainly the forested and mountainous areas with minimum human settlement, which is mainly used as a closed water supply

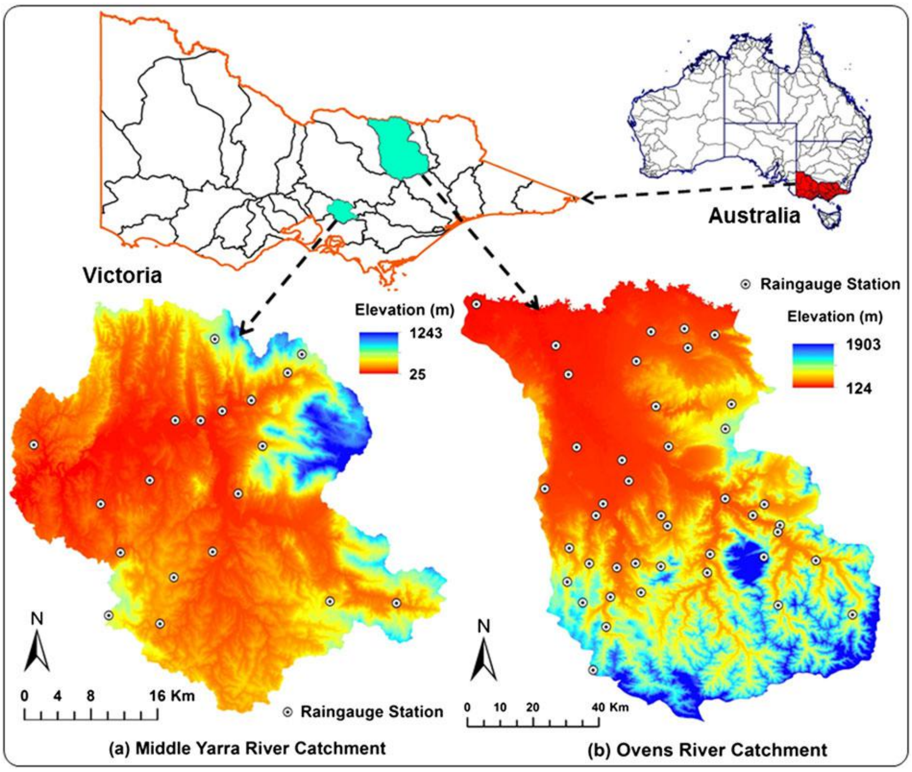

FIGURE 1 Location and topography of (a) Middle Yarra River catchment and (b) Ovens River catchment in Victoria with existing rain-gauge stations

catchment for Melbourne and has been reserved for more than 100 years for water supply purposes. The Middle Yarra segment, starting from the Warburton Gorge to 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, which is located downstream of Warrandyte, is mainly characterized by the urbanized floodplain areas of Melbourne city (Adhikary, Muttil, & Yilmaz, 2016b). Most of the land along rivers and creeks in the middle and lower segments has been cleared for the agricultural or urban development (Barua et al., 2012; Melbourne Water, 2015).

Rainfall varies significantly through different segments of the Yarra River catchment. The mean annual rainfall varies across the catchment from 600 mm in the Lower Yarra segment to 1,100 mm in the Upper Yarra segment (Daly et al., 2013). The Middle Yarra segment (part of the case study area in Figure 1) covers an area of 1,511 km^2 and consists of three storage reservoirs. Decreasing rainfall patterns in the catchment will reduce the streamflows, which in turn will lead to the reduction in reservoir inflows and hence impact the overall water availability in the catchment. Moreover, the reduced streamflows may cause increased risk of bushfires. Conversely, increasing rainfall patterns and the occurrence of extreme rainfall events (as reported in Yilmaz and Perera [2014] and Yilmaz, Hossain, and Perera [2014]) will result in excess amount of streamflows that may cause flash floods in the urbanized lower segment and makes it vulnerable and risk-prone. The urbanized lower part of the catchment is also dependent on the water supply from the storage reservoirs mainly located in the middle and upper segments of the catchment (Adhikary et al., 2015). Therefore, accurate spatial distribution of rainfall in the middle and upper segments of the catchment could be useful for accurate estimation of future streamflows for optimal reservoir operation and effective flood control in the urbanized lower part.

The Ovens River catchment in northeast Victoria is also considered as a part of the case study area in this study, which is shown in Figure 1. The catchment covers an area of 7,813 km^2 (Yu et al., 2013), which extends from the Great Dividing Range in the south to the Murray River in the north, with the Yarrawonga Weir forming the downstream boundary. It is considered to be one of the least modified catchments within the Murray-Darling basin. The catchment contributes approximately 14% to the average flows of the Murray River in spite of its relatively small size (0.75 percent of the total Murray-Darling Basin area; EPA Victoria, 2003; Yu et al., 2013). The Ovens River is the main river in the catchment, which originates on the northern edges of the Victorian Alps and flows in a north-westerly direction until its junction with the Murray River near Lake Mulwala. The riverine plains and alluvial flats are primarily cleared for agricultural use, while the hills and mountains are covered by forests with native plant species (Yu et al., 2013). Total average water use in the catchment is about 30,000 million litres per year, 64% of which is diverted from the Ovens River and its tributaries. A major part of this water use is irrigation, which constitutes more than 16,000 million litres annually (Schreider et al., 1996). The river itself provides natural conditions suitable for many significant native fish species, particularly the endangered Murray Cod (EPA Victoria, 2003). Thus, the catchment is considered to be important, not only at a regional scale, but also at the national scale in terms of its water supply volume for domestic and agricultural production, and high environmental value.

The climate of the Ovens River catchment varies considerably with topography and elevation (Yu et al., 2013). The average annual rainfall varies from 1,127 mm in the Alpine region at Bright to 636 mm on the alluvial plains in Wangaratta with most rainfall occurring in winter months (Yu et al., 2013). Approximately 45% of the annual rainfall occurs during the winter (June to September) whereas the summer is warm and dry. Winter snowfalls are common at altitudes above 1,000 m (EPA Victoria, 2003). Therefore, enhanced estimation of rainfall and its spatial distribution could be useful for the effective management of water supply and agricultural activities in the catchment.

2.2 | Data used

In this study, historical rainfall data from existing rain-gauge stations in the Middle Yarra River catchment and the Ovens River catchment (Figure 1) for 1980–2012 period are considered. There are 19 rain-gauge stations in the Middle Yarra River catchment, whereas the Ovens River catchment includes 42 rain-gauge stations operated by the Australian Bureau of Meteorology. Daily rainfall data are collected from the Scientific Information for Land Owners climate database (http://www.longpaddock.qld.gov.au/silo/) and compiled to generate monthly and annual rainfall, which are then used for the analysis. Summary statistics of monthly rainfall data are given in Table 1. The annual average rainfall in the Middle Yarra River catchment for the aforementioned period varies from 710 mm to 1422 mm with a mean value of 1082 mm. The southern and south-eastern part experiences the highest rainfall, whereas the lowest rainfall occurs in the north-western part in the study area. September is the wettest month (rainfall amount equals to 112.5 mm) with the highest variation in rainfall. The driest month is February (rainfall amount equals to 56.4 mm) with the second highest variability. On the other hand, the annual average rainfall for the same period in the Ovens River catchment varies from 231 to 2,473 mm with a mean value of 913 mm. The wettest month is July (rainfall amount equals to 112.9 mm) with the highest variability and February (rainfall amount equals to 51.3 mm) appears to be the driest month with the third highest variation in rainfall.

For the OCK analysis, a DEM of both the catchments with 10 m resolution (shown in Figure 1) is collected from the Geoscience Australia. The elevation of the Middle Yarra River catchment varies from 25 m (lowest-mainly in central, north-western, and western part) to 1,243 m (highest-mainly in northern, north-eastern, and eastern part) with a mean elevation of 621 m above mean sea level. Whereas, the elevation of the Ovens River catchment varies from 124 m (lowest-mainly in the upper north-western part) to 1903 m (highest-mainly in the lower southern, south-eastern and eastern part) with a mean elevation of 874 m above mean sea level. Monthly and annual rainfall generally tend to increase with the higher elevations caused by the orographic effect of mountainous terrain (Goovaerts, 2000). Several studies have revealed that rainfall usually shows good correlation with elevation. For example, Goovaerts (2000) showed that a good to significant correlation exists between the monthly rainfall and elevation,

TABLE 1 Summary statistics for monthly rainfall data of Middle Yarra River catchment and Ovens River catchment

Month	Mean (mm)	Minimum (mm)	Maximum (mm)	Standard deviation (mm)	Correlation coefficient (versus elevation)[a]
Middle Yarra River catchment					
January	67.3	1.8	201.4	31.23	0.79
February	56.4	0.0	269.5	53.40	0.69
March	66.3	9.2	217.8	36.89	0.77
April	84.7	15.2	246.0	45.81	0.74
May	88.3	10.2	239.7	42.47	0.67
June	106.4	13.8	300.2	46.18	0.70
July	102.1	17.9	303.9	48.95	0.73
August	108.6	21.1	289.7	50.67	0.72
September	112.5	25.3	350.3	56.98	0.67
October	101.6	4.0	333.9	50.86	0.69
November	96.1	15.3	258.7	47.60	0.77
December	91.3	8.2	301.2	52.88	0.74
Ovens River catchment					
January	55.8	0.0	364.7	51.41	0.49
February	51.3	0.0	421.9	65.66	0.26
March	53.7	0.3	418.1	51.01	0.49
April	54.5	1.0	234.6	38.63	0.72
May	74.6	2.0	360.7	56.71	0.61
June	95.4	1.4	457.0	62.99	0.65
July	112.9	6.0	622.4	71.92	0.60
August	105.2	5.8	468.0	68.94	0.61
September	86.3	5.2	484.1	55.44	0.65
October	73.4	0.1	410.8	61.61	0.67
November	73.6	0.6	290.8	47.70	0.68
December	64.1	0.2	374.6	54.79	0.68

[a]Linear correlation coefficient between rainfall and elevation data.

which varies from .33 to .83. Subyani and Al-Dakheel (2009) found that good correlation ranging from .34 to .77 exists between seasonal rainfall and elevation in the Southwest Saudi Arabia. Moral (2010) identified a good correlation ranging from .33 to .67 between monthly and annual rainfall and elevation in the southwest region of Spain.

As can be seen in Table 1, the correlation coefficient (CC) between the monthly average rainfall and elevation for the Middle Yarra River catchment varies from .67 to .79, where 8 months have the CC values greater than .70. This indicates that a strong correlation exists between the monthly rainfall and elevation in the catchment, suggesting that elevation may enhance the monthly rainfall estimates when used as a secondary variable in the OCK analysis. On the contrary, the CC between the monthly average rainfall and elevation for the Ovens River catchment varies from .26 to .72, where 6 months exhibit the CC values higher than .65. Apart from the driest month of February, the correlation ranges from .49 to .72 in the Ovens River catchment. Thus, it seems beneficial taking into account this exhaustive secondary variable (elevation in this study) into the enhanced estimation and mapping of rainfall in both the catchments. Goovaerts (1997) mentioned that use of multiple elevation data other than the colocated positions of rain-gauges can lead to unstable cokriging systems because the correlation between near elevation data is much greater than the correlation between distant rainfall data. Therefore, the colocated elevation data are used for the OCK analysis in this study, which are extracted at the same positions of rain-gauge stations from the DEM of the catchments.

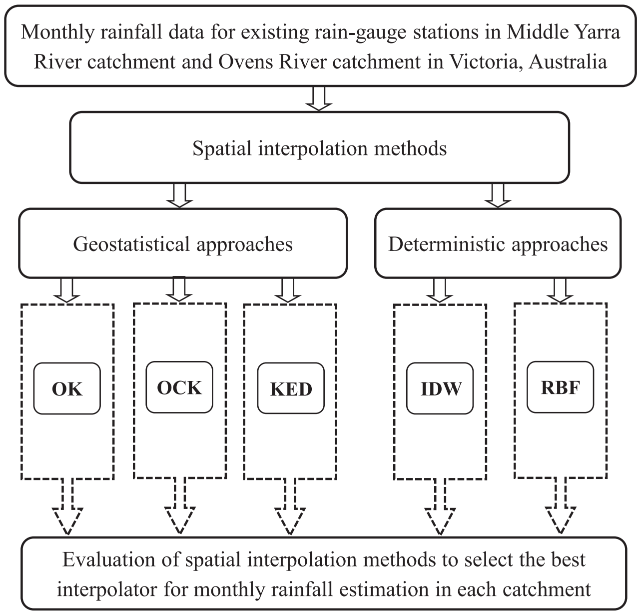

FIGURE 2 Methodological framework adopted in this study. IDW = inverse distance weighting; KED = kriging with an external drift; OCK = ordinary cokriging; OK = ordinary kriging; RBF = radial basis function

3 | METHODOLOGY

The methodological framework adopted in this study for spatial interpolation of rainfall includes three kriging-based geostatistical (OK, OCK, and KED) and two deterministic (IDW and RBF) interpolation methods, which is shown in Figure 2. A brief description of these methods is presented in this section. The variogram and its estimation technique are also summarised with each of the kriging methods because it is a key component of kriging. For a more detailed description of the methods used in the current study, readers are referred to several recent geostatistical textbooks including Journel and Huijbregts (1978); Isaaks and Srivastava (1989), Goovaerts (1997), Chilès and Delfiner (1999), Wackernagel (2003), and Webster and Oliver (2007).

3.1 | Ordinary kriging

Kriging refers to a family of generalized least-squares regression methods in geostatistics that estimate values at unsampled locations using the sampled observations in a specified search neighborhood (Goovaerts, 1997; Isaaks & Srivastava, 1989). OK is a geostatistical interpolation method based on spatially dependent variance, which gives unbiased estimates of variable values at target location in space using the known sampling values at surrounding locations. The unbiasedness in the OK estimates is ensured by forcing the kriging weights to sum to 1. Thus, the OK estimator may be stated as a linear combination of variable values, which is given by

$$\widehat{Z}_{OK}(s_0) = \sum_{i=1}^{n} \omega_i^{OK} Z(s_i) \qquad \text{with} \quad \sum_{i=1}^{n} \omega_i^{OK} = 1 \tag{1}$$

where $\widehat{Z}_{OK}(s_0)$ is the estimated value of variable Z (i.e., rainfall) at target (at which estimation is to be made) unsampled location s_0; ω_i^{OK} indicates the OK weights linked with the sampled location s_i with respect to s_0; and n is the number of sampling points used in estimation. While giving the estimation at target location, OK provides a variance measure to signify the reliability of the estimation.

OK is known as the best linear unbiased estimator (Isaaks & Srivastava, 1989). It is linear in the sense that it gives the estimation based on the weighted linear combinations of observed values. It is best in the sense that the estimate variance is minimized while interpolating the unknown value at desired location. And it is unbiased because it tries to have the expected value of the residual to be zero (Adhikary et al., 2016b). The weight constraint in Equation 1 ensures the unbiased estimation in OK. For OK, the kriging weights are determined to minimize the estimation variance and ensure the unbiasedness of the estimator.

The OK weights ω_i^{OK} can be obtained by solving a system of $(n + 1)$ simultaneous linear equations as follows:

$$\begin{aligned} &\sum_{i=1}^{n} \gamma(s_i - s_j)\omega_i^{OK} + \mu_1^{OK} = \gamma(s_j - s_0) \qquad for\ \ j = 1, \ldots\ldots\ldots, n \\ &\sum_{i=1}^{n} \omega_i^{OK} = 1 \end{aligned} \tag{2}$$

where $\gamma(s_i - s_j)$ is the variogram values between sampling locations s_i and s_j, $\gamma(s_j - s_0)$ is the variogram values between sampling location s_j and the target location, s_0, and μ_1^{OK} is the Lagrange multiplier parameter. Equation 2 indicates that OK highly depends on a mathematical function called variogram model that indicates the degree of spatial autocorrelation in datasets.

For OK interpolation of variables, first an experimental variogram $\hat{\gamma}(d)$ is derived by

$$\hat{\gamma}(d) = \frac{1}{2N(d)} \sum_{i=1}^{N(d)} [Z(s_i + d) - Z(s_i)]^2 \tag{3}$$

where $Z(s_i)$ and $Z(s_i + d)$ are the variable values at corresponding sampling locations s_i and $(s_i + d)$, respectively, located at d distance apart and $N(d)$ is the number of data pairs. A variogram cloud is initially generated using Equation 3 for observations at any two data points, in which all semivariance values are plotted against their separation distance. The experimental variogram is computed from the variogram cloud by subdividing it into a number of lags and taking an average of each lag interval (Johnston, VerHoef, Krivoruchko, & Lucas, 2001; Robertson, 2008). A variogram model $\gamma(d)$ is then fitted to the experimental variogram. A typical variogram cloud based on Equation 3 and a typical experimental variogram with a typical fitted model is shown in Figure 3.

Exponential, Gaussian, and spherical are the most commonly used variogram models for kriging applications in hydrology (Adhikary et al., 2015), which are also used to model the experimental variogram. The

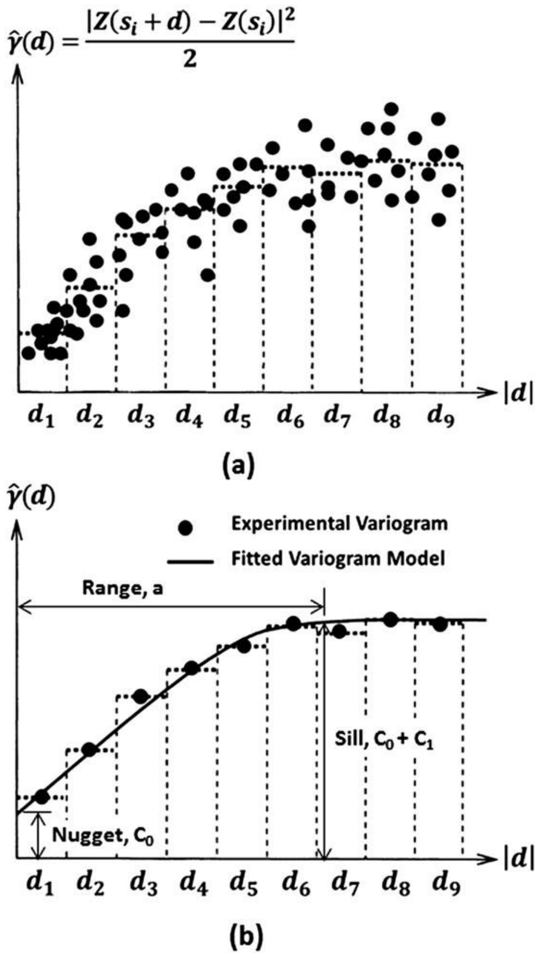

FIGURE 3 (a) a typical variogram cloud for a finite set of discrete lags, and (b) a typical experimental variogram based on the variogram cloud fitted by a typical variogram model with its parameters

functional forms these variogram models are given in Table 2. The three models are fitted to the experimental variogram using regression by noting the residual sum of squares (RSS) between the experimental $\hat{\gamma}(d_k)$ and modelled $\gamma(d_k)$ variogram values (Mair & Fares, 2011) with a trial-and-error approach for different lag sizes and lag intervals (Goovaerts, 1997) such that the RSS is minimum. RSS is given by

$$\text{RSS} = \sum_{k=1}^{K} [\hat{\gamma}(d_k) - \gamma(d_k)]^2 \tag{4}$$

RSS in Equation 4 provides an exact measure of how well the variogram model fits the experimental variogram (Robertson, 2008). Lag sizes and number of lags are varied based on a general rule of thumb, in which the lag size times the number of lags should be about half of the largest distance among all data pairs in the variogram cloud (Johnston et al., 2001, p. 66). The variogram parameters (nugget, sill, and range) are also iteratively changed to obtain the best fitted model. The model with its corresponding parameters that minimizes RSS is selected as the best variogram model and finally used in OK analysis. Variogram model fitting is performed using GS+ geostatistical software (Robertson, 2008) and OK is implemented through ArcGISv9.3.1 software (ESRI, 2009) and its geostatistical analyst extension (Johnston et al., 2001).

3.2 | Ordinary cokriging

OCK method is a modification of the OK method. The key advantage of OCK is that it can make use of more than one variable rather than using only a single variable in the estimation process. The OCK method is used to enhance the estimation of primary variable by using secondary variable assuming that the variables are correlated to each other (Isaaks & Srivastava, 1989). In this study, rainfall and elevation are considered, respectively, as the primary and secondary variables in the OCK method. Like OK method, the aim of the OCK method is to estimate the primary variable. The OCK estimator (Goovaerts, 1997) considering one secondary variable (i.e., elevation), which is cross-correlated with the primary variable (i.e., rainfall) may be written as

$$\hat{Z}_{OCK}(s_0) = \sum_{i_1=1}^{n} \omega_{i_1}^{OCK} Z(s_{i_1}) + \sum_{i_2=1}^{m} \omega_{i_2}^{OCK} V(s_{i_2}) \tag{5}$$

$$\text{with} \sum_{i_1=1}^{n} \omega_{i_1}^{OCK} = 1; \sum_{i_2=1}^{m} \omega_{i_2}^{OCK} = 0$$

where $\hat{Z}_{OCK}(s_0)$ is the estimated value of primary variable at target unsampled location s_0, $\omega_{i_1}^{OCK}$ and $\omega_{i_2}^{OCK}$ are the kriging weights associated with the sampling locations of the primary and secondary variables Z

TABLE 2 Commonly used positive-definite variogram models

Model name	Model equation
Exponential	$\gamma(d) = C_0 + C_1\left[1 - \exp\left(-\frac{3d}{a}\right)\right]$
Gaussian	$\gamma(d) = C_0 + C_1\left[1 - \exp\left(-\frac{3d^2}{a^2}\right)\right]$
Spherical	$\gamma(d) = C_0 + C_1\left[\frac{3}{2}\left(\frac{d}{a}\right) - \frac{1}{2}\left(\frac{d^3}{a^3}\right)\right], \quad d<a$ $= C_0 + C_1, \quad d \geq a$

C_0 = nugget coefficient, $C_0 + C_1$ = Sill, a = range of variogram model.
d = distance of separation between two locations.

and V, respectively, n and m are the number of sampling points for the primary and secondary variables.

The OCK weights are obtained by solving a system of (n + 2) simultaneous linear equations (Goovaerts, 1997) that can be given by

$$\begin{aligned}
&\sum_{i_1=1}^{n}\gamma_{zz}(s_{i_1}-s_{j_1})\omega_{i_1}^{OCK}+\sum_{i_2=1}^{m}\gamma_{zv}(s_{i_2}-s_{j_1})\omega_{i_2}^{OCK}+\mu_1^{OCK}=\gamma_{zz}(s_{j_1}-s_0) \quad \text{for } j_1=1,\ldots\ldots,n\\
&\sum_{i_1=1}^{n}\gamma_{vz}(s_{i_1}-s_{j_2})\omega_{i_1}^{OCK}+\sum_{i_2=1}^{m}\gamma_{vv}(s_{i_2}-s_{j_2})\omega_{i_2}^{OCK}+\mu_2^{OCK}=\gamma_{vz}(s_{j_2}-s_0) \quad \text{for } j_2=1,\ldots\ldots,m\\
&\sum_{i_1=1}^{n}\omega_{i_1}^{OCK}=1\\
&\sum_{i_2=1}^{m}\omega_{i_2}^{OCK}=0
\end{aligned} \tag{6}$$

where $\gamma_{zv}(s_{i_2}-s_{j_1})$ and $\gamma_{vz}(s_{i_1}-s_{j_2})$ are the cross-variogram values between sampled Z and V values, and μ_1^{OCK} and μ_2^{OCK} are the Lagrange multiplier parameters accounting for the two unbiased conditions.

The elementary step in the OCK method is to establish an appropriate model for cross continuity and dependency between the primary (rainfall) and secondary (elevation) variable. This positive correlation between variables is referred to as the cross-regionalization or coregionalization (Goovaerts, 1997; Wackernagel, 2003), which can be quantified by cross-variogram or cross-covariance. These models are used to define the cross continuity and dependency between two variables in the OCK method (Subyani & Al-Dakheel, 2009). The cross-variogram models between the primary (rainfall) and secondary (elevation) variables in the OCK method are obtained by fitting with an experimental cross-variogram that is given by

$$\widehat{\gamma}_{zv}=\widehat{\gamma}_{vz}=\frac{1}{2N(d)}\sum_{i=1}^{N(d)}[Z(s_i+d)-Z(s_i)]\times[V(s_i+d)-V(s_i)] \tag{7}$$

It is important to note that the variogram models must satisfy the positive-definite condition (PDC) in kriging. For a single variable (rainfall) in the OK method, the condition is satisfied by using the positive-definite variogram model functions given in Table 2. However, the OCK method considering two variables (rainfall and elevation) consists of one cross-variogram and two direct variograms, and additional requirement for satisfying the PDC (Goovaerts, 1999). In order to make sure that the cross-variogram model is positive-definite (all eigenvalues are positive), an indicator called the Cauchy-Schwarz inequality (Journel & Huijbregts, 1978; Phillips et al., 1992) must be satisfied for all distance values d, which is given by

$$\gamma_{zv}(d)\le[\gamma_{zz}(d)\gamma_{vv}(d)]^{\frac{1}{2}} \tag{8}$$

where $\gamma_{zv}(d)$ is the cross-variogram model between primary and secondary variables, and $\gamma_{zz}(d)$ and $\gamma_{vv}(d)$ are the direct variogram models for primary and secondary variables, respectively. Based on the indicator shown in Equation 8, Hevesi et al. (1992) suggested a graphical test of PDC for the fitted models as follows:

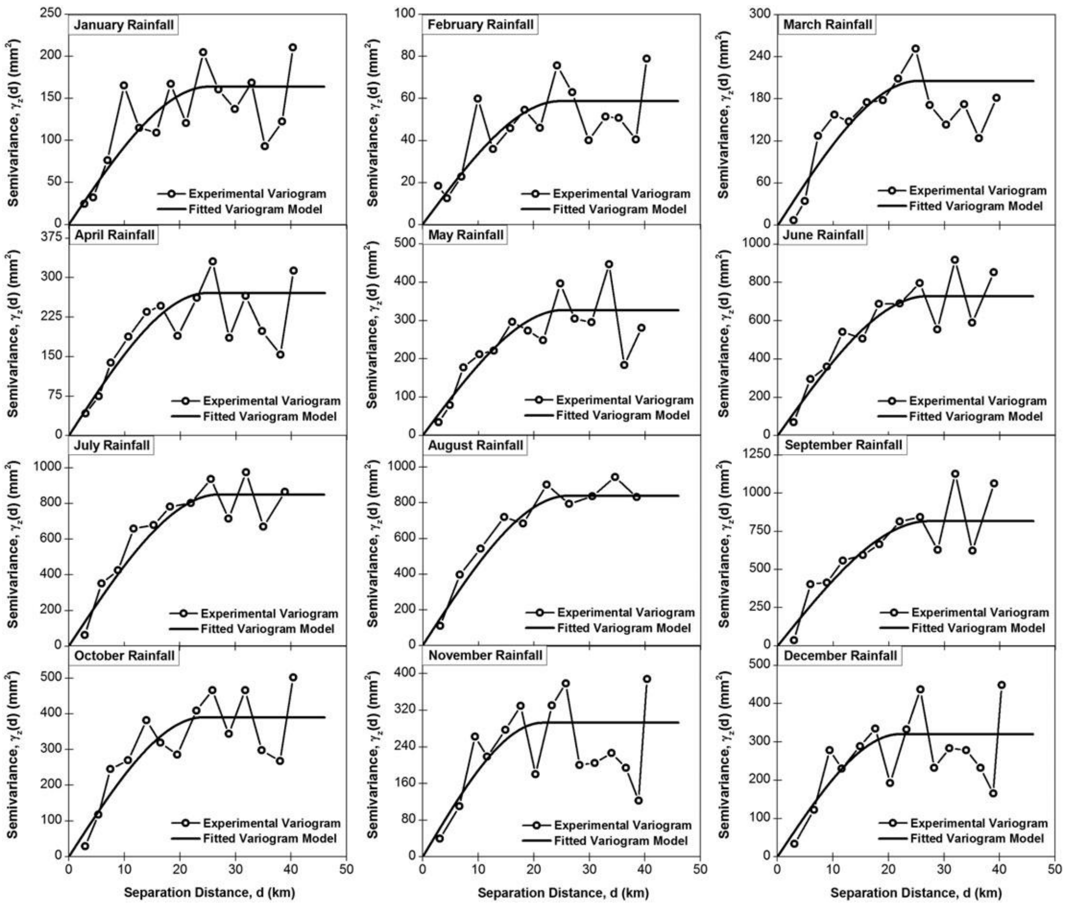

FIGURE 4 Experimental variograms and fitted variogram models for the monthly rainfall data of the Middle Yarra River catchment used in the ordinary kriging interpolation method

$$\mathrm{PDC} = [\gamma_{zz}(d)\gamma_{vv}(d)]^{\frac{1}{2}} \tag{9}$$

The model is considered positive-definite if the absolute value of the cross-variogram model $\gamma_{zv}(d)$ in Equation 8 is less than the slope and corresponding absolute value of PDC curve in Equation 9 for all distance values d. In the OCK method, direct and cross-variogram models are fitted as linear combination of the same set of basic models given in Table 2 such that the RSS value by Equation 4 is minimum under the requirement of PDC. Variogram model fitting is performed using GS+ geostatistical software (Robertson, 2008) and OCK is implemented through ArcGISv9.3.1 software (ESRI, 2009) and its geostatistical analyst extension (Johnston et al., 2001).

3.3 | Kriging with an external drift

KED is a particular type of universal kriging that gives the estimation of a primary variable Z, known only at a small number of locations in the study catchment, through a secondary variable V, exhaustively known in the same area (Feki et al., 2012). The trend or local mean of the primary variable is first derived using the secondary variable (Goovaerts, 1997; Wackernagel, 2003) and then simple kriging is carried out on residuals from the local mean. The KED estimator (Wackernagel, 2003) is generally given as

$$\widehat{Z}_{KED}(s_0) = \sum_{i=1}^{n} \omega_i^{KED} Z(s_i) \qquad \text{with} \quad \sum_{i=1}^{n} \omega_i^{KED} = 1 \tag{10}$$

The KED weights ω_i^{KED} can be obtained by solving a system of (n + 2) simultaneous linear equations as follows:

$$\begin{aligned} &\sum_{i=1}^{n} \gamma_R(s_i - s_j)\omega_i^{KED} + \mu_0^{KED} + \mu_1^{KED} V(s_j) = \gamma_R(s_j - s_0) \quad \text{for } j = 1, \ldots\ldots, n \\ &\sum_{i=1}^{n} \omega_i^{KED} = 1 \\ &\sum_{i=1}^{n} \omega_i^{KED} V(s_i) = V(s_0) \end{aligned} \tag{11}$$

where $\gamma_R(s_i - s_j)$ is the residual variogram values between sampling locations s_i and s_j, $\gamma_R(s_j - s_0)$ is the residual variogram values between sampling location s_j and the target location, s_0, and μ_0^{KED} and μ_1^{KED} are the Lagrange multiplier parameters.

OCK and KED differ in the way the secondary variable V is used. The secondary variable (e.g., elevation in this study) gives only the trend information in KED, whereas estimation with OCK is directly influenced by it (Delbari et al., 2013). In case of KED, the primary and secondary variables should exhibit a linear relationship. In addition, estimation with KED requires the secondary variable values at all the

TABLE 3 Results of fitted variogram models for monthly rainfall data for using in the OK interpolation method

Month	Model name	Variogram parameters			Cross-validation statistics	
		Nugget, C_0 (mm^2)	Sill, $C_0 + C_1$ (mm^2)	Range, a (km)	SM	SRMS
Middle Yarra River catchment						
January	Spherical	0.15	163.65	25.20	0.068	0.998
February	Spherical	0.20	58.70	24.75	0.059	1.000
March	Spherical	0.00	205.38	25.25	0.059	0.863
April	Spherical	1.25	270.75	24.85	0.054	0.897
May	Spherical	1.10	327.20	25.25	0.039	0.916
June	Spherical	0.10	727.60	26.85	0.041	0.847
July	Spherical	0.10	849.60	27.00	0.031	0.795
August	Spherical	4.00	839.00	26.11	0.035	0.851
September	Spherical	1.00	816.00	27.07	0.045	0.883
October	Spherical	0.00	390.00	24.18	0.049	0.903
November	Spherical	0.00	292.50	21.53	0.058	0.861
December	Spherical	0.00	320.00	22.00	0.060	0.863
Ovens River catchment						
January	Spherical	92.00	455.00	109.50	-0.048	0.992
February	Spherical	83.00	360.00	70.00	0.007	1.024
March	Spherical	85.00	501.00	107.10	-0.021	0.991
April	Gaussian	93.00	793.00	162.29	-0.060	1.000
May	Spherical	109.00	1016.00	108.60	-0.047	1.016
June	Spherical	50.00	1240.00	90.00	-0.039	1.001
July	Gaussian	180.00	2030.00	80.50	-0.083	0.992
August	Spherical	58.00	1420.00	65.00	-0.040	0.990
September	Spherical	25.00	937.00	67.00	-0.028	0.997
October	Spherical	102.00	519.00	75.40	-0.027	1.001
November	Spherical	19.00	430.00	68.00	0.001	0.988
December	Spherical	35.00	296.00	89.10	-0.028	0.988

Note. OK = ordinary kriging; SM = standardized mean error; SRMS = standardized root mean square error.

estimation grid nodes as well as all the sampling locations s_i. The residual variogram models are fitted based on the basic models in Table 2 such that the RSS value between the experimental and modelled variogram values by Equation 4 is minimum. Variogram model fitting and estimation with KED are performed using GS+ geostatistical software (Robertson, 2008).

3.4 | Inverse distance weighting

IDW interpolation method (ASCE, 1996) gives a linear weighted average of several neighbouring observations to estimate the variable value at target location. This method assumes that each observation point has local influence that diminishes with distance. IDW assigns greater weights to observation points near to the target location, and the weights diminish as a function of distance (Johnston et al., 2001). The estimation by IDW can be written as

$$\widehat{Z}(s_0) = \sum_{i=1}^{n} \omega_i Z(s_i) \qquad where \quad \omega_i = d_{i0}^{-k} / \sum_{i=1}^{n} d_{i0}^{-k} \tag{12}$$

where $\widehat{Z}(s_0)$ is the estimated rainfall value at desired location s_0, $Z(s_i)$ is the Z value at location s_i, ω_i is the weight assigned to observation points, d_{i0} is the distance between the sampling point at locations s_i and s_0, n is the number of sampling points, and k is a power, which is referred to as a control parameter.

As "k" approaches zero and the weights becomes more similar, IDW estimates approach the simple average of the surrounding observations. However, the effect of the farthest observations on the estimated value is diminished with the increase of k. The value of k ranges from 1 to 6 (Teegavarapu & Chandramouli, 2005). Several studies have investigated with variations in a power to examine its effects on the spatial distribution of information from rainfall observations (Chen & Liu, 2012). Therefore, k value is varied in the range of 1 to 6 with an increment of 0.1 in the current study. The optimal k value is selected based on the lowest RMSE value between the observed and estimated values. All rain-gauge stations are considered in the search neighbourhood in the estimation process. IDW interpolation is performed by ArcGISv9.3.1 software (ESRI, 2009) and its geostatistical analyst extension (Johnston et al., 2001).

3.5 | Radial basis function

RBF (Chilès & Delfiner, 1999) is an exact interpolation method, which stands for a diverse group of interpolation methods. The RBF estimator can be viewed as a weighted linear function of distance from grid point to data point plus a bias factor, which is given by

$$\widehat{Z}(s_0) = \sum_{i=1}^{n} \omega_i \varnothing(\|s_i - s_0\|) + \mu \tag{13}$$

where $\varnothing(r)$ is the radial basis function ($r = \|s_i - s_0\|$), r is the radial distance from target point s_0 to sampling points s_i, ω_i are the weights and μ is the Lagrangian multiplier. The weights are obtained by solving of a system of (n + 1) simultaneous linear equations.

The basis kernel functions in the RBF method are analogous to variograms in kriging, which makes it similar to geostatistical interpolation methods. However, RBF does not have the advantage of a prior analysis of spatial correlation unlike kriging. When interpolating a grid node, the basis kernel functions define the optimal set of weights to be used with the sampling points. There are several radial basis functions available (Johnston et al., 2001) However, thin plate spline is the most commonly used radial basis function for interpolation (e.g., Boer, de Beurs, & Hartkamp, 2001; Di Piazza et al., 2011; Hutchinson, 1995). In this study, thin plate spline is also used as the radial basis function, which is given by

$$\varnothing(r) = (cr)^2 \ln(cr) \tag{14}$$

where c is the smoothing parameter, which is obtained through cross-validation process. The optimal value of the smoothing parameter is selected based on the lowest RMSE value between the observed and estimated values. RBF interpolation method is performed by ArcGISv9.3.1 software (ESRI, 2009) and its geostatistical analyst extension (Johnston et al., 2001).

3.6 | Assessment of interpolation methods

The performance of different interpolation methods (OK, OCK, KED, IDW, and RBF) used in this study are evaluated and compared through cross-validation process. The cross-validation is a simple leave-one-out

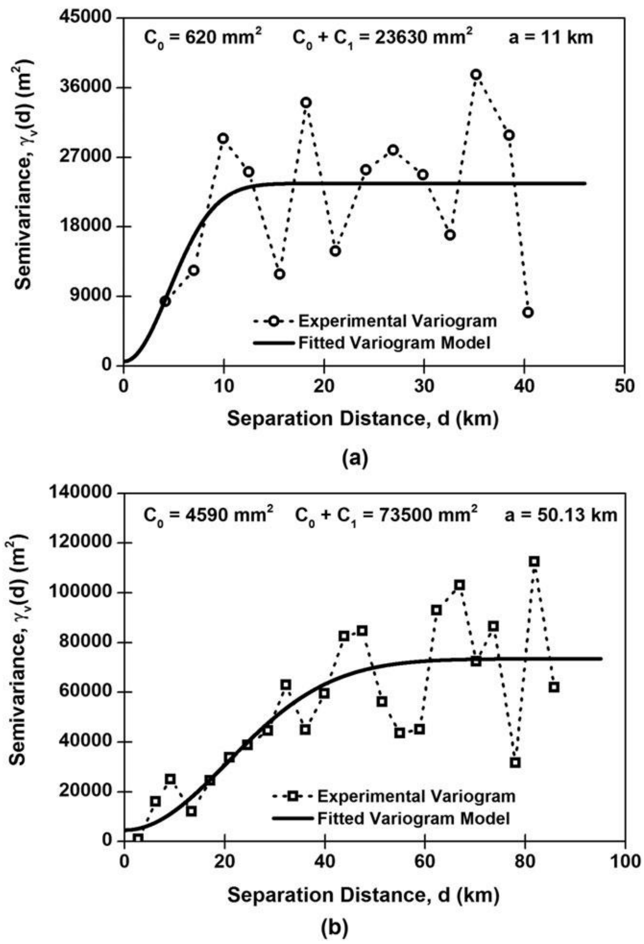

FIGURE 5 Experimental variograms and fitted variogram models based on the collocated elevation data for (a) Middle Yarra River catchment, and (b) Ovens River catchment

validation procedure (Haddad, Rahman, Zaman, & Shrestha, 2013) in which observations are removed one at a time from the dataset and then re-estimated from the remaining observations using the adopted model. Cross-validation provides important evidence of the performance measures for the interpolation methods. In this study, the performance of all the interpolation methods for rainfall estimation is compared based on mean bias error (MBE), RMSE, and coefficient of determination (R^2) values between the observed and estimated rainfall values, which are given by Equations 15–17.

$$\mathrm{MBE} = \frac{1}{n}\sum_{i=1}^{n}\left[Z(s_i) - \widehat{Z}(s_i)\right] \tag{15}$$

$$\mathrm{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left[Z(s_i) - \widehat{Z}(s_i)\right]^2} \tag{16}$$

$$R^2 = \frac{\left[\sum_{i=1}^{n}\{Z(s_i) - Z_m\}\left\{\widehat{Z}(s_i) - \widehat{Z}_m\right\}\right]^2}{\sum_{i=1}^{n}\{Z(s_i) - Z_m\}^2\sum_{i=1}^{n}\left\{\widehat{Z}(s_i) - \widehat{Z}_m\right\}^2} \tag{17}$$

In Equations 15–17, $Z(s_i)$ and $\widehat{Z}(s_i)$ are the observed and predicted values, Z_m and $\widehat{Z}_m$ are the mean of the observed and predicted values, and n is the number of sampled data points. The interpolation method with the lowest MBE and RMSE values and the highest R^2 value is chosen as the best interpolation method.

As has been mentioned earlier, kriging gives the prediction standard error while giving the estimation of unsampled variables, the adequacy of the variogram model for kriging and cokriging estimation should also be tested to produce correct interpolation results (Johnston et al., 2001; Phillips et al., 1992). Therefore, two additional standardized cross-validation statistics are used in this study, which are standardized mean error (SM) and standardized root mean square error (SRMS) as given by Equations 18–19

$$\mathrm{SM} = \frac{1}{n}\sum_{i=1}^{n}\frac{\left[Z(s_i) - \widehat{Z}(s_i)\right]}{\widehat{\sigma}(s_i)} \tag{18}$$

$$\mathrm{SRMS} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left[\frac{Z(s_i) - \widehat{Z}(s_i)}{\widehat{\sigma}(s_i)}\right]^2} \tag{19}$$

where $\widehat{\sigma}(s_i)$ is the prediction standard error for location s_i. SM should be close to 0 if the estimates using the adopted variogram model are unbiased. SRMS should be close to 1 if the estimation variances are consistent and the variability of the prediction is correctly assessed (Adhikary et al., 2015; Johnston et al., 2001).

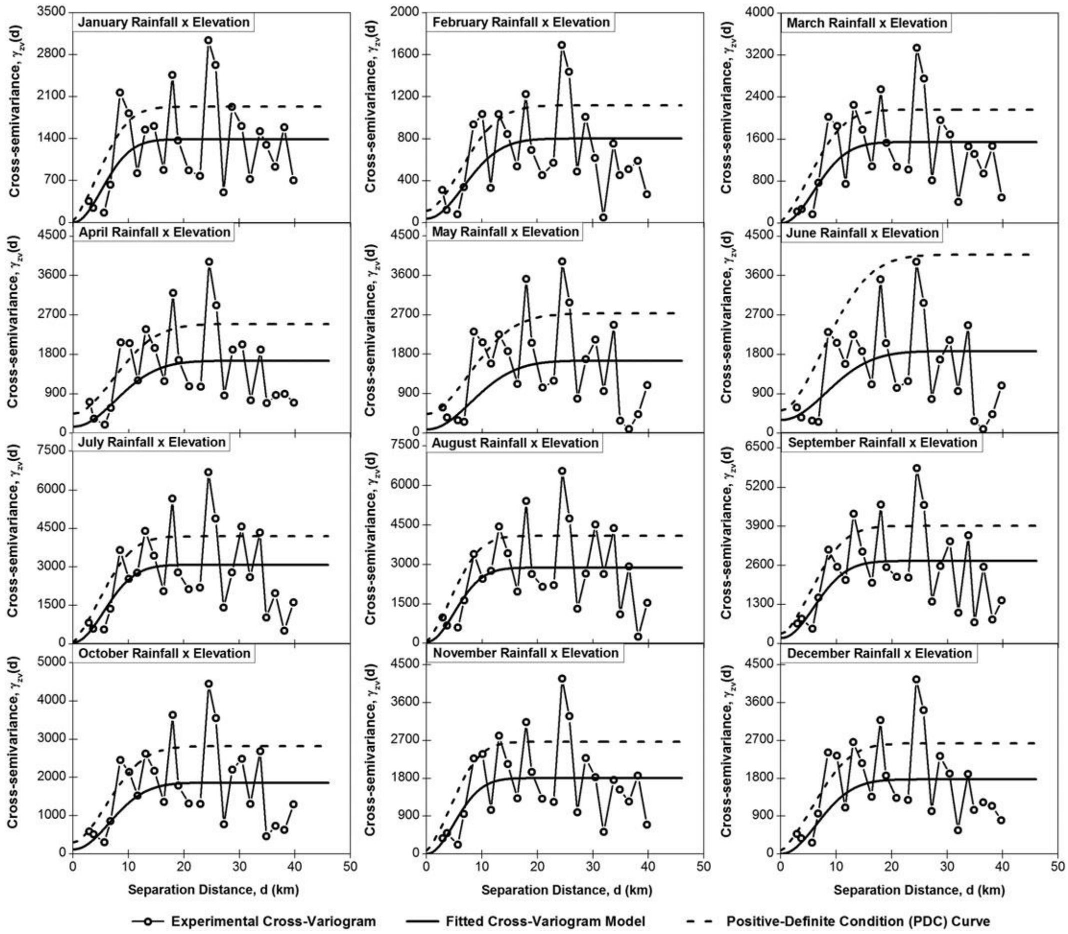

FIGURE 6 Experimental cross-variograms with the fitted cross-variogram models and positive-definite condition curve based on the monthly rainfall and collocated elevation data for the Middle Yarra River catchment used in the ordinary cokriging interpolation method

4 | RESULTS AND DISCUSSION

4.1 | Variogram models for OK analysis

The OK analysis requires the estimation of the direct variogram models for rainfall data. In this study, an isotropic experimental variogram is estimated from the rainfall dataset for each month assuming an identical spatial correlation in all directions and neglecting the influence of anisotropy on the variogram parameters. Isotropy is assumed for the methodological simplicity. Isotropy is a property of a natural process or data where directional influence is considered insignificant and spatial dependence (autocorrelation) changes only with the distance between two locations (Johnston et al., 2001). Under the isotropic condition, the semivariance is assumed the same for a given distance, regardless of direction. Initially, the directional experimental variograms are estimated from each monthly rainfall dataset. However, the directional variograms are found noisy because of the less number of rain-gauge stations in the study area. Therefore, the directional influence is ignored in the experimental variogram calculation. The experimental variogram is then fitted with three predefined variogram model functions (exponential, Gaussian, and spherical as given in Table 2) to obtain the variogram models for each monthly rainfall dataset.

For convenience in this study, results obtained for the Middle Yarra River catchment are presented and discussed elaborately and compared with that obtained for the Ovens River catchment. Figure 4 shows the experimental variograms and fitted variogram models with optimal variogram parameters (i.e., nugget, sill, and range) for monthly rainfall data of the Middle Yarra River catchment, which are used in OK analysis. The best fitted variogram models are selected based on the minimum RSS values using a trial-and-error process with different lag intervals. The variogram parameters are iteratively changed to get the best fitted model, which yields the minimum RSS. As can be seen from Figure 4, the spherical model is the best fitted variogram model for all months having spatial structure of $.66 < R < .97$ (results of R not shown in Figure). It can be also noted that 10 months (except November and December) have R values greater than .75. The optimal variogram parameters for each monthly rainfall dataset for both the catchments are provided in Table 3. As can be seen from Table 3, the ratio of nugget coefficient to sill of the variogram model is small for all months for both the catchments. This evidently indicates that a strong spatial correlation exists between the monthly mean rainfall and the spatial distribution of the rain-gauge stations over the study area. This supports the use of geostatistical interpolation methods such as OK, OCK, and KED, which consider

TABLE 4 Results of fitted cross-variogram models between monthly rainfall and elevation for using in the OCK interpolation method

Month	Model name	Variogram parameters			Cross-validation statistics	
		Nugget, C_0 (mm^2)	Sill, $C_0 + C_1$ (mm^2)	Range, a (km)	SM	SRMS
Middle Yarra River catchment						
January	Gaussian	1.0	1387.0	12.65	0.053	1.001
February	Gaussian	40.0	801.0	16.19	0.067	1.000
March	Gaussian	1.0	1545.0	14.36	0.055	1.006
April	Gaussian	150.0	1653.0	18.35	0.057	1.004
May	Gaussian	80.0	1651.0	19.40	0.042	1.019
June	Gaussian	300.0	1874.5	19.75	0.055	1.001
July	Gaussian	55.0	3081.2	13.50	0.008	1.002
August	Gaussian	57.5	2872.9	12.10	0.007	1.006
September	Gaussian	179.4	2746.0	14.25	0.042	1.002
October	Gaussian	114.5	1860.9	15.85	0.053	1.004
November	Gaussian	3.4	1804.4	12.00	0.034	1.000
December	Gaussian	1.7	1774.7	15.66	0.096	1.002
Ovens River catchment						
January	Gaussian	485.2	2197.9	60.98	-0.020	0.989
February	Gaussian	525.8	2540.3	50.56	0.001	1.119
March	Gaussian	635.1	2349.5	52.45	-0.012	1.093
April	Gaussian	373.7	2209.3	36.01	-0.019	0.904
May	Gaussian	595.0	3170.7	32.99	-0.015	1.090
June	Gaussian	342.6	3961.7	52.48	-0.022	1.117
July	Gaussian	652.2	4175.6	71.62	-0.030	1.028
August	Gaussian	919.2	4223.1	96.48	-0.024	1.095
September	Gaussian	733.6	3717.8	66.49	-0.018	1.000
October	Gaussian	256.0	2048.1	29.73	-0.010	1.010
November	Gaussian	413.5	1925.5	64.09	-0.013	1.050
December	Gaussian	123.5	1935.7	32.60	-0.021	1.007

Note. OCK = ordinary cokriging; SM = standardized mean error; SRMS = standardized root mean square error.

the spatial correlation in the estimation process. The range of influence and the sill of the variogram model vary from one month to another, but the variogram exhibit a same spherical structure in all months. This may be caused due to the control of the relief on the spatial distribution of rainfall (Delbari et al., 2013). The range of influence is lowest for November (21.53 km) and highest for September (27.07 km). Furthermore, the cross-validation statistics in Table 3 for both the catchments confirm that the fitted variogram models for all monthly rainfall data satisfy the unbiased condition and thus can be used for the OK analysis.

For elevation data, an isotropic experimental variogram is computed ignoring the directional influence. The experimental variogram is then fitted with the aforementioned three variogram model functions. The best fitted variogram model is selected using the same procedure described above. The Gaussian variogram model gives the best fitted model for both the catchments with the lowest RSS value, which is shown in Figure 5. The optimal variogram parameters for both the catchments are also shown in the figure. In order to avoid possible inconsistencies in the subsequent modelling of direct and cross-variograms in OCK analysis (Goovaerts, 1997, 2000), the colocated elevation data (see Figure 1) are used for estimating the variogram of elevation, not the entire DEM of the catchment. Therefore, the cokriging method adopted in this study is referred to as the colocated OCK method (Wackernagel, 2003).

4.2 | Cross-variogram models for OCK analysis

The OCK analysis requires the simultaneous estimation of the direct and cross-variogram models for the rainfall and elevation variables. The three variogram models are fitted as a linear combination of the same set of standard models given in Table 2 so that the RSS value is minimum under the constraints of PDC (Goovaerts, 1999). Figure 6 shows the experimental and fitted cross-variogram models for the Middle Yarra River catchment. The number of data pairs in each lag size is the same for all the three direct and cross-variogram models. Figure 6 also shows the PDC curve computed based on Equation 9 to examine the positive-definiteness criteria of the cross-variogram models obtained for the catchment. Additionally, the cross-validation statistics are used for identifying the adequacy and final selection of the adopted cross-variogram model for the OCK analysis. The cross-validation results obtained using all the adopted cross-variogram models for both the catchments are presented in Table 4. The results in Table 4 indicate that the cross-variogram models of all monthly datasets are suitable for the OCK analysis considering all neighbourhoods for both the catchments.

As can be seen from Figure 6, the Gaussian variogram model fits well for all monthly datasets of the Middle Yarra River catchment, which also satisfy the PDC criteria defined by Equation 9. Furthermore, the correlation between monthly rainfall and elevation for all

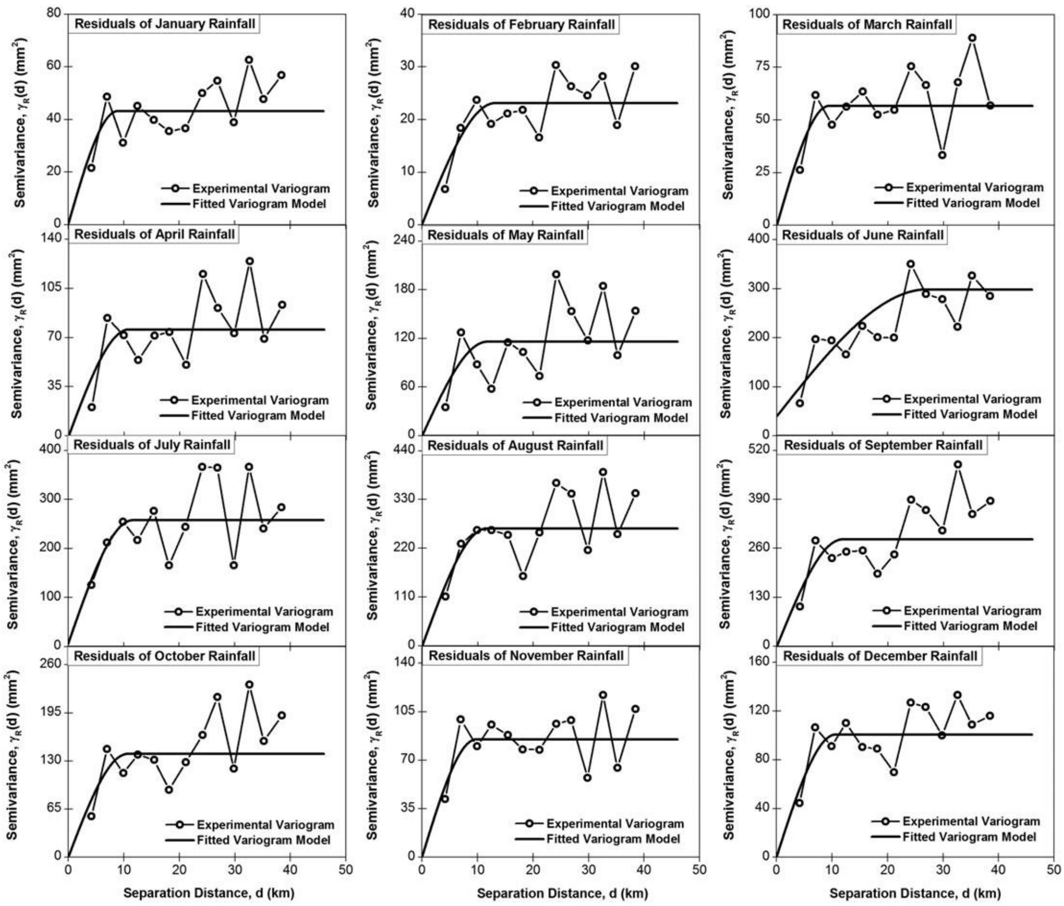

FIGURE 7 Experimental residual variograms and fitted residual variogram models for the Middle Yarra River catchment used in the kriging with an external drift interpolation method

months in Table 1 indicates that elevation will contribute to enhance the monthly rainfall estimation in the catchment. The figure also shows that the values of the sample cross-variogram increase for distances from 0 to 25 km (more than half of the maximum interstation distance) for almost all months. This indicates that a positive spatial cross-correlation exists between rainfall and elevation in the catchment. This wide ranges may be due to the high correlation (.67 < R < .79) between the monthly rainfall and elevation (Table 1). Such high correlation confirms that the monthly rainfall in the Middle Yarra River catchment is mainly caused by the orographic effects.

Figure 6 also shows the PDC curves, which are computed to examine the positive-definite conditions of the cross-variogram models of the catchment. It is worth pointing out that the PDC curve may give a qualitative indication for the degree of correlation. As can be observed from Figure 6, the plotted PDC curve for most of the months showed a close fit to the cross-variogram model for smaller distances with few exceptions (February, April, May, and June months). For example, the PDC curve is closer to the cross-variogram model in the case of January, March, July, November, and December months depending on the degree of correlation. This conclusion holds true based on the higher correlation for these months as given in Table 1.

4.3 | Residual variogram models for KED analysis

In order to implement the KED analysis, experimental residual variograms are estimated based on the residuals obtained from linear regression between rainfall and elevation data neglecting the influence of anisotropy on the variogram parameters. The experimental residual variogram is then fitted using the three standard models given in Table 2. Figure 7 shows the experimental and fitted residual variogram models for all monthly datasets of the Middle Yarra River catchment. It can be seen from the figure that the spherical model gives the best fitted model for all monthly datasets. The optimal variogram parameters and the corresponding cross-validation statistics of the selected residual variogram models for both the catchments are presented in Table 5. As can be also seen from Figure 7 and Table 5, the residual variogram models exhibit relatively smaller sills than those obtained from the actual rainfall datasets (see Figure 4) but they follow very similar structure. This is not unexpected because the residual variograms from the linear regression represents variation, which remains after removing the trend (Lloyd, 2005). The cross-validation statistics shown in Table 5 also indicate that the residual variogram models of all monthly datasets for both the catchments are satisfactory for the KED analysis.

TABLE 5 Results of fitted residual variogram models for using in the KED interpolation method

Month	Model name	Variogram parameters			Cross-validation statistics	
		Nugget, C_0 (mm^2)	Sill, $C_0 + C_1$ (mm^2)	Range, a (km)	SM	SRMS
Middle Yarra River catchment						
January	Spherical	0.10	43.14	8.86	0.029	1.022
February	Spherical	0.01	23.11	13.10	0.025	0.981
March	Spherical	0.10	56.69	9.35	0.060	0.999
April	Spherical	0.10	75.70	10.76	0.010	1.000
May	Spherical	0.10	115.80	11.61	-0.034	1.031
June	Spherical	39.90	298.70	27.03	0.005	0.981
July	Spherical	5.70	257.70	11.59	-0.024	0.990
August	Spherical	0.10	265.00	11.57	-0.013	0.984
September	Spherical	0.10	283.50	11.77	0.027	0.982
October	Spherical	0.10	139.40	10.60	-0.010	1.004
November	Spherical	0.10	84.94	9.69	0.054	0.992
December	Spherical	0.10	100.70	10.61	0.058	0.990
Ovens River catchment						
January	Spherical	38.90	239.40	42.90	-0.017	1.012
February	Spherical	60.10	320.90	32.72	0.004	1.029
March	Spherical	13.40	246.20	26.80	-0.011	0.996
April	Gaussian	67.20	188.00	103.05	-0.043	1.008
May	Spherical	217.00	597.00	108.40	-0.026	1.071
June	Spherical	60.00	841.00	70.20	-0.044	1.028
July	Spherical	5.00	1713.00	106.00	-0.092	1.005
August	Gaussian	250.00	2084.00	90.54	-0.045	1.001
September	Spherical	1.00	975.00	100.85	-0.027	1.029
October	Spherical	42.10	383.60	87.00	-0.018	1.041
November	Spherical	6.00	367.30	92.30	-0.012	0.989
December	Spherical	31.00	192.50	88.75	-0.028	0.993

Note. KED = kriging with an external drift; SM = standardized mean error; SRMS = standardized root mean square error.

4.4 | Spatial prediction of rainfall

In this study, different geostatistical and deterministic interpolation methods including OK, OCK, KED, IDW, and RBF are adopted to estimate the spatial distribution of monthly mean rainfall in the Middle Yarra River catchment and the Ovens River catchment in Australia. Several performance measures including MBE, RMSE, and R^2 are frequently used to indicate how accurately an interpolator predicts the observed data. Smaller values of MBE and RMSE with a higher R^2 value of an interpolator indicate better prediction by the corresponding method. In case of the scatter plot, the better prediction is that if all scattered points lay close to the 45^0 line with the highest R^2 value between the predicted and observed values (Adhikary et al., 2016a).

Table 6 presents the different performance measures of the adopted interpolation methods (or interpolators) for estimating monthly rainfall over both the study catchments. The different interpolation methods are quantitatively compared based on these performance measures in order to identify the best interpolator for each of the catchment. As can be seen from the table, geostatistical (OK, OCK, and KED) interpolation methods perform better than deterministic (IDW and RBF) interpolation methods for monthly rainfall estimation in the study area. The OCK method gives the best results for rainfall estimation over the study area for all months when considering all the performance measures. The KED method gives the second best results, which is very close to the performance of the OCK method but performs better than the OK method for both the catchments. IDW and RBF give similar performance with higher error in rainfall estimation over the study area. For the Middle Yarra River catchment, Table 6 also shows that in some months, the RBF method performs better than the OK method for rainfall estimation. However, no remarkable differences are seen between them when considering all the performance measures.

For OK, OCK, KED, IDW, and RBF methods, the average RMSE values (Table 6) for the Middle Yarra River catchment are 11.93, 10.29, 10.85, 12.66, and 12.22 mm, respectively, whereas the average RMSE values for the Ovens River catchment are 17.42, 16.15, 16.65, 17.54, and 23.41 mm, respectively. For OK, OCK, KED, IDW, and

TABLE 6 Performance of different interpolation (OK, OCK, KED, IDW, and RBF) methods for monthly rainfall estimation in the study area

	MBE (mm)					RMSE (mm)					R^2				
Month	OK	OCK	KED	IDW	RBF	OK	OCK	KED	IDW	RBF	OK	OCK	KED	IDW	RBF
Middle Yarra River catchment															
January	0.98	0.63	0.73	1.76	1.56	8.92	6.96	7.99	8.96	9.99	0.40	0.66	0.56	0.42	0.31
February	0.51	0.38	0.41	0.73	0.85	5.25	3.80	4.45	5.26	5.46	0.42	0.71	0.59	0.42	0.42
March	0.97	0.65	0.70	1.67	1.78	8.71	6.79	8.41	9.58	9.50	0.54	0.73	0.58	0.45	0.46
April	1.03	0.83	0.89	1.70	2.27	9.80	8.26	8.75	10.68	10.04	0.54	0.67	0.64	0.46	0.55
May	0.78	0.72	-0.40	1.43	2.41	11.55	10.78	11.98	11.58	11.80	0.49	0.56	0.51	0.49	0.55
June	1.19	1.26	1.17	2.25	3.29	15.15	13.77	13.44	17.26	15.15	0.61	0.67	0.69	0.49	0.63
July	0.89	0.11	-0.36	1.73	2.77	15.50	14.63	14.84	16.85	15.01	0.65	0.69	0.68	0.59	0.68
August	1.05	0.03	-0.20	2.08	3.04	16.75	15.94	14.67	18.19	16.12	0.59	0.63	0.69	0.52	0.64
September	1.38	0.87	0.92	2.69	3.52	16.79	15.73	15.38	18.46	17.17	0.57	0.62	0.63	0.48	0.57
October	1.07	0.79	-0.11	1.86	2.32	12.34	10.70	11.56	12.05	12.48	0.53	0.64	0.60	0.56	0.56
November	1.23	0.55	0.58	2.16	2.28	11.00	8.25	9.10	11.43	11.72	0.50	0.73	0.66	0.47	0.46
December	1.31	1.07	1.08	2.25	2.28	11.43	7.93	9.58	11.65	12.17	0.51	0.78	0.66	0.50	0.48
Average	1.03	0.66	0.45	1.86	2.36	11.93	10.29	10.85	12.66	12.22	0.53	0.67	0.62	0.49	0.54
Ovens River catchment															
January	-0.62	-0.56	-0.25	-1.78	-1.13	12.86	12.39	12.45	13.68	18.42	0.41	0.44	0.42	0.34	0.18
February	0.11	-0.01	0.08	-0.96	1.26	17.51	16.32	17.34	17.53	25.64	0.15	0.23	0.19	0.11	0.02
March	-0.28	-0.22	-0.15	-1.26	-0.25	13.13	12.30	12.87	14.44	17.14	0.44	0.59	0.51	0.33	0.32
April	-0.62	-0.38	-0.41	-0.53	-0.58	10.31	9.19	9.58	11.00	14.91	0.56	0.71	0.66	0.49	0.31
May	-1.07	-0.59	-0.60	-1.94	-1.56	23.18	21.16	22.89	23.27	34.19	0.18	0.29	0.25	0.15	0.01
June	-0.92	-0.87	-0.94	-1.19	-0.81	22.13	21.98	22.04	22.21	32.50	0.52	0.66	0.61	0.57	0.30
July	-1.94	-1.36	-2.01	-3.53	-2.09	23.16	22.89	22.92	23.01	31.35	0.62	0.68	0.65	0.64	0.44
August	-0.94	-0.92	-1.05	-1.35	-1.16	22.24	21.92	21.97	22.27	30.01	0.66	0.70	0.68	0.68	0.49
September	-0.62	-0.55	-0.62	-0.46	-0.87	20.39	20.01	20.08	20.44	28.12	0.50	0.57	0.54	0.56	0.32
October	-0.53	-0.31	-0.35	-0.63	-0.41	19.37	17.65	19.01	19.64	26.24	0.23	0.39	0.32	0.19	0.04
November	0.01	-0.33	-0.05	0.18	0.30	10.52	9.37	9.48	12.17	11.61	0.73	0.80	0.78	0.65	0.70
December	-0.27	-0.21	-0.11	-0.41	-0.22	9.67	8.59	9.12	10.82	10.83	0.63	0.71	0.69	0.54	0.61
Average	-0.64	-0.53	-0.54	-1.15	-0.63	17.42	16.15	16.65	17.54	23.41	0.47	0.57	0.53	0.44	0.31

Note. IDW = inverse distance weighting; KED = kriging with an external drift; MBE = mean bias error; OCK = ordinary cokriging; OK = ordinary kriging; *R2* = coefficient of determination; RBF = radial basis function; RMSE = root mean square error.

RBF methods, the average R^2 values (Table 6) for the Middle Yarra River catchment are .53, .67, .62, .49 and .54, respectively whereas the average R^2 values for the Ovens River catchment are .47, .57, .53, .44, and .31, respectively. The higher R^2 value in the OCK and KED methods indicate that using elevation as a secondary variable brings more information in the rainfall estimation process under the kriging-based geostatistical analysis framework.

As explained by Delbari et al. (2013), using elevation as a secondary variable may not always improve the prediction accuracy through the OCK analysis if the spatial continuity of elevation is weaker than that of rainfall despite a high correlation exists between rainfall and elevation. In this study, the relative nugget effect (i.e., ratio of nugget coefficient to sill) of the direct variogram models for rainfall (Table 3) and elevation (Figure 5), and the cross-variogram models for rainfall-elevation (Table 4) for both the catchments are found very small in all months. This results in the improvement in the rainfall estimation by the OCK method, which is thus selected as the best interpolator for the study area in this study. Therefore, the OCK method (the best interpolator) is used to generate a continuous rainfall dataset of the monthly average rainfall for each of the catchments, which are shown in Figure 8 and Figure 9. The created datasets are expected to be very useful in various hydrological and water resources planning studies within the Yarra River catchment and the Ovens River catchment in Australia.

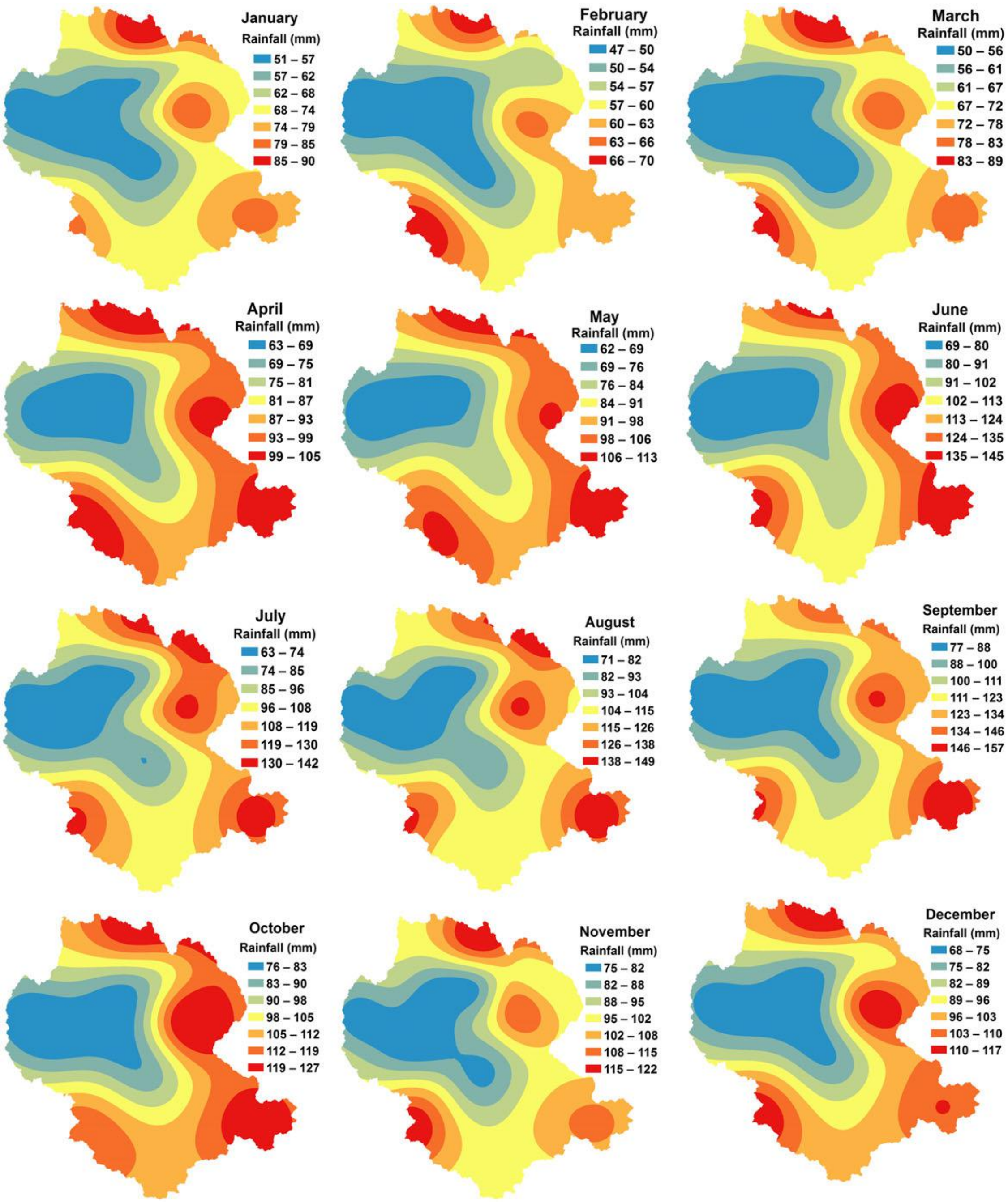

FIGURE 8 Spatial distribution of monthly rainfall in the Middle Yarra River catchment using the ordinary cokriging (the best interpolator in this study) interpolation method

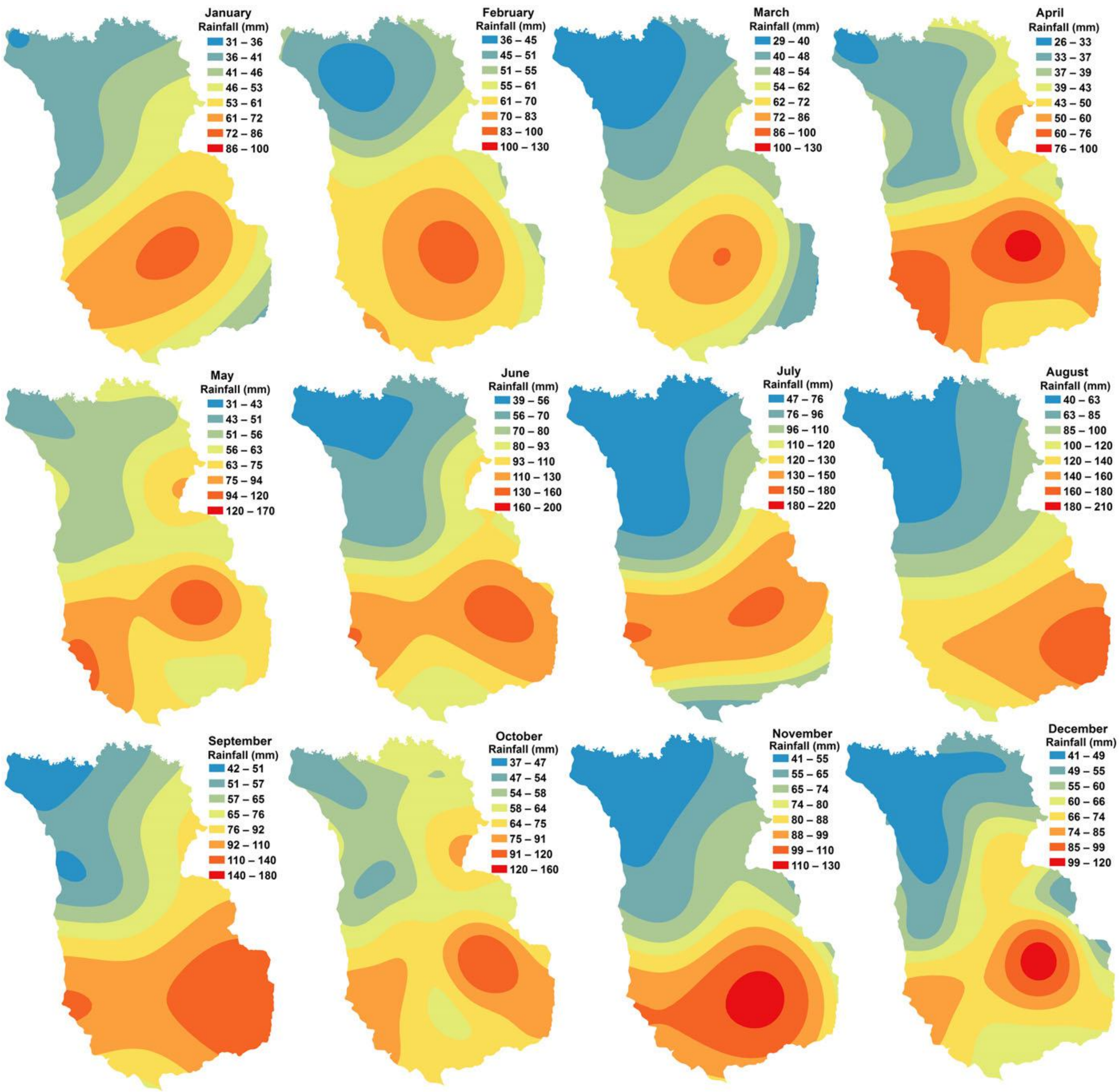

FIGURE 9 Spatial distribution of monthly rainfall in the Ovens River catchment using the ordinary cokriging (the best interpolator in this study) interpolation method

5 | CONCLUSIONS

In this study, three kriging-based geostatistical (OK, OCK, and KED) and two deterministic (IDW and RBF with thin plate spline) interpolation methods are used for estimating spatial distribution of monthly mean rainfall in the Middle Yarra River catchment and the Ovens River catchment in Victoria, Australia. The objective is to compare the performances of these interpolation methods to select the best interpolation method for generating a high quality continuous rainfall dataset in the form of a rainfall map for the study area. The elevation data obtained from a DEM of the study area is used as a supplementary variable in addition to rainfall records for the cokriging analysis using the ordinary cokriging and kriging with an external drift methods. Results show that the geostatistical methods outperform the deterministic methods for spatial interpolation of rainfall over the study area. Specifically, the performance of the cokriging methods (OCK and KED) is better than that of other geostatistical methods. The performance of the RBF with thin plate spline is found practically as good as the ordinary kriging method for rainfall estimation, whereas the IDW method is shown to have the worst results for the study area. OCK performs the best among all the interpolators and gives the improved estimates of rainfall in all months for both the catchments. It provides the lowest estimation errors and the highest correlations between the estimated and observed monthly average rainfall. Thus, ordinary cokriging is identified as the best interpolator in this study for estimating spatial distribution of rainfall in both the catchments. The obtained results indicate that making use of elevation as an auxiliary variable in addition to rainfall data can enhance the estimation of rainfall in a catchment with the mountainous and/or complex terrain. This study thus recommends the use of cokriging for the generation of continuous rainfall map especially in catchments with high spatial variation of rainfall as well as elevation.

ACKNOWLEDGEMENTS

The authors acknowledge the financial support for this study through an International Postgraduate Research Scholarship (IPRS) provided by the Government of Australia and Victoria University, Melbourne. The authors are thankful to the Australian Bureau of Meteorology (BoM) and Geoscience Australia for providing the necessary data used in this study. The authors also wish to thank the Editor and three anonymous reviewers for their valuable comments and suggestions, which have improved the quality of this paper.

REFERENCES

Adhikary, S. K., Yilmaz, A. G., & Muttil, N. (2015). Optimal design of raingauge network in the Middle Yarra River catchment, Australia. *Hydrological Processes*, *29*(11), 2582–2599. https://doi.org/10.1002/hyp.10389

Adhikary, S. K., Muttil, N., & Yilmaz, A. G. (2016a). Ordinary kriging and genetic programming for spatial estimation of rainfall in the Middle Yarra River catchment, Australia. *Hydrology Research*, *47*(6), 1182–1197. https://doi.org/10.2166/nh.2016.196

Adhikary, S. K., Muttil, N., & Yilmaz, A. G. (2016b). Genetic programming-based ordinary kriging for spatial interpolation of rainfall. *Journal of Hydrologic Engineering*, *21*(2) 04015062. https://doi.org/10.1061/(ASCE)HE.1943-5584.0001300

ASCE (1996). *Hydrology handbook* (2nd ed.). American Society of Civil Engineers (ASCE): Reston, VA.

Barua, S., Muttil, N., Ng, A. W. M., & Perera, B. J. C. (2012). Rainfall trend and its implications for water resource management within the Yarra River catchment, Australia. *Hydrological Processes*, *27*(12), 1727–1738. https://doi.org/10.1002/hyp.9311

Boer, E. P. J., de Beurs, K. M., & Hartkamp, A. D. (2001). Kriging and thin plate splines for mapping climate variables. *International Journal of Applied Earth Observation and Geoinformation*, *3*(2), 146–154. https://doi.org/10.1016/S0303-2434(01)85006-6

Chen, F. W., & Liu, C. W. (2012). Estimation of the spatial rainfall distribution using inverse distance weighting (IDW) in the middle of Taiwan. *Paddy and Water Environment*, *10*(3), 209–222. https://doi.org/10.1007/s10333-012-0319-1

Chilès, J. P., & Delfiner, P. (1999). *Geostatistics: modeling spatial uncertainty*. New York: John Wiley & Sons.

Daly, C. (2006). Guidelines for assessing the suitability of spatial climate data sets. *International Journal of Climatology*, *26*(6), 707–721. https://doi.org/10.1002/joc.1322

Daly, E., Kolotelo, P., Schang, C., Osborne, C. A., Coleman, R., Deletic, A., & McCarthy, D. T. (2013). *Escherichia coli* Concentrations and loads in an urbanised catchment: The Yarra River, Australia. *Journal of Hydrology*, *497*, 51–61. https://doi.org/10.1016/j.jhydrol.2013.05.024

Delbari, M., Afrasiab, P., & Jahani, S. (2013). Spatial interpolation of monthly and annual rainfall in northeast of Iran. *Meteorology and Atmospheric Physics*, *122*(1), 103–113. https://doi.org/10.1007/s00703-013-0273-5

Di Piazza, A., Conti, F. L., Noto, L. V., Viola, F., & Loggia, G. L. (2011). Comparative analysis of different techniques for spatial interpolation of rainfall data to create a serially complete monthly time series of precipitation for Sicily, Italy. *International Journal of Applied Earth Observation and Geoinformation*, *13*(3), 396–408. https://doi.org/10.1016/j.jag.2011.01.005

Dirks, K. N., Hay, J. E., Stow, C. D., & Harris, D. (1998). High resolution studies of rainfall on Norfolk Island part II: Interpolation of rainfall data. *Journal of Hydrology*, *208*(3–4), 187–193. https://doi.org/10.1016/S0022-1694(98)00155-3

EPA Victoria. (2003). Environmental condition of rivers and streams in the Ovens catchment. Environment Report of EPA Victoria, Publication No. 909. Available at: <http://www.epa.vic.gov.au/~/media/Publications/909.pdf> (Accessed on: 20 November, 2016).

ESRI (2009). *ArcGISv9.3.1 (computer software product of ESRI)*. Redlands, CA, USA: Environmental Systems Research Institute (ESRI).

Feki, H., Slimani, M., & Cudennec, C. (2012). Incorporating elevation in rainfall interpolation in Tunisia using geostatistical methods. *Hydrological Sciences Journal*, *57*(7), 1294–1314. https://doi.org/10.1080/02626667.2012.710334

Gittins, R. (1968). Trend surface analysis of ecological data. *Journal of Ecology*, *56*(3), 845–869. https://doi.org/10.2307/2258110

Goovaerts, P. (1997). *Geostatistics for natural resources evaluation*. New York: Oxford University Press.

Goovaerts, P. (1999). Geostatistics in soil science: State-of-the-art and perspectives. *Geoderma*, *89*, 1–46. https://doi.org/10.1016/S0016-7061(98)00078-0

Goovaerts, P. (2000). Geostatistical approaches for incorporating elevation into the spatial interpolation of rainfall. *Journal of Hydrology*, *228*(1–2), 113–129. https://doi.org/10.1016/S0022-1694(00)00144-X

Gyasi-Agyei, Y. (2016). Assessment of radar-based locally varying anisotropy on daily rainfall interpolation. *Hydrological Sciences Journal*, *61*(10), 1890–1902. https://doi.org/10.1080/02626667.2015.1083652

Haddad, K., Rahman, A., Zaman, M., & Shrestha, S. (2013). Applicability of Monte Carlo cross validation technique for model development and validation using generalized least squares regression. *Journal of Hydrology*, *482*, 119–128. https://doi.org/10.1016/j.jhydrol.2012.12.041

Hancock, P. A., & Hutchinson, M. F. (2006). Spatial interpolation of large climate data sets using bivariate thin plate smoothing splines. *Environmental Modelling & Software*, *21*(12), 1684–1694. https://doi.org/10.1016/j.envsoft.2005.08.005

Hevesi, J. A., Istok, J. D., & Flint, A. L. (1992). Precipitation estimation in mountainous terrain using multivariate geostatistics (part 1 and 2). *Journal of Applied Meteorology*, *31*(7), 661–688. https://doi.org/10.1175/1520-0450(1992)031<0661:PEIMTU>2.0.CO;

Hsieh, H. H., Cheng, S. J., Liou, J. Y., Chou, S. C., & Siao, B. R. (2006). Characterization of spatially distributed summer daily rainfall. *Journal of Chinese Agricultural Engineering*, *52*, 47–55.

Hutchinson, M. F. (1995). Interpolating mean rainfall using thin plate smoothing splines. *International Journal of Geographical Information Systems*, *9*(4), 385–403. https://doi.org/10.1080/02693799508902045

Isaaks, H. E., & Srivastava, R. M. (1989). *Applied geostatistics*. New York: Oxford University Press.

Jeffrey, S. J., Carter, J. O., Moodie, K. B., & Beswick, A. R. (2001). Using spatial interpolation to construct a comprehensive archive of Australian climate data. *Environmental Modelling & Software*, *16*(4), 309–330. https://doi.org/10.1016/S1364-8152(01)00008-1

Johnson, F., Hutchinson, M. F., The C, Beesley, C., & Green, J. (2016). Topographic relationships for design rainfalls over Australia. *Journal of Hydrology*, *533*, 439–451. https://doi.org/10.1016/j.jhydrol.2015.12.035

Johnston, K., VerHoef, J. M., Krivoruchko, K., & Lucas, N. (2001). *Using ArcGIS geostatistical analyst*. Redlands, CA, USA: ArcGIS Manual by ESRI.

Jones, D. A., Wang, W., & Fawcett, R. (2009). High-quality spatial climate data-sets for Australia. *Australian Meteorological and Oceanographic Journal*, *58*(4), 233–248.

Journel, A., & Huijbregts, C. (1978). *Mining geostatistics*. New York: Academic Press.

Li, J., & Heap, A. D. (2011). A review of comparative studies of spatial interpolation methods in environmental sciences: Performance and impact factors. *Ecological Informatics*, *6*(3–4), 228–241. https://doi.org/10.1016/j.ecoinf.2010.12.003

Li, M., & Shao, Q. (2010). An improved statistical approach to merge satellite rainfall estimates and rain-gauge data. *Journal of Hydrology*, *385*(1–2), 51–64. https://doi.org/10.1016/j.jhydrol.2010.01.023

Lloyd, C. D. (2005). Assessing the effect of integrating elevation data into the estimation of monthly precipitation in Great Britain. *Journal of Hydrology*, *308*(1–4), 128–150. https://doi.org/10.1016/j.jhydrol.2004.10.026

Ly, S., Charles, C., & Degré, A. (2011). Geostatistical interpolation of daily rainfall at catchment scale: The use of several variogram models in the Ourthe and Ambleve catchments, Belgium. *Hydrology and Earth System Sciences*, *15*(7), 2259–2274. https://doi.org/10.5194/hess-15-2259-2011

Mair, A., & Fares, A. (2011). Comparison of rainfall interpolation methods in a mountainous region of a tropical island. *Journal of Hydrologic Engineering*, *16*(4), 371–383. https://doi.org/10.1061/(ASCE)HE.1943-5584.0000330

Martínez-cob, A. (1996). Multivariate geostatistical analysis of evapotranspiration and precipitation in mountainous terrain. *Journal of Hydrology*, *174*(1–2), 19–35. https://doi.org/10.1016/0022-1694(95)02755-6

Melbourne Water. (2015). 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: 20 October, 2015).

Moral, F. J. (2010). Comparison of different geostatistical approaches to map climate variables: Application to precipitation. *International Journal of Climatology*, *30*(4), 620–631. https://doi.org/10.1002/joc.1913

Phillips, D. L., Dolph, J., & Marks, D. (1992). A comparison of geostatistical procedures for spatial analysis of precipitation in mountainous terrain. *Agricultural and Forest Meteorology*, *58*(1–2), 119–141. https://doi.org/10.1016/0168-1923(92)90114-J

Robertson, G. P. (2008). *GS+: geostatistics for the environmental sciences*. Gamma Design Software: Plainwell, Michigan, USA.

Schreider, S. Y., Jakeman, A. J., Pittock, A. B., & Whetton, P. H. (1996). Estimation of possible climate change impacts on water availability, extreme flow events and soil moisture in the Goulburn and ovens basins, Victoria. *Climate Change*, *34*(3–4), 513–546. https://doi.org/10.1007/BF00139304

Sharma, R. H., & Shakya, N. M. (2006). Hydrological changes and its impact on water resources of Bagmati watershed, Nepal. *Journal of Hydrology*, *327*(3–4), 315–322. https://doi.org/10.1016/j.jhydrol.2005.11.051

Subyani, A. M., & Al-Dakheel, A. M. (2009). Multivariate geostatistical methods of mean annual and seasonal rainfall in Southwest Saudi Arabia. *Arabian Journal of Geosciences*, *2*(1), 19–27. https://doi.org/10.1007/s12517-008-0015-z

Teegavarapu, R. S. V., & Chandramouli, V. (2005). Improved weighting methods, deterministic and stochastic data-driven models for estimation of missing precipitation records. *Journal of Hydrology*, *312*(1–4), 191–206. https://doi.org/10.1016/j.jhydrol.2005.02.015

Thiessen, A. H. (1911). Precipitation averages for large areas. *Monthly Weather Review*, *39*(7), 1082–1084. https://doi.org/10.1175/1520-0493(1911)39<1082b:PAFLA>2.0.CO;2

Wackernagel, H. (2003). *Multivariate geostatistics: An introduction with applications* (3rd ed.). Berlin: Springer-Verlag.

Webster, R., & Oliver, M. A. (2007). *Geostatistics for environmental scientists* (2nd ed.). Chichester, United Kingdom: John Wiley & Sons.

Woldemeskel, F. M., Sivakumar, B., & Sharma, A. (2013). Merging gauge and satellite rainfall with specification of associated uncertainty across Australia. *Journal of Hydrology*, *499*, 167–176. https://doi.org/10.1016/j.jhydrol.2013.06.039

Yang, X., Xie, X., Liu, D. L., Ji, F., & Wang, L. (2015). Spatial interpolation of daily rainfall data for local climate impact assessment over greater Sydney region. *Advances in Meteorology*, *2015*, 563629. https://doi.org/10.1155/2015/563629

Yilmaz, A. G., & Perera, B. J. C. (2014). Extreme rainfall nonstationarity investigation and intensity–frequency–duration relationship. *Journal of Hydrologic Engineering*, *19*(6), 1160–1172. https://doi.org/10.1061/(ASCE)HE.1943-5584.0000878

Yilmaz, A. G., Hossain, I., & Perera, B. J. C. (2014). Effect of climate change and variability on extreme rainfall intensity–frequency–duration relationships: A case study of Melbourne. *Hydrology and Earth System Sciences*, *18*, 4065–4076. https://doi.org/10.5194/hess-18-4065-2014

Yu, M. C. L., Cartwright, I., Braden, J. L., & de Bree, S. T. (2013). Examining the spatial and temporal variation of groundwater inflows to a valley-to-floodplain river using 222Rn, geochemistry and river discharge: The Ovens River, Southeast Australia. *Hydrology and Earth System Sciences*, *17*(12), 4907–4924. https://doi.org/10.5194/hess-17-4907-2013

How to cite this article: Adhikary SK, Muttil N, Yilmaz AG. Cokriging for enhanced spatial interpolation of rainfall in two Australian catchments. *Hydrological Processes*. 2017;31:2143–2161. https://doi.org/10.1002/hyp.11163

Chapter 5

Optimal Design of a Rain Gauge Network Using Kriging-Based Geostatistical Approach

5.1 Introduction

Rainfall is one of the most sought-after variables of hydrological processes. Rain gauge station is the important component of a comprehensive hydrometric network, which is used to collect rainfall data and related hydrological information (Wang et al., 2015). Rain gauge stations can be considered the ground-based sensing nodes in a hydrological sensor web (Chen et al., 2014), which is a reconfigurable and collaborative observation system (Zheng et al., 2012). Thus, the proper deployment of rain gauge stations through an optimal network configuration can improve the utilization of limited resources and promote cost savings in various water resources management tasks including flood control, drought management, reservoir operation, streamflow forecasting and disaster warning, and study of climate change impact on water resources. Industrial development and growing domestic and agricultural water demands are imposing increased pressure on water resources and on the environment. With the increasing pressure on water resources, more emphasis should be given on sustainable management of water resources. This can be achieved through an optimal hydrometric network (Mishra and Coulibaly, 2009), which can provide the adequate information of hydrological data with appropriate temporal and spatial resolution.

In recent decades, it has been largely documented that climate change could have adverse regional effects on hydrologic extremes such as floods and droughts in terms of their frequency and severity. Impacts of climate change on various water resources sectors have been extensively investigated as well (Intergovernmental Panel on Climate Change, 2001). Climate change during the late 20th century was significant (Stewart et al., 2005), and the projected changes in climate over the 21st century predict a climate that would result in greater shifting in seasonality of streamflows, with more rainfall, more frequent flooding and earlier loss of mountain snowpack (Dettinger et al., 2004; Knowles and Cayan, 2004). Climate variability affects the spatial and temporal distribution of rainfall over the catchment, and the influence of the spatial variability of rainfall on storm runoff has been addressed well by several researchers (e.g., Wilson et al., 1979; Troutman, 1983; Krajewski et al., 1991; 2003). Climatic variability also changes the streamflow regime of river systems (e.g., Poff et al., 1997). For example, large-scale inter-annual and inter-decadal climate variations related to the ENSO and to the SAM (Bureau of Meteorology, 2016) account for high variability in climate and streamflow systems in Australia (Dutta et al., 2006; Chowdhury and Beecham, 2010). All the aforementioned findings indicate the importance of having high quality hydrological information to sustain the researchers' efforts in deriving appropriate models for flood warning and drought mitigation. This highlights the increasing requirement of optimal network design in the context of climate and land use changes. However, the spatial and temporal effect of hydrometeorological observations used for different water resources studies changes under changing climate conditions (Mishra and Coulibaly, 2009). WMO (1994) recommends that hydrometric networks should be reviewed from time to time to take into account the ‘‘reduction in hydrological uncertainty brought about by the data since the last network analysis’’ and any changes related to funding, data needs, and logistics, etc. Therefore, due to nonstationary climate conditions, the hydrometric networks should be evaluated and reviewed periodically to account for climate change and land use changes for the efficient management of water resources systems.

Rain gauge network is usually installed to get direct measurements of rainfall data. However, many of the water resources systems are large in spatial extent and often consist of a rain gauge network that is very sparse due to large cost involvement, logistics, and geological factors. This results in considerable uncertainty in the collected rainfall data from these sparse rain gauge networks (Zealand et al., 1999; Goovaerts, 2000). Therefore, an optimal rain gauge network should be established, which is capable of providing the high quality rainfall estimates needed for the effective hydrological analysis and design of water resources projects and is thus regarded as an indispensable component of any hydrological study. In this chapter, a simple and effective rain gauge network design methodology using the kriging-based geostatistical approach under the variance reduction framework is presented in order to achieve an optimal rain gauge network for the case study catchment.

In rain gauge network design studies, it is a common practice to identify and select the best network configuration having optimal number and locations of rain gauge stations. Adequate density as well as location of stations in the network equally plays a vital role in determining whether the network is optimal and sufficient information is gained. These issues of rain gauge network design can be suitably handled by the kriging-based geostatistical approach. Therefore, this approach was adopted for optimal design of rain gauge network demonstrated in this study, which finds wide applications in the rain gauge network design across the world (e.g., Shamsi et al., 1988; Kassim and Kottegoda, 1991; Loof et al., 1994; Papamichail and Metaxa, 1996; Pardo-Igúzquiza, 1998; Tsintikidis et al., 2002; Chen et al., 2008; Cheng et al., 2008; Yeh et al., 2011; Shaghaghian and Abedini, 2013; Aziz et al., 2016; Chang et al., 2017; Feki et al., 2017). An important feature of this approach is the provision of kriging standard error (KSE) (obtained through square root of the kriging variance) that forms the basis for rain gauge network design and evaluation. Based on this approach, the optimal network configuration is achieved through minimizing the KSE values of the network, which involves a methodical search for the optimal number and locations (or optimal combination) of rain gauge stations in the network producing the minimum kriging standard error of areal average and/or point rainfall estimates.

An optimal rain gauge network is essentially a balanced or ideal network, which should neither be suffered from lack of rain gauge stations nor be over-saturated with redundant rain gauge stations (Mishra and Coulibaly, 2009; Shaghaghian and Abedini, 2013). In other words, the optimal rain gauge network should consist of 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 address such problem either by expanding the network through installation of additional rain gauge stations (network augmentation) to enhance the rainfall estimation with reduced uncertainty or by rationalizing the network through eliminating redundant rain gauge stations (network rationalization) from the network to minimize the operation and maintenance cost of the station and/or network (Mishra and Coulibaly, 2009).

In most of the past studies, network expansion with additional rain gauge stations to minimize the network variance has been the underlying criterion to achieve the optimal network. However, an existing rain gauge network may consist of redundant stations, which have little or no contribution to the network performance for providing quality rainfall estimates. As a solution this issue, both additional as well as redundant stations were considered to achieve the optimal rain gauge network design. In this chapter, the optimal rain gauge network for the case study catchment was demonstrated and established through optimal positing of additional rain gauge stations (network augmentation process) in addition to eliminating and/or optimally relocating of existing redundant rain gauge stations (network rationalization process). The underlying principle is that optimal positioning of additional and redundant rain gauge stations in the high variance zones will reduce the KSE of the network and hence the improved performance of the network for enhanced estimation of rainfall can be achieved. Applying this principle repetitively, a certain stage will come when the optimal network configuration with optimal combination of all rain gauge stations can be achieved to form the optimal rain gauge network, which exhibits the best network performance with minimum KSE. A flow chart of the methodological framework adopted for the optimal design of a rain gauge network in this study is shown in Figure 5-1.

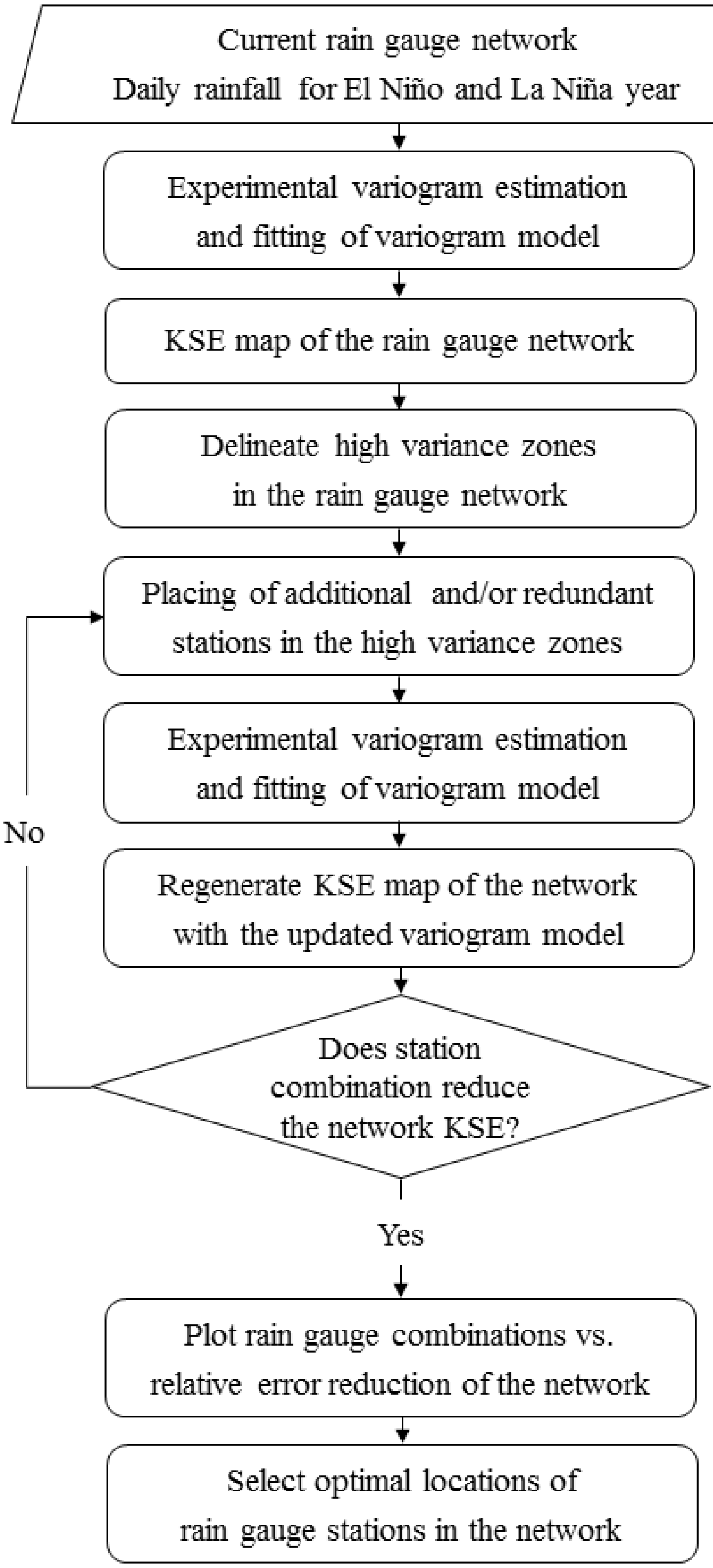

Figure 5-1. The methodological framework adopted for optimal rain gauge network design in this study

In Australia, high climate variability usually results in the high variation of rainfall (Chiew and McMahon, 2003; Ruiz et al., 2007). In particular, the spatial variability of rainfall in southeastern Australia (where the case study catchment is located) is strongly influenced by the El Niño and La Niña processes of the ENSO effect (Murphy and Ribbe, 2004; Dutta et al., 2006; Pittock et al., 2006; Chowdhury and Beecham, 2010; Mekanik et al., 2013). As a solution to this issue, the spatial variability of rainfall caused by the ENSO effect was taken into consideration in the rain gauge network design process. Use of rainfall records from both the El Niño and La Niña periods in the rain gauge network design can offer a better representation of the high rainfall variability experienced in the Middle Yarra River catchment of Victoria, Australia (the case study catchment). Therefore, the rain gauge network was designed independently based on the El Niño and La Niña rainfall records and demonstrated in this chapter. The network that gave the enhanced estimates of areal average and point rainfalls for both the El Niño and La Niña periods was chosen as the optimal rain gauge network for the case study catchment. The procedure of using rainfall records from the El Niño and La Niña periods enables the obtained optimal rain gauge network to take the spatial variability of local rainfall patterns over the case study catchment into account, which can yield the enhanced estimation of rainfall in the study catchment.

This chapter consists of the following journal paper.

1. **Adhikary SK**, Yilmaz AG, Muttil N. 2015. Optimal design of rain gauge network in the Middle Yarra River catchment, Australia. *Hydrological Processes* 29(11): 2582-2599. DOI: 10.1002/hyp.10389. (SCImago Journal Rank indicator: Q1; Impact Factor: 3.014).

VICTORIA UNIVERSITY
MELBOURNE AUSTRALIA

GRADUATE RESEARCH CENTRE

5.2 DECLARATION OF CO-AUTHORSHIP AND CO-CONTRIBUTION: PAPERS INCORPORATED IN THESIS BY PUBLICATION

This declaration is to be completed for each conjointly authored publication and placed at the beginning of the thesis chapter in which the publication appears.

1. PUBLICATION DETAILS (to be completed by the candidate)

Title of Paper/Journal/Book: Optimal Design of Rain Gauge Network in the Middle Yarra River Catchment, Australia

Surname: Adhikary First name: Sajal Kumar

College: College of Engineering & Science Candidate's Contribution (%): 80

Status:

Accepted and in press: ☑ Date: 28/10/2014

Published: ☑ Date: 03/12/2014

2. CANDIDATE DECLARATION

I declare that the publication above meets the requirements to be included in the thesis as outlined in the HDR Policy and related Procedures – policy.vu.edu.au.

Signature	Date
[redacted]	20/02/2017

3. CO-AUTHOR(S) DECLARATION

In the case of the above publication, the following authors contributed to the work as follows:

The undersigned certify that:

1. They meet criteria for authorship in that they have participated in the conception, execution or interpretation of at least that part of the publication in their field of expertise;
2. They take public responsibility for their part of the publication, except for the responsible author who accepts overall responsibility for the publication;
3. There are no other authors of the publication according to these criteria;
4. Potential conflicts of interest have been disclosed to a) granting bodies, b) the editor or publisher of journals or other publications, and c) the head of the responsible academic unit; and

5. The original data will be held for at least five years from the date indicated below and is stored at the following **location**(**s**):

College of Engineering and Science, Victoria University, Melbourne, Australia

Name(s) of Co-Author(s)	Contribution (%)	Nature of Contribution	Signature	Date
Sajal Kumar Adhikary	80	Conceptual ideas, data analysis, model simulations, paper writing	[redacted]	20/02/2017
Abdullah Gokhan Yilmaz	10	Critical comments, discussion on conceptual ideas	[redacted]	20/02/2017
Nitin Muttil	10	Critical comments, discussion on conceptual ideas	[redacted]	20/02/2017

Updated: June 2015

HYDROLOGICAL PROCESSES
Hydrol. Process. **29**, 2582–2599 (2015)
Published online 3 December 2014 in Wiley Online Library
(wileyonlinelibrary.com) DOI: 10.1002/hyp.10389

5.3

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 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

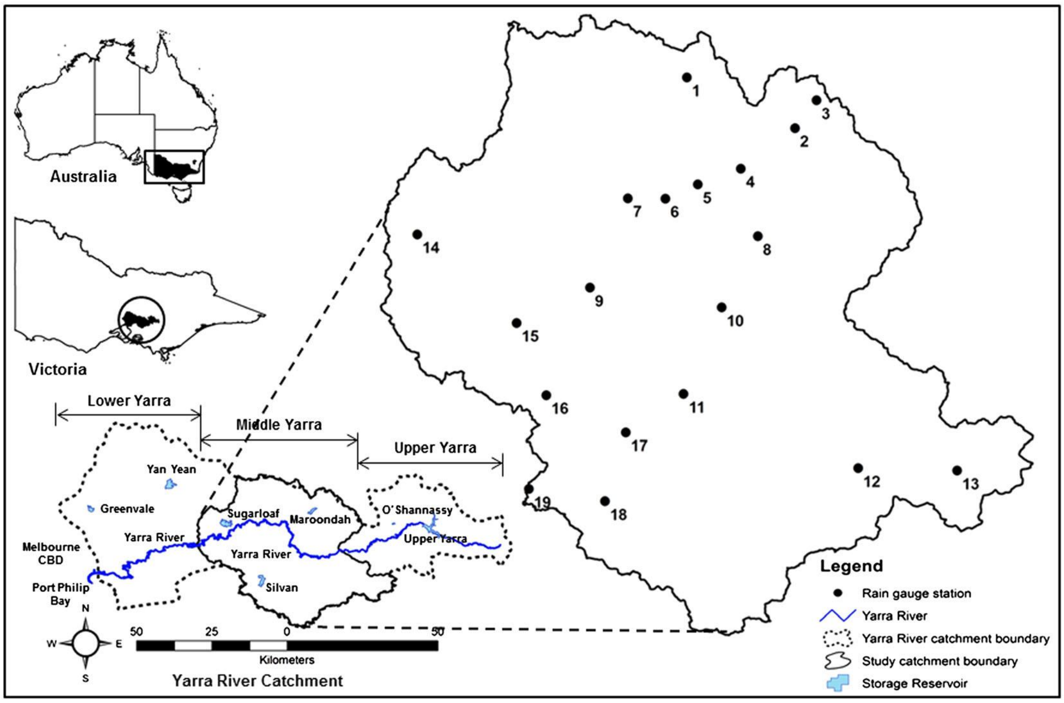

Figure 1. Yarra River catchment showing the study area with rain gauge stations

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.

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.

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

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

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}).

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.
$KS_{17,\ 0.05} = 0.3180$.
$KS_{19,\ 0.05} = 0.3010$.

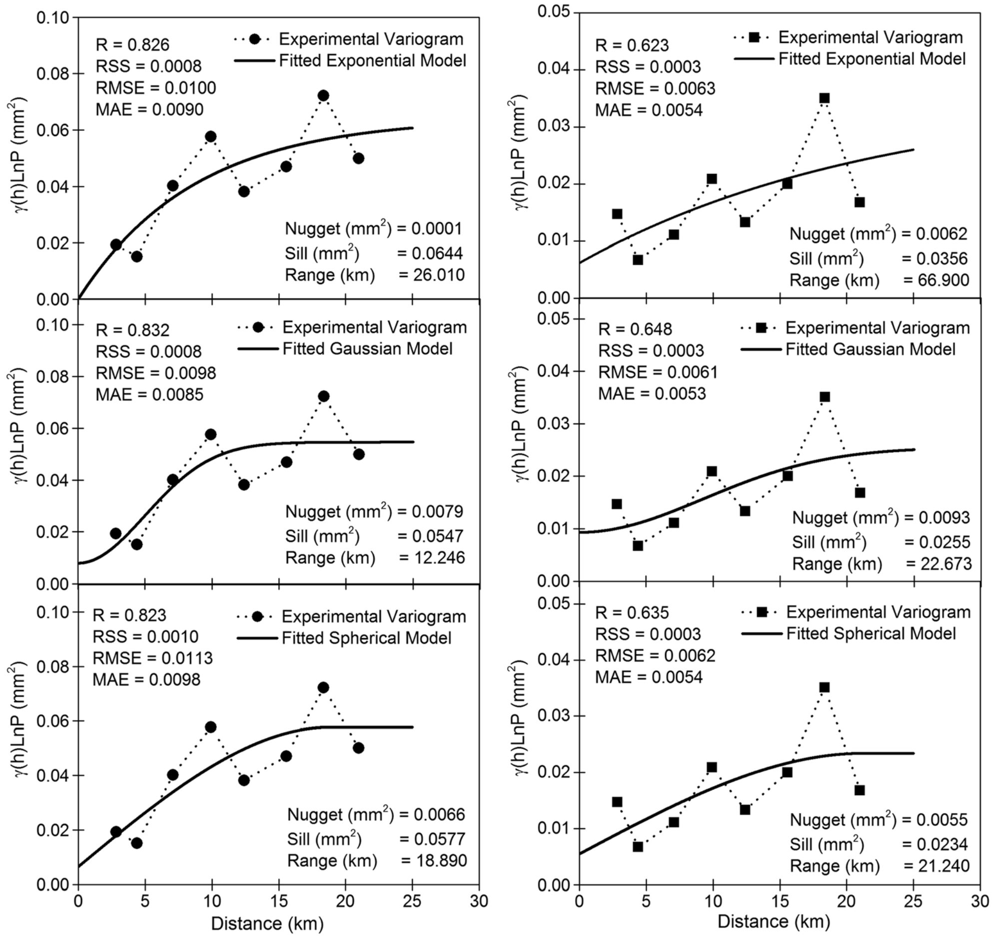

Figure 2. Fitted variogram models for mean daily rainfall of (a) El Niño year and (b) La Niña year for the network before 2011 (Bureau of Meteorology's base network with 17 rain gauge stations)

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

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

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

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

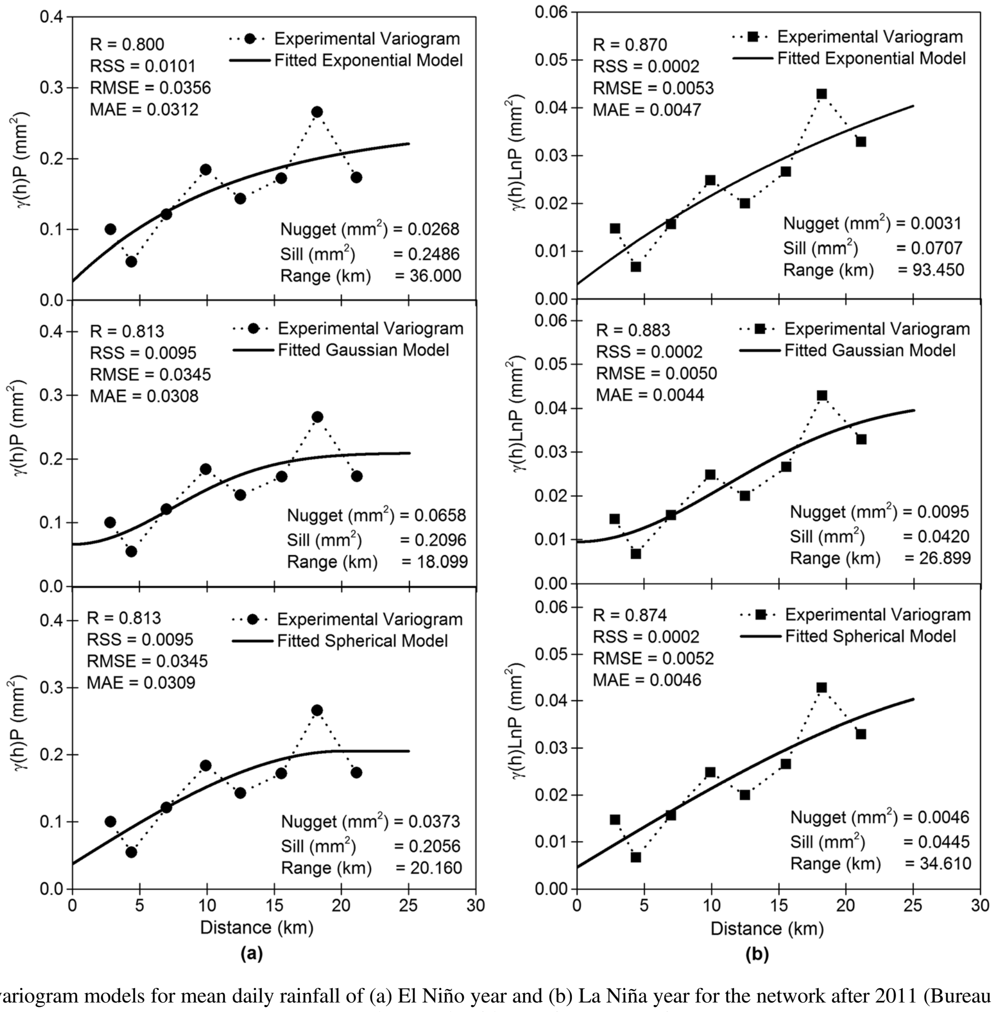

Figure 3. Fitted variogram models for mean daily rainfall of (a) El Niño year and (b) La Niña year for the network after 2011 (Bureau of Meteorology's augmented network with 19 rain gauge stations)

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

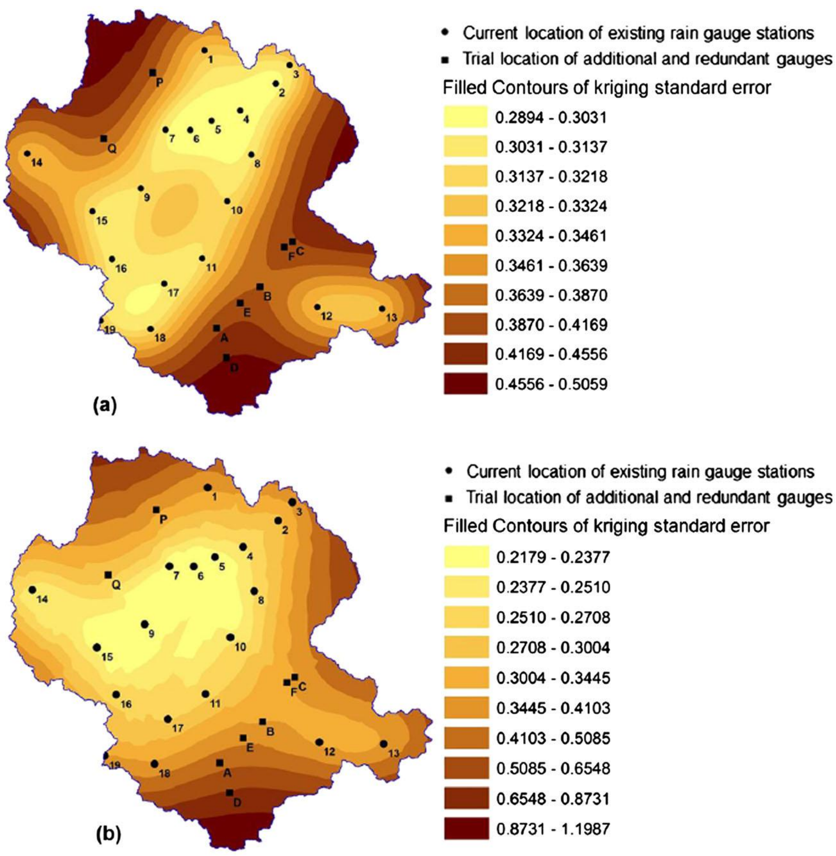

Figure 4. Kriging standard error (KSE) map for (a) El Niño year and (b) La Niña year with trial locations of additional stations (stations 18 and 19) and redundant stations (stations 5 and 6) in the high variance zones of the KSE map

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.

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.

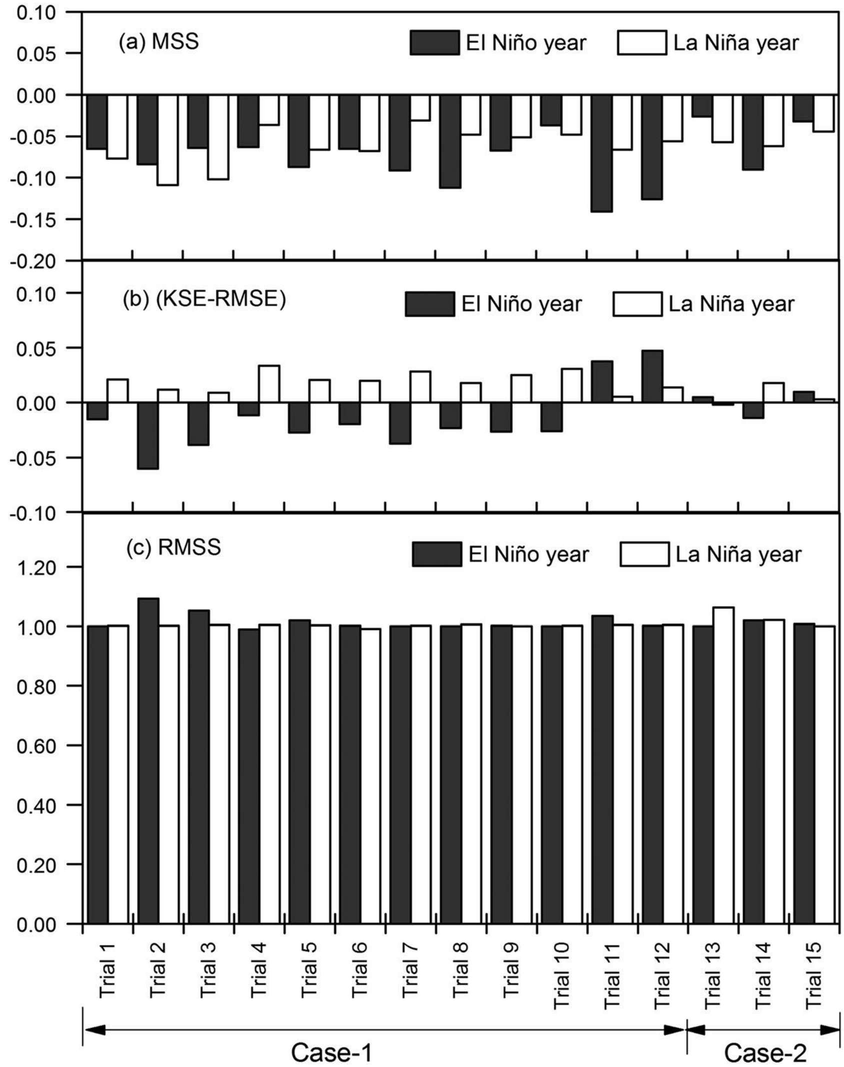

Figure 5. Cross-validation statistics (a) mean standardized prediction error (MSS), (b) difference of kriging standard error (KSE) and root mean square error (RMSE) and (c) root mean square standardized (RMSS) for El Niño and La Niña years. MSS and (KSE-RMSE) values close to 0 indicate an accurate model,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

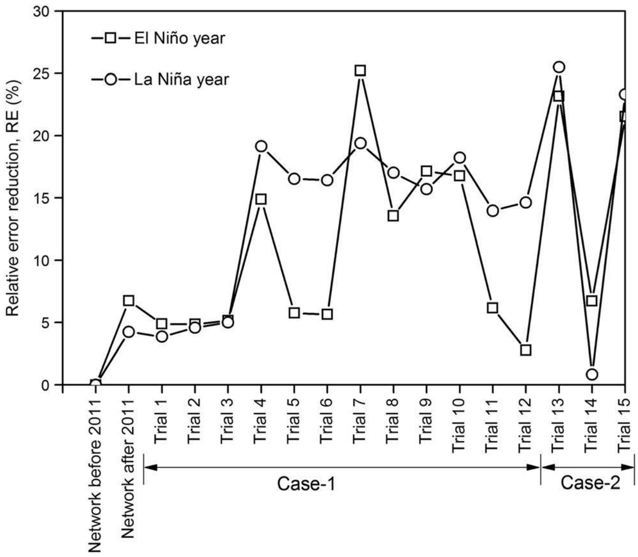

Figure 6. Relative error (RE) reduction for different combinations of stations in the network after 2011

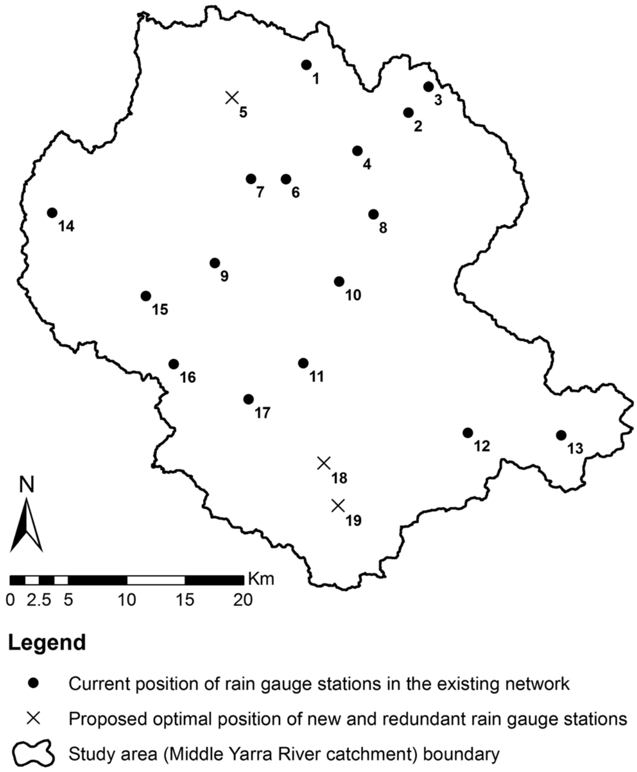

Figure 7. Devised optimal rain gauge network in the Middle Yarra River catchment

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

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

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.

Chapter 6

Enhanced Streamflow Forecasting Using Optimal Rain Gauge Network-Based Input

6.1 Introduction

In general, rainfall is considered independent of streamflow simulation and forecasting in most of the hydrological investigations. For example, rain gauge network is frequently designed to achieve the optimal rain gauge network for enhanced estimation of point rainfall at ungauged locations and/or areal average rainfall over a catchment or study area (e.g., Bras and Rodriguez-Iturbe, 1976; Bastin et al., 1984; Papamichail and Metaxa, 1996; Pardo-Igúzquiza, 1998; Tsintikidis et al., 2002; Cheng et al., 2008; Shaghaghian and Abedini, 2013; Aziz et al., 2016; Haggag et al., 2016; Feki et al., 2017). But this does not allow one to focus on the strength and weakness of the established optimal rain gauge network that really matter when rainfall data from the optimal network are fed into the streamflow forecasting models (Andréassian et al., 2001). Therefore, a rain gauge network should be designed to provide a satisfactory solution to the specific needs such as the enhanced streamflow forecasting in this study for which the network is being designed and/or established. Since streamflow is a consequence of rainfall, correct rainfall input is of great importance for accurate and enhanced streamflow forecasting. Hence, an optimal rain gauge network should be established, which can provide the high quality rainfall input that can be used in streamflow forecasting models in order to achieve the enhanced streamflow forecasting.

In the existing literature, studies on impact of rainfall input uncertainty and rain gauge network density and distribution on the performance of streamflow simulation and forecasting are abundantly found (e.g., Faurès et al., 1995; St-Hilaire et al., 2003; Dong et al., 2005; Anctil et al., 2006; Bárdossy and Das, 2008; Ekström and Jones, 2009; Moulin et al., 2009; Xu et al., 2013; Tsai et al., 2014; Kar et al., 2015; Zeinivand, 2015; Chen et al., 2017). In those studies, streamflow forecasting models employed rainfall data obtained directly from the existing rain gauge network, which might not be an optimal network and hence might not provide the quality rainfall estimates. This ultimately affects the performance of the streamflow forecasting accuracy and results in the less accurate streamflow forecasting. However, the direct use of rainfall input from an optimal rain gauge network for streamflow simulation and forecasting is rarely documented. Hence, the development of an ANN-based enhanced streamflow forecasting approach using the optimal rain gauge network-based input is detailed in this chapter. In the proposed ANN-based enhanced streamflow forecasting approach, the rainfall input from an optimal rain gauge network was used in the ANN-based streamflow forecasting model instead of using the rainfall input directly from an existing non-optimal rain gauge network in order to achieve the enhanced streamflow forecasting.

In this investigation, the first streamflow forecasting model was developed using the rainfall input from the existing rain gauge network and the second streamflow forecasting model was established using the rainfall input from an optimal rain gauge network (details of the optimal rain gauge network design demonstrated in Chapter 5) over the case study catchment. Then, the third streamflow forecasting model was developed using the rainfall input from an augmented optimal rain gauge network over the case study catchment. This chapter provides a comparison of the performances of the aforementioned three streamflow forecasting models to test the performance and robustness of the proposed ANN-based enhanced streamflow forecasting approach. Furthermore, the issues associated with the application of ANN-based input selection technique for selecting significant input variables are detailed in this chapter. Based on the input significance obtained by the ANN-based input selection technique, the chapter finally presents an indirect way of identifying the optimal locations of rain gauge stations in the final optimal rain gauge network, which can give the best streamflow forecasting performance of the network. Furthermore, this is the first study of this kind undertaken in Australia to incorporate the optimal rain gauge network-based rainfall input within the ANN-based data-driven streamflow forecasting framework for enhanced streamflow forecasting.

As was indicated in Section 1.1, floods and droughts induced water management problems have become the major concerns in Australia, particularly in southeastern Australia. Australia's population and agricultural production are highly concentrated in this part of the country where water resources play a crucial role in the region's economic development. The main aim of the reservoir operation in Australia is to store as much water as possible to meet water demands during droughts while keeping provision for flood control during floods. Lower rainfall usually causes the reduction in streamflows, which obviously results in the shortage of reservoir inflows and affects the overall water supply to the urban areas. In addition, reduction in streamflows may cause the increased risks of bushfires, which is not unusual in the southeastern Australia during droughts. On the other hand, the occurrence of extreme rainfall results in the excess amount of streamflows that often causes flash floods in the urbanized catchments located along the coastlines of the region and makes it vulnerable and risk-prone. Therefore, accurate and enhanced streamflow forecasting is of great importance for optimal operation of reservoirs and drought management, and effective flood control and management in Australia. Accordingly, the proposed enhanced streamflow forecasting approach and the obtained results detailed in this study could be very much supportive to address the aforementioned floods and droughts induced water management challenges in an efficient manner, specifically in southeastern Australia.

This chapter consists of the following journal paper.

1. **Adhikary SK**, Muttil N, Yilmaz AG. 2017b. Improving streamflow forecast using optimal rain gauge network-based input to artificial neural network models. *Hydrology Research* (Revised based on the reviewers' comments and resubmitted to the journal). (SCImago Journal Rank indicator: Q1; Impact Factor: 1.779).

GRADUATE RESEARCH CENTRE

6.2 DECLARATION OF CO-AUTHORSHIP AND CO-CONTRIBUTION: PAPERS INCORPORATED IN THESIS BY PUBLICATION

This declaration is to be completed for each conjointly authored publication and placed at the beginning of the thesis chapter in which the publication appears.

1. PUBLICATION DETAILS (to be completed by the candidate)

Title of Paper/Journal/Book: Improving Streamflow Forecast Using Optimal Rain Gauge Network-Based Input to Artificial Neural Network Models

Surname: Adhikary

First name: Sajal Kumar

College: College of Engineering & Science

Candidate's Contribution (%): 80

Status:

Accepted and in press: ☑ Date:

Published: ☑ Date:

2. CANDIDATE DECLARATION

I declare that the publication above meets the requirements to be included in the thesis as outlined in the HDR Policy and related Procedures – policy.vu.edu.au.

[redacted]	15/03/2017
Signature	**Date**

3. CO-AUTHOR(S) DECLARATION

In the case of the above publication, the following authors contributed to the work as follows:

The undersigned certify that:

1. They meet criteria for authorship in that they have participated in the conception, execution or interpretation of at least that part of the publication in their field of expertise;

2. They take public responsibility for their part of the publication, except for the responsible author who accepts overall responsibility for the publication;

3. There are no other authors of the publication according to these criteria;

4. Potential conflicts of interest have been disclosed to a) granting bodies, b) the editor or publisher of journals or other publications, and c) the head of the responsible academic unit; and

5. The original data will be held for at least five years from the date indicated below and is stored at the following **location**(s):

College of Engineering and Science, Victoria University, Melbourne, Australia

Name(s) of Co-Author(s)	**Contribution (%)**	**Nature of Contribution**	**Signature**	**Date**
Sajal Kumar Adhikary	80	Conceptual ideas, data analysis, model simulations, paper writing	[redacted]	15/03/2017
Nitin Muttil	10	Critical comments, discussion on conceptual ideas		15/03/2017
Abdullah Gokhan Yilmaz	10	Critical comments, discussion on conceptual ideas		15/03/2017

Manuscript Submitted to "Hydrology Research" Journal

(Revised based on the reviewers' comments and resubmitted)

6.3 Improving streamflow forecast using optimal rain gauge network-based input to artificial neural network models

Sajal Kumar Adhikary[a, b*], Nitin Muttil[a, b], Abdullah Gokhan Yilmaz[c]

[a] *College of Engineering and Science, Victoria University, P.O. Box 14428, Melbourne, Victoria 8001, Australia*

[b] *Institute for Sustainability and Innovation, Victoria University, P.O. Box 14428, Melbourne, Victoria 8001, Australia*

[c] *College of Engineering, University of Sharjah, P.O. Box 27272, Sharjah, United Arab Emirates*

[*]Correspondence to: **Sajal Kumar Adhikary**, College of Engineering and Science, Victoria University, Footscray Park Campus, P.O. Box 14428, Melbourne, Victoria 8001, Australia. Email: sajal.adhikary@live.vu.edu.au, Telephone: +61 3 9919 5041.

ABSTRACT

Accurate streamflow forecasting is of great importance for the effective management of water resources systems. In this study, an improved streamflow forecasting approach using the optimal rain gauge network-based input to artificial neural network (ANN) models is proposed and demonstrated through a case study (the Middle Yarra River catchment in Victoria, Australia). First, the optimal rain gauge network is established based on the current rain gauge network in the catchment. Rainfall data from the optimal and current rain gauge networks together with streamflow observations are used as the input to train the ANN. Then, the best subset of significant input variables relating to streamflow at the catchment outlet is identified by the trained ANN. Finally, one-day-ahead streamflow forecasting is carried out using ANN models formulated based on the selected input variables for each rain gauge network. The results indicate that the optimal rain gauge network-based input to ANN models gives the best streamflow forecasting results for the training, validation and testing phases in terms of various performance evaluation measures. Overall, the study concludes that the proposed approach is highly effective to achieve the enhanced streamflow forecasting and could be a viable option for streamflow forecasting in other catchments.

KEYWORDS: *Streamflow forecasting, Artificial neural network, Optimal rain gauge network, Input selection, Yarra River catchment.*

1. INTRODUCTION

Streamflow is one of the key variables in hydrology. Accurate forecasting of streamflow is essential for many of the activities associated with the efficient planning and operation of the components of risk-based water resources systems. Particularly, flood control and operational river management systems highly depend on accurate and reliable forecasting of streamflow. The analysis and design of dams and bridges, management of extreme events including floods and droughts, optimal operation of reservoir encompassing irrigation, hydropower generation, domestic and industry water supply objectives are a few examples where information regarding short-term and long-term streamflow forecasting is vital (Londhe and Charhate, 2010). Hence, there is a growing need to improve the short-term and long-term streamflow forecasting for the efficient optimization of water resources systems (Akhtar et al., 2009).

The approaches used for streamflow forecasting cover a wide range of methods from completely black box (data-driven or machine learning) models to detailed conceptual or physically-based models (Porporato and Ridolfi, 2001). The conceptual or physically-based models usually require extensive data and huge computational efforts, and are influenced by the effects of overparameterization and parameter redundancy (Linares-Rodriguez et al., 2015). Furthermore, such models could not be applied to a slightly different system. As a result of these limitations, data-driven methods have been increasingly preferred for hydrological modelling and forecasting (Khu et al., 2001; Yilmaz & Muttil, 2014). In particular, a data-driven method that has gained significant attention to researchers in recent years is the artificial neural network (ANN)-based streamflow forecasting technique (e.g., Zealand et al., 1999; Dibike and Solomatine, 2001; Birikundavyi et al., 2002; Huang et al., 2004; Kumar et al., 2004; Wu et al., 2005; Kişi, 2007; Srinivasulu and Jain, 2009; Londhe and Charhate, 2010; Abrahart et al., 2012; Sivapragasam et al., 2014; Linares-Rodriguez et al., 2015; Taormina et al., 2015).

The majority of the aforementioned studies have confirmed that ANN is able to outperform traditional statistical methods. ANN is perhaps the most popular machine learning method with flexible mathematical structure, which is capable of identifying a direct mapping between inputs and outputs without detailed consideration of the internal structure of the physical process (Maier and Dandy, 2000; Dibike and Solomatine, 2001). ANN models are computationally fast and reliable, and yield results comparable to conceptual models. These models can extract the complex non-linear relationships between the inputs and outputs of a process without the physics being explicitly provided. Furthermore, ANN models for streamflow forecasting require only a limited number of input variables, such as rainfall and flow data (e.g., Talei et al., 2010; Londhe and Charhate, 2010; Yilmaz and Muttil, 2014), which makes them suitable for forecasting applications in practice. For a detailed description of ANNs with their modelling processes and applications in hydrology and water resources, readers are referred to Govindaraju and Rao (2000), ASCE Task Committee (2000a,b), Dawson and Wilby (2001), Maier et al. (2010), and Tayfur (2012).

This study mainly focuses on the important hydrological aspects of rainfall input for streamflow simulation within the framework of ANN-based streamflow forecasting models. Rainfall is one of the most important inputs in the development of ANN models for streamflow forecasting. Since streamflow is a consequence of rainfall, using accurate rainfall input to ANN models is vital in order to achieve the enhanced streamflow forecasting. However, many of the water resources systems are large in spatial extent and often consist of a rain gauge network that is very sparse due to economic, geological and logistic factors. This may cause inaccuracy in the collected rainfall information (Zealand et al., 1999). Therefore, it is necessary to establish an optimal rain gauge network, which can give high quality rainfall estimates for accurate streamflow forecasting. An optimal rain gauge network refers to a balanced or ideal network that never suffers from station shortages, or from over saturations caused by redundant stations (Mishra and Coulibaly, 2009; Shaghaghian and Abedini, 2013; Adhikary et al., 2015). If rainfall information can be more accurately estimated through the optimal network and used in ANN-based streamflow forecasting models, it is likely that enhanced streamflow forecasting can be achieved, a conclusion supported by the works of Andréassian et al. (2001), who tested the sensitivity of watershed models to the imperfect knowledge of rainfall input.

Rainfall is often considered independent of streamflow forecasting in many hydrological studies such as average areal rainfall estimation over a catchment (e.g., Bras and Rodriguez-Iturbe, 1976; Bastin et al., 1984; Seed and Austin, 1990; Adhikary et al., 2016a, 2017) or the design of rain gauge networks (e.g., Papamichail and Metaxa, 1996; Pardo-Igúzquiza, 1998; Tsintikidis et al., 2002; Chen et al., 2008; Cheng et al., 2008; Adhikary et al., 2015; Feki et al., 2017). However, this does not allow one to focus on the strength and weakness of an established network that really matter when rainfall data are fed into a streamflow forecasting model. Furthermore, Bras (1979) and Storm et al. (1989) emphasized that watersheds act as low-pass filters, attenuating the rainfall variability. It is thus necessary to take this filter into account to determine the quality and quantity of rainfall data required to achieve a certain degree of accuracy in streamflow forecasting. Hence, it is logical to design a rain gauge network for providing a satisfactory solution to the specific needs (enhanced streamflow forecasting in the current study) for which the network is being established. Based on the aforementioned considerations, it is thus hypothesized that use of the optimal rain gauge network-based input to streamflow forecasting models can contribute to the improved streamflow forecasting.

To date, many studies have been devoted to the impact of rainfall input, varying rain gauge network density and distribution on the performance of streamflow forecasting (e.g., Faurès et al., 1995; St-Hilaire et al., 2003; Dong et al., 2005; Anctil et al., 2006; Xu et al., 2006; Bárdossy and Das, 2008; Ekström and Jones, 2009; Moulin et al., 2009; Volkmann et al., 2010; Xu et al., 2013; Tsai et al., 2014; Linares-Rodriguez et al., 2015). However, none of these studies used rainfall input from the optimal rain gauge network for streamflow forecasting. Therefore, the objective of this study is to use rainfall information from an optimally designed rain gauge network in combination with streamflow observations as the input to ANN-based streamflow forecasting models for enhanced streamflow forecasting. The specific focus is to evaluate the effectiveness of integrating an optimal rain gauge network within the framework of ANN models to achieve the improved streamflow forecasting. The experimental approach is planned in two phases and demonstrated through an application to the Middle Yarra River catchment in Victoria, Australia. First, the optimal rain gauge network is established from the current operational rain gauge network in the catchment by using the well-known Kriging-based geostatistical technique (presented in Adhikary et al., 2015). Next, streamflow forecasting is undertaken one day in advance at the catchment outlet based on the selected significant input variables (rainfall and streamflow) for each of the current and optimal rain gauge networks. Such an approach could be scalable to other catchments contingent upon addressing the local contextual issues, which is expected to be a viable option to achieve the enhanced streamflow forecasting.

The remainder of the paper is structured as follows. First, the study area and dataset used are described in details. This is followed by the detailed description of the methodology adopted in this study. The results are summarized next and finally, the conclusions drawn from the study are presented.

2. STUDY AREA AND DATASET USED

2.1 The study area

In the current study, the middle segment of the Yarra River catchment (referred to as the Middle Yarra River catchment) located in Victoria, Australia, is selected as the case study area. Approximate location of the catchment is shown in Figure 1. The catchment is located at northeast of Melbourne, which covers an area of 4044 km^2. The catchment is home to more than one-third of Victoria's population (approximately 1.8 Million). 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, which makes it a very productive catchment. The Yarra River thus plays a key role in the way Melbourne has developed and grown (Adhikary et al., 2016b).

The Yarra River catchment is divided into three distinctive sub-catchments (as shown in Figure 1), namely Upper Yarra, Middle Yarra, and Lower Yarra segments based on the different land use patterns. The Upper Yarra segment of the catchment consists of mainly forested and mountainous areas with minimum human settlement. Approximately 70% of Melbourne's drinking water supply comes from this pristine upper segment (Barua et al., 2012). The Middle Yarra segment is distinguished as the only part of the catchment with an extensive flood plain, which is mainly used for agricultural activities. The Lower Yarra segment is mainly characterized by the urbanized floodplain areas of Melbourne city. The average annual rainfall varies across the Yarra River catchment from about 1100 mm in the Upper Yarra segment to 600 mm in the lower Yarra segment (Daly et al., 2013). Hence, water resources management in the catchment is of great importance considering the diverse water use activities and high variability in rainfall.

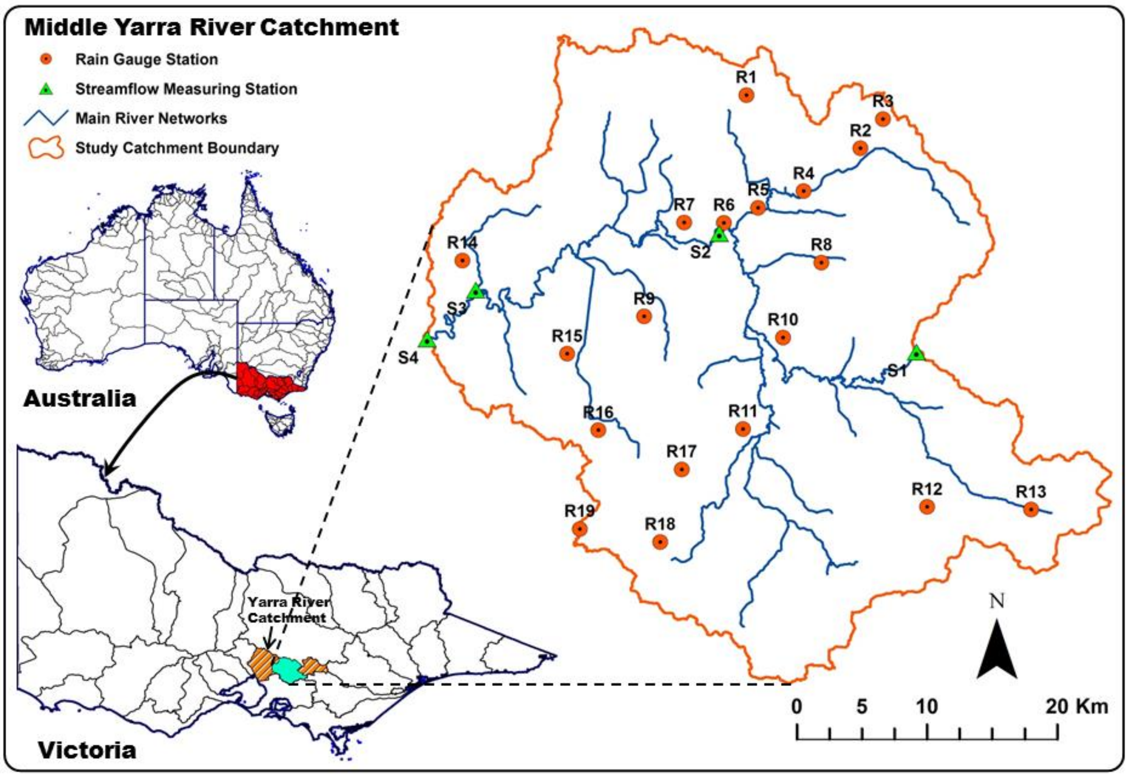

Figure 1. Location of the study area (Middle Yarra River catchment) with hydrometric stations

The Middle Yarra segment (the case study area as shown in Figure 1) covers an area of 1511 km^2. There are three storage reservoirs, namely, Maroondah, Silvan, and Sugarloaf in the study area that supports water supply for a range of activities including urban and agricultural activities. The main aim of the reservoir operation in Australia is to store as much water as possible to meet water demands during droughts while keeping provision for flood control during floods. Lower rainfall causes reduction in streamflows, which obviously results in the shortage of reservoir inflows and affects the overall water availability. In addition, reduction in streamflows may cause increased risks of bushfires. On the other hand, the occurrence of higher or extreme rainfall results 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 also depends on the water supply from the storage reservoirs mainly located in the middle and upper segments of the catchment (Adhikary et al., 2015). Therefore, accurate streamflow forecasting is of great significance for optimal operation of storage reservoirs, and planning for effective flood control and mitigation measures, particularly in the urbanized lower segment of Yarra River catchment.

2.2 Dataset used

Available literature suggests that many different variables are used as input to ANN models. Rainfall and antecedent streamflow are the most frequently used inputs for ANN-based streamflow forecasting models. The antecedent streamflow acts indirectly as a descriptor of the moisture state in the watershed (Anctil et al., 2004). The input also consists of air temperature or potential evapotranspiration in combination with rainfall information. However, some studies have shown that model results are nearly insensitive to the potential evapotranspiration or temperature and thus their usage as input are unnecessary (e.g., Oudin et al., 2005, 2006; Xu et al., 2006). Therefore, rainfall data together with streamflow observations are used as the necessary input to develop ANN-based streamflow forecasting models in the current study.

In the current study, dataset are based on the historical rainfall records from the rain gauge network of Australian Bureau of Meteorology (BoM) and streamflow observations from the streamflow measuring network of Melbourne Water Corporation (MWC). Spatial location of the hydrometric stations within the study area is shown in Figure 1. There are nineteen rain gauge stations (indicated with R1 to R19) in the BoM's current network and four streamflow measuring stations (indicated with S1 to S4) along the main course of the Yarra River in the study area. Table 1 presents the particulars of the hydrometric (rain gauge and streamflow) stations.

Table 1. Summary of the hydrometric stations in the Middle Yarra River catchment

Station no.	Station Details			
	Station ID	**Name of Station**	**Easting (m)**	**Northing (m)**
Rain gauge stations				
R1	86142	Toolangi (Mount St Leonard DPI)	367665	5840620
R2	86366	Fernshaw	376433	5836534
R3	86009	Black Spur	378165	5838779
R4	86070	Maroondah Weir	372048	5833250
R5	86385	Healesville (Mount Yule)	368559	5831973
R6	86363	Tarrawarra	365931	5830821
R7	86364	Tarrawarra Monastery	362905	5830845
R8	86219	Coranderrk Badger Weir	373425	5827770
R9	86383	Coldstream	359825	5823625
R10	86229	Healesville (Valley View Farm)	370480	5822015
R11	86367	Seville	367398	5815000
R12	86358	Gladysdale (Little Feet Farm)	381535	5809020
R13	86094	Powelltown DNRE	389545	5808810
R14	86059	Kangaroo Ground	345855	5827920
R15	86066	Lilydale	353900	5820765
R16	86076	Montrose	356285	5814905
R17	86106	Silvan	362717	5811901
R18	86072	Monbulk (Spring Road)	361051	5806323
R19	86266	Ferny Creek	354874	5807326
Streamflow measuring stations				
S1	229212	Yarra River at Millgrove	380730	5820906
S2	229653	Yarra River at Yarra Grange	365590	5830000
S3	229608	Watsons Creek at Kangaroo Ground South	346900	5825660
S4	229200	Yarra River at Warrandyte	343157	5821896

Station no. are same as in Figure 1.

Station ID for rain gauges is as defined by the Bureau of Meteorology (BoM), Australia at http://www.bom.gov.au/climate/data/stations/ and Station ID for streamflow gauges is as defined by the Melbourne Water Corporation (MWC).

Thirty years of daily meteorological and hydrological data (from 1980 to 2009) including rainfall and streamflow are used in this study. The choice of this study period is based on the availability of high quality of data with no missing records for an extended period. Daily rainfall data of all nineteen rain gauge stations are collected from the Scientific Information for Land Owners (SILO) climate database (http://www.longpaddock.qld.gov.au/silo/). The SILO database has been selected for this study because SILO data are quality controlled and completely free from missing records. The missing records in this database are filled up during quality control process based on the ordinary kriging and thin plate spline interpolation techniques using available records in the nearby surrounding stations. The SILO database gives an additional benefit of data drill opportunity using the aforementioned interpolation techniques by which one can get the necessary rainfall data at any ungauged location in the catchment (Jeffrey et al., 2001). Streamflow data of all four streamflow measuring stations are collected from the MWC database. The average annual rainfall in the study area during the 1980-2009 period varies from 710 mm to 1422 mm with a mean rainfall of 1063 mm. Approximately 60% of the mean rainfall occurs in the winter (June-August) and spring (September-November) seasons, which contributes mostly to streamflow.

3. METHODOLOGY

This study presents an approach of streamflow forecasting in an attempt to achieve the enhanced streamflow forecasting using the optimal rain gauge network-based input to ANN models. The methodological framework of the proposed approach is shown in Figure 2, which is demonstrated through an application to the Middle Yarra River catchment in Victoria, Australia. As can be seen from the figure, the framework has two parts and in the first part of the framework, an optimal and an augmented rain gauge network are established from the BoM's current operational rain gauge network. The second part consists of streamflow forecasting, which focuses on the impact of optimal rain gauge network-based input on the performance of streamflow forecasting. In general, the framework is implemented through following four steps: (i) optimal rain gauge network design, (ii) augmentation of the optimal rain gauge network, (iii) ANN-based input variable selection, and (iv) streamflow forecasting and assessment. These steps are described in the following subsections.

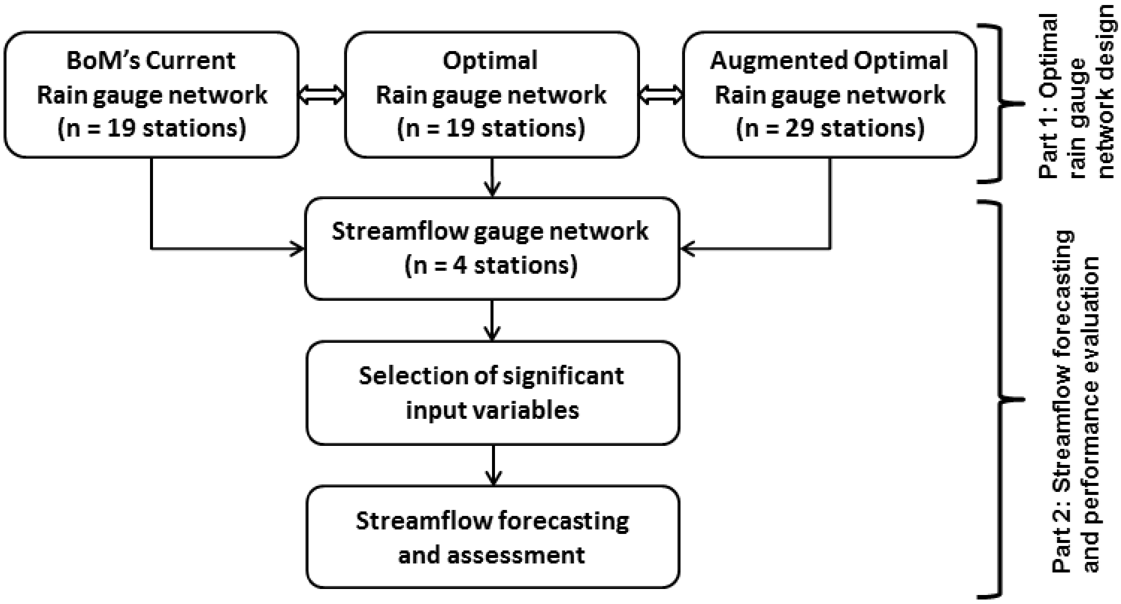

Figure 2. Framework of methodology adopted in this study

3.1 Optimal Rain Gauge Network Design

An optimal network should essentially consist of sufficient number of rain gauge stations with suitable locations in such a way that the network can provide optimum rainfall information with minimum uncertainty and cost. Adequate station density as well as location in the network equally plays a vital role in determining if the rain gauge network is optimal and sufficient information is gained (Adhikary et al., 2015). Thus, the optimal network is achieved through optimal positioning of additional stations (i.e., network extension) together with redundant stations or simply removing redundant stations (i.e., network rationalization) (St-Hilaire et al., 2003; Mishra and Coulibaly, 2009). In this study, the kriging-based geostatistical technique is used for optimal rain gauge network design. Kriging is a well-known stochastic interpolation technique that provides unbiased estimates of a variable at unsampled locations based on the sampled values at surrounding locations as well as kriging variance of estimation. The optimal rain gauge network is achieved through minimizing the kriging variance of the current network under the framework of variance reduction principle.

The principle demonstrates that optimal positioning of additional as well as redundant stations in the high variance zones of the network reduces network variance and thus improves the network performance.

Details of the optimal rain gauge network design in the Middle Yarra River catchment can be found in an earlier study conducted by Adhikary et al. (2015). The optimal network in that study was established through a methodical search for the optimal number and locations of stations in the current network using the network extension and rationalization procedures. The optimal network established in this way for the study catchment is shown in Figure 3. As can be seen from the figure, the optimal network consists of nineteen rain gauge stations including sixteen original stations (stations R1-R4, R6, R7-R17) in their current positions, two additional stations (stations R18a and R19a), and a redundant station (station R5b) in their corresponding new optimal positions. The rainfall data at the identified optimal locations of the additional and redundant stations (stations R18a, R19a, and R5b) in the optimal network is also obtained from the SILO database through their data drill option based on the ordinary kriging technique (Jeffrey et al., 2001). A major finding in the study of Adhikary et al. (2015) was that the established optimal network provides more accurate areal average and point rainfall estimates in the Middle Yarra River catchment. Now, the objective of the current study is to answer the questions whether the optimal network-based rainfall information could produce enhanced streamflow forecasting.

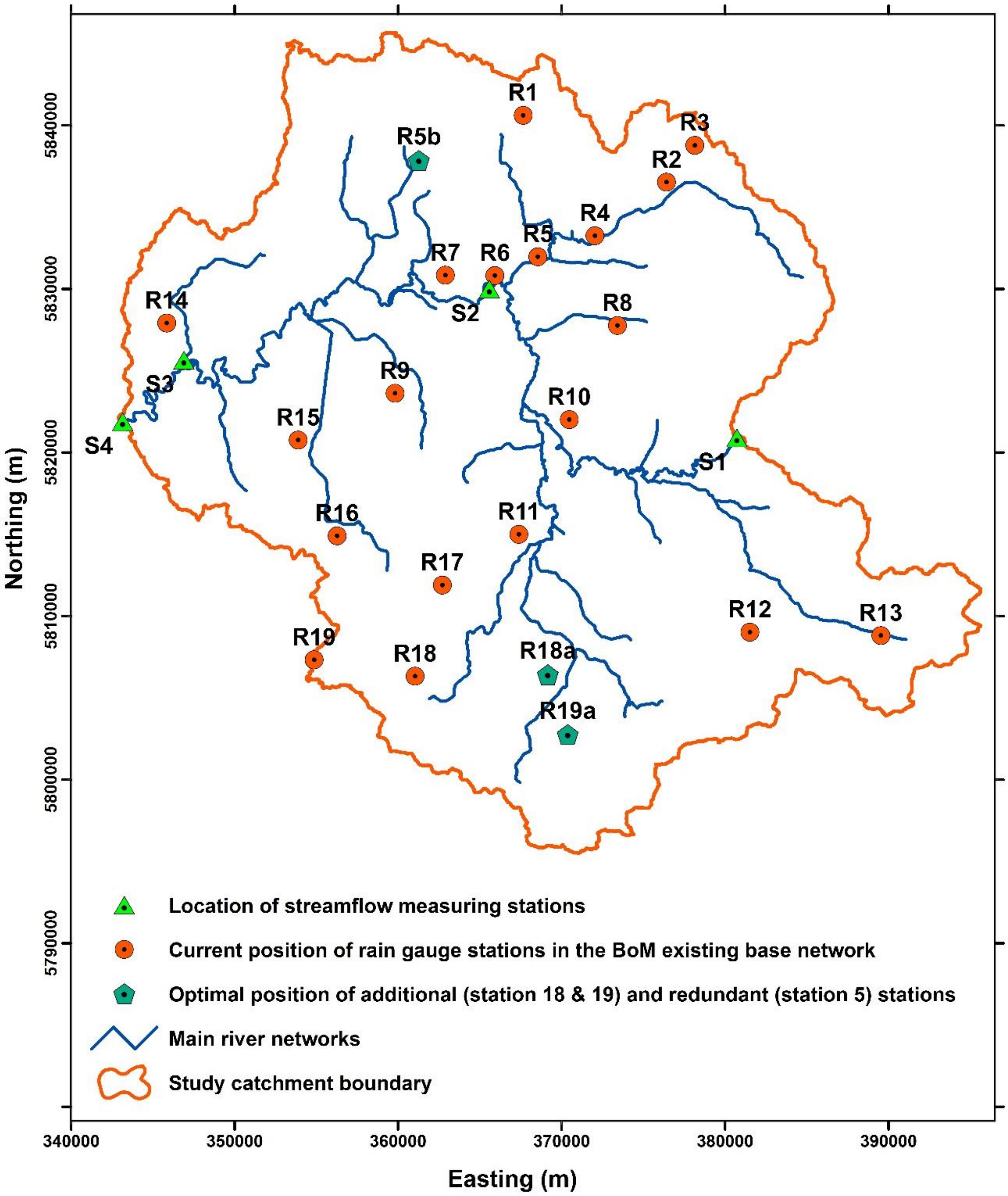

Figure 3. Optimal rain gauge network as presented in Adhikary et al. (2015) for the Middle Yarra River catchment

3.2 Augmentation of Optimal Rain Gauge Network

In rain gauge network design, it is commonly believed that a denser network with more rain gauge stations causes reduction of network variance and thus results in the improved estimate of areal average or point rainfalls in a catchment (e.g., Papamichail and Metaxa, 1996; Cheng et al., 2008). Furthermore, the network density often influences the quality of flow simulations (St-Hilaire et al., 2003). It is worth mentioning that unlike the past studies, no additional fictitious rain gauge stations to increase the network density were considered for optimal rain gauge network design presented in Adhikary et al. (2015). Considering these factors, additional fictitious stations are incorporated to augment the optimal network of Adhikary et al. (2015) to increase the network density, which will be called as the augmented optimal rain gauge network in the current study. The main intention is to investigate the potential of an augmented or dense network in enhancing the performance of streamflow forecasting. This strategy facilitates exploring the impact of a relatively denser network on the streamflow forecasting accuracy. This also helps to identify the locations of key fictitious stations in addition to rain gauge stations in the optimal network, which have greater influence on the accurate streamflow forecasting.

In order to augment the optimal network presented in Adhikary et al. (2015), the study catchment is first delineated into a number of sub-catchments based on the digital elevation model using the ArcGIS software. Additional fictitious stations are then placed in such a way that each sub-catchment comprises at least one rain gauge station. Ten additional fictitious stations are considered for the network augmentation. Thus, the resulting augmented optimal network consists of twenty-nine rain gauge stations, which is shown in Figure 4. The rainfall data at the locations of fictitious stations (stations P1 to P10) in the augmented optimal network are also collected from the SILO database. The data are generated through the data drill option based on the ordinary kriging technique (Jeffrey et al., 2001). For further details of the rainfall estimation at ungauged locations using the ordinary kriging technique, readers are referred to Adhikary et al. (2016b).

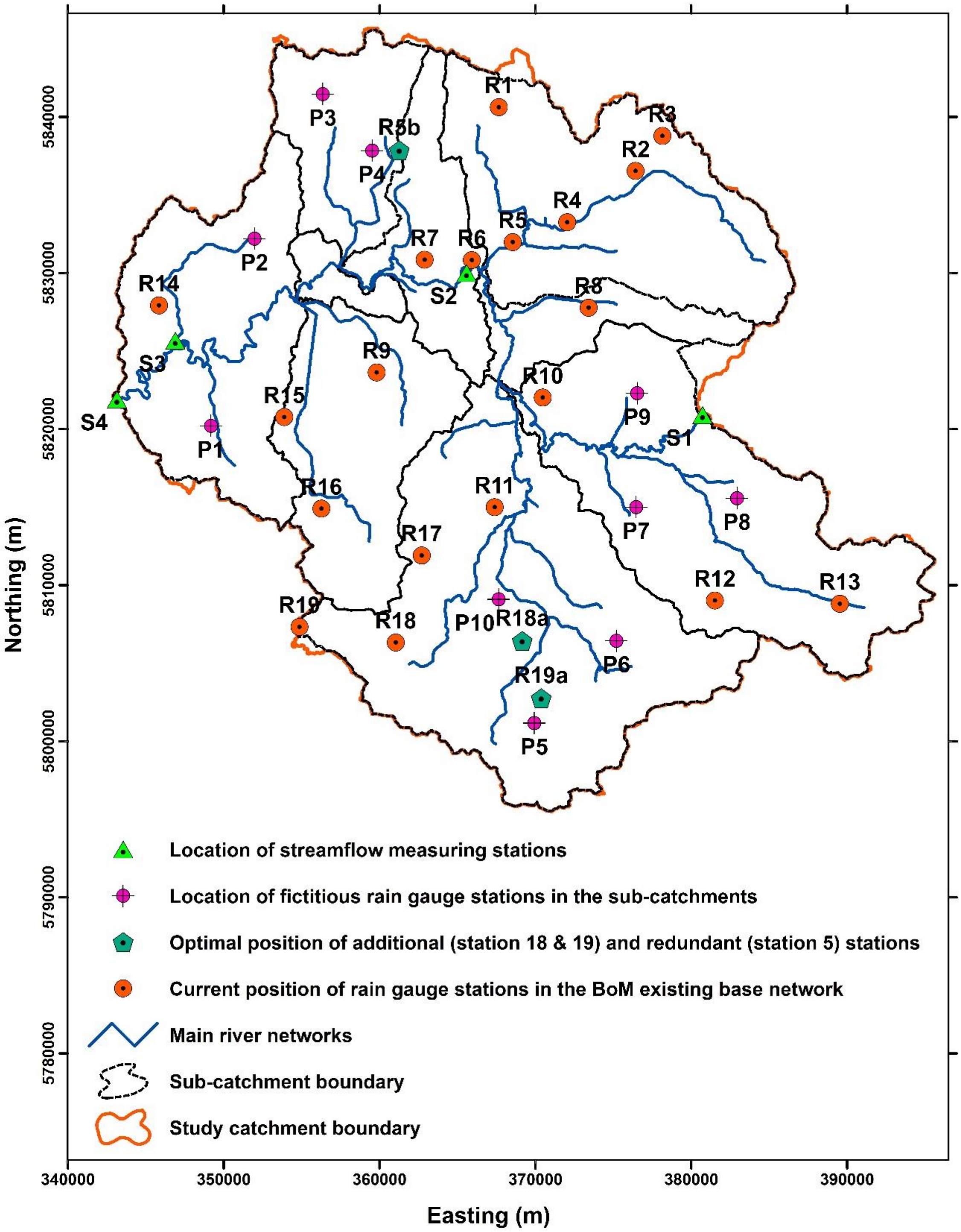

Figure 4. The augmented optimal rain gauge network with additional fictitious stations in the study area.

3.3 ANN-Based Input Variable Selection

3.3.1 ANN model

Artificial neural networks (ANNs) are biologically inspired general computational models that have been roughly based on the functioning of human brain. ANN is highly beneficial over conventional hydrological models because it has flexible structures that are able to simulate not only the linear but also the complex nonlinear hydrologic relationship between a model's input and output variables. In addition, ANN is capable of adapting itself to changing conditions leading to the enhanced model performance, shorter computation times and faster model development (Yilmaz and Muttil, 2014). Once trained properly, ANN model can be used to make forecasting of a future output for a set of given inputs. Detailed background of the ANN theory can be found in Govindaraju and Rao (2000) and Tayfur (2012).

An ANN is characterized by its architecture, training, or learning algorithm and by its activation function. The ANN model constructed in this study is the feed-forward multilayer perceptron (MLP), which is the most commonly used network topology in hydrological forecasting (ASCE Task Committee, 2000a, b). The MLP is organized as layers of computing elements, known as neurons, connected between layers via weights. A single hidden layer is considered in this study because a single hidden layer with sufficient neurons is often sufficient in many cases to fit multi-dimensional mapping problems well (Wu et al., 2005). Thus, the resulting MLP network configuration, as shown in Figure 5, consists of an input layer that receives inputs from the environment, an intermediate hidden layer, and an output layer that produces the network's response (Muttil and Chau, 2006, 2007). The number of neurons in the hidden layer depends on the problem complexity, number of input and output variables. Having a large number of hidden neurons usually gives the network flexibility to solve complex systems but this may cause overfitting. Therefore, it is essential to identify the optimal number of nodes in the hidden layer, which greatly influences the performance of the trained network. In this study, the optimum number of neurons in the hidden layer is identified using a trial and error approach by varying the number of hidden layer neurons.

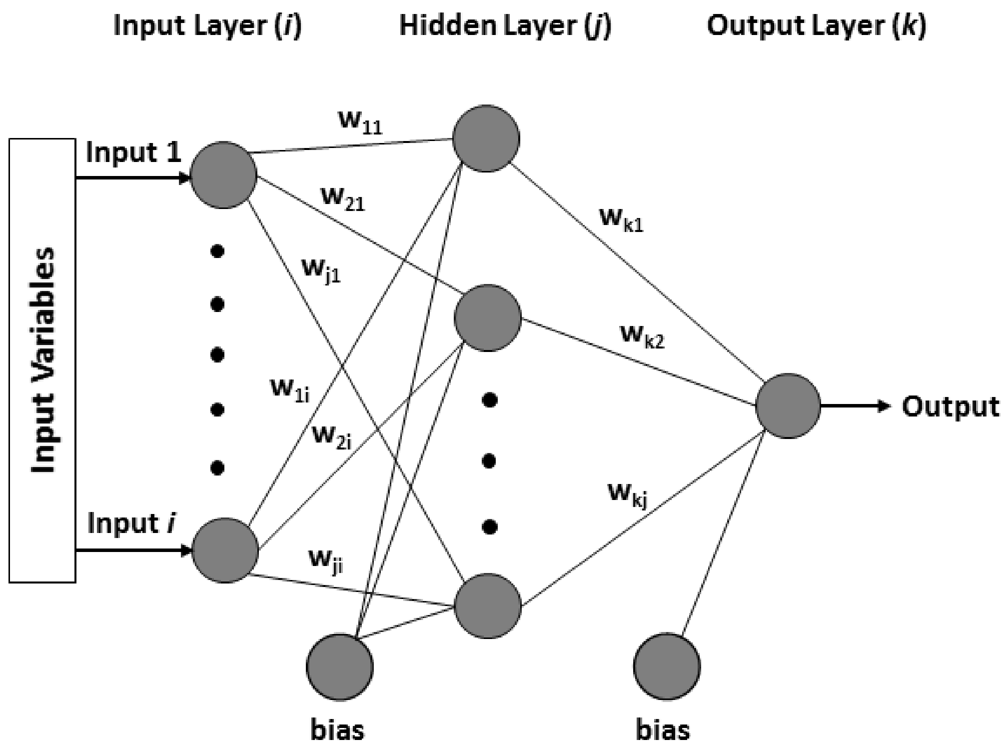

Figure 5. Configuration of a three-layer feed-forward multilayer perceptron (MLP) neural network architecture

In the MLP network, processing in neurons is done from the input layer through hidden layers to the output layer. Nonlinearity of the system is captured with activation functions in the ANN model. Amongst many types of activation functions, the sigmoid and the hyperbolic-tangent activation functions are the most commonly used functions in hydrological modelling (Dawson and Wilby, 2001). In this study, the sigmoid activation function is used in the hidden layer and a linear activation function is used in the output layer.

A backpropagation algorithm is used to train the ANN model, which is a supervised learning algorithm that adjusts the connection weights and biases in the backward direction. A number of training algorithms have been developed for error backpropagation learning. In this study, the Levenberg–Marqurdt (LM) backpropagation algorithm is used. The LM algorithm is more reliable than any other backpropagation variants because it has the fastest convergence among all algorithms and is also able to obtain the lowest mean square error in many cases (Linares-Rodriguez et al., 2015). The ANN model is implemented through the MATLAB Neural Network Toolbox.

A common practice in ANN modelling is to split the input dataset into appropriate training, validation, and testing subsets. This often helps to avoid overfitting problems and guarantee generalization capability of ANN (Linares-Rodriguez et al., 2015). Thus, the sampled dataset (i.e., 9667) of this study is divided according to the proportions 70% (i.e., 6767), 15% (i.e., 1450), and 15% (i.e., 1450) for training, validation, and testing datasets, respectively. More data (two-third of total data) are considered in the training set because in an ideal situation a larger-input dataset is preferable for training an ANN model. This approach often helps to achieve a better calibrated ANN model by capturing all the maximum and minimum values in the data series. The training dataset is used to train the ANN model. The validation dataset is used during the training process to confirm that the model does not cause an overtraining problem. In other words, when validation error increases for a specified number of iterations, the training is stopped. Finally, the performance of the trained ANN model is tested using the testing datasets. ANN weights and biases are also initialized using a fixed random seed value so that the same ANN model structure can reproduce the same network response at all times. The backpropagation training of the ANN is terminated after 1000 epochs, which is expected to be satisfactory in this study.

3.3.2 Identification of significant input variables based on ANN weights

One of the most important steps in ANN modelling is the identification of an appropriate set of input variables that essentially defines the output of a system (Muttil and Chau, 2006, 2007). If relevant input variables cannot be accurately identified, it is likely that the desired input-output relationships cannot be accurately captured by the ANN model. On the contrary, when excessive numbers of variables are used as the input, the highly correlated variables dominate the model and hence it is not possible to use information from all the measurements available. In addition, too many inputs may cause overparameterization problems (Akhtar et al., 2009; Linares-Rodriguez et al., 2015). This is usually addressed by different pre-processing and/or input selection techniques that attempt to reduce the input space by selecting the most significant input variables. The commonly used input selection techniques include correlation-based analysis, mutual information analysis, data mining techniques (e.g., principal component analysis, cluster analysis), and forward selection and backward elimination techniques (Bowden et al., 2005; Muttil and Chau, 2007).

In the recent past, an ANN-based input selection technique has been demonstrated by Muttil and Chau (2006, 2007) to identify the most significant input variables, which offers several advantages. Since ANN itself is used for significant input variable selection, no further analytical procedures are necessary for the same. A major advantage of ANN model is that it is able to learn problems involving very non-linear and complex data. Therefore, the model can identify correlated patterns between input data and corresponding target values. The ANN-based input selection technique overcomes some of the limitations associated with the aforementioned commonly used input selection techniques. For example, ANN can take into account the interaction amongst variables in the input space and thus identify variables that may not be significant by itself, but are significant in combination with other variables (Muttil and Chau, 2007). Thus, the ANN-based technique is ideally suited for identifying significant input variables for streamflow forecasting.

In this study, a two-step procedure is adopted to identify the most significant input variables. First, a set of candidate inputs is prepared based on a priori knowledge of the system being modelled. Before training, it is often useful to scale the inputs and targets so that they always fall within a specified range. In this study, input and output data are standardized between 0 and 1. The MLP network is then trained with the standardized data and the ANN-based input selection technique is used to select the best set of input variables that predominantly describes the streamflow at the catchment outlet. According to this technique, an interpretation of the connection weights along the paths from the input layer to the hidden layer of the trained network is undertaken. The inputs with the largest weight values indicate the most significant input variables. An input significance measure, known as the contribution factor, is used to determine the relative predictive importance of the independent variables in predicting the network's output. The contribution factor of the n^{th} variable, CF_n is defined by Equation (1) as:

$$CF_n = \frac{\sum_{j=1}^{nH} ABS(w_{jn})}{\sum_{i=1}^{nG}\sum_{j=1}^{nH} ABS(w_{ji})} \times 100 \tag{1}$$

Where nG is the number of input variables, nH indicates the number of hidden nodes, w_{ji} are the weights from input layer i to the hidden layer j (as shown in Figure 5) and ABS refers to the absolute function. The summation of absolute values of network weights is used because some weight values may be positive and others are negative (Muttil and Chau, 2007).

3.4 Streamflow Forecasting and Assessment

In the current study, streamflow forecasting is achieved through developing the rainfall-runoff (R-R) relationship between the future streamflow at the catchment outlet, and rainfall and streamflow records available up to the current time t. Mathematically, the R-R relationship can be expressed as:

$$Q_{t+V\Delta t} = f(R_t, R_{t-\Delta t}, \cdots, R_{t-U\Delta t}, Q_t, Q_{t-\Delta t}, \cdots, Q_{t-U\Delta t}) \tag{2}$$

where Q is the streamflow (m^3/s), R is the rainfall (mm), V (with V = 1, 2, 3,……) denotes how far into the future the streamflow forecasting is sought, U (with U = 1, 2, 3,……) indicates how far back the recorded data in the time series likely affect the streamflow forecasts while Δt stands for time interval. The neural network structure for the ANN model as generalized in Figure 5 is used to forecast the one-day-ahead streamflow at the catchment outlet. It is important to note that for a simple demonstration of the proposed methodology, only one-day-ahead streamflow forecasting is undertaken in the current study and thus seven-days-ahead and/or seasonal forecast of streamflows are not the scope of this work.

In the current study, three different ANN-based streamflow forecasting models are formulated, which are described below:

- **ANN model-1:** This ANN model includes the current rain gauge network-based rainfall data together with streamflow observations as the input (see Figure 1). This model is designated as the base case for comparison in order to test the robustness and efficacy of the proposed approach.
- **ANN model-2:** This ANN model uses the optimal rain gauge network-based rainfall data along with streamflow observations as the input (see Figure 3). This model is indicated as the test case-1 wherein no additional fictitious stations are incorporated in the optimal network design.
- **ANN model-3:** This ANN model includes the augmented optimal rain gauge network-based rainfall data in combination with streamflow observations as the input (see Figure 4). This model is designated as the test case-2, in which additional fictitious stations are considered to augment the optimal rain gauge network.

The performance of each ANN model for streamflow forecasting is assessed and compared using four different evaluation metrics given in Equation (3)-(6): normalized root mean squared error (NRMSE), mean absolute error (MAE), Nash-Sutcliffe coefficient of efficiency (NSCE), correlation coefficient (CC). Further details on these metrics can be found in Dawson et al. (2007) and Moriasi et al. (2007).

$$\mathrm{NRMSE} = \sqrt{\frac{\sum_{t=1}^{N}[Q_{obs}(t) - Q_{est}(t)]^2}{\sum_{t=1}^{N}[Q_{obs}(t) - \bar{Q}_{obs}]^2}} \tag{3}$$

$$\mathrm{MAE} = \frac{\sum_{t=1}^{N}\|Q_{obs}(t) - Q_{est}(t)\|}{N} \tag{4}$$

$$\mathrm{NSCE} = 1 - \frac{\sum_{t=1}^{N}[Q_{obs}(t) - Q_{est}(t)]^2}{\sum_{t=1}^{N}[Q_{obs}(t) - \bar{Q}_{obs}]^2} \tag{5}$$

$$\mathrm{CC} = \frac{\sum_{t=1}^{N}[Q_{obs}(t) - \bar{Q}_{obs}] \times [Q_{est}(t) - \bar{Q}_{est}]}{\sqrt{\sum_{t=1}^{N}[Q_{obs}(t) - \bar{Q}_{obs}]^2} \times \sqrt{\sum_{t=1}^{N}[Q_{est}(t) - \bar{Q}_{est}]^2}} \tag{6}$$

where $Q_{obs}(t)$ is the observed streamflow at time t, $Q_{est}(t)$ is the forecasted streamflow at time t, $\bar{Q}_{obs}$ and $\bar{Q}_{est}$ are the mean value of observed and forecasted streamflow, respectively, and N is the number of observations in the time series data.

4. RESULTS AND DISCUSSION

4.1 Current and Optimal Rain Gauge Network-based Input for ANN Model

The three-layer feed-forward MLP neural network, as generalized in Figure 5, is first trained to formulate the ANN model-1using inputs that comprise data from the current rain gauge network-based rainfall and available streamflow records in the Middle Yarra River catchment (Figure 1). The neural network is then trained to formulate the ANN model-2 using inputs that include data from the optimal rain gauge network-based rainfall and available streamflow records in the catchment (Figure 3). As mentioned earlier, there are 19 rain gauges and 4 streamflow measuring stations in the current rain gauge network (Figure 1 and Table 1). The optimal rain gauge network as described in Adhikary et al. (2015) also consists of the same number of rain gauge and streamflow measuring stations (Figure 3) because no additional fictitious stations were considered for the optimal network design in that study. A major advantage of the optimal network is that stations are optimally located in the optimal network and hence it provides improved rainfall estimates (please see Adhikary et al., 2015 for details).

According to the Bransby-Williams formula (Wanielista et al., 1997), it is estimated that the catchment has a time of concentration of approximately 3 days. Rainfall occurring within a duration equal to time of concentration would exhibit the greatest influence on streamflows. In addition, streamflow values from the preceding duration provide the antecedent flow information prior to the onset of a rainfall event (Wu et al., 2005). Therefore, a time lag of 3 days (t, t-1 and t-2) is adopted in this study to obtain the time-lagged input (rainfall and streamflow) values for forecasting (t+1) streamflows at the catchment outlet. Hence, for each of the current and optimal networks, rainfall and streamflow data from 19+4 (=23) stations gives a total of 23*3 (= 69) inputs from which the significant input variables are to be selected to formulate the ANN model-1 and ANN model-2.

The input layer of the neural network for both the current and optimal networks consists of 69 nodes based on the 69 inputs. The output layer consists of a single node, which is streamflow at the catchment outlet that is to be forecasted. The neural networks are then trained with the training details and data division described earlier. The backpropagation training of the neural networks is terminated after 1000 epochs, which is found to be sufficient in this study. In order to find the optimum number of hidden nodes, a trial and error procedure is adopted in the training of neural networks by gradually varying the number of nodes in the hidden layer from 2 to 10. The optimal number of hidden nodes is found to be 6 for both the current and optimal rain gauge networks. Hence, the resulting neural network based on the current and optimal rain gauge network-based input has 69-6-1 structure.

ANN weights for each of the trained neural networks with 69-6-1 structure are obtained from the simulation. The ANN weights are inserted in Equation (1) to calculate the contribution factor of each of the 69 inputs for both the current and optimal rain gauge networks, which are presented in Table 2. The sum of the contribution factors of all the 69 input variables should be 100%, which can be seen in Table 2. The sum of the contribution factors of all the 69 input variables should be 100%, which can be seen in Table 2. As explained earlier, the definition of the contribution factor demonstrates that the higher its value for an input variable, the more that input contributes to the streamflow forecasting. In other words, if all input variables are considered to have equal significance, then each input exhibits a significance of 1/69 (equivalent to contribution factor of 1.45%) of the total contribution factor (= 100%) of all input variables. Thus, the input variables with a contribution factor greater than 1.45% are considered as the relatively more significant variables, which are indicated with the shaded color in Table 2.

Table 2. Contribution factor based on the trained ANN weights for the current and optimal rain gauge networks for one-day-ahead streamflow forecasting

Sl. no.	Current rain gauge network (see Figure 1) (BoM's existing base network)					Optimal rain gauge network (see Figure 3) (Additional fictitious stations are not considered in the network design)				
	Input Variables	Contribution factor* (*CFn*) of the input variables (%)			Sum	Input Variables	Contribution factor* (*CFn*) of the input variables (%)			Sum
		(t)	(t-1)	(t-2)			(t)	(t-1)	(t-2)	
1	R1	**2.30**	1.18	0.86	4.35	R1	**2.82**	1.41	**1.64**	5.88
2	R2	1.32	**1.84**	0.91	4.06	R2	**2.60**	1.38	0.82	4.80
3	R3	**2.09**	**1.57**	1.32	4.99	R3	**1.93**	**1.68**	1.42	5.03
4	R4	1.18	1.05	**1.50**	3.74	R4	1.04	1.24	0.83	3.12
5	R5	0.81	**2.28**	**1.58**	4.68	**R5b*****	**1.64**	1.01	0.97	3.61
6	R6	0.89	0.86	0.53	2.28	R6	1.28	0.97	0.86	3.10
7	R7	**2.15**	0.71	1.42	4.29	R7	**2.08**	1.00	1.22	4.31
8	R8	**1.79**	**2.52**	**1.55**	5.86	R8	**2.39**	**2.19**	**1.84**	6.43
9	R9	**1.61**	1.42	**1.51**	4.54	R9	1.09	1.02	**1.76**	3.87
10	R10	**2.49**	1.15	**1.79**	5.42	R10	1.16	1.39	0.89	3.44
11	R11	0.96	0.86	1.22	3.04	R11	**1.83**	**1.58**	**1.52**	4.93
12	R12	1.34	0.69	0.61	2.63	R12	1.32	0.92	0.78	3.02
13	R13	**1.90**	1.44	0.73	4.06	R13	0.79	1.40	**1.55**	3.74
14	R14	1.28	**2.22**	1.27	4.76	R14	0.84	0.79	1.16	2.79
15	R15	**2.05**	1.44	0.90	4.39	R15	1.22	0.68	**1.46**	3.36
16	R16	**1.56**	0.65	1.09	3.30	R16	**2.06**	0.64	1.24	3.94
17	R17	**1.86**	0.70	1.21	3.76	R17	**1.78**	1.08	0.76	3.63
18	R18	1.34	**1.98**	0.98	4.30	**R18a****	**2.10**	**1.62**	0.63	4.35
19	R19	**2.21**	0.56	0.92	3.70	**R19a****	**1.91**	0.83	1.34	4.08
20	S1	**1.76**	**2.53**	**1.58**	5.88	S1	**1.45**	**1.72**	**1.51**	4.68
21	S2	**2.68**	**2.83**	**1.76**	7.27	S2	**2.79**	**2.44**	**2.35**	7.58
22	S3	**2.91**	1.06	1.23	5.19	S3	**2.62**	1.30	1.03	4.96
23	S4	1.37	1.11	1.03	3.51	S4	**2.49**	**2.29**	0.57	5.35
		Sum of contribution of all variables =			100		Sum of contribution of all variables =			100

*Shaded cell shows variables having a contribution factor greater than 1/69 = 1.45%

**Optimal position of additional rain gauge stations (stations 18 and 19, see Figure 3) as identified by Adhikary *et al.* (2015)

***Optimal re-located position of redundant rain gauge station (station 5, see Figure 3) as identified by Adhikary *et al.* (2015)

It is evident from Table 2 that the influence of the significant input variables decreases in most cases with an increase of time lag for both the current and optimal rain gauge networks. It is also seen from the table that the optimally located stations in the optimal rain gauge network have higher contribution factor than that given by them in the current rain gauge network. In other words, the significant input variables based on the optimal network describes the outlet streamflow relatively better than what the current network does. This indicates the significance of incorporating the optimal rain gauge network-based input for accurate streamflow forecasting in a catchment, which is the main focus of the current study.

4.2 Augmented Optimal Rain Gauge Network-Based Input for ANN Model

The MLP neural network, as generalized in Figure 5, is also trained using inputs that include data from the augmented optimal rain gauge network-based rainfall and available streamflow values in the study catchment to formulate the ANN model-3. The augmented optimal network consists of 29 rain gauges and 4 streamflow measuring stations as shown in Figure 4. Therefore, rainfall and streamflow data from 19+10+4 (= 33) stations in the augmented optimal network gives a total of 33*3 (= 99) inputs considering the adopted 3 day time lag, from which the significant input variables are to be selected for the ANN model-3. Thus, the input layer of the neural network for the augmented optimal network comprises 99 nodes and the output layer consists of a single node based on the outlet streamflow that is to be forecasted. The optimal number of nodes in the hidden layer is found to be 4 based on the trial and error process by gradually varying the number of hidden nodes from 2 to 10. Thus, the resulting neural network has 99-4-1 structure for the augmented optimal network, which is trained using the same training specification and data division explained earlier. ANN weights of the trained neural network with 99-6-1 structure are then obtained from the simulation. The contribution factor of each of the 99 input variables is calculated using Equation (1) based on the ANN weights for the augmented optimal network, which is presented in Table 3.

Table 3. Contribution factor based on the trained ANN weights for the augmented optimal rain gauge network for one-day-ahead streamflow forecasting

Sl. no.	Augmented optimal rain gauge network (see Figure 4) (Additional fictitious stations are considered in the optimal network design)				
	Input Variables	Contribution factor* (*CFn*) of the input variables (%)			Sum
		(t)	(t-1)	(t-2)	
1	R1	0.82	0.84	0.87	2.53
2	R2	**2.36**	**1.11**	0.54	4.01
3	R3	**1.83**	**1.21**	0.87	3.91
4	R4	0.78	**1.46**	**1.43**	3.68
5	**R5b*** **	**1.22**	0.88	0.63	2.74
6	R6	0.60	0.64	**1.02**	2.26
7	R7	**1.79**	0.73	0.98	3.51
8	R8	**1.44**	**1.61**	**2.13**	5.18
9	R9	0.87	**1.28**	0.33	2.47
10	R10	0.57	**1.10**	0.59	2.26
11	R11	0.95	**1.51**	0.85	3.31
12	R12	**2.25**	0.64	0.55	3.44
13	R13	**1.35**	0.36	0.55	2.26
14	R14	0.25	0.24	0.98	1.46
15	R15	0.85	0.60	0.87	2.33
16	R16	**1.35**	0.26	0.58	2.20
17	R17	0.75	0.76	0.30	1.81
18	**R18a** **	0.74	**1.03**	**1.14**	2.91
19	**R19a** **	**1.61**	0.72	**1.06**	3.39
20	S1	**2.22**	**2.02**	**1.62**	5.86
21	S2	**2.55**	**2.55**	**2.00**	7.10
22	S3	**1.10**	**1.18**	0.89	3.17
23	S4	**1.41**	0.62	0.94	2.98
24	P1	**1.01**	0.46	**1.30**	2.78
25	P2	**1.40**	0.41	**1.22**	3.02
26	P3	0.73	0.55	**1.09**	2.37
27	P4	0.49	0.72	**1.02**	2.24
28	P5	0.32	**1.09**	0.38	1.80
29	P6	0.97	0.99	0.91	2.87
30	P7	0.39	0.80	**1.03**	2.22
31	P8	0.80	0.61	**1.22**	2.64
32	P9	0.49	0.84	0.82	2.14
33	P10	**1.30**	**1.20**	0.68	3.18
			Sum of contribution of all variables =		100

*Shaded cell shows variables having a contribution factor greater than 1/99 = 1.01%

**Optimal position of additional rain gauge stations (stations 18 and 19, see Figure 3) as identified by Adhikary *et al*. (2015)

***Optimal re-located position of redundant rain gauge station (station 5, see Figure 3) as identified by Adhikary *et al*. (2015)

The sum of the contribution factors of all the 99 input variables should be 100%, which can be seen in Table 3. In general, if all input variables are considered to have equal significance, then each input has a significance of 1/99 (equivalent to contribution factor of 1.01%) of the total contribution factor (= 100%) of all input variables. Thus, the input variables with a contribution factor greater than 1.01% are considered as the relatively more significant variables in this case, which are indicated with the shaded dark color in Table 3. As can be seen from the table, apart from the selected other significant input variables, some additional fictitious stations in the augmented optimal network are seen to have influence on the outlet streamflows. This indicates that the optimal locations of rain gauge stations should be decided in the final operational network after satisfying the objectives of accurate rainfall estimations as well as enhanced streamflow forecasting simultaneously.

4.3 Streamflow Forecasting with Current and Optimal Rain Gauge Network-Based Input

In order to forecast streamflow, the ANN-based streamflow forecasting models (i.e., ANN model-1 and ANN model-2) as explained by Equation (2) are formulated using the identified significant inputs for both the current and optimal rain gauge networks. It can be seen from Table 2 that 29 significant inputs are identified for the current rain gauge network whereas 30 significant inputs are identified for the optimal rain gauge network. The neural networks are trained once again with the training details and data division described earlier using the selected significant inputs. The optimum number of hidden neurons for the neural networks is also obtained through the trial and error process described earlier. The optimal number of hidden nodes is found to be 2 for both the ANN model-1 and ANN model-2. Hence, the ANN model-1 consists of 29-2-1 structure whereas the ANN model-2 has 30-2-1 structure. The neural network simulations of the ANN model-1 and ANN model-2 are then carried out to generate one-day-ahead streamflow forecasting at the catchment outlet. For both the ANN model-1 and ANN model-2, different performance evaluation measures are computed for the observed and predicted streamflows, which are presented in Table 4.

Table 4. Comparison of different performance measures of ANN modelling for the one-day-ahead streamflow forecasting

Rain gauge network and ANN models used	**Training phase**				**Validation phase**				**Testing phase**			
	NRMSE	**MAE**	**NSCE**	**CC**	**NRMSE**	**MAE**	**NSCE**	**CC**	**NRMSE**	**MAE**	**NSCE**	**CC**
Current rain gauge network (BoM's base network): **ANN model-1**	0.251	2.236	0.937	0.968	0.296	1.511	0.913	0.955	0.284	0.946	0.919	0.961
Optimal rain gauge network considering no additional fictitious stations: **ANN model-2**	0.190	1.620	0.964	0.982	0.248	1.327	0.939	0.969	0.264	0.923	0.930	0.969
Augmented optimal rain gauge network considering additional fictitious stations: **ANN model-3**	0.183	1.425	0.967	0.983	0.250	1.115	0.938	0.968	0.232	0.658	0.946	0.974

NRMSE - Normalized Root Mean Squared Error; MAE – Mean Absolute Error; NSCE – Nash-Sutcliffe Coefficient of Efficiency; CC – Correlation Coefficient

It is evident from the results presented in the table that the ANN model-2 produces more accurate streamflow forecasts than that produced by the ANN model-1. The improvement is significant in terms of the four performance evaluation measures (NRMSE, MAE, NSCE, and CC) in which NRMSE and MAE are decreased by 7.1% and 2.4%, respectively whereas NSCE is increased from 0.919 to 0.930 and CC is improved from 0.961 to 0.969 for the testing dataset. Scatter plots of the ANN model-1 and ANN model-2 forecasting results in the testing phase are presented in Figures 6a-6b. As can be seen from the figures, scatter points for the ANN model-2 forecasted values are located more closely with the 45^0 calibration line and thus, shows a relatively better agreement between the observed and predicted streamflows than that given by the ANN model-1 forecast. Furthermore, time series plots of the observed and simulated streamflows in the testing phase for the ANN model-1 and ANN model-2 are shown in Figures 7a-7b. It is evident from the figures that ANN model-2 exhibits better agreement between the observed and streamflow time series than ANN model-1. Figures 7a-7b (with the largest peak values zoomed) also indicate that the ability of ANN model-2 to capture the peak values are better than the ANN model-1. This conclusively proves that the optimal rain gauge network-based input (ANN model-2) produces better streamflow forecasts than the current rain gauge network-based inputs (ANN model-1) produce. All these findings reveal the effectiveness of using the optimal rain gauge network-based input in improving the streamflow forecasting of a catchment.

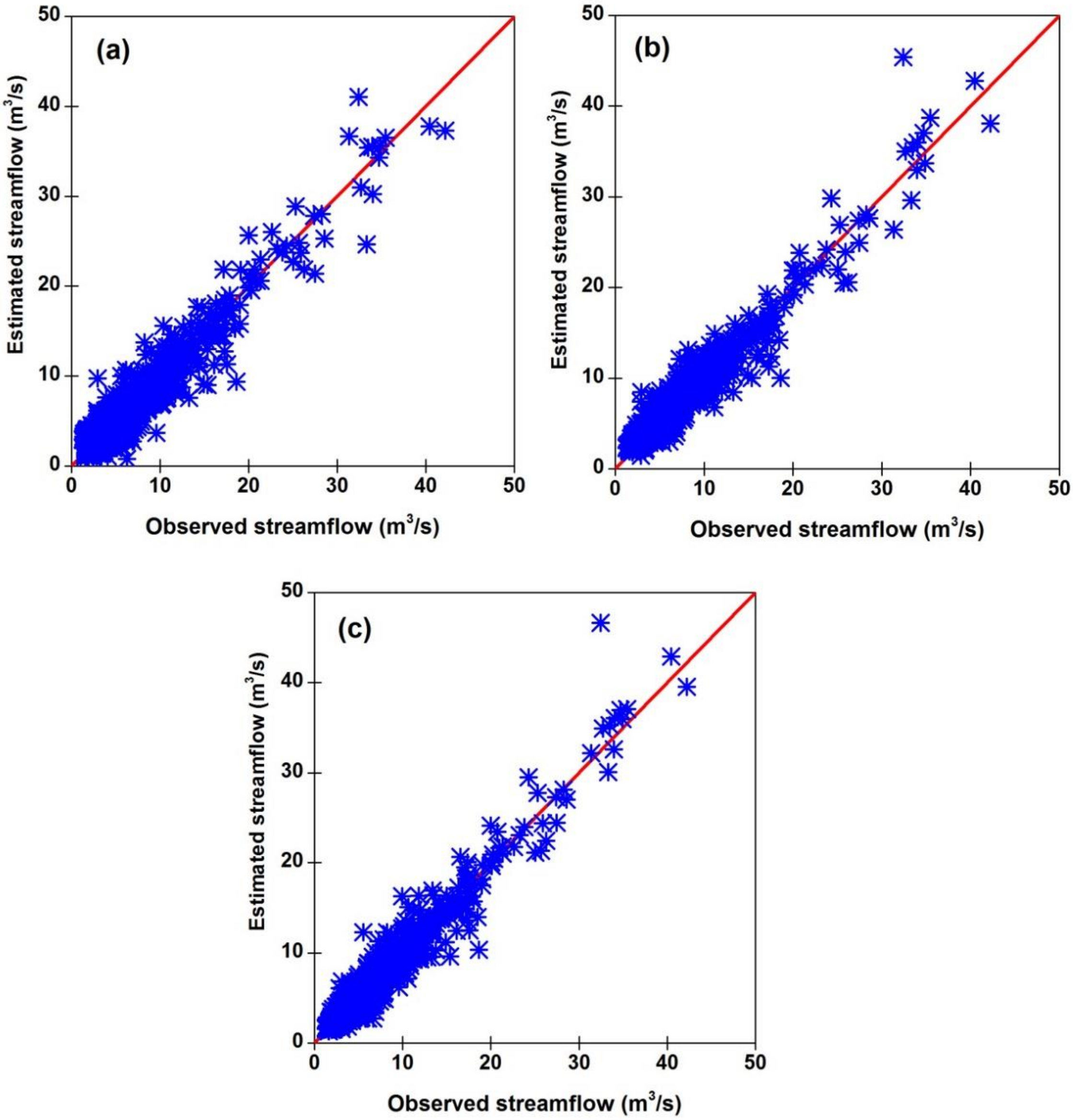

Figure 6. Scatter plots for the testing phase for (a) ANN model-1 based on the BoM's current rain gauge network, (b) ANN model-2 based on the optimal rain gauge network considering no additional fictitious stations (c) ANN model-3 based on the augmented optimal rain gauge network considering additional fictitious stations

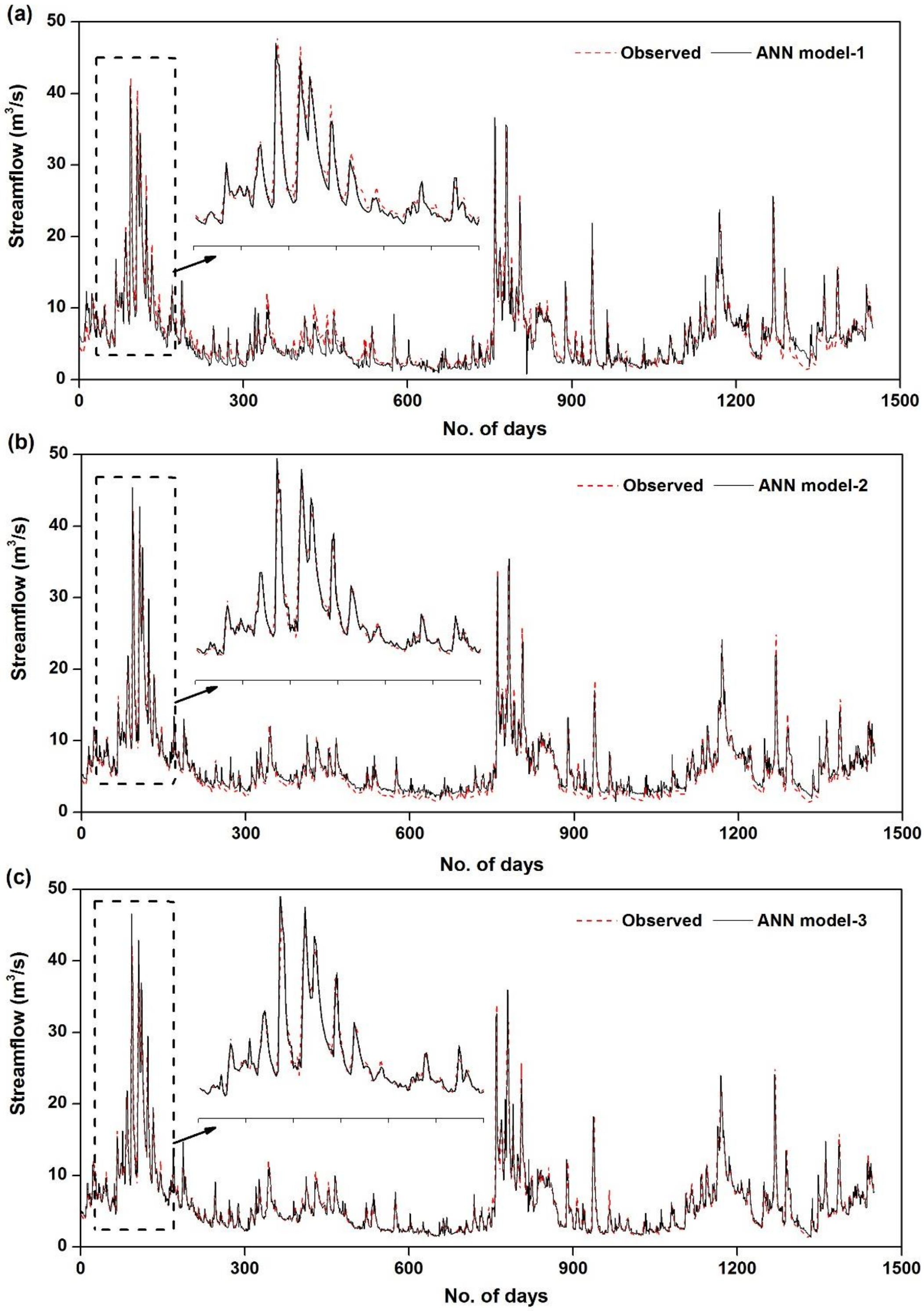

Figure 7. Time series plots of the observed streamflow vs simulated streamflow in the testing phase for (a) ANN model-1 based on the BoM's current rain gauge network, (b) ANN model-2 based on the optimal rain gauge network considering no additional fictitious stations (c) ANN model-3 based on the augmented optimal rain gauge network considering additional fictitious stations

4.4 Streamflow Forecasting with Augmented Optimal Rain Gauge Network-Based Input

ANN model-3 as explained by Equation (2) is formulated using the identified significant inputs for the augmented optimal rain gauge network. Table 3 shows that 42 inputs are selected as the significant inputs for the augmented optimal network. The neural networks are trained once again with the training details and data division described earlier using the 42 selected significant inputs. The optimum number of hidden neurons for the neural network is obtained through the trial and error process described earlier, which is found to be 2 for the ANN model-3. Thus, the ANN model-3 has 42-2-1 structure, which is then used to generate one-day-ahead streamflow forecasting at the catchment outlet. Different performance evaluation measures are also computed for the observed and the ANN model-3 forecasting results, which is shown in Table 4. As can be seen from the table, the ANN model-3 outperforms all the models to generate accurate streamflow forecasting. The improvement is significant when compared to ANN model-1, in terms of the four performance evaluation measures (NRMSE, MAE, NSCE, and CC) in which NRMSE and MAE are reduced by 18.3% and 30.4%, respectively whereas NSCE is improved from 0.919 to 0.946 and CC is improved from 0.961 to 0.974 for the testing dataset. Although the improvement by the ANN model-3 is not significant compared to the ANN model-2, this gives an important insight about the usage of an augmented rain gauge network for the enhanced streamflow forecasting.

For the ANN model-3 forecasting results, the scatter plot in the testing phase as shown in Figure 6c also shows that the best results is achieved through this model. It is also evident from the time series plot for the ANN model-3 in the testing phase as shown in Figure 7c that the observed and simulated streamflows have the best agreement. As can be also seen from the Figures 7a-7c that the ability of ANN model-3 to capture the peak values are better than the ANN model-1 and ANN model-2. These results indicate that improved streamflow forecasting can be achieved through an augmented rain gauge network. In other words, these results demonstrate that one should obtain the final operational rain gauge network from this augmented network, which is able to provide accurate rainfall estimates (rain gauge network design objective) as well as give the enhanced streamflow forecasting (flow forecasting objective) simultaneously. However, it is emphasized that if cost is a concern, the optimal network, which provides significant improvement in streamflow forecasting over the current network, should be used in practice for flow forecasting since the optimal network consists of the optimal number of stations. The reason is that unlike the previous studies, the optimal network used in this study was designed without incorporating the additional fictitious rain gauge stations. Again, if forecasting accuracy is taken as the primary objective, the augmented network is recommended for flow forecasting in practice.

5. Conclusions

Four conclusions can be drawn based on the findings of the current study:

- The proposed approach of using the rainfall input to ANN-based streamflow forecasting models from the optimal rain gauge network appears to be effective for the enhanced streamflow forecasting, particularly when the current operational rain gauge network is not optimal. The study conclusively proves the significance of the optimal location of rain gauge station in a catchment for enhanced streamflow forecasting.
- The optimal locations of rain gauge stations in the final operational optimal network should be established after satisfying the accurate rainfall estimations and improved streamflow forecasting objectives simultaneously. The network design based on only accurate rainfall estimations objective may not always guarantee accurate streamflow forecasting.
- Further improvement of forecasting performance can be achieved through expansion or augmentation of the rain gauge network considering additional fictitious rain gauge stations. In fact, the best forecasting performance are achieved in this study when the augmented rain gauge network-based input is used in the ANN-based streamflow forecasting models.
- ANN-based input variable selection offers an indirect way of identifying the optimal locations of rain gauge stations in the final operational rain gauge network. The optimal locations of rain gauge stations can be identified from

an augmented or expanded network by checking the selected significant input variables.

ACKNOWLEDGEMENTS

The authors acknowledge the financial support from the Australian Government and Victoria University, Melbourne through an International Postgraduate Research Scholarship (IPRS) scheme to carry out this study. The authors are also thankful to five anonymous reviewers for their valuable comments and suggestions, which have improved the quality of the paper.

REFERENCES

Abrahart, R.J., Anctil, F., Coulibaly, P., Dawson, C.W., Mount, N.J., See, L.M., Shamseldin, A.Y., Solomatine, D.P., Toth, E. & Wilby, R.L. 2012 Two decades of anarchy? Emerging themes and outstanding challenges for neural network river forecasting. *Prog. Phy. Geo.* **36(4),** 480-513. DOI: 10.1177/0309133312444943.

Adhikary, S.K., Yilmaz, A.G. & Muttil, N. 2015 Optimal design of rain gauge network in the Middle Yarra River catchment, Australia. *Hydrol. Process.* **29(11),** 2582-2599. DOI: 10.1002/hyp.10389.

Adhikary, S.K., Muttil, N. &Yilmaz, A.G. 2016a Ordinary kriging and genetic programming for spatial estimation of rainfall in the Middle Yarra River catchment, Australia. *Hydrol. Res.* **47(6),** 1182-1197. DOI: 10.2166/nh.2016.196.

Adhikary, S.K., Muttil, N. & Yilmaz, A.G. 2016b Genetic programming-based ordinary kriging for spatial interpolation of rainfall. *J. Hydrol. Eng.* **21(2),** 04015062. DOI: 10.1061/(ASCE)HE.1943-5584.0001300.

Adhikary, S.K., Muttil, N. &Yilmaz, A.G. 2017 Cokriging for enhanced spatial interpolation of rainfall in two Australian catchments. *Hydrol. Process.* **31(12),** 2143-2161. DOI: 10.1002/hyp.11163.

Akhtar, M.K., Corzo, G.A., van Andel, S.J. & Jonoski, A. 2009 River flow forecasting with artificial neural networks using satellite observed precipitation pre-processed with flow lengths and travel time information: case study of the Ganges River basin. *Hydrol. Earth Syst. Sci.* **13,** 1607-1618. DOI: 10.5194/hess-13-1607-2009.

Anctil, F., Michel, C., Perrin, C. & Andréassian, V. 2004 A soil moisture index as an auxiliary ANN input for streamflow forecasting. *J. Hydrol.* **286(1-4),** 155-167. DOI: 10.1016/j.jhydrol.2003.09.006.

Anctil, F., Lauzon, N., Andréassian, V., Oudin, L. & Perrin, C. 2006 Improvement of rainfall-runoff forecasts through mean areal rainfall optimization. *J. Hydrol.* **328(3-4),** 717-725. DOI: 10.1016/j.jhydrol.2006.01.016.

Andréassian, V., Perrin, C., Michel, C., Usart-Sanchez, I. & Lavabre, J. 2001 Impact of imperfect rainfall knowledge on the efficiency and the parameters of watershed models. *J. Hydrol.* **250(1-4),** 206-223. DOI: 10.1016/S0022-1694(01)00437-1.

ASCE Task Committee. 2000a Artificial neural networks in hydrology. I: Preliminary concepts. *J. Hydrol. Eng.* **5(2),** 115-137. DOI: 10.1061/(ASCE)10840699(2000)5:2(115).

ASCE Task Committee. 2000b Artificial neural networks in hydrology. II: Hydrologic applications. *J. Hydrol. Eng.* **5(2),** 115-123. DOI: 10.1061/(ASCE)10840699(2000)5:2(124).

Bárdossy, A. & Das, T. 2008 Influence of rainfall observation network on model calibration and application. *Hydrol. Earth Syst. Sci.* **12,** 77-89. DOI: 10.5194/hess-12-77-2008.

Barua, S., Muttil, N., Ng, A.W.M. & Perera, B.J.C. 2012 Rainfall trend and its implications for water resource management within the Yarra River catchment, Australia. *Hydrol. Process.* **27(12),** 1727-1738. DOI: 10.1002/hyp.9311.

Bastin, G., Lorent, B., Duqué, C. & Gevers, M. 1984 Optimal estimation of the average areal rainfall and optimal selection of rain gauge locations. *Water Resour. Res.* **20(4),** 463-470. DOI: 10.1029/WR020i004p00463.

Birikundavyi, S., Labib, R., Trung, H.T. & Rousselle, J. 2002 Performance of neural networks in daily streamflow forecasting. *J. Hydrol. Eng.* **7(5),** 392-398. DOI: 10.1061/(ASCE) 1084-0699(2002)7:5(392).

Bowden, G.J., Dandy, G.C. & Maier, H.R. 2005 Input determination for neural network models in water resources applications. Part 1 – background and methodology. *J. Hydrol.* **301(1-4),** 75-92. DOI: 10.1016/j.jhydrol.2004.06.021.

Bras, R.L. & Rodriguez-Iturbe, I. 1976 Network design for the estimation of areal mean of rainfall events. *Water Resour. Res.* **12(6),** 1185-1195. DOI: 10.1029/WR012i006 p01185.

Bras, R.L. 1979 Sampling of interrelated random fields: the rainfall-runoff case. *Water Resour. Res.* **15(6),** 1767–1780. DOI: 10.1029/WR015i006p01767.

Chen, Y.C., Wei, C. & Yeh, H.C. 2008 Rainfall network design using kriging and entropy. *Hydrol. Process.* **22,** 340-346. DOI: 10.1002/hyp.6292.

Cheng, K.S., Lin, Y.C. & Liou, J.J. 2008 Rain-gauge network evaluation and augmentation using geostatistics. *Hydrol. Process.* **22,** 2554-2564. DOI: 10.1002/ hyp.6851.

Daly, E., Kolotelo, P., Schang, C., Osborne, C.A., Coleman, R., Deletic, A. & McCarthy, D.T. 2013 Escherichia coli concentrations and loads in an urbanised catchment: The Yarra River, Australia. *J. Hydrol.* **497,** 51-61. DOI: 10.1016/j.jhydrol.2013.05.024.

Dawson, C.W. & Wilby, R.L. 2001 Hydrological modelling using artificial neural networks. *Prog. Phy. Geo.* **25(1),** 80-108. DOI: 10.1177/030913330102500104.

Dawson, C.W., Abrahart, R.J. & See, L.M. 2007 HydroTest: A web-based toolbox of evaluation metrics for the standardised assessment of hydrological forecasts. *Env. Model. Soft.* **22(7),** 1034-1052. DOI: 10.1016/j.envsoft.2006.06.008.

Dibike, Y.B. & Solomatine, D.P. 2001 River flow forecasting using artificial neural networks. *Phy. Chem. Earth, Part B: Hydrol. Ocean. Atmosph.* **26(1),** 1-7. DOI: 10.1016/S1464-1909(01)85005-X.

Dong, X., Dohmen-Janssen, M. & Booij, M.J. 2005 Appropriate spatial sampling of rainfall for flow simulation. *Hydrol. Sci. J.* **50(2),** 279-298. DOI: 10.1623/hysj.50.2.279.61801.

Ekström, M. & Jones, P.D. 2009 Impact of rainfall estimation uncertainty on streamflow estimations for catchments Wye and Tyne in the United Kingdom. *Int. J. Clim.* **29,** 79-86. DOI: 10.1002/joc.1665.

Faurès, J.M., Goodrich, D., Woolhiser, D.A. & Sorooshian, S. 1995 Impact of small-scale spatial rainfall variability on runoff modelling. *J. Hydrol.* **173,** 309-326. DOI: 10.1016/0022-1694(95)02704-S.

Feki, H., Slimani, M. & Cudennec, C. 2017 Geostatistically based optimization of a rainfall monitoring network extension: case of the climatically heterogeneous Tunisia. *Hydrol. Res*. **48(2),** 514-541. DOI: 10.2166/nh.2016.256.

Govindaraju, R.S. & Rao, A.R. (ed). 2000 *Artificial Neural Networks in Hydrology.* Kluwer Academic Publishers: Boston.

Huang, W., Xu, B. & Chan-Hilton, A. 2004 Forecasting flows in Apalachicola River using neural networks. *Hydrol. Process.* **18,** 2545-2564. DOI: 10.1002/hyp.1492.

Jeffrey, S.J., Carter, J.O., Moodie, K.B. & Beswick, A.R. 2001 Using spatial interpolation to construct a comprehensive archive of Australian climate data. *Env. Model. Soft.* **16,** 309-330. DOI: 10.1016/S1364-8152(01)00008-1.

Khu, S., Liong, S.Y., Babovic, V., Madsen, H. & Muttil, N. 2001 Genetic programming and its application in real-time runoff forecasting. *J. Am. Water Resour. Assoc.* **37(2),** 439-451. DOI: 10.1111/j.1752-1688.2001.tb00980.x.

Kişi, Ö. 2007 Streamflow forecasting using different artificial neural network algorithms. *J. Hydrol. Eng.* **12(5),** 532-539. DOI: 10.1061/(ASCE)10840699(2007) 12:5(532).

Kumar, D.N., Raju, K.S. & Sathish, T. 2004 River flow forecasting using recurrent neural networks. *Water Resour. Manage.* **18(2),** 143-161. DOI: 10.1023/B:WARM. 0000024727.94701.12.

Linares-Rodriguez, A., Lara-Fanego, V., Pozo-Vazquez, D. & Tovar-Pescador, J. 2015 One-day-ahead streamflow forecasting using artificial neural networks and a meteorological mesoscale model. *J. Hydrol. Eng.* **20(9),** 05015001. DOI: 10.1061/(ASCE)HE.19435584. 0001163.

Londhe, S. & Charhate, S. 2010 Comparison of data-driven modelling techniques for river flow forecasting. *Hydrol. Sci. J.* **55(7),** 1163-1174. DOI: 10.1080/02626667. 2010.512867.

Maier, H.R. & Dandy, G.C. 2000 Neural networks for the prediction and forecasting of water resources variables: a review of modelling issues and applications. *Env. Model. Soft.* **15(1),** 101-124. DOI: 10.1016/S1364-8152(99)00007-9.

Maier, H.R., Jain, A., Dandy, G.C. & Sudheer, K.P. 2010 Methods used for the development of neural networks for the prediction of water resource variables in river systems: current status and future directions. *Env. Model. Soft.* **25(8),** 891-909. DOI: 10.1016/j.envsoft. 2010.02.003.

Mishra, A.K. & Coulibaly, P. 2009 Developments in hydrometric network design: a review. *Rev. Geophy.* **47**, RG2001. DOI: 10.1029/2007RG000243.

Moriasi, D.N., Arnold, J.G., van Liew, M.W., Bingner, R.L., Harmel, R.D. & Veith, T.L. 2007 Model evaluation guidelines for systematic quantification of accuracy in watershed simulations. *Tran. ASABE* **50(3),** 885-900. DOI: 10.13031/2013.23153.

Moulin, L. Gaume, E. & Obled, C. 2009 Uncertainties on mean areal precipitation: assessment and impact on streamflow simulations. *Hydrol. Earth Syst. Sci.* **13,** 99-114. DOI: 10.5194/hess-13-99-2009.

Muttil, N. & Chau, K.W. 2006 Neural network and genetic programming for modelling coastal algal blooms. *Int. J. Env. Pollut.* **28(3-4),** 223-238. DOI: 10.1504/IJEP. 2006.011208.

Muttil, N. & Chau, K.W. 2007 Machine learning paradigms for selecting ecologically significant input variables. *Eng. App. Artifi. Intell.* **20,** 735-744. DOI: 10.1016/ j.engappai.2006.11.016.

Oudin, L., Michel, C. & Anctil, F. 2005 Which potential evapotranspiration input for a lumped rainfall-runoff model?: Part 1 - Can rainfall-runoff models effectively handle detailed potential evapotranspiration inputs? *J. Hydrol.* **303(1-4),** 275-289. DOI: 10.1016/j.jhydrol. 2004.08.025.

Oudin, L., Perrin, C., Mathevet, T., Andréassian, V. & Michel, C. 2006 Impact of biased and randomly corrupted inputs on the efficiency and the parameters of watershed models. *J. Hydrol.* **320(1-2),** 62-83. DOI: 10.1016/j.jhydrol.2005.07 .016.

Papamichail, D.M. & Metaxa, I.G. 1996 Geostatistical analysis of spatial variability of rainfall and optimal design of rain gauge network. *Water Resour. Manage.* **10(2),** 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. *J. Hydrol.* **210(1-4),** 206-220. DOI: 10.1016/S0022-1694(98)00188-7.

Porporato, A. & Ridolfi, L. 2001 Multivariate nonlinear prediction of river flows. *J. Hydrol.* **248(1-4),** 109-122. DOI: 10.1016/S0022-1694(01)00395-X.

Seed, A.W. & Austin, G.L. 1990 Sampling errors for rain gauge-derived mean areal daily and monthly rainfall. *J. Hydrol.* **118(1-4),** 163-173. DOI: 10.1016/00221694 (90)90256-W.

Shaghaghian M, Abedini M. 2013. Rain gauge network design using coupled geostatistical and multivariate techniques. *Scientia Iranica* **20(2)**: 259-269. DOI: 10.1016/j.scient.2012.11.014.

Sivapragasam, C., Vanitha, S., Muttil, N., Suganya, K., Suji, S., Selvi, M.T. & Sudha, S.J. 2014 Monthly flow forecast for Mississippi River basin using artificial neural networks. *Neu. Comp. App.* **24(7),** 1785-1793. DOI: 10.1007/s00521-013-1419-6.

Srinivasulu, S. & Jain, A. 2009 River flow prediction using an integrated approach. *J. Hydrol. Eng.* **14(1),** 75-83. DOI: 10.1061/(ASCE)1084-0699(2009)14:1(75).

Storm, B., Høgh, J.K. & Refsgaard, J.C. 1989 Estimation of catchment rainfall uncertainty and its influence on runoff prediction. *Hydrol. Res.* **19(2),** 77-88.

St-Hilaire, A., Ouarda, T.B.M.J., Lachance, M., Bobee, B., Gaudet, J. & Gignac, C. 2003 Assessment of the impact of meteorological network density on the estimation of basin precipitation and runoff: a case study. *Hydrol. Process.* **17(18),** 3561-3580. DOI: 10.1002/hyp.1350.

Talei, A., Chua, L.H.C. & Wong, T.S.W. 2010 Evaluation of rainfall and discharge inputs used by Adaptive Network-based Fuzzy Inference Systems (ANFIS) in rainfall-runoff modeling. *J. Hydrol.* **391(3-4),** 248-262. DOI: 10.1016/j.jhydrol. 2010.07.023.

Taormina, R., Chau, K.W. & Sivakumar, B. 2015 Neural network river forecasting through baseflow separation and binary-coded swarm optimization. *J. Hydrol.* **529,** 1788-1797. DOI: 10.1016/j.jhydrol.2015.08.008.

Tayfur, G. 2012 *Soft Computing in Water Resources Engineering: Artificial Neural Networks, Fuzzy Logic and Genetic Algorithms*. WIT Press: Southampton, UK.

Tsai, M.J., Abrahart, R.J., Mount, N.J. & Chang, F.J. 2014 Including spatial distribution in a data-driven rainfall-runoff model to improve reservoir inflow forecasting in Taiwan. *Hydrol. Process.* **28(3),** 1055-1070. DOI: 10.1002/ hyp.9559.

Tsintikidis, D., Georgakakos, K.P., Sperfslage, J.A., Smith, D.E. & Carpenter, T.M. 2002 Precipitation uncertainty, rain gauge network design within Folsom Lake watershed. *J. Hydrol. Eng.* **7(2),** 175-184. DOI: 10.1061/(ASCE)10840699(2002)7: 2(175).

Volkmann, T.H.M., Lyon, S.W., Gupta, H.V. & Troch, P.A. 2010 Multicriteria design of rain gauge networks for flash flood prediction in semiarid catchments with complex terrain. *Water Resour. Res.* **46(11),** W11554. DOI: 10.1029/2010WR 009145.

Wanielista, M., Kersten, R. & Eaglin, R. 1997 *Hydrology: Water Quantity and Quality Control*, (2nd Ed). John Wiley & Sons: New York.

Wu, J.S., Han, J., Annambhotla, S. & Bryant, S. 2005 Artificial neural networks for forecasting watershed runoff and streamflows. *J. Hydrol. Eng.* **10(3),** 216-222. DOI: 10.1061/(ASCE)1084-0699(2005)10:3(216).

Xu, C.Y., Tunemar, L., Chen, Y.D. & Singh, V.P. 2006 Evaluation of seasonal and spatial variations of conceptual hydrological model sensitivity to precipitation data errors. *J. Hydrol.* **324(1-4),** 80-93. DOI: 10.1016/j.jhydrol.2005.09.019.

Xu, H., Xu, C.Y., Chen, H., Zhang, Z. & Li, L. 2013 Assessing the influence of rain gauge network density and distribution on hydrological model performance in a humid region of China. *J. Hydrol.* **505,** 1-12. DOI: 10.1016/j.jhydrol.2013.09.004.

Yilmaz, A.G. & Muttil, N. 2014 Runoff estimation by machine learning methods and application to the Euphrates Basin in Turkey. *J. Hydrol. Eng.* **19(5),** 1015–1025. DOI: 10.1061/(ASCE)HE.1943-5584.0000869.

Zealand, C.M., Burn, D.H. & Simonovic, S.P. 1999 Short-term streamflow forecasting using artificial neural networks. *J. Hydrol.* **214(1-4),** 32-48. DOI: 10.1016/ S00221694(98) 00242-X.

Chapter 7

Summary, Conclusions and Recommendations for Future Study

7.1 Summary

Australia is called the land of 'drought and flooding rains'. This contrast was more evident than ever in the last two decades, when the extensive 1997-2009 Millennium Drought was followed by the occurrence of a series of large-scale devastating floods in 2010-2011. These caused severe havoc in eastern and southeastern Australia where the country's population and agricultural production are highly concentrated. Water resources play a vital role in the economic development of the region. The region's explosive population growth and resulting new demands on limited water resources require efficient management of existing water resources rather than building new facilities to meet the challenge. These issues have created great challenges to effective management of water resources particularly in eastern and southeastern Australia. In the water management communities, it is well-known that to address the aforementioned water management challenges, maximizing water management efficiency based on streamflow forecasting is crucial.

Streamflow forecasting is of vital importance to flood control and mitigation and water resources planning and management. Streamflow forecasting is usually challenging due to the complexity of hydrologic systems. However, a key challenge remains to achieve the enhanced accuracy in streamflow forecasting. The accuracy of streamflow forecasting mainly depends on the input data, wherein the catchment rainfall causes the largest impact on streamflow. Since streamflow is a consequence of rainfall, uncertainty associated with rainfall can adversely affect the accuracy of streamflow forecasting. This highlights the importance of using accurate rainfall data in streamflow forecasting models to achieve the enhanced streamflow forecasting. It is now widely recognized in the hydrological communities that an optimal rain gauge network is able to provide high quality rainfall estimates. Thus, it can be hypothesized that incorporating the rainfall input from an optimal rain gauge network rather than from an existing rain gauge network (which may not be an optimal network) in the streamflow forecasting models would enhance the accuracy in streamflow forecasting. Therefore, this thesis focused on the optimal design of a rain gauge network and then use of the optimal network-based rainfall input along with streamflow values in streamflow forecasting models in order to improve the accuracy in streamflow forecasting.

This study concentrates on investigating and providing potential solutions to the following issues associated with, rain gauge network design, spatial estimation of rainfall, and streamflow forecasting: (1) potential of using universal function approximation techniques in variogram modelling and kriging interpolation of rainfall, (2) performance of different univariate and multivariate kriging methods for enhanced spatial interpolation of rainfall, and potential of using auxiliary information (elevation) in addition to rainfall in kriging method for improved estimation of rainfall, (3) use of network augmentation (using additional rain gauge stations) as well as network rationalization (eliminating or re-locating redundant rain gauge stations) in the optimal rain gauge network design, and (4) potential of incorporating rainfall input from an optimal rain gauge network in the ANN-based streamflow forecasting models to achieve the enhanced streamflow forecasting.

The aforementioned aims of the study were demonstrated through a case study area, which consists of the middle part of the Yarra River catchment (referred to as the 'Middle Yarra River catchment' in this thesis) in Victoria, Australia. The extensive clearing of lands due to agricultural and urban development in the middle and lower segment of the Yarra River catchment often results in high river flows that causes frequent flash flooding in the study area and the urbanized lower segment where Melbourne city is located. The study area also contains three major storage reservoirs, which support water supplies to domestic, agricultural, industrial and environmental purposes. Hence, the enhanced estimation of accurate rainfall and future streamflows specifically in the middle and upper segments of the Yarra River catchment was identified as an absolute and timely need. In this study, 19 rain gauge stations and 4 streamflow measuring stations of the Yarra River located within the study area were considered for the analysis. All rainfall and streamflow time series data were collected at daily time steps and streamflow forecasting models detailed in this study were developed using the ANN-based data-driven modelling technique.

Hydrological investigations often require the estimates of hydrologic variables such as rainfall at ungauged locations in a catchment where no direct observations are readily available. Variance dependent stochastic interpolation methods such as kriging are the most commonly used methods for this purpose. In traditional kriging, accurate kriging results highly depend on the fitting and estimation of a correct variogram model. Selection of an appropriate variogram model, finding the optimal variogram parameters (i.e., nugget, sill and range coefficients) and the associated computational burdens are some difficulties involved in the traditional kriging. As a potential solution to these issues, a new universal function approximation-based kriging was developed using GP in this study where GP was used as a universal function approximator to derive the variogram model. This new variant of kriging is referred to as the GPOK method in which the standard parametric variogram models (i.e., exponential, gaussian, spherical models) in traditional ordinary kriging were replaced by the GP-derived variogram model. The performance of the GPOK method was compared with the traditional and another universal function approximation technique such as ANN-based ordinary kriging methods. It was found that the GPOK method demonstrated in this study overcomes the limitations associated with the traditional ordinary kriging and outperforms in estimating rainfalls at ungauged locations. It was also found that variogram modelling using GP offers several advantages. For example, GP does not require a pre-defined mathematical form or architecture unlike ANN to generate the functional variogram models. In addition, GP can generate variogram models, which consists of similar mathematical structure as having with the standard variogram models. Furthermore, GP-derived variogram model does not require identifying the variogram parameters in advance, unlike the standard parametric variogram models in traditional kriging. Therefore, the GPOK was found to be completely free from the tedious trial and error process of estimating the variogram parameters. The function approximation capability of GP generates the best fitted GP-derived variogram models compared to the standard models, which were found to be the best functional variogram models. Thus, the GP-derived variogram models offer a viable alternative to the existing standard variogram models for kriging interpolation of rainfall.

A key advantage of kriging over deterministic and other conventional interpolation methods is that while providing the kriging variance for rain gauge network design, it is capable of complementing the sparsely sampled primary variable (such as rainfall) by the correlated densely sampled auxiliary information (such as elevation) to improve the estimation accuracy of primary variable. This multivariate extension of kriging is referred to as the cokriging method, which was investigated in this study for enhanced spatial interpolation of rainfall. Based on the rainfall-topography relationship of the case study catchment, the elevation was used as an auxiliary variable in addition to rainfall in the cokriging methods. However, it is often challenging to select the best interpolation method arbitrarily from a wide variety of available kriging and deterministic interpolation methods for estimating the spatial distribution of rainfall for a particular area because the performance of an interpolation method depends on many factors. As a solution to this issue, a comparative evaluation of a range of univariate and multivariate kriging and deterministic interpolation methods was performed in this study to identify the most appropriate interpolator for enhanced spatial interpolation of rainfall and generation of continuous rainfall maps. Following a comparison of performances of all interpolation methods, it was found that the ordinary cokriging using elevation information along with rainfall outperformed the other methods and was found to be the most suitable interpolator for enhanced estimation of spatial distribution of rainfall in the Middle Yarra River catchment.

Conventionally, increasing network density through network augmentation using additional rain gauge stations to reduce the kriging variance of the network is generally used for the design of a rain gauge network. However, it is likely that a rain gauge network may consist of redundant stations, which have little or no contribution to the network performance for providing quality rainfall estimates. Since operation and maintenance of a rain gauge station involves large costs, removal of redundant stations can contribute to substantial cost reductions. Therefore, for optimal design of a rain gauge network using the kriging-based geostatistical approach based on the variance reduction principle, it is essential to consider both additional and redundant stations simultaneously in order to achieve the variance reduction and cost-effectiveness objectives of the optimal network. As a solution to this issue, a simple and effective rain gauge network design technique considering both additional and redundant stations, which involves a methodical search for the optimal number and locations of rain gauge stations in the network that minimize the kriging variance of the network. The technique allows achieving the optimal rain gauge network for the case study catchment through optimal positioning of additional stations (network augmentation) as well as removing and/or optimally relocating of existing redundant stations (network rationalization). Furthermore, the spatial variability of rainfall in the case study catchment is mainly caused by the El Niño and La Niña processes of ENSO phenomenon. As a solution to this issue, the rain gauge network was designed separately based on rainfall records obtained for both El Niño and La Niña periods and the network that gave the enhanced estimates of areal average and point rainfall values was chosen as the optimal rain gauge network. Following a comparison of performances between the optimal and existing rain gauge networks, it was found that the optimal network outperformed the existing one based on the spatiotemporal estimation of rainfalls.

In general, rainfall is considered independent of streamflow forecasting in many hydrological studies such as the optimal design of a rain gauge network for estimation of areal average rainfall over a catchment or area. However, this does not allow one to know the strength and weakness of the designed optimal network when the optimal network-based rainfall data are used in streamflow forecasting models. Therefore, it is highly rational to design an optimal rain gauge network for providing a satisfactory solution to the specific need of enhanced streamflow forecasting for which the network is being designed. In practice, rainfall data are fed into streamflow forecasting models directly from the existing rain gauge network that might exhibit high variance and hence not be the optimal one. This ultimately adversely affects the performance of streamflow forecasting models and thus results in the less accurate streamflow forecasts. As a solution to this issue, an ANN-based enhanced streamflow forecasting approach was developed and demonstrated in this study, which incorporated the optimal rain gauge network-based input instead of using existing non-optimal rain gauge network-based input in order to achieve the enhanced accuracy in streamflow forecasting. The proposed approach was found to be highly effective in improving the accuracy in streamflow forecasting, particularly when the current operational rain gauge network is not an optimal one. Further improvement in streamflow forecasting was achieved through the expansion or augmentation of the optimal rain gauge network considering additional fictitious rain gauge stations. The catchments were divided into a number of sub-catchments based on the digital elevation model and then the fictitious stations were placed in sub-catchments (where there is no station) and streamflow forecasting was performed. The ANN-based input selection technique that was also demonstrated in this study offers a viable technique for selecting significant input variables for data-driven modelling as this technique is capable of learning problems involving very non-linear and complex data. Furthermore, the input contribution factor of selected significant input variables obtained from the technique gives an indirect measure by which optimal locations of stations from a rain gauge network can be identified, giving an indirect way of rain gauge network design for enhanced streamflow forecasting.

7.2 Conclusions

The following conclusions are drawn based on the findings obtained from this study:

- In a rain gauge network design exercise, network augmentation (using additional stations) as well as network rationalization (eliminating or re-locating redundant stations) and identification of optimal locations of both additional and redundant stations are seen as a potential way to achieve the optimal rain gauge network. In other words, the additional fictitious stations that contribute to reduce kriging variance of the network can be selected. And the redundant stations that have little or no contribution to the network variance reduction can be either eliminated or optimally located in the high variance areas of the network.

- In this study, it was found that relocation and optimal placement of only redundant stations in the high variance areas of the rain gauge network gives the maximum reduction of the kriging variance and the best estimates of areal average and point rainfalls. This concludes that optimal placement of redundant stations in addition to additional stations is vital to achieve an optimal rain gauge network.
- The enhanced streamflow forecasting approach developed in this study incorporates the input from an optimal rain gauge network. This approach was found to be highly effective in improving the streamflow forecasting accuracy, particularly when the current operational rain gauge network is not an optimal one.
- The ANN-based input selection technique that was employed in this study for streamflow forecasting offers a viable technique for significant input variables selection. This technique is capable of learning problems involving very non-linear and complex data.
- Further improvement in streamflow forecasting was achieved through expansion or augmentation of the optimal rain gauge network by incorporating additional fictitious rain gauge stations. The fictitious stations were added in sub-catchments that were delineated based on the digital elevation model.
- The genetic programming-based ordinary kriging (GPOK) method demonstrated in this study, which uses the GP-derived variogram model to replace the standard variogram models (i.e., exponential, gaussian, spherical etc.) in traditional kriging, was found to be the most suitable technique for improved estimation of rainfall at ungauged locations. This study conclusively proves that the fusion of geostatistical (ordinary kriging) and data-driven (genetic programming) techniques has a high potential for spatial surface interpolation.

- Use of GP as a universal function approximator in variogram modelling is beneficial as GP-based variogram modelling does not require a pre-defined mathematical form or architecture unlike other universal function approximator such as ANN to generate the functional variogram models. Furthermore, the GP-derived variogram models were found to give the best functional variogram models that can be used as a viable alternative to the standard variogram models. However, it is important to note that one should be cautious before using the GP-derived variogram model because it may not exhibit positive-definite function to produce a unique and stable solution for the kriging weights, which is an important pre-requisite of kriging. Hence, it is emphasized that in case of the GP-derived variogram models, two mandatory criteria including cross-validation and positive definiteness condition must be satisfied simultaneously to obtain an authorized GP-derived variogram model.
- Making use of auxiliary variables that are highly correlated with rainfall can enhance the spatial estimation of rainfall in a catchment. The ordinary cokriging using elevation information in combination with rainfall was found to be the most suitable interpolator in this study for estimating enhanced spatial distribution of rainfall.

7.3 Recommendations for Future Study

Based on the present study, the following future studies are recommended.

- This study mainly focuses on the improvement of streamflow forecasting by using the optimal rain gauge network-based input in the ANN-based data-driven modelling framework. The streamflow forecasting methodology presented in this study could be extended by incorporating the input from an optimally designed integrated hydrometric network (combined rain gauge and stream gauge network), which is recommended as a potential investigation for future studies.

- Study of potential improvements to the optimal rain gauge network design presented in this study through multi-objective optimization is considered as a useful investigation in future. Since large cost is involved with a rain gauge network (e.g., cost of installation of stations, cost of monitoring, cost associated with redundant station), inclusion of cost reduction objective together with variance reduction objective in the multi-objective optimization scheme of the network design is seen as a way to achieve the cost-efficient optimal rain gauge network.

- The simulation and forecasting of streamflows using a conceptual or physically-distributed model based on the input from an optimal rain gauge network along with other hydrological and environmental variables (e.g., evaporation, temperature, humidity, soil moisture etc.) is considered as another potential investigation in future.

- Assessment of uncertainty in rainfall estimates without having an optimal rain gauge (or hydrometric) network and its impact on streamflow forecasting accuracy is also recommended for future studies.

- Use of gauge-radar rainfall ensembles as input to the streamflow forecasting models in data-driven or conceptual modelling framework is seen as another potential investigation in future studies. For this purpose, merging of rain gauge-radar rainfalls for accurate quantitative precipitation estimation with high spatial resolution can be achieved through different cokriging models presented in this study.

References

Note that the following references are cited within the text, in Chapters 1 to 6 of this thesis. References cited within the journal papers included in different chapters of this thesis can be found at the end of each journal paper.

ABARES. 2012. Land Use and Management Information for Australia - the Australian Collaborative Land Use and Management Program (ACLUMP), Australian Bureau of Agricultural and Resource Economics and Sciences (ABARES). Available online at: <http://www.agriculture.gov.au/abares/aclump/land-use/data-download>. (Accessed on: 17th August 2016).

Abo-Monasar A, Al-Zahrani MA. 2014. Estimation of rainfall distribution for the southwest region of Saudi Arabia. *Hydrological Sciences Journal* **59(2)**: 420-431. DOI: 10.1080/02626667.2013.872788.

Adib A, Moslemzadeh M. 2016. Optimal selection of number of rainfall gauging stations by kriging and genetic algorithm methods. *International Journal of Optimization in Civil Engineering* **6(4)**: 581-594.

Ahmed S. 2004. Geostatistical estimation variance approach to optimize an air temperature monitoring network. *Water, Air, and Soil Pollution* **158(1)**: 387-399. DOI: 10.1023/B:WATE.0000044861.84822.7b.

AIDR. 2015. Flood – Victoria 2011. Australian Institute for Disaster Resilience (AIDR). Available online at:

<https://www.emknowledge.org.au/resource/507/2011/flood---victoria-2011>. (Accessed on: 7th October 2015).

Akhtar MK, Corzo GA, van Andel SJ, Jonoski A. 2009. River flow forecasting with artificial neural networks using satellite observed precipitation pre-processed with flow lengths and travel time information: case study of the Ganges River basin. *Hydrology and Earth System Sciences* **13**: 1607-1618. DOI: 10.5194/hess-13-1607-2009.

Alpaslan N, Harmancioglu NB, Singh VP. 1992. The role of the entropy concept in design and evaluation of water quality monitoring networks. In *Entropy and Energy Dissipation in Water Resources*, Singh V. P. and Fiorentino M. (eds.), Kluwer Academic Publishers. Water Science and Technology Library: Dordecht. pp. 261-282.

Al-Zahrani M, Husain T. 1998. An algorithm for designing a precipitation network in the south-western region of Saudi Arabia. *Journal of Hydrology* **205(3)**: 205-216. DOI: 10.1016/S0022-1694(97)00153-4.

Anctil F, Lauzon N, Andréassian V, Oudin L, Perrin C. 2006. Improvement of rainfall-runoff forecasts through mean areal rainfall optimization. *Journal of Hydrology* **328(3-4)**: 717-725. DOI: 10.1016/j.jhydrol.2006.01.016.

Andréassian V, Perrin C, Michel C, Usart-Sanchez I, Lavabre J. 2001. Impact of imperfect rainfall knowledge on the efficiency and the parameters of watershed models. *Journal of Hydrology* **250(1-4)**: 206-223. DOI: 10.1016/S0022-1694(01)00437-1.

Arnaud P, Bouvier C, Cisneros L, Dominguez R. 2002. Influence of rainfall spatial variability on flood prediction. *Journal of Hydrology* **260(1-4)**: 216-230. DOI: 10.1016/S0022-1694(01)00611-4.

ASCE Task Committee. 1990a. Review of geostatistics in geohydrology, I: Basic concepts. *Journal of Hydraulic Engineering* **116(5)**: 612-632. DOI: 10.1061/(ASCE)0733-9429(1990)116:5(612).

ASCE Task Committee. 1990b. Review of geostatistics in geohydrology, II: Applications. *Journal of Hydraulic Engineering* **116(5)**: 633-658. DOI: 10.1061/(ASCE)0733-9429(1990)116:5(633).

Ashraf M, Loftis JC, Hubbard KG. 1997. Application of geostatistics to evaluate partial weather station networks. *Agricultural and Forest Meteorology* **84(3-4)**: 255-271. DOI: 10.1016/S0168-1923(96)02358-1.

Awadallah AG. 2012. Selecting optimum locations of rainfall stations using kriging and entropy. *International Journal of Civil & Environmental Engineering* **12(1):** 36-41.

Aziz MKBM, Yusof F, Daud ZM, Yusop Z, Kasno MA. 2014. *Redesigning rain gauge network in Johor using geostatistics and simulated annealing*. Proceedings of 2nd International Science Postgraduate Conference (ISPC 2014) Universiti Teknologi Malaysia. pp 1-15.

Aziz MKBM, Yusof F, Daud ZM, Yusop Z, Kasno MA. 2016. Optimal design of rain gauge network in Johor by using geostatistics and particle swarm optimization. *International Journal of GEOMATE* **11(25)**: 2422-2428.

Banzhaf W, Nordin P, Keller RE, Francone FD. 1997. *Genetic Programming: An Introduction* (1st Edition). Morgan Kaufmann Publishers: San Francisco, CA.

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(1)**: 375-386. DOI: 10.1007/s10661-007-0046-z.

Barnes RJ, Johnson TB. 1984. *Positive kriging*, In: Verly G, David M, Journel AG, Marechal A. (Eds.), Geostatistics for natural resources characterization, Part 1. Reidel, Dordrecht, Holland, pp. 231-244.

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(12)**: 1727-1738. DOI: 10.1002/hyp.9311.

Barua S. 2010. *Drought Assessment and Forecasting Using a Nonlinear Aggregated Drought Index*. Ph.D. Thesis, College of Engineering and Science, Victoria University.

Bárdossy A, Das T. 2008. Influence of rainfall observation network on model calibration and application. *Hydrology and Earth System Sciences* **12**: 77-89. DOI: 10.5194/hess-12-77-2008.

Basalirwa CPK, Ogallo LJ, Mutua FM. 1993. The design of regional minimum rain gauge network. *International Journal of Water Resources Development* **9(4)**: 411-424. DOI: 10.1080/07900629308722598.

Bastin G, Lorent B, Duqué C, Gevers M. 1984. Optimal estimation of the average areal rainfall and optimal selection of rain gauge locations. *Water Resources Research* **20(4)**: 463-470. DOI: 10.1029/WR020i004p00463.

Beven K, Hornberger GM. 1982. Assessing the effect of spatial pattern of precipitation in modelling streamflow hydrographs. *Journal of the American Water Resources Association* **18(5)**: 823-829. DOI: 10.1111/j.1752-1688.1982.tb00078.x.

Birikundavyi S, Labib R, Trung HT, Rousselle J. 2002. Performance of neural networks in daily streamflow forecasting. *Journal of Hydrologic Engineering* **7(5)**: 392-398. DOI: 10.1061/(ASCE)1084-0699(2002)7:5(392).

BITRE, 2008. Bureau of Infrastructure, Transport and Regional Economics (BITRE) analysis of the Emergency Management Australia database, Table 30, pp. 44, published in BITRE, 2008. About Australia's Regions, Department of Infrastructure, Transport, Regional Development and Local Government, Australian Government, Canberra.

Bonaccorso B, Cancelliere A, Rossi G. 2003. Network design for drought monitoring by geostatistical techniques. *European Water* **3(4)**: 9-15.

Bradley AA, Peters-Lidard C, Nelson BR, Smith JA, Young CB. 2002. Rain gauge network design using NEXRAD precipitation estimates. *Journal of the American Water Resources Association* **38(5)**: 1393-1407. DOI: 10.1111/j.1752-1688.2002.tb04354.x.

Bras RL, Rodriguez-Iturbe I. 1976. Network design for the estimation of areal mean of rainfall events. *Water Resources Research* **12(6)**: 1185-1195. DOI: 10.1029/WR012i006p01185.

Bureau of Meteorology. 2016. Australian climate influences. Available online at: http://www.bom.gov.au/climate/about/ (Accessed on 24th December 2016).

Bureau of Meteorology. 2011. Improving Water Information Program Progress Report: Advances in water information made by the Bureau of Meteorology in 2011, Australia.

Cameron K, Hunter P. 2002. Using spatial models and kriging techniques to optimize long-term groundwater monitoring networks: a case study. *Environmetrics* **13(5-6)**: 629-656. DOI: 10.1002/env.582.

Canadian Water Resources, A., Mitchell B, Shrubsole D. 1994. *Canadian Water Management: Visions for Sustainability*. Canadian Water Resources Association (Association Canadienne Des Ressources Hydriques).

Carty R. Pierotti S. 2010. *Yarra River Application Project - Source Catchments Hydrology Calibration Report*. eWater Cooperative Research Centre, Canberra, Australia.

Chang CH, Wu SJ, Hsu CT, Shen JC, Lien HC. 2017. An evaluation framework for identifying the optimal raingauge network based on spatiotemporal variation in quantitative precipitation estimation. *Hydrology Research* **48(1)**: 77-98. DOI: 10.2166/nh.2016.169.

Chang CL, Lo SL, Yu SL. 2005. Applying fuzzy theory and genetic algorithm to interpolate precipitation. *Journal of Hydrology* **314(1–4)**: 92–104. DOI: 10.1016/j.jhydrol.2005.03.034.

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(8)**: 665-676. DOI: 10.1061/(ASCE)HE.1943-5584.0000353.

Chen FW, Liu CW. 2012. Estimation of the spatial rainfall distribution using inverse distance weighting (IDW) in the middle of Taiwan. *Paddy and Water Environment* **10(3)**: 209-222. DOI: 10.1007/s10333-012-0319-1.

Chen N, Wang K, Xiao C, Gong J. 2014. A heterogeneous sensor web node meta-model for the management of a flood monitoring system. *Environmental Modelling & Software* **54**: 222-237. DOI: 10.1016/j.envsoft.2014.01.014.

Chen T, Ren L, Yuan F, Yang X, Jiang S, Tang T, Liu Y, Zhao C, Zhang L. 2017. Comparison of spatial interpolation schemes for rainfall data and application in hydrological modeling. *Water* **9(5)**: 342. DOI: 10.3390/w9050342.

Chen YC, Wei C, Yeh HC. 2008. Rainfall network design using kriging and entropy. *Hydrological Processes* **22(3)**: 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(14)**: 2554-2564. DOI: 10.1002/hyp.6851.

Chiew FH, McMahon TA. 2003. El Niño/Southern Oscillation and Australian rainfall and streamflow. *Australasian Journal of Water Resources* **6(2)**: 115-129. DOI: 10.1080/13241583.2003.11465216.

Chiew FHS, Potter NJ, Vaze J, Petherham C, Zhang L, Teng J, Post D. 2014. Observed hydrologic non-stationarity in far south-eastern Australia: Implications for modelling and prediction. *Stochastic Environmental Research and Risk Assessment* **28(1)**: 3-15. DOI: 10.1007/s00477-013-0755-5.

Chowdhury RK, Beecham S. 2010. Australian rainfall trends and their relation to the southern oscillation index. *Hydrological Processes* **24(4)**: 504-514. DOI: 10.1002/hyp.7504.

Cole SJ, Moore RJ. 2008. Hydrological modelling using raingauge and radar-based estimators of areal rainfall. *Journal of Hydrology* **358(3-4)**: 159-181. DOI: 10.1016/j.jhydrol.2008.05.025.

Dai Q, Bray M, Zhuo L, Islam T, Han D. 2017. A scheme for rain gauge network design based on remotely sensed rainfall measurements. *Journal of Hydrometeorology* **18(2)**: 363-379. DOI: 10.1175/JHM-D-16-0136.1.

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.

Davis JC. 1973. *Statistics and Data Analysis in Geology*. John Wiley & Sons: New York.

Delbari M, Afrasiab P, Jahani S. 2013. Spatial interpolation of monthly and annual rainfall in northeast of Iran. *Meteorology and Atmospheric Physics* **122(1)**: 103-113. DOI: 10.1007/s00703-013-0273-5.

Delhomme JP. 1978. Kriging in the hydrosciences. *Advances in Water Resources* **1(5)**: 251-266. DOI: 10.1016/0309-1708(78)90039-8.

Dettinger MD, Cayan DR, Meyer MK, Jeton AE. 2004. Simulated hydrologic responses to climate variations and change in the Merced, Carson, and American River basins, Sierra Nevada, California, 1900–2099. *Climate Change* **62(1-3)**: 283-317, DOI:10.1023/B:CLIM.0000013683.13346.4f.

Deutsch CV. 1996. Correcting for negative weights in ordinary kriging. *Computers & Geosciences* **22(7)**: 765-773. DOI: 10.1016/0098-3004(96)00005-2.

Dibike YB, Solomatine DP. 2001. River flow forecasting using artificial neural networks. *Physics and Chemistry of the Earth, Part B: Hydrology, Oceans and Atmosphere* **26(1)**: 1-7. DOI: 10.1016/S1464-1909(01)85005-X.

Dirks KN, Hay JE, Stow CD, Harris D. 1998. High resolution studies of rainfall on Norfolk Island part II: interpolation of rainfall data. *Journal of Hydrology* **208(3-4)**: 187-193. DOI: 10.1016/S0022-1694(98)00155-3.

Di Piazza A, Conti FL, Noto LV, Viola F, Loggia GL. 2011. Comparative analysis of different techniques for spatial interpolation of rainfall data to create a serially complete monthly time series of precipitation for Sicily, Italy. *International Journal of Applied Earth Observation and Geoinformation* **13(3)**: 396-408. DOI: 10.1016/j.jag.2011.01.005.

Dong X, Janssen CMD, Booij M. 2005. Appropriate spatial sampling of rainfall or flow simulation. *Hydrological Sciences Journal* **50(2)**: 279-298. DOI: 10.1623/hysj.50.2.279.61801.

Dutta SC, Ritchie JW, Freebairn DM, Abawi GY. 2006. Rainfall and streamflow response to El Niño Southern Oscillation: A case study in a semiarid catchment, Australia. *Hydrological Sciences Journal* **51(6)**: 1006-1020. DOI: 10.1623/hysj.51.6.1006.

Ekström M, Jones PD. 2009. Impact of rainfall estimation uncertainty on streamflow estimations for catchments Wye and Tyne in the United Kingdom. *International Journal of Climatology* **29**: 79-86. DOI: 10.1002/joc.1665.

EPA Victoria. 1999. *Protecting the Environmental Health of Yarra Catchment Waterways Policy Impact Assessment*. Environmental Protection Agency (EPA) of Victoria Report No. 654, Melbourne, Australia.

Faurès J-M, Goodrich DC, Woolhiser DA, Sorooshian S. 1995. Impact of small-scale spatial variability on runoff modeling. *Journal of Hydrology* **173(1-4)**: 309-326. DOI: 10.1016/0022-1694(95)02704-S.

Feki H, Slimani M, Cudennec C. 2012. Incorporating elevation in rainfall interpolation in Tunisia using geostatistical methods. *Hydrological Sciences Journal* **57(7)**: 1294-1314. DOI: 10.1080/02626667.2012.710334.

Feki H, Slimani M, Cudennec C. 2017. Geostatistically based optimization of a rainfall monitoring network extension: case of the climatically heterogeneous Tunisia. *Hydrology Research* **48(2)**: 514-541; DOI: 10.2166/nh.2016.256.

Gardner B. 1994. *Agriculture in the Yarra Catchment*. Discussion Paper No. 5, Melbourne, Australia.

Georgakakos KP, Bae DH, Cayan DR. 1995. Hydroclimatology of continental watersheds: 1. Temporal analyses. *Water Resources Research* **31(3)**: 655-675. DOI: 10.1029/94WR02375.

Ghahraman B, Sepaskhah AR. 2001. Autographic rain-gage network design for Iran by kriging. *Iranian Journal of Science and Technology Transaction B-Engineering* **25(B4)**: 653-660.

Golding BW. 2009. Uncertainty propagation in a London flood simulation. *Journal of Flood Risk Management* **2(1)**: 2-15. DOI: 10.1111/j.1753-318X.2008.01014.x.

Goovaerts P. 1997. *Geostatistics for Natural Resources Evaluation*. Oxford University Press: New York.

Goovaerts P. 2000. Geostatistical approaches for incorporating elevation into the spatial interpolation of rainfall. *Journal of Hydrology* **228(1-2)**: 113-129. DOI: 10.1016/S0022-1694(00)00144-X.

Grant SB, Fletcher TD, Feldman D, Saphores JD, Cook PLM, Stewardson M, Low K, Burry K, Hamilton AJ. 2013. Adapting urban water systems to a changing climate: Lessons from the Millennium Drought in Southeast Australia. *Environmental Science & Technology* **47**: 10727-10734. DOI: 10.1021/es400618z.

Haggag M, Ali A, Awadallah A. 2016. Evaluation of rain gauge network in arid regions using geostatistical approach: case study in northern Oman. *Arabian Journal of Geosciences* **9(9)**: 552. DOI: 10.1007/s12517-016-2576-6.

Hamlin MJ. 1983. The significance of rainfall in the study of hydrological processes at basin scale. *Journal of Hydrology* **65(1-3)**: 73-94. DOI: 10.1016/0022-1694(83)90211-1.

Hancock PA, Hutchinson MF. 2006. Spatial interpolation of large climate data sets using bivariate thin plate smoothing splines. *Environmental Modelling & Software* **21(12)**: 1684-1694. DOI: 10.1016/j.envsoft.2005.08.005.

Harmancioglu NB, Yevjevich V. 1987. Transfer of hydrologic information among rivers points. *Journal of Hydrology* **91(1-2)**: 103-118. DOI: 10.1016/0022-1694(87)90131-4.

Harmancioglu NB, Alpaslan N. 1992. Water quality monitoring network design: A problem of multi-objective decision making. *Water Resources Bulletin* **28(1)**: 179-192. DOI: 10.1111/j.1752-1688.1992.tb03163.x.

Harmancioglu NB, Alpaslan N, Singh VP. 1994. Assessment of the entropy principle as applied to water quality monitoring network design. In: *Stochastic and Statistical Methods in Hydrology and Environmental Engineering*, Hipel, K.W. et al.(eds), Kluwer Academic Publishers: Netherlands. pp.135-148.

Harmancioglu NB, Singh VP. 1998. Entropy in Environmental and Water Resources. In: *Encyclopaedia of Hydrology and Water Resources*, Herscly R. W. and Fairbridge R. W. (eds.), Kluwer Publishers: Dordrecht.

Harmanciouglu NB, Fistikoglu O, Ozkul SD, Singh VP, Alpaslan MN. 1999. *Water Quality Monitoring Network Design*. Water Science and Technology Library, Vol. 33, Kluwer Academic Publishers: The Netherlands.

Hevesi JA, Istok JD, Flint AL. 1992. Precipitation estimation in mountainous terrain using multivariate geostatistics (part 1 and 2). *Journal of Applied Meteorology* **31(7)**: 661-688. DOI: 10.1175/1520-0450(1992)031<0661:PEIMTU>2.0.CO;2.

Hsieh HH, Cheng SJ, Liou JY, Chou SC, Siao BR. 2006. Characterization of spatially distributed summer daily rainfall. *Journal of Chinese Agricultural Engineering* **52**: 47-55.

Huang W, Xu B, Chan-Hilton A. 2004. Forecasting flows in Apalachicola River using neural networks. *Hydrological Processes* **18**: 2545-2564. DOI: 10.1002/hyp. 1492.

Huang Z, Wang H, Zhang R. 2012. An improved kriging interpolation technique based on SVM and its recovery experiment in oceanic missing data. *American Journal of Computational Mathematics* **2**: 56-60. DOI: 10.4236/ajcm.2012.21007.

Husain T. 1989. Hydrologic uncertainty measure and network design. *JAWRA Journal of the American Water Resources Association* **25(3)**: 527-534. DOI: 10.1111/j.1752-1688.1989.tb03088.x.

Hutchinson MF. 1995. Interpolating mean rainfall using thin plate smoothing splines. *International Journal of Geographical Information Systems* **9(4)**: 385-403. DOI: 10.1080/02693799508902045.

Intergovernmental Panel on Climate Change. 2001. *Special Report on the Regional Impacts of Climate Change: An Assessment of Vulnerability* (Edited by RT Watson et al.). Cambridge University Press: Cambridge, United Kingdom.

Isaaks HE, Srivastava RM. 1989. *An Introduction to Applied Geostatistics*. Oxford University Press: Ann Arbor.

Isbell RF. 2002. *The Australian Soil Classification*. Australian Soil and Land Survey Handbook Series 4, Revised Edition, CSIRO Publishing, Australia.

IUCN. 1980. *World Conservation Strategy: Living Resource Conservation for Sustainable Development*. International Union for Conservation of Nature (IUCN), Gland, Switzerland.

Jain A, Kumar AM. 2007. Hybrid neural network models for hydrologic time series forecasting. *Applied Soft Computing* **7(2)**: 585-592. DOI: 10.1016/j.asoc.2006.03.002.

Jaynes ET. 1957a. Information theory and statistical mechanics I. *Physics Reviews* **106(4)**: 620-630. DOI: 10.1103/PhysRev.106.620.

Jaynes ET. 1957b. Information theory and statistical mechanics II. *Physics Reviews* **108(2)**: 171-190. DOI: 10.1103/PhysRev.108.171.

Jeffrey SJ, Carter JO, Moodie KB, Beswick AR. 2001. Using spatial interpolation to construct a comprehensive archive of Australian climate data. *Environmental Modelling & Software* **16(4)**: 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.

Jones DA, Wang W, Fawcett R. 2009. High-quality spatial climate data-sets for Australia. *Australian Meteorological and Oceanographic Journal* **58(4)**: 233-248.

Jones PD, Lister DH, Wilby RL, Kostopoulou E. 2006. Extended river flow reconstructions for England and Wales, 1865-2002. *International Journal of Climatology* **25**: 219-231. DOI: 10.1002/joc.1252.

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

Júnez-Ferreira H, González J, Reyes E, Herrera GS. 2016. A geostatistical methodology to evaluate the performance of groundwater quality monitoring networks using a vulnerability index. *Mathematical Geosciences* **48(1)**: 25-44. DOI: 10.1007/s11004-015-9613-y.

Kar AK, Lohani AK, Goel NK, Roy GP. 2015. Rain gauge network design for flood forecasting using multi-criteria decision analysis and clustering techniques in lower Mahanadi river basin, India. *Journal of Hydrology: Regional Studies* **4**: 313-332. DOI: 10.1016/j.ejrh.2015.07.003.

Karimi-Hosseini A, Haddad OB, Mariño MA. 2011. Site selection of raingauges using entropy methodologies. *Proceedings of the ICE-Water Management* **164(7)**: 321-333. DOI: 10.1680/wama.2011.164.7.321.

Kassim AHM, Kottegoda N. 1991. Rainfall network design through comparative kriging methods. *Hydrological Sciences Journal* **36(3)**: 223-240. DOI: 10.1080/02626669109492505.

Keating J. 1992. *The Drought Walked Through: A History of Water Shortage in Victoria*. Department of Water Resources Victoria, Melbourne, Australia.

Kholghi M, Hosseini SM. 2009. Comparison of groundwater level estimation using neuro-fuzzy and ordinary kriging. *Environmental Modeling & Assessment* **14**: 729-737. DOI: 10.1007/s10666-008-9174-2.

Kişi Ö. 2007. Streamflow forecasting using different artificial neural network algorithms. *Journal of Hydrologic Engineering* **12(5)**: 532-539. DOI: 10.1061/(ASCE)1084-0699(2007)12:5(532).

Knowles N, Cayan DR. 2004. Elevational dependence of projected hydrologic changes in the San Francisco estuary and watershed. *Climate Change* **62(1-3)**: 319 – 336. DOI:10.1023/B:CLIM.0000013696.14308.b9.

Koza JR. 1992. *Genetic Programming: On the Programming of Computers by Means of Natural Selection*. MIT Press: Cambridge, MA.

Krajewski WF, Lakshmi V, Georgakakos P, Jain SC. 1991. A Monte Carlo study of rainfall sampling effect on a distributed catchment model. *Water Resources Research* **27(11)**: 119-128. DOI: 10.1029/90WR01977.

Krajewski WF, Ciach GJ, Habib E. 2003. An analysis of small-scale rainfall variability in different climatic regimes. *Hydrological Sciences Journal* **48(2)**: 151-162. DOI: 10.1623/hysj.48.2.151.44694.

Krstanovic P, Singh VP. 1992a. Evaluation of rainfall networks using entropy: I. Theoretical development. *Water Resources Management* **6(4)**: 279-293. DOI: 10.1007/BF00872281.

Krstanovic P, Singh VP. 1992b. Evaluation of rainfall networks using entropy: II. Application. *Water Resources Management* **6(4)**: 295-314. DOI: 10.1007/BF00872282.

Kumar DN, Raju KS, Sathish T. 2004. River flow forecasting using recurrent neural networks. *Water Resources Management* **18(2)**: 143-161. DOI: 10.1023/B:WARM.0000024727.94701.12.

Kumar S, Sondhi SK, Phogat V. 2005. Network design for groundwater level monitoring in Upper Bari Doab Canal Tract, Punjab, India. *Irrigation and Drainage* **54(4)**: 431-442. DOI: 10.1002/ird.194.

Langbein WB. 1979. Overview of conference on hydrologic data networks. *Water Resources Research* **15(6)**: 1867 –1871, DOI:10.1029/WR015i006p01867.

Li C, Singh VP, Mishra AK. 2012. Entropy theory-based criterion for hydrometric network evaluation and design: maximum information minimum redundancy. *Water Resources Research* **48(5)**: W05521. DOI: 10.1029/2011WR011251.

Li J, Heap AD. 2011. A review of comparative studies of spatial interpolation methods in environmental sciences: performance and impact factors. *Ecological Informatics* **6(3-4)**: 228-241. DOI: 10.1016/j.ecoinf.2010.12.003.

Li M, Shao Q. 2010. An improved statistical approach to merge satellite rainfall estimates and rain-gauge data. *Journal of Hydrology* **385(1-2)**: 51-64. DOI: 10.1016/j.jhydrol.2010.01.023.

Linares-Rodriguez A, Lara-Fanego V, Pozo-Vazquez D, Tovar-Pescador J. 2015. One-day-ahead streamflow forecasting using artificial neural networks and a meteorological mesoscale model. *Journal of Hydrologic Engineering* **20(9)**: 05015001. DOI: 10.1061/(ASCE)HE.1943-5584.0001163

Lloyd CD. 2005. Assessing the effect of integrating elevation data into the estimation of monthly precipitation in Great Britain. *Journal of Hydrology* **308(1-4)**: 128–150. DOI: 10.1016/j.jhydrol.2004.10.026.

Londhe S, Charhate S. 2010. Comparison of data-driven modelling techniques for river flow forecasting. *Hydrological Sciences Journal* **55(7)**: 1163-1174. DOI: 10.1080/02626667.2010.512867.

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(4)**: 497-513. DOI: 10.1080/07900629408722649.

Loucks DP, van Beek E, Stedinger JR, Dijkman JPM, Villars MT. 2005. *Water Resources Systems Planning and Management: An Introduction to Methods, Models and Applications*. UNESCO: Paris.

Low KG, Grant SB, Hamilton AJ, Gan K, Saphores JD, Arora M, Feldman DL. 2015. Fighting drought with innovation: Melbourne's response to the Millennium Drought in Southeast Australia. *Wiley Interdisciplinary Reviews: Water* **2(4)**: 315-328. DOI: 10.1002/wat2.1087.

Lu L, Hedley D. 2004. The impact of the 2002–03 drought on the economy and agricultural employment. Economic Roundup, pp. 25–43. Available online at: https://archive.treasury.gov.au/documents/817/HTML/docshell.asp?URL=03_article_2.asp (Accessed on 18th January 2017).

Mahmoudi-Meimand H, Nazif S, Abbaspour RA, Sabokbar HF. 2015. An algorithm for optimisation of a rain gauge network based on geostatistics and entropy concepts using GIS. *Journal of Spatial Science* **61(1)**: DOI: 10.1080/14498596.2015.1030789.

Maier HR, Dandy GC. 2000. Neural networks for the prediction and forecasting of water resources variables: a review of modelling issues and applications. *Environmental Modelling & Software* **15(1)**: 101-124. DOI: 10.1016/S1364-8152(99)00007-9.

Mair A, Fares A. 2011. Comparison of rainfall interpolation methods in a mountainous region of a tropical island. *Journal of Hydrologic Engineering* **16(4)**: 371-383. DOI: 10.1061/(ASCE)HE.1943-5584.0000330.

Makkeasorn A, Chang NB, Zhou X. 2008. Short-term streamflow forecasting with global climate change implications – A comparative study between genetic programming and neural network models. *Journal of Hydrology* **352(3-4)**: 336-354. DOI: 10.1016/j.jhydrol.2008.01.023.

Maskey S, Guinot V, Price RK. 2004. Treatment of precipitation in rainfall-runoff modelling: a fuzzy set approach. *Advances in Water Resources* **27(9)**: 889-898. DOI: 10.1016/j.advwatres.2004.07.001.

McMillan H, Jackson B, Clark M, Kavetski D, Woods R. 2011. Rainfall uncertainty in hydrological modelling: An evaluation of multiplicative error models. *Journal of Hydrology* **400(1-2)**: 83-94. DOI: 10.1016/j.jhydrol.2011.01.026.

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. 2015. Yarra Catchment, Melbourne Water Corporation. Available online at: <http://www.melbournewater.com.au/>. (Accessed on 20th November 2015).

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

Mishra AK, Coulibaly P. 2009. Developments in hydrometric network design: A review. *Reviews of Geophysics* **47(2)**: RG2001. DOI: 10.1029/2007RG000243.

Mishra AK. 2013. Effect of rain gauge density over the accuracy of rainfall: a case study over Bangalore, India. *SpringerPlus* **2**: 311. DOI: 10.1186/2193-1801-2-311.

Mogheir Y, Singh VP. 2002. Application of information theory to groundwater quality monitoring networks. *Water Resources Management* **16(1)**: 37-49. DOI: 10.1023/A:1015511811686.

Mogheir Y, de Lima JLMP, Singh VP. 2009. Entropy and multi-objective based approach for groundwater quality monitoring network assessment and redesign. *Water Resources Management* **23(8)**: 1603-1620. DOI: 10.1007/s11269-008-9343-8.

Moore RJ, Jones DA, Cox DR, Isham VS. 2000. Design of the HYREX raingauge network. *Hydrology and Earth System Sciences* **4(4)**: 521-530. DOI: 10.5194/hess-4-521-2000.

Moral FJ. 2010. Comparison of different geostatistical approaches to map climate variables: application to precipitation. *International Journal of Climatology* **30(4)**: 620-631. DOI: 10.1002/joc.1913.

Moulin L, Gaume E, Obled C. 2009. Uncertainties on mean areal precipitation: assessment and impact on streamflow simulations. *Hydrology and Earth System Sciences* **13**: 99-114. DOI: 10.5194/hess-13-99-2009.

Murphy BF, Ribbe J. 2004. Variability of southeastern Queensland rainfall and climate indices. *International Journal of Climatology* **24(6)**: 703-721. DOI: 10.1002/joc.1018.

Murphy BF, Timbal B. 2008. A review of recent climate variability and climate change in southeastern Australia. *International Journal of Climatology* **28(7)**: 859-879. DOI: 10.1002/joc.1627.

Muttil N, Barua S, Ng AWM, Perera BJC. 2009. Spatio-temporal analysis to detect climate changes within the Yarra River catchment. *Proceedings of Australia's National Water Conference and Exhibition (Ozwater 09)*, Melbourne, Australia, 16-18 March 2009.

Narany TS, Ramli MF, Fakharian K, Aris AZ, Sulaiman WNA. 2015. Multi-objective based approach for groundwater quality monitoring network optimization. *Water Resources Management* **29(14)**: 5141-5156. DOI: 10.1007/s11269-015-1109-5.

National Climate Centre. 2011. Frequent Heavy Rain Events in Late 2010/Early 2011 Lead to Widespread Flooding across Eastern Australia. National Climate Centre, Bureau of Meteorology, Australia. *Interaction* **39(1)**: 29-34. Available online at: <http://search.informit.com.au/documentSummary;dn=845118025535278;res=IELHSS>. (Accessed on: 5th December 2016).

Nour MH, Smit DW, El-Din MG. 2006. Geostatistical mapping of precipitation: implications for rain gauge network design. *Water Science & Technology* **53(10)**: 101-110. DOI: 10.2166/wst.2006.303.

Oliver MA, Webster R. 2014. A tutorial guide to geostatistics: computing and modelling variograms and kriging. *Catena* **113**: 56-69. DOI: 10.1016/j.catena.2013.09.006.

Ozkul S, Harmancioglu NB, Singh VP. 2000. Entropy-based assessment of water quality monitoring networks. *Journal of Hydrologic Engineering* **5(1)**: 90-100. DOI: 10.1061/(ASCE)1084-0699(2000)5:1(90).

Papamichail DM, Metaxa IG. 1996. Geostatistical analysis of spatial variability of rainfall and optimal design of a rain gauge network. *Water Resources Management* **10(2)**: 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(1)**: 206-220. DOI: 10.1016/S0022-1694(98)00188-7.

Phillips DL, Dolph J, Marks D. 1992. A comparison of geostatistical procedures for spatial analysis of precipitation in mountainous terrain. *Agricultural and Forest Meteorology* **58(1-2)**: 119-141. DOI: 10.1016/0168-1923(92)90114-J.

Pittock J, Connell D. 2010. Australia demonstrates the planet's future: water and climate in the Murray-Darling basin. *Water Resources Development* **26(4)**: 561-578. DOI: 10.1080/07900627.2010.519522.

Pittock B, Abbs D, Suppiah R, Jones R. 2006. Climatic background to past and future floods in Australia. *Advances in Ecological Research* **39**: 13-39. DOI: 10.1016/S0065-2504(06)39002-2.

Poff NL, Allan JD, Bain MB, Karr JR, Prestegaard KL, Richter BD, Sparks RE, Stromberg JC. 1997. The natural flow regime: A paradigm for river conservation and restoration. *BioScience* **47(11)**: 769-783. DOI: 10.2307/1313099.

Ponce V, Lohani A, Huston P. 1997. Surface Albedo and water resources: hydroclimatological impact of human activities. *Journal of Hydrological Engineering* **2(4)**: 197–203. DOI: 10.1061/(ASCE)1084-0699(1997)2:4(197).

Putthividhya A, Tanaka K. 2013. Optimal rain gauge network design and spatial precipitation mapping based on geostatistical analysis from colocated elevation and humidity data. *Chiang Mai Journal of Science* **40(2)**: 187-197.

Retnam MTP, Williams BJ. 1988. Input errors in rainfall-runoff modelling. *Mathematics and Computers in Simulation* **30(1-2)**: 119-131. DOI: 10.1016/0378-4754(88)90114-0.

Ridolfi E, Montesarchio V, Russo F, Napolitano F, Bartzokas A. 2011. An entropy approach for evaluating the maximum information content achievable by an urban rainfall network. *Natural Hazards & Earth System Sciences* **11(7)**: 2075-2083. DOI: 10.5194/nhess-11-2075-2011.

Robertson GP. 2008. *GS+: Geostatistics for the Environmental Sciences*. Gamma Design Software: Plainwell, Michigan, USA.

Ruiz JE, Cordery I, Sharma A. 2007. Forecasting streamflows in Australia using the tropical Indo-Pacific thermocline as predictor. *Journal of Hydrology* **341(3-4)**: 156-164. DOI: 10.1016/j.jhydrol.2007.04.021.

Schilling W, Fuchs L. 1986. Errors in stormwater modelling – A quantitative assessment. *Journal of Hydraulic Engineering* **112(2)**: 111-123. DOI: 10.1061/(ASCE)0733-9429(1986)112:2(111).

Seed AW, Austin GL. 1990. Sampling errors for rain gauge-derived mean areal daily and monthly rainfall. *Journal of Hydrology* **118(1-4)**: 163-173. DOI: 10.1016/0022-1694(90)90256-W.

Shafiei M, Ghahraman B, Saghafian B, Pande S, Gharari S, Davary K. 2014. Assessment of rain gauge networks using a probabilistic GIS based approach. *Hydrology Research* **45(4-5)**: 551-562. DOI: 10.2166/nh.2013.042.

Shaghaghian M, Abedini M. 2013. Rain gauge network design using coupled geostatistical and multivariate techniques. *Scientia Iranica* **20(2)**: 259-269. DOI: 10.1016/j.scient.2012.11.014.

Shamseldin AY. 2004. *Hybrid neural network modelling solutions*. In RJ Abrahart, PE Kneale, LM See (Eds.) Neural Networks for Hydrological Modeling. A.A. Balkema Publishers: Leiden. pp. 61-79.

Shamsi U, Quimpo R, Yoganarasimhan G. 1988. An application of kriging to rainfall network design. *Nordic Hydrology* **19(3)**: 137-152. DOI: 10.2166/nh.1988.010.

Shannon CE. 1948a. A mathematical theory of communications. *Bell System Technical Journal* **27(3)**: 379-423. DOI: 10.1002/j.1538-7305.1948.tb01338.x.

Shannon CE. 1948b. A mathematical theory of communication. III and IV. *Bell System Technical Journal* **27(4)**: 623-656. DOI: 10.1002/j.1538-7305.1948.tb00917.x.

Sharma RH, Shakya NM. 2006. Hydrological changes and its impact on water resources of Bagmati watershed, Nepal. *Journal of Hydrology* **327(3-4)**: 315-322. DOI: 10.1016/j.jhydrol.2005.11.051.

Singh VP. 1997. The use of entropy in hydrology and water resources. *Hydrological Processes* **11(6)**: 587-626. DOI: 10.1002/(SICI)1099-1085(199705)11:6<587: :AID-HYP479>3.0.CO;2-P.

Singh VP. 2013. *Entropy Theory and Its Application in Environmental and Water Engineering*. Wiley-Blackwell, John Wiley & Sons: Chichester, United Kingdom.

Sitharam TG, Samui P, Anbazhagan P. 2008. Spatial variability of rock depth in Bangalore using geostatistical, neural network and support vector machine models. *Geotechnical and Geological Engineering* **26(5)**: 503-517. DOI: 10.1007/s10706-008-9185-4.

Sivapragasam C, Vanitha S, Muttil N, Suganya K, Suji S, Selvi MT, Sudha SJ. 2014. Monthly flow forecast for Mississippi River basin using artificial neural networks. *Neural Computing and Applications* **24(7)**: 1785-1793. DOI: 10.1007/s00521-013-1419-6.

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.

Srinivasulu S, Jain A. 2009. River flow prediction using an integrated approach. *Journal of Hydrologic Engineering* **14(1)**: 75-83. DOI: 10.1061/(ASCE)1084-0699(2009)14:1(75).

Stewart IT, Cayan DR, Dettinger MD. 2005. Changes toward earlier streamflow timing across Western North America. *Journal of Climate* **18(8)**: 1136-1155. DOI: 10. 1175/JCLI3321.1.

St-Hilaire A, Ouarda TBMJ, Lachance M, Bobée B, Gaudet J, Gignac C. 2003. Assessment of the impact of meteorological network density on the estimation of basin precipitation and runoff: a case study. *Hydrological Processes* **17**: 3561-3580. DOI: 10.1002/hyp.1350.

Subyani AM, Al-Dakheel AM. 2009. Multivariate geostatistical methods of mean annual and seasonal rainfall in southwest Saudi Arabia. *Arabian Journal of Geosciences* **2(1)**: 19-27. DOI: 10.1007/s12517-008-0015-z.

Sun X, Manton MJ, Ebert EE. 2003. *Regional rainfall estimation using double-kriging of rain gauge and satellite observations*. BMRC Research Report No. 94. Bureau of Meteorology Research Centre (BMRC), Australia.

Szidarovszky F, Baafi EY, Kim YC. 1987. Kriging without negative weights. *Mathematical Geology* **19(6)**: 549-559. DOI: 10.1007/BF00896920.

Talei A, Chua LHC, Wong TSW. 2010. Evaluation of rainfall and discharge inputs used by Adaptive Network-based Fuzzy Inference Systems (ANFIS) in rainfall-runoff modeling. *Journal of Hydrology* **391(3-4)**: 248-262. DOI: 10.1016/j.jhydrol.2010.07.023.

Tan, K.S. and Rhodes, B.G., 2008. Implications of the 1997-2006 drought on water resources planning for Melbourne. *Proceedings of 31st Hydrology and Water Resources Symposium (HWRS 08)*, Adelaide, Australia, pp. 2016-2027.

Taormina R, Chau KW, Sivakumar B. 2015. Neural network river forecasting through baseflow separation and binary-coded swarm optimization. *Journal of Hydrology* **529**: 1788-1797. DOI: 10.1016/j.jhydrol.2015.08.008.

Teegavarapu RSV. 2007. Use of universal function approximation in variance-dependent surface interpolation method: An application in hydrology. *Journal of Hydrology* **332(1-2)**: 16-29. DOI: 10.1016/j.jhydrol.2006.06.017.

Teegavarapu RSV, Tufail M, Ormsbee L. 2009. Optimal functional forms for estimation of missing precipitation records. *Journal of Hydrology* **374(1-2)**: 106-115. DOI: 10.1016/j.jhydrol.2009.06.014.

Triki I, Zairi M, Dhia HB. 2013. A geostatistical approach for groundwater head monitoring network optimization: case of the Sfax superficial aquifer (Tunisia). *Water and Environment Journal* **27(3)**: 362-372. DOI: 10.1111/j.1747-6593.2012.00352.x.

Troutman BM. 1983. Runoff prediction errors and bias in parameter estimation induced by spatial variability of precipitation. *Water Resources Research* **19(3)**: 791-810. DOI: 10.1029/WR019i003p00791.

Tsai MJ, Abrahart RJ, Mount NJ, Chang FJ. 2014. Including spatial distribution in a data-driven rainfall-runoff model to improve reservoir inflow forecasting in Taiwan. *Hydrological Processes* **28(3)**: 1055-1070. DOI: 10.1002/hyp.9559.

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

van den Honert RC, McAneney J. 2011. The 2011 Brisbane Floods: Causes, Impacts and Implications. *Water* **3**: 1149-1173. DOI: 10.3390/w3041149.

van der Made JW. 1988. *Analysis of Some Criteria for Design and Operation of Surface Water Gauging Networks*. Rijkswaterstaat: The Hague.

van Dijk AIJM, Beck HE, Crosbie RS, de Jeu RAM, Liu YY, Podger GM, Timbal B, Viney NR. 2013. The Millennium Drought in southeast Australia (2001-2009): natural and human causes and implications for water resources, ecosystems, economy, and society. *Water Resources Research* **49(2)**: 1040-1057. DOI: 10.1002/wrcr.20123.

Victorian Government. 2011. The Review of the 2010-11 Flood Warnings & Response – Victorian Floods Review, Final Report. State Government of Victoria.

Victorian Government Department of Sustainability and Environment. 2006. *Strengthening the Management of the Yarra and Maribyrnong Rivers: A Background Report for Future Water Quality Management*. Department of Sustainability and Environment (DSE), Melbourne, State Government of Victoria.

Victorian Government Department of Sustainability and Environment. 2008. *Climate Change in Victoria Summary Report*. Department of Sustainability and Environment (DSE), Melbourne, State Government of Victoria.

Vieux BE. 2001. *Distributed Hydrologic Modelling using GIS*. Water Science and Technology Library, Kluwer Academic Publishers: Dordrecht, Netherlands.

Vivekanandan N, Jagtap R. 2012. Evaluation and selection of rain gauge network using entropy. *Journal of the Institution of Engineers (India): Series A* **93(4)**: 223-232. DOI 10.1007/s40030-013-0032-0.

Volkmann THM, Lyon SW, Gupta HV, Troch PA. 2010. Multicriteria design of rain gauge networks for flash flood prediction in semiarid catchments with complex terrain. *Water Resources Research* **46(11)**: W11554. DOI: 10.1029/2010WR 009145.

Wackernagel H. 2003. *Multivariate Geostatistics: An Introduction with Applications* (3rd Edition). Springer-Verlag: Berlin, Germany.

Wallis P, Birrell R, Griggs D, Healy E, Langford J, Stanley J. 2009. *Melbourne's Water Situation: The Opportunity for Diverse Solutions*. Monash Sustainability Institute Report 09/2, Melbourne, Victoria.

Wang W. 2006. *Stochasticity, Nonlinearity and Forecasting of Streamflow Processes*. IOS Press: Amsterdam.

Wang K, Chen N, Tong D, Wang K, Wang W, Gong J. 2015. Optimizing precipitation station locations: a case study of the Jinsha River Basin. *International Journal of Geographical Information Science* **30(6)**: 1207-1227. DOI: 10.1080/13658816.2015.1119280.

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

Wei C, Yeh HC, Chen YC. 2014. Spatiotemporal scaling effect on rainfall network design using entropy. *Entropy* **16(8)**: 4626-4647. DOI: 10.3390/e16084626.

Werstuck C, Coulibaly P. 2016. Hydrometric network design using dual entropy multi-objective optimization in the Ottawa River basin. *Hydrology Research* (Article in press, published online): nh2016344. DOI: 10.2166/nh.2016.344.

Whetton PH, Fowler AM, Haylock MR, Pittock AB. 1993. Implications of climate change due to the enhanced greenhouse effect on floods and droughts in Australia. *Climatic Change* **25(3)**: 289-317. DOI: 10.1007/BF01098378.

Wilson CB, Valdes JB, Rodriguez-Iturbe I. 1979. On the influence of the spatial distribution of rainfall on storm runoff. *Water Resources Research* **15(2)**: 321-328. DOI: 10.1029/WR015i002p00321.

Woldemeskel FM, Sivakumar B, Sharma A. 2013. Merging gauge and satellite rainfall with specification of associated uncertainty across Australia. *Journal of Hydrology* **499**: 167-176. DOI: 10.1016/j.jhydrol.2013.06.039.

WMO. 1994. *Guide to Hydrological Practices* (5th Edition). World Meteorological Organization (WMO) Publication No. 168.

Wu JS, Han J, Annambhotla S, Bryant S. 2005. Artificial neural networks for forecasting watershed runoff and streamflows. *Journal of Hydrologic Engineering* **10(3)**: 216-222. DOI: 10.1061/(ASCE)1084-0699(2005)10:3(216).

Xu CY, Tunemar L, Chen YD, Singh VP. 2006. Evaluation of seasonal and spatial variations of conceptual hydrological model sensitivity to precipitation data errors. *Journal of Hydrology* **324(1-4)**: 80-93. DOI: 10.1016/j.jhydrol.2005.09.019.

Xu H, Xu CY, Chen H, Zhang Z, Li L. 2013. Assessing the influence of rain gauge network density and distribution on hydrological model performance in a humid region of China. *Journal of Hydrology* **505**: 1-12. DOI: 10.1016/j.jhydrol.2013. 09.004.

Xu H, Xu CY, Saelthun NR, Xu Y, Zhou B, Chen H. 2015. Entropy theory based multi-criteria resampling of rain gauge networks for hydrological modelling – a case study of humid area in southern China. *Journal of Hydrology* **525**: 138-151. DOI: 10.1016/j.jhydrol.2015.03.034.

Xu P, Wang D, Singh VP, Wang Y, Wu J, Wang L, Zou X, Chen Y, Chen X, Liu J, Zou Y, He R. 2017. A two-phase copula entropy-based multiobjective optimization approach to hydrometeorological gauge network design. *Journal of Hydrology* (Article in press): DOI: 10.1016/j.jhydrol.2017.09.046.

Yang FG, Cao SY, Liu XN, Yang KJ. 2008. Design of groundwater level monitoring network with ordinary kriging. *Journal of Hydrodynamics* **20(3)**: 339-346. DOI: 10.1016/S1001-6058(08)60066-9.

Yang X, Xie X, Liu DL, Ji F, Wang L. 2015. Spatial interpolation of daily rainfall data for local climate impact assessment over greater Sydney region. *Advances in Meteorology* **2015**: 563629. DOI: 10.1155/2015/563629.

Yeh HC, Chen YC, Wei C, Chen RH. 2011. Entropy and kriging approach to rainfall network design. *Paddy and Water Environment* **9(3)**: 343-355. DOI: 10.1007/s10333-010-0247-x.

Yilmaz AG, Muttil N. 2014. Runoff estimation by machine learning methods and application to the Euphrates Basin in Turkey. *Journal of Hydrologic Engineering* **19(5)**: 1015–1025. DOI: 10.1061/(ASCE)HE.1943-5584.0000869.

Yilmaz AG, Perera BJC. 2014. Extreme rainfall nonstationarity investigation and intensity-frequency-duration relationship. *Journal of Hydrologic Engineering* **19(6)**: 1160–1172. DOI: 10.1061/(ASCE)HE.1943-5584.0000878.

Yilmaz AG, Hossain I, Perera BJC. 2014. Effect of climate change and variability on extreme rainfall intensity-frequency-duration relationships: a case study of Melbourne. *Hydrology and Earth System Sciences* **18**: 4065–4076. DOI: 10.5194/hess-18-4065-2014.

Yoo C, Jung K, Lee J. 2008. Evaluation of rain gauge network using entropy theory: Comparison of mixed and continuous distribution function applications. Journal of Hydrologic Engineering 13(4): 226-235.

Zealand CM, Burn DH, Simonovic SP. 1999. Short term streamflow forecasting using artificial neural networks. *Journal of Hydrology* **214(1-4)**: 32-48. DOI: 10.1016/S0022-1694(98)00242-X.

Zeinivand H. 2015. Comparison of interpolation methods for precipitation fields using the physically based and spatially distributed model of river runoff on the example of the Gharesou basin, Iran. *Russian Meteorology and Hydrology* **40(7)**: 480-488. DOI: 10.3103/S1068373915070079.

Zheng Z, Chen N, Li P, Wang W. 2012. Integration of hydrological observations into a spatial data infrastructure under a sensor web environment. *International Journal of Digital Earth* **6**: 22-40. DOI: 10.1080/17538947.2012.743605.

Zhu D, Peng DZ, Cluckie ID. 2013. Statistical analysis of error propagation from radar rainfall to hydrological models. *Hydrology and Earth System Sciences* **17**: 1445-1453. DOI: 10.5194/hess-17-1445-2013.