# Development of cloud-based rainfall–run-off model using Google Earth Engine

**Sukant Jain[1,*], R. K. Jaiswal[1], A. K. Lohani[2] and Ravi Galkate[1]**

[1]Central India Hydrology Regional Center, National Institute of Hydrology, Bhopal 462 016, India
[2]National Institute of Hydrology, Jal Vigyan Bhavan, Roorkee 247 667, India

**The capability of open-data sources in cloud computing was explored in rainfall–run-off modelling through the Soil Conservation Services curve number (SCS CN) model. The Google Earth Engine (GEE) has a petabytes catalogue of global remote sensing and GIS datasets, numerous functions and algorithms to manipulate and visualize datasets rapidly. In this study, an algorithm has been developed to prepare dynamic CN maps in GEE using OpenLandMap Soil Texture and MODIS land use/land cover (LULC) data through ternary function and climate hazards group infrared precipitation rainfall collection for input rainfall and creation of antecedent moisture condition. The capabilities of the developed algorithm were demonstrated for Shipra, Kuttiyadi and Bah river catchments in India. However, it can be used with different satellite data for estimating the run-off and impact of LULC change on run-off for any part of the world and any desired period. The developed algorithm not only utilizes GEF and the public archive database for estimating the run-off at basin/subbasin scales for the planning of water resources, but also provides a quick evaluation of the impact of LULC change on run-off over time.**

**Keywords:** Cloud computing platform, curve number, river basin, rainfall–run-off modelling, water resources.

RUN-OFF is the most important hydrological component and a variety of models have been developed, which range from empirical to conceptual and physically based models for planning water resources. Binnie's percentage is one of the oldest documented rainfall–run-off relations developed in 1872 for Indian watersheds. In 1912, Barlow derived a similar coefficient which accounts for different classes of catchment based on their physical and topographical attributes. Similar to these, other relationships like Strange tables, Inglis and Desouza formula, Khosla's formula, etc. were developed to understand the rainfall–run-off relationships at the regional scale.

In 1933, the Soil Conservation Services (SCS) were formed and entrusted to set up soil conservation practices for projects. In the late 40s, Sherman[1] proposed the idea of setting up a relation between rainfall and run-off by unit hydrograph theory. Later Mockus[2] suggested the inclusion of parameters like land use/land cover (LULC) type, soil, antecedent rainfall, spatial distribution, etc. to estimate the surface run-off. Based on infiltration rate, Musgrave[3] grouped soils into four classes or groups, namely A, B, C and D. This classification was a considerable achievement for the SCS curve number (CN) method[4] that provided an empirical relationship to estimate run-off which considers initial abstraction based on LULC and soil type known as SCS CN method. Due to its simplicity, it became one of the most popular models to estimate run-off[5,6]. Some of the widely used models like SWAT[7] and HEC-HMS[8] also use the CN method to estimate surface run-off. Originally, the SCS model was developed for estimation of direct surface run-off from small rainfall events in small watersheds in USA. Later, it was endorsed for several regions, climatic conditions and land uses[9,10]. Over the past decades, due to rapid urbanization the LULC pattern has drastically changed that influences CN and hydrological parameters influencing run-off from the catchment[11]. Advancement in geospatial technologies helps include these parameters for estimating run-off spatially and temporally. Various remote sensing products from satellites like IRS, Landsat, NOAA, TRMM, etc. can be used to obtain reliable input parameters like rainfall, land use, soil type, etc. from regional to global scale[12].

Several tools, software have been developed to capture and process remote sensing and GIS data for rainfall–run-off modelling. However, either they are expensive or require large computer hardware support[13]. Few free and open-source platforms are available for water resources studies[14], but require advanced skills. Presently no model can include the temporal changes or dynamics of LULC change into the process, and changes can only be accommodated by repeating the entire simulation process. This problem becomes more severe if the data to be analysed are spatially and temporally larger or have multiple parameters.

The lack of methodology for handling, computing, collaborating and defining large spatial datasets is being felt for a long time, and the development of safe, reliable, efficient, and advance cloud-computing platforms and methodologies are the need of the hour[15,16]. There are some efficient cloud-computing platforms available to store, access and analyse large datasets, like AWS (Amazon Web Services) and Azure, which are paid services. AWS

*For correspondence. (e-mail: sukantjain1@gmail.com)

provides a dedicated remote-sensing platform containing numerous datasets and machine learning algorithms. Azure by Microsoft is an artificial intelligence-based cloud-computing platform.

Google Earth Engine (GEE)[17] is another cloud-based computing platform that combines the computational infrastructure of Google and openly available remote sensing and GIS datasets. Any user can access GEE via any web browser. With automatic parallel processing, it can efficiently handle big data and is freely available to all. For example, it takes 100 h to process 654,178 Landsat-7 scenes (707 terabytes data) to produce a global forest map without downloading any datasets or installing any software[18].

In the hydrology domain, GEE is being used for several purposes, including monitoring surface water dynamics[19], reservoir or lake monitoring[20], coastal studies[21], snow and glacier studies[22], etc. A literature review shows that no study has been carried out till now to estimate the surface run-off using big data on GEE. To overcome this, it was proposed to employ GEE to estimate surface run-off using the SCS CN method for any catchment considering the LULC dynamics. The present study demonstrates the use of public archive database and geospatial cloud computation technology of GEE (written in JavaScript) for estimating the run-off at basin/sub-basin scales through a CN-based rainfall–run-off algorithm developed in JavaScript. The developed algorithm can be applied to any river basin for a quick estimate of run-off and three catchments have been selected for demonstration.

## Study area

The CN method is widely used for medium and small ungauged watersheds that require minimum parameters. The changes in LULC and heterogeneity of soils in the watershed can be incorporated due to the availability of a large dataset on the cloud in GEE, which has the power of cloud computing and the availability of a huge data repository[23]. The performance of the developed code for computation of surface run-off in GEE has been applied to three catchments, i.e. Shipra, Bah and Kuttiyadi. Shipra river originates from the Kakribardi hills in the Vindhya range and flows north across the Malwa Plateau to join the Chambal River in Madhya Pradesh. Observed discharge data of Shipra basin up to Ujjain (catchment area: 2095 sq. km) was used to assess the performance of the model. Kuttiyadi river originates from South Wayanad and joins the Arabian sea near Badayan town in Kerala. Results were compared with records of the Central Water Commission discharge site at Kuttiyadi in Kozhikode district of Kerala (catchment area: 238 sq. km). Bah river (catchment area: 896 sq. km) is the tributary of Betwa, which originates from Bhopal and a discharge site is located in Vidisha district, Madhya Pradesh.

## Materials and methodology

To develop the SCS CN model using cloud data and the GEE server, different sources of data availability were examined. The methodology was finalized with the help of a flowchart to use dynamic LULC (MCD12Q1.006 MODIS Land Cover)[24], rainfall (Climate Hazards Group InfraRed Precipitation (CHIRPS) with Station Data)[25], Global Soil data (OpenLandMap)[26], and other data for SCS CN model (Figure 1).

### *SCS CN model*

The traditional CN method is an event-based, lumped rainfall–run-off model and based on the combination of the following water budget equation

$$P = I_a + F + Q. \tag{1}$$

Two proportional equality hypotheses are expressed as as follows

$$\frac{Q}{P - I_a} = \frac{F}{S}, \tag{2}$$

$$I_a = \lambda S, \tag{3}$$

where $P$ is the daily rainfall, $I_a$ the initial abstraction, $F$ the actual retention, $Q$ the direct surface run-off, $S$ the potential maximum retention and $\lambda$ is the initial abstraction coefficient. Combining the above two equations, the popular SCS CN equation is given below.

$$Q = \frac{(P - I_a)^2}{P - I_a + S} \text{ for } P > I_a, \tag{4}$$

$$Q = 0 \text{ for } P \le I_a,$$

where $S$ is a function of CN and can be computed using the following equation

$$S = 25.4\left(\frac{1000}{\text{CN}} - 10\right), \tag{5}$$

where $S$ is expressed in millimetres and CN is a dimensionless run-off coefficient that depends on land use, soil and antecedent moisture condition (AMC)[27].

Antecedent moisture is the relative dryness or wetness of a catchment which changes continuously and has a significant effect on the run-off process[28]. AMC can be divided into three classes. AMC I is considered for the dry condition with five-day antecedent rainfall, i.e. AMC is less than 13 mm. When AMC is more than 28 mm, it may be a wet condition (AMC III) and when 13 mm ≤

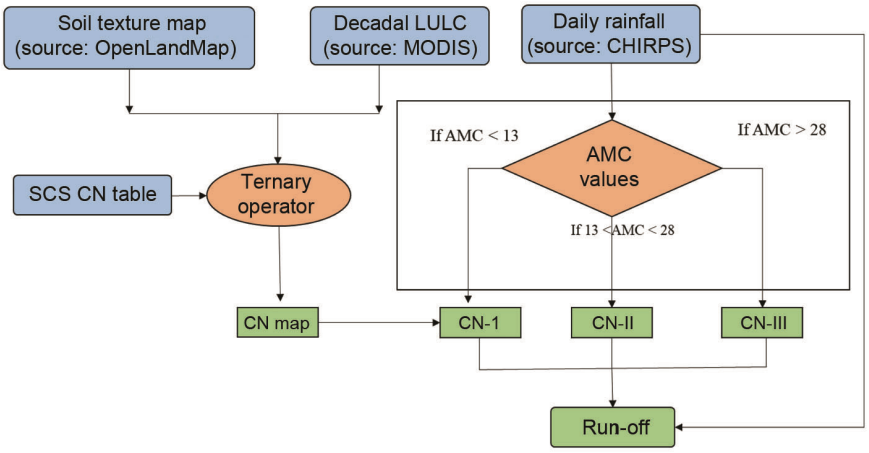


**Figure 1.** Flowchart of workflow.


**Figure 2.** Components of Earth Engine Editor.

AMC < 28 mm, it can be considered as average (AMC II)[29]. For AMC II CNs have been proposed based on a combination of LULC and soil group conditions. For AMC I and AMC III, CN can be derived by the following equations

$$\mathrm{CN(I)} = \frac{\mathrm{CN(II)}}{2.281 - 0.0128\mathrm{CN(II)}}, \tag{6}$$

$$\mathrm{CN(III)} = \frac{\mathrm{CN(II)}}{0.427 - 0.00573\mathrm{CN(II)}}, \tag{7}$$

## *Google Earth Engine*

GEE is a cloud-based platform that can be accessed through the web for planetary-scale geospatial analysis to solve a variety of high-impact societal issues[17]. Figure 2 presents the components of the Earth Engine Editor.

GEE is a state-of-the-art technology to perform LULC mapping, disaster management, and earth sciences applications on a global scale[30–32]. However, no studies are available in the literature on rainfall–run-off modelling using GEE.

Figure 3 shows a simple version of GEE system architecture. The Earth Engine Code Editor uses Python and JavaScript programming languages to interact with the system through REST API. The request is handled by front-end servers which forward the complex sub-queries to a pool of servers parallelly to get the outputs.

## *SCS CN model in GEE*

With the huge pool of spatial data available in GEE, the selection of input data becomes more convenient. Numerous filter functions help the users to shortlist their datasets from huge sets of image collections. GEE is also

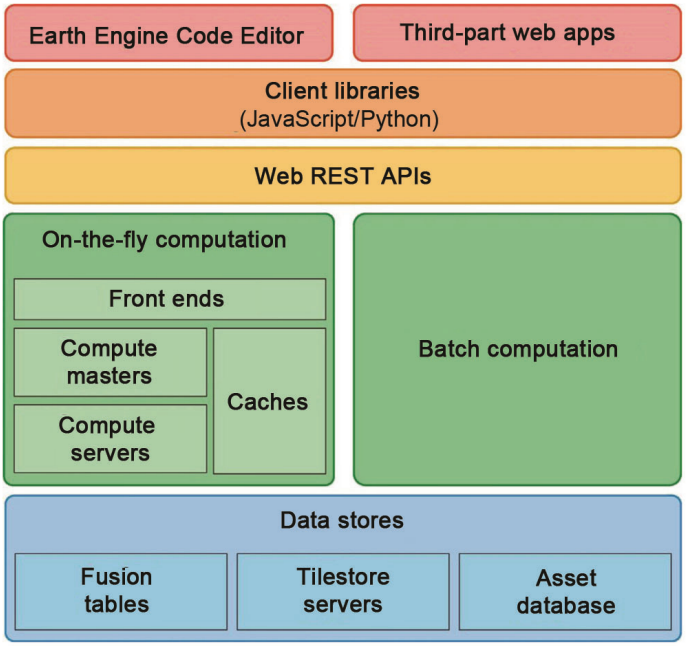


**Figure 3.** A simplified system architecture diagram[17].

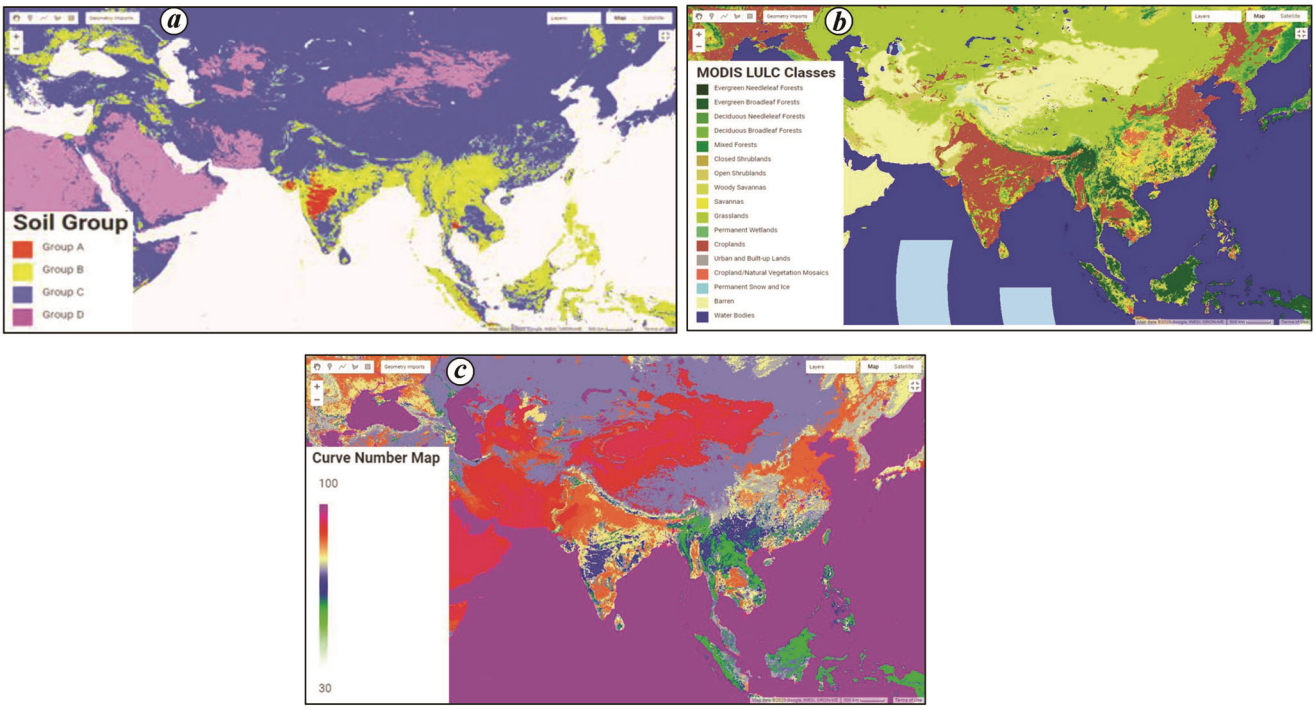


**Figure 4.** ***a***, Global soil hydrologic group map. ***b***, Global modis land use/land cover. ***c***, Global curve number II map (outputs obtained from GEE).

capable of handling dynamic data, which is otherwise not possible using the presently available models. First, the soil texture map is converted to four hydrologic soil groups, A, B, C and D using the ternary operator in image expression. Then it is added to the downscaled LULC MODIS data as a second band. The curve number (CN II) map is then prepared using the conditional statement for all the combinations of 17 MODIS LULC data classes and four soil groups. Based on CN II, the CN I and CN III maps are generated referring to eqs (6) and (7). The $S$ image is the function of the CN image and depends on AMC. Hence, the $S$ images corresponding to CN maps

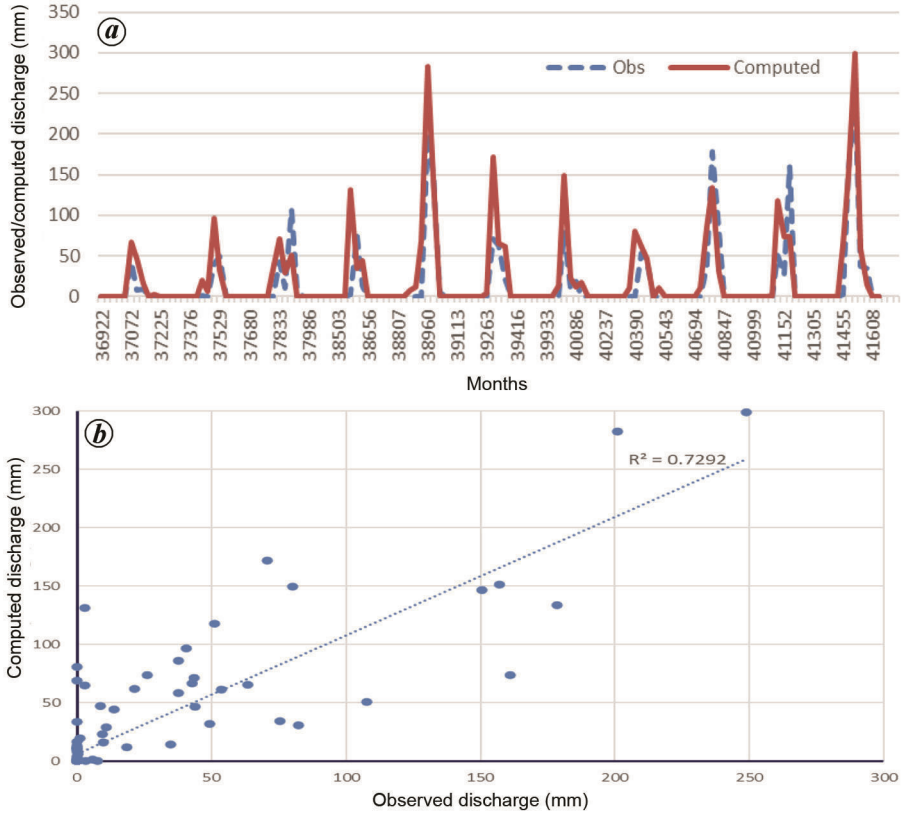


**Figure 5.** ***a***, Monthly observed versus computed discharge time series at Shipra river. ***b***, Scatter plot for observed versus computed monthly discharge data.

are created during the final $Q$ calculation. To reduce computational time for all CN conditions, $S$ images are created as a global variable in the script.

Daily AMC images are created using daily rainfall data images. The sum of the previous five days, including same-day rainfall is used to develop an AMC collection for each day of the study period. To estimate run-off ($Q$), a function based on eq. (4) is defined. Initially, the range of AMC values for each pixel and each day is checked and the S-II image is created for CN-II as the base image. Pixels having AMC less than 13 mm are replaced by the S-I image computed from CN I and those having AMC greater than or equal to 28 mm are replaced by the S-III image computed from CN III. All the three images are merged to get a single $S$ image for each day according to the AMC condition for all pixels. After applying the equation for computation of run-off, another condition is checked – if rainfall of a single pixel on a single day is less than $I_a$, the resulted run-off will be zero; otherwise, the run-off is estimated by a previously defined function. This function is mapped over the entire collection of AMC and rainfall to get the run-off images of each date.

To visualize the plot of rainfall and run-off in a single graph, both parameters are combined in a single image for each date. To visualize the image, the sum image of rainfall and run-off is produced through one of the reducer functions, i.e. imagecollection.sum() and the final time-series graph is produced using the image.series() function. The inspector tab can be used to visualize the time series of any pixel.

## Result and discussion

The rainfall–run-off model based on the CN method was developed in GEE, a cloud-computing platform. All the basic datasets, i.e. soil maps, LULC maps and rainfall data were imported in GEE through JavaScript API. Initially, the soil texture map was converted to a soil group map (Figure 4 *a*) and combined with LULC (Figure 4 *b*) to prepare the CN II maps (Figure 4 *c*). AMC data were defined by the summation of rainfall of the previous five days. Finally, the daily run-off was derived based on AMC conditions of each pixel. The model is capable of computing surface run-off from any extent of area, but for demonstration the run-off was computed using rainfall data of Shipra, Bah and Kuttiyadi rivers situated in different parts of India. The rainfall and run-off have shown a similar trend, i.e. both started increasing in the middle of June, were maximum in August and started declining by October end.

Using the inspector tool, the graph can be plotted at any pixel to check the response of CN over any specific

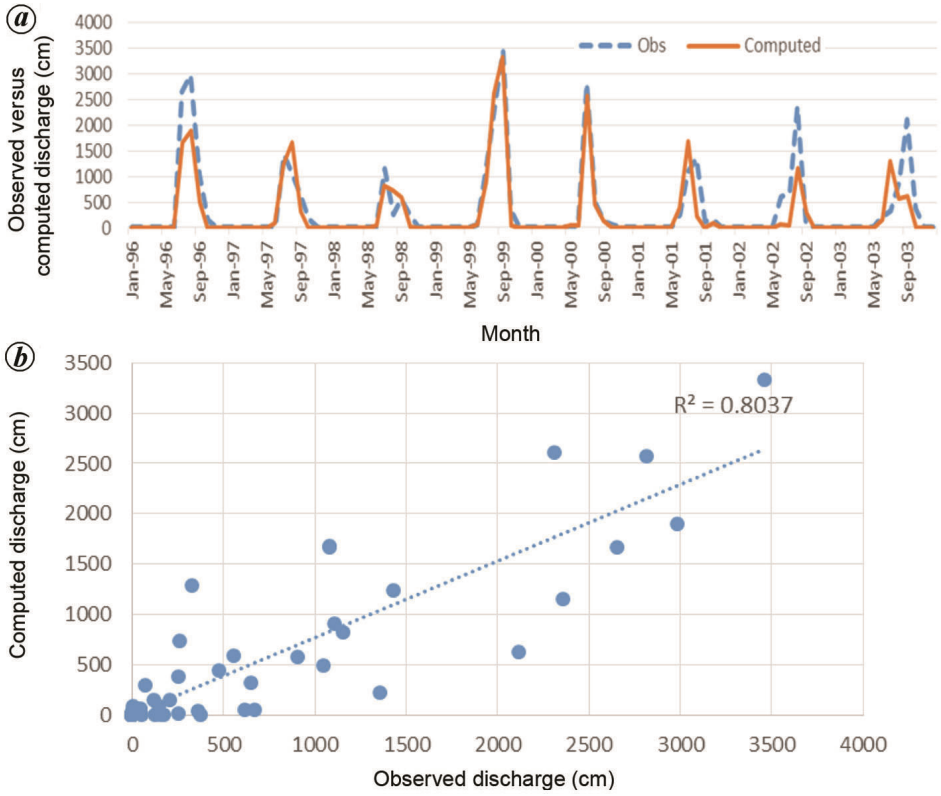


**Figure 6.** ***a***, Monthly observed versus computed discharge time series at Bah river. ***b***, Scatter plot for observed versus computed monthly discharge data.

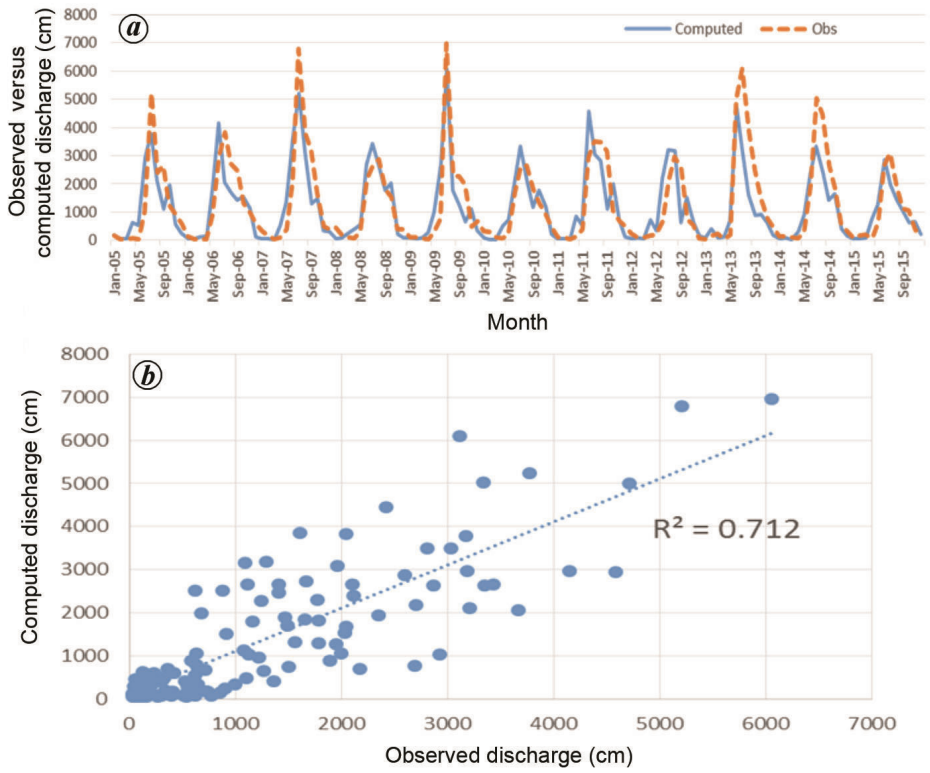


**Figure 7.** ***a***, Monthly observed versus computed discharge time series at Kuttiyadi river. ***b***, Scatter plot for observed versus computed monthly discharge data.

**Table 1.** Performance of the developed model based on efficiency criteria

| Nash–Sutcliff efficiency | Performance |
|---|---|
| 0.75–1.0 | Very good |
| 0.65–0.75 | Good |
| 0.50–0.65 | Satisfactory |
| 0.40–0.50 | Acceptable |
| Less than 0.40 | Unsatisfactory |

LU class. The results showed less run-off over forest areas due to more initial abstraction, whereas maximum amount of rainfall was converted to run-off in urban areas due to high CN value. For water bodies, all the rainfall is treated as run-off because CN for a water body is considered as 100. Results can also be visually enhanced through mapping of different products like rainfall or run-off during a specific period.

The performance of the simulated model was examined with observed surface run-off. The straight-line method was used to remove base flow. The developed CN model does not incorporate routing to estimate run-off. Therefore, daily run-off data of all the catchments were upscaled to monthly data, and goodness-of-fit parameters like the coefficient of determination ($R^2$) and Nash–Sutcliff efficiency were estimated with the observed data. Figures 5–7 show the observed versus simulated discharge plots of Shipra, Bah and Kuttiyadi rivers respectively. The $R^2$ and Nash–Sutcliff efficiency were found to be 0.73 and 59.2% for Shipra, 0.80 and 79.3% for Bah and 0.71 and 70.7% for Kuttiyadi rivers respectively. The performance of the SCS CN model based on Table 1 can be categorized as satisfactory for Shipra, very good for Bah, and good for Kuttiyadi basins.

## Conclusion

GEE has emerged as a powerful tool for water resources management and hydrological studies due to its handling capacity of a large sets of data through cloud computing. The presently available rainfall–run-off models and technology executed on the local computer use only static datasets for LULC, topography and soils, and are not capable of handling large datasets. They also require more time to assimilate the dynamics of changing temporal and spatial data. The methodology presented here enables the use of a larger set of dynamic data as input to a well-known rainfall–run-off model (SCS CN model) suitable for gauged as well as ungauged catchments and regions across the globe. The developed code can model rainfall–run-off relationships on a daily timescale for medium catchments, but the routing module is not included in the algorithm. It can be used for coarse time intervals like monthly, which is more useful for planning instead of daily. The developed model for the computation of run-off can be used for quick assessment of the rainfall–run-off process and the impact of LU change on run-off for any region of the world. It will be a handy tool for policymakers to make informed decisions on water availability to cater to the huge population for domestic use and the economics of our nation dominated by agriculture. The developed algorithm has all the limitations of the SCS technique like inability to model snow-fed rivers and large catchments. It can be accessed via the web link https://github.com/sukantjain/Google-Earth-Engine/blob/main/JavaScript/CurveNumberModel.js


1. Sherman, L. K., The unit hydrograph method. *Physics of the Earth*, 1949, 514–525.
2. Mockus, V., *Estimation of Total Surface Runoff for Individual Storms*, Exhibit A of Appendix B, Interim Survey Report Grand (Neosho) River Watershed, USDA, 1 December 1949.
3. Musgrave, G. W., How Much of the Rain Enters the Soil? *Yearbook of Agriculture 1955: Water*, USDA, Washington, DC, USA, 1955.
4. Woodward, D. E., Hawkins, R. H. and Quan, Q. D., Curve number method: origins, applications and limitations. In Hydrologic Modeling for the 21st Century: 2nd Federal Interagency Hydrologic Modeling Conference, Las Vegas, NV, USA, 2002.
5. Fu, S., Zhang, G. Wang, N. and Luo, L., Initial abstraction ratio in the SCS-CN method in the Loess Plateau of China. *Trans. ASABE*, 2011, **54**, 163–169.
6. Verma, S., Verma, R. K., Mishra, S. K., Singh, A. and Jayaraj, G. K., A revisit of NRCS-CN inspired models coupled with RS and GIS for run-off estimation. *Hydrol. Sci. J.*, 2017, **62**, 1891–1930.
7. Arnold, J. G. and Allen, P. M., Estimating hydrologic budgets for three Illinois watersheds. *J. Hydrol.*, 1996, **176**, 57–77.
8. Anderson, M. L., Chen, Z.-Q., Kavvas, M. L. and Feldman, A., Coupling HEC-HMS with atmospheric models for prediction of watershed run-off. *J. Hydrol. Eng.*, 2002, **7**, 312–318.
9. Romero, P., Castro, G., Gómez, J. A. and Fereres, E., Curve number values for olive orchards under different soil management. *Soil Sci. Soc. Am. J.*, 2007, **71**, 1758–1769.
10. Soulis, K. X. and Valiantzas, J. D., Identification of the SCS-CN parameter spatial distribution using rainfall-run-off data in heterogeneous watersheds. *Water Resour. Manage.*, 2013, **27**, 1737–1749.
11. Garg, V., Nikam, B. R., Thakur, P. K., Aggarwal, S. P., Gupta, P. K. and Srivastav, S. K., Human-induced land use land cover change and its impact on hydrology. *HydroResearch*, 2019, **1**, 48–56.
12. Schultz, G. A., Remote sensing in hydrology. *J. Hydrol.*, 1988, **100**, 239–265.
13. Ranade, R., Garg, A., Jain, S. and Pandey, K., Satellite Image Enhancement Toolbox (SIET) – an open source image enhancement implementation. In *Open Source Geospatial Tools in Climate Change Research and Natural Resources Management* (ed. OSGEO-India), 2015, pp. 8–13.
14. Dixit, A., Thakur, P. K., Aggarwal, S. P. and Jain, S., Open source geospatial tools for hydrological modeling. In OSGEO – India: FOSS4G 2015 – Second National Conference 'Open Source Geospatial Tools in Climate Change Research and Natural Resources Management', Roorkee, 8–10 June 2015, pp. 1–5.
15. Chi, M., Plaza, A., Benediktsson, J. A., Sun, Z., Shen, J. and Zhu, Y., Big data for remote sensing: challenges and opportunities. *Proc. IEEE*, 2016, **104**, 2207–2219.
16. Ma, Y. *et al.*, Remote sensing big data computing: challenges and opportunities. *Future Gener. Comput. Syst.*, 2015, **51**, 47–60.
17. Gorelick, N., Hancher, M., Dixon, M., Ilyushchenko, S., Thau, D. and Moore, R., Google Earth Engine: planetary-scale geospatial

analysis for everyone. *Remote Sensing Environ.*, 2017, **202**, 18–27.
18. Hansen, M. C. *et al*., High-resolution global maps of 21st century forest cover change. *Science*, 2013, **342**, 850–853.
19. Nguyen, U. N. T., Pham, L. T. H. and Dang, T. D., An automatic water detection approach using Landsat 8 OLI and Google Earth Engine cloud computing to map lakes and reservoirs in New Zealand. *Environ. Monit. Assess.*, 2019, **191**, 235.
20. Murphy, S., Wright, R. and Rouwet, D., Color and temperature of the crater lakes at Kelimutu volcano through time. *Bull. Volcanol.*, 2018, **80**, 2.
21. Vos, K., Harley, M. D., Splinter, K. D., Simmons, J. A. and Turner, I. L., Sub-annual to multi-decadal shoreline variability from publicly available satellite imagery. *Coast. Eng.*, 2019, **150**, 160–174.
22. Snapir, B., Momblanch, A., Jain, S. K., Waine, T. W. and Holman, I. P., A method for monthly mapping of wet and dry snow using Sentinel-1 and MODIS: application to a Himalayan river basin. *Int. J. Appl. Earth Obs. Geoinf.*, 2019, **74**, 222–230.
23. Chatterjee, C., Jha, R., Lohani, A. K., Kumar, R. and Singh, R., Runoff curve number estimation for a basin using remote sensing and GIS. *Asian-Pac. Remote Sensing GIS J.*, 2001, **14**, 1–7.
24. Friedl, M. and Sulla-Menashe, D., MCD12Q1 MODIS/Terra+ aqua land cover type yearly L3 global 500 m SIN grid V006 (data set). NASA EOSDIS L. Process. DAAC, 2015, **10**; https://doi.org/10.5067/MODIS/MCD12Q1.006.
25. Funk, C. *et al.*, The climate hazards infrared precipitation with stations – a new environmental record for monitoring extremes. *Sci. Data*, 2015, **2**, 150066.
26. Hengl, T., Soil texture classes (USDA system) for 6 soil depths (0, 10, 30, 60, 100 and 200 cm) at 250 m, 2018; 10.5281/zenode.1475451.
27. Amutha, R. and Porchelvan, P., Estimation of surface run-off in Malattar sub-watershed using SCS-CN method. *J. Indian Soc. Remote Sensing*, 2009, **37**, 291–304.
28. Mishra, S. K. and Singh, V. P., A relook at NEH-4 curve number data and antecedent moisture condition criteria. *Hydrol. Process.*, 2006, **20**, 2755–2768.
29. Hjelmfelt, Jr A. T., Kramer, K. A. and Burwell, R. E., Curve numbers as random variables. *Rainfall-Runoff Relationships*, Water Resources Publication, Littleton, CO, USA, 1982, pp. 365–370.
30. Goldblatt, R., You, W., Hanson, G. and Khandelwal, A. K., Detecting the boundaries of urban areas in India: a dataset for pixel-based image classification in Google Earth Engine. *Remote Sensing*, 2016, **8**.
31. Campos-Taberner, M. *et al.*, Global estimation of biophysical variables from Google Earth Engine platform. *Remote Sensing*, 2018, **10**, 1167.
32. Aguilar, R., Zurita-Milla, R., Izquierdo-Verdiguier, E. and Rolf A. de By, A cloud-based multi-temporal ensemble classifier to map smallholder farming systems. *Remote Sensing*, 2018, **10**, 729.

Received 31 July 2020; revised accepted 8 October 2021

doi: 10.18520/cs/v121/i11/1433-1440