Water Resources Management (2019) 33:3011–3024
https://doi.org/10.1007/s11269-019-02283-y

A Multi-Model Nonstationary Rainfall-Runoff Modeling Framework: Analysis and Toolbox

Mojtaba Sadegh[1,2] • Amir AghaKouchak[1] • Alejandro Flores[3] • Iman Mallakpour[1] • Mohammad Reza Nikoo[4]

Received: 17 September 2018 / Accepted: 12 May 2019 /
Published online: 1 June 2019

Abstract

We present a framework and toolbox for multi-model (one at a time) nonstationary modeling of rainfall-runoff (RR) transformation. The designed time-varying nature of the five available conceptual RR models in the toolbox allows for modeling processes that are nonstationary in essence. Nonstationary Rainfall-Runoff Toolbox (NRRT) delivers insights about underlying watershed processes through interactive tuning of model parameters to reflect temporal nonstationarities. The toolbox includes a number of performance metrics, along with visual graphics to evaluate the goodness-of-fit of the model simulations. Our analysis shows that the proposed time-varying RR modeling framework successfully captures the nonstationary behavior of the Wights catchment in Australia. A multi-model analysis of this catchment, that has endured deforestation, provides insights on the functionality of different conceptual modules of RR models, and their representation of the real-world.

Keywords Nonstationarity · Rainfall-runoff modeling · Time-varying models · NRRT toolbox

1 Introduction

Rainfall-Runoff (RR) transformation has been traditionally represented by physical, conceptual, and statistical models (AghaKouchak et al. 2013; Singh and Woolhiser 2002). RR models simplify nonlinear, and spatially/temporally varying processes at the watershed scale (Pathiraja

Electronic supplementary material The online version of this article (https://doi.org/10.1007/s11269-019-02283-y) contains supplementary material, which is available to authorized users.

✉ Amir AghaKouchak
amir.a@uci.edu

[1] Department of Civil and Environmental Engineering, University of California, Irvine, CA, USA

[2] Department of Civil Engineering, Boise State University, Boise, ID, USA

[3] Department of Geosciences, Boise State University, Boise, ID, USA

[4] Department of Civil and Environmental Engineering, Shiraz University, Shiraz, Iran

et al. 2016b). Conceptual lumped RR models, for example, employ an aggregate representation of complex and distributed vegetation, surface and subsurface processes through transfer functions that route water between several serial and parallel storage compartments (buckets) (Nash et al. 1960; Schaake et al. 1996; Singh and Woolhiser 2002).

The traditional assumption in most hydrologic modeling practices is the stationarity of underlying processes (Sadegh et al. 2018). This convenient assumption, however, has been challenged by several studies (Leclerc and Ouarda 2007; Milly et al. 2008; Salas and Obeysekera 2013). Adopting a stationary model structure and parameterization disregards the natural and anthropogenic changes to watershed conditions, and limits their temporal transferability (Pathiraja et al. 2016b). While adopting time-varying model structure might be rather challenging (see Marshall et al. 2006 for a "hierarchical mixtures of experts" framework for time varying model structure), especially for simulation of continuous hydrologic variables, it is feasible to introduce temporal transition functions for model parameters. Time varying model parameters serve as a surrogate for physical changes in the watershed.

In the absence of a clear physical correspondence between conceptual model parameters and measurable geophysical properties, parameters are commonly inferred from the observed system response. This renders the model parameters highly dependent on the calibration data (Sadegh et al. 2018), and consequently shadows the usefulness of derived optimal model realizations when watershed conditions change (Pathiraja et al. 2016b). Changes in the watershed properties might be due to a plethora of natural and anthropogenic causes, including but not limited to: climate change, deforestation, afforestation, vegetation and land use change, urbanization and wildfires (see Pathiraja et al. 2016a and references therein).

Time varying system analysis has been widely used in different fields of science (Mohammadpour and Scherer 2012; Niedzwiecki 2000; Richards 1983; Grenier 1983), and has attracted the attention of hydrologists in the recent decades. For example, Brown et al. (2006) used a flow duration curve model with parameters that were dependent on land cover to analyze the impacts of afforestation on streamflow magnitude and frequency. Ouarda and El-Adlouni (2011) and Cheng et al. (2014) used a GEV model with time varying parameters for nonstationary frequency analysis. Application of such an approach in watershed modeling, however, has only recently being adopted. Westra et al. (2014) allowed for temporal variation of the maximum capacity of the production store in the GR4J model, which significantly improved the streamflow prediction for the Scott Creek catchment in South Australia. Efstratiadis et al. (2015) employed separate hydrologic units with time varying surface areas, as well as a time varying lumped conceptual model to simulate the response of Ferson Creek basin that has been urbanizing over the past 30 years. Pathiraja et al. (2016a, b) documented that adopting time varying model parameters improve model predictions, specifically when extrapolating in time. They also scrutinized the potential of data assimilation techniques to derive the temporal patterns in hydrologic model parameters from nonstationary streamflow observations.

In this paper, we introduce the Nonstationary Rainfall-Runoff Toolbox (NRRT) as a general multi-model (one at a time) framework that permits time-varying realizations of hydrologic models to predict hydrologic response of nonstationary processes in watersheds. We argue that physical changes in a watershed transform the underlying processes, which are manifested in temporal variation of parameter values in a conceptual model. Deforestation, for example, decreases the aggregate evaporation of a watershed and increases the overall streamflow response. This can be translated into conceptual models by decreasing the upper storage compartment capacity, which in turn leads to greater direct flow response and lower evaporative capacity.

2 Hydrologic Models

We briefly describe the structure of the lumped conceptual models used in NRRT, and refer interested readers to the source papers for a detailed description. A short description of model equations is provided in Table S1, and feasible ranges of the model parameters are presented in Table 1.

2.1 GR4J, GR5J, and GR6J

GR4J is a four parameter parsimonious lumped conceptual RR model that translates precipitation (P) and potential evapotranspiration (PET) inputs to streamflow estimates at the

Table 1 Description of parameters of conceptual rainfall-runoff models

Model name	Parameter name	Description	Unit	Range	Reference
GR4J	$S1_{max}$	Maximum capacity of the production store	mm	[0, 1500]	Perrin et al. (2003)
	Exch	Groundwater exchange coefficient	mm	[−10, 10]	
	$S2_{max}$	One day ahead maximum capacity of the routing store	mm	[1, 500]	
	UHB	Time base of unit hydrograph UH1	days	[0.5, 8]	
GR5J	$Exch_T$	Groundwater exchange threshold (for long-term memory)	–	[−4, 4]	Le Moine (2008)
GR6J	$S3_{max}$	Maximum capacity of additional routing store (parallel to $S2_{max}$)	mm	[0.5, 20]	Pushpalatha et al. (2011)
HyMod	C_{max}	Maximum soil moisture reservoir	mm	[1, 500]	Boyle et al. (2000); Gharari et al. (2013)
	Bexp	Spatial variability of soil moisture storage	–	[0.1, 2]	
	Alpha	Distribution factor between fast and slow routing reservoirs	–	[0.1, 0.99]	
	R_s	Residence time of slow flow reservoir	days	[0.001, 0.1]	
	R_q	Residence time of quick flow reservoirs	days	[0.1, 0.99]	
HBV	PWP	Soil permanent wilting point	mm	[90, 180]	Aghakouchak and Habib (2010)
	FC	Field capacity (soil moisture)	mm	[100, 200]	
	Beta	Shape coefficient	–	[1, 7]	
	L	Threshold water level (surface flow)	mm	[2, 5]	
	K_0	Near surface flow storage coefficient	1/day	[0.05, 0.2]	
	K_1	Interflow storage coefficient	1/day	[0.01, 0.1]	
	K_2	Baseflow storage coefficient	1/day	[0.01, 0.05]	
	K_{perc}	Percolation storage coefficient	1/day	[0.01, 0.05]	
	C	Potential evapotranspiration adjustment coefficient	1/°C	[0.01, 0.07]	
Snow module	DD	Degree-day factor	mm/°C	[0.01, 7]	Aghakouchak and Habib (2010)

catchment outlet (Perrin 2000). GR4J consists of one production and one routing storage compartments, two unit hydrograph (UH) functions that address the time lag between precipitation and streamflow generation, and a transfer function between UH2 and the routing store. GR4J partitions the inputs into net precipitation and net evapotranspiration, by subtracting PET from P. If the PET demand is not met from precipitation, ET can extract some portion of the production store. Net precipitation, if exists, partially contributes to the production store as well as the routing UH functions. Output of the production store (called percolation) also contributes to the UH routing. Inputs to the UH functions are distributed between UH1 (90%) and UH2 (10%), which then either directly (UH2) or indirectly (UH1, through a groundwater exchange term "f") contribute to the routing store. Outlet of the routing store and the net aggregate flow from UH1 and groundwater exchange then constitute the streamflow at the catchment outlet.

In an effort to improve the low flow prediction of GR4J, Le Moine (2008) introduced an extra parameter $Exch_T$ to the groundwater exchange term, "f", to account for long-term memory of the catchment. This parameter could be viewed as a new storage compartment in the model for long-term memory of the system. Moreover, Pushpalatha et al. (2011) in an attempt to further improve the low flow prediction, added a parallel routing storage, $S3_{max}$, to that of the GR5J and constructed GR6J. Figure 1a presents a schematic overview of the GR6J model structure, which can be simplified to GR5J and GR4J with the removal of some elements.

2.2 HyMod

Boyle (2001) introduced HyMod as a five parameter parsimonious lumped conceptual RR model. HyMod warrants a near sufficient process representation by including slow and fast streamflow responses along with a nonlinear soil moisture storage compartment (Gharari et al. 2013), while maintaining a low number of parameters. This model consists of five storage compartments, including one soil moisture store, three serial fast flow reservoirs and one slow flow reservoir. Fast and slow flow reservoirs work in parallel, after being serially fed from the soil moisture storage compartment. Precipitation and evapotranspiration are the drivers of the soil moisture compartment, which is controlled by two parameters, namely storage capacity, Cmax, and spatial variability parameter of soil moisture, B_{exp}. Output of the soil moisture reservoir is partitioned into two segments feeding the three fast and one slow flow storage compartments. These fast and slow flow reservoirs act as routing functions, and their outlets will form the streamflow at the catchment outlet. A schematic presentation of the HyMod model structure is presented in Fig. 1d.

2.3 HBV

HBV is originally developed by the Swedish Meteorological and Hydrological Institute (AghaKouchak et al. 2013), and its first version was introduced at the Nordic Hydrological Conference in Sandefjord in 1972 (Bergström 1992). Since then, multiple versions of this model are developed with different levels of complexity and spatial configuration (lumped and distributed). We adopt the lumped version of Aghakouchak and Habib (2010) as a ten parameter conceptual RR model, a schematic illustration of which is available in Fig. 1b. We secluded the snow module from the HBV framework, which reduces its total number of model parameters to nine. The snow module is discussed in Section 2.4. HBV consists of four

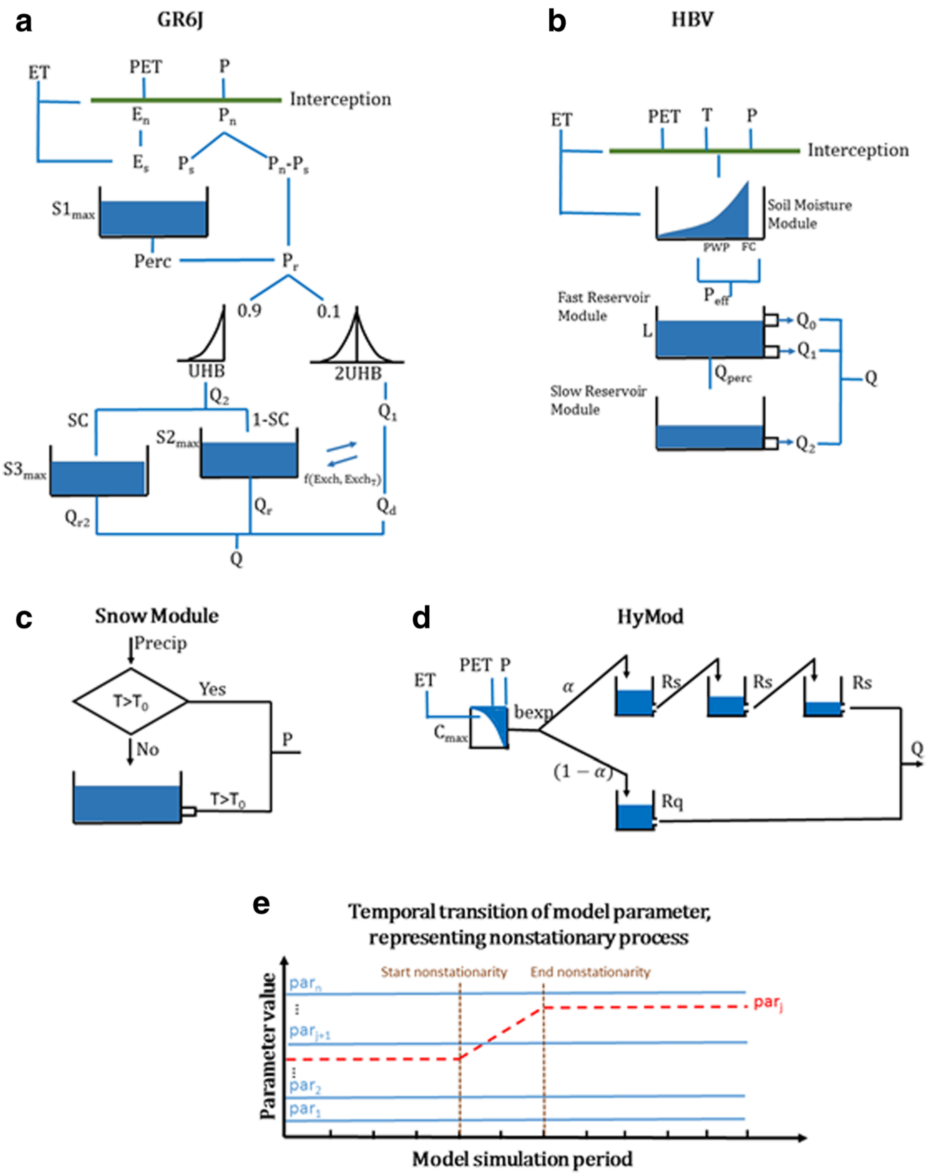

Fig. 1 Schematic structure of GR6J (**a**), HBV (**b**), snow module (**c**), and HyMod (**d**) models. GR6J simplifies to GR5J if the additional routing store ($S3_{max}$) is removed, and further simplifies to GR4J if parameter $Exch_T$ is omitted. Temporal transition of model parameter to represent process nonstationarity is shown in subfigure **c**. Observed precipitation is first filtered through the snow module to determine potential snow accumulation or melt. Freezing threshold is set at $T_0 = 0$ °C

serial storage compartments, namely soil moisture, surface runoff, interflow, and baseflow reservoirs. The surface runoff and interflow modules contribute to the near surface flow, and the last reservoir is the source of baseflow (AghaKouchak et al. 2013). Precipitation and potential evapotranspiration drive the soil moisture reservoir, which in turn feeds the surface runoff reservoir. A constant percolation rate connects the surface runoff, interflow and baseflow reservoirs. All these compartments contribute to streamflow generation. Aside from the structural difference of HBV from other models, it also differs in treating the PET forcing. While all models of this paper adopt a long term climatic average of PET as model input, HBV exclusively adjust PET based on divergence between the observed and long term monthly mean temperatures (AghaKouchak et al. 2013; Aghakouchak and Habib 2010).

2.4 Snow Module

This module, which originally was one routine of the HBV model of Aghakouchak and Habib (2010), determines snow accumulation and melt according to observed temperature. Snow accumulates, if temperature remains below freezing point, T_0, and melts if temperature exceeds T_0. Snowmelt is estimated based on

$$S_m = DD(T - T_0), \tag{1}$$

in which S_m is snow melt, DD signifies degree-day factor, and T is observed temperature. Degree-day factor governs the snowmelt volume triggered by temperatures remaining 1 °C above T_0 for one day. It is a common practice to consider a constant value between 0.7 and 9 mm/°C/day for DD (Aghakouchak and Habib 2010).

We separated this module from the HBV model of Aghakouchak and Habib (2010), and applied it to all models (including HBV). In other words, we first apply the snow module to precipitation and temperature observation and determine whether precipitation occurs in the form of rainfall or snow. If temperature is above freezing point, aggregate of snow melt and rainfall is fed to the subsequent modules of the RR models.

2.5 Temporally Transitioning Model Parameters

The RR models discussed in this section assume watershed processes are stationary, and hence a constant model parameter could be used to simulate the RR transformation. The fixed model parameters could be tuned through optimization against observed streamflow. However, watersheds may undergo nonstationarity (such as deforestation or urbanization), which renders the assumption of stationarity unrealistic. One can model such transformations in the watersheds through time-varying model parameters. In this study, we use a linear transition function for one model parameter at a time, which allows for nonstationary modeling of streamflow in a watershed that endured physical change. Figure 1e schematically presents the fixed (blue) and transitioning (red) model parameters in the simulation period. The transitioning parameter holds a fixed value before the nonstationarity starts, linearly shifts to a new value at the end of nonstationarity period, and remains fixed afterwards. The assumption of linear transition for the time-varying parameter is a pragmatic assumption adopted in this study, but in general any transition function could be used (such as exponential or polynomial of any order). Using a synthetic case study in the Supplementary Information (SI), we show that this assumption can sufficiently represent the changes in watershed processes for practical nonstationarity hypothesis testing purposes.

The underlying implicit assumption in this analysis is that watershed can be treated as stationary before any modification is applied. Watershed then undergoes physical changes in a period of time and reach a different stationary state after physical changes are exerted. Each state is assumed to be associated with one optimal parameter realization, which can sufficiently represent the watershed characteristics. However, there might be several sets of parameters that are equally good (equifinality) (Beven and Freer 2001). Moreover, there are other parameter sets that are acceptable, which can be used to determine the predictive uncertainty ranges of the model simulations (Sadegh et al. 2018). See SI for uncertainty analysis of model parameters.

3 Nonstationary Rainfall-Runoff Toolbox (NRRT)

Figure S1 (in SI) depicts the graphical user interface of NRRT, which composes of two steps: 1. calibration, and 2. analysis. The calibration step is used to tune model parameters to the under-study watershed. This step commences with the selection of model and provision of the inputs, including the period of calibration (start and end date), as well as watershed area and forcing/calibration data. Note that each model uses 1 year of spin up, and hence forcing should include at least 1 year of data before the calibration start date to allow model states to tune out. Forcing data (daily) can be either provided as a text file [8 tab delimited columns: year, month, day, precipitation (mm), potential evapotranspiration (PET: mm), streamflow (mm), min and max daily temperature (°C)], or as regular format of the MOPEX data set (".dly" extension).

The calibration step utilizes a gradient-based "interior-point" optimization algorithm (Byrd et al. 2000; Waltz et al. 2006) to find the "optimal"/"best" parameter set that minimizes the "sum of squared residuals" between model simulated streamflow and its observed counterpart. This approach is specifically amenable to conditional (bounded) parameter spaces, and finds the optimal parameter set by searching the interior (feasible) region. The Hessian matrix in this approach is estimated through a dense quasi-Newton approximation (Sadegh et al. 2017). The "interior-point" optimization algorithm, however, like any other local optimization approach, is susceptible to finding local optimum. It is suggested that the calibration step is repeated multiple times to ensure the global optimum solution is found. In the NRRT toolbox, we refrain from using global optimization approaches due to their computationally extensive and time consuming nature. Our experience shows HyMod and HBV are more prone to local optima, due to their complex and rigged response surfaces, and GR4J has shown less susceptibility in this sense. Although uncertainty analysis is beyond the scope of NRRT, one can also derive posterior distribution of model parameters. See SI for several case studies analyzed with Markov Chain Monte Carlo (MCMC) simulation. MCMC numerically solves the Bayes equation and returns the posterior distribution of model parameters (Sadegh et al. 2017). MCMC codes are also provided as part of the toolbox package for future users.

Upon selection of the model in NRRT, its parameter names will appear on the analysis section and their calibrated values will be assigned. The analysis section, then, allows for out-of-sample evaluation of the model given the derived parameters. The package permits only "one" parameter to vary in time in order to simulate nonstationary behavior of the watershed. In essence, more parameters can be time-varying; however, interaction between multiple parameters can dilute the impacts of each on the streamflow simulation, and this might complicate the interpretation of the results. The user should then select the time-varying parameter, assign the period that watershed experienced physical changes, and set the time-varying parameter at the start and finish points of watershed nonstationarity period.

In both calibration and analysis steps, the user can conveniently observe model simulations and contrast model response with the observed values within the NRRT's GUI. Some performance metrics, including Nash-Sutcliffe efficiency (NSE), root mean square error (RMSE), and percent bias (PBIAS) are also reported to the user on the screen.

4 Case Study

We focus on an experimental watershed, namely Wights catchment, in Western Australia to illustrate the efficacy of the proposed framework and toolbox. This small basin has a drainage area of 0.94 km^2, with cold, wet winters and hot, dry summers (Pathiraja et al. 2016a). As a study watershed, it was deforested and converted to pasture approximately three years after commencement of a monitoring system (Mroczkowski et al. 1997). The catchment underwent deforestation in the period of Nov. 01, 1976 to Mar. 10, 1977, which coincides with summer in the Southern Hemisphere. Daily precipitation, temperature, streamflow and potential evapotranspiration observations between 04/24/1974 and 04/06/1990 are available for this catchment. Refer to Figure 1 of Pathiraja et al. (2016a) for a map of the Wights catchment, and to Bettenay et al. 1980 for a detailed description of the watershed characteristics. Note that we also have analyzed the Skykomish watershed in WA, USA (USGS station # 12134500), results of which are provided in the SI (Figure S6).

5 Application and Results

In this section, we focus on simulating the streamflow response of the Wights catchment to climatic drivers (precipitation and potential evapotranspiration) using GR4J model. Analysis of this watershed using HyMod and HBV models are provided in the SI (figures S7-S12 for HyMod and S13-S18 for HBV). The Wights catchment has experienced three phases: 1. forested landscape (before Nov. 31, 1976), 2. transitional phase of deforestation (Nov. 1, 1976 to Mar. 10, 1977), and 3. pasture landscape (after Mar. 11, 1977). We first tune the GR4J model parameters against 5 years of observed data (Mar. 11, 1977 to Mar. 10, 1982) in the third phase (pasture landscape). Figure 2 shows the observed precipitation (blue bars) and snow accumulation/melt (green bar) (4A), potential evapotranspiration (black dots) and GR4J simulated actual evapotranspiration (red line) (4B), and GR4J simulated streamflow (red line) against its observed counterpart (blue dots) (4C). There is no snow accumulation in the period of observation due to the above freezing temperatures. The GR4J simulations nicely track the observed response of the watershed, with acceptable performance metrics of NSE = 0.776 (−), RMSE = 0.009 m^3/s, and PBIAS = −11.9% (Fig. 4c).

Evaporation module in the GR4J model accurately represents the real world processes in the Wights catchment, which plays a vital role in the success of this parsimonious model to precisely simulate streamflows. In GR4J, precipitation is first absorbed by the interception layer, and excess precipitation (coined as net precipitation) is divided between the production store ($S1_{max}$) and the routing unit hydrographs. This is well aligned with the response of the Wights catchment to climatic drivers, as sparse and occasional precipitation in the summer months (Dec-Feb) rarely result in streamflow at the catchment outlet. In fact, the precipitation event of Feb. 10–11, 1980 is a good illustrative case for this argument. A relatively large precipitation of 38 mm on Feb. 10, 1980 followed by 13 mm on Feb. 11, 1980 resulted in almost zero observed runoff response at the catchment outlet, which is adequately captured by the model.

We now use the calibrated model (with parameter values: $S1_{max} = 117.555$, Exch = −9.994, $S2_{max} = 63.933$, UHB = 1.404, DD = 3.499) to simulate the Wights catchment's streamflow in the period of Apr. 25, 1975 to Mar. 10, 1982. Notice that this is a partially out-of-sample period for model evaluation; and covers all three phases of the catchment (forested, transitional, and

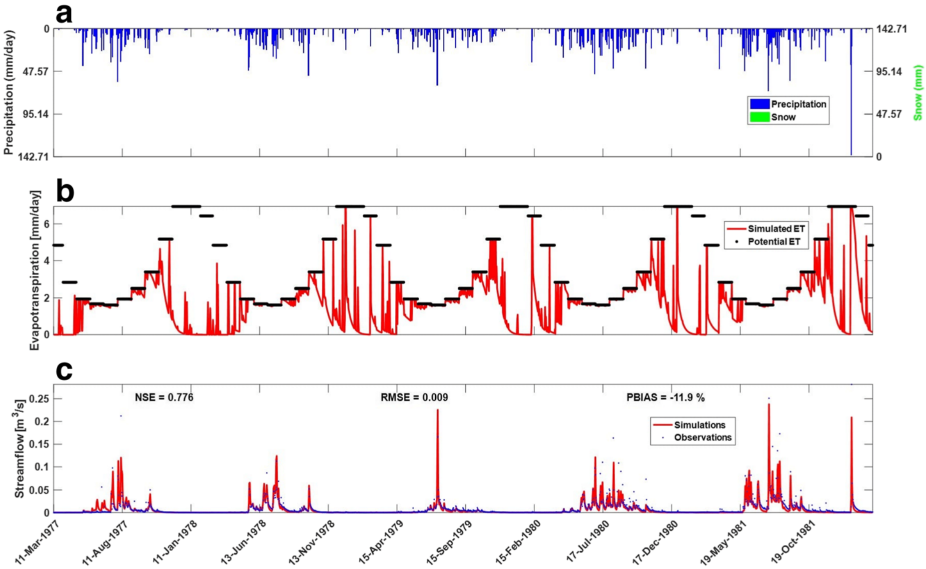

Fig. 2 Observed precipitation (**a**), potential and simulated actual evapotranspiration (**b**), and observed and simulated streamflow for the Wights catchment (**c**). This figure shows GR4J simulations for the period of calibration (Mar. 11, 1977 to Mar. 10, 1982)

pasture landscape). Figure 3c depicts the simulated and observed streamflow response of the under study catchment. A sharp decline in the NSE value in the evaluation period compared to

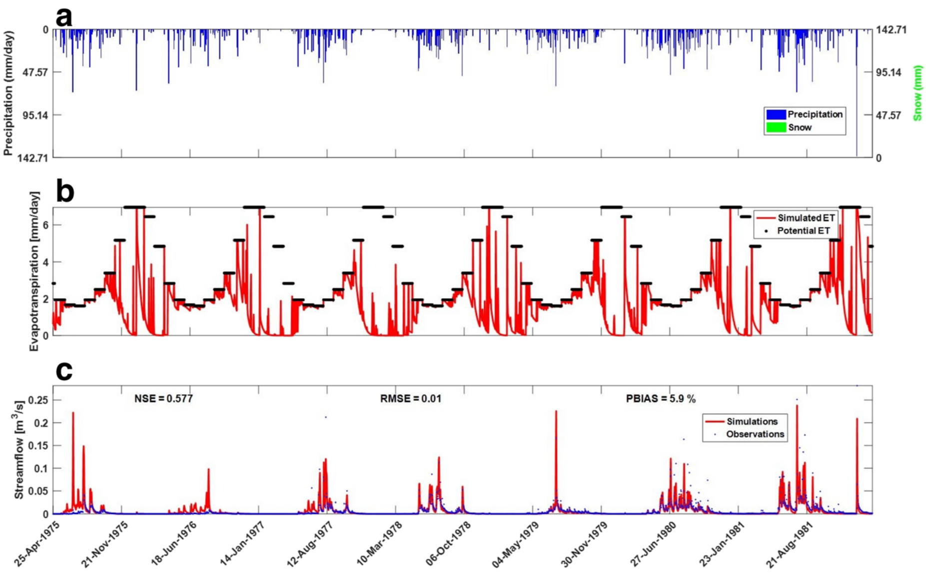

Fig. 3 Observed precipitation (**a**), potential and simulated actual evapotranspiration (**b**), and observed and simulated streamflow for the Wights catchment (**c**). This figure shows evaluation of the calibrated stationary GR4J model in the period of Apr. 25, 1975 to Mar. 10, 1982. This includes an out-ofsample period of Apr. 25, 1975 to Mar. 11, 1977 during which the Wights catchment was forested, and a transitional deforestation period, before reaching a new equilibrium of pasture landscape

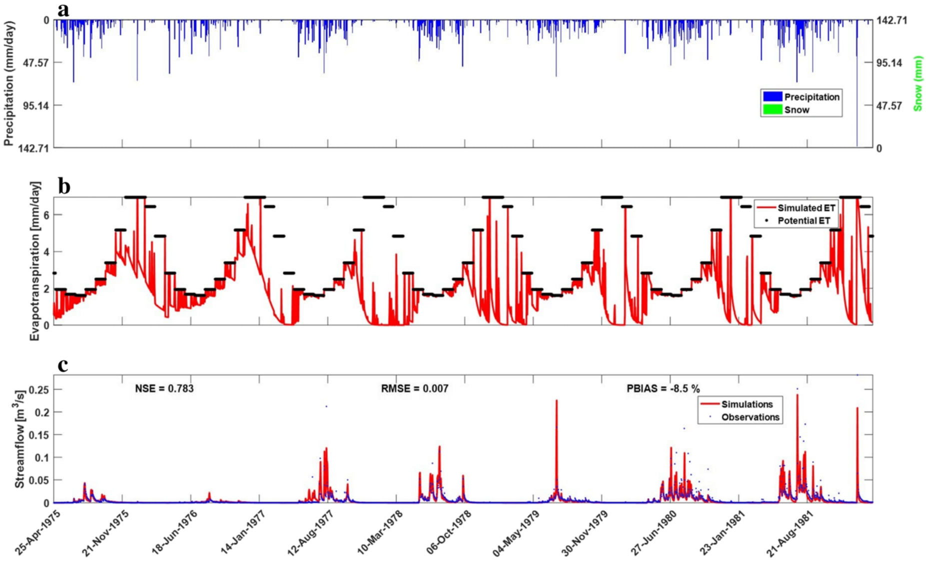

Fig. 4 Observed precipitation (**a**), potential and simulated actual evapotranspiration (**b**), and observed and simulated streamflow for the Wights catchment (**c**). This figure shows evaluation of the calibrated GR4J model in the period of Apr. 25, 1975 to Mar. 10, 1982, with a time-varying S1maxparameter that linearly decreases from 650 (mm) to 117.555 (mm) over the period of Nov. 1, 1976 to Mar. 10, 1977

that of the calibration interval (0.577 vs 0.776) shows that the model performance has deteriorated for the out-of-sample period. More specifically, model simulations in the period of Jul.-Sep. of 1975 and Jun.-Aug. of 1976 significantly overpredict observed streamflows. This is rather expected due to the conflicting nature of the model assumptions (pasture land cover) and the physical state (forested land cover) of the watershed. The GR4J model with a fixed structure assumes that watershed is stationary, whereas the Wights catchment has experienced three different physical states in the evaluation period.

Before Nov. 31, 1976 the Wights catchment had a forested landscape, which was converted to a pasture landscape in the transitional period of Nov. 1, 1976 to Mar. 10, 1977. The GR4J model parameters are, however, calibrated against the pasture landscape conditions. One obvious insight in comparing the forested versus pasture landscape is the higher storage capacity of the forest due to interception and near surface storage of the root zone. Forested landscape also withdraws soil moisture through the root zone and impacts the runoff generation. One would expect the pasture landscape to generate a higher runoff compared to the forested landscape in response to a similar rainfall. This is in fact the driver of the over prediction of streamflow by the GR4J model calibrated against pasture conditions when evaluated against the forested state (see periods of 20 Jul.-Sep. of 1975 and Jun.-Aug. of 1976 in Fig. 3c).

The changes in the interception and near surface storage capacities can be vicariously addressed through a time-varying parameter of maximum production store capacity ($S1_{max}$). We can assign the nonstationarity period (Nov. 1, 1976 to Mar. 10, 1977) and select the $S1_{max}$ parameter of the GR4J model as time-varying in NRRT to improve streamflow simulations. Next step is to define the starting and final value of $S1_{max}$. Selection of the final value is straight-forward as the period of Mar. 11, 1977 onward coincides with the calibration period. Therefore, we set the final $S1_{max}$ value to 117.555 (mm) as inferred in the calibration step. The starting value of $S1_{max}$, however, should be inferred through trial-and-error, considering that it is bounded by a lower threshold of 117.555 (mm).

We estimated a starting $S1_{max}$ value of 650 (mm) or greater warrants adequate representation of the rainfall-runoff generation process for the Wights catchment. Figure 4c presents the GR4J simulated streamflow in the period of Apr. 25, 1975 to Mar. 10, 1982 with a time-varying $S1_{max}$ that linearly reduces from 650 (mm) to 117.555 (mm) during the period of Nov. 1, 1976 to Mar. 10, 1977. Other GR4J parameters are set to their calibrated values during the entire simulation period. Updated model simulations can indeed track the observed streamflow in all three phases of the watershed conditions. The satisfactory performance metrics of NSE = 0.783, RMSE = 0.007, and PBIAS = −8.5% corroborate our hypothesis that a time-varying model realization can adequately represent the nonstationary behavior of the Wights catchment.

Mass conservation is the founding pillar of most conceptual rainfall-runoff models, including GR4J, and hence the precipitation input should leave the watershed either in the form of streamflow or evapotranspiration. The winter and early spring (Jul.-Sep.) precipitation of 1975 (forested phase) is partially routed through the watershed as streamflow, and a significant portion of the precipitation is cached in the production store. The stored water in turn contributes to the evapotranspiration in the late spring and summer months (Nov.-Feb. 1975) as vividly depicted in Fig. 4b. These months only experience a few precipitation events (Fig. 4a), whereas the simulated evapotranspiration of the catchment is very high in this period. Indeed, comparison of Figs. 3b and 4b shows that the higher value of maximum production store capacity allows the model to cache some of the winter precipitation to satisfy the high summer evapotranspirative demand. This is similar to the true behavior of a forested watershed.

Selection of the time-varying parameter should correspond to the physical changes of the watershed. In the case of the Wights catchment, loss of near surface storage due to deforestation is adequately represented by a decrease in the maximum capacity of the production store ($S1_{max}$) of the GR4J model. If, however, a non-relevant parameter such as the maximum capacity of the routing store ($S2_{max}$) was to be chosen as the time-varying parameter, the simulated streamflow wouldn't improve sufficiently. Figure 5c shows the simulated streamflow by the GR4J model if $S2_{max}$ varies from 200 (mm) to 63.933 (mm) in the nonstationarity period. Such modification of the model realization fail to adequately represent the underlying physical changes of the Wights catchment. Performance metrics of this model configuration (NSE = 0.625, RMSE = 0.01, and PBIAS = 16.3%) are expectedly inferior to those of the time-varying $S1_{max}$ configuration.

Some final remarks are worth noting here:

1. Due to the temperature distribution in the Wights catchment, snow accumulation/melt module is not activated, and hence the DD parameter value is not activated.
2. The user should be aware of the parameter interactions/compensations in the conceptual rainfall-runoff models. In the NRRT toolbox, we have allowed for one parameter to vary at a time to minimize such interactions. In essence, any number of parameters can vary in time.
3. It is possible that modification of a seemingly non-relevant parameter yields some level of improvement in the simulated response of the watershed. It is therefore suggested that the user relates the parameter role to the changes of the watershed in a physically meaningful way.
4. In this analysis, we have a clear understanding of the physical changes of the watershed, and their onset and offset time. However, this might not be the case for some watersheds. In that case, the user should play around with the nonstationarity period and parameter start and final values simultaneously, in search of a model realization that sufficiently

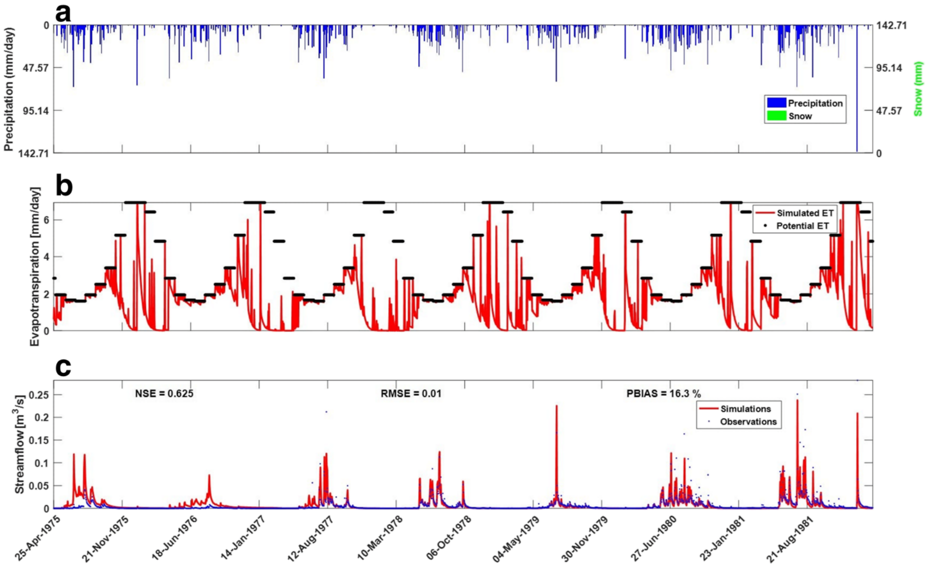

Fig. 5 Observed precipitation (**a**), potential and simulated actual evapotranspiration (**b**), and observed and simulated streamflow for the Wights catchment (**c**). This figure shows evaluation of the calibrated GR4J model in the period of Apr. 25, 1975 to Mar. 10, 1982, with a time-varying S2maxparameter that linearly decreases from 200 (mm) to 63.933 (mm) over the period of Nov. 1, 1976 to Mar. 10, 1977

describes the rainfall-runoff transformation of the watershed. Literature recommends using data assimilation techniques (Pathiraja et al. 2016a, b) and approximate Bayesian computation (Sadegh et al. 2015) to detect the onset and offset time of the nonstationarity period.

5. We have only shown the GR4J model simulations in this manuscript, and results of other models are presented in the SI. Similar conclusions are drawn using HyMod and HBV models. Model comparison is beyond the scope of this paper, but we argue that explicit representation of interception in the GR4J model better imitate the real world processes of the Wights catchment compared to HyMod and HBV.
6. Focus of this paper is on introducing the NRRT toolbox and hypothesis testing framework. The user can select any parameter in the model of choice to vary in time, and use the NRRT toolbox to test different hypotheses. Selected parameter might refer to a change in groundwater-surface water interactions, for example, and need not necessarily address land cover change, as discussed herein.

6 Conclusions

Milly et al. (2008) argue that "*stationarity is dead*" (contested in Lins and Cohn (2011), see also Koutsoyiannis (2006, 2011); Koutsoyiannis and Montanari (2015); Sadegh et al. (2015) and references therein). International Association of Hydrological Sciences (IAHS) named the decade of 2013–2022 as "*Panta Rhei – Everything Flows*" to echo the importance of this concept. In the face of an environment that is changing over time due to natural and anthropogenic disturbances, a time-invariant conceptual model structure/parameterization, as employed traditionally, may not be sufficient for predicting the behavior of a nonstationary system.

This paper presents the Nonstationary Rainfall-Runoff Toolbox, (NRRT), for multi-model (one at a time) nonstationary modeling of rainfallrunoff (RR) processes. The designed time-varying nature of the available conceptual RR models (GR4J, GR5J, GR6J, HyMod, and HBV) in the NRRT toolbox allows for simulating processes that are nonstationary in essence. NRRT translates the potentially time-varying changes in parameters into model simulations in order to imitate the nonstationary nature of the watershed processes. We have shown that a time-varying parameterization of conceptual RR models, if corresponds to the physical changes of a watershed, is indeed capable of modeling the response of a nonstationary watershed to climate forcing. In our case study, the Wights catchment, loss of near surface storage due to deforestation is adequately represented by a decrease in the maximum capacity of the production store ($S1_{max}$) of the GR4J model.

This toolbox is freely available to public, and available for download from http://coen.boisestate.edu/hydroclimate/softwares/.

Acknowledgements This study was partially supported by the United States National Science Foundation Award No. *CMMI* −1635797, ARO's Environmental Sciences Division Award No. W911NF-14-1-0684, and National Oceanic and Atmospheric Administration (NOAA) Award No. NA14OAR4310222.

Compliance with Ethical Standards

Conflict of Interest None.

References

Aghakouchak A, Habib E (2010) Application of a conceptual hydrologic model in teaching hydrologic processes. Int J Eng Educ 26:963–973

AghaKouchak A, Nakhjiri N, Habib E (2013) An educational model for ensemble streamflow simulation and uncertainty analysis. Hydrol Earth Syst Sci 17:445–452

Bergström S (1992) The HBV model: its structure and applications. Swedish Meteorological and Hydrological Institute, Report 4, Norrköping, Sweden

Bettenay E, Russel WRG, Hudson DR, Gilkes RJ (1980) A description of experimental catchments in the collie area, Western Australia, tech. Pap. 7. Land Resour. Manage., Perth

Beven K, Freer J (2001) Equifinality, data assimilation, and uncertainty estimation in mechanistic modelling of complex environmental systems using the GLUE methodology. J Hydrol 249:11–29

Boyle DP (2001) Multicriteria calibration of hydrologic models, PhD Thesis, Department of Hydrology and Water Resources Engineering, The University of Arizona

Boyle DP, Gupta HV, Sorooshian S (2000) Toward improved calibration of hydrologic models: combining the strengths of manual and automatic methods. Water Resour Res 36:3663–3674

Brown AE, Podger P, Davidson A, Dowling T, Zhang L (2006) A methodology to predict the impact of changes in forest cover on flow duration curves. CSIRO Land and Water Science Report 8/06, Canberra

Byrd RH, Gilbert JC, Nocedal J (2000) A trust region method based on interior point techniques for nonlinear programming. Math Program 89:149–185

Cheng L, AghaKouchak A, Gilleland E, Katz RW (2014) Non-stationary extreme value analysis in a changing climate. Clim Chang 127:353–369

Efstratiadis A, Nalbantis I, Koutsoyiannis D (2015) Hydrological modelling of temporally-varying catchments: facets of change and the value of information. Hydrol Sci J 60:1438–1461

Gharari S, Hrachowitz M, Fenicia F, Savenije H (2013) An approach to identify time consistent model parameters: sub-period calibration. Hydrol Earth Syst Sci 17:149–161

Grenier Y (1983) Time-dependent ARMA modeling of nonstationary signals. IEEE Trans Acoust Speech Signal Process 31:899–911

Koutsoyiannis D (2006) Nonstationarity versus scaling in hydrology. J Hydrol 324:239–254

Koutsoyiannis D (2011) Hurst-Kolmogorov dynamics and uncertainty1. J Am Water Resour Assoc 47:481–495

Koutsoyiannis D, Montanari A (2015) Negligent killing of scientific concepts: the stationarity case. Hydrol Sci J 60:1174–1183
Le Moine N (2008) Le bassin versant de surface vu par le souterrain: une voie d'amélioration des performances et du réalisme des modèles pluie-débit?, Ph.D. thesis, Paris 6
Leclerc M, Ouarda TB (2007) Non-stationary regional flood frequency analysis at ungauged sites. J Hydrol 343: 254–265
Lins HF, Cohn TA (2011) Stationarity: wanted dead or alive? J Am Water Resour Assoc 47:475–480
Marshall L, Sharma A, Nott D (2006) Modeling the catchment via mixtures: issues of model specification and validation. Water Resour Res 42:W11409
Milly PCD, Betancourt J, Falkenmark M, Hirsch RM, Kundzewicz ZW, Lettenmaier DP, Stouffer RJ (2008) Stationarity is dead: whither water management? Science 319:573–574. https://doi.org/10.1126/science.1151915, http://science.sciencemag.org/content/319/5863/573
Mohammadpour J, Scherer CW (2012) Control of linear parameter varying systems with applications. Springer Science & Business Media, Boston
Mroczkowski M, Raper PG, Kuczera G (1997) The quest for more powerful validation of conceptual catchment models. Water Resour Res 33:2325–2335
Nash JE et al (1960) A unit hydrograph study, with particular reference to British catchments. Proc Inst Civ Eng 17(3):249–282
Niedzwiecki M (2000) Identification of time-varying processes. Wiley, New York
Ouarda T, El-Adlouni S (2011) Bayesian nonstationary frequency analysis of hydrological variables. J Am Water Resour Assoc 47:496–505
Pathiraja S, Marshall L, Sharma A, Moradkhani H (2016a) Detecting non-stationary hydrologic model parameters in a paired catchment system using data assimilation. Adv Water Resour 94:103–119
Pathiraja S, Marshall L, Sharma A, Moradkhani H (2016b) Hydrologic modeling in dynamic catchments: a data assimilation approach. Water Resour Res 52:3350–3372
Perrin C (2000) Vers une amélioration d'un modele pluie-débit au travers d'une approche comparative, Ph.D. thesis, Ph. D. Thesis, INP Grenoble/Cemagref Antony, France
Perrin C, Michel C, Andréassian V (2003) Improvement of a parsimonious model for streamflow simulation. J Hydrol 279:275–289
Pushpalatha R, Perrin C, Le Moine N, Mathevet T, Andréassian V (2011) A downward structural sensitivity analysis of hydrological models to improve low-flow simulation. J Hydrol 411:66–76
Richards JA (1983) Analysis of periodically time-varying systems. Springer Science & Business Media, New York
Sadegh M, Vrugt JA, Xu C, Volpi E (2015) The stationarity paradigm revisited: hypothesis testing using diagnostics, summary metrics, and DREAM (ABC). Water Resour Res 51:9207–9231
Sadegh M, Ragno E, AghaKouchak A (2017) Multivariate copula analysis toolbox (MvCAT): describing dependence and underlying uncertainty using a Bayesian framework. Water Resour Res 53(6):5166–5183
Sadegh M, Majd MS, Hernandez J, Haghighi AT (2018) The quest for hydrological signatures: effects of data transformation on Bayesian inference of watershed models. Water Resour Manag 32(5):1867–1881
Salas JD, Obeysekera J (2013) Revisiting the concepts of return period and risk for nonstationary hydrologic extreme events. J Hydrol Eng 19:554–568
Schaake JC, Koren VI, Duan Q-Y, Mitchell K, Chen F (1996) Simple water balance model for estimating runoff at different spatial and temporal scales. J Geophys Res Atmos 101:7461–7475
Singh VP, Woolhiser DA (2002) Mathematical modeling of watershed hydrology. J Hydrol Eng 7:270–292
Waltz RA, Morales JL, Nocedal J, Orban D (2006) An interior algorithm for nonlinear optimization that combines line search and trust region steps. Math Program 107:391–408
Westra S, Thyer M, Leonard M, Kavetski D, Lambert M (2014) A strategy for diagnosing and interpreting hydrological model nonstationarity. Water Resour Res 50:5090–5113