Catchment Yield Prediction in an Ungauged Basin Using PyTOPKAPI

B. S. Fatoyinbo, D. Stretch, O. T. Amoo, D. Allopi

Abstract—This study extends the use of the Drainage Area Regionalization (DAR) method in generating synthetic data and calibrating PyTOPKAPI stream yield for an ungauged basin at a daily time scale. The generation of runoff in determining a river yield has been subjected to various topographic and spatial meteorological variables, which integers form the Catchment Characteristics Model (CCM). Many of the conventional CCM models adapted in Africa have been challenged with a paucity of adequate, relevance and accurate data to parameterize and validate the potential. The purpose of generating synthetic flow is to test a hydrological model, which will not suffer from the impact of very low flows or very high flows, thus allowing to check whether the model is structurally sound enough or not. The employed physically-based, watershed-scale hydrologic model (PyTOPKAPI) was parameterized with GIS-pre-processing parameters and remote sensing hydro-meteorological variables. The validation with mean annual runoff ratio proposes a decent graphical understanding between observed and the simulated discharge. The Nash-Sutcliffe efficiency and coefficient of determination (R^2) values of 0.704 and 0.739 proves strong model efficiency. Given the current climate variability impact, water planner can now assert a tool for flow quantification and sustainable planning purposes.

Keywords—Ungauged Basin, Catchment Characteristics Model, Synthetic data, GIS.

I. INTRODUCTION

PREDICTING runoff in most ungauged catchment territories is crucial to many viable applications in drainage design infrastructure, runoff forecasting, and many catchment management tasks [1]. However, how benevolent this task is, due to inadequate and lack of stream gauges or historical flow data, poses relative difficultly in many regions; thereby, making catchment yield and runoff prediction using alternative method or information, a recourse challenge. A yield is that portion of precipitation on a catchment that can be collected for use. Catchment yield is necessary to check if sufficient water is available at a chosen site. The most widely employed method in generating data in data scarce regions has been regionalization [2]-[5].

Recently coupled physically based models in the GIS environment with remote sensed data have aided physical meaning on basin characteristics and hydrological processes at the catchment scale [6], [7]. However, the accuracy of the application these hydrological models at an ungauged basin is still based on the proximity of the two or several sites that shared similar physical or climatic hydrological characteristics [8].

This paper presents a regionalization approach in calibrating a CCM named 'Python implementation of TOPographic Kinematic Approximation and Integration' (PyTOPKAPI) model as a tool for evaluating water yield and runoff generation at a daily time step in the ungauged Mhlanga Basin, KwaZulu-Natal, South Africa. This study improves and helps eliminate uncertainty associated with low flow over estimation and safe yield prediction recorded in previous studies [17]. The versatility of the PyTOPKAPI model for an ungauged basin streamflow simulation is tested and validated with DAR synthetic data. Also, the main problem with the use of synthetic data is that the model is tested with respect to the assumptions made in generating the synthetic data. Thus, the present study shows the behavioral characteristics of the catchment response and sensitivity of the model's parameter.

The 'PyTOPKAPI' model is a physically-based distributed hydrological model developed as an improvement to the earlier TOPKAPI model [9]. It is coded in python programming language and can be accessed directly through an interactive PYTHON environment in a computer operating system before use. The model was tested on Liebenbergsvlei catchment in South Africa to simulate river discharge at 6-hour time-step. The initial TOPKAPI model, consist of five modules (soils, overland, channel, evapotranspiration and snow) [10]. However, the infiltration process was not properly accounted for at the initial. Consequently, an improved PyTOPKAPI which incorporates a true representation of infiltration module (Green Ampt) was developed to produce a quick overland runoff when exposed to high rainfall and vice versa [9]. The DAR method has found application in several studies [8], [11]-[13]. Its approach of transposing data from donor to target catchment for river discharge prediction in an unknown catchment have been found to be relatively simple with minimal data requirement compared to other regionalization methods [12]. DAR works by simulating hydrologic response basin to geographic proximity of the ungauged watershed of interest [14]. Mohamoud [13] suggests choosing a basin with the closest stream gauge, while Smakhtin [15] suggests that several reference stream gauges by proxy should be used in order to smooth out any timing-related issues between the ungauged and reference locations.

B. S. Fatoyinbo is with Centre for Research in Environmental, Coastal & Hydrological Engineering, School of Engineering University of Kwa-Zulu-Natal, Durban 4041, South Africa. (e-mail: bukfat2009@gmail.com).

O. T. Amoo, is with Department of Civil Engineering and Surveying, Durban, University of Technology, Durban, South Africa.

D. Stretch is with Centre for Research in Environmental, Coastal & Hydrological Engineering, School of Engineering University of Kwa-Zulu-Natal, Durban 4041, South Africa.

D. Allopi is with Department of Civil Engineering and Surveying, Durban, University of Technology, Durban, South Africa.

World Academy of Science, Engineering and Technology
International Journal of Environmental and Ecological Engineering
Vol:11, No:3, 2017

Factors like main channel slope, precipitation intensity, in addition to drainage area, contribute significantly to runoff generation in a river catchment [16]. This gives a better streamflow result and corrects the bias in the residuals for the two groups of stations used. Previous studies conducted in the study area (Mhlanga catchment) have used various methods to generate river flow considering that flow records are either not available or are limited in availability [17], [18]. This paper uses a regionalization drainage-area ratio approach in calibrating a CCM Model.

The next section includes the detailed methodology followed during the study, focusing on the work objectives. Section IV provides the results and discussion on the use of DAR for PYTOPKAPI calibration, while Section IV summarizes the main findings and provides a concise recommendation for future research.

II. STUDY AREA

The area under study (Mhlanga) is located around 29º 42 '9"S and 31º 6' 0"E east coast of South Africa with average slope of 0.6%. The area is in quaternary sub-catchment U30B, located within the Mvoti-Umzimkulu water management area. It has a draining area of 80 km² with a mean annual precipitation of 1000 mm, mean annual evapotranspiration of 1210 mm and mean annual runoff of 0.4 m³/s, as obtainable in WR90 [18].
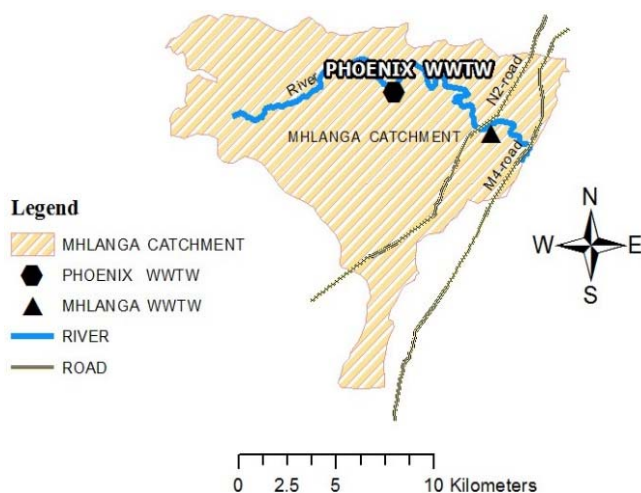

Fig 1 Mhlanga catchment map

The study area is a typical smaller and ungauged catchment in South Africa, characterized with no streamflow data as against that usually available for Quaternary Catchments (QCs); it is confounded by missing streamflow, inaccurate measurement and recording of data, as well as inadequate record length for analysis, which are just some of the problems associated with the assessment of the region's water resources. Despite ever increasing efforts since 1960 to automate and record daily streamflow, most of the automatic records are characterized with frequently gaps or inaccuracies, especially after extreme flood events when gauging plates require re-calibration and are temporarily out of operation. In-filling or extending daily streamflow records using interpolation techniques and reference gauging stations in catchments with highly variable river systems, or where there are ephemeral rivers, is problematic [19].The study area is as represented in Fig. 1.

III. MATERIALS AND METHODS

Hydro-metrological data over a period of 16 years (1999-2014) consisting of daily rainfall and potential-evapotranspiration) was collected from SASRI (South African Sugarcane Research Institute) weather web. The data was subjected to trend analysis and further processing in PyTOPKAPI model. Similarly, daily historical streamflow records for the Mdloti gauging river station (U3H001), a neighboring catchment to the study area, were obtained from the Department of Water and Sanitation (DWS). This data were to be transposed to the other basin to synthesize its discharge.

The daily plot of the rainfall, discharge and evapotranspiration distribution pattern for the study period is presented in Figs. 2 and 3.
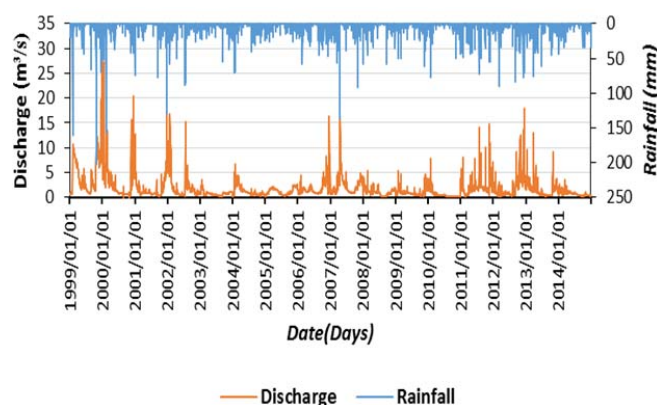

Fig. 2 Plot of daily Rainfall and discharge measured at gauging station U3R001
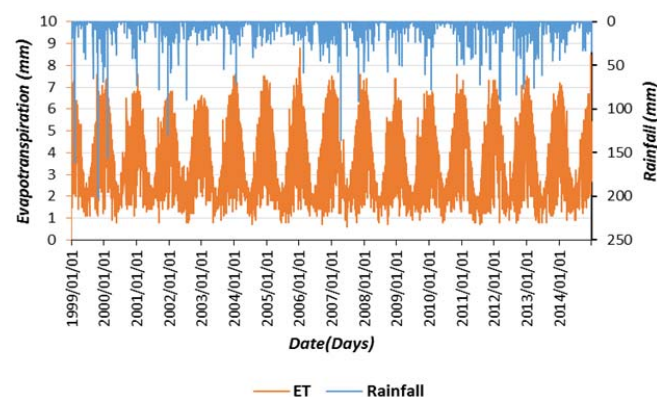

Fig. 3 Plot of daily Rainfall and discharge measured at gauging station SASRI

A. Drainage Area Regionalization

In order to check for the efficiency of the DAR, regionalization method was used to transpose flow data from a

World Academy of Science, Engineering and Technology
International Journal of Environmental and Ecological Engineering
Vol:11, No:3, 2017

nearby station with station number U3H001 and 3.865 km distance apart. The mean annual runoff for the period of 16 years estimated from the method was compared with the reference natural mean runoff. The natural mean annual runoff value was used as a threshold for the estimated DAR method.

The conventional DAR method is the common and easiest method for transposing data [13], [20], [21]. This method is implemented by multiplying the drainage area of an ungauged area to that of a nearby gauging station (1). This method performs best when the proportion of source to the interested site drainage area is within the range 0.5-1.5 [20]. In this study, the proposed method by Mohamoud [13], which addresses the earlier limitation (account for area ratio less than 0.5), was used. This method is referred to in (1) and (2):

$$Q_u = Q_g \left(A_u / A_g \right) \tag{1}$$

$$Q_u = Q_g \tan \left(A_u / A_g \right) \tag{2}$$

DAR was used for generating synthetic flow to test and validate the PyTOPKAPI hydrological model. DAR can effectively predict runoff at a daily time-step. Nearby gauged Stream flow data (Mdloti station) was used to estimate the discharge at the Mhlanga (ungauged) basin. These data were utilized for the model simulation processes (calibration and validation).

B. PyTOPKAPI: Data Requirements and Analysis

A PYTOPKAPI model is data driven, which requires several data types ranging from topographic, land use, soil, and climatic data as input parameters. The topographic data entails a finer resolution (30m) ASTER DEM (Advanced Space-borne Thermal Emission and Reflection Radiometer) sourced from United States Geological Survey (USGS) [31].

The DEM was resampled to 0.5km resolution and further processed for the extraction of flow direction, flow accumulation, stream network for watershed delineation of the catchment and terrain slope. These data were used to develop a true representation of the catchment. Fig. 4 shows the topographic processed maps required for the study.
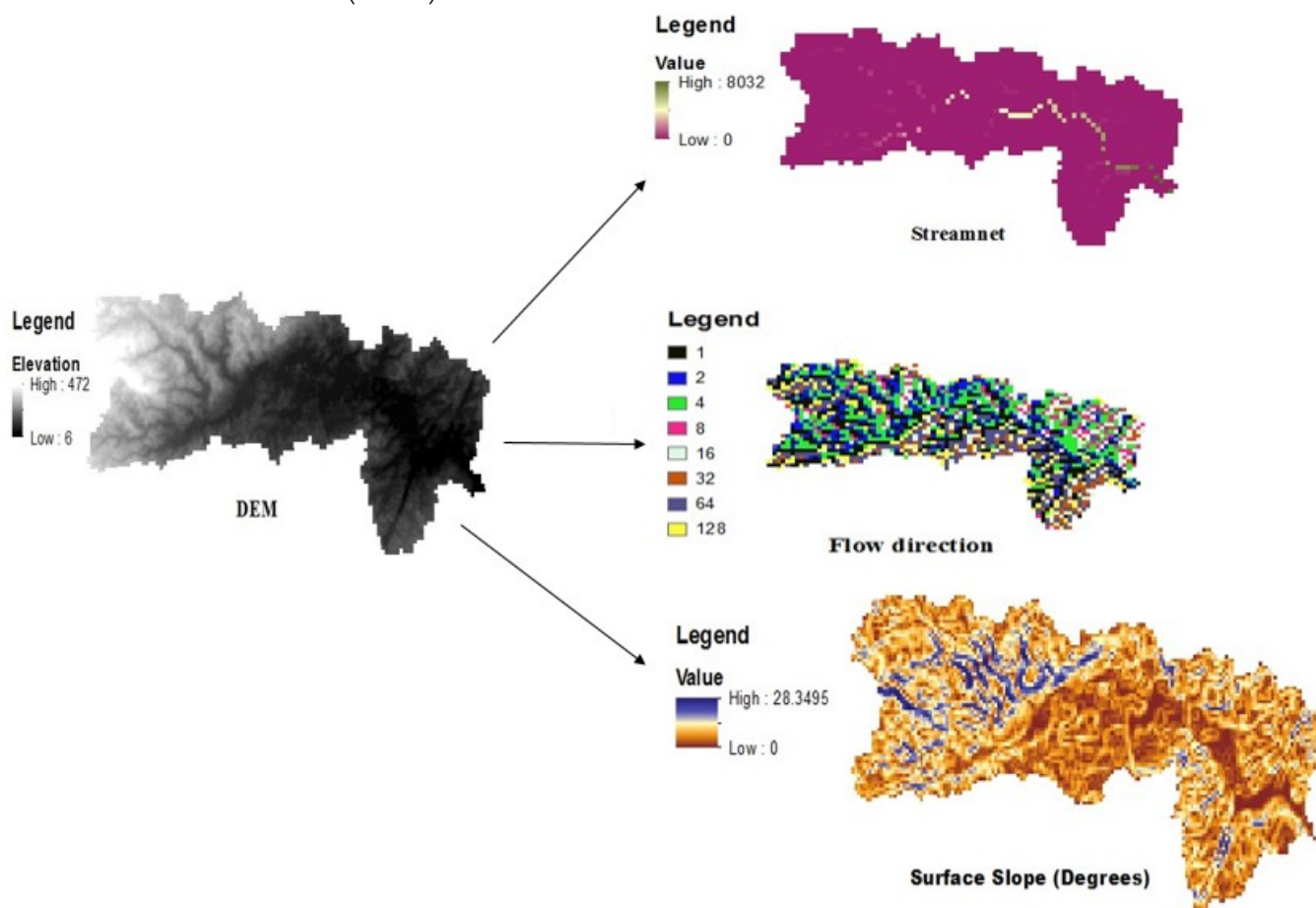
Fig. 4 Topographic data maps for PyTOPKAPI model input

World Academy of Science, Engineering and Technology
International Journal of Environmental and Ecological Engineering
Vol:11, No:3, 2017

C. Land Cover/ Use Data

Land cover data were used to estimate the Manning's roughness coefficient which determines the velocity of the overland flow [22]. The Mhlanga catchment was extracted from a larger Africa land use raster-based at a resolution of 1 km2, obtained from the USGS website [23], and thereafter, downscaled to the same resolution (0.5 km) for use. The land cover map for the catchment is presented in Fig. 5.

D. Soil

The soil map was sourced at 1 km spatial resolution from the Harmonized World Soil database (HWSD) [24]. This was further downscaled to 0.5 km for input into PyTOPKAPI model. Other soil parameters/data were sourced from the literature [25]. Fig. 6 presents the catchment's soil classification map legend as sourced from HWSD database, while Fig. 7 relates the soil texture to its properties, as required by PyTOPKAPI model.
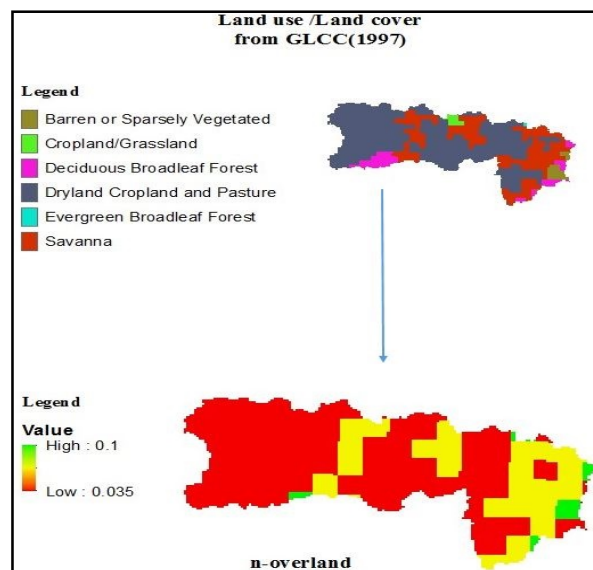

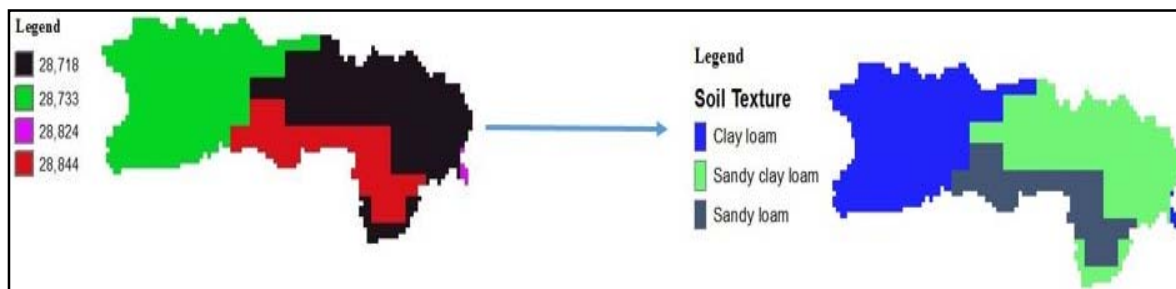Fig. 5 Land cover/use map from [23] and the inferred Manning's roughness map

Fig. 6 Soil map of Mhlanga catchment. The legend identifies the soil in each cell to its textural class from the Harmonized World Soil database (HWSD)
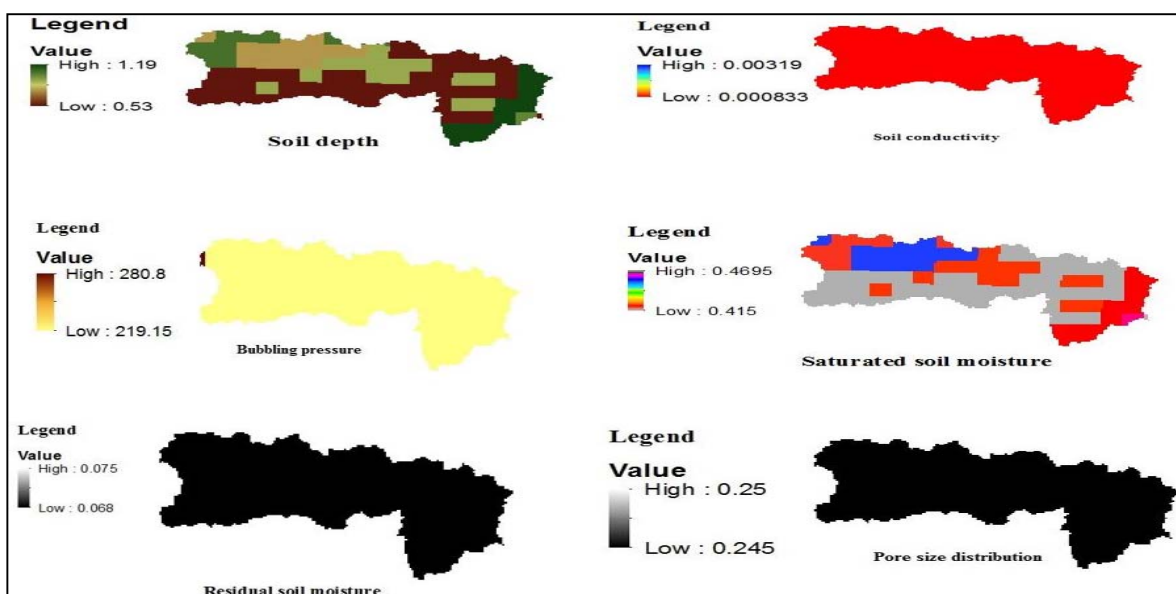

Fig. 7 Maps that relate soil texture to its properties required by PYTOPKAPI

World Academy of Science, Engineering and Technology
International Journal of Environmental and Ecological Engineering
Vol:11, No:3, 2017

TABLE I
PyTOPKAPI PARAMETERS INITIAL AND POST-CALIBRATION MULTIPLYING FACTOR VALUES

Spatially distributed parameters	Parameter range	Source	Post-calibration Multiplying factor value
Ground Slope tangent $tan\beta$	0.0018 - 0.1717	DEM [31]	
Channel slope tangent $tan\beta c$	0.00044 - 0.024	DEM [31]	
Soil layer depth (m) L	1-0.1	Soil type map [26]	Fac_L 2.85.
Saturated hydraulic conductivity (m/s) Ks	6.38E-4 - 7.19E-3	Soil textural map [25]	Fac_K 1.2
Residual soil moisture content (cm³/cm³) θr	41E-2 -75E-2		
Saturated soil moisture content (cm³/cm³) θs	33E-2- 41.2E-2		
Manning's roughness coeff. (surface) no	3E-2 - 12E-2	Land-use map [23], [27]	Fac_{no} 1.0
Manning's roughness coeff.(channel) nc	0 – 5E-2	[28]	Fac_{nc} 1.0
Soil pore size λ	19.4E-2 - 32E-2	Soil textural map [25]	
Soil bubbling pressure $psib$	146.6 - 280.8		
Global parameters			
Cell Dimension(m)	500	DEM	
Max. Channel width (m)	25	Aerial photograph	
Min. Channel width(m)	5	-	
Channel Area(m) A $threshold$	25000000		
Pore size distribution αs	2.5	[28]	
Power coefficient αo & αc	1.667		
time step Δt	86400	-	

E. PyTOPKAPI Model Set-Up

The model was setup by inputting the above process information - topography, soil characteristics, land use, and data obtained from literature. The model set up requires geo-processed input cell, global and forcing files as parameters. These entails catchment boundary, DEM, soil depth, surface slope, saturated and residual moisture content, soil conductivity, manning overland pore size index and bubbling pressure. These parameters constitute the cell parameters required for PyTOPKAPI.

The global parameter file entails geometric characteristics of the channel or grid cell values in the model. These parameters include lateral dimension of the grid cell (x), model time step (Δt), pore size distribution(α_s), power coefficient from manning equation (α_o & α_c) and area with which the cell initiates a river channel ($A_{threshold}$), as well as its maximum and minimum channel width (W_{min} & W_{max}).

The forcing file contains rainfall, reference and actual evapotranspiration data in an HDF5 binary file. This is stored in a 2-D array, each row representing a single time step and each column a single model cell. Table I summarizes the PyTOPKAPI data input requirements, while Figs. 2, 3, and 4 depict the process input files necessary for calibration. The catchment data are necessary to be adjusted due to several inherent errors or implementation before being used as the default input parameters [10].

IV. RESULTS AND DISCUSSION

The extrapolated trend result for rainfall, discharge flow and evaporation, shows the distribution pattern of the datasets. It can be observed that the discharge pattern follow the rainfall and evapotranspiration pattern regardless of the different magnitudes. It can be deduced that both rainfall and evapotranspiration contribute to the discharge magnitude.

The mean annual runoff generated by the drainage area method was 0.386 m³/s; favorable when compared with the actual mean runoff that must not exceed the threshold of 0.4 m³/s [17]. Several factors could lead to the success of DAR used to transpose data. These factors are based primarily on the fact that the region is relatively small and hydrologically homogenous; this is largely because the soil, climate, topography, and basin characteristics are broadly similar throughout the region, creating a fairly predictable hydrologic response.

Calibration was done by varying these parameters through trial and error based on graphical matching between simulated and observed stream flow. This was to obtain the final parameters values without losing its physical representation. Successful model calibration gives better understanding of the catchment behaviors. The model calibration was done for a 10-year period (1999-2008), while validation was done for a six-year period (2009-2014), out of the total 16-year period data. The choice of 10 years was considered to accommodate the span of wet and dry season conditions in the area. This was also in agreement with Foglia, Hill [29] and Li, Wang [30] which suggest that a calibration data series with a span of at least eight years is sufficient to give more consistent optimal parameter values in a more consistent simulation. All other initial parameters values were suitable and retained. The initial mean soil moisture over the catchment was adjusted during the calibration process. The optimal value of 55% was retained. Simulation was carried out for dataset period of six years (2009-2014) using the final parameter values obtained during calibration in Table I to validate the model. The agreement between the observed and simulated was quantified by statistical metrics such as Nash-Sutcliffe modeling efficiency (Nash) and coefficient of determination. The calibration and validation efficiency statistics summary for the model is shown in Table II.

TABLE II
THE CALIBRATION AND VALIDATION EFFICIENCY STATISTICS

	Years	Mean annual runoff (m³/s)
Drainage area ratio	16	0.380
WR90 report[17]	-	0.400
PYTOPKAPI	16	0.359

Efficiency	Calibration(10yrs.)	Validation (6 yrs.)
Nash	0.706	0.704
Coefficient of determination	0.708	0.739

The result of the model simulation and observed synthetic flow both for calibration and validation is shown in Fig. 8 and

World Academy of Science, Engineering and Technology
International Journal of Environmental and Ecological Engineering
Vol:11, No:3, 2017

Fig. 9, while a scatter plot between observed and simulated for validation is presented in Fig. 10 and Fig. 11. It can be deduced from the simulation results that for low, moderate and peak flows, the absolute amount and timing of streamflow variation are well reproduced. Also, in both the calibration and validation period, the results show a rise and decrease in the streamflow value following the rainfall pattern, and thus, the validity of spatial distribution and behavior is assured.
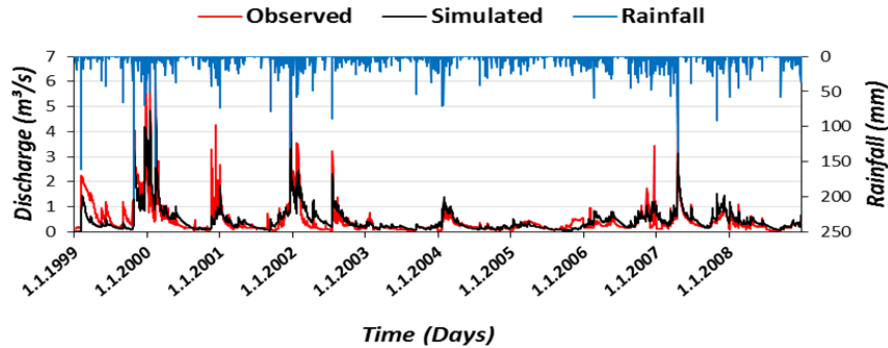
Fig. 8 Plot of simulated and observed discharge for calibration period 1999/01/01 to 2008/12/31
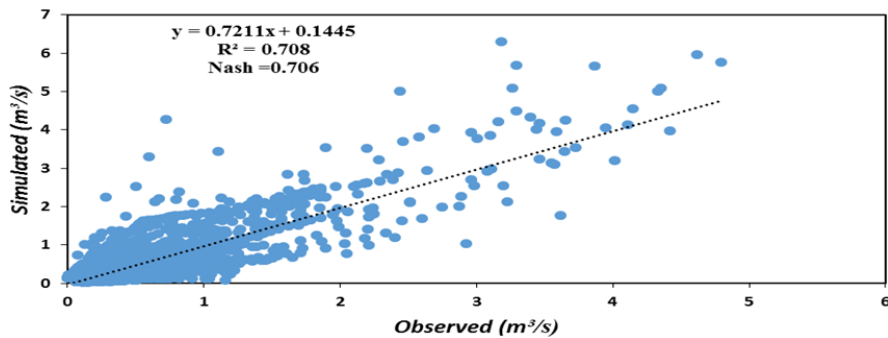

Fig. 9 Scatter plot of simulated and observed daily discharge for the calibration period

In Fig. 8, the observed and calibrated flow, with its rainfall event, were well represented. The good graphical agreement results in a higher value of the Coefficient of determination (R^2) of 0.708 and the Nash-Sutcliffe efficiency of 0.706. A scatter plot between the observed and simulated for validation in Fig. 9 shows a high performance of the model.

Fig. 10 shows a good correspondence between the simulation and observed data set for the period. The model validation performance is shown in Fig. 11 with Coefficient of determination (R^2) and Nash-Sutcliffe efficiency values of 0.739 and 0.704, respectively. The above approach gives satisfactory results.

The results indicate that the PyTOPKAPI model is an effective catchment management that can be applied to an ungauged catchment. Its ability to reproduce high flows will be useful in flood alert application. For runoff prediction at many ungauged rivers, water planners can now assert a tool for flow quantification planning purposes.
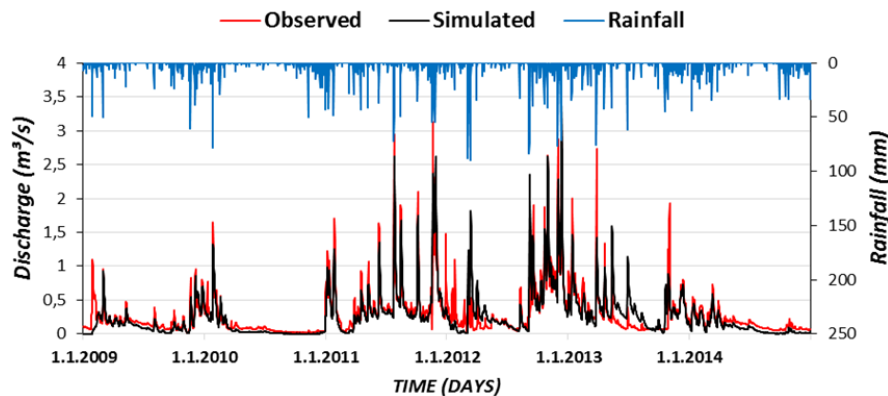

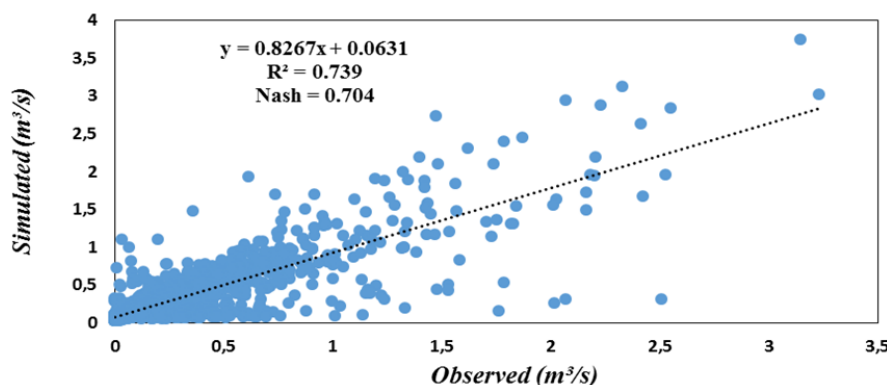Fig. 10 Plot of simulated and observed discharge catchment for validation period 2009/01/01 to 2014/12/31

Fig. 11 Scatter plot of simulated and observed daily discharge for the validation period

V. CONCLUSIONS

In this study, we examined the possibility of calibrating a physically-based hydrological model (PyTOPKAPI) at an ungauged basin to simulate accurately and recreate daily discharge flow. This was achieved by comparing the estimate of mean annual runoff ratio developed for the study area with the PyTOPKAPI model mean annual runoff estimates.

Results of model performance evaluation using- Nash-Sutcliffe efficiency and coefficients of determination (0.704 and 0.739, respectively) also indicate a good simulation performance in reproducing the synthetic daily observed discharge of the Mhlanga catchment. The performance of the DAR was assessed by comparing its mean annual flow to the referenced mean annual flow of the Mhlanga catchment which shows a good agreement.

The results confirm that catchments which are hydrologically similar exhibit similar behaviors. The model was found suitable for simulating the hydrologic configuration of the study area based on the behavioral characteristics of its neighboring catchment, and that the mean annual runoff ratio method and PYTOPKAPI is a useful tool in simulating hydrological processes in ungauged catchments or in catchments with poor or missing datasets. For prediction at many ungauged rivers, the water planner can now assert this tool for flow quantification planning purposes.

REFERENCES

[1] G. Blöschl, Runoff prediction in ungauged basins: synthesis across processes, places and scales. 2013: *Cambridge University* Press.
[2] H. Hellebrand, et al. Identification of hydrologically similar basins by means of parameter transfer with a view to regionalization. in *EGU General Assembly Conference Abstracts*. 2015.
[3] T. Razavi and P. Coulibaly, Streamflow prediction in ungauged basins: review of regionalization methods. *Journal of Hydrologic Engineering*, 2012. 18(8): p. 958-975.
[4] J. Song, et al., Streamflow prediction in ungauged basins by regressive regionalization: a case study in Huai River Basin, China. *Hydrology Research*, 2015: p. nh2015155.
[5] I.K. Westerberg, et al., 2016 Uncertainty in hydrological signatures for gauged and ungauged catchments. *Water Resources Research*, 2016.
[6] Y. Chen, GIS and remote sensing in hydrology, water resources and environment. 2004: *International Assn of Hydrological* Sciences.
[7] S. Khatami, and B. Khazaei, Benefits of Gis ApplicAtion in HydroloGicAl ModelinG: A Brief Summary. VATTEN–*Journal of Water Management and Research*, 2014. 70(1): p. 41-49.
[8] C.C. Gianfagna, et al., Watershed area ratio accurately predicts daily streamflow in nested catchments in the Catskills, *New York. Journal of Hydrology*: Regional Studies, 2015. 4: p. 583-594.
[9] S. Sinclair, and G.G.S. Pegram, *A* sensitivity assessment of the TOPKAPI model with an added infiltration module. *Journal of Hydrology*, 2013. 479: p. 100-112.
[10] Z. Liu, M.L. Martina, and E. Todini, Flood forecasting using a fully distributed model: application of the TOPKAPI model to the Upper Xixian Catchment. *Hydrology and Earth System Sciences* Discussions, 2005. 9(4): p. 347-364.
[11] A. Dahl, D. Emerson, and A. Veccia, Evaluation of Drainage-Area Ratio Method Used to Estimate Streamflow for the Red River of the North Basin, *North Dakota and Minnesota.* 2006.
[12] K. Ergen, and E. Kentel, An integrated map correlation method and multiple-source sites drainage-area ratio method for estimating streamflows at ungauged catchments: A case study of the Western Black Sea Region, Turkey. *Journal of environmental management*, 2016. 166: p. 309-320.
[13] Y.M. Mohamoud, Prediction of daily flow duration curves and streamflow for ungauged catchments using regional flow duration curves. *Hydrological Sciences Journal*, 2008. 53(4): p. 706-724.
[14] S.A. Archfield, and R.M. Vogel, Map correlation method: Selection of a reference streamgage to estimate daily streamflow at ungaged catchments. *Water Resources Research*, 2010. 46(10).
[15] V.Y. Smakhtin, Generation of natural daily flow time-series in regulated rivers using a non-linear spatial interpolation technique. Regulated Rivers: *Research & Management*, 1999. 15(4): p. 311-323.
[16] D.G.Emerson, A.V. Vecchia, and A.L. Dahl, Evaluation of drainage-area ratio method used to estimate streamflow for the Red River of the North Basin, North Dakota and Minnesota. 2005: *US Department of the Interior*, US Geological Survey.
[17] R.A. Lawrie,D.D.Stretch, and R. Perissinotto, The effects of wastewater discharges on the functioning of a small temporarily open/closed estuary. *Estuarine, Coastal and Shelf Science*, 2010. 87(2): p. 237-245.
[18] D. Stretch, and I. Zietsman, Hydrodynamics: Flows, Residence Times, Water Levels and Mouth Dynamics Contributions to Information

World Academy of Science, Engineering and Technology
International Journal of Environmental and Ecological Engineering
Vol:11, No:3, 2017

Requirements for the Implementation Of Resource Directed Measures for Estuaries, in Response of the Biological Communities to Flow Variation and Mouth State in Two KwaZulu-Natal Temporarily Open/closed Estuaries. 2004, *JB Adams Pretoria*, South Africa.

[19] R.,Schulze, and A.Pike, Development and Evaluation of an Installed Hydrological Modelling System. *Water Research Commission, Pretoria,* South Africa, 2004.

[20] L. Fry, et al., Identifying streamgage networks for maximizing the effectiveness of regional water balance modeling. *Water Resources Research*, 2013. 49(5): p. 2689-2700.

[21] K. Nruthya, and V. Srinivas, Evaluating Methods to Predict Streamflow at Ungauged Sites using Regional Flow Duration Curves: *A Case Study.* Aquatic Procedia, 2015. 4: p. 641-648.

[22] V.T., Chow, Open channel flow. *MacGraw-Hill Book Co. Inc*.: New York, 1959.

[23] GLCC, Global Land Cover Characteristics. 1997: United States Geological Survey (USGS).Available: http://edcdaac.usgs.gov/glcc. (Accessed 22 may 2016).

[24] F. Nachtergaele, et al., Harmonized World Soil Database (HWSD). *Food and Agriculture Organization of the United Nations,* Rome, 2009.

[25] W.D. Rawls,. Brakensiek, and K. Saxtonn, Estimation of soil water properties. *Transactions of the ASAE,* 1982. 25(5): p. 1316-1320.

[26] I. FAO, and I. ISRIC, JRC (2012) Harmonized World Soil Database (version 1.2*). FAO, Rome*, Italy and IIASA, Laxenburg, Austria, 2014.

[27] K.O., Asante, et al., Technical manual for the geospatial stream flow model (GeoSFM*)*. 2008*, U. S. Geological Survey.*

[28] The Théo Vischel, et al., Implementation of the TOPKAPI model in South Africa:Initial results from the Liebenbergsvlei catchment.

[29] L. Foglia, et al., Sensitivity analysis, calibration, and testing of a distributed hydrological model using error-based weighting and one objective function. *Water Resources Research*, 2009. 45(6).

[30] C.-z. Li, et al., Effect of calibration data series length on performance and optimal parameters of hydrological model. *Water Science and Engineering*, 2010. 3(4): p. 378-393.

[31] USGS website, 2016. Available: http://earthexplorer.usgs.gov (Accessed 22 may 2016).