



A comparative study of the estimation methods for NRCS curve number of natural arid basins and the impact on flash flood predications

Mohammed M. Farran¹ · Amro Elfeki · Mohamed Elhag^{1,2,3} · Anis Chaabani¹

Received: 25 August 2020 / Accepted: 4 December 2020 / Published online: 17 January 2021
© Saudi Society for Geosciences 2021

Abstract

There are various methods to estimate the curve number (CN) for flood studies. In the ungauged basins, hydrologists rely on the use of an NRCS-CN table called (CN_{design}). The CN_{design} , in this study, is estimated using remote sensing techniques and geographical information systems based on alluvium-rock-vegetation classification of natural basins. However, in gauged basins, it is common to use rainfall-runoff data through the application of the least-squares method (LSM) to get the best CN value (CN_{obs}), or the asymptotic fitting method (AFM) to obtain asymptotic CN (CN_{∞}). A comparison between these methods is made under the effect of changing both the coefficient of abstraction ratio, λ , and the effect of data sorting techniques to find out the best estimation of CN for reliable prediction of floods. A methodology has been developed to convert the NRCS-CN table values at $\lambda = 0.2$ to $\lambda = 0.01$ for arid basins. The relationship between the observed CN and the NRCS-CN table shows that estimating runoff using $\lambda = 0.2$ is best made by CN of the impervious area ($CN_{imp} = 90$) instead of 98 (for urban areas) used in the literature. The highest value of CN between the methods is the CN_{design} , then CN_{obs} . CN_{∞} shows the lowest value. Therefore, for a safe design of the hydraulic structures, it is recommended to use CN_{design} . However, for the simulation of the rainfall (P)-runoff (Q) process in the natural basins, it is recommended to use CN_{obs} at the natural sorting of data pairs (P : Q). The root mean square error (RMSE) of CN is reduced from 11 at $CN_{imp} = 98$ to 7 at $CN_{imp} = 90$. This value reflects the infiltration process in the rocks due to the high density of fractures and fissures in the mountains in the area. The developed NRCS-CN table at $\lambda = 0.01$ reduces the RMSE of the estimated runoff depth by 57%.

Keywords Curve number (CN) · Asymptotic fitting method (AFM) · Least-squares method (LSM) · Geographic information system (GIS) · Arid basins · Saudi Arabia

Introduction

Saudi Arabia is one of the arid regions that lack specific studies related to the estimation of CN from field data since runoff

stations are absent in most of the basins. Therefore, flood mitigation studies rely on the soil conservation service curve number (NRCS-CN) procedure. The NRCS-CN method is an acceptable procedure for estimating the volume of runoff in basins for various rainfall-runoff events worldwide USDA-SCS (1985). This procedure uses tables that are easily accessible and experimental formulas to create representative values for CN. The tables are based on $\lambda = 0.2$ and derived mainly NRCS-CN for agricultural basins. Therefore, there is a need to transfer these tables to $\lambda = 0.01$ for arid basins (Farran and Elfeki 2020a). This point is addressed in this paper. There are various methods to estimate CN value for basins. These methods are NRCS-CN table USDA-SCS (1985), least square method (LSM) (e.g., Hawkins et al. 2008; Mishra and Singh 2013; Farran and Elfeki 2020a), and asymptotic fitting method (AFM) (e.g., Hawkins 1993; D'Asaro and Grillone 2012; Farran and Elfeki 2020b). Hawkins (1975) showed the

Editorial Responsibility: Broder J. Merkel

✉ Mohammed M. Farran
mohamadalfaran@yahoo.com

¹ Department of Hydrology and Water Resources Management, Faculty of Meteorology, Environment and Arid Land Agriculture, King Abdulaziz University, Jeddah, Saudi Arabia

² Institute of Remote Sensing and Digital Earth (RADI), Chinese Academy of Science (CAS), 100094 Beijing, China

³ Department of Applied Geosciences Faculty of Science, German University of Technology in Oman, 1816 Muscat, Oman

importance of the estimation of an accurate CN. Therefore, the research question in this study is regarding which of these methods provides an accurate estimation of CN and consequently a reliable runoff prediction in natural arid basins? To the best of the authors' knowledge, there is no answer to this question in the available literature. In the next paragraphs, a review of the methods for CN estimation is presented followed by the main objectives of the research.

To calculate the curve number based on the NRCS method, two main components, namely hydrologic soil group (HSG) and land use, are needed. Most research on literature uses the NRCS-CN procedure to evaluate the generation of runoff. For example, in many river basins in Italy, statistical models have been used with CN methods to find runoff volumes (Kottegoda et al. 2000). In addition, Lin et al. (2007) have used the CN procedure to see how runoff changed due to land-use changes. The runoff has been significantly affected by changes in land use. Moreover, Hamdan et al. (2007) and Hamad et al. (2012) have studied the effect of land-use changes on runoff. Furthermore, Isik et al. (2013) have investigated how land-use changes impact runoff using the CN procedure and artificial neural networks. In West Bank, Shadeed and Almasri (2010) utilized GIS with CN procedure to predict runoff and CN values. Also, Hamad et al. (2012) have utilized GIS to assess how land-use change affects the water budget of the Gaza Strip. Additionally, Al-Juaidi (2018) has assessed the effect of land-use change on potential runoff volumes in the Gaza Strip using the NRCS-CN method and GIS with HEC-GeoHMS. Therefore, the mentioned studies among many others use the common NRCS-CN table to estimate runoff from rainfall events for ungagged basins to provide a design value for CN. However, there are other approaches to determining CN based on rainfall-runoff measurements. For example, one of the methods applied in the literature is the least-squares method (LSM). Walker (1970) is believed to be the first to use LSM, where he applied an experience-and error LSM to the data of runoff from many small basins in the state of Utah in the USA. Several studies have used LSM to reach the potential maximum retention values, e.g., (Simanton et al. 1973; Bales and Betson 1981; Cooley and Lane 1982; Curtis et al. 1983). Hanson et al. (1981) used LSM to find CNs on twenty-five small basins in Montana, Wyoming, and South Dakota and reached a general agreement with the NRCS handbook values. Hawkins et al. (2002 and 2008) used LSM to determine the maximum retention, S , of the basin, and the abstraction ratio, λ , of the runoff equation.

Hawkins (1993) used the asymptotic fitting method (AFM) for the asymptotic estimation of CN values and observed three different types of CN behaviors resulting from CN and rainfall relationship named standard, complacent, and violent behaviors. Thereafter, several studies followed Hawkins's method of using AFM. Bonta (1997) proposed the development of Hawkins' technique for the asymptotic estimation of CN

values from measured data in violent and standard basins using the derived frequency distributions. He discovered that the derived distribution method is promising for CN estimation when rainfall-runoff data is limited. Hjelmfelt et al. (2001) provided complimentary examples of basins that highlight similar behaviors. Fennessey and Hawkins (2001) have compared the CN values for 60 catchments in the USA resulting from three methods and they named them CN_{best} (LSM based), CN_{∞} (AFM based), and CN_{design} determined from the CN_{table} . D'Asaro and Grillone (2012) assessed CN from rainfall-runoff events of 61 Sicilian basins by three various methods: the least-squares method, National Engineering Handbook-4 (NEH4) method, and the asymptotic fitting method (AFM). The results showed that the value of the initial abstraction ratio (λ) studied for basins with the standard CN behavior is well below the original value ($\lambda = 0.2$) mentioned in the USDA-NRCS reports. The NEH4 method was unable to determine the CN value of the correct basins. As AFM indicates a significant response to the standard CN behavior in about 77% of the basins studied versus a complacent behavior (18%) and a violent behavior (5%).

The geographic information system (GIS) has turned into a basic tool in hydrological modeling since it can deal with a vast measure of spatial information. Also, remote sensing (RS) can supply estimations of several hydrologic variables utilized in hydrologic modeling applications. In recent years, satellite remote sensing and GIS have emerged as powerful tools for collecting information on land use and land cover of large areas (Shih 1988; Subudhi et al. 1989). The information on land use, land cover, and hydrologic soil type can be integrated into a GIS environment for a quick and accurate estimation of runoff curve numbers (Stuebe and Johnston 1990). In addition, Still and Shih (1991) utilized remote sensing data to develop the runoff index at the level of the basin and showed how successfully can the remotely sensed data be utilized to track changes in surface runoff caused by land-use changes. Moreover, Saleh, and Al-Hatrushi (2009), and Youssef et al. (2011) have used remote sensing (RS) and GIS procedures as effective tools in flood risk evaluation.

In the past few years, several flood assessment studies have been conducted in Saudi Arabia mainly using GIS and RS procedures (e.g., Metwaly et al. 2010; Al Saud 2010; Dawod and Koshak 2011; Dawod et al. 2011; Al-Hasan and Mattar 2014; Abdulrazzak et al. 2017; Abdulrazzak et al. 2018). Additionally, Al-Subai (1992), and Şen and Al-Subai (2002) have used the traditional NRCS-CN tables for modeling rainfall-runoff in the southwestern region of Saudi Arabia. Furthermore, Dawod et al. (2013) have recommended using the NRCS-CN method as an ideal method for estimating the flood in Makkah, Saudi Arabia. Moreover, Mahmoud et al. (2014) have used GIS-based land use, hydrological soil group, and slope to estimate runoff in the Al-Baha region, the Kingdom of Saudi Arabia (KSA). The above studies

considered only the traditional NRCS-CN tables (USDA-SCS, 1985) because they deal with ungauged basins and there are no runoff measurements.

In recent times, some quantitative studies have been conducted on CN in Saudi Arabia using rainfall-runoff data. For example, Alagha et al. (2016) studied the CN behavior of the Yiba basin in KSA, and the results showed a standard behavior with an approached value of fifty-two. Farran and Elfeki (2020a) used LSM to obtain the optimum range of values for λ and CN based on three estimation methods. The analysis showed that the range of CN varies between 45 and 85 at $\lambda = 0.2$. They concluded that the initial abstraction ratio λ of 0.01 is more representative to the arid basins. Also, Farran and Elfeki (2020b) used AFM to estimate the asymptotic CN (CN_{∞}). The analysis showed that the dominant CN- P relationship is the standard behavior and CN_{∞} shows high variability over the studied basins concerning three values of λ (0.01, 0.2 and 0.3) and the two sorting methods.

Therefore, the aims of this study are the following: (1) estimation of the curve number using the traditional (NRCS-CN) method based on land use and land cover evaluated by using GIS and RS techniques, (2) transforming the NRCS-CN table at $\lambda = 0.2$ into CN table at $\lambda = 0.01$ and apply the resulted values on flood predications, and (3) comparison between the different estimation methods, namely LSM, AFM, and SCS-CN table methods for flood evaluation. The objective of this comparison is to see how the values of the traditional (SCS-CN) method (used in ungauged basins) are related to the values obtained in both the LSM and the AFM (used in gauged basins). The results of this research will guide hydrologists and engineers during the estimation of the CN value of ungauged basins for flood mitigation studies in KSA.

Study area and data collection

The study has taken place on the eastern cliff of Asir region in the south western part of Saudi Arabia. The elevation the cliff of Asir is up to 3000 m and extends west to the Red Sea coast. Geologically, the study region is located in the Asir escarpment which is part of the Arabian Shield (Farran and Elfeki 2020b). The main geology of the region consists of volcanic rocks (Precambrian) and sedimentary rocks (Elfeki et al. 2020). The region is characterized as being very arid with extremist temperature differences, high evaporation rates, and low annual rainfall and random rainfall events are causing flash floods (Abdulrazzak et al. 2020). Also, annual rainfall is closely related to altitude, with total rainfall ranging from 30 to 100 mm on the coastal plain of the Red Sea and up to 450 mm at altitudes over 2000 m (Wheater et al. 1991).

The study included five main basins which are Al-Lith, Tabalah, Yiba, Liyyah, and Habawnah. It is divided into sub-basins (19 sub-basins in total: 4 subbasins in Al-Lith, 3

subbasins in Tabalah, 4 subbasins in Yiba, 3 subbasins in Liyyah, and 5 subbasins in Habawnah). Al-Lith, Yiba, and Liyyah basins drain to the west towards the Red Sea, while Habawnah and Tabalah basins drain to the east towards the Rub al Khali. The study depended on the historical data from measurements of storms (rainfall and runoff events) reaching 161 events collected over a 4-years period (1984 to 1987) by Saudi Arabian Dames and Moore (1988). The historical data was found in a hard copy format, and it has to be transformed into a soft copy format (digital format) to be used. Therefore, Get Data Graph Digitizer software is necessary for this process. Then, the data is transformed to the Excel program. The Watershed Modeling System (WMS) is used to extract basin characteristics and the stream network through the delineation process. Figure 1 shows the general layout of the basins and the locations of the runoff stations at each outlet of the sub-basins and the stream networks.

Methodology

The methodology can be summarized in the following steps:

1. Estimation of CN based on the LSM for natural and ordered sorting of rainfall-runoff data pairs to get CN_{obs} for both $\lambda = 0.2$ and 0.01,
2. Estimation of CN based on the AFM for natural and ordered sorting of rainfall-runoff data pairs to get CN_{∞} for both $\lambda = 0.2$ and 0.01,
3. Developing a formula to transfer the corresponding NRCS-CN tables at $\lambda = 0.2$ to tables for $\lambda = 0.01$,
4. Estimation of CN using NRCS-CN tables for both $\lambda = 0.2$ and the developed tables in step 3 for $\lambda = 0.01$ using GIS and RS for land use and land cover classifications. These values are called CN_{design} .

The detailed methodology is explained in the following sections.

The NRCS-CN method equations which are introduced by the USDA-SCS (1985) are given by:

$$Q = \begin{cases} \frac{(P-\lambda S)^2}{P + (1-\lambda)S}, & P > \lambda S \\ 0, & P \leq \lambda S \end{cases} \quad (1)$$

Where: P is the rainfall depth,

Q is the direct runoff depth,

λ is the initial abstraction ratio, and

S is the potential maximum retention of the basin ($S = 25,400/CN - 254$), where CN is the curve number.

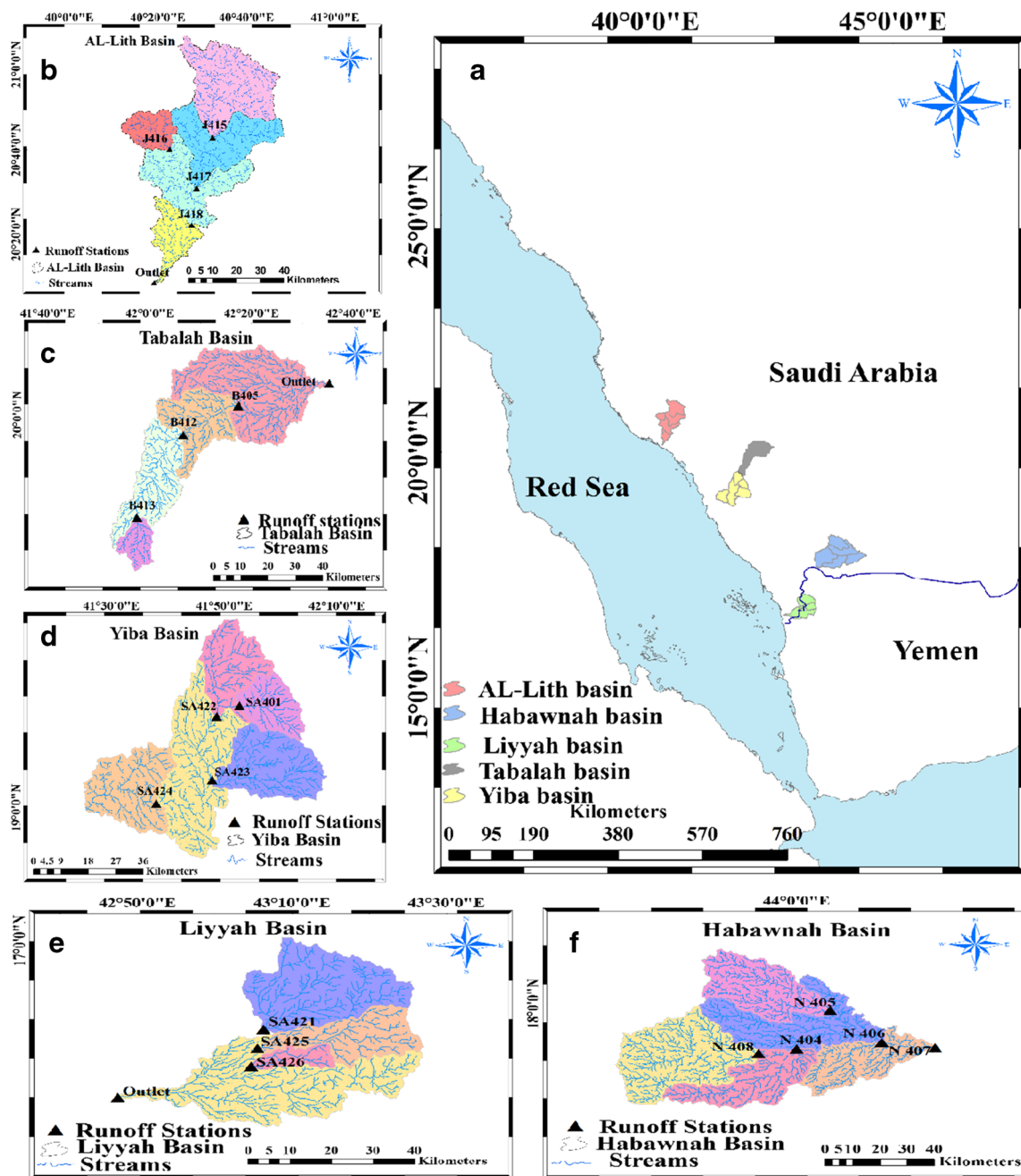


Fig. 1 Locations of the study basins in KSA. **a** General layout of the basins, **b** AL-Lith basin, **c** Tabalah basin, **d** Yiba basin, **e** Liyyah basin, and **f** Habawnah basin

From rainfall and runoff measurements of individual storms, and for a given value of λ , one may calculate CN by combing Eq. 1 with the equation for S to get the equation of CN as,

$$CN = \frac{25400}{\frac{P}{\lambda} + \frac{(1-\lambda)Q - \sqrt{(1-\lambda)^2 Q^2 + 4\lambda PQ}}{2\lambda^2} + 254} \quad (2)$$

In the following sections, the methods used to estimate CN is reviewed and a brief description of each method is presented.

The least squares method

The LSM for CN estimation is introduced by Hawkins et al. (2008). The method determines the parameters λ and CN of the well-known runoff equation. The LSM is a minimization of the sum of squared differences between the observed direct runoff (Q) and the estimated runoff (\hat{Q}). It gives the best estimation of the parameters. The estimated runoff \hat{Q} is obtained by Mishra and Singh (2013),

$$\widehat{Q} = \frac{[P-254\lambda(\frac{100}{CN}-1)]^2}{P + 254(1-\lambda) \left(\frac{100}{CN}-1\right)} \tag{3}$$

The minimization problem is formulated by the objective function as given by Hawkins et al. (2008),

$$f(CN, \lambda) = \min \sum_{i=1}^n \left(Q_i - \widehat{Q}_i\right)^2 \tag{4}$$

Where: Q_i is the observed runoff of an event i (mm), \widehat{Q}_i is the estimated runoff of the same event i (mm), and n is the no. of events (observations).

The values of CN and λ obtained by LSM is called CN_{best} and λ_{best} . Farran and Elfeki (2020a) have used the MLS to obtain CN_{best} and λ_{best} for these basins and their sub-basins and even for each rainfall-runoff event. The best value for λ was found to be 0.01. In the current study, both the value of $\lambda=0.01$ and the common value of $\lambda=0.2$ (USDA-SCS 1985) are used for comparison reason. Therefore, the authors use the terminology CN_{obs} instead of CN_{best} to represent the value of the curve number obtained by substituting the rainfall depth, the corresponding runoff depth, and the value of λ (0.01 or 0.2) at each event in Eq. 2 to get CN_{obs} at $\lambda=0.01$ and $\lambda=0.2$ respectively in the rest of the paper.

Asymptotic fitting method

The asymptotic fitting method (AFM) is a technique to estimate the behavior of the precipitation and the CN value (the so-called CN- P relationship). As the

precipitation approaches infinity ($P \rightarrow \infty$), the CN approaches an asymptotic value called CN_{∞} . Hawkins (1993) showed three different behavior between CN and P . These behaviors are as follows: (1) the complacent behavior, (2) the standard behavior, and (3) the violent behavior. For more details about these behaviors, a reference is made to Hawkins (1993). The formula of the complacent behavior which represents the lower limit of the CN- P relationship is given by Hawkins (1993),

$$CN_O = \frac{25400}{\left(\frac{P}{\lambda}\right) + 254} \tag{5}$$

Where CN_O is the complacent behavior that represents the curve number at $P=Ia$ (the initial abstraction, $Ia = \lambda S$), and therefore, $Q=0$ (a reference is made to Eqs. 1 and 2).

The formula of the standard behavior is given by Hawkins (1993),

$$CN(P) = CN_{\infty} + (100 - CN_{\infty}) \exp\left(-\frac{P}{b}\right) \tag{6}$$

Where CN_{∞} is the asymptotic value ; and b is a constant .

The formula of violent behavior is not considered in this context because it is beyond the scope of the study. It does not show up in the analysis of these basins (Farran and Elfeki 2020b). An extensive study of CN_{∞} on the same basins is presented in Farran and Elfeki (2020b). The estimation of the fitting parameters (CN_{∞} and b) is made through the minimization of the squared difference of CN for the given P , Q , and λ as,

$$f(CN_{\infty}, b) = \min \sum_{i=1}^n \left[\left(\frac{25400}{\frac{P_i}{\lambda} + \frac{(1-\lambda)Q_i - \sqrt{(1-\lambda)^2 Q_i^2 + 4\lambda P_i Q_i}}{2\lambda^2} + 254} \right) - \left(CN_{\infty} + (100 - CN_{\infty}) \exp\left(-\frac{P_i}{b}\right) \right) \right]^2 \tag{7}$$

Where P_i and Q_i are the precipitation and the runoff depths respectively of the event i .

The results of this technique are discussed in the “[results and discussions](#)” section.

Remote sensing and geographical information system for CN estimation

In this approach, the CN is estimated from land use and land cover based on the NRCS table at $\lambda=0.2$ (NRCS

2004) and not from rainfall-runoff events as explained in the previous methods. The procedure to implement this approach is given in the following steps.

RS-based land cover classification approach

Data collection

Data were collected freely from the open-source of the Earth Explorer website (<https://earthexplorer.usgs.gov/>). Images of

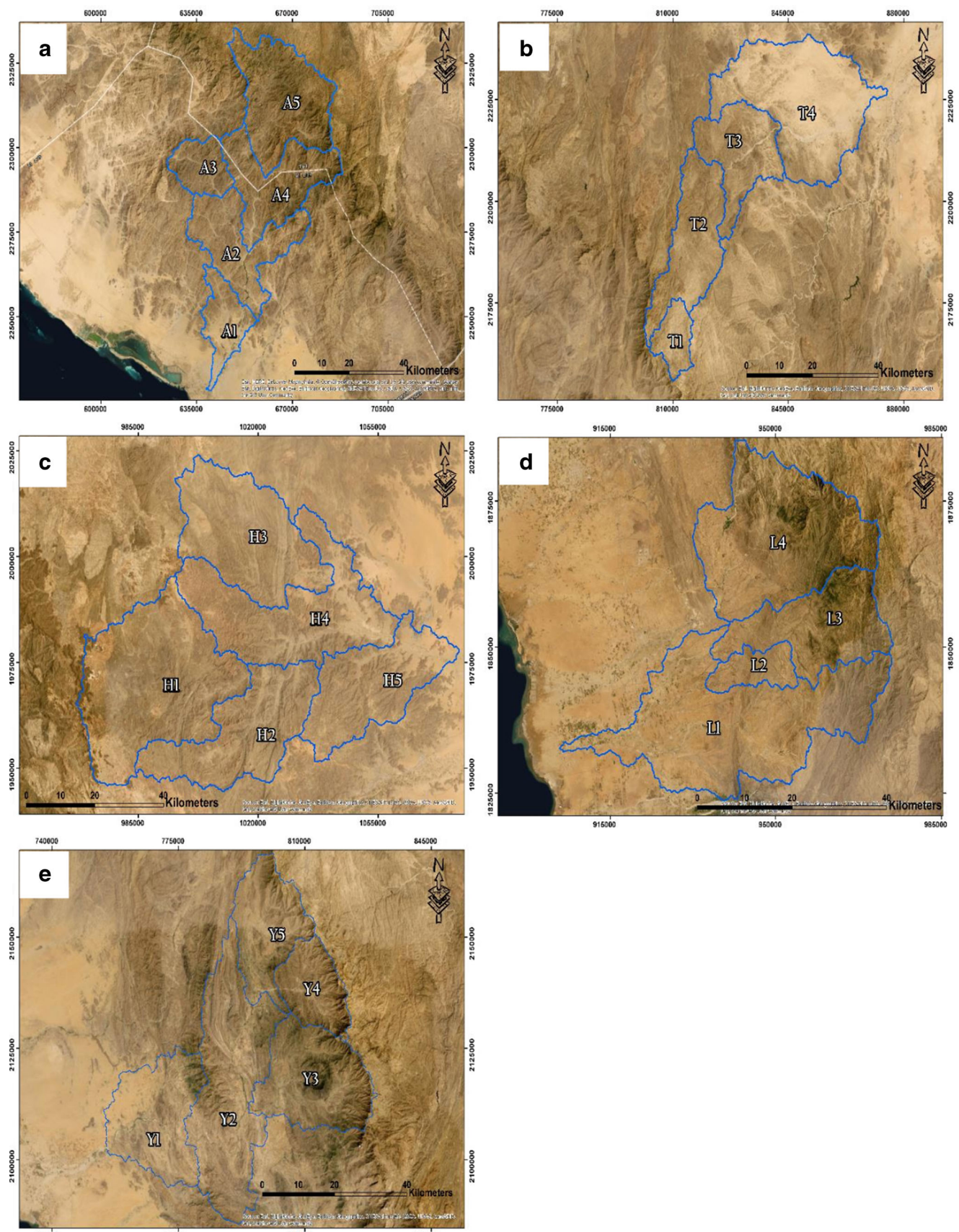


Fig. 2 Maps of the study basins. a AL-Lith, b Tabalah, c Habawnah, d Liyyah, and e Yiba

Operational Land Imager (OLI) were used in the current study. OLI consists of 11 bands including three bands in the visible region, one band in the InfraRed, two bands in the Shortwave Infrared (30-m resolution), one Panchromatic band (15 m resolution), and two bands in the Thermal Infrared (100 m resolution) in addition to two bands targeting the Ultra blue and Cirrus.

Eight OLI scenes were collected coherently during November 2017 to minimize the variation of the atmospheric conditions. The scenes were preprocessed for geometric, radiometric, and atmospheric correction (Itten and Meyer 1993; Baboo and Devi 2011) then mosaicked to cover the whole study area.

Image classification process

The classification process is an information extraction procedure that detects these spectral signatures and then assigns pixels into classes according to the similarity of their spectral signatures (Briem et al. 2002). Image classification is an essential procedure in the field of remote sensing, digital image analysis, and pattern recognition (Kloer 1994; Richards and Jia 1999).

Supervised classification

Supervised classification is a computer-based process where each pixel is assigned to a category based on a preferable decision rule (Swain and Davis 1981). Supervised classification starts initially with the selection of the training sites for different terrain categories. Based on previous scholarly work on classification accuracies (Elhag and Boteva 2016; Elhag 2016), two different supervised classification algorithms are exercised in the current research study, maximum likelihood—ML and support vector machine—SVM. Maximum likelihood classification is achieved according to the following equation:

$$g_i(x) = \ln P(\omega_i)^{-1} / 2 \ln |\Sigma_i| - \frac{d}{2} \ln 2\pi - (x - m_i)^T \Sigma_i^{-1} (x - m_i) \quad (8)$$

Where:

- i is the class number,
- x is the d -dimensional data (where d is the number of bands),
- $P(\omega_i)$ is the probability that class ω_i occurs in the image and is assumed the same for all classes,
- $|\Sigma_i|$ is the determinant of the covariance matrix of the data in class ω_i ,
- Σ_i^{-1} is the inverse of the covariance matrix, and
- m_i is the mean vector of class i ,

Support vector machine is performed according to the two-layer neural net using the following inner product kernel,

$$K(x_i, x_j) = \tanh(\alpha x_i^T x_j + \beta) \quad (9)$$

Where:

$K(x_i, x_j)$ is the inner product kernel function,

- α is the gamma term in the kernel function for all kernel types except linear, and
- β is the bias term in the kernel function for the polynomial and sigmoid kernels.

This approach is employed using Erdas software. With this technique, three land cover classes were identified that contain rocks, vegetation, and soil as shown in Fig. 2 and the details of estimation CN, based on this approach, will be explained in the next section.

Classification accuracy assessment

Accuracy assessment was realized based on 100 points of randomly distributed points in each basin collected from Google Earth, the points representing the major land use land cover in the designated study area. The points were evenly distributed in each scene. The points were converted into 50-m² polygons under the GIS environment for approachability reasons.

Validation points were individually assigned to three different land cover categories: rocks, alluvium, and vegetation. Points were used to compute users, producers, and overall accuracies.

Producer’s accuracy is calculated as follows:

$$\text{Producer accuracy\%} = \frac{C_{ii}}{C_{*i}} 100 \quad (10)$$

Where,

- C_{ii} are the elements at position i th row and i th column.
- C_{*i} is the column sums.

User’s accuracy is calculated using the formula,

$$\text{User accuracy\%} = \frac{C_{ii}}{C_{i*}} 100 \quad (11)$$

- C_{ii} is the element at position i th row and i th column, and
- C_{i*} is the row sums.

The overall accuracy is estimated by the following formula,

$$\text{overall accuracy\%} = \frac{\sum_{j=1}^U C_{jj}}{O} 100 \quad (12)$$

Where,

- C_{jj} is the element at position j th row and j th column, and

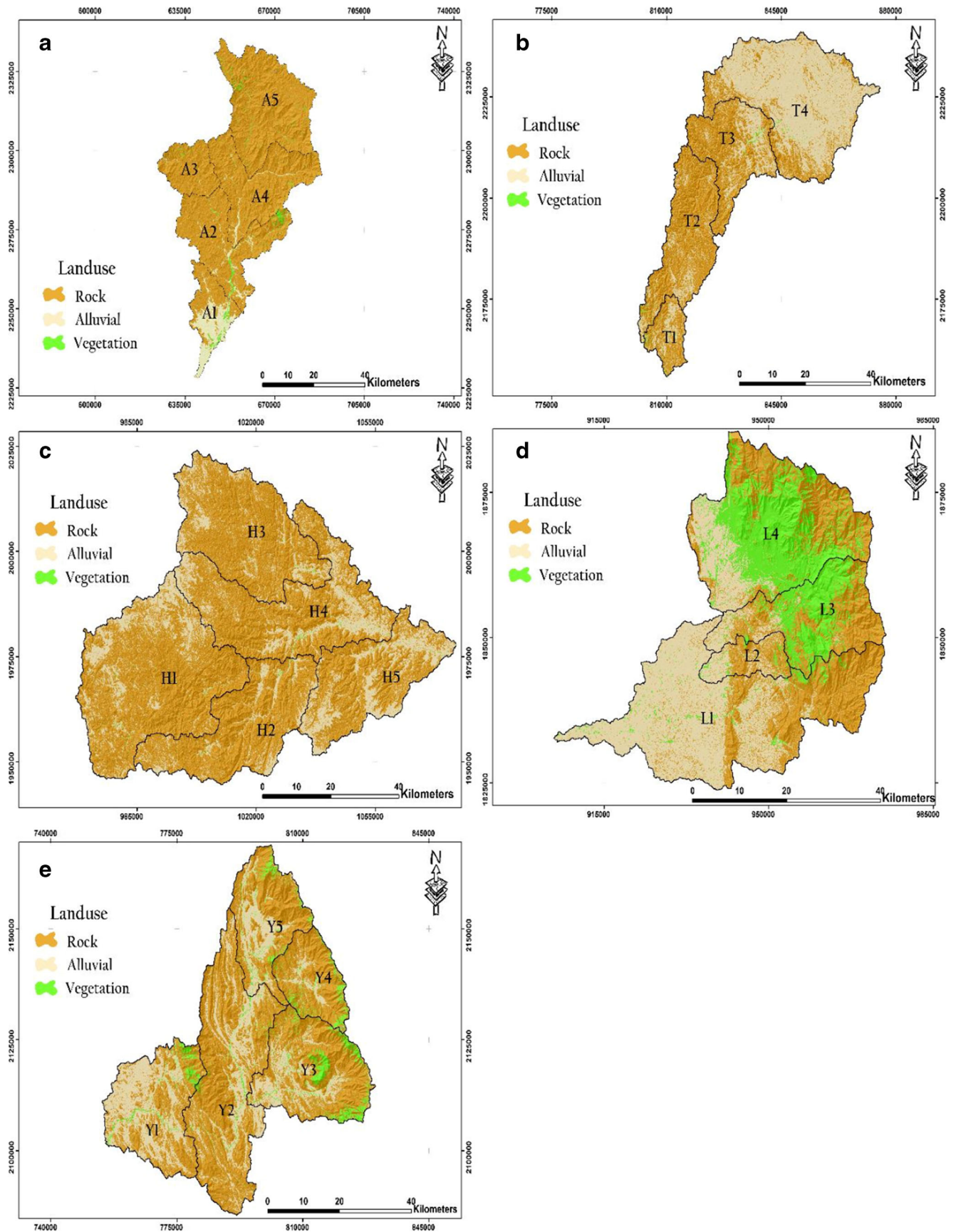


Fig. 3 Land cover (LC) map of the study basins. **a** AL-Lith, **b** Tabalah, **c** Habawnah, **d** Liyyah, and **e** Yiba. Three classifications are identified, namely rock, alluvium, and vegetation

O and U are the total number of pixels and classes respectively.

Matching of user’s and producer’s accuracies delivers accurateness to the classification and assures a robust liability of the implemented accuracy assessment (Cohen (1960), Congalton et al. (1983), and Congalton (1991)).

\hat{K} statistics is a second measure accuracy agreement. This measure of agreement is based on Congalton and Mead (1983) findings. It is defined as the maximum likelihood estimation from the multinomial distribution. \hat{K} is calculated using the following equation:

$$\hat{K} = \frac{\sum_{i=1}^r x_{ii} - \sum_{i=1}^r (x_{i+} \cdot x_{+i})}{N^2 - \sum_{i=1}^r (x_{i+} \cdot x_{+i})} \tag{13}$$

Where

r is the number of rows and columns in the error matrix,

x_{ii} is the number of observations in row i and column i (the diagonal elements),

x_{i+} is the marginal total of row i ,

x_{+i} is the marginal total of column i , and

N is the total number of observations in the matrix.

To increase the accuracy assessment of the vegetation land cover class, Normalized Difference Vegetation Index (NDVI) was exercised according to Pettorelli et al. (2005) where the NDVI layer was overlaid over the slope layer under GIS environment to examine the vegetation cover in steeply areas. NDVI was carried out as follows:

$$NDVI = (NIR - RED) / (NIR + RED) \tag{14}$$

Where:

NDVI is the Normalized Difference Vegetation Index.

NIR and RED are the amounts of near-infrared and red light, respectively, reflected by the vegetation and captured by the sensor of the satellite.

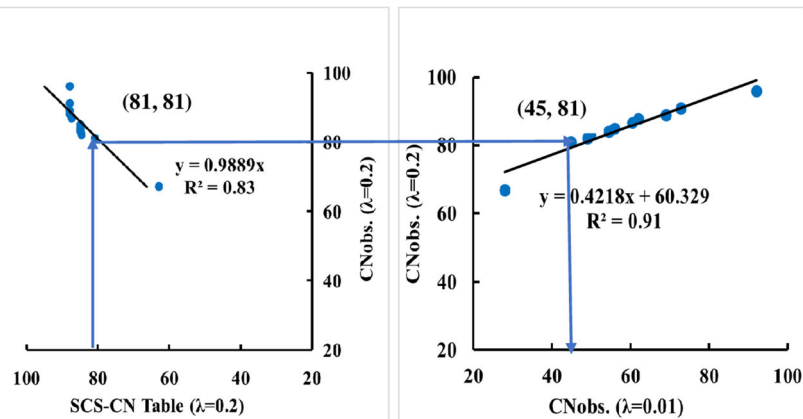
Figure 2 shows the study basins projected on the base map to show the land cover in the study area. Since these basins are natural basins, therefore the only thing one may see is the rocks and the alluvium. Figure 3 shows the basins based on

Table 1 Area of alluvium, vegetation, and rocks and the estimated CN values of sub-basins based on NRCS-CN table at $\lambda = 0.2$ (USDA-NRCS 2004)

Basin	Stations	Code	Area alluvium km ²	Area alluvium %	Area rock km ²	Area rock %	Area Vegetation km ²	Area Vegetation %	Hydrologic condition ^a	Hydrologic soil group			
										A	B	C	D
Yiba	SA401	Y4	32.3	10.6	246.1	80.4	27.5	9.0	Poor	63	77	85	88
	SA422	Y5	124.6	27.1	306.8	66.8	27.9	6.1	Poor	63	77	85	88
	SA423	Y3	138.3	22.6	403.7	65.9	70.5	11.5	Poor	63	77	85	88
	SA424	Y2	240.8	24.8	708.4	73.1	20.4	2.1	Poor	63	77	85	88
AL-Lith	J415	A5	3.1	0.3	951.6	97.5	20.9	2.1	Poor	63	77	85	88
	J416	A3	3.3	1.1	293.5	98.0	2.8	0.9	Poor	63	77	85	88
	J417	A4	18.7	2.5	711.0	96.6	6.4	0.9	Poor	63	77	85	88
	J418	A2	36.9	5.1	664.8	91.8	22.5	3.1	Poor	63	77	85	88
Tabalah	B405	T3	128.6	27.1	342.7	72.3	2.9	0.6	Poor	63	77	85	88
	B412	T2	86.3	15.4	470.90	84.0	3.2	0.6	Poor	63	77	85	88
	B413	T1	43.2	73.9	12.1	20.7	3.2	5.5	Good	49	68	79	84
Liyyah	SA421	L4	129.5	17.7	244.3	33.5	356.1	48.8	Poor	63	77	85	88
	SA425	L3	44.9	11.7	179.2	46.7	160.0	41.7	Poor	63	77	85	88
	SA426	L2	43.2	40.3	56.9	53.1	7.1	6.6	Fair	55	72	81	86
Habawnah	N404	H2	119.6	14.8	689.2	85.0	1.8	0.2	Poor	63	77	85	88
	N405	H3	174.1	18.5	765.0	81.4	1.3	0.1	Poor	63	77	85	88
	N406	H4	278.1	25.9	791.2	73.8	2.8	0.3	Poor	63	77	85	88
	N407	H5	253.2	34.0	490.0	65.8	1.0	0.1	Fair	55	72	81	86
	N408	H1	302.0	21.5	1102.8	78.3	3.3	0.2	Poor	63	77	85	88

^a Poor: < 30% ground cover (litter, grass, and brush overstory), fair: 30 to 70% ground cover, good: > 70% ground cover.

Fig. 4 The developed regression equations to convert NRCS-CN table values at $\lambda = 0.2$ to the developed CN table values at $\lambda = 0.01$ suited for the study basins



the above classification, and three different classifications are distinguished, namely the alluvium, the vegetation, and the rocks. Since most of the basins contain rock and alluvium, therefore, these classifications are reduced to alluviums and rocks to simplify the estimations. Therefore, it has been decided that the vegetation area located in the alluvium is added to the alluvium area and the vegetation located in the rock is added to the rock area. Table 1 shows the area of the vegetation, the rock, and the alluvium and the percentage of the vegetation, the rock, and the alluvium for each sub-basin. According to the NRCS-CN table, the

hydrologic condition is estimated for each sub-basin based on the percentage of grass in the ground cover to poor: < 30%, fair: 30 to 70%, and good: > 70%. Based on this classification, the hydrologic conditions of the sub-basins are determined. The table shows that most of the sub-basins have a poor hydrologic condition. Table 1 shows also the hydrologic soil groups A, B, C, and D, with group A corresponding to high infiltration, while group D corresponds to low infiltration rate. The values of CN assigned to these groups are estimated based on the common landscape of the arid and the semi-arid basins as

Table 2 Classification of the observed CN from rainfall-runoff measurements into hydrologic soil groups A, B, C, and D (NRCS-CN table at $\lambda = 0.2$, USDA-NRCS 2004) for the sub-basins

Basin	Stations	Code	Hydrologic condition	Hydrologic soil group (HSG)				CNobs. at $\lambda = 0.2$	Tabulated CN value (HSG) close to CNobs ($\lambda = 0.2$)	Classification of CNobs. at $\lambda = 0.2$ based on HSG
				A	B	C	D			
Yiba	SA401	Y4	Poor	63	77	85	88	84	85	C
	SA422	Y5	Poor	63	77	85	88	83	85	C
	SA423	Y3	Poor	63	77	85	88	89	88	D
	SA424	Y2	Poor	63	77	85	88	91	88	D
AL-Lith	J415	A5	Poor	63	77	85	88	82	85	C
	J416	A3	Poor	63	77	85	88	84	85	C
	J417	A4	Poor	63	77	85	88	86	85	C
	J418	A2	Poor	63	77	85	88	89	88	D
Tabalah	B405	T3	Poor	63	77	85	88	82	85	C
	B412	T2	Poor	63	77	85	88	87	88	D
	B413	T1	Good	49	68	79	84	85	84	D
Liyah	SA421	L4	Poor	63	77	85	88	82	85	C
	SA425	L3	Poor	63	77	85	88	67	63	A
	SA426	L2	Fair	55	72	81	86	77	81	C
Habawnah	N404	H2	Poor	63	77	85	88	91	88	D
	N405	H3	Poor	63	77	85	88	88	88	D
	N406	H4	Poor	63	77	85	88	88	88	D
	N407	H5	Fair	55	72	81	86	83	81	C
	N408	H1	Poor	63	77	85	88	96	88	D

The bold italic face is the value close to the CN_{obs}

Table 3 Comparison between runoff curve numbers for arid and semiarid rangelands $CN_{T(0.2)}$ and estimated $CN_{T(0.01)}$ in the current study

Cover description		$CN_{T(0.2)}$ (NRCS 2004)				$CN_{T(0.01)}$ (current study)			
		Hydrologic soil group				Hydrologic soil group			
Cover type	Hydrologic condition ^a	A ^b	B	C	D	A	B	C	D
Herbaceous—a mixture of grass, weeds, and low-growing brush, with a brush the minor element	Poor	-	80	87	93	-	49	66	80
	Fair	-	71	81	89	-	27	51	71
	Good	-	62	74	85	-	6	35	61
Oak-aspen—mountain brush mixture of oak brush, aspen, mountain mahogany, bitterbrush, maple, and other brush	Poor	-	66	74	79	-	15	35	47
	Fair	-	48	57	63	-	-	-	8
	Good	-	30	41	48	-	-	-	-
Pinyon-juniper—pinyon, juniper, or both; grass understory	Poor	-	75	85	89	-	37	61	71
	Fair	-	58	73	80	-	-	32	49
	Good	-	41	61	71	-	-	3	27
Sage-grass—sage with an understory of grass	Poor	-	67	80	85	-	18	49	61
	Fair	-	51	63	70	-	-	8	25
	Good	-	35	47	55	-	-	-	-
Desert shrub—major plants include saltbush, greasewood, creosote bush, black bush, bursage, paloverde, mesquite, and cactus	Poor	63	77	85	88	8	42	61	68
	Fair	55	72	81	86	-	30	51	63
	Good	49	68	79	84	-	20	47	59

^a Poor: < 30% ground cover (litter, grass, and brush overstory), fair: 30 to 70% ground cover, good: > 70% ground cover

^b Curve numbers for group A have been developed only for the desert shrub

given in the table of NRCS-CN (NRCS 2004), namely desert shrub-major plants include saltbush, greasewood, creosote bush, black bush, bur sage, paloverde, mesquite, and cactus. Therefore, the values in the tables are obtained from the NRCS-CN table at $\lambda = 0.2$ (NRCS, 2004).

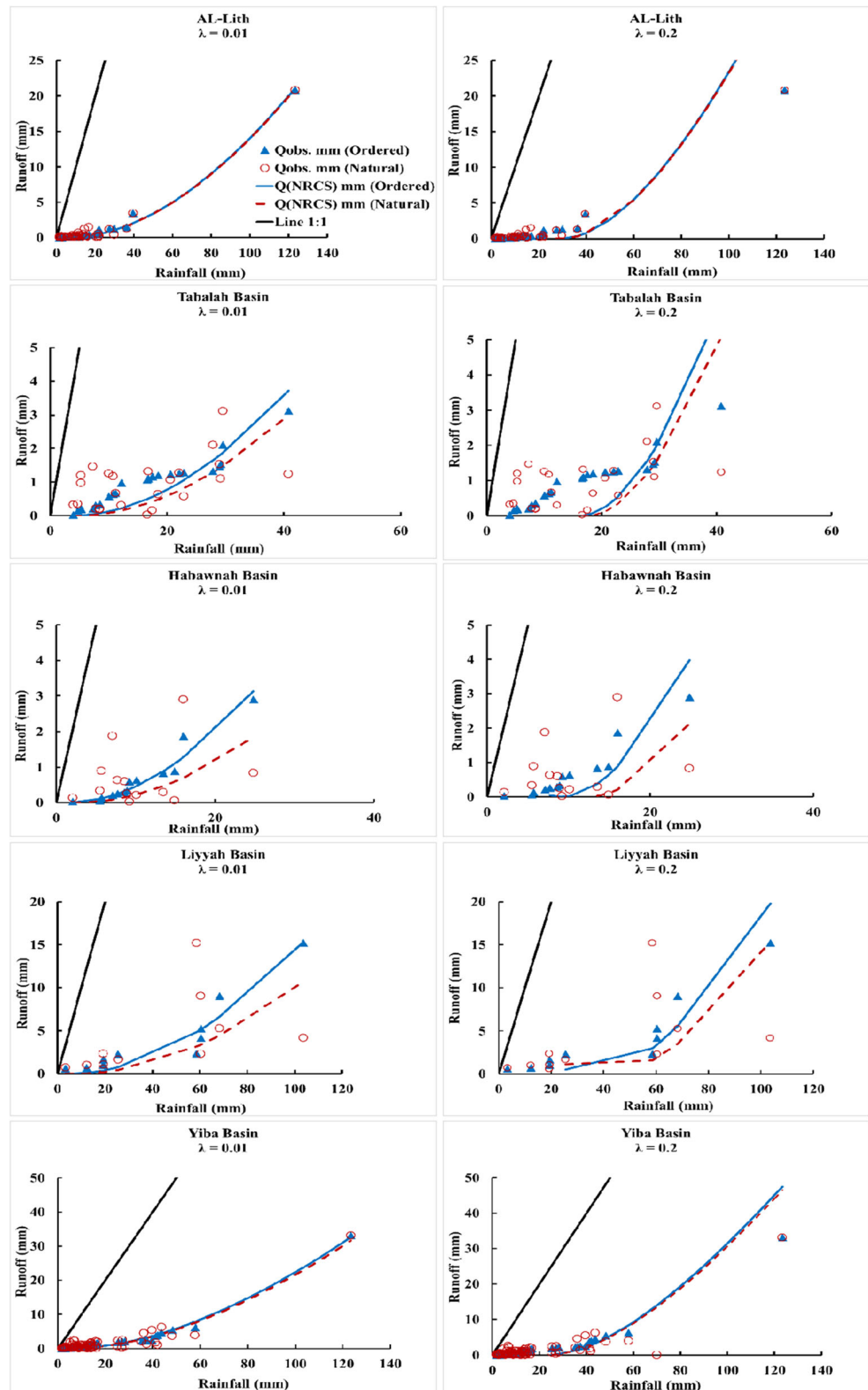
Design CN value based on RS-GIS and NRCS-CN table: the updated traditional approach

In this section, the traditional approach to estimate CN based on the NRCS-CN table is reviewed. However, the traditional approach has been updated by two new aspects. The first aspect is the use of the RS-GIS techniques, as explained in the previous section in the classification of the land cover of the basins into alluvium, rock, and vegetation and estimation of the proportion of each of these classes in the basins. The second aspect is that the NRCS-CN table values are estimated based on the abstraction ratio, $\lambda = 0.2$ (USDA-NRCS 2004). However, in the current study, a value of $\lambda = 0.01$ is also considered in the analysis since it has been shown to best represent the arid basins in Saudi Arabia (Farran and Elfeki, 2020a and b). Therefore, there is a need to convert the NRCS-CN table values from $\lambda = 0.2$ to $\lambda = 0.01$ to test the effect of λ on the results. Figure 4 shows the regression equations that have been developed to make such a transformation. The figure can be expanded as follows. The left image shows the relationship between CN estimated from the rainfall-runoff observations at $\lambda = 0.2$ based on Eq. 2 (CN values are also

presented in Table 2 under column CN_{obs}) and plotted on the vertical axis. The CN values from the NRCS-CN table estimated from the values in Table 2 (the bold italic face in the table under HSG, they are presented in column CN value (HSG) close to CN_{obs}) and plotted on the horizontal axis. A best-fit line is made between the percentiles of both CN values as shown in Fig. 4 (left) to find the relationship between the observed and tabulated CN at $\lambda = 0.2$. The developed relationship is shown in the graph. It is almost a one to one relationship (linear relation with $R^2 = 0.83$). It is worth mentioning that the CN_{obs} at $\lambda = 0.2$ can be classified as soil C and D which indicates relatively low infiltration. However, this seems to be in the contrary to the actual soil condition in these basins which normally have high infiltration during the phenomenon of transmission losses. This point will be discussed later in the “results and discussions” section. Figure 4 (right) shows a percentile plot between the observed CN at $\lambda = 0.2$ on the vertical axis and the observed CN at $\lambda = 0.01$. A best-fit line is made through the data to find a relationship between CN_{obs} at $\lambda = 0.2$ and at CN_{obs} at $\lambda = 0.01$. This relation takes the form in the figure (linear relation with $R^2 = 0.91$). Combining the two formulas in Fig. 4 and rearranging, one may obtain a relationship between NRCS-CN table at $\lambda = 0.2$, $CN_{T(0.2)}$, and the developed CN table at $\lambda = 0.01$, $CN_{T(0.01)}$ as,

$$CN_{T(0.01)} = 2.4CN_{T(0.2)} - 143, CN_{T(0.2)} > 60 \tag{15}$$

Fig. 5 Comparison between the NRCS-CN theory and the observed rainfall-runoff data under two scenarios, namely the ordered versus the natural sorting of data pairs ($P: Q$): Left column ($\lambda = 0.01$) and right column ($\lambda = 0.2$)



Equation 15 is then used to construct a table for CN at $\lambda = 0.01$. It should be noted that Eq. 15 is only valid for $CN_{T(0.2)}$ larger than 60. The reason is that CN is directly proportional to λ

(Farran and Elfeki 2020a) and therefore when λ decreases, the value of CN decreases as well; however, CN cannot be negative, so the limit is $CN_{T(0.2)} = 60$, which corresponds to $CN_{T(0.01)} = 1$.

Table 3 shows a comparison between $CN_{T(0.2)}$ and the corresponding $CN_{T(0.01)}$ for different hydrologic conditions and HSG. It is obvious from the table that the $CN_{T(0.01)}$ is lower than the $CN_{T(0.2)}$. The low values of $CN_{T(0.01)}$ reflect the effect of transmission losses in these basins. Both values of $CN_{T(0.2)}$ and $CN_{T(0.01)}$ are going to be used to estimate design CN values for the basins as will be explained in the coming section.

The common approach of CN estimation in ungagged basins is made using the NRCS-CN table (USDA-NRCS 2004). Hydrologists normally use a weighted average CN (called the composite CN, and is abbreviated as CN_c) over the basin through the application of the following equation,

$$CN_c = \frac{\sum_{i=1}^n CN_i A_i}{\sum_{i=1}^n A_i} \tag{16}$$

Where,

- CN_c is the composite runoff curve number,
- A_i is the area of class i ,
- CN_i is the curve number of class i , and
- n is the number of classes.

Since the natural basins in the current study are classified as containing only two classes, namely rock and alluvium, therefore, for practical applications point of view, Eq. 16 can be simplified to read (USDA-NRCS 2004),

$$CN_c = CN_p + \left(\frac{P_{im}}{100}\right)(CN_{imp} - CN_p) \tag{17}$$

where:

- CN_p is the pervious runoff curve number (in case of natural basins is the alluvium),
- P_{im} is the percent imperviousness (in the case of the natural basins in the rock), and
- CN_{imp}

Table 4 R^2 and RMSE of the runoff, Q , estimated by the LSM at the two values of λ under both natural and ordered sorting of data

Basins	Natural sorting				Ordered sorting			
	CN ($\lambda = 0.01$)		CN ($\lambda = 0.2$)		CN ($\lambda = 0.01$)		CN ($\lambda = 0.2$)	
	R^2	RMSE mm	R^2	RMSE mm	R^2	RMSE mm	R^2	RMSE mm
AL-Lith	0.98	0.42	0.79	4.31	0.98	0.29	0.79	4.35
Tabalah	0.25	0.81	0.04	1.19	0.88	0.40	0.23	1.10
Habawnah	0.04	0.91	0.00	1.07	0.94	0.20	0.66	0.63
Liyyah	0.14	4.56	0.02	6.31	0.92	1.30	0.72	2.81
Yiba	0.92	1.07	0.83	2.79	0.98	0.51	0.83	2.80

is the curve number of the impervious area (rock) which is equal to 98 based on (USDA-NRCS 2004).

Equation 16 is mainly used in the literature to study the runoff in urban areas with CN_{imp} used to describe the runoff curve number in paved regions with a value of 98. This equation is considered to estimate the design CN (which is called $CN_{design} = CN_c$) based on the traditional approach used by hydrologists and engineers to estimate the CN of a basin.

The aforementioned methodologies will lead to three values of CN for each λ , namely CN_{obs} which is estimated from the observation, CN_{∞} which is the asymptotic CN, and the estimated CN from the NRCS-CN table (CN_{design}). The results are discussed in the following section.

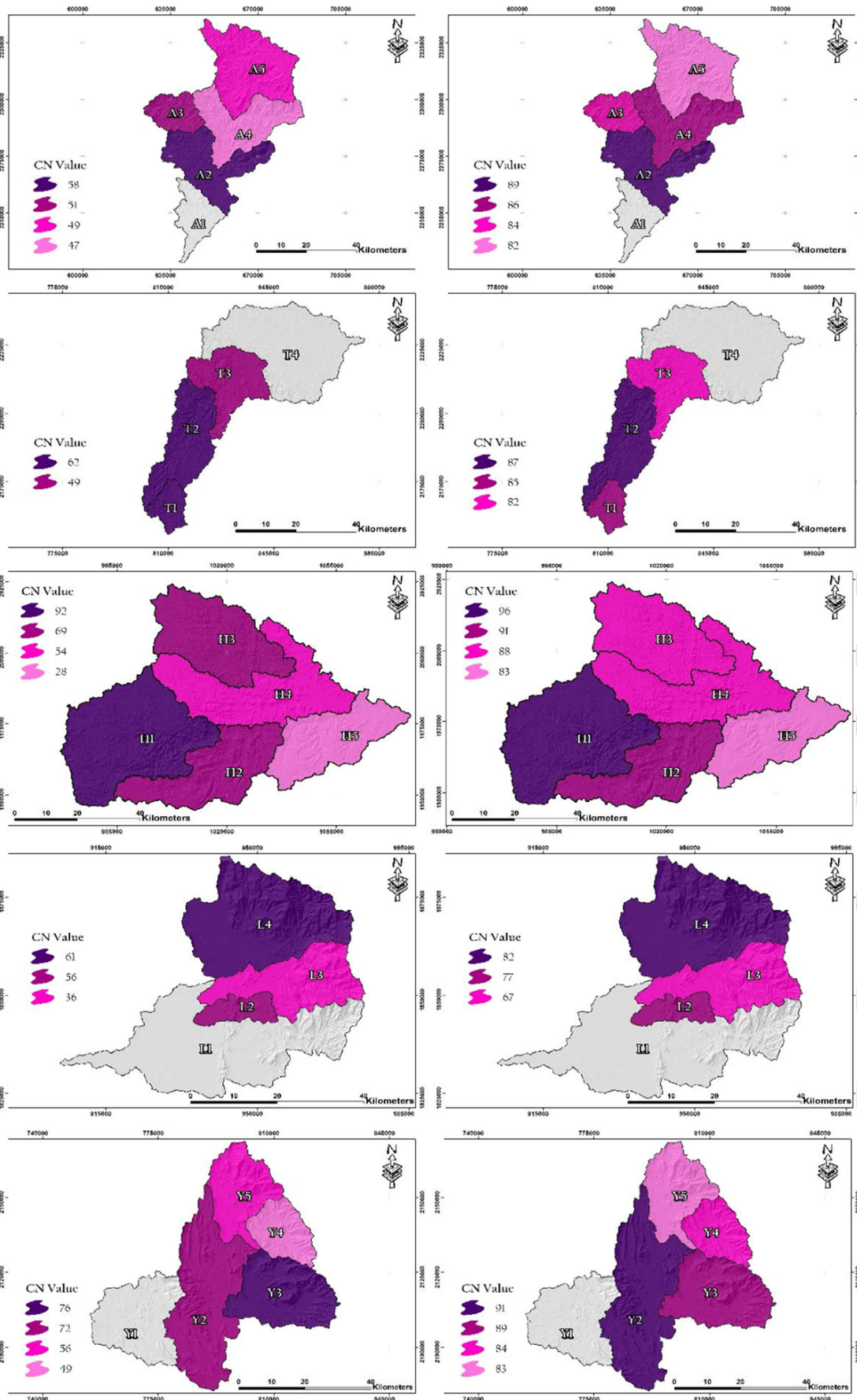
Results and discussions

Analysis of the results based on LSM

Figure 5 shows a comparison between the NRCS-CN theory and the observed rainfall-runoff data under two scenarios, namely the natural versus the ordered sorting of rainfall-runoff data pairs ($P: Q$). The natural sorting of data is used in the literature by, e.g., D’Asaro and Grillone (2012). The pairs of rainfall and the corresponding runoff are combined to estimate CN. This method is consistent since the generated runoff is coming from the same rainfall that caused it. However, in the ordered data sorting method (Hjelmfelt 1980), the rainfall and runoff data are independent and are rearranged by ranking as one may do in frequency analysis of rainfall. This ordered method is also called the frequency matching technique since a new group of pairs of $P: Q$ is generated (in this case, Q is not associated with its causative P). Both techniques are considered in the analysis. The left column of Fig. 5 displays the results for both techniques at $\lambda = 0.01$ and the right column displays the results at $\lambda = 0.2$. The natural data appeared scattered in the figure with the red circles. However, the ordered data are less scattered and presented in the blue triangles. The NRCS-CN theory given by

Table 5 Summary of the results of CN based on LSM (CN_{obs}) at the two values of λ under both natural and ordered sorting of data

Basins	Natural sorting		Ordered sorting	
	CN ($\lambda = 0.01$)	CN ($\lambda = 0.2$)	CN ($\lambda = 0.01$)	CN ($\lambda = 0.2$)
	AL-Lith	32	64	32
Tabalah	38	75	43	76
Habawnah	51	81	63	85
Liyyah	25	56	32	60
Yiba	43	69	44	69



◀ **Fig. 6** Mapping CN_{obs} values over the basins based on $\lambda=0.2$ (right column) and $\lambda=0.01$ (left column). Zones with gray color have no values because there are no runoff stations at the outlet of these basins to estimate CN

Eq. 1 is plotted through the data for both cases (i.e., natural and ordered) at both values of λ . Visual inspection of the figure shows that the theory fits better the ordered data rather than the natural data. The reason is that the ranking process leads to less scattered data and therefore most of the literature studies use

ordered data. Another important observation is that at $\lambda=0.01$, the data is better represented by the NRCS-CN theory in comparison with $\lambda=0.2$ (the common value). This confirms that in the Saudi arid environment, it is recommended to use $\lambda=0.01$ rather than 0.2. Table 4 presents a quantitative analysis of the results presented in Fig. 5. Both the coefficient of determination, R^2 , and the root mean square error (RMSE) of the fitting are estimated for both cases at both values of λ . The table shows that R^2 is the highest at $\lambda=0.01$ under ordered sorting of data and provides the minimum RMSE. This

Fig. 7 AFM applied to the rainfall data under the two values of $\lambda=0.01$ and 0.2 respectively, and the natural (left column) and ordered (right column) sorting of data pairs ($P: Q$)

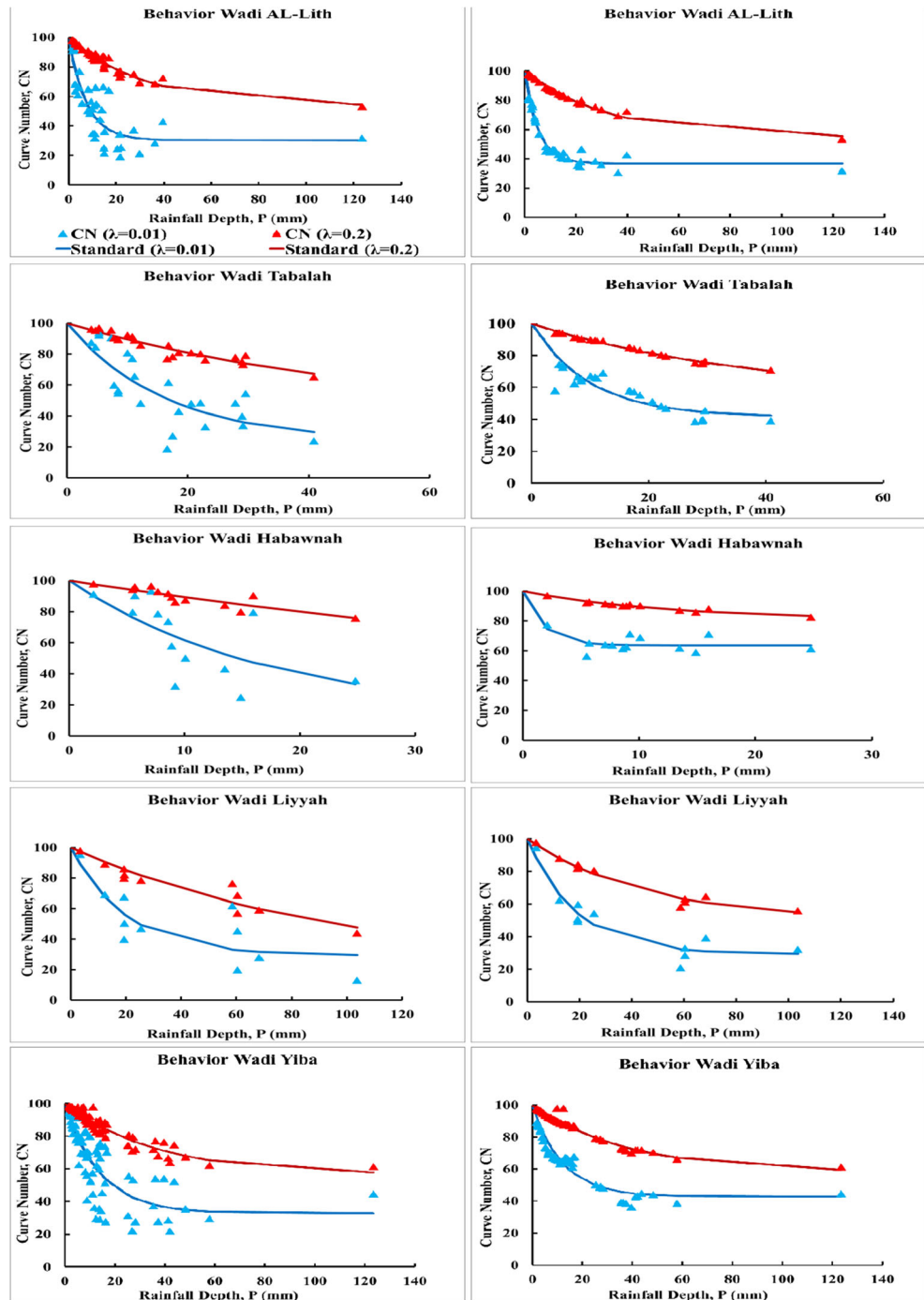


Table 6 Summary of the developed equations for the standard behavior under both natural and ordered sorting of data pairs with both values of λ

Basins		Standard ($\lambda = 0.01$)	R^2	Standard ($\lambda = 0.2$)	R^2
AL-Lith	Natural	$CN(P) = 30.3 + 69.7\exp(-P/7.5)$	0.71	$CN(P) = 53.3 + 46.7\exp(-P/32.2)$	0.95
	Ordered	$CN(P) = 36.8 + 63.2\exp(-P/5.2)$	0.96	$CN(P) = 54.5 + 45.5\exp(-P/32.2)$	0.99
Tabalah	Natural	$CN(P) = 23.7 + 76.3\exp(-P/16.1)$	0.66	$CN(P) = 37.5 + 62.5\exp(-P/54.8)$	0.91
	Ordered	$CN(P) = 41.5 + 58.5\exp(-P/9.9)$	0.72	$CN(P) = 53.4 + 46.6\exp(-P/40)$	0.99
Habawnah	Natural	$CN(P) = 12.1 + 87.9\exp(-P/17.5)$	0.47	$CN(P) = 9.5 + 90.5\exp(-P/80.2)$	0.78
	Ordered	$CN(P) = 63.6 + 36.4\exp(-P/1.7)$	0.36	$CN(P) = 80.5 + 19.5\exp(-P/12.7)$	0.95
Liyah	Natural	$CN(P) = 29.1 + 70.9\exp(-P/20.3)$	0.67	$CN(P) = 19.7 + 80.3\exp(-P/98.4)$	0.89
	Ordered	$CN(P) = 29.2 + 70.8\exp(-P/18.7)$	0.91	$CN(P) = 49.9 + 50.1\exp(-P/44.6)$	0.97
Yiba	Natural	$CN(P) = 32.6 + 67.4\exp(-P/14.1)$	0.63	$CN(P) = 56.1 + 43.9\exp(-P/36.8)$	0.91
	Ordered	$CN(P) = 42.7 + 57.3\exp(-P/12.1)$	0.77	$CN(P) = 57.9 + 42.1\exp(-P/38)$	0.96

result is previously supported by Farran and Elfeki (2020a). Table 5 displays the results of CN based on LSM. The table shows that the values of CN are lower at $\lambda = 0.01$ when compared with $\lambda = 0.2$. This is due to the significance of the transmission losses in these basins which is a common phenomenon in arid regions. Also, one may notice that the CN values in the case of the natural data are relatively lower than the case of ordered data. The reason is that in the natural data, the runoff is generated from its rainfall; therefore, the obtained CN is a real representation of the event. It accounts for the actual transmission losses from that specific event. However, in the ordered data, it is not the case. The rainfall and runoff are not from the same event and therefore the CN does not reflect the actual transmission losses.

Figure 6 shows the mapping of the CN_{obs} values over the basins based on $\lambda = 0.2$ (right column) and $\lambda = 0.01$ (left column). The figure displays the magnitude of CN for each subbasin with different colors noticing that the subbasin with gray color has no data since there are no measurements of runoff at the outlet of such subbasin (see Table 1 and Fig. 1 for each runoff station for each subbasin). It is quite obvious that the values of CN are varying between each subbasin according to the rainfall-runoff measurement at the outlet station. The values of CN for $\lambda = 0.2$ is relatively higher than the values of CN for $\lambda = 0.01$. This observation is because there is a direct proportion between the CN and λ as explained by Farran and Elfeki (2020a). However, the values of CN at $\lambda = 0.01$ seem more realistic to represent these basins since it accounts for the transmission losses which is a significant phenomenon in these mountainous arid basins (Farran and Elfeki, 2020a and c). The values of CN in these maps are useful to study the rainfall-runoff simulations in these basins for flood studies.

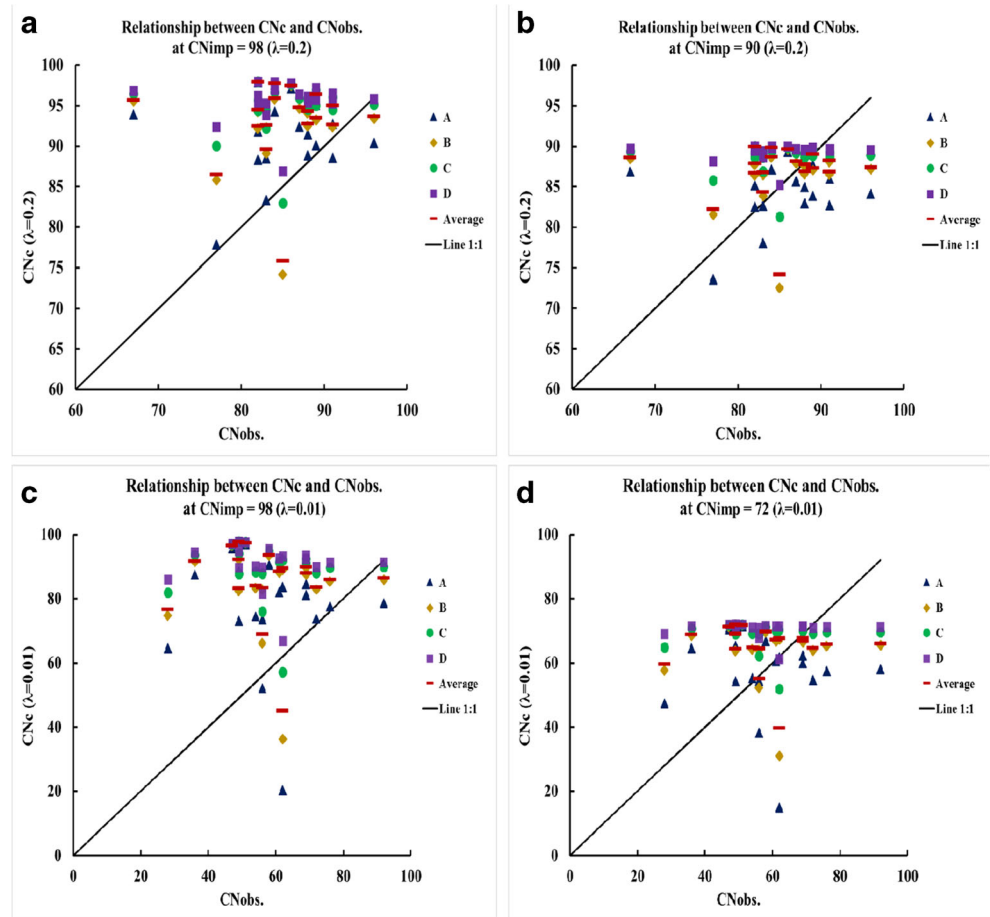
Analysis of the results of the AFM

Figure 7 shows the results of the application of the AFM to the rainfall data under the two values of $\lambda = 0.01$ and 0.2 and under the natural (left column) and ordered (right column) sorting of data pairs ($P: Q$). The standard behavior is the dominant pattern in the basins. The figure shows a relatively higher scatter for the natural sorting of data when compared with the ordered sorting data. Another observation is that the scatter is relatively high for $\lambda = 0.01$ when compared with $\lambda = 0.2$; however, the asymptotic value, CN_{∞} , is reached for most of the basins within the size of the rainfall data for $\lambda = 0.01$ and not reached for $\lambda = 0.2$.

The developed $CN-P$ relationship for the basins is displayed in Table 6. The fitting parameters (CN_{∞} and b) of the $CN-P$ equations are given in the table. The CN_{∞} varies from 12.1 to 32.6 in the case of $\lambda = 0.01$ and from 9.5 to 56.1 for $\lambda = 0.2$ under the natural sorting of data. The decay coefficient, b , varies between 7.5 and 20.3 mm in the case of $\lambda = 0.01$ and from 32.2 to 98.4 mm for $\lambda = 0.2$ under the natural sorting of data. This means that low CN_{∞} is reached for $\lambda = 0.01$ with a fast decay when compared with $\lambda = 0.2$ which again reflects the high transmission losses in these basins. The high transmission losses have been supported by an earlier study on the Yiba basin (Elfeki et al. 2015). They showed that the transmission losses, in some reaches in the Yiba basin, can reach up to 84% of the inflow hydrograph.

On the other hand, the CN_{∞} varies between 29.2 and 63.6 in the case of $\lambda = 0.01$ and from 49.9 to 80 for $\lambda = 0.2$ under the ordered sorting of data. The decay coefficient, b , varies between 1.7 and 18.7 mm in the case of $\lambda = 0.01$ and from 12.7 to 44.6 mm for $\lambda = 0.2$ under the ordered sorting of data. In this case, the CN_{∞} and b have the same features as the natural case; however, the CN_{∞} is somewhat higher due to the ordering of the data and also the scatter is reduced. The decay is faster to reach CN_{∞} in the sorting case while the decay at $\lambda = 0.01$ is faster than at $\lambda = 0.2$.

Fig. 8 Comparison between composite CN (CN_c) obtained from Eq. (17) and observed CN (CN_{obs}) at various values of λ and CN_{imp} at different HSG (A, B, C, and D) and the average value of CN. **a** $\lambda = 0.2$ and $CN_{imp} = 98$, **b** $\lambda = 0.2$ and $CN_{imp} = 90$, **c** $\lambda = 0.01$ and $CN_{imp} = 98$, and **d** $\lambda = 0.01$ and $CN_{imp} = 72$



A quantitative measure of the discrepancies between the cases is displayed in Table 6. The coefficient of determination, R^2 , is relatively higher in the case of $\lambda = 0.2$ than in the case of $\lambda = 0.01$. However, this does not mean that $\lambda = 0.2$ is better to represent the data than $\lambda = 0.01$ because the NRCS-CN theory has proven that $\lambda = 0.01$ is the best to fit the data (see the results in Table 4). Therefore, AFM seems to be elusive in testing the validity of the NRCS-CN theory. Since the results of AFM is on the contrary of the LSM in the previous section. Therefore, it is not recommended to be used to represent CN for these natural arid basins.

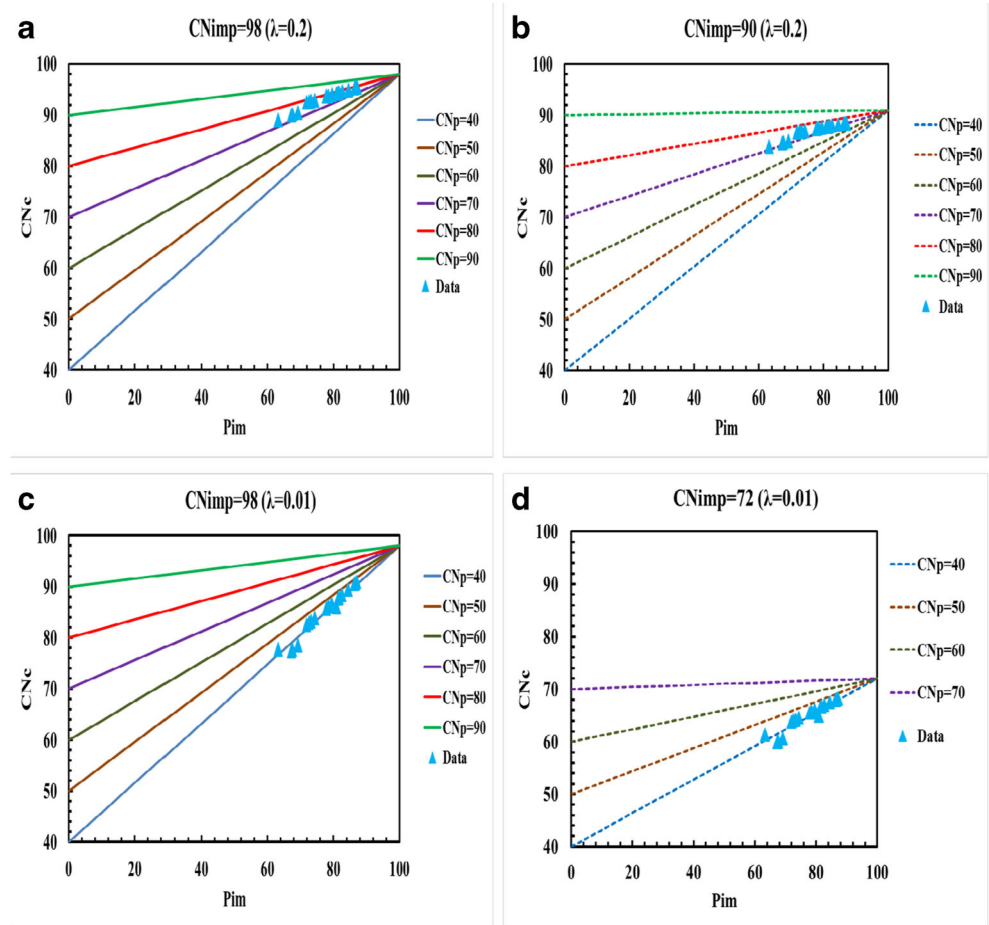
Comparison of the CN_c (CN_{desgin}) estimated from NRCS table and CN estimated from the observation (CN_{obs})

Figure 8 shows a scatter plot between the composite CN (CN_c) obtained from Eq. 17 and observed CN (CN_{obs}) at various values of λ for different values of CN_{imp} at different HSG (A, B, C, and D) as shown in Table 1. The average value of CN of the various values of HSG (A, B, C, and D) is also plotted. In this figure, a tuning with the parameter CN_{imp} is made to get the scattered data points around line 1:1. In Fig. 8a at $\lambda = 0.2$, and $CN_{imp} = 98$ (it represents impervious pavements in urban areas), the scatter points are mostly located above the line of perfect fit (1:1 line). This means that the tabulated NRCS-CN values (Table 3 $CN_{T(0.2)}$, obtained from the reference NRCS 2004) overestimate the observations at $\lambda = 0.2$ by using $CN_{imp} = 98$. However, by tuning the CN_{imp} to a value of 90 at the same $\lambda = 0.2$, the line of perfect fit is somewhat located in the middle of the scattered data (Fig. 8b). This implies that when analyzing runoff in these basins at $\lambda = 0.2$, the best value for CN_{imp} should be 90 instead of 98. Table 7 shows the estimation of the RMSE for both cases at $\lambda = 0.2$. It shows that the RMSE in the CN is reduced

Table 7 Comparison between RMSE of CN observed (CN_{obs}) and composite runoff CN (CN_c) at different values of CN_{imp} and λ

λ	RMSE between CN_{obs} and CN_c		
	$CN_{imp} = 98$	$CN_{imp} = 90$	$CN_{imp} = 72$
0.2	11	7	–
0.01	34	–	18

Fig. 9 Comparison between CN_c obtained from Eq. 17 CN_c from the data as a function of the percentage of the impervious area (P_{im}) at various values of λ and CN_{imp} . **a** $\lambda = 0.2$ and $CN_{imp} = 98$, **b** $\lambda = 0.2$ and $CN_{imp} = 90$, **c** $\lambda = 0.01$ and $CN_{imp} = 98$, and **d** $\lambda = 0.01$ and $CN_{imp} = 72$



from 11 at $CN_{imp} = 98$ to 7 at $CN_{imp} = 90$. The interpretation of these results is that the mountainous rocks in the study area contain a relatively high density of fractures and fissures, due to weathering conditions, that contribute to the infiltration process and consequently lead to lower CN_{imp} . Figure 8c shows the same relationship between the composite CN (CN_c) obtained from Eq. 17 and the observed CN (CN_{obs}) at $\lambda = 0.01$ and $CN_{imp} = 98$. The figure shows the overestimation of the CN_c when compared with CN_{obs} . The CN_c is estimated

based on the results in Table 3 under the column of $CN_{T(0.01)}$, based on the developed equation (Eq. 15). However, CN_c , in this case, is relatively much more scattered in the vertical direction than that of Fig. 8a. This scatter is reduced by tuning the CN_{imp} to a value equal to 72 as shown in Fig. 8d. Table 7 shows a reduction in RMSE of CN from 34 to 18 for $CN_{imp} = 98$ to $CN_{imp} = 72$ respectively. Again, this reflects the infiltration from these rocks due to fissures and cracks in these rocks. Therefore, it is recommended to use $\lambda = 0.2$ with $CN_{imp} = 90$

Table 8 Summary of the percentage of pervious (alluvium) and impervious (rocks) and the corresponding CN (CN_p) for alluvium, CN_{imp} for rocks, and the composite CN (CN_c) at different values of λ

Basin	P %	Pim %	$\lambda = 0.2$					$\lambda = 0.01$				
			CN_p	CN_{imp}	CN_c	CN_{imp}	CN_c	CN_p	CN_{imp}	CN_c	CN_{imp}	CN_c
AL-Lith	27.39	72.61	70	98	92	90	86	40	98	82	72	63
Tabalah	17.76	82.24	70	98	93	90	87	40	98	85	72	65
Habawnah	21.26	78.74	70	98	94	90	88	40	98	86	72	66
Liyyah	18.87	81.13	70	98	93	90	87	40	98	85	72	65
Yiba	24.47	75.53	70	98	94	90	87	40	98	86	72	66

or $\lambda = 0.01$ with $CN_{imp} = 72$ in analyzing runoff in such basins. In both cases, the error in runoff estimation will be minimal.

Relationship between CN_c from Eq. (17) and CN_c from the data as a function of the percentage of the impervious area (P_{im})

Figure 9 displays the theoretical relationship between CN_c obtained from Eq. 17 and the CN_c from the data as a function of the percentage of the impervious area (P_{im}) at various values of λ and CN_{imp} . The figure shows four scenarios. The first scenario is when $\lambda = 0.2$ and $CN_{imp} = 98$; this case is presented in Fig. 9a. It shows that the data for the basins are located at the line of $CN_p = 70$ which means that the alluvium

area has $CN = 70$ which seems relatively high. The same observation can be noticed in the second scenario when $\lambda = 0.2$ and $CN_{imp} = 90$; the CN data for the alluvium is also located at the line with $CN_p = 70$. The third scenario is presented in Fig. 7c where $\lambda = 0.01$ and $CN_{imp} = 98$. In this scenario, the data is located on the line with $CN_p = 40$. This leads to an estimation of the CN of the alluvium in these basins to be about 40. This seems more realistic for these basins since it would account for the transmission losses in the alluvium. The last scenario that appeared in Fig. 9d is similar to the previous scenario with $\lambda = 0.01$, while with $CN_{imp} = 72$. Table 8 summarizes the parameters ($P\%$, $Imp\%$, CN_p , CN_{imp} , and CN_c) used for fitting the data for all scenarios for these basins. These parameters can be used for future studies of rainfall-runoff simulations in these basins.

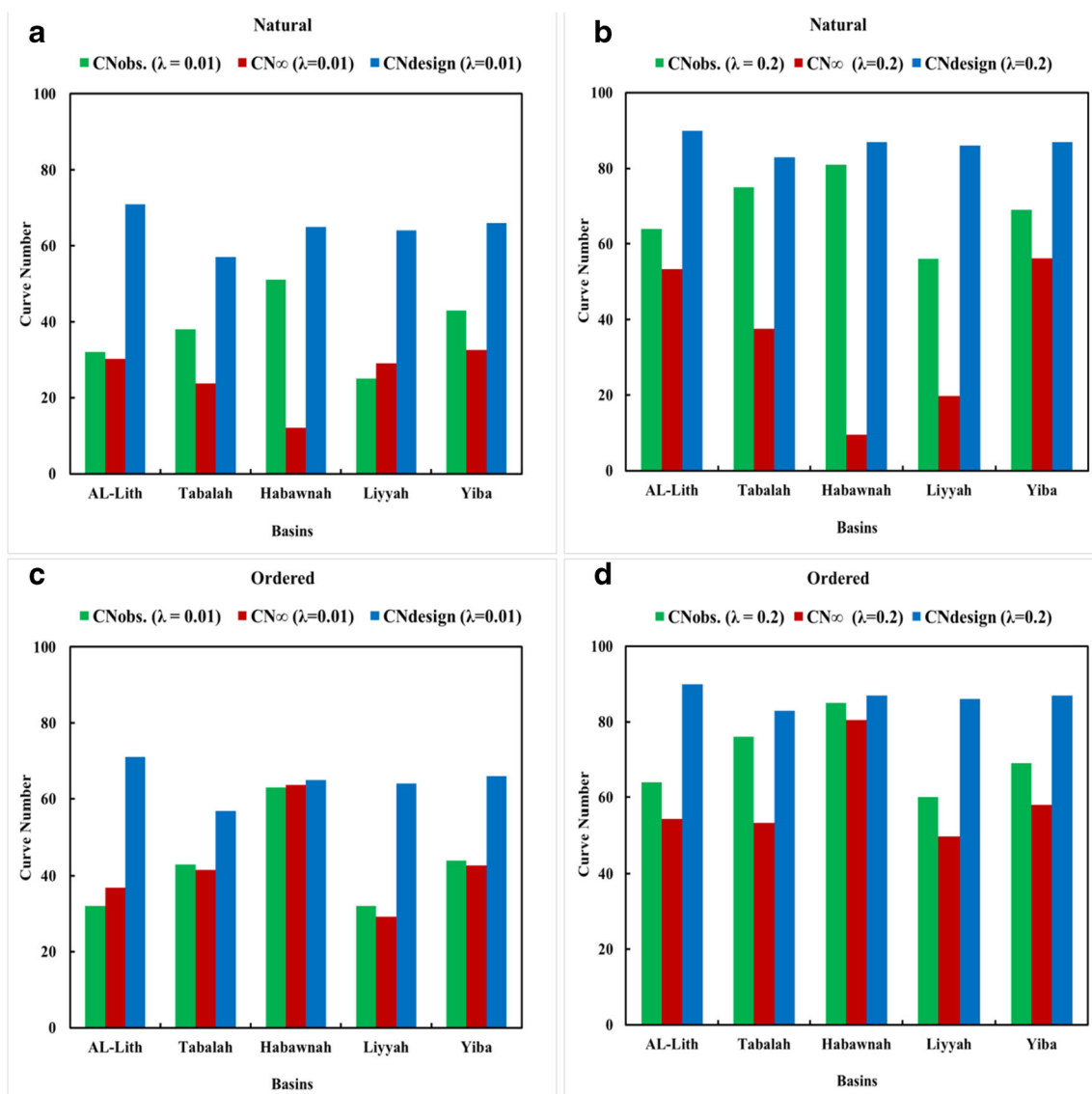
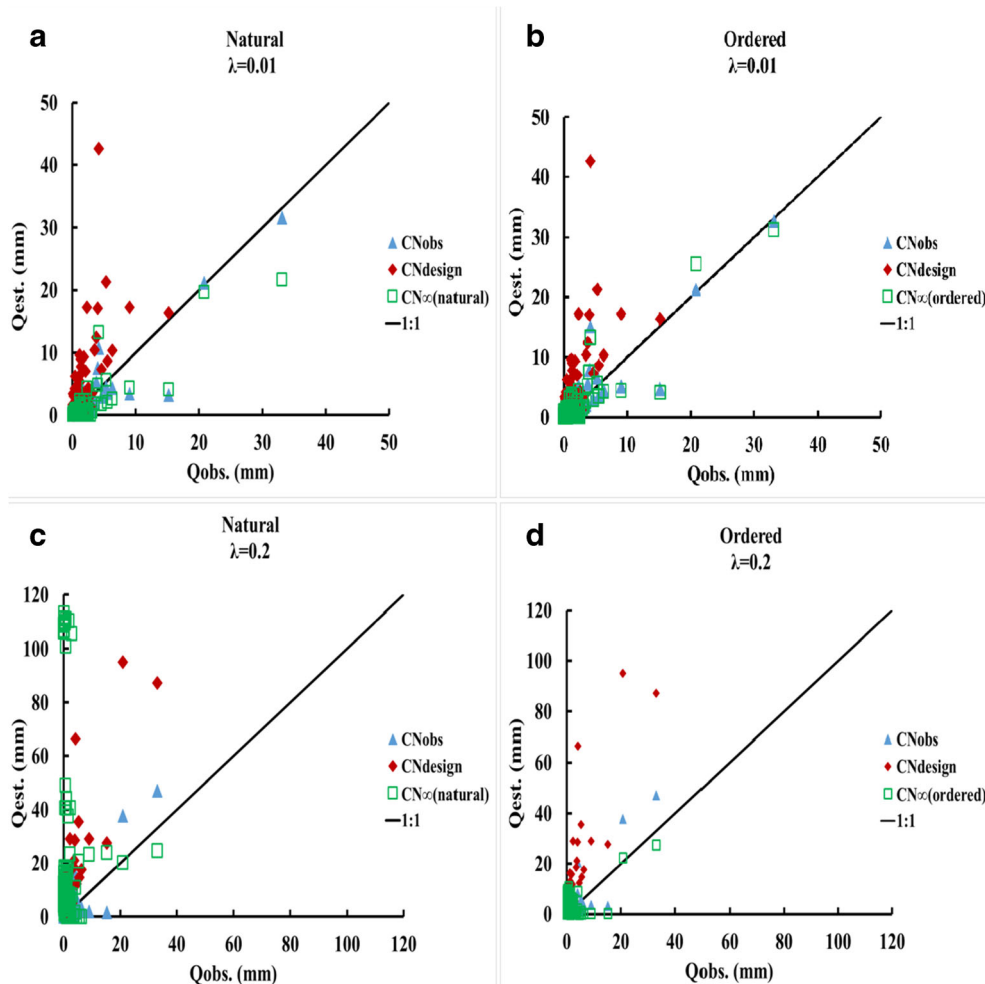


Fig. 10 Comparison between various methods for CN estimation (CN_{obs} , CN_{∞} , and CN_{design}) at $\lambda = 0.01$ and $\lambda = 0.2$. **a** Natural sorting of data pairs ($P: Q$) at $\lambda = 0.01$, **b** natural sorting of data pairs ($P: Q$) at $\lambda = 0.2$, **c**

ordered sorting of data pairs ($P: Q$) at $\lambda = 0.01$, and **d** ordered sorting of data pairs ($P: Q$) at $\lambda = 0.2$

Fig. 11 Validation of the various methods for CN estimation (CN_{obs} , CN_{∞} , and CN_{design}) at $\lambda = 0.01$ and $\lambda = 0.2$. **a** Natural sorting of data pairs ($P: Q$) at $\lambda = 0.01$, **b** ordered sorting of data pairs ($P: Q$) at $\lambda = 0.01$, **c** natural sorting of data pairs ($P: Q$) at $\lambda = 0.2$, and **d** ordered sorting of data pairs ($P: Q$) at $\lambda = 0.2$



Comparison of various methods of CN estimation

Figure 10 summarizes a comparison between the various estimation methods of CN, namely CN_{obs} , CN_{∞} , and CN_{design} , for the basins under ordered and natural sorting of $P: Q$ pairs and at $\lambda = 0.01$ and $\lambda = 0.2$. Figure 10a and b are for the natural sorting of data at $\lambda = 0.01$ and 0.2 respectively, while Figs. 10b and c are for ordered sorting of data at $\lambda = 0.01$ and 0.2 respectively. The figures show that CN_{design} is the highest for all basins for both values of λ and at both types of data sorting. The second highest value is CN_{obs} (that is based on

LSM), while the lowest value is for CN_{∞} (based on AFM). The design CN is made through the NRCS-CN table for $\lambda = 0.2$ which is commonly used by hydrologists and engineers worldwide. However, the design CN for $\lambda = 0.01$ is somewhat lower than the CN at $\lambda = 0.2$. The low values of CN_{design} for these arid basins are obtained from the developed equation (Eq. 15) which accounts for the transmission losses in these basins. It is also obvious that the values based on ordered sorting of data are relatively higher than the natural sorting of data for both values of λ and for both CN_{obs} and CN_{∞} . However, CN_{design} does not depend on the order because it

Table 9 Summary of the validation of the CN estimation methods. Comparison between estimated and observed runoff depth (mm)

CN type	Natural sorting				Ordered sorting			
	$\lambda = 0.01$		$\lambda = 0.2$		$\lambda = 0.01$		$\lambda = 0.2$	
	R^2	RMSE (mm)	R^2	RMSE (mm)	R^2	RMSE (mm)	R^2	RMSE (mm)
CN_{obs}	0.83	1.47	0.63	3.33	0.81	1.55	0.64	3.37
CN_{∞}	0.78	1.76	0.00	32.54	0.81	1.55	0.16	5.08
CN_{design}	0.74	5.89	0.72	10.42	0.74	5.89	0.72	10.42

is based on the NRCS table. These results show high variability between the various methods. However, the variability is relatively reduced in the case of ordered sorting of $P: Q$ data pairs, since ordering provides a relatively higher correlation between the $P: Q$ pairs. The ordered sorting is commonly used in consideration for design purposes, while the natural sorting of data is mainly used for natural simulation of the rainfall-runoff process in the basins.

Validation of various methods of CN estimation

Figure 11 shows a comparison between the observed and the estimated runoff depths (mm) using the values of CN estimated from the various methods (CN_{obs} , CN_{∞} , and CN_{design}) at both $\lambda = 0.2$ and 0.01 under the case of ordered and sorted data pairs ($P: Q$). The figure shows that CN_{design} overestimated the observed values for all cases. This indicates that the use of the NRCS-CN table (Table 3 at $CN_{(0.2)}$) at $\lambda = 0.2$ or the modified NRCS-CN table (Table 3 at $CN_{(0.01)}$) at $\lambda = 0.01$ produces high values for CN estimation that in turn leads to a safer design of the flood mitigation structures. However, CN_{obs} and CN_{∞} are around the line of perfect fit (1:1 line) for the case of $\lambda = 0.01$ under both sorted and ordered data points (see Figs. 11 a and b). This result suggests the use of both CN_{obs} and CN_{∞} for the simulation of the hydrological processes in the basins since it matches the observed runoff values. The case of $\lambda = 0.01$, in Figs. 11 c and d, shows however a highly scatter data especially at the natural sorting of data pairs ($P: Q$). This confirms that $\lambda = 0.2$ is not suitable to be used in the Saudi arid environment (Farran and Elfeki 2020c). Table 9 summarizes a quantification of Fig. 11. Both R^2 and RMSE are estimated for the aforementioned cases and the results are tabulated. The table indicated that the highest $R^2 = 0.83$ and minimum RMSE = 1.47 mm at CN_{obs} with $\lambda = 0.01$ at the natural sorting of data. While the lowest $R^2 = 0.0$ and a maximum RMSE = 32.54 mm at CN_{∞} with $\lambda = 0.2$ at the natural sorting of data. This result shows that CN_{∞} is not a good estimation of the CN in these arid basins which is also confirmed by Farran and Elfeki (2020b). The CN_{design} for both $\lambda = 0.2$ and $\lambda = 0.01$ shows relatively the same $R^2 = 0.72$ and 0.74 respectively; however, the RMSE is minimal at $\lambda = 0.01$ (5.98 mm) in comparison with $\lambda = 0.2$ (10.42 mm). CN_{design} does not depend on data sorting or ordering. This last result concludes that the NRCS-CN table at $\lambda = 0.01$ should be used in the design of mitigation structures rather than the NRCS-CN at $\lambda = 0.2$ since it reduces the RMSE.

Summary and conclusions

A comparison between various methods for CN estimation has been carried out. These methods are the NRCS-CN table

(CN_{design}), the least-squares method (LSM) which leads to the so-called CN_{obs} , and the asymptotic fitting method (AFM) which leads to obtaining the so-called asymptotic CN, CN_{∞} . A comparison between these methods is carried out under the effect of changing the coefficient of abstraction ratio, λ , and under the effect of sorting the data pairs ($P: Q$). Moreover, a methodology has also been developed to convert the NRCS-CN table values from $\lambda = 0.2$ to $\lambda = 0.01$ for arid basins. The following conclusion can be drawn from the study:

1. The values of CN_{design} , CN_{obs} , and CN_{∞} are relatively lower in the case of $\lambda = 0.01$ when compared with $\lambda = 0.2$. This reflects the transmission losses occurring in these basins.
2. The highest CN value obtained is the CN_{design} , then CN_{obs} , while CN_{∞} shows the lowest value. Therefore, for a safe design of the hydraulic structures, it is recommended to use CN_{design} . However, for the simulation of the rainfall-runoff process in the natural basins, it is recommended to use CN_{obs} at the natural sorting of data pairs ($P: Q$). CN_{∞} is not a good estimation of CN in the Saudi arid environment since the asymptotic value is not reached within the rainfall record in most of the basins.
3. The relationship between the observed CN and the NRCS-CN table shows that estimating runoff using $\lambda = 0.2$ is best made by CN of the impervious areas ($CN_{imp} = 90$) instead of $CN_{imp} = 98$ used in the literature. The RMSE of CN is reduced from 11 at $CN_{imp} = 98$ to 7 at $CN_{imp} = 90$. This value reflects the infiltration process in the rocks of the mountains (impervious region of the basin) due to the high density of fractures and fissures in these mountains.
4. The developed NRCS-CN table at $\lambda = 0.01$ should be used in the design of mitigation structures in the natural basins in the Saudi arid environment rather than the NRCS-CN at $\lambda = 0.2$ since the developed table reduces the RMSE in the estimated runoff by 57%.

The results of this research recommend the use of $\lambda = 0.01$ for flood studies in the Saudi arid environment for natural arid basins together with the developed NRCS-CN table at $\lambda = 0.01$.

Funding This work was supported by the Deanship of Scientific Research (DSR), King Abdulaziz University, Jeddah, under grant No. (DF-719-155-1441). The authors, therefore, gratefully acknowledge the DSR technical and financial support.

References

- Abdulrazzak M, Al-Shabani A, Noor K, Elfeki A, Kamis A (2017) Integrating hydrological and hydraulic modelling for flood risk management in a high resolution urbanized area: case study Taibah

- University campus, KSA. In: Paper presented at the Euro-Mediterranean Conference for Environmental Integration
- Abdulrazzak M, Elfeki A, Kamis AS, Kassab M, Alamri N, Noor K, Chaabani A (2018) The impact of rainfall distribution patterns on hydrological and hydraulic response in arid regions: case study Medina, Saudi Arabia. *Arab J Geosci* 11(21):679
- Abdulrazzak MJ, Oikonomou PD, Grigg NS (2020) Transboundary groundwater cooperation among countries of the Arabian Peninsula. *J Water Resour Plan Manag* 146(1):05019023
- Al Saud M (2010) Assessment of flood hazard of Jeddah area 2009, Saudi Arabia. *J Water Resour Prot* 2(9):839–847
- Alagha MO, Gutub SA, Elfeki AM (2016) Estimations of NRCS curve number from watershed morphometric parameters: a case study of Yiba watershed in Saudi Arabia. *Int J Civil Eng Technol* 7(2):247–265
- Al-Hasan AAS, Mattar YE-S (2014) Mean runoff coefficient estimation for ungauged streams in the Kingdom of Saudi Arabia. *Arab J Geosci* 7(5):2019–2029
- Al-Juaidi AE (2018) A simplified GIS-based SCS-CN method for the assessment of land-use change on runoff. *Arab J Geosci* 11(11):269
- Al-Subai K (1992) Erosion-sedimentation and seismic considerations for dam siting in the central Tihamat Asir region. Unpublished Ph. D. thesis, King Abdulaziz University, Faculty of Earth Sciences, Kingdom of Saudi Arabia, Jeddah, Saudi Arabia
- Baboo SS, Devi MR (2011) Geometric correction in recent high resolution satellite imagery: a case study in Coimbatore, Tamil Nadu. *Int J Comput Appl* 14(1):32–37
- Bales J, Betson R (1981) The curve number as a hydrologic index. *Rainfall Runoff Relationship* 371–386
- Bonta JV (1997) Determination of watershed curve number using derived distributions. *J Irrig Drain Eng* 123(1):28–36
- Briem GJ, Benediktsson JA, Sveinsson JR (2002) Multiple classifiers applied to multisource remote sensing data. *IEEE Trans Geosci Remote Sens* 40(10):2291–2299
- Cohen J (1960) A coefficient of agreement for nominal scales. *Educ Psychol Meas* 20(1):37–46
- Congalton RG (1991) A review of assessing the accuracy of classifications of remotely sensed data. *Remote Sens Environ* 37(1):35–46
- Congalton RG, Mead RA (1983) A quantitative method to test for consistency and correctness in photointerpretation. *Photogramm Eng Remote Sens* 49(1):69–74
- Congalton RG, Oderwald RG, Mead RA (1983) Assessing Landsat classification accuracy using discrete multivariate analysis statistical techniques. *Photogramm Eng Remote Sens* 49(12):1671–1678
- Cooley KR, Lane LJ (1982) Modified runoff curve numbers for sugarcane and pineapple fields in Hawaii. *J Soil Water Conserv* 37(5):295–298
- Curtis JG, Whalen PJ, Emenger RS, Reade DS (1983) A catalog of intermountain watershed curve numbers. Jupiter Pluvius Press (Utah State University, Watershed Science), in cooperation with State of Utah, Division of Oil, Gas, and Mining, p 58
- D'Asaro F, Grillone G (2012) Empirical investigation of curve number method parameters in the Mediterranean area. *J Hydrol Eng* 17(10):1141–1152
- Dawod GM, Koshak NA (2011) Developing GIS-based unit hydrographs for flood management in Makkah metropolitan area, Saudi Arabia. *J Geogr Inf Syst* 3(02):160–165
- Dawod GM, Mirza MN, Al-Ghamdi KA (2011) GIS-based spatial mapping of flash flood hazard in Makkah City, Saudi Arabia. *J Geogr Inf Syst* 3(3):217
- Dawod GM, Mirza MN, Al-Ghamdi KA (2013) Assessment of several flood estimation methodologies in Makkah metropolitan area, Saudi Arabia. *Arab J Geosci* 6(4):985–993
- Elfeki AM, Ewea HAR, Bahrawi JA, Al-Amri NS (2015) Incorporating transmission losses in flash flood routing in ephemeral streams by using the three-parameter Muskingum method. *Arab J Geosci* 8:5153–5165. <https://doi.org/10.1007/s12517-014-1511-y>
- Elfeki AM, Masoud M, Basahi J, Zaidi S (2020) A unified approach for hydrological modeling of arid catchments for flood hazards assessment: case study of Wadi Itwad, southwest of Saudi Arabia. *Arab J Geosci* 13(12):1–21. <https://doi.org/10.1007/s12517-020-05430-7>
- Elhag M (2016) Detection of temporal changes of eastern coast of Saudi Arabia for better natural resources management
- Elhag M, Boteva S (2016) Mediterranean land use and land cover classification assessment using high spatial resolution data. In: Paper presented at the IOP Conference Series, Earth and Environmental Science
- Farran MM, Elfeki AM (2020a) Statistical analysis of NRCS curve number (NRCS-CN) in arid basins based on historical data. *Arab J Geosci* 13(1):1–15. <https://doi.org/10.1007/s12517-019-4993-9>
- Farran MM, Elfeki AM (2020b) Variability of the asymptotic curve number in mountainous undeveloped arid basins based on historical data: case study in Saudi Arabia. *J Afr Earth Sci* 162:103697. <https://doi.org/10.1016/j.jafrearsci.2019.103697>
- Farran MM, Elfeki AM (2020c) Evaluation and validity of the antecedent moisture condition (AMC) of Natural Resources Conservation Service-Curve Number (NRCS-CN) procedure in undeveloped arid basins. *Arab J Geosci* 13:275 (2020). <https://doi.org/10.1007/s12517-020-5242-y>
- Fennessey L, Hawkins R (2001) The NRCS curve number, a new look at an old tool. Paper presented at the Proc. of Pennsylvania Stormwater Management Symp., Villanova Uni
- Hamad JT, Eshtawi TA, Abushaban AM, Habboub MO (2012) Modeling the impact of land-use change on water budget of Gaza Strip. *J Water Resource Prot* 4(06):325–333
- Hamdan SM, Troeger U, Nassar A (2007) Stormwater availability in the Gaza Strip, Palestine. *Int J Environ Health* 1(4):580–594
- Hanson CL, Neff EL, Doyle JT, Gilbert TL (1981) Runoff curve numbers for Northern Plains rangelands. *J Soil Water Conserv* 36(5):302–305
- Hawkins RH (1975) The importance of accurate curve numbers in the estimation of storm runoff. *J Am Water Resour Assoc* 11(5):887–891
- Hawkins RH (1993) Asymptotic determination of runoff curve numbers from data. *J Irrig Drain Eng* 119(2):334–345
- Hawkins RH, Jiang R, Woodward D (2002) Application of curve number method in watershed hydrology. In: Paper presented at the Presentation at American Water Resources Association Annual Meeting, Albuquerque NM
- Hawkins, R. H., Ward, T. J., Woodward, D. E., Van Mullem, J. A. (2008). Curve number hydrology: state of the practice
- Hjelmfelt AT (1980) Empirical investigation of curve number technique. *J Hydraul Div* 106(9):1471–1476
- Hjelmfelt AT, Woodward D, Conaway G, Plummer A, Quan Q, Van Mullen J, Rietz P (2001) Curve numbers, recent developments. In: Paper presented at the proceedings of the congress-international association for hydraulic research
- Wheater HS, Butler AP, Stewart EJ, Hamilton GS (1991) A multivariate spatial-temporal model of rainfall in S.W. Saudi Arabia. I. Data characteristics and model formulation. *J Hydrol* 125:175–199
- Isik S, Kalin L, Schoonover JE, Srivastava P, Lockaby BG (2013) Modeling effects of changing land use/cover on daily streamflow: an artificial neural network and curve number based hybrid approach. *J Hydrol* 485:103–112
- Itten KI, Meyer P (1993) Geometric and radiometric correction of TM data of mountainous forested areas. *IEEE Trans Geosci Remote Sens* 31(4):764–770
- Kloer BR (1994) Hybrid parametric/non-parametric image classification. In: Paper presented at the paper to be presented at the ACSM-ASPRS Annual Convention. April, Reno, Nevada

- Kottegoda N, Natale L, Raiteri E (2000) Statistical modelling of daily streamflows using rainfall input and curve number technique. *J Hydrol* 234(3–4):170–186
- Lin Y-P, Hong N-M, Wu P-J, Wu C-F, Verburg PH (2007) Impacts of land use change scenarios on hydrology and land use patterns in the Wu-Tu watershed in Northern Taiwan. *Landsc Urban Plan* 80(1–2): 111–126
- Mahmoud SH, Mohammad F, Alazba A (2014) Determination of potential runoff coefficient for Al-Baha region, Saudi Arabia using GIS. *Arab J Geosci* 7(5):2041–2057
- Metwaly M, El-Awadi E, Al-Arifi N (2010) Flooding risk analysis of the central part of western Saudi Arabia using remote sensing data. Paper presented at the Proceedings of the Fifth National GIS Symposium in Saudi Arabia. Al-Khobar, April
- Mishra SK, Singh V (2013) Soil conservation service curve number (SCS-CN) methodology (vol 42). Springer Science & Business Media
- Pettorelli N, Vik JO, Mysterud A, Gaillard J-M, Tucker CJ, Stenseth NC (2005) Using the satellite-derived NDVI to assess ecological responses to environmental change. *Trends Ecol Evol* 20(9):503–510
- Richards JA, Jia X (1999) Remote sensing digital image analysis (vol 3). Springer, Berlin
- Salehi A, Al-Hatrushi S (2009) Torrential flood hazards assessment, management, and mitigation, in Wadi Aday, Muscat area, Sultanate Of Oman, a GIS and RS approach. *Egypt J Remote Sens Space Sci* 12(2009):71–86
- Saudi Arabian Dames, Moore (1988) Representative basins study for Wadi: Yiba, Habwnah, Tabalah, Liyyah and Al-Lith (main report) Kingdom of Saudi Arabia, Ministry of Agriculture and Water, Water Resource Development Department
- Şen Z, Al-Suba'i K (2002) Hydrological considerations for dam siting in arid regions: a Saudi Arabian study. *Hydrol Sci J* 47(2):173–186
- Shadeed S, Almasri M (2010) Application of GIS-based SCS-CN method in West Bank catchments, Palestine. *Water Sci Eng* 3(1):1–13
- Shih SF (1988) Satellite data and geographic information system for land use classification. *J Irrig Drain Eng* 114(3):505–519
- Simanton J, Renard K, Sutter N (1973) Procedure for identifying parameters affecting storm runoff volumes in a semiarid environment. *USDA Agr Res*
- Still D, Shih S (1991) Satellite data and geographic information system in runoff curve number prediction. In: Paper presented at the Proceeding of the International Conference on Computer Application in Water Resources. Taipei, Taiwan, ROC
- Stuebe MM, Johnston DM (1990) Runoff volume estimation using GIS techniques 1. *JAWRA Journal of the American Water Resources Association* 26(4):611–620
- Subudhi A, Sharma N, Mishra D (1989) Use of Landsat thematic mapper for urban land use/land cover mapping. *J Indian Soc Remote Sens*
- Swain PH, Davis SM (1981) Remote sensing: the quantitative approach. *IEEE Trans Pattern Anal Mach Intell* (6):713–714
- USDA-NRCS (2004) National Engineering Handbook: section 4: hydrology, soil conservation service. USDA, Washington, DC
- USDA-SCS (1985) National Engineering Handbook Section 4: hydrology, soil conservation service. USDA, Washington, DC
- Walker, C. H. (1970). Estimating the rainfall-runoff characteristics of selected small Utah watersheds
- Youssef AM, Pradhan B, Hassan AM (2011) Flash flood risk estimation along the St. Katherine road, southern Sinai, Egypt using GIS based morphometry and satellite imagery. *Environ Earth Sci* 62(3):611–623