Building a Flood-Warning Framework for Ungauged Locations Using Low Resolution, Open-Access Remotely Sensed Surface Soil Moisture, Precipitation, Soil, and Topographic Information

Seokhyeon Kim, Kyungrock Paik, Fiona M. Johnson, and Ashish Sharma

***Abstract*—Soil moisture (SM) plays an important role in determining the antecedent condition of a watershed, while topographic attributes define how and where SM and rainfall interact to create floods. Based on this principle, we present a method to identify flood risk at a location in a watershed by using remotely sensed SM and open-access information on rainfall, soil properties, and topography. The method consists of three hydrologic modules that represent the generation, transfer, and accumulation of direct runoff. To simplify the modeling and provide timely warnings, the flood risk is ascertained based on frequency of exceedance, with warnings issued if above a specified threshold. The simplicity of the method is highlighted by the use of only three parameters for each watershed of interest, with effective regionalization allowing use in ungauged watersheds. For this proof-of-concept study, the proposed model was calibrated and tested for 65 hydrologic reference stations in the Murray–Darling Basin in Australia over a 35-year study period by using satellite-derived surface SM. The three model parameters were first estimated using the first ten-year data and then the model performance was evaluated through flood threshold exceedance analyses over the remaining 25-year study period. The results for estimated parameters and skill scores showed promise. The three model parameters can be regionalized as a function of watershed characteristics, and/or representative values estimated from neighboring watersheds, allowing use in ungauged basins everywhere.**

Manuscript received December 3, 2017; revised December 29, 2017; accepted January 2, 2018. Date of publication January 30, 2018; date of current version February 12, 2018. This work was undertaken as part of a Discovery Project (DP140102394) funded by the Australian Research Council. The work of S. Kim was supported by the University of New South Wales Tuition Fee Scholarship and School Postdoctoral Writing Fellowship. The work of K. Paik was supported by the Korea University Grant. *(Corresponding author: Ashish Sharma.)*

S. Kim, F. M. Johnson, and A. Sharma are with the School of Civil and Environmental Engineering, University of New South Wales, Sydney, NSW 2052, Australia (e-mail: seokhyeon.kim@unsw.edu.au; f.johnson@unsw.edu.au; a.sharma@unsw.edu.au).

K. Paik is with the School of Civil, Environmental, and Architectural Engineering, Korea University, Seoul 02841, South Korea (e-mail: paik@korea.ac.kr).

This paper has supplementary downloadable material available at http://ieeexplore.ieee.org. The following supplementary materials contain Table S1 showing a list of the 65 hydrologic reference stations (HRS) in the Murray-Darling basin (MDB) used in this study, along with descriptions of a set of scripts written in MATLAB (hereafter refer to as Parameter Optimization and Simulation Toolkit (POST)) for the parameter optimization, simulation over the 25-year validation period and plotting results from six selected HRSs as examples (marked with asterisks in Table S1). The total size of the file is 105 KB.

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/JSTARS.2018.2790409

***Index Terms*—European Space Agency Climate Change Initiative (ESA CCI), flood warning, parameter regionalization, remote sensing, soil moisture active passive (SMAP), soil moisture (SM), ungauged basins.**

I. Introduction

FLOOD-RELATED disasters take a significant number of human lives globally, besides causing substantial economic damage amounting to hundreds of millions of dollars per event [1]. More than 150000 people have died from floods in the past two decades (1996–2015), with more than 80% of these deaths occurring in low- or middle-income countries [2]. Heavy casualties in such countries are largely due to underdeveloped flood-warning systems, which play a crucial role in providing sufficient time for evacuation [3]. Flood warnings are generally based on numerical simulations using well calibrated rainfall–runoff models, which require real time and past hydrological data over a watershed. Monitoring and acquisition of appropriate hydrological data are often insufficient in developing countries, which makes flood warning difficult. Therefore, to overcome such limitations, flood-warning methodologies for ungauged or poorly gauged basins need to be devised, such that they can be adopted requiring no or minimal on-site measurements of flows and other hydrological data.

As a viable alternative to ground-based information, satellite remote sensing can help detect and monitor spatiotemporal evolution of flooding events. For example, the development of the well-studied Pakistan flood in 2010 is appropriately captured by various satellite images (see Fig. 1). This event left roughly 20% of the total area of the country under water, brought severe economic damage, and led to the loss of more than 2000 lives and displacement of 20 million people [4].

Both antecedent soil moisture (SM) and rainfall are important factors that lead to the formation of direct runoff and trigger flooding [5]. Runoff flows to lower elevations driven by gravity, in directions that can be estimated from topographic information. All such information (SM, rainfall, and topography) are remotely available today at reasonable spatiotemporal resolutions. As presented in Fig. 1(a)–(d), we can observe the SM before the event (10 July), at the peak of flooding (28 July–4 August), and after the event (13 August) using satellite-derived information. As the event progressed,

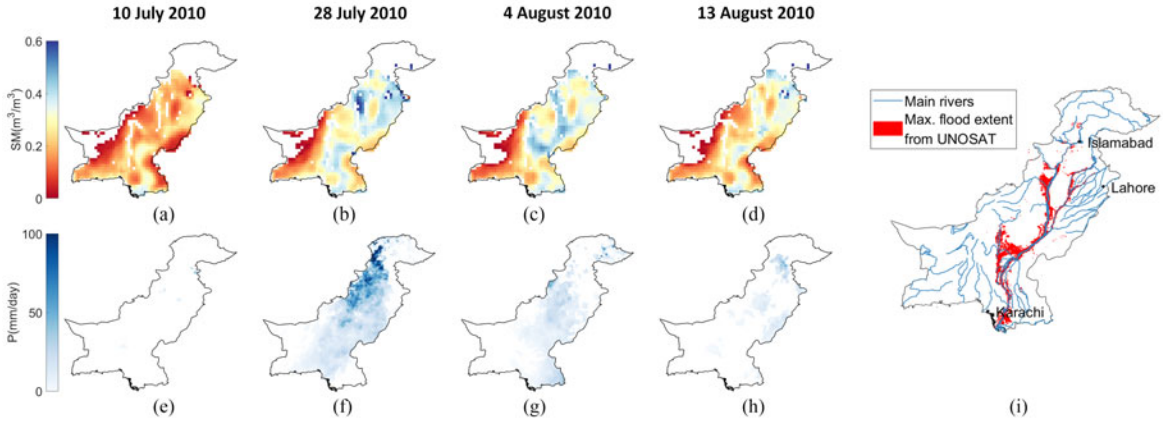


Fig. 1. Spatiotemporal distribution of (a)–(d) SM derived from the Advanced Microwave Scanning Radiometer Earth Observing System at 0.25°, (e)–(h) rainfall from the CHIRPS final product at 0.05°, and (i) maximum flood extent records from United Nations Institute for Training and Research's Operational Satellite Applications Programme (http://www.unitar.org/unosat/).

SM levels steadily increased, leading to more precipitation becoming runoff and ultimately producing extensive inundation [see Fig. 1(e)–(h) and (i)]. This mechanism is consistent with the considerable literature on this topic, which states that antecedent SM affects runoff in a significant way [6]–[8].

The case from the Pakistan flood suggests that using remotely sensed information on SM can provide evidence of flooding as well as the likelihood of flooding into the future. The use of remotely sensed data for hydrologic prediction over remote or ungauged regions has been investigated in many studies across the world. For example, satellite-derived SM data have been used to determine the antecedent watershed conditions in rainfall–runoff modeling applications [9]–[13]. SM data have been assimilated into hydrological models for correcting biases in precipitation, initial conditions, or both [14]–[16]. Lumped models have been widely used for such hydrologic modeling because of their simpler model structure. However, such lumped models are unable to account for the spatial heterogeneity of SM and rainfall, and hence perform best in watersheds where the spatial heterogeneity is negligible [5], [15], [17], [18]. While distributed models can be used to explicitly include the spatial variation of precipitation and SM, they have high computational costs and require large amounts of data and parameters to accurately represent the processes involved [19], [20]. To make the matter worse, such data are not fully available for sparsely gauged or ungauged watersheds where flood warnings are most needed.

A new approach is proposed in this study to address these limitations. A simple flood-warning method was developed that provides a qualitative assessment of near real-time flood risk across a watershed by combining readily available satellite-derived SM data with open-access rainfall, topographic and soil data at a coarse spatial resolution. The motivation here was to demonstrate whether effective warnings are possible using open-access information alone, the flood-warning model being calibrated and tested over multiple gauged basins with various sizes for a 35-year study period. Consequently, the parameters for the model were first calibrated by using the first ten-year data. Then, a possibility of parameter regionalization was investigated in terms of watershed characteristics to assess applicability over ungauged basins.

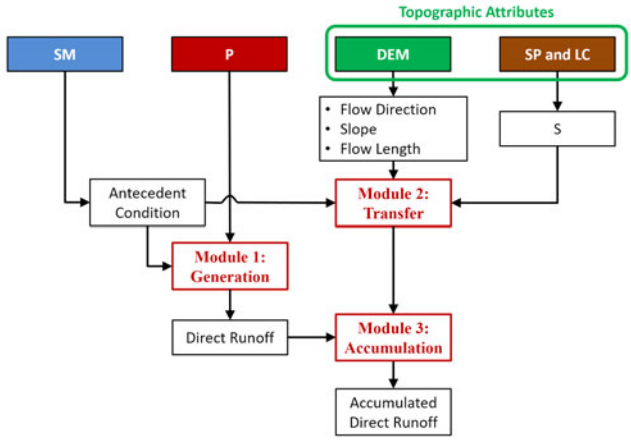


Fig. 2. Flow chart of the proposed method consisting of three modules: generation, transfer, and accumulation for combing effects from three data layers: SM, rainfall, and topographic attributes, where SM is the soil moisture, P is the rainfall, DEM is the digital elevation model, SP is the soil property, LC is the land cover, and *S* is the potential retention, respectively.

Details of the method are presented in Section II. The study area, data used in this study, parameter optimization, and model evaluation strategy are described in Section III, and then the results for parameter optimization and model evaluation are shown in Section IV. Sections V and VI discuss the results and suggest future research directions along with a summary of this study.

II. Method

It should be noted that the proposed method is designed for flood-warning applications instead of a quantitative estimation of streamflow. For this, the model considers where and how much rainfall occurs, and then calculates the transfer and accumulation of the rainfall excess using just SM, rainfall, and relevant topographic attributes. The proposed approach consists of three major modules for runoff generation, transfer, and accumulation (see Fig. 2).

The first module generates direct runoff using the Soil Moisture Accounting (SMA) based National Resources Conservation

Service Curve Number (NRCS-CN) method [21] from the rainfall data at each grid location with the antecedent conditions being determined by the remotely sensed SM data. Next, spatiotemporally varying isochrones are generated by the second module using a modified version of the NRCS lag time equation [22]. The third module accumulates flux to direct runoff on a daily basis based on the isochrones [23]–[26]. In this study, the accumulated direct runoff is regarded as a flood-warning indicator, which is based on the SM and rainfall during last few days. Details of each module are presented below. Note that the calculated direct runoff is expressed as a relative measure, here termed the proxy flow ($\tilde{Q}$) because only its variability and timing are being considered for assessing flood risk. The rationale for doing so is to utilize the established distributional similarity between observed and simulated values, without requiring a correspondence between the moments of the derived flows. Detailed information on a software toolkit for the three modules with example data is available in Supplementary materials.

A. Module 1: Generation

The original NRCS-CN is a rainfall–runoff model developed by the Soil Conservation Service (now NRCS) of the U.S. Department of Agriculture (USDA) [27]. In the model, the direct runoff (Q) is simply calculated using an explicit function of rainfall (P) and the potential maximum retention (S), which is transformed from the dimensionless CN. The simplicity of this approach has resulted in widespread applications since it was first developed in 1949. As it is an event-based rainfall–runoff modeling process, there have been difficulties in determining the antecedent SM condition of the watershed in relation to the maximum potential retention S [21]. In the original NRCS-CN method, the CN-derived S represents Antecedent Moisture Condition (AMC) II, and it varies from AMC I (drier) to AMC III (wetter). Although the five-day antecedent precipitation index (API_5) had been commonly used to determine the AMC level, the need for alternatives has been suggested due to poor results derived from API_5 [5], [15], [17], [28], [29].

As an alternative to the original method, Michel *et al.* [21] proposed a simple method incorporating SM into the NRCS-CN formulation, namely the SMA model. In this method, the initial SM condition is represented as an SM store level (V_0) within the total soil storage, which equals the sum of the maximum potential retention (S) and threshold for runoff generation (S_a). The amount of direct runoff (Q) generated depends on the relative differences between V_0, S_a, and P, which are classified into three typical cases [see Fig. 3(a)–(c)]. Case 1 [see Fig. 3(a)] represents the condition in which the sum of V_0 and P is lower than S_a, and then P is completely retained within the S_a column. In Case 2 [see Fig. 3(b)], where V_0 is lower than S_a but $P + V_0$ is greater than S_a, the initial amount of P is first captured in S_a and then the remaining part (i.e., $P + V_0 - S_\mathrm{a}$) contributes to direct runoff (Q). In Case 3 [see Fig. 3(c)] when V_0 is higher than S_a, Q depends on relative differences in fluxes of P and Q into and out of the total soil column (i.e., $S + S_\mathrm{a}$). Further details for these three cases can be found in Michel *et al.* [21].

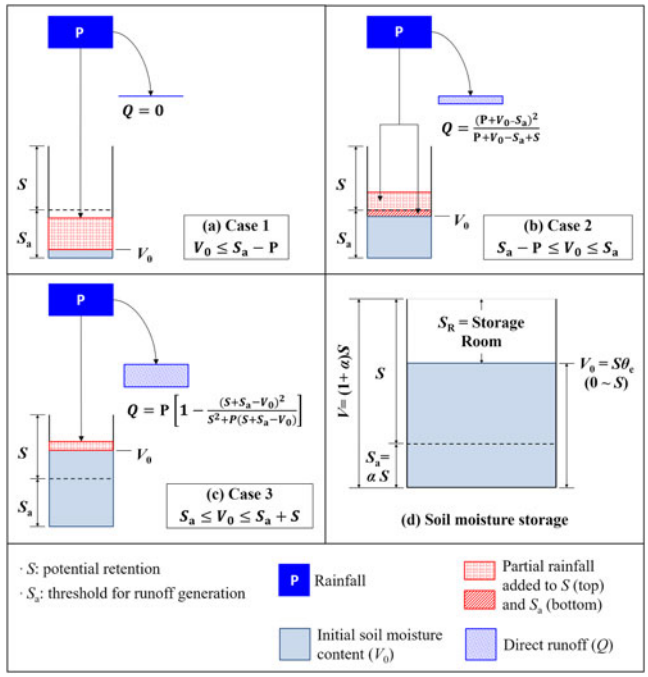


Fig. 3. Schematic diagrams of (a)–(c) Module 1: Generation, incorporating SMA procedure into NRCS-CN model for three typical cases (Michel *et al.*, 2005), and (d) SM storage used in this study.

The main issue for the SMA model is to estimate V_0, which was not explicitly described before. In this study, V_0 was directly assessed from S in proportion to the degree of saturation (θ_e) [5], [15], [17], [29] as

$$V_0 = \theta_\mathrm{e} \cdot S \tag{1}$$

where θ_e is a scaled value (0 to 1) of SM (θ) at each grid cell with respect to the maximum (θ_max) and minimum observed SM (θ_min) [17] over the study period. θ_max and θ_min were separately obtained from each SM product used in this study and then applied to each period, respectively.

In this study, S_a was regarded as dependent on S such as $S_\mathrm{a} = \alpha S$ [21], and the ratio (α) was optimized to improve the model calibration as presented in Fig. 3(d). Once V_0 and P are known at a grid cell, direct runoff at the grid cell is generated depending on the given condition of S_a.

B. Module 2: Transfer

The NRCS time lag method was originally developed to calculate the time of concentration by considering flow length, slope, and water retention by S over a watershed [22]. In this study, transfer of the generated direct runoff is based on grid to grid temporal lag (T_L), which represents grid to grid delay of the flood. Adopting the NRCS time lag method, we calculated T_L (hour) as

$$T_\mathrm{L} = \frac{l^{0.8} {S_\mathrm{R}}^{0.7}}{\beta Y^{0.5}} \tag{2}$$

where $l =$ flow length (km), $S_\mathrm{R} =$ soil storage room (mm), $Y =$ average watershed land slope (%), and $\beta =$ time coefficient.

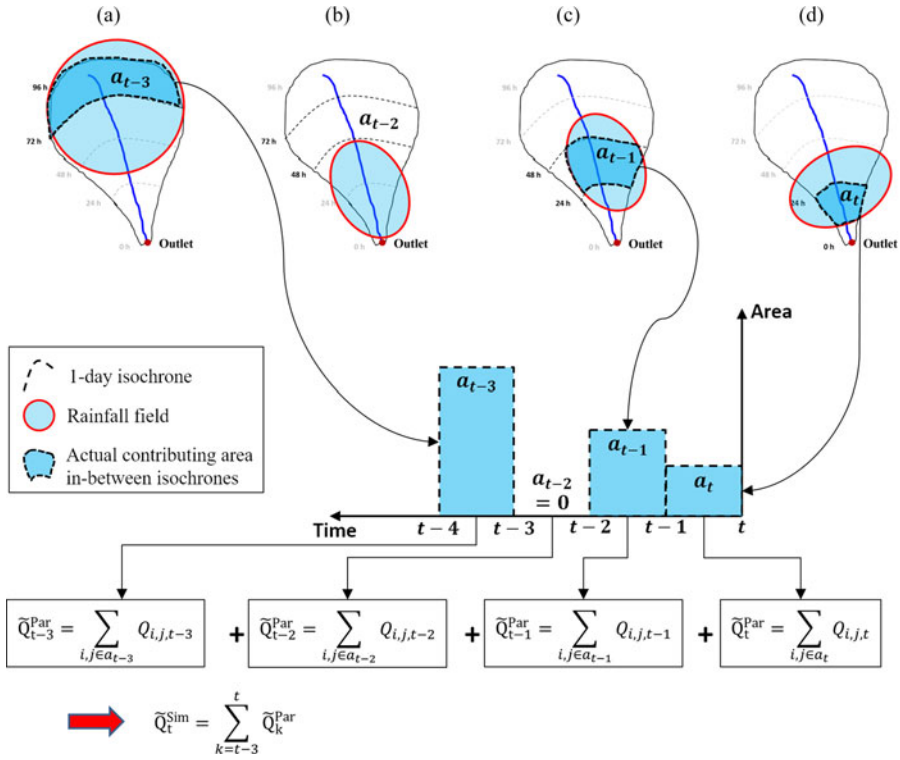


Fig. 4. Schematic diagram for spatiotemporally varying time-area calculation by Module 3 in case of $p = 4$. (a) $k = t - 3$. (b) $k = t - 2$. (c) $k = t - 1$. (d) $k = t$.

Replacing S in the original equation of the NRCS time lag method, the soil storage room (S_{R}), defined as $S_{\mathrm{R}} = S + S_{\mathrm{a}} - V_0$ in Fig. 3(d), was used to calculate T_{L} by the SMA procedure incorporated in Module 1. T_{L} varies in space and time because it is calculated using S_{R}, which varies in time due to the rainfall and in space due to the different soil properties at each location. Here, the time coefficient β is the second parameter to be optimized in this study.

C. Module 3: Accumulation

The direct runoff generated during the recent p days (including today) accumulates at a watershed outlet on a given day (t). This can be calculated as a routing procedure, determining the time and magnitude of the accumulated direct runoff. For demonstration purposes, an example with p equal to four days is shown in Fig. 4, with a detailed explanation for Module 3 following.

The storage room ($S_{\mathrm{R}}(i, j, k)$) and direct runoff ($Q(i, j, k)$) matrices for recent four days based on Module 1 are prepared in advance using rainfall and SM maps. The size of the matrices is $m \times n \times p$ of which spatial coordinates i are for rows (1 to m) and j for columns (1 to n), and k for time steps ($t - 3$ to t). One can generate isochrones with the resulting S_{R} matrix and results from the terrain analyses (i.e., flow direction, slope and flow length) obtained through Module 2. When $k = t - 3$, T_{L} is first calculated on a grid to grid basis by using temporally averaged $S_{\mathrm{R}}(:, :, t - 3 : t)$ because the direct runoff generated at $k = t - 3$ has been transferred to the outlet until $k = t$. Then, the total T_{L} from a grid cell to the outlet is calculated by accumulating each grid to grid T_{L} along the flow direction until the outlet is reached. Once calculations over all grid cells in the watershed are completed, isochrones are formed by which the watershed area is divided into sections with discrete 24-h units because the temporal resolution of rainfall and SM data is one day.

At $k = t - 3$ [see Fig. 4(a)], the section of interest is the area between 72 and 96 h, which represents the total travel time to the outlet of three and four days, respectively. Then, the area overlapping with the rainfall field is regarded as the actual contributing area (a_{t-3}) where direct runoff is generated and integrated to $\tilde{Q}_{t-3}^{\mathrm{Par}}$ partially contributing to the total flowrate at the outlet at $k = t$. Here, $\tilde{Q}_{t-3}^{\mathrm{Par}}$ is generally expressed as (3) and defined as the summation of $\tilde{Q}_{i,j,k}$ or the direct runoff at a grid cell (i, j) within the actual contributing area at time $k(a_k)$.

$$\tilde{Q}_k^{\mathrm{Par}} = a_k \sum_{i,j \in a_k} \tilde{Q}_{i,j,k}. \tag{3}$$

In the same manner, $\tilde{Q}_{t-2}^{\mathrm{Par}}$, $\tilde{Q}_{t-1}^{\mathrm{Par}}$, and $\tilde{Q}_t^{\mathrm{Par}}$ are sequentially calculated, and then the values are exponentially weighted and averaged to $\tilde{Q}_t^{\mathrm{Sim}}$ as in (4) and (5). Note that $\tilde{Q}_{t-2}^{\mathrm{Par}}$ is zero in Fig. 4 because a_{t-2} is zero even though it rains on the day. In this way, Module 3 is implemented for each day during the study period.

$$\tilde{Q}_t^{\mathrm{Sim}} = \frac{\sum_{k=t-p+1}^{t} \tilde{Q}_k^{\mathrm{Par}} w_k}{\sum_{k=t-p+1}^{t} w_k} \tag{4}$$

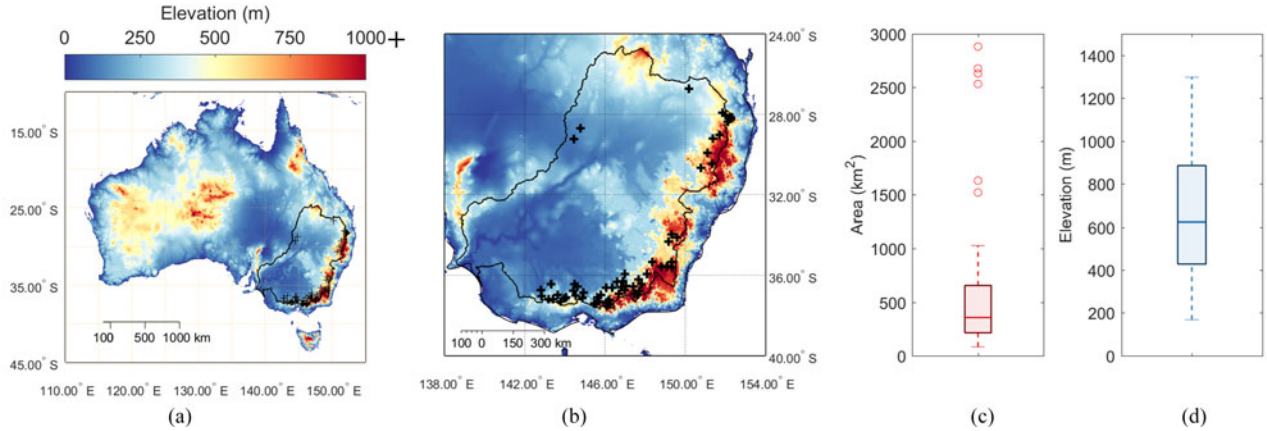


Fig. 5. Locations (+) at (a) continental scale, (b) regional scale, (c) watershed area, and (d) mean elevation distributions of 65 HRSs in MDB used for this study.

where

$$w_k = e^{-\frac{t-k}{p-1}}. \tag{5}$$

An important consideration in Module 3 is the correct length for p to calculate $\tilde{Q}_t^{\mathrm{Sim}}$, which is the third and final parameter to be optimized in this study. Note that the proposed model can then be applied at any point within the watershed using the three data layers of SM, rainfall, and topographic attributes (see Fig. 2)

III. Study Area and Data

A. Study Area

The Murray–Darling basin (MDB) was selected as the study area because it has almost complete streamflow data over multiple decades. Although such streamflow data are seldom available in more remote settings, this offers a proof-of-concept to demonstrate that the proposed method is applicable to ungauged basins elsewhere.

The MDB consists of three large rivers, the Darling river (2740 km), the Murray river (2520 km), and the Murrumbidgee river (1575 km), and it is a large semiarid basin covering 14% of Australia's area (1.06×10^6 km^2). The basin is characterized by high natural hydroclimatic variability, and is regarded as Australia's food bowl, supplying one third of the national food supply with agricultural area covering more than 80% of the basin [30].

The MDB has been highly regulated with structures such as dams, locks, and weirs for irrigation and flood control [31]. To deal with natural hydrological processes, the proposed flood-warning method was tested at multiple hydrologic reference stations (HRS) in the MDB that have minimal regulation. The HRS network includes 222 stations over Australia and can be used for identifying long-term trends in streamflow variability across all Australian hydroclimatic regions because they have been selected due to high hydrological data quality as well as minimal regulation and land use change (http://www.bom.gov.au/water/hrs/index.shtml) [32]. Although there are 75 HRSs within the MDB, only those stations whose temporal coverage is 50% at least during the ten-year calibration period were used. An additional constraint was that the watershed area is larger than the 25 km^2, which is the spatial resolution of the gridded input data. As a result, only 65 stations were selected, as shown in Fig. 5. Areas of the 65 HRSs range from 83 to 35239 km^2 and mean elevations from 168 to 1298 m (above sea level), as shown in Fig. 5(c) and (d), respectively. Note that the three stations with areas bigger than 3000 km^2 were not included in Fig. 5(c) for better visualization. Further details for the 65 HRSs are available in Table S1 of Supplementary materials.

B. Data and Processing

In this study, three data types have been used: first, SM to represent antecedent conditions preceding the rainfall event; second, rainfall; and, third, topographic attributes defining how and where the SM and rainfall interact. The required datasets covering 35-year study period from January 1, 1981 to December 31, 2015 are summarized in Table I. The first ten-year of data were used for model calibration and the remaining data of 25 years for model validation. In addition, an independent two-year period (April 1, 2015–March 31, 2017) using a recently available satellite SM product was also used for evaluating that the proposed method is applicable in near real time as described in the next section.

A spatial resolution of 0.05° (approximately 5 km) was used as the common spatial resolution in this study. All data were resampled to this resolution using bilinear interpolation unless otherwise mentioned. The streamflow data for the 65 HRS stations were obtained from the Australian Bureau of Meteorology (http://www.bom.gov.au/waterdata/).

1) Satellite SM Data: Currently, there is no satellite SM data from satellites that are operational in near real time and cover multiple decades required to validate the model performance. For this reason, the European Space Agency Climate Change Initiative (ESA CCI) SM was used in this study. The ESA CCI has released daily surface SM products at a spatial resolution of 0.25° using seven passive and three active microwave spaceborne instruments covering 37 years from November 1, 1978 to December 31, 2015 (http://www.esa-soilmoisture-cci.org/)

TABLE I
SUMMARY OF DATA USED IN THIS STUDY

Data		Source	Product	Temporal coverage	Resolution (Temp./Spa.)	Units
SM		ESA CCI	Active/passive combined SM	Nov 1979–Dec 2015	Daily/0.25°	m^3/m^3
		SMAP	Level 3 SM (SPL3SMP_E)	Apr 2015–Present	Daily/9 km	m^3/m^3
Rainfall		AWAP	Gridded daily rainfall	Jan 1900–Present	Daily/0.05°	mm
Topographic attributes	Soil Property	HWSD	Topsoil proportion		–/0.0083°	%
	DEM	SRTM	DEM V4	–	–/1 km	m (above sea level)
	LC	MODIS	MCD12C1	–	Yearly/0.05°	–

Temp. = temporal, Spa. = spatial.

[33], [34]. Three products are available in the CCI SM: active, passive, and active–passive combined data, which have been comprehensively validated with promising performances at the global scale [35], [36]. The active–passive combined data (version 03.2) were used to specify antecedent conditions during the multidecadal simulation in this study. To ensure high data quality, grid cells where topographic complexity (TC) or wetland fraction (WF) is greater than 10% were masked [35]. The TC and WF datasets are also available from the EAS CCI.

In addition to the CCI SM, this paper also considered a surface SM product from the soil moisture active passive (SMAP) mission [37] to evaluate whether the proposed method is applicable in near real time for an independent two-year period (April 1, 2015–March 31, 2017). Since April 2015, the SMAP has provided various SM products at spatial resolutions of 3, 9, and 36 km at both ascending (6:00 P.M.) and descending (6:00 A.M.) overpasses using an *L*-band radiometer (passive, 1.41 GHz) and a rotating reflector radar (active, 1.26 GHz). However, because the radar stopped transmitting in July 2015 due to a component failure, only radiometer-derived products are currently available. The SMAP Level 39-km Equal-Area Scalable Earth Grid SM, Version 1 (SPL3SMP_E) at the descending overpass was used, which has a spatial resolution of 9 km $\times$ 9 km [38]. The descending SM product are generally considered to be of higher quality because the equilibrium thermal conditions of vegetation, and near surface soil in the early morning is more favorable for the SM retrieval using passive microwave observations [39]. Only quality assured SMAP SM data were used, for which vegetation water content is lower than $5\,\mathrm{kg/m^2}$, and freeze–thaw fraction and static water body fraction are lower than 10% [38].

Although gaps in the satellite SM products are inevitable due to satellite swath patterns, vegetation, and frozen conditions [40], the proposed model requires fully filled SM data in space and time. The gaps in time and space over the study period were filled by a three-dimensional (3-D) gap filling method developed by Garcia [41]. This is a penalized least square regression based on 3-D discrete cosine transform (DCT-PLS) that solely relies on a smoothing parameter *s* (higher *s*, smoother result). In this study, the SM data (i.e., CCI and SMAP) over Australia were filled by the DCT-PLS method using *s* of 10^{-6} [40].

2) Rainfall Data: The 0.05° gridded rainfall data product from the Australian water availability project (AWAP) [42] was used as the rainfall forcing data. The method for retrieving the gridded AWAP precipitation data from the ground stations over Australia was developed by the Bureau of Meteorology National Climate Centre for AWAP (http://www.csiro.au/awap/). The method first uses a thin-plate smoothing spline to interpolate the monthly rainfall climatology and implements a successive correction method to interpolate anomalies of daily rainfall [43]. The AWAP rainfall data is available from 1900 onward as operational daily maps over Australia with a day of latency and has been successfully applied in various long-term hydroclimatological studies [44]–[46].

3) Topographic Attributes: The 1-km digital elevation model (DEM) product derived from the 90-m shuttle radar topographic mission (SRTM, http://www.cgiar-csi.org/) [47] was used for defining the topographic attributes after resampling to 0.05°. The method proposed by Wang and Liu [48] was first applied to the resampled DEM to fill surface depressions, which can lead to disconnected stream-flow patterns. The MATLAB-based TopoToolbox (https://topotoolbox.wordpress.com/) [49] was used to delineate watersheds through terrain analyses including calculation of gradient, flow direction detection by the steepest descent method [50], and calculation of flow accumulation numbers at each grid.

As presented in Section II, direct runoff was estimated using an NRCS-CN method. This study used the CN-derived potential retention (*S*) that requires land cover (LC) and soil type data [51]. The most recent MODIS LC product at 0.05° spatial resolution (MCD12C1) (data from 2012) [52] was used. Soil attribute maps of silt, clay, and sand at 90-m spatial resolution were obtained from the Harmonized World Soil Database (HWSD) data (v1.2) [53], [54]. A CN map was prepared by classifying each 0.05° grid cell into 12 texture classes of the USDA soil classification system [55] by using the soil attributes maps. Then, the USDA 12 texture classes were simply categorized into four hydrologic soil groups (HSG) A, B, C, and D, and the CN of each grid cell was selected from the standard lookup tables [22] according to the HSG classification and MODIS LC data. The estimated CN map was assumed to represent the maximum potential retention (*S*), as shown in Fig. 3(d).

C. Optimization of Model Parameters

There are three parameters to be optimized.

1) α that determines the threshold for runoff generation ($S_\mathrm{a} = \alpha S$).

2) β in (2) calculating T_L.

3) p in (4) calculating the accumulated direct runoff to the outlet.

As previously mentioned, this study aims to provide a qualitative flood warning based on the calculated direct runoff, rather than quantitatively predicting flowrate at a location. For this reason, its variability and timing are important considerations, and thus the three parameters were optimized to maximize the Pearson correlation coefficient (R) between two time series of simulation ($\tilde{Q}_t^{\mathrm{Sim}}$) and observation ($Q_t^{\mathrm{Obs}}$) at each station [56] as

$$\text{Maximize } R = \mathrm{f}(\alpha, \beta, p) = \frac{E\left[\left(\tilde{Q}_t^{\mathrm{Sim}} - \mu^{\mathrm{Sim}}\right)\left(Q_t^{\mathrm{Obs}} - \mu^{\mathrm{Obs}}\right)\right]}{\sigma^{\mathrm{Sim}}\sigma^{\mathrm{Obs}}}$$
$$\text{Subject to } 0 \le \alpha \le 2,\ 1 \le \beta \le 30,$$
$$1 \le p \le 2p^0 \ \ (\text{integer}) \tag{6}$$

where μ^{Sim} and μ^{Obs} are the mean values, and σ^{Sim} and σ^{Obs} are the standard deviations of $\tilde{Q}_t^{\mathrm{Sim}}$ and Q_t^{Obs}, respectively. The upper bounds of the constraints were set as double the published recommended values [5], [21], [22]. For example, the maximum S_a is twice S, β is limited to 30, which is nearly twice the value of 16.88 by only the unit conversion of the original NRCS equation for calculating the time of concentration, and finally p^0 is the maximum number of days calculated by the original NRCS equation using the CN-derived S. Note that the first ten-year CCI SM and corresponding rainfall data (January 1, 1981–December 31, 1990) were used for parameter optimization, and the estimated parameters of the 65 HRSs were used for validation over the remaining 25 years as described in the next section.

D. Model Evaluation Strategy

The proposed model was designed for providing the flood warning by correctly allocating each simulated value to a probability of occurrence in its cumulative density function (CDF), and then identifying extreme values possibly causing flooding conditions. In case the simulation is perfect in this point of view, both percentiles of simulated and observed values completely match. Regarding this, the model performance was tested through CDF-based threshold exceedance analyses by only using discriminating events above a fixed threshold [57].

The method provides a relative flood warning by considering the exceedance probability of a flowrate over time at that location. For example, as presented in Fig. 6, the historical floods (x, Floods #1–3) in the watershed are first mapped onto the empirical CDF ($F(x)$) of observed flowrate (red line). The warning threshold corresponding to Flood #1 is the exceedance probability $(1 - F(x_1))$ for this event. This exceedance probability is then transferred to the CDF of simulated flowrate (blue line) and the equivalent simulation flow accumulation threshold x_1 is ascertained.

The strategy shown in Fig. 7(a) and (b) was used to determine a suitable flood occurrence threshold and evaluate the model simulations. A high exceedance probability of 1% of observed daily direct runoff ($Q_{1\%}^{\mathrm{Obs}}$) was arbitrarily chosen and used as the flood threshold and the value of the simulation with the same exceedance probability ($\tilde{Q}_{1\%}^{\mathrm{Sim}}$) was found for each watershed.

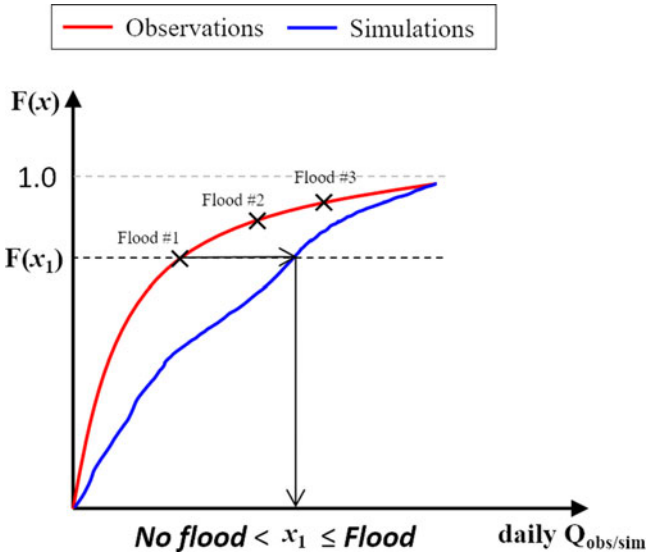


Fig. 6. Schematic for flood warning in a watershed based on matching the observed probability of a flood event with the simulated flow CDF.

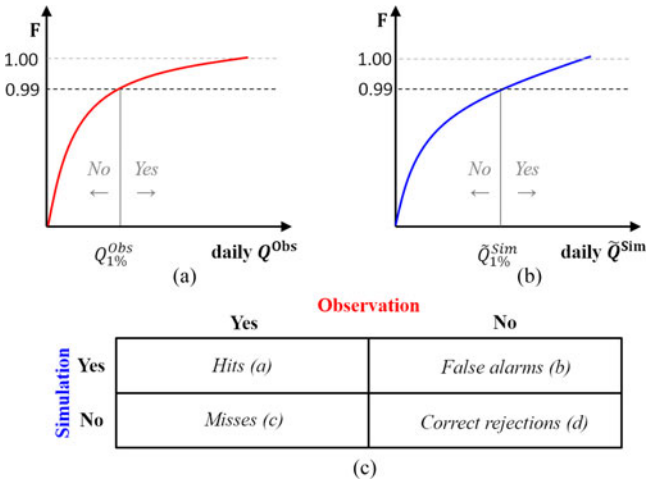


Fig. 7. Cumulative density function (CDFs) of daily (a) observations and (b) simulations, showing (c) 2 × 2 contingency table based forecast evaluations in terms of hits (a), false alarms (b), misses (c) and correct rejection (d), respectively.

This threshold was used to forecast whether a flood occurs or not. A 2 × 2 contingency table was established at each station for the model evaluation, as shown in Fig. 7(c), consisting of hits (a), false alarms (b), misses (c), and correct rejections (d). If the simulation is in perfect linear correlation with the observation, both b and c are zeros in the table.

The contingency tables were used to assess the model performance using a number of performance attributes [58]. The proportion correct (PC) represents the model accuracy and is calculated as

$$\mathrm{PC} = \frac{a+d}{a+b+c+d} \ \text{range}\ [0,1]\,, \text{best} : 1. \tag{7}$$

In cases where the number of events forecast is substantially less than the nonoccurrence flood events, the threat score (TS)

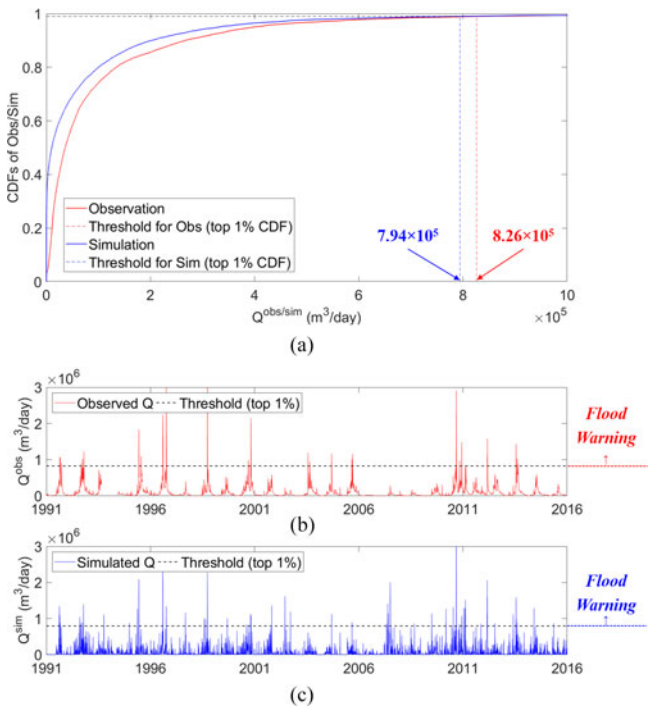


Fig. 8. Example of flood warning during the 25-year validation period at a HRS (403232: Morses Creek at Wandiligong). (a) CDFs of observation and simulation, and time series of (b) observation and (c) simulation.

is particularly useful. This is computed as

$$\mathrm{TS} = \frac{a}{a+b+c} \text{ range } [0,1], \text{best} : 1. \tag{8}$$

Together with PC and TS, other metrics considered include the frequency of hits (FOH), probability of detection (POD), and Peirce's skill score (PSS) as shown as (9), (10), and (11), respectively, for evaluating the proposed flood-warning method

$$\mathrm{FOH} = \frac{a}{a+b} \text{ range } [0,1], \text{best} : 1, \tag{9}$$

$$\mathrm{POD} = \frac{a}{a+c} \text{ range } [0,1], \text{best} : 1, \tag{10}$$

$$\mathrm{PSS} = \mathrm{POD} - \frac{b}{b+d} \text{ range } [-1,1], \text{best} : 1. \tag{11}$$

Two validation cases were used. The first (named V1) evaluates the floods predicted in the 25-year period using the CCI SM. The second validation case (V2) considers floods in the most two-year period using the near real-time SMAP SM. In both the cases, the top 1% threshold value (i.e., $F = 0.99$ in Fig. 7) and three optimized parameters from the ten-year CCI SM period were used to predict floods in the two validation cases.

An example of how the flood warning works is presented in Fig. 8. The selected station is one of 65 HRSs (403232: Morses Creek at Wandiligong) and the evaluation covers the 25-year validation period. First, both observation and simulation thresholds are determined by using the 1% exceedance probability based on the two CDFs, as shown in Fig. 8(a), and the two thresholds are used to define whether flood warning are issued at the location as presented in Fig. 8(b) and (c). Hits, false alarms, misses, and correct rejections are calculated based on the 2×2 contingency table and the five skill scores (i.e., PC, TS, FOH, POD, and PSS) can be estimated.

IV. Results

A. Parameter Optimization Results

During the ten-year calibration period using the CCI SM, the parameters were optimized for each of the 65 HRS stations. Fig. 9 presents the results of the correlation (R) between $\tilde{Q}_t^{\mathrm{Sim}}$ and the observed streamflow for the calibration period, and the calibrated values for each of three parameters (α, β, and p).

The focus here is on the spatial and statistical distributions of the parameters. The correlations of the parameters with three watershed characteristics: area, mean slope, and mean elevation, are also considered to ascertain if regionalization of the parameters could allow the method to be used in ungauged basins. Here, the mean slope is defined as the ratio of the elevation difference between the farthest point from the outlet and the outlet to its flow-path length, and the mean elevation was calculated by arithmetically averaging DEM values within each watershed boundary.

As presented in Fig. 9(a), the correlations between the model simulations and observed streamflow are generally moderate to strong, with mean R of 0.55 ± 0.09. Poorly performing locations tended to be watersheds where the mean elevation is high and the mean slope is small, which are illustrated as yellowish marks in Fig. 9(a).

To consider the potential for regionalization, the correlation of the optimized parameter with the four watershed characteristics were calculated. For α that determines the runoff threshold in the SMA based NRCS-CN model, there were no significant relationships with any of the watershed characteristics. However, the values for α were generally around 0.2–0.3 with only a few outliers more than 0.7 [see Fig. 9(b)], suggesting that it may be possible to use a single representative value for α as proposed by Michel *et al.* [21]. This was tested by using the median of all 65 α values ($\alpha_{\mathrm{median}} = 0.26$) and the overall model performance was only slightly degraded with mean R of 0.51 ± 0.09.

The β values were found to be significantly correlated with mean slope (Pearson correlation = –0.53), which reflects the influence of β on calculating the correct value for T_{L}. Flatter slopes lead to longer T_{L} and result in larger β values. Finally, the number of days over which rainfall should be accumulated (p) ranges from 2 to 8 except for the two large watersheds in the western MDB [see Fig. 9(d)]. The number of days of accumulation is closely related to watershed area and mean slope, with Pearson correlation coefficients of 0.94 and –0.48, respectively.

B. Model Evaluation Results

Evaluation results using the top 1% thresholds for flood warnings are presented as box plots in Fig. 10 in terms of the 2×2 contingency table-based skill scores (i.e., PC, TS, FOH, POD, and PSS). The box plots show the R between observations and

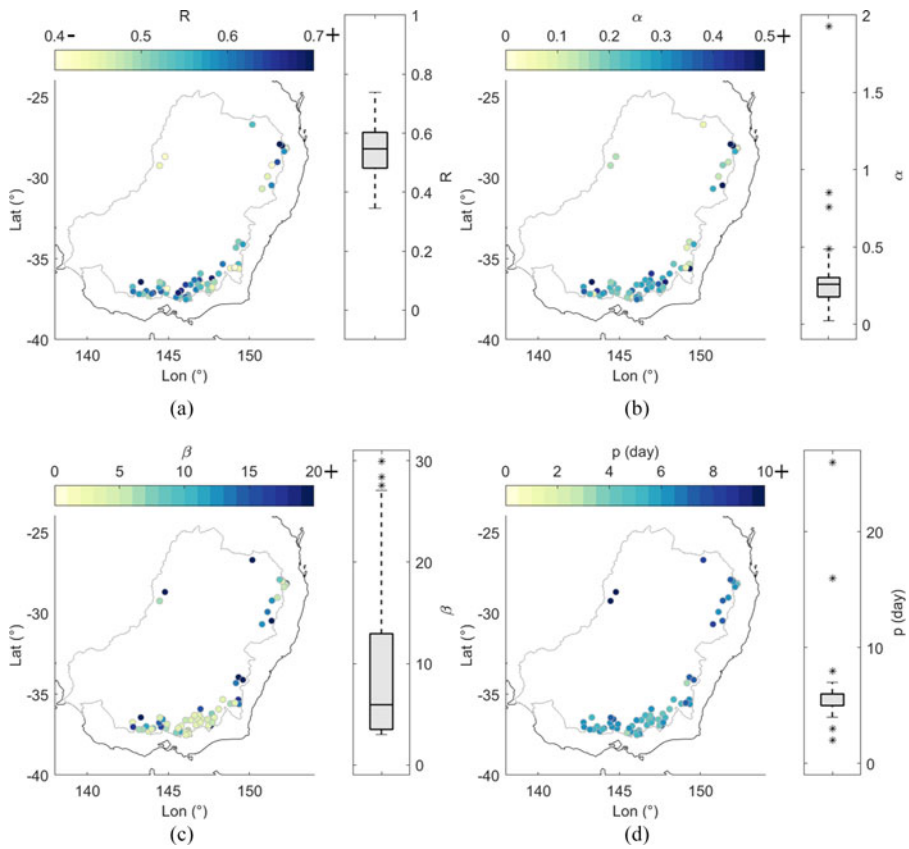


Fig. 9. Parameter optimization results over 65 HRSs during ten-year calibration period, presented as maps and boxplots. (a) Maximized temporal correlation (R) between simulations and observations, (b) α for determining the threshold for runoff generation ($S_{\mathrm{a}} = \alpha S$), (c) β for calculating the grid to grid temporal lag (T_{L}), and (d) necessary last p days for calculating $\tilde{Q}_t^{\mathrm{Sim}}$ to the outlet.

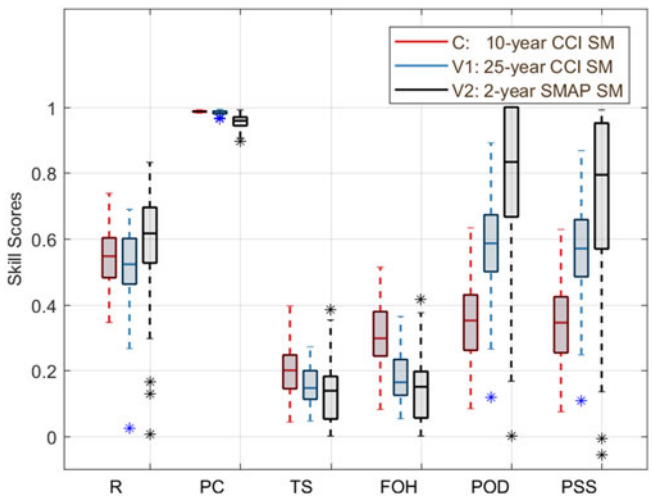


Fig. 10. Box plot presenting R between observations and simulations, and estimated 2×2 contingency table based five skill scores for C (red), V1 (blue) and V2 (black). The red circle for R marks largely degraded stations for V1 and V2.

simulations, and the five skill scores over the 65 HRSs for the calibration (C) and two validation periods (V1 and V2). Note that two stations were excluded for V2 because their daily flowrate observations do not cover the two-year validation period (April 1, 2015–March 31, 2017), and the skill scores from the SMAP SM are not necessarily representative due to small numbers of cases, especially for hits, false alarms, and misses, resulting from the short two-year period.

Promising results were obtained in validation (both V1 and V2), as shown in the first column in Fig. 10. This suggests that the proposed simple model is reasonably robust compared to the R values of the calibration period. Moderate correlations were achieved for the V1 validation period showing that the model provides a good measure of the temporal variability of flows. Moreover, the fact that there were moderate to strong R values in V2 suggests that the SMAP SM can be effectively used for the flood warning in near real time. However, there were degraded R values in both V1 and V2, and it was found that large degradations in R values are closely related with α values. This is because α is dependent on SM data used for the model. When the SM store level (V_0) is estimated by the SM data, α adjusts the runoff threshold in the SMA model. By doing so, a higher α induces lower direct runoff, but the lower α works to increase water flux at an outlet. In this regard, a large degradation in R can occur in the case of an α value overfitted to specific SM data. For example, the largest degradations in both V1 and V2 arose from at the same HRS (422319B: Dalrymple Creek at Allora) for which α was 1.93, optimized using the ten-year calibration period. The maximized R of the HRS in the calibration period was 0.70 but was significantly degraded to 0.02 for V1 and 0.01 for V2, respectively, meaning the α overfitted to the calibration

TABLE II
OVERALL PERCENTAGES OF FOUR COMPONENTS OF 2 × 2 CONTINGENCY TABLE FOR THE CALIBRATION (C) AND BOTH VALIDATION PERIODS (V1 AND V2)

Period	2 × 2 contingency table based components (%)					#Obs
	H	FA	M	CR	Total	
C	0.31	0.70	0.60	98.39	100	237380
V1	0.29	1.45	0.23	98.02	100	593515
V2	0.60	4.07	0.22	95.11	100	46053

C = Calibration, V1 = Validation1, V2 = Validation2, H = hits, FA = False alarms, M = Misses, CR = Correct rejections, #Obs = Total number of daily observations.

period and did not work properly for both validation periods and accordingly caused low skill scores.

The calculated skill scores provide guidance on different aspects of the model's performance. The PC values for C, and both V1 and V2 were close to one (the second column in Fig. 10). This is expected because the flood threshold is calculated with reference to all daily flows so that most days do not have flooding, and these are easy to predict. However, there are two notable differences between the calibration and validation periods, which are significant increases in false alarms and decreases in misses as presented in Table II. These features are shown in Fig. 8(b) and (c) in which high values of simulated direct runoff were more frequently noted than the observations, implying that higher α values are required for both V1 and V2 to reduce the false alarms. Even though the percentages of false alarms and misses are relatively small compared to the full data case, they sensitively work by excluding the correct rejections which make up an absolute majority when calculating the other skill scores as presented in (8)–(11).

For TS and FOH (the third and fourth columns in Fig. 10), the case of C generally shows better performances than V1 and V2. This is because more false alarms were made in V1 and V2 than C. For the same reason, this tendency was intensified for FOH, which does not consider misses in the calculation (9). On the other hand, the effect of increased false alarms and decreased misses in the validation periods positively affects POD and PSS, which were significantly improved as presented in the fifth and sixth columns in (10). This is because POD only takes the decreased misses into consideration against the number of hits, and PSS is almost same as POD due to the substantial portion of correct rejections (11).

Although it is not feasible for this study to quantitatively check whether the obtained scores are competitive, it seems that the *R* and PSS values are comparable with results from the hydrologic model evaluation results for Global Flood Awareness System (GloFAS) [57]. The hydrological modeling for GloFAS was evaluated by comparing simulated versus observed flowrate during a 21-year period for 620 stations over the world, and then the PSS calculated based on a threshold exceedance analysis using the 90th percentile as the threshold. Among these results, seven stations within the MDB showed mean *R* of 0.44 ± 0.08 and mean PSS of 0.38 ± 0.12 that are comparable with the results presented in the first and sixth columns in Fig. 10.

V. DISCUSSIONS

This proof-of-concept study focused on building a framework for flood warnings over ungauged basins based on the parameter optimization and model evaluation results over the 65 gauged HRSs for the multiple decades. We now discuss implications for regionalization of the three calibrated parameters, which will allow application to any ungauged watershed. Further observations on model performance in a regionalized setting are also provided.

A. Parameter Regionalization

The optimized β and p parameters showed significant correlations with the watershed area, mean slope, and mean elevation, and a median value of α provided a similar simulation, resulting in the use of the calibrated α values. All these findings suggest that the three parameters can be regionalized as a function of the watershed characteristics and/or mean values of parameters estimated from neighboring watersheds. This provides some confidence that the method can be used for flood warnings in ungauged basins. However, extended analyses over wider hydroclimatic coverages (i.e., continental or global scales) are required for building a more representative framework for the parameter regionalization, which will be a part of future works. In addition, there is another possibility for the parameter optimization without using the flowrate data. The results presented above rely on the availability of observed flow data for calibrating model parameters and assessing model accuracy. In an ideal situation, the proposed method should be applicable when no such data is available. In this study, the parameter regionalization was previously suggested for this, but another possible way is to use a near real-time satellite-derived signal product from the Global Flood Detection System (http://www.gdacs.org/flooddetection/) [59]. The product is for identifying major floods on daily basis and derived from satellite-based passive microwave observations of surface water extent and floodplains, which can be used for estimating parameters of hydrologic models. For example, Khan *et al*. [60] calibrated model parameters by directly comparing two daily exceedance frequencies of the M/C ratio datasets and flowrate simulated from a distributed hydrologic model, and showed promising results. This process can be directly applied for the proposed method in this study.

B. Model Performance

Although the maximized *R* following parameter optimization for both cases were generally moderate to strong, weak *R* (<0.3) were observed over a few stations, which resulted in relatively low skill scores. It was found that α is highly dependent on SM data used for the model and the large degradation in *R* occurred when using α values overfitted to the SM data of the validation period. These results suggest the need for using consistent SM data in the calibration and modeling. Besides, the parameter optimization using the two-year SMAP SM showed higher *R* values than those of the ten-year CCI SM. A mean *R* of 0.79 ± 0.15 was obtained when the SMAP SM was independently used for the parameter optimization during the two-year period (i.e.,

April 1, 2015–March 31, 2017). This result supports that improvements in SM data can result in an improved flood-warning system, with an improvement if only one of the two are changed to a better product. For example, the daily SMAP SM product is the near real-time data used for the validation in this study after simply filling gaps [41] and then resampling to 0.05° with bilinear interpolation, which offers large possibilities for improvement in terms of quality [61]–[63] and spatial resolution [64]–[67].

In addition, it should be also noted that, even though the AWAP rainfall data used here have a high spatial resolution of 0.05° and is available in near real time, it only covers the Australia domain. For global applications of the proposed flood-warning method, the availability of rainfall data is essential and a number of possibilities can be considered. For example, the 0.05° gridded daily precipitation data from the Climate Hazards Group Infrared Precipitation with Stations (CHIRPS) [68] could be used for applications over other regions. The CHIRPS precipitation data are available from 1981 onward with the quasi-global coverage of 50°N–50°S, and the latest version of data, Version 2.0, was released in February 2015 (http://chg.geog.ucsb.edu/data/chirps/). However, the preliminary product using sparse gauge data generally requires a two-day latency, which needs to be overcome for forecasting applications. Some rainfall products from satellite missions (e.g., Integrated Multi-satellitE Retrievals for Global Precipitation Mission) are available in near real time and provide potentials with betterments in the accuracy and spatial resolution [69], [70]. In addition, improvements in characterizing rainfall extremes using the satellite rainfall products [71] offer a hope that better rainfall products with low latency will be made available in the near future. In addition to the SM and rainfall data, other factors affecting the model performance are likely to be in estimating the potential retention and resampling the DEM described in Section III. In this study, the HWSD soil attribute maps and MODIS LC data in 2012 were consistently used for estimating the potential retention, but various sources of soil datasets and especially the temporal change of LC data should be considered for further improvements.

The proposed method, however, suffers from a number of limitations, which can be addressed through further development. First, further sensitivity analyses are necessary by applying various combinations of datasets (i.e., SM, rainfall, soil, and LC) over watersheds having different attributes to those in MDB and different calibration and validation periods. This will shed a greater light on the need for more sophisticated regionalization alternatives than the simplistic nearest watershed approach used here. Second, improved SM and rainfall data needs to be incorporated into the proposed method, as and when new products are released. Finally, model refinements are needed when extending to regions having markedly different flood-causing mechanisms (such as snowmelt for instance).

VI. Summary and Conclusion

In this study, a method was presented for flood warning in a watershed by only using satellite-derived SM and readily available rainfall products, open-access soil property, and topographic information at coarse spatial scales. The proposed method consists of three modules for generation, transfer, and accumulation of direct runoff, and uses an accumulated direct runoff for indicating flood risk at a location. This provides a flood warning based on the assessment of an exceedance probability, rather than predicting actual flood magnitudes that require greater model complexity and input accuracy.

For the model development, three parameters in the model were optimized by maximizing the temporal correlation between simulated values and observed flowrate from 65 HRS in the MDB over the ten-year calibration period. Then, the model performances were evaluated in terms of contingency table based skill scores over the 25-year validation. The outcomes for the parameter estimation and model performances showed promise, with possibilities of extension to completely ungauged locations where no past flow observations exist.

Acknowledgment

This study was initiated from the Synthesis Workshop on "*Dynamics of Structure and Functions of Complex Networks*" held at Korea University in 2015. The authors thank the teams from ESA CCI, NASA, Australian Bureau of Meteorology, Climate Hazards Group and Food and Agriculture Organization of the United Nations for making their data publicly available. Particularly, they would like to thank Dr. L. Alfieri for providing the evaluation results of the hydrological modeling for GloFAS over Australia.

References

[1] D. Guha-Sapir, P. Hoyois, P. Wallemacq, and R. Below, Annual Disaster Statistical Review: Numbers and Trends. Retrieved from Brussels, Belgium, 2016. [Online]. Available: http://emdat.be/sites/default/files/adsr_2016.pdf

[2] CRED/UNISDR, "Poverty and death: Disaster mortality 1996–2015," Centre for Research on the Epidemiology of Disasters, Brussels, Belgium 2016.

[3] K. Carsell, N. Pingel, and D. Ford, "Quantifying the benefit of a flood warning system," *Natural Hazards Rev.*, vol. 5, pp. 131–140, 2004.

[4] J. P. Syvitski and G. R. Brakenridge, "Causation and avoidance of catastrophic flooding along the Indus river, Pakistan," *GSA Today*, vol. 23, pp. 4–10, 2013.

[5] P. Durán-Barroso, J. González, and J. B. Valdés, "Improvement of the integration of soil moisture accounting into the NRCS-CN model," *J. Hydrol.*, vol. 542, pp. 809–819, Nov. 2016.

[6] V. Castillo, A. Gomez-Plaza, and M. Martınez-Mena, "The role of antecedent soil water content in the runoff response of semiarid catchments: A simulation approach," *J. Hydrol.*, vol. 284, pp. 114–130, 2003.

[7] S. Pathiraja, S. Westra, and A. Sharma, "Why continuous simulation? The role of antecedent moisture in design flood estimation," *Water Resources Res.*, vol. 48, 2012, Art. no. W06534.

[8] R. Ruggenthaler, F. Schöberl, G. Markart, K. Klebinder, A. Hammerle, and G. Leitinger, "Quantification of soil moisture effects on runoff formation at the hillslope scale," *J. Irrigation Drainage Eng.*, vol. 141, 2015, Art. no. 05015001.

[9] P. R. Houser, W. J. Shuttleworth, J. S. Famiglietti, H. V. Gupta, K. H. Syed, and D. C. Goodrich, "Integration of soil moisture remote sensing and hydrologic modeling using data assimilation," Water Resources Research, vol. 34, no. 12, pp. 3405–3420, 1998, doi:10.1029/1998WR900001.

[10] J. Komma, G. Blöschl, and C. Reszler, "Soil moisture updating by Ensemble Kalman Filtering in real-time flood forecasting," *J. Hydrol.*, vol. 357, pp. 228–242, Aug. 2008.

[11] L. Brocca *et al.*, "Improving runoff prediction through the assimilation of the ASCAT soil moisture product," *Hydrol. Earth Syst. Sci.*, vol. 14, pp. 1881–1893, 2010.

[12] F. Chen, W. T. Crow, P. J. Starks, and D. N. Moriasi, "Improving hydrologic predictions of a catchment model via assimilation of surface soil moisture," *Adv. Water Res.*, vol. 34, pp. 526–536, Apr. 2011.

[13] C. Alvarez-Garreton, D. Ryu, A. W. Western, W. T. Crow, and D. E. Robertson, "The impacts of assimilating satellite soil moisture into a rainfall-runoff model in a semi-arid catchment," *J. Hydrol.*, vol. 519, pp. 2763–2774, Nov. 2014.

[14] W. T. Crow and D. Ryu, "A new data assimilation approach for improving runoff prediction using remotely-sensed soil moisture retrievals," *Hydrol. Earth Syst. Sci.*, vol. 13, pp. 1–16, 2009.

[15] C. Massari, L. Brocca, T. Moramarco, Y. Tramblay, and J.-F. Didon Lescot, "Potential of soil moisture observations in flood modelling: Estimating initial conditions and correcting rainfall," *Adv. Water Res.*, vol. 74, pp. 44–53, Dec. 2014.

[16] N. Wanders, D. Karssenberg, A. de Roo, S. M. de Jong, and M. F. P. Bierkens, "The suitability of remotely sensed soil moisture for improving operational flood forecasting," *Hydrol. Earth Syst. Sci.*, vol. 18, pp. 2343–2357, 2014.

[17] L. Brocca, F. Melone, and T. Moramarco, "On the estimation of antecedent wetness conditions in rainfall–runoff modelling," *Hydrol. Process.*, vol. 22, pp. 629–642, 2008.

[18] W. Shi, M. Huang, K. Gongadze, and L. Wu, "A modified SCS-CN method incorporating storm duration and antecedent soil moisture estimation for runoff prediction," *Water Res. Manag.*, vol. 31, pp. 1713–1727, 2017.

[19] F. Francés, J. I. Vélez, and J. J. Vélez, "Split-parameter structure for the automatic calibration of distributed hydrological models," *J. Hydrol.*, vol. 332, pp. 226–240, Jan. 2007.

[20] H. Ajami, U. Khan, N. K. Tuteja, and A. Sharma, "Development of a computationally efficient semi-distributed hydrologic modeling application for soil moisture, lateral flow and runoff simulation," *Environ. Model. Softw.*, vol. 85, pp. 319–331, Nov. 2016.

[21] C. Michel, V. Andréassian, and C. Perrin, "Soil Conservation Service Curve Number method: How to mend a wrong soil moisture accounting procedure?" *Water Res. Res.*, vol. 41, 2005, Art. no. W02011.

[22] NRCS "Chapter 15 time of concentration," Part 630 Hydrology National Engineering Handbook, Natural Resources Conservation Serv., United States Dept. Agriculture, Washington, DC, USA, 2010.

[23] B. Saghafian, P. Y. Julien, and H. Rajaie, "Runoff hydrograph simulation based on time variable isochrone technique," *J. Hydrol.*, vol. 261, pp. 193–203, Apr. 2002.

[24] C. Clark, "Storage and the unit hydrograph," in *Proc. Amer. Soc. Civil Eng.*, 1945, pp. 1333–1360.

[25] M. L. Terstriep and J. B. Stall, "The Illinois urban drainage area simulator, ILLUDAS," Illinois State Water Survey, Champaign, IL, USA, Bull. 58, 1974.

[26] L. Watkins, "*Department of Scientific and Industrial Research Road Research Laboratory Technical Paper No.55—The Design of Urban Sewer Systems—Research Into the Relation Between Rate of Rainfall and the Rate of Flow in Sewers*. London, U.K.: Her Majesty's Stationery Office, 1962.

[27] NRCS "Chapter 10 estimation of direct runoff from storm rainfall," Part 630 Hydrology National Engineering Handbook, Natural Resources Conservation Service, United States Dept. Agriculture, Washington, DC, USA, 2004.

[28] S. K. Mishra, M. K. Jain, R. P. Pandey, and V. P. Singh, "Catchment area-based evaluation of the AMC-dependent SCS-CN-based rainfall–runoff models," *Hydrol. Process.*, vol. 19, pp. 2701–2718, 2005.

[29] H. E. Beck, R. A. M. de Jeu, J. Schellekens, A. I. J. M. van Dijk, and L. A. Bruijnzeel, "Improving curve number based storm runoff estimates using soil moisture proxies," *IEEE J. Select. Topics Appl. Earth Obs. Remote Sens.*, vol. 2, no. 4, pp. 250–259, Dec. 2009.

[30] M. Leblanc, S. Tweed, A. Van Dijk, and B. Timbal, "A review of historic and future hydrological changes in the Murray-Darling Basin," *Global Planet. Change*, vol. 80–81, pp. 226–246, Jan. 2012.

[31] R. T. Kingsford "Ecological impacts of dams, water diversions and river management on floodplain wetlands in Australia," *Austral Ecol.*, vol. 25, pp. 109–127, 2000.

[32] M. Turner, M. Bari, G. Amirthanathan, and Z. Ahmad, "Australian network of hydrologic reference stations—Advances in design, development and implementation," in *Proc. Hydrol. Water Res. Symp.*, 2012, pp. 1555–1564.

[33] Y. Y. Liu *et al.*, "Developing an improved soil moisture dataset by blending passive and active microwave satellite-based retrievals," *Hydrol. Earth Syst. Sci.*, vol. 15, pp. 425–436, 2011.

[34] Y. Y. Liu *et al.*, "Trend-preserving blending of passive and active microwave soil moisture retrievals," *Remote Sens. Environ.*, vol. 123, pp. 280–297, Aug. 2012.

[35] W. A. Dorigo *et al.*, "Evaluation of the ESA CCI soil moisture product using ground-based observations," *Remote Sens. Environ.*, vol. 162, pp. 380–395, 2014.

[36] W. Dorigo *et al.*, "ESA CCI Soil Moisture for improved Earth system understanding: State-of-the art and future directions," *Remote Sens. Environ.* , vol. 203, pp. 185–215, 2017.

[37] D. Entekhabi *et al.*, "The soil moisture active passive (SMAP) mission," *Proc. IEEE*, vol. 98, no. 5, pp. 704–716, May 2010.

[38] P. E. O'Neill, S. Chan, E. G. Njoku, T. Jackson, and R. Bindlish, "SMAP enhanced L3 radiometer global daily 9 km EASE-Grid soil moisture, version 1 (SPL3SMP_E)," NASA National Snow and Ice Data Center Distributed Active Archive Center, Boulder, CO, USA, 2016.

[39] R. A. M. De Jeu, W. Wagner, T. R. H. Holmes, A. J. Dolman, N. C. Giesen, and J. Friesen, "Global soil moisture patterns observed by space borne microwave radiometers and scatterometers," *Surv. Geophy.*, vol. 29, pp. 399–420, Oct. 2008.

[40] G. Wang, D. Garcia, Y. Liu, R. De Jeu, and A. J. Dolman, "A three-dimensional gap filling method for large geophysical datasets: Application to global satellite soil moisture observations," *Environ. Model. Softw.*, vol. 30, pp. 139–142, 2012.

[41] D. Garcia, "Robust smoothing of gridded data in one and higher dimensions with missing values," *Comput. Statist. Data Anal.*, vol. 54, pp. 1167–1178, Apr. 2010.

[42] M. Raupach, P. Briggs, V. Haverd, E. King, M. Paget, and C. Trudinger, "Australian water availability project (AWAP): CSIRO marine and atmospheric research component: Final report for phase 3," Centre for Australian Weather and Climate Research, Bureau of meteorology and CSIRO, Melbourne, VIC, Australia, 2009, vol. 67.

[43] C. Beesley, A. Frost, and J. Zajaczkowski, "A comparison of the BAWAP and SILO spatially interpolated daily rainfall datasets," in *Proc. 18th World IMACS/MODSIM Congr.,* Cairns, QLD, Australia, 2009, pp. 3886–3892.

[44] J. Gergis *et al.*, "On the long-term context of the 1997–2009 'Big Dry' in South-Eastern Australia: Insights from a 206-year multi-proxy rainfall reconstruction," *Clim. Change*, vol. 111, pp. 923–944, 2012.

[45] A. D. King, L. V. Alexander, and M. G. Donat, "The efficacy of using gridded data to examine extreme rainfall characteristics: a case study for Australia," *Int. J. Climatol.*, vol. 33, pp. 2376–2387, 2013.

[46] T. L. Delworth and F. Zeng, "Regional rainfall decline in Australia attributed to anthropogenic greenhouse gases and ozone levels," *Nature Geosci.*, vol. 7, pp. 583–587, 2014.

[47] A. Jarvis, H. I. Reuter, A. Nelson, and E. Guevara, "Hole-filled seamless SRTM data V4," International Centre for Tropical Agriculture, Cali, Colombia, 2008.

[48] L. Wang and H. Liu, "An efficient method for identifying and filling surface depressions in digital elevation models for hydrologic analysis and modelling," *Int. J. Geograph. Inf. Sci.*, vol. 20, pp. 193–213, 2006.

[49] W. Schwanghart and D. Scherler, "Short communication: Topotoolbox 2-MATLAB-based software for topographic analysis and modeling in Earth surface sciences," *Earth Surf. Dyn.*, vol. 2, pp. 1–7, 2014.

[50] J. F. O'Callaghan and D. M. Mark, "The extraction of drainage networks from digital elevation data," *Comput. Vis., Graph., Image Process.*, vol. 28, pp. 323–344, 1984.

[51] Y. Hong and R. Adler, "Estimation of global SCS curve numbers using satellite remote sensing and geospatial data," *Int. J. Remote Sens.*, vol. 29, pp. 471–477, 2008.

[52] NASA-LP-DAAC, "Land Cover Type Yearly L3 Global 0.05Deg CMG," Version 051 ed: NASA LP DACC, USGS EROS Center, Sioux Falls, South Dakota, 2012. [Online]. Available: https://lpdaac.usgs.gov, 2012.

[53] F. Nachtergaele and N. Batjes, Harmonized world soil database," Food and Agriculture Organization, Quebec City, QC, Canada, 2012.

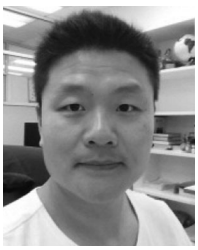
[54] W. R. Wieder, J. Boehnert, G. B. Bonan, and M. Langseth, "Regridded harmonized world soil database v1.2. data set," Oak Ridge National Laboratory Distributed Active Archive Center, Oak Ridge, TN, USA, 2014.

[55] R. Cronshey "Urban hydrology for small watersheds," Eng. Div., Soil Conservation Serv., U.S. Dept. Agriculture, Washington, DC, USA, 1986.

[56] D. R. Legates and G. J. McCabe, "Evaluating the use of "goodness-of-fit" Measures in hydrologic and hydroclimatic model validation," *Water Resources Res.*, vol. 35, pp. 233–241, 1999.

[57] L. Alfieri *et al.*, "GloFAS -global ensemble streamflow forecasting and flood early warning," *Hydrol. Earth Syst. Sci.*, vol. 17, pp. 1161–1175, 2013.

[58] D. S. Wilks, *Statistical Methods in the Atmospheric Sciences*, vol. 100, Orlando, FL, USA: Academic, 2011.
[59] T. De Groeve "Flood monitoring and mapping using passive microwave remote sensing in Namibia," *Geomatics, Natural Hazards Risk*, vol. 1, pp. 19–35, Mar. 2010.
[60] S. I. Khan *et al.*, "Microwave satellite data for hydrologic modeling in ungauged basins," *IEEE Geosci. Remote Sens. Lett.*, vol. 9, no. 4, pp. 663–667, Jul. 2012.
[61] S. Kim, R. M. Parinussa, Y. Y. Liu, F. M. Johnson, and A. Sharma, "Merging alternate remotely-sensed soil moisture retrievals using a non-static model combination approach," *Remote Sens.*, vol. 8, pp. 43–62, Jun. 2016.
[62] M. Enenkel *et al.*, "Combining satellite observations to develop a global soil moisture product for near-real-time applications," *Hydrol. Earth Syst. Sci.*, vol. 20, pp. 4191–4208, 2016.
[63] S. Kim, R. M. Parinussa, Y. Y. Liu, F. M. Johnson, and A. Sharma, "A framework for combining multiple soil moisture retrievals based on maximizing temporal correlation," *Geophys. Res. Lett.*, vol. 42, pp. 6662–6670, Aug . 2015.
[64] S. Kim, K. Balakrishnan, Y. Liu, F. Johnson, and A. Sharma, "Spatial disaggregation of coarse soil moisture data by using high-resolution remotely sensed vegetation products," *IEEE Geosci. Remote Sens. Lett.*, vol. 14, no. 9, pp. 1604–1608, Sep. 2017.
[65] K. C. Kornelsen "Downscaling satellite microwave observations to facilitate high resolution hydrological modelling," Geography Earth Sci., McMaster Univ., Hamilton, ON, Canada, 2015.
[66] B. Fang and V. Lakshmi, "Soil moisture at watershed scale: Remote sensing techniques," *J. Hydrol.*, vol. 516, pp. 258–272, Aug. 2014.
[67] N. N. Das *et al.*, "Soil moisture active passive (SMAP) project assessment report for the L2SMSP beta-release data products," Jet Propulsion Lab., Calif. Inst. Technol. Rep. JPL D-56549, Pasadena, CA, USA, Nov. 2017.
[68] C. Funk *et al.*, "The climate hazards infrared precipitation with stations—A new environmental record for monitoring extremes," *Sci. Data*, vol. 2, Dec. 2015, Art. no. 150066.
[69] Z. Duan, J. Liu, Y. Tuo, G. Chiogna, and M. Disse, "Evaluation of eight high spatial resolution gridded precipitation products in Adige Basin (Italy) at multiple temporal and spatial scales," *Sci. Total Environ.*, vol. 573, pp. 1536–1553, 2016.
[70] G. J. Huffman, D. T. Bolvin, and E. J. Nelkin, "Integrated Multi-satellitE Retrievals for GPM (IMERG) technical documentation," NASA Goddard Space Flight Center Code 612, Greenbelt, MD, USA, p. 47, 2015.
[71] A. Libertino, A. Sharma, V. Lakshmi, and P. Claps, "A global assessment of the timing of extreme rainfall from TRMM and GPM for improving hydrologic design," *Environ. Res. Lett.*, vol. 11, 2016, Art. no. 054003.



Seokhyeon Kim received the B.E. and M.E. degrees from Korea University, Seoul, South Korea, and the Ph.D. degree from the University of New South Wales (UNSW), Sydney, NSW, Australia, in 2001, 2008, and 2017, respectively, all in civil engineering.

He is currently a Research Associate with the School of Civil and Environmental Engineering, UNSW. His research interests include the satellite remote sensing based hydrologic applications, improvements of remote sensing data in terms of accuracy and spatial resolution, computational hydrologic modeling and optimization, and big data analyses.

Kyungrock Paik received the B.E. degree in civil engineering and M.E. degree in water resources engineering from Korea University, Seoul, South Korea, and the Ph.D. degree in civil engineering from the University of Illinois at Urbana–Champaign, Champaign, IL, USA, in 1995, 2001, and 2006, respectively.

From 2006 to 2009, he was a Faculty with the School of Environmental Systems Engineering, University of Western Australia, Perth, WA, Australia. In 2009, he moved back to South Korea and is currently a Professor with the School of Civil, Environmental, and Architectural Engineering, Korea University. His research interests include physical hydrology and fluvial geomorphology. In particular, he has investigated a theory to describe shaping of Earth's terrestrial landscapes, simulation methods for dynamic Earth surface environment, self-organized networks of natural or human-nature coupled systems, surface-atmosphere-vegetation feedback, and river dynamics in hydrogeomorphic perspectives.

Fiona M. Johnson received the B.E. and Ph.D. degrees in civil engineering from the University of New South Wales (UNSW), Sydney, NSW, Australia, in 2002 and 2010, respectively.

She is currently a Senior Lecturer with the School of Civil and Environmental Engineering, UNSW. Prior to joining UNSW, she was with the Bureau of Meteorology working on the revision of intensity-frequency-duration design rainfall data for Australia. Her research interests include climate change impacts on water resources systems, in particular developing methods to bias correct General Circulation Model simulations to improve their representation of statistics important for engineering design, the analysis of extreme rainfall events and understanding the types of future changes that may lead to changes in flood risk.

Ashish Sharma received the Bachelor's degree in civil engineering from the University of Roorkee, Roorkee, India, in 1989, the Master's degree in water resources engineering from the Indian Institute of Technology Delhi, India in 1991, and the Ph.D. degree in water resources engineering from Utah State University, Logan, UT, USA, in 1996.

He is currently a Professor with the School of Civil and Environmental Engineering, University of New South Wales, Sydney, NSW, Australia. His research interests include finding ways of meaningfully assessing impact of climate change on hydrology and water resources, assessing (and reducing) modeling uncertainty, estimating design floods in a more meaningful way, issuing seasonal forecasts for water resources management, developing better models that simulate both hydrology and ecology in an increasingly warming world, and a bunch of other things mostly aligned around his strengths in using statistical analysis and methods for a range of hydrologic problems.