

# Design of sampling locations for river water quality monitoring considering seasonal variation of point and diffuse pollution loads

Vikas Varekar · Subhankar Karmakar ·  
Ramakar Jha · N. C. Ghosh

Received: 29 March 2014 / Accepted: 5 May 2015 / Published online: 26 May 2015  
© Springer International Publishing Switzerland 2015

**Abstract** The design of a water quality monitoring network (WQMN) is a complicated decision-making process because each sampling involves high installation, operational, and maintenance costs. Therefore, data with the highest information content should be collected. The effect of seasonal variation in point and diffuse pollution loadings on river water quality may have a significant impact on the optimal selection of sampling locations, but this possible effect has never been addressed in the evaluation and design of monitoring networks. The present study proposes a systematic approach for siting an optimal number and location of river water quality sampling stations based on seasonal or monsoonal variations in both point and diffuse pollution loadings. The proposed approach conceptualizes water quality monitoring as a two-stage process; the first stage of which is to consider

all potential water quality sampling sites, selected based on the existing guidelines or frameworks, and the locations of both point and diffuse pollution sources. The monitoring at all sampling sites thus identified should be continued for an adequate period of time to account for the effect of the monsoon season. In the second stage, the monitoring network is then designed separately for monsoon and non-monsoon periods by optimizing the number and locations of sampling sites, using a modified Sanders approach. The impacts of human interventions on the design of the sampling net are quantified geospatially by estimating diffuse pollution loads and verified with land use map. To demonstrate the proposed methodology, the Kali River basin in the western Uttar Pradesh state of India was selected as a study area. The final design suggests consequential pre- and post-monsoonal changes in the location and priority of water quality monitoring stations based on the seasonal variation of point and diffuse pollution loadings.

---

V. Varekar · S. Karmakar (✉)  
Centre for Environmental Science and Engineering, Indian  
Institute of Technology Bombay, Powai, Mumbai 400076,  
India  
e-mail: subhankar.karmakar@gmail.com

S. Karmakar  
Interdisciplinary Program in Climate Studies, Indian Institute  
of Technology Bombay, Powai, Mumbai 400076, India

R. Jha  
Department of Civil Engineering, National Institute of  
Technology Rourkela, Rourkela, Orissa 769008, India

N. C. Ghosh  
Ground Water Hydrology Division, National Institute of  
Hydrology, Roorkee, Uttarakhand 247667, India

**Keywords** Diffuse pollution · Kali River · Monsoonal  
variation · Sampling locations · Seasonal variation ·  
Water quality monitoring network

## Introduction

The degradation of freshwater quality is one of the major concerns of aquatic ecosystems on a worldwide scale. Human interference and natural hydrological

processes play very important roles in water quality degradation. The degradation of water quality is a strong motivation for water quality management and determination of the ambient water quality, the extent of the impact of pollution by anthropogenic activities, and possible control measures to restore water quality. Water quality monitoring is a collection of representative and quantitative information on physical, biological, and chemical water quality characteristics with respect to space and time (Sanders et al. 1983; Ward et al. 1989; Strobl and Robillard 2008). According to recent reviews (Dixon and Chiswell 1996; Strobl and Robillard 2008; Mishra and Coulibaly 2009, 2010), the design of a water quality monitoring network (WQMN) consists of a set of tasks: (a) determination of the number and spatial distribution of monitoring stations, (b) selection of a sampling frequency, and (c) selection of water quality parameters to be monitored. Defining monitoring objectives and budgetary constraints has been found essential while designing water quality monitoring network.

In a global scale, various countries, viz. European Union member states Norway and the European Commission (Water Framework Directives of the European Community 2003), USA (United States Environmental Protection Agency (USEPA) 2003; United States Geological Survey 2005), India (Ministry of Environment and Forest of the Government of India 2005; Central Pollution Control Board of India 2007), Canada (Canadian Council of Ministers of the Environment 2006), and Australia (Department of Water of the Government of Western Australia 2009), have constituted their own protocols and guidelines for design of water quality monitoring networks. Most recently, USEPA published a report on National Rivers and Streams Assessment 2013–2014, in which sampling sites were selected randomly, based on the statistical survey design to represent the population of the streams and rivers in their ecological region, i.e., the geographic area which consist of similar ecological features, climate, and plant and animal communities (USEPA 2014).

Apart from these guidelines and frameworks, various approaches to the design of WQMN have been proposed by previous researchers. Sharp (1971) proposed a systematic approach to uniform sampling of rivers and streams. In this method, optimum sampling locations are selected by dividing the stream network into successive halves by identifying the centroids. Sanders et al. (1983) proposed a modified form of Sharp's method in which tributaries were replaced by pollution loadings and the number of outfalls. Both the Sharp and Sanders

approaches are based on the topology of the river basin. The Sanders et al. (1983) modification of Sharp's approach has been used by various researchers for selection of sampling locations (Varekar et al. 2012; Do et al. 2011, 2012) and sub-basins (Cetinkaya and Harmancioglu 2012). Water quality parameters are inherently stochastic in nature (Sanders et al. 1983; Harmancioglu and Alphasan 1992, 1994; Harmancioglu et al. 1999; Ozkul et al. 2000). Therefore, various statistical techniques, viz. statistical entropy (Harmancioglu and Alphasan 1992, 1994; Harmancioglu et al. 1999; Ozkul et al. 2000; Karamouz et al. 2009a; Alameddine et al. 2013), statistical kriging (Lo et al. 1996; Karamouz et al. 2009b; Hudak and Loaiciga 1993), and multivariate statistical techniques (Ouyang 2005; Varol et al. 2012; Wang et al. 2012; Musthafa et al. 2012, 2014), have been used effectively by researchers for the design and assessment of sampling sites. Optimization approaches have also been applied as effective tools for the design of sampling locations (Telci et al. 2009; Cetinkaya and Harmancioglu 2012; Asadollahfardi et al. 2014). Apart from these approaches, fuzzy logic and genetic algorithms on geographic information system (GIS) platforms (Park et al. 2006; Strobl et al. 2006a, b), fuzzy theory with multiple criteria analysis (Chang and Lin 2014a), and VIKOR method (Chang and Lin 2014b) have also used for the design of water quality monitoring networks (Table 1).

The anthropogenic activities are the key elements responsible for alternation of surface water quality. These activities are accounted in the terms of pollution loads. Thereby, consideration of the pollution potential of anthropogenic activities becomes vital for effective water quality management. Hence, the principal objective of proposed river water quality monitoring design is evaluating the potential of pollution load entering into the river and the need for respective pollution control measures. In addition, present monitoring design also suits with the monitoring objectives defined by Central Pollution Control Board (CPCB 2007) of India, which are listed as follows: (i) rational planning of pollution control strategies; (ii) identification of nature and magnitude of pollution control measures; (iii) effectiveness evaluation of existing pollution control efforts; (iv) identification of the mass flow of contaminants in surface water and effluents; (v) formulation of standards and permit requirements; (vi) testing of compliance with standards and classifications for waters; and (vii) early warning and detection of pollution. The reviews that have been conducted on the guidelines and frameworks for water

**Table 1** Summary of literature on design of water quality monitoring network

Sr. no.	Methodology	Author/s	Results
01	<p>Method based on topology: Sanders approach (please refer to “Design of water quality sampling locations for monsoon and non-monsoon periods” section)</p>	<p>Sharp (1971) Sanders et al. (1983) Do et al. (2011, 2012)</p>	<p>The proposed approach was satisfactorily applied to the Edisto River in South Carolina for development of uniform sampling plan The proposed approach was demonstrated for the Massachusetts portion of the Connecticut river. All the three design approaches, viz. number of tributaries, pollution loadings, and number of outfalls, were successfully tested Proposed a methodology for the selection of sampling locations, taking into account human-induced and agricultural diffuse pollution loading in a GIS framework. In this methodology, potential sampling sites are selected by applying the Sanders approach (Sanders et al. 1983). However, diffuse pollution loading is estimated without accounting for the hydrological characteristics of the river basin, i.e., considering distance zones, implying a major assumption of a “linear ground surface,” which was a limitation of their study. This methodology was applied to design of nutrient monitoring points upstream of the Fetsui reservoirs, Taipei, Taiwan. The optimum number of sampling locations is four</p>
02	<p>Statistical entropy: 2.1 The concept of statistical entropy has been applied to the design and assessment of surface water quality monitoring networks. “Entropy” is a measure of the degree of uncertainty, which is derived from information theory, as described by Shannon and Weaver (1949). In this approach, a water quality monitoring network is treated as an information system, and the entropy is calculated and considered in the design and assessment of sampling sites, the frequency of sampling, and the selection of parameters, in terms of cost-effectiveness and efficiency of the network</p> <p>2.2 Maximum entropy-based hierarchical spatiotemporal Bayesian model: This model is used for optimization of existing water quality monitoring network by identifying and retaining the sampling locations delivering maximum information; whereas, new sampling locations are also identified for modification and extension of a network. The concept of entropy is used to account the system uncertainties. The entropy-based criteria are as follows: dissolved oxygen standard violation entropy, total system entropy, and chlorophyll-<i>a</i> standard violation entropy are considered for design of an estuarine monitoring network. A multiple attribute decision-making framework is proposed to consider three competing design criteria, which results in an optimal monitoring network</p>	<p>Hamancioglu and Alphasan (1992) Ozkul et al. (2000) Karamouz et al. (2009a) Alameddine et al. (2013)</p>	<p>The proposed entropy method was demonstrated for assessing the efficiency and the benefits of the existing water quality monitoring network of the highly polluted Porsuk river in Turkey. The significant shortcomings, major strengths, and recommendations for possible use of methodology were discussed in brief A newly introduced approach for the assessment of combined spatial/temporal frequencies of monitoring networks was tested for observed water quality data of the Mississippi River in Louisiana. The spatial, temporal, and space/time design of network based on the information content of the system was successfully demonstrated The proposed entropy-based approach was applied to the Karoon river system located in the southeastern of Iran. The optimum number of sampling locations was selected based on the measure of transinformation. The optimal number of sampling locations is 10 out of 28 potential sampling sites The proposed approach was implemented for the optimization of the water quality monitoring network of Neuse River Estuary in North Carolina, USA. Results obtained shows that the high priority areas identified by the dissolved oxygen violation entropy and total system entropy criteria are largely coincident</p>

**Table 1** (continued)

Sr. no.	Methodology	Author/s	Results
03	<p>Statistical kriging: Kriging is a geospatial method (Cressie and Christopher 2011) for regionalizing spatiotemporal variations, which has been widely used in various fields, such as hydrology, surveying, mining, geology, etc. In this approach, the spatiotemporal variation of water quality is determined by estimating variance. The number of sampling points is optimized by selecting the sampling locations based on the estimated variance</p>	<p>Lo et al. (1996)</p> <p>Karamouz et al. (2009b)</p>	<p>The kriging method was applied for design of water quality monitoring network of the Keelung River in Northern Taiwan. The optimal number of sampling locations is 21. More sampling locations are needed in the downstream reaches</p> <p>The model consists of kriging method and analytical hierarchy process was applied for the design of water quality monitoring network of the Karoon River, from Gotvand Dam to Persian Gulf, Iran. The optimum number of sampling locations is 35</p>
04	<p>Multivariate statistical techniques, such as principal component analysis (PCA) and principal factor analysis (PFA), have been used effectively for assessment of water quality monitoring networks by various researchers. These techniques are used to identify important components or factors that explain most of the variance of the system. Those factors are designed to reduce the number of variables to a small number of indices while attempting to preserve relationships present in the data. The task of selecting important monitoring stations, detecting changes in the characteristics of water quality parameters, and data reduction and interpretation are accomplished using PCA and PFA techniques</p>	<p>Ouyang (2005)</p> <p>Musthafa et al. (2012, 2014)</p>	<p>The PCA/PFA was applied for assessment of the existing water quality monitoring network of the main stem of lower St. Johns River in Florida, USA with 22 sampling stations. The obtained result shows that three monitoring stations were non-principal because they failed to explain the annual variance of the data set and therefore could be less important</p> <p>The assessment of existing water quality network of the Kabbini River basin of Kerala, India was carried out using multivariate statistical techniques (PCA/PFA). The results show that there is a need to redesign the network seasonally</p>
05	<p>Optimization</p> <p>5.1 Fuzzy logic with genetic algorithm: a methodology based on GIS and a genetic algorithm was proposed for the design of WQMN. The planning objectives were identified, and subsequently, the corresponding fitness functions were defined. The genetic algorithm was then connected with a GIS database to identify the optimal sampling locations</p> <p>5.2 Fuzzy logic with hydrological simulation model: Introduced a methodology for identifying the critical sampling points (CSP) within a watershed using an approach that integrated a hydrological simulation model and concepts of fuzzy logic in a GIS platform. In this approach, hydrological factors and factors pertaining to soils, vegetation, topography, and land use are taken into consideration in the selection of critical sampling locations</p>	<p>Park et al. (2006)</p> <p>Strobl et al. (2006a, b)</p>	<p>The proposed approach was applied to the Nakdong River system, Korea for the design of water quality monitoring network. The results obtained show that out of 110 currently in operation sampling sites, only 35 are coinciding with the optimally designed network</p> <p>The proposed CSP methodology was validated for a small experimental Pennsylvania watershed. The total phosphorous (TP) data from number of single storm event was available for different sampling locations of a given watershed. The comparison of the ratios of the predicted and observed TP loads between various sampling locations shows that the CSP methodology works well</p>
5.3	<p>Linear programming: The proposed approach was entitled "optimal water quality monitoring network for river systems." The optimization model considers two decision variables, the first to minimize the average detection time of the contamination event and the second to maximize the reliability of the monitoring system</p>	<p>Telci et al. (2009)</p>	<p>To demonstrate the proposed methodology, Altamaha in the state of Georgia, USA was selected as a case study. The obtained results show that the proposed model can be effectively used for the optimal design of river water quality monitoring network</p>
5.4	<p>Multiple criteria analysis: This approach was used to check the suitability of the water quality monitoring design. Seven criteria, viz. percentage of farmland area, amount of diffuse source pollution, green cover ratio, percentage of built-up area, ratio of landslide area, density of water quality monitoring stations, and overutilization on hillsides, were selected for the multiple criteria analysis. These criteria are weighted and normalized. The weighted method is applied to score the sub-basins. The density of water quality monitoring stations is depends on the score of sub-basins. The prioritization of the need of a higher density water quality monitoring stations is carried out by fuzzy theory</p>	<p>Chang and Lin (2014a)</p>	<p>The proposed approach was used to check the suitability of the water quality monitoring design in the Taipei Water Resource Domain in northern Taiwan. The obtained result shows that few sub-basins require more water sampling locations</p>

**Table 1** (continued)

Sr. no.	Methodology	Author/s	Results
5.5	VIKOR method: “VIKOR” a multiple criteria analysis was applied to evaluate the design of the water quality monitoring network. For multiple criteria analysis, five criteria were selected, viz. overutilization area ratio of hillsides, green cover ratio, landslide area ratio, density of water quality monitoring stations, and diffuse source pollution	Chang and Lin (2014b)	The proposed approach was used to evaluate the design of water quality monitoring in the Taipei Water Resource Domain in Northern Taiwan. The result shows that the BeiShih Creek Basin requires more monitoring stations
5.6	Dynamic programming and Sanders approach: A dynamic programming approach for assessment of water quality sampling sites was proposed. In this assessment, Sanders et al. (1983) modification of Sharp’s approach was used for the selection of sub-basins, and dynamic programming was used to select the optimum number of sampling stations	Cetinkaya and Harmancioglu (2012)	The proposed methodology was demonstrated for the design of a water quality monitoring network in Gediz River basin along the Aegean coast of Turkey. The obtained results show that DPA can also be an effective tool in optimizing the number of stations
		Asadollahfardi et al. (2014)	The dynamic programming approach has been applied to evaluate the existing water quality sampling stations of the Sefid-Rud River, located in the north of Iran. The results shows that eight and seven stations out of 21 existing monitoring stations need to be retained on the upstream and downstream

quality monitoring network in various countries and on the numerical and statistical approaches to the design of such networks have shown that there is no universally accepted unique approach to design of water quality monitoring networks (Strobl and Robillard 2008). The guidelines are mostly based on experts’ judgments and do not consider the analyses of existing water quality inventory data and watershed and river characteristics, and they also may lack a mathematical basis for design, while the numerical and statistical approaches may be too theoretical and mathematically rigorous. Moreover, statistical approaches are more efficient for assessment of existing monitoring networks than for designing new networks. Most of the above-discussed approaches consider the effect of point or discrete sources and the existing status of river water quality. However, intensive use of fertilizers on agricultural land may generate enormous amounts of nutrient-rich agricultural runoff that enters into the river as a result of diffuse source of pollution (Jha et al. 2005). In tropical countries, where monsoons are predominant, river water quality may vary significantly during the monsoon (June to September) and non-monsoon (October to May) seasons (Tsirkunov et al. 1992; Bhangu and Whitfield 1997; Vega et al. 1998; Hanrahana et al. 2003; Singh et al. 2004; Ouyang et al. 2006; Chang 2008; Altin et al. 2009), mainly due to the impact of diffuse pollution loads. Hence, both point and diffuse pollution sources must be considered to achieve a realistic design for surface water quality monitoring network.

Based on a comprehensive review conducted in the present study, it has been established that the design of a water quality monitoring network should consider (1) the location and magnitude of both point and diffuse pollution loadings and (2) their temporal variations, i.e., during the monsoon (June to September) and non-monsoon (October to May) seasons, which motivates us in the present study. To account for the effect of seasonal/temporal variability, i.e., the monsoon and non-monsoon seasons, the annual cycle is divided into two seasons in Indian applications (Jain et al. 2007; Central Ground Water Board, Government of India 2004). It has also been found that a periodic assessment of the monitoring network is required to consider this seasonal variation, which may result in *pre-monsoon* (before the monsoon season starts) and *post-monsoon* (immediately at the end of monsoon season) design of the network. In this regard, GIS, geographic positioning systems (GPS), and remote sensing tools are found useful in exploiting various types of data, such as data on the hydraulic and hydrological characteristics of

river basins, land use practices, agricultural cropping patterns, use of fertilizers, the relative locations of pollution sources and monitoring stations, etc. Such a realistic design approach helps a watershed manager to design a network with a user-friendly interface. The present study is the first effort to address the effect of monsoonal and non-monsoonal variations in the design of a water quality monitoring network, considering both point and diffuse sources of pollution. The proposed approach conceptualizes water quality monitoring as a two-stage process, the first stage of which involves considering all potential water quality sampling sites selected on the basis of existing guidelines or frameworks and the locations of both point and diffuse pollution sources. The monitoring at all sampling sites should be continued for a sufficiently long period of time to account for the effect of the monsoon season on water quality. The monitoring network is then designed separately for the monsoon and non-monsoon periods by optimizing the number of sampling sites, subject to budgetary constraints and the importance of each individual station, which is quantified using optimal hierarchy levels (Sanders et al. 1983; Do et al. 2011, 2012), based on the binary tree structure of graph theory (Knuth 1968). The second stage involves establishment of a monitoring program with optimized monitoring of the network based on the network structure in the monsoon and non-monsoon seasons. The network monitoring structure should be modified or redesigned in subsequent years with updated water quality data. The Kali River basin in western Uttar Pradesh, India (Jha et al. 2005, 2007) was selected as the study area. The potential sampling sites were identified based on the locations of points of entry of point and diffuse pollution into the river, using geospatial techniques. The modified Sanders approach (Sanders et al. 1983), which is well accepted among researchers (Do et al. 2011, 2012; Varekar et al. 2012), was applied to the design of sampling sites.

## Materials and methods

### Proposed methodology

The proposed methodology consists of three components: (I) field investigations, primary data collection and geospatial analysis; (II) pollution load estimation; and (III) design of water quality sampling locations for monsoon and non-monsoon periods. Figure 1 shows the organization and flow of the proposed method.

Field investigations, primary data collection, and geospatial analysis

The proposed methodology introduces a realistic approach to the selection of sampling locations based on field survey and the use of geospatial techniques to account for the impact of human activities. The river system considered in this study is the Kali River basin in Uttar Pradesh, India (Fig. 2). The Kali River is a tributary of the Hindon River. Its origin is near the Saharanpur district, and it travels a length of 125 km before merging with the Hindon River. The catchment area of the Kali River is 1,475.50 km<sup>2</sup>. Kali River has a significant socioeconomic value, i.e., people residing in the watershed have been depending on the river water for drinking, bathing, washing of clothes, and also for agriculture. But the river water quality is gradually degrading due to disposal of municipal wastewater and significant volume of agricultural runoff entering into the river (Jha et al. 2007). Therefore, the proposed monitoring network design framework was applied to the Kali River system, for evaluating the potential of pollution load entering into the river and the need of pollution control measures. The sampling locations were selected based on the location of mouth of sub-watersheds, as the water quality monitored at mouth should represent the characteristics of the respective sub-watershed. Hence, in the current study, the 16 sampling locations were the representative monitoring sites, both spatially and temporally, and not restricted to any particular nodal information. Topo sheets (survey of India: 53 G/9, 53 G/10, 53 G/11, 53 G/12, 53 G/13, 53 G/14), a location map, district map, and political map, were used to identify the exact location of the catchment area. A digital elevation map (DEM) with 20-m resolution was obtained from the Indian Space Research Organization (ISRO). The watershed was delineated using ArcGIS 9.3. The main stream of the Kali River was delineated using the terrain processing tool in Arc-Hydro by defining the threshold value. Field visits were conducted to collect details of the cropping pattern, fertilizer utilization, and the locations of point pollution sources (estimated using GPS) and to check the accessibility of sampling sites. A total of 16 sampling stations were identified, based on the available guidelines for water quality monitoring (CPCB 2007), the locations of sub-watershed outlets, and the relative locations of point pollution sources. Of these 16 monitoring stations, six are located downstream from point sources, merging

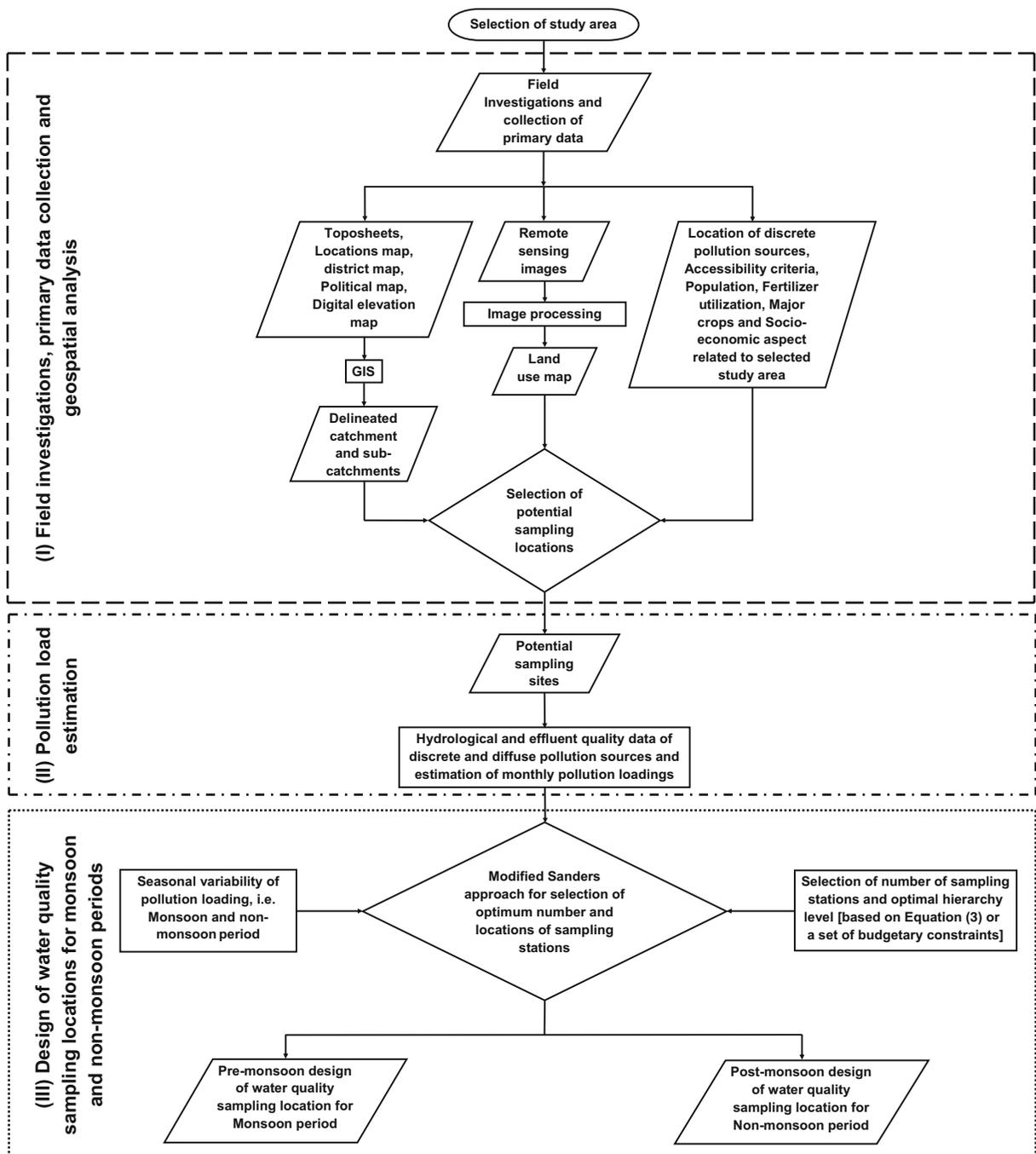
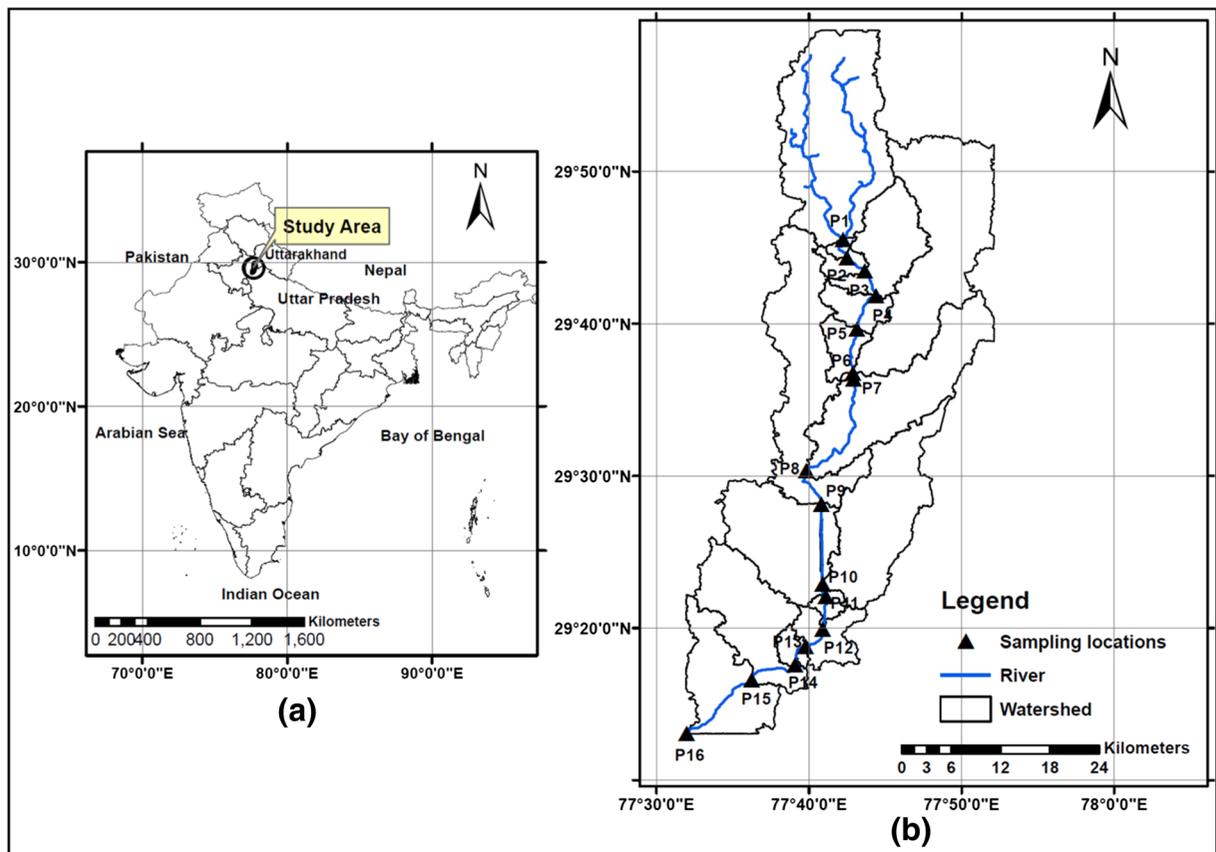


Fig. 1 Framework of proposed methodology

with the Kali River in the form of natural open drains. The effects of diffuse pollution sources are captured by all 16 monitoring stations, as they are located at the outlets of sub-watersheds. All 16 sampling sites (the details of which are tabulated in Table 2) are accessible throughout the year. The land use map (shown in

Fig. 3a) was generated from satellite images of the study area (Source, United States Geological Survey (USGS) Earth Explorer) using the image processing software ERDAS IMAGINE 9.1, following a supervised classification. The land use map shows that 80.11 % of the watershed area is used for agricultural purposes (Fig. 3).



**Fig. 2** a Location map of the study area. b Location of potential sampling sites in Kali River basin

### Pollution load estimation

Extensive sampling from the 16 locations on the river was conducted. As mentioned in the “[Introduction](#)” section, the proposed method consists of two stages of monitoring. The initial monitoring in the first year should be more comprehensive than the subsequent seasonal monitoring. Therefore, the hydraulic and water quality characteristics of the river were estimated for 1 year, from March 1999 to February 2000. Sampling was conducted on the 10th and/or the 11th of each month (three times daily) during the non-monsoon season, November to June. During the monsoon season (July to October), sampling was conducted based on the occurrence of storm events. It should be noted that the Kali River basin is located in northern India, where the monsoon approaches later than in other parts of country. The “grab samples are single samples collected at a specific spot at a site over a short period of time (typically seconds or minutes)” (APHA 2000), and the sampling method is called as

grab sampling. The grab sampling method was used to collect the river water samples. The samples were collected at a depth of 15 cm from 3 points across the location of sampling (1/3, 1/2, and 2/3 distance along river cross section). The cross-sectional area of the river was estimated at each location using a measuring tape and a leveling staff. The velocity and depth of the water were measured using a current meter (electromagnetic current meter WTW 197, Germany) and a leveling staff, respectively. The cross-sectional area, depth, and velocity were used to estimate the discharge at various locations along the river. The water samples collected were analyzed to determine their pH, temperature, electrical conductivity, and chemical properties, i.e., their nitrate, phosphate, biochemical oxygen demand (BOD), and dissolved oxygen (DO) contents, following standard methods (APHA/AWWA/WEF 2000). For example, the azide modification method has been used for determination of dissolved oxygen (APHA/AWWA/WEF 2000). Portable meters (Hach sensION1 pH meter and Hach

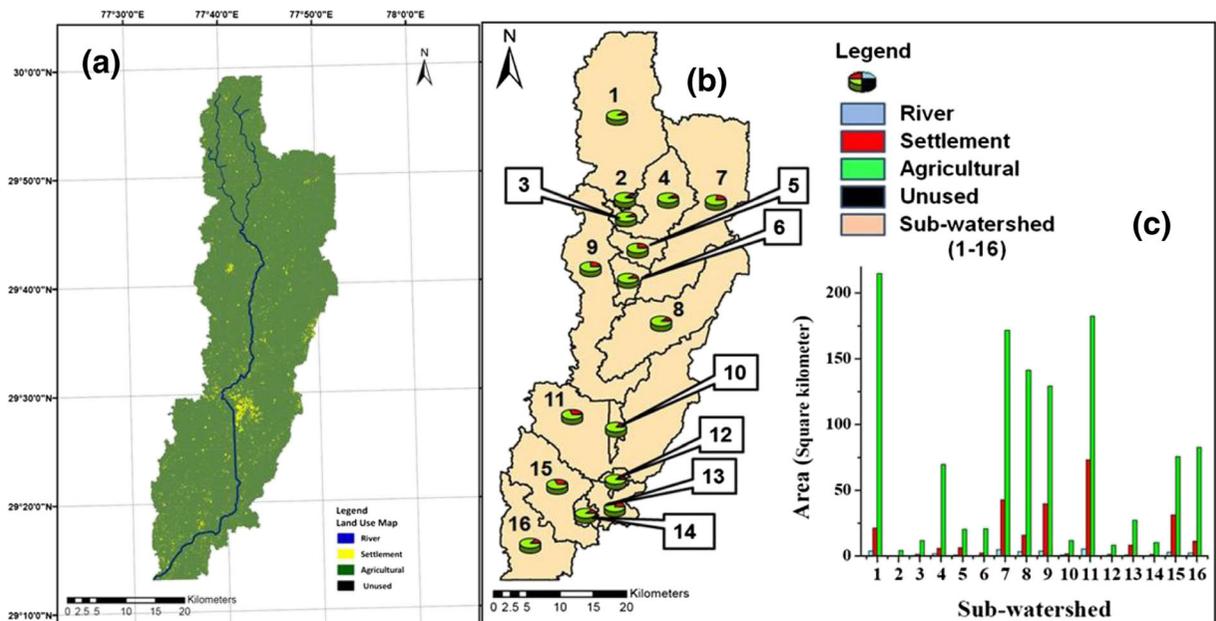
**Table 2** Details of sampling stations and river reaches

Sampling station (P)	Sampling station designation	River reach (R)	Pollution source
P <sub>1</sub>	Rastam	NA	NP
P <sub>2</sub>	Diwaleri	R <sub>1</sub> : Rastam–Diwaleri (P <sub>1</sub> –P <sub>2</sub> )	NP
P <sub>3</sub>	Chetanpuri	R <sub>2</sub> : Diwaleri–Chetanpuri (P <sub>2</sub> –P <sub>3</sub> )	NP
P <sub>4</sub>	Miragpur	R <sub>3</sub> : Chetanpuri–Miragpur (P <sub>3</sub> –P <sub>4</sub> )	NP
P <sub>5</sub>	Sanpla (u/s)	R <sub>4</sub> : Miragpur–Sanpla (u/s) (P <sub>4</sub> –P <sub>5</sub> )	P
P <sub>6</sub>	Sanpla (d/s)	R <sub>5</sub> : Sanpla (u/s)–Sanpla (d/s) (P <sub>5</sub> –P <sub>6</sub> )	NP
P <sub>7</sub>	Malira (u/s)	R <sub>6</sub> : Sanpla (d/s)–Malira (u/s) (P <sub>6</sub> –P <sub>7</sub> )	P
P <sub>8</sub>	Malira (d/s)	R <sub>7</sub> : Malira (u/s)–Malira (d/s) (P <sub>7</sub> –P <sub>8</sub> )	NP
P <sub>9</sub>	Khanjanpur (u/s)	R <sub>8</sub> : Malira (d/s)–Khanjanpur (u/s) (P <sub>8</sub> –P <sub>9</sub> )	P
P <sub>10</sub>	Khanjanpur (d/s)	R <sub>9</sub> : Khanjanpur (u/s)–Khanjanpur (d/s) (P <sub>9</sub> –P <sub>10</sub> )	NP
P <sub>11</sub>	Vihelna (u/s)	R <sub>10</sub> : Khanjanpur (d/s)–Vihelna (u/s) (P <sub>10</sub> –P <sub>11</sub> )	P
P <sub>12</sub>	Vihelna (d/s)	R <sub>11</sub> : Vihelna (u/s)–Vihelna (d/s) (P <sub>11</sub> –P <sub>12</sub> )	NP
P <sub>13</sub>	Mansurpur (u/s)	R <sub>12</sub> : Vihelna (d/s)–Mansurpur (u/s) (P <sub>12</sub> –P <sub>13</sub> )	P
P <sub>14</sub>	Mansurpur (d/s)	R <sub>13</sub> : Mansurpur (u/s)–Mansurpur (d/s) (P <sub>13</sub> –P <sub>14</sub> )	NP
P <sub>15</sub>	Ancholi	R <sub>14</sub> : Mansurpur (d/s)–Ancholi (P <sub>14</sub> –P <sub>15</sub> )	P
P <sub>16</sub>	Pitlokar	R <sub>15</sub> : Ancholi–Pitlokar (P <sub>15</sub> –P <sub>16</sub> )	NP

NP diffuse source of pollution, P point source of pollution, u/s upstream, d/s downstream, NA not applicable]

sensION5 conductivity meter) were used to measure the physical parameters in the field. The remaining parameters were analyzed in the laboratory. The diffuse source BOD ( $L_d$ ) inflow to the river was found to be difficult to determine. Therefore, in the present

analysis, a mass balance approach was used to estimate the  $L_d$  values (Jha et al. 2007) from the available data. To understand continuous spatiotemporal variation of river flow and water quality characteristics, a set of nonparametric cubic splines (Stasinopoulos and

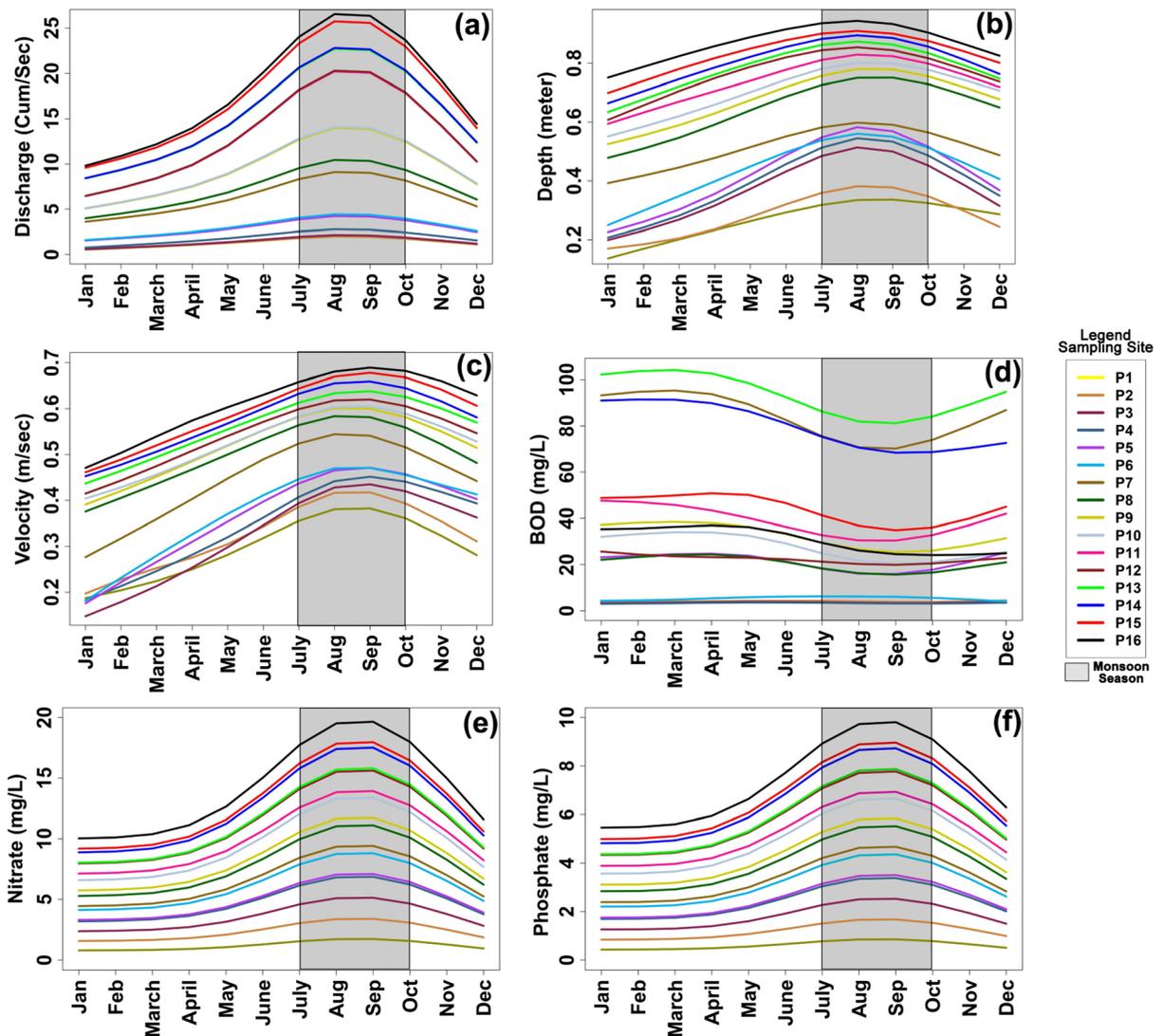


**Fig. 3** a Land use map of the study area. b Relative fraction of land use for each sub-watershed. c Area under different land use category for each sub-watershed (in square kilometer)

Rigby 2007) was fitted to the observed data and represented in Fig. 4. The exact locations of six-point pollution sources were determined during the field visits, and the BOD loadings of these point sources were simulated using hydraulic and water quality data for the downstream river water (Table 3). The diffuse source loadings (nitrate and phosphate) were also estimated from data collected at the 16 sampling locations (Table 4). The procedures followed to estimate the discharge from the pollution sources and to simulate the characteristics of the effluent are discussed in Appendix 1.

Design of water quality sampling locations for monsoon and non-monsoon periods

As mentioned in the “Introduction” section, the proposed approach conceptualizes water quality monitoring as a two-stage process. In the first stage, all 16 sampling locations were identified as potential sampling sites, based on the guidelines for water quality monitoring (CPCB 2007), the locations of sub-watershed outlets, and the relative locations of point pollution sources. After completion of 1 year of comprehensive water quality monitoring, a modified Sanders approach was used to



**Fig. 4** Spatiotemporal variation of flow and water quality characteristics using nonparametric cubic splines: **a** discharge, **b** depth, **c** velocity, **d** BOD, **e** nitrate, and **f** phosphate

**Table 3** Seasonal pollution loading of point sources

Point source	River reach (R)	Monsoon			Non- monsoon		
		$\Delta Q_D^{R_i}$ (m <sup>3</sup> /s)	BOD concentration (mg/l)	BOD load (kg/day)	$\Delta Q_D^{R_i}$ (m <sup>3</sup> /s)	BOD concentration (mg/l)	BOD load (kg/day)
1	R <sub>4</sub>	1.32	369.92	42,296.34	0.75	583.64	37,846.53
2	R <sub>6</sub>	3.13	1,124.94	303,830.65	1.59	1,499.77	206,290.02
3	R <sub>8</sub>	2.40	125.45	25,968.48	0.78	445.37	30,095.90
4	R <sub>10</sub>	4.57	158.67	62,713.91	0.94	487.26	39,666.18
5	R <sub>12</sub>	2.17	1,378.24	258,498.85	1.87	1,870.39	302,959.78
6	R <sub>14</sub>	2.26	264.18	51,556.49	0.98	690.21	58,488.51

$\Delta Q_D^{R_i}$ : discharge of point pollution source

select the optimal locations and number of sampling sites, taking into consideration the seasonal variation in the point and diffuse source pollution loadings.

*Determination of optimum locations for sampling*

The Sanders approach (Sanders et al. 1983) is based on Sharp’s (1971) sampling method. In Sharp’s method, a river network is divided into a number of interior and

exterior links. An *exterior link* is mainly tributary, has a minimum mean discharge, and is not fed by other defined streams. An exterior link or tributary is also called a *first-order tributary*. An *interior link* is not a tributary; it is formed by the intersection of two exterior tributaries and is called a *second-order tributary* (Sanders et al. 1983). Each exterior tributary or link that contributes to the main stretch of a river is assigned a magnitude or weight of one. The magnitude or weight of an interior

**Table 4** Seasonal pollution loading of diffuse sources

Sampling station designation	Sampling station (P)	Nitrate (kg/day)		Phosphate (kg/day)	
		Monsoon	Non-monsoon	Monsoon	Non-monsoon
Rastam	P <sub>1</sub>	332.42	72.41	164.04	39.14
Diwaleri	P <sub>2</sub>	6.31	1.35	3.11	0.73
Chetanpuri	P <sub>3</sub>	18.31	3.90	9.02	2.09
Miragpur	P <sub>4</sub>	107.30	22.88	52.85	12.27
Sanpla (u/s)	P <sub>5</sub>	33.50	9.06	16.87	5.28
Sanpla (d/s)	P <sub>6</sub>	32.37	7.13	15.99	3.88
Malira (u/s)	P <sub>7</sub>	277.97	71.61	139.31	41.10
Malira (d/s)	P <sub>8</sub>	219.69	48.72	108.58	26.52
Khanjanpur (u/s)	P <sub>9</sub>	213.12	58.04	107.43	33.89
Khanjanpur (d/s)	P <sub>10</sub>	18.06	4.19	8.96	2.32
Vihelna (u/s)	P <sub>11</sub>	308.84	91.00	157.03	54.34
Vihelna (d/s)	P <sub>12</sub>	13.10	2.94	6.48	1.61
Mansurpur (u/s)	P <sub>13</sub>	44.85	12.05	22.59	7.02
Mansurpur (d/s)	P <sub>14</sub>	15.95	3.57	7.89	1.95
Ancholi	P <sub>15</sub>	128.68	38.11	65.46	22.78
Pitlokar	P <sub>16</sub>	129.28	29.60	64.08	16.32

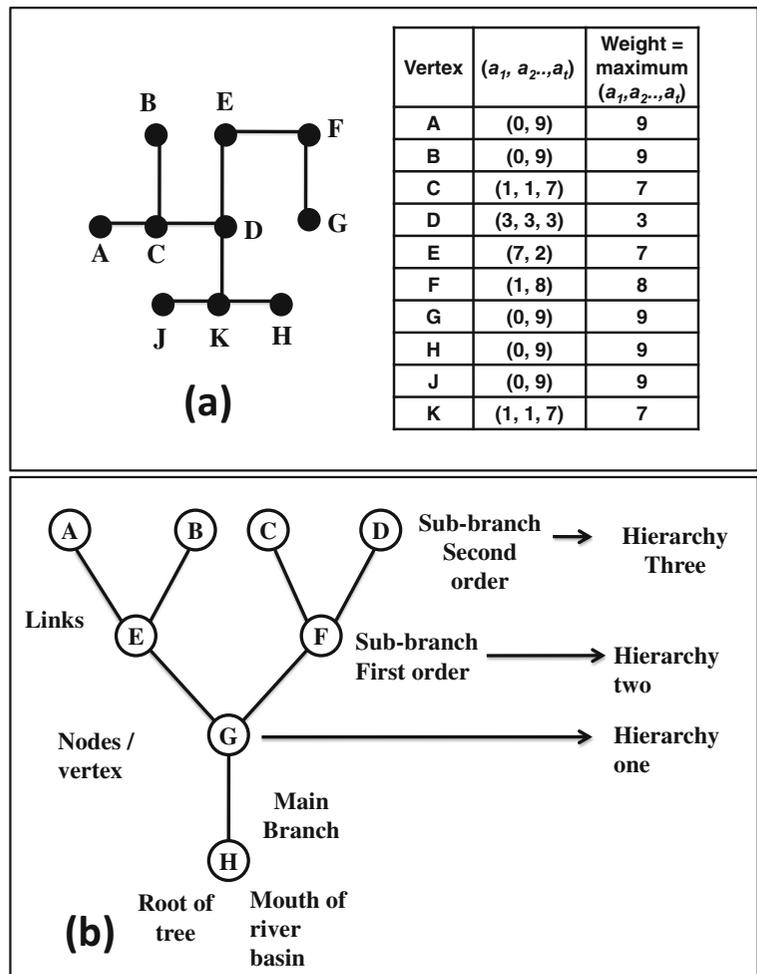
link is equal to the sum of the magnitudes of intersecting exterior links. The magnitude at the mouth of a river watershed is equal to the number of contributing exterior tributaries. After numbering the entire river network, the optimal sampling sites are selected based on the centroids of the river network.

*The centroid of river network*

According to Sharp’s approach, a river network is analogous to a binary tree structure in graph theory. The tree is a vital nonlinear structure used in computer programming (Knuth 1968). “Tree structure means the branching relations between nodes (data point) much as that found in the trees of nature” (Knuth 1968). The start (origin) and end points of

links (i.e., points of intersection of links and the terminating points of links) are termed vertices, and links joining vertices are the edges of a tree. The centroid is the vertex with the minimum weight (i.e., the vertex for which leading sub-trees have the minimum number of vertices). In other words, the centroid of a tree is the vertex that has approximately an equal number of vertices on its upstream and downstream ends. The tree structure and its centroid are as shown in Fig. 5a. For example, the hypothetical tree shown in Fig. 5a, conceptualized from Knuth (1968)), has 10 vertices (A to K),  $t$  denotes the number of leading sub-trees for a given vertex, and  $a_1, a_2, \dots, a_t$  are the number of vertices in the respective sub-trees. The weight of each vertex is estimated. The vertex  $D$  has the minimum weight, 3.

**Fig. 5** a Centroid of tree. b Modified Sander’s approach for a hypothetical river tree



Hence, vertex  $D$  is the centroid of the given hypothetical tree. In 1971, Sharp used the concept of a centroid (Knuth 1968) to divide a river network into approximately equal halves. He proposed the mathematical expression for determining the centroidal link, i.e., the link that divides the network into two equal halves. In nature, it has been observed that in general, each river tributary has two sub-branches. Hence, the bifurcation ratio for a river system is considered to be 2. Therefore, a river system can be considered analogous to a binary tree structure. The Sanders approach is actually a modification of Sharp's method, in which tributaries are replaced by the number of outfalls and pollution loadings (Sanders et al. 1983).

*Modified Sanders approach*

In the present analysis, a modified Sanders approach, derived by Do et al. (2011), was used to determine the optimum locations of the sampling sites. The potential sampling locations are the nodes (data points)/vertices of a river tree, as shown in Fig. 5b. The magnitude of each node is equal to the pollution loading at the respective sampling point. The mouth of the watershed is the root of a tree, and its magnitude is equal to the summation of the pollution loads at each node/vertex within the watershed. The first centroid, i.e., the first-order station (first hierarchy), is the node/sampling point with a magnitude closest to  $M_h$  (Eq. 1):

$$M_h = \left[ \frac{N_o + 1}{2} \right] \tag{1}$$

where  $M_h$  is the magnitude of the node/sampling location at the  $h$ th hierarchy, and  $N_o$  is the total pollution load at the mouth of river basin. The first sampling location is placed at the first centroid, identified using Eq. 1, which indicates the main branch of the tree [the link joining the centroidal node (data point) to the root of the tree]. The first centroid is the unique node with a magnitude closest to half of the whole river network's magnitude. Hence, the first centroid is at the first hierarchy level or the principal branch of a tree and deserves the highest priority in sampling. The successive centroids at the different hierarchy levels are estimated using Eq. 2. For example, there are two centroids at

the second hierarchy level or first-order sub-branch of tree, with the second highest priority in sampling.

$$M_{h+1} = \left[ \frac{M_h + 1}{2} \right] \tag{2}$$

where  $M_{h+1}$  is the magnitude of the node/sampling location at the  $(h+1)$ th hierarchy. The river system is a binary tree structure; hence, the maximum number of centroids used to determine sampling locations at various hierarchy levels is estimated using Eq. 3:

$$n = 2^h - 1 \tag{3}$$

where  $n$  indicates the number of sampling stations at the  $h$ th hierarchy level.

*Determination of optimum number of sampling sites*

After delineation of the sampling locations, the optimum number of sampling sites is estimated. The availability of funding is the governing factor in the selection of the optimum number of sampling locations. If the budget is known, then the optimum number of sampling locations is determined by the total available funding divided by the operating cost of a single monitoring site. If the budget is unknown, then  $M_h$  values are first calculated as a function of the pollution loading (Eqs. 1 and 2). Do et al. (2011) proposed that the sampling should be stopped at  $M_h$  (the  $h$ th hierarchy), which is much smaller than  $M_1$  (the first centroid), and that the optimum number of sampling locations correspond to the hierarchy level  $M_h$  (Eq. 3). Another approach to selection of the optimum number of sampling sites is to consider the effluent disposal standards of wastewater. According to this approach, the number of sites corresponding to the sampling locations violating the effluent disposal standards would be the optimum number of sampling sites.

**Results and discussion**

The proposed approach was applied to Kali River basin, located in western Uttar Pradesh, India. Two distinct networks of sampling locations were designed by accounting the seasonal variation of both point and diffuse pollution loads, i.e., for the monsoon and non-monsoon seasons. The design process involves selection of the optimum number and the locations of sampling points. The design of the sampling locations for the monsoon

period loading is termed the pre-monsoon design, while that for the non-monsoon period loading, the design is termed the post-monsoon design.

#### Design of sampling locations based on point pollution load

The major point sources of pollution in the study area are six open drains carrying municipal and industrial wastewater. The sampling locations downstream of these sources are considered potential sampling sites, i.e., P<sub>5</sub>, P<sub>7</sub>, P<sub>9</sub>, P<sub>11</sub>, P<sub>13</sub>, and P<sub>15</sub>, as shown in Table 2. These six sampling sites were analyzed for the design of the water quality sampling locations. The objective of the design process was to select the optimum number and locations of sampling points from the six potential sampling sites.

First, the optimum number of sampling locations was selected. For the present study, the budget was unknown, and effluent disposal standards were not defined. Hence, the optimum number of sampling

locations was determined on the basis of pollution loadings. The number of sampling locations for each hierarchy level was estimated using Eq. 3. For the point pollution sources, the analysis was carried out up to the sixth hierarchy level because there were six potential sampling locations.  $M_h$  values (Eqs. 1 and 2) were calculated for  $h=1$  to 6 for both the monsoon and non-monsoon seasons. The estimated  $M_h$  values and the number of sampling locations are shown in Table 5. Do et al. (2011) proposed that the threshold value for selection of the optimum hierarchy level,  $M_h$ , should be around one fifth of the first hierarchy,  $M_1$  (i.e.,  $M_h=0.2 M_1$ ). Hence, in the present analysis, hierarchy level  $h_3$  was selected as the design hierarchy level for both seasons. The significant locations and number of sampling sites were selected based on BOD loadings at the six potential sampling sites, and the  $M_h$  values were estimated up to the third hierarchy level. The significant sampling locations at hierarchy  $h$  are the locations with the smallest differences between the estimated  $M_h$  values and the measured BOD loadings. The optimum

**Table 5** Estimation of  $M_h$  values

Hierarchy levels (h)	No. of sampling stations (n)	$M_h$ values	Pollution loadings					
			BOD		Nitrate		Phosphate	
			Monsoon $N_o=744,864.72$ kg/day	Non-monsoon $N_o=675,346.91$ kg/day	Monsoon $N_o=1,899.77$ kg/day	Non-monsoon $N_o=476.58$ kg/day	Monsoon $N_o=949.70$ kg/day	Non-monsoon $N_o=271.24$ kg/day
1	1	$M_1$	372,432.90	337,674.00	950.38	238.79	475.35	136.12
2	3	$M_2$	186,216.90	168,837.50	475.69	119.90	238.18	68.56
3	7	$M_3$	93,108.97	84,419.24	238.35	60.45	119.59	34.78
4	15	$M_4$	46,554.98	42,210.12	119.67	30.72	60.29	17.89
5	31	$M_5$	23,277.99	21,105.56	60.34	15.86	30.65	9.44
6	63	$M_6$	NA	NA	30.67	8.43	15.82	5.22
7	127	$M_7$	NA	NA	15.83	4.72	8.41	3.11
8	255	$M_8$	NA	NA	8.42	2.86	4.71	2.06
9	511	$M_9$	NA	NA	4.71	1.93	2.85	1.53
10	1,023	$M_{10}$	NA	NA	2.85	1.46	1.93	1.26
11	2,047	$M_{11}$	NA	NA	1.93	1.23	1.46	1.13
12	4,095	$M_{12}$	NA	NA	1.46	1.12	1.23	1.07
13	8,191	$M_{13}$	NA	NA	1.23	1.06	1.12	1.03
14	16,383	$M_{14}$	NA	NA	1.12	1.03	1.06	1.02
15	32,767	$M_{15}$	NA	NA	1.06	1.01	1.03	1.01
16	65,535	$M_{16}$	NA	NA	1.03	1.01	1.01	1.00

NA not applicable

number and locations of sampling sites and their priority for both seasons are shown in Table 6.

*Monsoonal design of sampling locations based on point pollution loads*

The monsoonal design of the sampling locations is illustrated in Fig. 6a. Four significant sampling locations in the monsoon period are P<sub>7</sub>, P<sub>11</sub>, P<sub>13</sub>, and P<sub>15</sub>. Here, the term “significant” refers to the sampling locations at hierarchy *h* with the smallest differences between the estimated *M<sub>h</sub>* values and the measured BOD loadings. The sampling location P<sub>7</sub> is at the first hierarchy level, P<sub>13</sub> is at the second, and P<sub>11</sub> and P<sub>15</sub> are at the third. The significant sampling locations are selected based on the pollution loadings. In the present study, all point pollution sources are of municipal origin; hence, their pollution potential is measured in terms of BOD (organic) loadings. These sources are in the form of open drains, the hydraulic behavior of which is different from that of closed conduits. The open drains carry both surface runoff and wastewater. Hence, the pollution load is considerably affected by dilution. Untreated wastewater is directly discharged into the river by the point sources of pollution, and as a result, the pollution load is quite high. The dilution effect is significant during the monsoon season, which results in higher discharge and lower BOD concentration at the point sources, as shown in Table 3.

The point source of pollution just upstream of P<sub>7</sub> consists of both municipal and industrial wastewater discharges, and the sub-watershed of P<sub>7</sub> has a settlement area of 42.67 km<sup>2</sup> (19.50 % of the total sub-watershed area; Fig. 3). The BOD loadings from both the point

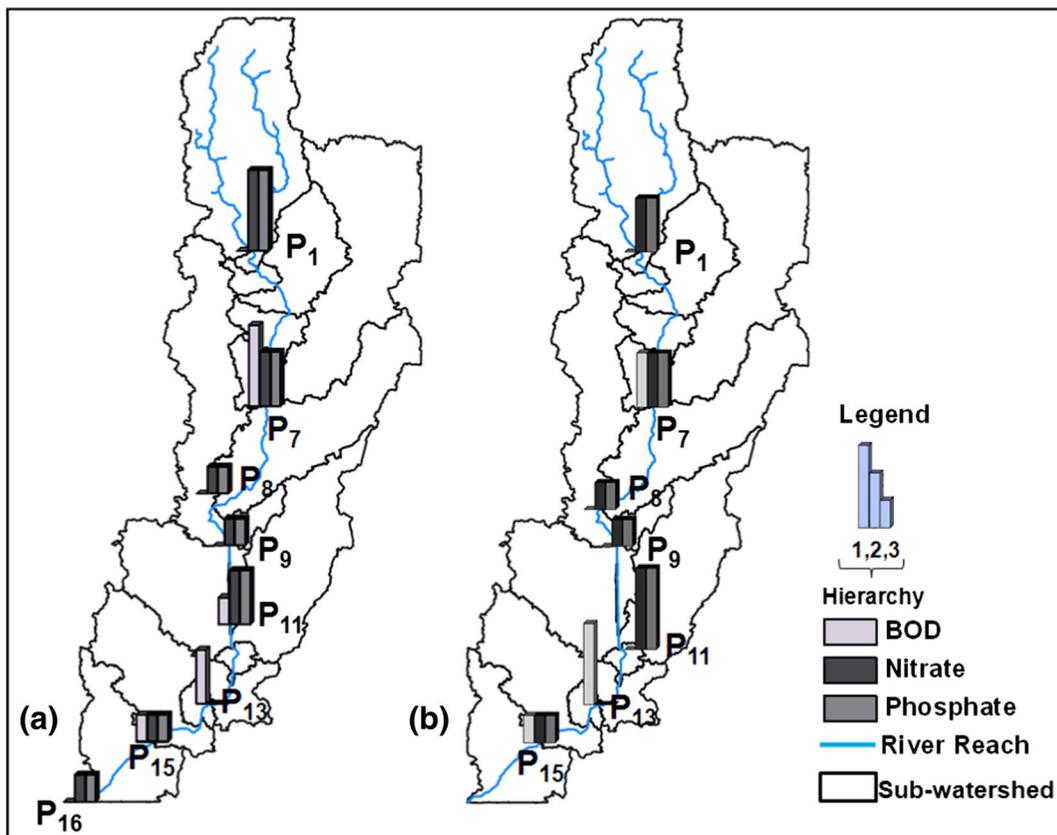
source and overland flow from the settlement area flow through the open drain, and therefore, P<sub>7</sub> has a significant BOD concentration of 1,124.94 mg/l (Table 3). The runoff during the monsoon season through the open drain confluence at P<sub>7</sub> is 3.13 m<sup>3</sup>/s because of the large area of the contributing sub-watershed, 218.81 km<sup>2</sup> (Fig. 3), which results in a considerable BOD discharge of 303,830.65 kg/day (Table 3), so P<sub>7</sub> is placed at the first hierarchy level (Fig. 6a and Table 6). Similarly, the point source upstream of P<sub>13</sub> carries both industrial and municipal wastewater, with the effluent load from industrial wastewater predominating. The sub-watershed of P<sub>13</sub> has a settlement area of 7.99 km<sup>2</sup> (22.21 % of the total sub-watershed area; Fig. 3). Therefore, the effluent discharge just upstream of P<sub>13</sub> has a considerable BOD concentration of 1,378.24 mg/l (Table 3). The contributing sub-watershed area of P<sub>13</sub> is 35.96 km<sup>2</sup> (Fig. 3), which is less than the area of the sub-watershed of P<sub>7</sub>. Consequently, the flow through the open drain at P<sub>13</sub> during the monsoon season is 2.17 m<sup>3</sup>/s (Table 3), which is less than the flow at P<sub>7</sub>. This results in a lower BOD load of 258,498.85 kg/day and places P<sub>13</sub> at the second hierarchy level (Fig. 6a and Table 6).

The sampling location P<sub>11</sub> has the largest contributing sub-watershed area of 261.33 km<sup>2</sup>, and the runoff and subsequent flow during the monsoon season through the open drain at P<sub>11</sub> is found to be highest, with a magnitude of 4.57 m<sup>3</sup>/s (Table 3). Although the sub-watershed has a larger settlement area of 73.36 km<sup>2</sup> (approximately 28.07 % of the total sub-watershed area, as shown in Fig. 3), the effluent BOD loading is approximately 62,713.91 kg/day because of the higher dilution of the polluting loads. As a result, the monitoring station P<sub>11</sub> is placed at the third hierarchy level (Fig. 6a). Again,

**Table 6** Design of sampling locations for point source of pollution based on BOD loading

Sampling station	River reach	Monsoon period			Non-monsoon period		
		BOD loadings	<i>M<sub>h</sub></i> values	Hierarchy level	BOD loadings	<i>M<sub>h</sub></i> values	Hierarchy level
P <sub>5</sub>	R <sub>4</sub>	42,296.34	NA	NA	37,846.53	NA	NA
P <sub>7</sub>	R <sub>6</sub>	303,830.65	372,432.90	1	206,290.02	168,837.5	2
P <sub>9</sub>	R <sub>8</sub>	25,968.48	NA	NA	30,095.90	NA	NA
P <sub>11</sub>	R <sub>10</sub>	62,713.91	93,108.97	3	39,666.18	NA	NA
P <sub>13</sub>	R <sub>12</sub>	258,498.85	186,216.9	2	302,959.78	337,674	1
P <sub>15</sub>	R <sub>14</sub>	51,556.49	93,108.97	3	58,488.51	84,419.24	3

NA not applicable



**Fig. 6** Effect of seasonal variation on design of sampling locations. **a** Monsoon season. **b** Non-monsoon season

the monitoring station at  $P_{15}$  is placed at the third hierarchy level due to its lower BOD load of 51,556.49 kg/day (Tables 3 and 6).

The rest of the monitoring stations that receive BOD from point sources, i.e.,  $P_5$  and  $P_9$ , have lower levels of BOD load (approximately 42,296.34 and 25,968.48 kg/day for  $P_5$  and  $P_9$ , respectively) and do not fulfill the optimality criterion for being in the monitoring network during the monsoon season. These two stations are found to be redundant and may be eliminated from the network for monitoring during the monsoon season.

#### *Non-monsoonal design of sampling locations based on point pollution loads*

The design of the sampling locations for the non-monsoon season is illustrated in Fig. 6b. Three significant sampling locations during the non-monsoon season are  $P_{13}$ ,  $P_7$ , and  $P_{15}$ , which are at the first, second, and third hierarchy levels, respectively. The significant sampling locations are selected based on BOD loadings. During the non-monsoon season, the dilution factor (the effect of surface runoff) is

insignificant, and the pollution loading depends upon the discharge and BOD concentration of each point source of pollution. The point source of pollution just upstream of  $P_7$  carries municipal and industrial wastewater. Due to the combined effect of the municipal and industrial wastewater, the effluent BOD load (206,290.02 kg/day) is significant, so  $P_7$  is placed at the second hierarchy level (Table 6). The point source just upstream of  $P_{13}$  also carries municipal wastewater and a significant amount of industrial wastewater. The resulting effluent BOD load is 302,959.78 kg/day which is the highest BOD discharge in the river watershed, so  $P_{13}$  is placed at the first hierarchy level (Table 6).

The point source upstream of sampling location  $P_{15}$  carries municipal wastewater and has a considerable effluent BOD load of 58,488.51 kg/day (Table 3). Hence,  $P_{15}$  is placed at the third hierarchy (Table 6 and Fig. 6b). The point sources upstream of  $P_5$ ,  $P_9$ , and  $P_{11}$  carry municipal wastewater and make up approximately 40 % of the unsewered settlement area (field study) of their respective sub-watersheds. The point source effluent BOD loadings upstream of  $P_5$ ,  $P_9$ , and  $P_{11}$  are 37,846.53, 30,095.90, and 39,666.18 kg/day, respectively

(Table 3), which are insignificant in comparison to the loadings derived at hierarchy level  $M_1$  (Table 5). Thus, sampling locations  $P_5$ ,  $P_9$ , and  $P_{11}$  failed to meet the optimality criterion and are removed from the design of sampling locations for the non-monsoonal season (Fig. 6b).

#### *Effect of seasonal variation on the design of sampling locations for point sources of pollution*

The BOD loads of point sources were taken into account in the design of the sampling locations. As mentioned in “Monsoonal design of sampling locations based on point pollution loads” section, the point sources are actually in the form of open drains. Therefore, surface runoff during the monsoon season has a significant impact on the effluent BOD load of the point pollution sources. The BOD loads of the point sources depend on the surface runoff and BOD concentration of the effluent, and the runoff is directly proportional to the area of the sub-watershed. Therefore, the seasonal effect brings two major types of change to a network: changes in the priority order of the sampling locations and a reduction in the number of sampling locations.

During the monsoon season, sampling location  $P_7$  is placed at the first hierarchy level and  $P_{13}$  is placed at the second hierarchy level, while during the non-monsoon season,  $P_{13}$  is placed at the first hierarchy level and  $P_7$  is placed at the second hierarchy level (Fig. 6). The sub-watershed area contributing to sampling location  $P_7$  (218.81 km<sup>2</sup>) is larger than that of  $P_{13}$  (35.96 km<sup>2</sup>). Therefore, the resulting discharge of the point source just upstream of  $P_7$  (3.13 m<sup>3</sup>/s) is greater than that of  $P_{13}$  (2.17 m<sup>3</sup>/s; Table 3). Hence, during the monsoon season, the effluent BOD loading upstream of  $P_7$  (303,830.65 kg/day) is greater than that upstream of  $P_{13}$  (258,498.85), and  $P_7$  is placed at the first hierarchy level, while  $P_{13}$  is placed at the second hierarchy level (Table 6).

As mentioned in “Non-monsoonal design of sampling locations based on point pollution loads” section, during the non-monsoon season, the discharge and effluent concentrations of individual point sources are significant, in contrast to surface runoff during the monsoon season. The point sources just upstream of  $P_7$  and  $P_{13}$  carry both industrial and municipal wastewater, and both of them discharge considerable BOD loads into the river. However, the effluent load from industrial wastewater is very predominant at  $P_{13}$ . Therefore, during the non-monsoon season, the point source upstream of  $P_{13}$  has a higher BOD load (302,959.78 kg/day) than that

upstream of  $P_7$  (206,290.02 kg/day; Table 3). Hence,  $P_{13}$  is positioned at the first hierarchy level, while  $P_7$  is positioned at the second hierarchy level.

The second major impact of seasonal variation is a reduction in the number of sampling sites. The sampling location  $P_{11}$  is significant during the monsoon season but insignificant during the non-monsoon season. During the monsoon season, the pollution loading is at a maximum due to the effect of surface runoff. During the non-monsoon season, the discharge and effluent characteristics of the point pollution sources determine the pollution loadings. The sub-watershed area of  $P_{11}$  is the largest (261.33 km<sup>2</sup>; Fig. 3), which results in the maximum discharge (4.57 m<sup>3</sup>/s; Table 3) by a point source upstream of  $P_{11}$ . Therefore,  $P_{11}$  is at the third hierarchy level during monsoon season and is insignificant during the non-monsoon season. The optimum number of sampling locations is reduced from four to three. Hence, in the present analysis, a new and cost-effective network design approach for selection of sampling locations is proposed. The effect of seasonal variation was taken into account and resulted in a reduction in the number of sampling locations during the non-monsoon season, which improves the cost-effectiveness of the design process.

The present analysis shows the cause-and-effect relationship between the results obtained and anthropogenic activities in the study area. The pollution loading is the optimality criterion, so the first priority of sampling is to select the sampling locations downstream of major pollution sources. The effect of surface runoff was successfully accounted in the design of the sampling locations. Municipal and industrial wastewater is discharged into the river without treatment. Rapid industrialization and urbanization in the study area have had an adverse impact on the river ecosystem and have resulted in considerable degradation of the river water quality. Therefore, the results obtained reflect the actual site condition, which makes the present analysis more realistic.

#### Design of sampling locations based on diffuse pollution loads

The study area consists of 16 sub-watersheds that are considered principal diffuse sources of pollution. The outlets of these 16 sub-watersheds are the potential sampling sites, i.e.,  $P_1$  to  $P_{16}$ . These 16 sampling sites were analyzed to select the optimum number and locations of water quality sampling locations.

First, the optimum number of sampling locations was selected by following the procedure described in “Design of sampling locations based on point pollution load” section.  $M_h$  values were calculated using Eqs. 1 and 2 for  $h=1$  to 16 and for both the monsoon and non-monsoon seasons. For diffuse pollution sources, the analysis was carried out up to the 16th hierarchy level because the potential sampling locations are 16. The estimated  $M_h$  values and the number of sampling locations are shown in Table 5. According to the optimality criterion proposed by Do et al. (2011), the design hierarchy level is  $h_3$ . The optimum number and locations of sampling sites and their priority order for both seasons are shown in Table 7.

*Monsoonal design of sampling locations based on diffuse pollution load*

The design of the sampling locations for the monsoon season is illustrated in Fig. 6a. The significant sampling locations for diffuse pollution loadings during the monsoon period are  $P_1, P_7, P_8, P_9, P_{11}, P_{15},$  and  $P_{16}$ . The sampling location  $P_1$  is at the first hierarchy level,  $P_7$  and

$P_{11}$  are at the second, and  $P_8, P_9, P_{15},$  and  $P_{16}$  are at the third (Fig. 6a). The significant sampling locations are selected based on nutrient pollution loadings, given that 80.11 % of the watershed area is used for agricultural purposes (Fig. 3). Nitrate and phosphate are the principal nutrients required for growth of plants and are the key ingredients in fertilizers. These fertilizers are used excessively to increase crop yields, resulting in nutrient-rich agricultural runoff. Hence, during the monsoon season, huge amounts of nutrients are carried out to the river by agricultural runoff. The pollution potential of diffuse sources is measured in terms of nitrate and phosphate (nutrient) loadings. The surface runoff is directly proportional to the watershed area. The nutrient concentration of the surface runoff is affected by the use of the watershed area for agricultural activities. Therefore, during the monsoon season, the surface runoff and agricultural area of a given sub-watershed are the key factors governing the nutrient loading and priority order of sampling locations. The presence of a point pollution source in a sub-watershed also has a significant impact on nutrient loadings and the design of sampling locations.

**Table 7** Design of sampling locations for diffused source of pollution based on Nitrate and Phosphate loading

Sampling station	River reach	Nitrate				Phosphate			
		Monsoon period		Non-monsoon period		Monsoon period		Non-monsoon period	
		Nitrate loadings	Hierarchy level	Nitrate loadings	Hierarchy level	Phosphate loadings	Hierarchy level	Phosphate loadings	Hierarchy level
$P_1$	$R_1$	332.42	1	72.41	2	164.04	1	39.14	2
$P_2$	$R_2$	6.31	NA	1.35	NA	3.11	NA	0.73	NA
$P_3$	$R_3$	18.31	NA	3.90	NA	9.02	NA	2.09	NA
$P_4$	$R_4$	107.30	NA	22.88	NA	52.85	NA	12.27	NA
$P_5$	$R_5$	33.50	NA	9.06	NA	16.87	NA	5.28	NA
$P_6$	$R_6$	32.37	NA	7.13	NA	15.99	NA	3.88	NA
$P_7$	$R_7$	277.97	2	71.61	2	139.31	2	41.10	2
$P_8$	$R_8$	219.69	3	48.72	3	108.58	3	26.52	3
$P_9$	$R_9$	213.12	3	58.04	3	107.43	3	33.89	3
$P_{10}$	$R_{10}$	18.06	NA	4.19	NA	8.96	NA	2.32	NA
$P_{11}$	$R_{11}$	308.84	2	91.00	1	157.03	2	54.34	1
$P_{12}$	$R_{12}$	13.10	NA	2.94	NA	6.48	NA	1.61	NA
$P_{13}$	$R_{13}$	44.85	NA	12.05	NA	22.59	NA	7.02	NA
$P_{14}$	$R_{14}$	15.95	NA	3.57	NA	7.89	NA	1.95	NA
$P_{15}$	$R_{15}$	128.68	3	38.11	3	65.46	3	22.78	3
$P_{16}$	$R_{16}$	129.28	3	29.60	NA	64.08	3	16.32	NA

NA not applicable

The area of the sub-watershed contributing to sampling location  $P_1$  is 239.59 km<sup>2</sup>, of which 214.83 km<sup>2</sup> (i.e., 89.67 %) is used for farming (Fig. 3). The nutrient pollution load at sampling location  $P_1$  is 332.42 kg of nitrate/day and 164.04 kg of phosphate/day (Table 4). Agricultural land use of the area of the sub-watershed contributing to  $P_1$  is significant. Hence, during the monsoon season, sampling location  $P_1$  discharges the maximum nutrient load in the basin and is at the first hierarchy level for both types of nutrient loadings (Fig. 6a and Table 7).

The sub-watershed area contributing to sampling location  $P_7$  is 218.81 km<sup>2</sup>, of which 171.59 km<sup>2</sup> (78.42 %) is used for farming (Fig. 3). The point pollution source just upstream of  $P_7$  discharges both municipal and industrial wastewater. Thus, due to the combined effect of point and diffuse pollution sources, the nutrient load at sampling location  $P_7$  is 277.97 kg of nitrate/day and 139.31 kg of phosphate/day (Table 4). Therefore, sampling location  $P_7$  is significant and is located at the second hierarchy level for both nutrients during the monsoon season (Fig. 6a and Table 7). Similarly, sampling location  $P_{11}$  has a considerable agricultural area, i.e., 182.26 km<sup>2</sup> (69.74 %) in a sub-watershed of 261.33 km<sup>2</sup> (Fig. 3). The point pollution source just upstream of  $P_{11}$  carries municipal wastewater, which results in a combined effect of point and diffuse pollution sources. The nutrient pollution loads at  $P_{11}$  are 308.84 kg of nitrate/day and 157.03 kg of phosphate/day (Table 4). Hence,  $P_{11}$  is placed at the second hierarchy level for both nutrient loadings during the monsoon period (Fig. 6a and Table 7).

The sampling locations  $P_8$ ,  $P_9$ ,  $P_{15}$ , and  $P_{16}$  are placed at the third hierarchy level (Fig. 6a and Table 7) as they discharge significant nutrient pollution loads for both nutrients. The sub-watershed area of sampling location  $P_8$  is 160.65 km<sup>2</sup>, of which 141.36 km<sup>2</sup> (87.99 %) is used for farming (Fig. 3). The nutrient loads at  $P_8$  are 219.69 kg of nitrate/day and 108.58 kg of phosphate/day (Table 4). The sub-watershed of  $P_9$  has an area of 173.11 km<sup>2</sup>, of which 129.26 km<sup>2</sup> (74.67 %) is used for farming (Fig. 3). The point pollution source carrying municipal wastewater discharges just upstream of  $P_9$ . Hence, a combined effect of both point and diffuse sources occurs at  $P_9$ , and the resulting nutrient loads are 213.12 kg of nitrate/day and 107.43 kg of phosphate/day (Table 4). The sub-watershed area contributing to sampling location  $P_{15}$  is 109.54 km<sup>2</sup>, of which 75.81 km<sup>2</sup> (i.e., 69.21 %) is used for farming (Fig. 3). The point source upstream of  $P_{15}$  discharges municipal wastewater. Hence,  $P_{15}$  also has the joint

effect of both point and diffuse pollution loads, and it is located at the third hierarchy level (Table 7). The nutrient loads at  $P_{15}$  are 128.68 kg of nitrate/day and 65.46 kg of phosphate/day (Table 4). The sampling location  $P_{16}$  is placed at the third hierarchy level due to the considerable nutrient loads at that point, i.e., 129.28 kg of nitrate/day and 64.08 kg of phosphate/day, during the monsoon season (Table 7). The sub-watershed area contributing to sampling location  $P_{16}$  is 96.11 km<sup>2</sup>, of which 82.49 km<sup>2</sup> (85.83 %) is used for agricultural purpose (Fig. 3). Therefore, sampling location  $P_{16}$  is significant because of the nutrient-rich agricultural runoff from its contributing sub-watershed. Sampling locations  $P_2$ ,  $P_3$ ,  $P_4$ ,  $P_5$ ,  $P_6$ ,  $P_{10}$ ,  $P_{12}$ , and  $P_{13}$  discharge nutrient loads that are insignificant compared to  $M_1$  (Tables 5 and 7). These sampling locations do not meet the optimality criterion defined in “Design of sampling locations based on diffuse pollution load” section. Hence, monitoring at these locations is not necessary during the monsoon season.

#### *Non-monsoonal design of sampling locations based on diffuse pollution load*

The design of sampling locations for the non-monsoon season is illustrated in Fig. 6b. The significant sampling locations during the non-monsoon period are  $P_1$ ,  $P_7$ ,  $P_8$ ,  $P_9$ ,  $P_{11}$ , and  $P_{15}$ . The sampling location  $P_{11}$  is at the first hierarchy level,  $P_1$  and  $P_7$  are at the second hierarchy level, and  $P_8$ ,  $P_9$ , and  $P_{15}$  are at the third hierarchy level for both nitrate and phosphate loadings. The significant sampling locations are selected based on the nutrient (i.e., nitrate and phosphate) loadings. In the non-monsoon season, surface runoff is insignificant, and the nutrients are mainly transported to the river by sub-surface flow. The point sources are also key sources of nutrient pollution. The sub-watershed contributing to  $P_{11}$  has an agricultural area of 182.26 km<sup>2</sup> (69.74 % of the total; Fig. 3), and the point pollution source carries municipal wastewater. Therefore, considerable amounts of nutrients, i.e., 91.00 kg of nitrate/day and 54.34 kg of phosphate/day, are discharged at  $P_{11}$ , and it is placed at the first hierarchy level (Fig. 6b and Table 7).

The sub-watershed contributing to sampling location  $P_1$  has an agricultural area of 214.83 km<sup>2</sup> (Fig. 3). The nutrient loads at  $P_1$  during the non-monsoon season are 72.41 kg of nitrate/day and 39.14 kg of phosphate/day (Table 4). The significant amounts of nutrients are

carried out to  $P_1$  by sub-surface flow, and therefore, it is located at the second hierarchy level during the non-monsoon season (Fig. 6b and Table 7).

The sub-watershed contributing to sampling location  $P_7$  has an agricultural area of 171.59 km<sup>2</sup> (Fig. 3). The point pollution source just upstream of  $P_7$  discharges both municipal and industrial wastewater. Thus, due to the joint effect of point source pollution and nutrients transported by sub-surface flow, the nutrient loadings at sampling location  $P_7$  are 71.61 kg of nitrates/day and 41.10 kg of phosphate/day (Table 4). Therefore, sampling location  $P_7$  is significant and is placed at the second hierarchy level for both nutrients during the non-monsoon season (Fig. 6b and Table 7).

Sampling locations  $P_8$ ,  $P_9$ , and  $P_{15}$  are at the third hierarchy level (Fig. 6b and Table 7) as they discharge significant amounts of both nutrients. The sub-watershed contributing to sampling location  $P_8$  has an agricultural area of 141.36 km<sup>2</sup> (Fig. 3). Hence, the sub-surface flow carries considerable amounts of nutrients during the non-monsoon period. The nutrient loads at  $P_8$  are 48.72 kg of nitrate/day and 26.52 kg of phosphate/day (Table 4). The sub-watershed of  $P_9$  has an agricultural area of 129.26 km<sup>2</sup> (Fig. 3). The point pollution source discharges municipal wastewater just upstream of  $P_9$ . Hence, a combined effect of both point source pollution and nutrient transport by sub-surface flow occurs at  $P_9$ , and the resulting nutrient loads are 58.04 kg of nitrate/day and 33.69 kg of phosphate/day (Table 4). The sub-watershed area contributing to sampling locations  $P_{15}$  has an agricultural area of 75.81 km<sup>2</sup> (Fig. 3). The point source just upstream of  $P_{15}$  discharges municipal wastewater. Hence, a considerable nutrient load occurs at  $P_{15}$  due to the combined effect of nutrient transport by the point source and sub-surface flow. The nutrient loads at  $P_{15}$  are 38.11 kg of nitrate/day and 22.78 kg of phosphate/day, so  $P_{15}$  is located at the third hierarchy level (Table 7). The sampling locations  $P_2$ ,  $P_3$ ,  $P_4$ ,  $P_5$ ,  $P_6$ ,  $P_{10}$ ,  $P_{12}$ ,  $P_{13}$ , and  $P_{16}$  discharge the nutrient loads, which are insignificant compared to  $M_1$ . Hence, these sampling locations do not meet the optimality criterion and do not need to be monitored during the non-monsoonal season (Fig. 6b).

#### *Effect of seasonal variation on design of sampling locations for diffuse sources of pollution*

The nutrient load (i.e., nitrate and phosphate) of diffuse pollution sources was accounted for in the design of the

sampling locations. The nutrient-rich agricultural runoff and point sources carrying municipal and industrial wastewater are the major sources of river pollution due to nutrients. Hence, as mentioned in “[Monsoonal design of sampling locations based on diffuse pollution loads](#)” section, in the monsoon season, the nutrients are transported to the river by nutrient-rich surface runoff from agricultural areas. In the non-monsoon season, the surface runoff is insignificant as there is no rainfall. However, some amount of nutrients goes underground through deep percolation and is finally transported to the river by sub-surface flow. The point sources play an important role in nutrient transport in both seasons. The seasonal effect brings two major changes to the network: a change in the priority ordering of the sampling locations and a reduction in the number of sampling locations.

In the monsoon season, sampling location  $P_1$  is at the first hierarchy level and  $P_{11}$  is at the second hierarchy level, while in the non-monsoon season,  $P_{11}$  is at the first hierarchy level and  $P_1$  is at the second hierarchy level (Fig. 6 and Table 7). The sub-watershed contributing to sampling location  $P_1$  has an agricultural area of 214.83 km<sup>2</sup>, which is greater than the agricultural area of the sub-watershed contributing to  $P_{11}$ , i.e., 182.26 km<sup>2</sup> (Fig. 3). In the monsoon season, the resulting surface runoff at  $P_1$  has greater nutrient potential than that at  $P_{11}$ . Therefore, in the monsoon season, the nutrient loads at  $P_1$  (332.42 kg of nitrate/day and 164.04 kg of phosphate/day) are higher than those at  $P_{11}$  (308.84 kg of nitrate/day and 157.03 kg of phosphate/day), and  $P_1$  is placed at the first hierarchy level, while  $P_{11}$  is placed at the second hierarchy level (Fig. 6 and Table 7).

As mentioned in “[Non-monsoonal design of sampling locations based on diffuse pollution loads](#)” section, in the non-monsoon season, nutrients are transported by sub-surface flow instead of surface runoff. Sampling stations  $P_1$  and  $P_{11}$  receive nutrient loads through sub-surface flow. However, the point pollution source just upstream of  $P_{11}$  discharges municipal wastewater. Hence, sampling location  $P_{11}$  is affected by both point and diffuse pollution sources and has higher nutrient loadings (91.00 kg of nitrate/day and 54.34 kg of phosphate/day) than  $P_1$  (72.41 kg of nitrate/day and 39.14 kg of phosphate/day; Table 4). Hence, in the non-monsoon season,  $P_{11}$  is placed at the first hierarchy level, while  $P_1$  is placed at the second hierarchy (Fig. 6a and Table 7).

The second major impact of seasonal variation is reduction in the number of sampling sites. Sampling

location  $P_{16}$  is significant in the monsoon season but insignificant in the non-monsoon season. In the monsoon season, the nutrient loading is highest due to the effect of surface runoff, while in non-monsoon season, the surface flow is insignificant and the nutrients are transported by sub-surface flow. The nutrient loads at  $P_{16}$  are significant (129.28 kg of nitrate/day and 64.08 kg of phosphate/day) in monsoon due to effect of surface runoff, while in the non-monsoon season, the nutrient loads are insignificant (29.60 kg of nitrate/day and 16.32 kg of phosphate/day) in comparison to  $M_1$  (Tables 5 and 7). Hence,  $P_{16}$  is located at the third hierarchy level during the monsoon season, and it is insignificant during the non-monsoon season. The optimum number of sampling locations is reduced from seven to six. Hence, the new network design approach for sampling locations proposed in the present study improves the cost-effectiveness of the design by taking the effect of seasonal variation into account.

As mentioned in “Effect of seasonal variation on the design of sampling locations for point sources of pollution” section, the present analysis shows a cause-and-effect relationship between the results obtained and the anthropogenic activities in the study area. The pollution loading is the optimality criterion, so the sampling locations discharging considerable nutrient loads have the highest priority for sampling. The excessive use of chemical fertilizers, improper agricultural practices, and untreated municipal and industrial wastewater disposal into the river have adverse effects on the river ecosystem and result in continuous degradation of river water quality. The effect of surface runoff is taken into account in the design of the sampling locations. Therefore, the results obtained reflect the actual site condition, which makes the present analysis more realistic.

## Conclusions

The present study investigates the effect of seasonal variations in point and diffuse pollution loadings on river water quality and quantifies their impact on the optimal design of sampling locations. The approach is obviously cost-effective, as it suggests the optimal number of monitoring stations for the pre-monsoon and post-monsoon seasons. An effect-based monitoring network is delineated directly through estimation of pollution loadings, and the optimal sampling locations are selected based on the optimal hierarchy level of a binary tree structure from

graph theory. The study shows that the number and locations (hierarchy/priority) of sampling sites may vary seasonally, which justifies pre- and post-monsoonal change or redesign of the monitoring stations. As discussed in the “Introduction” section, numerous frameworks/guidelines proposed by different countries and mathematical approaches proposed by various researchers have been applied for design of sampling locations. Most of the framework/guidelines are simple, not case study specific, and easy to locate the sampling locations by accounting the effect of pollution sources and accessibility of sampling sites, whereas the mathematical approaches are comparatively complicated, case study specific, and are capable of optimally allocate sampling sites. The present study realizes importance of both the approaches, and a first attempt has been made to combine the existing guidelines and the modified Sanders approach for design of water quality sampling locations. The proposed approach was applied to the 1,475.50-km<sup>2</sup> catchment area of the Kali River in northern India, which channelizes a 125-km stretch of polluted river.

In this study, the impacts of agriculture and urban development on river water quality are quantified in a GIS framework. The optimal sampling locations for BOD and nutrient loadings for the monsoon season are four and seven, respectively, while for the non-monsoon season, they are three and six, respectively. Excessive use of fertilizers within the river watershed, 80.11 % of the area of which is used for agricultural purposes, results in massive amount of nutrients being discharged into the river through surface and sub-surface flows. Untreated domestic and industrial wastewater is discharged into the river through open drains. There is a need for wastewater treatment before disposal in the river and a need to create awareness among farmers concerning the proper use of fertilizers. The results obtained show that the optimal number of sampling locations is reduced from four to three and from seven to six in the non-monsoon season for BOD and nutrient loadings, respectively. The diffuse pollution load is a function of the surface runoff, which is quite high, especially during the monsoon season. Hence, to capture the effect of surface runoff, more sampling locations are required in the monsoon season than in the non-monsoon season. The estimation of seasonal pollution loads and determination of the optimal number of monitoring stations is definitely cost-effective, as the process of water quality sampling is expensive. This approach might be more effective for a larger river with

significant spatiotemporal variability in water quality. The effect of seasonal variability of pollution loads on design of sampling sites is successfully demonstrated in the present study considering both point and diffuse sources of pollution, which is applicable to any part of the world with prevalent monsoon season.

Further, the major shortcoming of the proposed approach is the extensive watershed data requirement, which has been mentioned in the “Introduction” section. The data sets include hydraulic and hydrological characteristics of river basins, land use practices, agricultural cropping patterns, use of fertilizers, and relative locations of pollution sources and monitoring stations. Again, the proposed approach needs expertise in various disciplines, viz. remote sensing and GIS applications, water quality testing, instrumentation, watershed hydrology, etc. Also, implementation of modified Sanders approach for a large river basin with more number of sampling locations might also be too complicated. Hence, a computational tool/software might facilitate a systematic and easier implementation of the proposed framework. The other design components of monitoring network, e.g., sampling frequency and parameters, may further be considered to evaluate the impact of seasonal variability of river water quality.

### Appendix 1

#### Estimation of discharge from pollution sources

Along the entire stretch of the Kali River, a total of six drains empty into the river, carrying municipal and industrial wastewater. These drains are considered major point sources of pollution. The river stretch is divided into 15 reaches ( $R_i$ ) based on the relative locations of 16 sampling sites (see Table 2 and Fig. 2). These 15 reaches are classified into two categories based on the locations of the point sources of pollution. The first category of the reaches ( $\tilde{R}_i^D$ ), which do not contain point sources of pollution (along the Kali River, a total of nine reaches fall into this category, i.e.,  $R_1, R_2, R_3, R_5, R_7, R_9, R_{11}, R_{13}$ , and  $R_{15}$ ). The second category of the reaches ( $R_i^D$ ) contains point sources of pollution (along the Kali River, a total of six reaches fall into this category, i.e.,  $R_4, R_6, R_8, R_{10}, R_{12}$ , and  $R_{14}$ ). Therefore, a river reach ( $R_i$ ) must fall into one of these two categories,  $\tilde{R}_i^D$  or  $R_i^D$ . The ground-water table in the study area was found during the field

study to be quite high (approximately 5.49 to 6.10 m below the ground surface), which indicates the possibility of losses from the river and the addition of discharges to the river as subsurface flow and precipitation. It was also found that the discharges from point sources coming through highly sedimented natural open drains were discontinuous, resulting in difficulty in direct measurement. Apart from this, of the six-point sources constituting natural open drains, four were found to be inaccessible. Therefore, the discharges from the point sources were routed using a mass balance approach. In general, the inflows and outflows of the intermediate river reaches can be estimated using the following equation:

$$Q_j^{P_{i+1}} = Q_j^{P_i} \pm \Delta Q_j^{R_i} \tag{4}$$

where  $Q_j^{P_{i+1}}$  ( $m^3/s$ ) is the river discharge at sampling location  $P_{i+1}$ ;  $i$  is the sampling location ( $i=1$  to  $15$ ), i.e., downstream of river reach  $R$  in month  $j$  (where  $j=1$  to  $12$ , i.e., January to December);  $Q_j^{P_i}$  ( $m^3/s$ ) is the river discharge at sampling point  $P_i$ , i.e., upstream of river reach  $R_i$  in the  $j$ th month;  $\Delta Q_j^{R_i}$  ( $m^3/s$ ) are the losses (which appear as negative quantities, e.g., evaporation from the river reach) and/or additions (which appear as positive quantities, e.g., discharges from aquifers adjacent to the river) of discharges to the river, estimated using the following generalized equation:

$$\Delta Q_j^{R_i} = \Delta Q_{D_j}^{R_i} \pm \Delta q_j^{R_i} L_{R_i} \tag{5}$$

where  $\Delta q_j^{R_i}$  ( $m^3/s/unit$  length of river reach) are additions and losses to the river discharge in the form of evaporation, sub-surface flow or seepage from the river, flow from diffuse sources, precipitation, etc.;  $\Delta Q_{D_j}^{R_i}$  ( $m^3/s$ ) is the addition of discharge by the point pollution source in reach  $R_i$ ; and  $L_{R_i}$  (in m) is the length of reach  $R_i$ . There may be two situations, depending upon the presence of point pollution sources in a river reach.

Situation 1: for river reaches in the first category (i.e., those with no point sources of pollution,  $\Delta Q_{D_j}^{R_i} = 0$ ), Eq. 4 becomes

$$Q_j^{P_{i+1}} = Q_j^{P_i} \pm \Delta q_j^{\tilde{R}_i^D} L_{R_i} \tag{6}$$

where  $\Delta q_j^{\tilde{R}_i^D}$  is the difference between  $Q_j^{P_{i+1}}$  and  $Q_j^{P_i}$  per unit length of reach  $R_i$ , which has no point source of pollution ( $\tilde{R}_i^D$ ).

Situation 2: for river reaches in the second category (i.e., those with point sources of pollution), Eq. 4 becomes

$$Q_j^{P_{i+1}} = Q_j^{P_i} \pm \Delta q_j^{R_i} L_{R_i} + \Delta Q_{D_j}^{R_i} \tag{7}$$

where  $R_i^D$  denotes river reach  $R_i$  with a point source of pollution. The discharge from point sources of pollution ( $\Delta Q_{D_j}^{R_i}$ ) is unknown and needs to be estimated. In this study, a stretch of 125 km of the Kali River was considered, and it was assumed that there was no significant variation in the weather and topographical conditions among the river reaches. Hence, the monthly average losses and additions per unit length of river ( $\Delta q_{avg_j}$ )

were estimated as the arithmetic means of  $\Delta q_j^{R_i}$  (Eq. 6) for nine river reaches with no point sources of pollution. The monthly losses and additions ( $\Delta q_j^{R_i}$ ) in Eq. 7 are replaced by  $\Delta q_{avg_j}$  to estimate the discharge from point sources of pollution. For the river reaches with point sources of pollution ( $R_4, R_6, R_8, R_{10}, R_{12},$  and  $R_{14}$ ), monthly point source discharge values ( $\Delta Q_{D_j}^{R_i}$ ) were estimated by substituting the river discharge value at each ends of each river reach  $R_i^D$ , i.e.,  $Q_j^{P_{i+1}}$  and  $Q_j^{P_i}$ , and the estimated value of  $\Delta q_{avg_j}$  in Eq. 7. The monsoonal and non-monsoonal discharges from each point source, i.e., the arithmetic averages of the discharge ( $\Delta Q_{D_j}^{R_i}$ ) values for the monsoon season (when  $j=7$  to  $10$ ) and non-monsoon seasons (when  $j=1$  to  $6$  and  $11$  to  $12$ ), respectively, is then calculated from the set of monthly values ( $\Delta Q_{D_j}^{R_i}$ ) and are tabulated in Table 3. As mentioned earlier, the sampling sites are located at the outlets of the sub-watersheds; therefore, the difference between the upstream and downstream river discharge measured at each sampling site (Eq. 4) yields the discharge contribution of each sub-watershed and corresponds to the diffuse source of pollution from that sub-watershed. The flows from all point and diffuse sources are routed according to the procedure discussed above and are used to estimate the pollution load through simulation.

### Simulation for estimating effluent characteristics

The modified BOD model (Jha et al. 2007) and refined diffuse source model (Jha et al. 2005) were applied to

simulate effluent characteristics from point (BOD) and diffuse (nitrate and phosphate) sources. Both the models were successfully calibrated and validated using the same data set used by Jha et al. (2005, 2007), which is the primary reason for choosing these models to simulate the effluent characteristics of the point and diffuse sources. The modified BOD model (Eq. 8) is based on a model for the first-order decay of organic matter proposed by Streeter and Phelps (1925), and the concept of Jolankai (1997) model was used in mass flux calculations, considering the diffuse sources of pollution and benthic oxygen demand. The hydraulic and water quality characteristics of the river water for  $R_i^D$  ( $R_4, R_6, R_8, R_{10}, R_{12},$  and  $R_{14}$ ) were used as inputs to the modified BOD model to simulate the BOD concentrations of the river water upstream and downstream of each point source. The effluent BOD (Table 3) of the point sources was then estimated by applying a BOD mass balance calculation for the point of mixing of effluent from each point source with river water, using the simulated values of river water BOD (at the point of mixing) and the estimated discharges of the point sources and river water (at the point of mixing) for a given month. The pollution loading was then estimated from the BOD concentration and the discharges of the point sources in each month. The monsoonal and non-monsoonal BOD loadings from each point source (i.e., the arithmetic averages of BOD loadings for the monsoon and non-monsoon seasons, respectively) were then calculated from the set of monthly values and are tabulated in Table 3.

$$C_j = C_{j_0} e^{-(K_1+K_2) t} + \frac{C_{j_d} \Delta q_j^{R_i} l (1 - e^{-(K_1+K_2) t})}{(K_1 + K_2) (Q_j^{P_i} + \Delta q_j^{R_i} l)} + \frac{B_j Q_j^{P_i} (1 - e^{-(K_1+K_2) t})}{(K_1 + K_2) (Q_j^{P_i} + \Delta q_j^{R_i} l)} \tag{8}$$

where  $C_j$  (mg/l) is the river water BOD at a distance  $l$  (in meters) from the origin of a river reach in a given month  $j$  ( $j=1$  to  $12$ ),  $C_{j_0}$  is the initial BOD concentration of the river water at the origin of the river reach,  $K_1$  is the reaction rate constant for biochemical decomposition of organic matter ( $\text{day}^{-1}$ ),  $K_2$  is the reaction rate constant for BOD removal by sedimentation ( $\text{day}^{-1}$ ),  $C_{j_d}$  is the lateral inflow BOD concentration to the stream (mg/l),  $\Delta q_j^{R_i}$  is

the lateral inflow rate ( $\text{m}^3/\text{s}/\text{unit length of river reach}$ ),  $Q_j^{P_i}$  is the flow rate at the river reach origin,  $B_j$  is the benthic oxygen demand ( $\text{mg}/\text{l}$ ), and  $t$  is the time of travel (day).

The refined diffuse source model (Jha et al. 2005), as expressed in Eq. 9, is used to estimate nitrate and phosphate loadings.

$$NL_j = \left( Q_j^{P_{i+1}} N_j^{P_{i+1}} - Q_j^{P_i} N_j^{P_i} e^{-kt} \right) \quad (9)$$

where  $NL_j$  is the diffuse/non-point source load for a given month  $j$ ;  $Q_j^{P_{i+1}}$  ( $\text{m}^3/\text{s}$ ) and  $N_j^{P_{i+1}}$  ( $\text{mg}/\text{l}$ ) are the discharge and nutrient concentration, respectively, at the downstream of river reach  $R_i$  for month  $j$ ;  $Q_j^{P_i}$  ( $\text{m}^3/\text{s}$ ) and  $N_j^{P_i}$  ( $\text{mg}/\text{l}$ ) are the discharge and nutrient concentration, respectively, at the upstream of river reach  $R_i$  for month  $j$ ;  $k$  is a decay constant ( $\text{day}^{-1}$ ); and  $t$  is the time of travel (day). The rates of decay for phosphate and nitrate were 0.22 and 0.10, respectively, at 20 °C (Jha et al. 2005). A temperature correction was applied for estimation of the decay constants. The measured discharges and effluent concentrations (nitrates and phosphates) at all river reaches were used as model inputs to estimate the monthly diffuse pollution loading. The seasonal nutrient loading was then estimated from the nutrient concentration and discharge of diffuse sources. The monsoonal and non-monsoonal nutrient loadings from each diffuse source were then calculated from the set of monthly values and are tabulated in Table 4.

**Acknowledgments** The authors sincerely thank the editor and the anonymous reviewers for their valuable suggestions. The authors also thank the Industrial Research and Consultancy Centre (IRCC), Indian Institute of Technology Bombay, Powai, Mumbai 400076, India for financially supporting the research work through a seed grant (Project code: 07IR053) available with the second author.

## References

Alameddine, I., Karmakar, S., Qian, S. S., Paerl, H. W., & Reckhow, K. H. (2013). Optimizing an estuarine water quality monitoring program through an entropy-based hierarchical spatiotemporal Bayesian framework. *Water Resources Research*, 49, 1–13.

Altin, A., Filiz, Z., & Iscen, C. F. (2009). Assessment of seasonal variations of surface water quality characteristics for Porsuk Stream. *Environmental Monitoring and Assessment*, 158, 51–65.

American Public Health Association (APHA). (2000). Standard methods for the examination of water and wastewater. 21st Ed., Washington, D.C.

Asadollahfardi, G., Asadi, M., Nasrinasrabadi, M., & Faraji, A. (2014). Dynamic programming approach (DPA) for assessment of water quality network: a case study of the Sefid-Rüd River. *Water Practice & Technology*, 9(2), 135–139.

Bhangu, I., & Whitfield, P. H. (1997). Seasonal and long-term variations in water quality of the Skeena river at Usk, British Columbia. *Water Research*, 31(9), 2187–2194.

Canadian Council of Ministers of the Environment, Water Quality Task Group, Canada (CCME). (2006). A Canada-wide framework for water quality monitoring. [http://www.ccme.ca/assets/pdf/wqm\\_framework\\_1.0\\_e\\_web.pdf](http://www.ccme.ca/assets/pdf/wqm_framework_1.0_e_web.pdf) (assessed on 16th April, 2015).

Central Ground Water Board Kerala region, Thiruvananthapuram, Government of India. (2004). Ground water estimation methodology—1997. <http://cgwb.gov.in/kr/computation.htm> (assessed on 16th April, 2015).

Central pollution control board, India (CPCB). (2007). Guidelines for water quality monitoring. <http://mpcb.gov.in/envtdata/GuidelinesforWQMonitoring%5B1%5D.pdf> (assessed on 16th April, 2015).

Cetinkaya, C. P., & Harmancioglu, N. B. (2012). Assessment of water quality sampling sites by a dynamic programming approach. *Journal of Hydrologic Engineering*, 17(2), 305–317.

Chang, H. (2008). Spatial analysis of water quality trends in the Han river basin, South Korea. *Water Research*, 42, 3285–3304.

Chang, C., & Lin, Y. (2014a). A water quality monitoring network design using fuzzy theory and multiple criteria analysis. *Environmental Monitoring and Assessment*, 186, 6459–6469.

Chang, C., & Lin, Y. (2014b). Using the VIKOR method to evaluate the design of a water quality monitoring network in a watershed. *International Journal of Environmental Science and Technology*, 11, 303–310.

Cressie, N., & Christopher, K. W. (2011). *Statistics for spatio-temporal data*. Hoboken: Wiley.

Department of water, Government of Western Australia. (2009). Water quality monitoring program design: a guideline for field sampling for surface water quality monitoring programs. <http://www.water.wa.gov.au/PublicationStore/first/87153.pdf> (assessed on 16th April, 2015).

Dixon, W., & Chiswell, B. (1996). Review of aquatic monitoring program design. *Water Research*, 30(9), 1935–1948.

Do, H. T., Lo, S. L., Chiueh, P. T., Thi, L. A., & Shang, W. T. (2011). Optimal design of river nutrient monitoring points based on an export coefficient model. *Journal of Hydrology*, 406, 129–135.

Do, H. T., Lo, S., Chiueh, P., & Thi, L. (2012). Design of sampling locations for mountainous river monitoring. *Environmental Modeling & Software*, 27(28), 62–70.

European Union Member States, Norway and the European Commission. (2003). Guidance on monitoring for the water framework directive. [http://www.nve.no/PageFiles/4661/Monitoring\\_Guidance.pdf?epslanguage=no](http://www.nve.no/PageFiles/4661/Monitoring_Guidance.pdf?epslanguage=no) (assessed on 16th April, 2015).

Hanrahana, G., Gledhilla, M., Houseb, W., & Worsfold, P. J. (2003). Evaluation of phosphorus concentrations in relation to annual

- and seasonal physico-chemical water quality parameters in a UK chalk stream. *Water Research*, 37, 3579–3589.
- Harmancioglu, N. B., & Alphasan, N. (1992). Water quality monitoring network design: a problem of multi-objective decision making. *Water Resources Bulletin*, 28(1), 179–192.
- Harmancioglu, N. B., & Alphasan, N. (1994). Basic approach in design of water quality monitoring networks. *Water Science and Technology*, 30(10), 49–56.
- Harmancioglu, N. B., Fistikoglu, O., Ozkul, S. D., Singh, V. P., & Alpasan, M. N. (1999). *Water quality monitoring network design*. *Water science and technology library*. London: Kluwer academic publishers.
- Hudak, P. F., & Loaiciga, H. (1993). An optimization method for monitoring network design in multilayered groundwater flow system. *Water Resources Research*, 29(8), 2835–2845.
- Jain, S. K., Agarwal, P. K., & Singh, V. P. (2007). *Hydrology and water resources of India*. *Water science and technology library*. Dordrecht: Springer.
- Jha, R., Ojha, C. S. P., & Bhatia, K. K. S. (2005). Estimating nutrient outflow from agricultural watersheds to the river Kali in India. *Journal of Environmental Engineering*, 131(12), 1706–1715.
- Jha, R., Ojha, C. S. P., & Bhatia, K. K. S. (2007). Development of refined BOD and DO models for highly polluted Kali River in India. *Journal of Environmental Engineering*, 133(8), 839–852.
- Jolankai, G. (1997). Basic river water quality models, IHP-V. Technical documents in hydrology No. 13, UNESCO, Paris.
- Karamouz, M., Nokhandan, A. K., Kerachian, R., & Maksimovic, C. (2009a). Design of on-line water quality monitoring system using the entropy theory: a case study. *Environmental Monitoring and Assessment*, 155, 63–85.
- Karamouz, M., Nokhandan, A. K., Kerachian, R., & Maksimovic, C. (2009b). Design of river water quality monitoring networks: a case study. *Environmental Monitoring and Assessment*, 14, 705–714.
- Knuth, D. E. (1968). *The art of computer programming, first edition, vol. 1. Fundamental algorithms*. California: Addison-Wesley.
- Lo, S. L., Kuo, J. T., & Wang, S. M. (1996). Water quality monitoring network design of Keelung river, Northern Taiwan. *Water Science and Technology*, 34(12), 49–57.
- Ministry of Environment and Forest, Government of India (MoEF). (2005). Uniform protocol on water quality monitoring order, 2005. <http://envfor.nic.in/legis/2151e.pdf> (assessed on 16th April, 2015).
- Mishra, K., & Coulibaly, P. (2009). Developments in hydrometric network design: a review. *Reviews of Geophysics*, 47, 1–24.
- Mishra, K., & Coulibaly, P. (2010). Hydrometric network evaluation for Canadian watersheds. *Journal of Hydrology*, 380, 420–437.
- Musthafa, O. M., Subhankar Karmakar., and Harikumar, P. S. (2012). Evaluation of river water quality monitoring network: a multivariate statistical approach to Kabbini river catchment. Proceedings of IWA World Water Congress and Exhibition 2012, International Water Association, Busan, Korea held on September 16–21, 2012.
- Musthafa, O. M., Karmakar, S., & Harikumar, P. S. (2014). Assessment and rationalization of water quality monitoring network: a multivariate statistical approach to the Kabbini River (India). *Environmental Science and Pollution Research*, 21, 10045–10066.
- Ouyang, Y. (2005). Evaluation of river water quality monitoring stations by principal component analysis. *Water Research*, 39, 2621–2635.
- Ouyang, Y., Nkedi-Kizzab, P., Wuc, Q. T., Shindeb, D., & Huang, C. H. (2006). Assessment of seasonal variations in surface water quality. *Water Research*, 40, 3800–3810.
- Ozkul, S., Harmancioglu, N. B., & Singh, V. P. (2000). Entropy-based assessment of water quality monitoring networks. *Journal of Hydrological Engineering*, 5(1), 90–100.
- Park, S., Choi, J. H., Wang, S., & Park, S. S. (2006). Design of a water quality monitoring network in a large river system using the genetic algorithm. *Ecological Modeling*, 99, 289–297.
- Sanders, T.G., Ward, R.C., Loftis, J.C., Steele, T.D., Adrian, D.D. and Yevjevich, V. (1983). Design of networks for monitoring water quality. Water Resources Publications, Highlands Ranch, Colorado, USA.
- Shannon, C. E., & Weaver, W. (1949). *The mathematical theory of communication*. Urbana: University of Illinois Press.
- Sharp, W. E. (1971). A topologically optimum water-sampling plan for rivers and streams. *Water Resources Research*, 7(6), 1641–1646.
- Singh, K. P., Malik, A., Mohana, D., & Sinha, S. (2004). Multivariate statistical techniques for the evaluation of spatial and temporal variations in water quality of Gomti river (India)—a case study. *Water Research*, 38, 3980–3992.
- Stasinopoulos, D. M., & Rigby, R. A. (2007). Generalized additive models for location scale and shape (GAMLSS) in R. *Journal of Statistical Software*, 23(7), 1–46.
- Streeter, H.W., and Phelps, E.B. (1925). A study of the pollution and natural purification of the Ohio river. Public Health Bulletin no. 146, Public Health Service, Washington, D. C.
- Strobl, R. O., & Robillard, P. D. (2008). Network design for water quality monitoring of surface freshwaters: a review. *Journal of Environmental Management*, 87, 639–648.
- Strobl, R. O., Robillard, P. D., Day, R. L., Shanon, R. D., & McDonnell, A. J. (2006a). A water quality monitoring network design methodology for the selection of critical sampling points: part II. *Environmental Monitoring and Assessment*, 122, 319–334.
- Strobl, R. O., Robillard, P. D., Day, R. L., Shanon, R. D., & McDonnell, A. J. (2006b). A water quality monitoring network design methodology for the selection of critical sampling points: part I. *Environmental Monitoring and Assessment*, 122, 137–158.
- Telci, I. T., Nam, K., Guan, J., & Aral, M. M. (2009). Optimal water quality monitoring network design for river systems. *Journal of Environmental Management*, 90, 2987–2998.
- Tsirunov, V. V., Nikanorov, A. M., Laznik, M. M., & Zhu, D. (1992). Analysis of long-term and seasonal river water quality changes in Latvia. *Water Research*, 26(9), 1203–1216.
- United States Environmental Protection Agency (USEPA). (2003). Elements of a state water monitoring and assessment program. [http://www.epa.gov/owow/monitoring/elements/elements03\\_14\\_03.pdf](http://www.epa.gov/owow/monitoring/elements/elements03_14_03.pdf) (assessed on 16th April, 2015).
- United States Environmental Protection Agency (USEPA). (2014). National Rivers and Streams Assessment 2013–2014. [http://water.epa.gov/type/rs/monitoring/riversurvey/upload/NRSA201314\\_FactSheetforCommunities\\_v3.pdf](http://water.epa.gov/type/rs/monitoring/riversurvey/upload/NRSA201314_FactSheetforCommunities_v3.pdf) (assessed on 16th April, 2015).

- United States Geological Survey (USGS). (2005). National field manual for the collection of water-quality data. <http://water.usgs.gov/owq/FieldManual> (assessed on 16th April, 2015).
- United States Geological Survey (USGS). Earth explorer. <http://earthexplorer.usgs.gov> (assessed on 16th April, 2015).
- Varekar, V., Karmakar, S., and Jha, R. (2012). Seasonal evaluation and redesign of surface water quality monitoring network: an application to Kali river basin, India, AOGS-AGU joint assembly 2012, Resorts World Convention Centre, Singapore. ISBN 978-981 -07-2049-0. Held on August 13–17, 2012.
- Varol, M., Gokot, B., Bekleyen, A., & Sen, B. (2012). Spatial and temporal variations in surface water quality of the dam reservoirs in the Tigris River basin, Turkey. *Catena*, 92, 11–21.
- Vega, M., Pardom, R., Barrado, E., & Deban, L. (1998). Assessment of seasonal and polluting effects on the quality of river water by exploratory data analysis. *Water Research*, 32(12), 3581–3592.
- Wang, X., Cai, Q., Lin, Y., and Qu, X. (2012). Evaluation of spatial and temporal variation in stream water quality by multivariate statistical techniques: a case study of the Xiangxi River basin, China. *Quaternary International*, xxx, 1–8.
- Ward, R. C., Loftis, J. C., & McBride, G. B. (1989). *Design of water quality monitoring systems*. New York: Van Nostrand Reinhold.