International Congress on Environmental Modelling and Software

1st International Congress on Environmental Modelling and Software - Lugano, Switzerland - June 2002

Jul 1st, 12:00 AM

Development of a Cell-Based Parsimonious Catchment Model

Okan Fistikoglu

Nilgun B. Harmancioglu

Development of a Cell-Based Parsimonious Catchment Model

Okan Fistikoglu[a] and Nilgun B. Harmancioglu[a]

[a] *Dokuz Eylul University, Faculty of Engineering, Buca, Izmir, Turkey (okan.fistikoglu@deu.edu.tr) (nilgun.harmancioglu@deu.edu.tr)*

Abstract: One of the crucial problems in conceptual rainfall runoff models is over-parameterization. In most cases, as the number of parameters to be calibrated increases, model performance either does not improve or, it may even decrease due to poorly defined parameters. Simple daily-based rainfall-runoff models are generally lumped models that average catchment heterogeneity. To increase the spatial resolution of such models, the computational resolution may be changed from catchment scale to a cell-based scale. This change may require a modification of the model structure by addition of new parameters. Such an approach may lead to two problems: increase in the complexity of model structure and over-parameterization. In the study presented, a cell-based, parsimonious, distributed, continuous, conceptual daily rainfall-runoff model (DRRSM) with a minimum amount of parameters is introduced. The data needs of the developed model include only daily rainfall, pan evaporation, DEM (digital elevation map), land cover distribution and soil properties which can be obtained easily from various sources through the Internet. The model has proved to be successful in cases where available data on catchment characteristics are insufficient to meet the calibration needs of over-parameterized complex models.

Keywords: Rainfall-runoff models; Parsimony; Cell-based model; DRRSM

1. INTRODUCTION

Current developments in data collection and manipulation systems such as digital sampling equipment, remote sensing technologies (RS) and geographical information systems (GIS) enable the current hydrologic modeling systems to integrate them with the new complex algorithms and computation techniques. As new developments occur, new problems arise, such as over parameterization of the models and the difficulty in the use of complex models due to comprehensive data requirements.

At present, one of the crucial problems in conceptual rainfall runoff models is over-parameterization. In most cases, as the number of parameters to be optimized increases, model performance either does not improve or, it may even decrease due poorly defined parameters. Larger number of parameters and increased complexity in model structures give a better fit to observed data in the calibration process due to increased degrees of freedom. Yet, in the verification phase, a comparison of complex models and models with simple structures and limited numbers of parameters, shows that simple models achieve results as good as those of more complex models [Perrin et al., 2001; Gan et al.1997].

On the other hand, one of the shortcomings of hydrologic simulations is that they typically do not consider the spatial distribution of different land surface features. Instead, they employ spatially averaged, or "lumped" parameters, which represent the generalized characteristics of the basin, although soil properties, slope, and land use/land cover vary spatially within a basin. To increase the spatial resolution of such models, the computational resolution may be changed from catchment scale to a cell-based scale. This change may require a modification of the model structure by addition of new parameters. Such an approach may lead to two problems: increase in the complexity of model structure and over-parameterization. Yet, to more accurately simulate the movement of water across the landscape, the relative spatial locations of surface features must be considered. The development of spatially distributed hydrologic models by having the hydrologic process model operate within a simulation environment, which can accurately represent spatial location of surface features, namely a Geographical Information System (GIS), makes this possible.

Another problem in complex modeling systems relates to data requirements. Particularly in developing countries, the use of advanced hydrologic models is highly limited due to the lack of sufficient hydrometeorological data and land-use information. Not only the quantity, but also the quality and reliability of available data prohibit the use of such models. National research institutions and universities develop and apply new models, which are suitable for use with the current data sets of the country [Fistikoglu, 2002].

In view of the above problems, the study presented herein was developed with the intent to: (a) develop a simple, parsimonious hydrologic simulation model which needs less hydrometeorological data than the complex ones, and which considers spatial distribution of land use, soil data, and input variables by using GIS integration; (b) develop a method to integrate GIS and a hydrologic model, which is a current issue in hydrologic research. The study aims to develop a simple hydrologic model for simulating daily runoff by using the available daily rainfall and evaporation data as inputs. The developed model is called DRRSM (Daily Rainfall Runoff Simulation Model) and uses the spatially distributed watershed data such as land cover and soil types by integrating GIS algorithms. The application of the DRRSM is demonstrated on Demirci watershed, which is a sub-watershed of the Gediz River basin along the Aegean coast of Turkey. The results of the application show that DRRSM can be used to simulate daily mean discharges of a watershed where comprehensive data do not exist.

2. DAILY RAINFALL RUNOFF SIMULATION MODEL (DRRSM)

2.1 Introduction

DRRSM is a distributed, conceptual, continuous watershed model that is developed to estimate daily mean runoff from medium-sized rural watersheds where precipitation, land cover and soil properties are spatially distributed. DRRSM can simulate evapotranspiration, surface detention, surface runoff, sub-surface storage, sub-surface runoff, groundwater storage and groundwater runoff by evaluating cell-based precipitation, land cover and soil properties of the watershed. It uses raster data and parameter layers in order to consider the spatial variability of both data and watershed properties such as the areal distributions of precipitation, land cover, and soil properties. As a medium size watershed model, 1-5 km resolution of land properties is satisfactory in the DRRSM.

DRRSM consists of three components which govern the simulations. One of them is the Hydrologic Response Units (HRU) component, which evaluates watershed properties such as precipitation, land cover and soil properties, considering their spatial variations. HRU component generates hydrologically homogeneous areas to be used in the simulations. The second component is the Vertical Water Budget (VWB) component, which runs water fluxes from surface, sub-surface and groundwater zones within each hydrological response unit (HRU). The last one is the Data and Parameter (D&P) component that organizes which HRU uses which precipitation and evaporation records as inputs, and which parameter sets belong to which HRU.

Runoff generations are based on simple linear discharge-storage relationships, which may be changed to a non-linear structure by the user since the DRRSM is developed by means of Object Oriented Programming techniques. The current equations are developed as linear structures in order to test the performance of the model in using spatially distributed data. The data needs of the developed model include only rainfall, pan evaporation, DEM (digital elevation map), land cover distribution and soil properties which can be obtained easily from various sources through the Internet. The main input of the model DRRSM is daily rainfall in which the duration and intensities of rainfall are not considered in the simulations. The model runs with only one-day time step; thus, it uses the summed inputs (daily rainfall) and generates averaged outputs (daily mean runoff).

In the DRRSM, all runoff components such as surface, sub-surface and groundwater runoff are linear functions of relevant storages. Thus, well-known linear runoff-storage relationships are the basic equations of the runoff generation process. For generating surface and sub-surface runoff, there is one assumption that runoff occurs only when the maximum storage capacities of surface and sub-surface storages are exceeded. All the excess water reaches the lower storage system or the stream network, and then the outlet of the watershed in one time step. Thus, in large watersheds, the watershed must be divided into medium size watersheds. The groundwater system continuously releases water depending on underground storage, which is also considered as a linear system. Furthermore, runoff routing in the channel is not considered in the DRRSM; the model computes only the surface, sub-surface and groundwater runoff components of the watersheds at the watershed outlet.

2.2 DRRSM Components

DRRSM is a cell-based or a raster-based simulation model where computations are carried out on a cell-based pattern. The number of cells affects the computation time and the use of computer memory. If the watershed area is large or cell dimensions are too small, the number of cells increases so that the computational time increases and the free memory of computers decreases. To cope with this problem, one computation for a group of homogeneous cells is preferred instead of doing the same computations for similar cells. The main problem then is to define homogeneous cells or areas which receive the same precipitation, and have the same soil properties and the same land cover attributes. In the case presented here, water balance computations were carried out for 13 homogeneous groups instead of repeating them for each 1kmx1km cell. Thus, the speed of simulations was increased. A GIS analysis technique is applied to raster data layers to define homogeneous areas.

DRRSM considers 3 spatially distributed watershed properties such as rainfall distribution, soil type distribution, and land cover distribution as digital raster data layers. The HRU component reads these 3 raster data layers and creates homogeneous hydrological land segments (HRU). Figure 1 shows DRRSM's vertical water budget component, which is the main component of the model. This component computes storages and runoff of each HRU, accounting for the spatially and temporally distributed parameters and inputs. After the number of HRUs and their precipitation, soil and land cover properties are determined, vertical water budget component gets this information and reads the related inputs and parameters, such as precipitation and evaporation data. Then, DRRSM runs the hydrologic simulation described in Fig.1, using the input data. Here, precipitation is the major input of the vertical water budget component.

In the DRRSM, daily potential evapotranspiration is calculated by using only daily pan evaporation (EPAN) and a coefficient (EC) which depends on land cover characteristics. Equation (1) defines the estimation of daily potential evapotranspiration with respect to daily pan evaporation and land cover types:

$$PET_{i,t} = EC_{i,mon}\ EPAN_{i,t} \tag{1}$$

where $PET_{i,t}$ is the daily potential evapotranspiration (mm/day) of the i^{th} HRU at the t^{th} day; $EC_{i,mon}$ is the evapotranspiration coefficient of the i^{th} HRU, that depends on land cover properties and varies between 0 and 1. This coefficient changes on a monthly scale; thus, it stays constant during a specific monthly interval. $EPAN_{i,t}$ is the pan evaporation (mm/day) of the i^{th} HRU at the t^{th} day. $EC_{i,mon}$ is a parameter to be calibrated by the user if potential evapotranspiration is calculated by Equation (1). In the model, evapotranspiration is extracted first from the surface zone (ET1), which stores water on the vegetation and the soil surface, and then from the sub-surface zone (ET2), which has minimum and maximum water content boundaries (SMIN, SMAX). When a HRU receives precipitation, the precipitation is diverted into the detention storage. Detention storage is a surface zone storage volume, which depends on land cover and has a maximum capacity (DETCAP). If the amount of water in the surface zone storage (S_{SUR}) exceeds its capacity (DETCAP), a proportion of that water (α) is diverted as surface flow and the remaining percent of water ($1-\alpha$) is diverted to subsurface zone storage (S_{SUB}) as infiltration. Maximum capacity of surface detention (DETCAP) depends on land cover properties. Surface runoff and infiltration occur only when surface zone storage (S_{SUR}) exceeds that capacity value, which changes monthly in the model. DETCAP is defined by the user and is another model parameter to be calibrated. Surface runoff occurs only when surface zone storage (S_{SUR}) exceeds the surface detention capacity (DETCAP) and is assumed to be a linear function of the amount of excess water as defined in Fig. 1. Infiltration depends on the surface zone water storage as in the case of surface runoff. If the surface zone storage exceeds the detention capacity, infiltration occurs. Infiltration rate is again assumed to be a linear function of the water exceeding detention capacity. Since α represents the surface runoff portion of the excess water, ($1-\alpha$) must be used as an infiltration coefficient in respect of continuity. Sub-surface runoff depends on the sub-surface zone water storage (S_{SUB}) as sketched in Fig. 1. As in the case of surface flow, sub-surface flow is assumed to be a linear function of the sub-surface zone water storage (S_{SUB}).

Sub-surface zone has a minimum and maximum storage capacity (SMIN, SMAX). Storage in the sub-surface zone changes during the simulations between these boundary values. Minimum and maximum values of the storage depend on soil properties of the HRU. If the amount of water stored in the surface zone is not enough for the daily evapotranspiration, the deficit is covered by the sub-surface storage if there is available water there (ET2). Then, the water content of the sub-surface zone is reduced by evapotranspiration. If the water deficit due to evapotranspiration is more than the water content of sub-surface zone, the water content of sub-surface zone is reduced to SMIN, and evapotranspiration has to be less than potential. After infiltration occurs, water content of the sub-surface zone (S_{SUB}) is increased by adding infiltration. If the sub-surface water content exceeds the maximum capacity of the sub-surface zone storage (SMAX), sub-surface runoff and percolation occur, depending on the coefficient β. Groundwater runoff depends on groundwater zone water storage (S_{GRO}). It is simulated as a linear function of the groundwater storage. Water infiltrating through the surface and percolating from the sub-surface zone storage to the groundwater storage may flow to active groundwater storage or may be lost by deep percolation. Active groundwater eventually reappears as baseflow, but deep percolation is considered lost from the simulated system.

In DRRSM, runoff routing is considered as the process of finding the total runoff at the watershed outlet. Since the simulation time step is longer than hours, and since the time distribution of the precipitation in a day is not considered, runoff routing in the model has to be lumped. As the runoff parameters α, β, γ yield the amount of water leaving the watershed as portions of the previous storages, the sum of the runoff amounts already represents the routing itself. Consequently, the total runoff leaving the watershed is defined by adding all runoff volumes that come from surface, sub-surface, and groundwater storages of each HRU. However, sub-surface and groundwater runoff of each HRU may be considered to be slower than surface runoff. To account for the time lag between surface, sub-surface, and groundwater runoff, DRRSM can use time lag parameters, which must be expressed in integers such as 1 day or 2 days, etc.

Data and parameter components of the DRRSM organize input data and model parameters in order to obtain proper usage by the HRUs. Since DRRSM is a daily-based rainfall-runoff simulation model, model inputs, i.e., daily total rainfall and evaporation, must be introduced as daily time series. Parameters of DRRSM which depend on land cover properties change monthly and those that depend on soil properties are constant during the simulations. Data and parameter component recognizes the days of the simulation and calls input data which are daily rainfall (P) and evaporation (EPAN) for each HRU from the input data file, considering the type of the HRU. Then, the evapotranspiration parameters (EC) and surface runoff parameters (α) are called from the parameter file, considering the month of the simulated day. Finally, data and parameter component reads the soil properties of each HRU and calls the sub-surface runoff parameter (β), maximum and minimum sub-surface zone storage capacities (SMIN, SMAX), and groundwater runoff parameters (γ and DEEP). After all data and parameters are defined the data and parameter component forwards them to the vertical water budget component.

2.3 Calibration Process

In DRRSM, three manual calibration techniques, which are commonly used in similar models, are preferred due to their ease of use. The first approach is to find the minimum value of the sum of the squared differences between observed and simulated runoffs as described in Equation 2:

$$D = \sum_{t=1}^{N}\left(Q_{t,\,observed} - Q_{t,\,model}\right)^2 \tag{2}$$

where D is the sum of the square of the differences between observed and simulated runoff, $Q_{t,observed}$ is the observed runoff at the t^{th} day, $Q_{t,model}$ is the simulated runoff at the t^{th} day, and N is the number of days used in the calibration process. D is not the basic calibration parameter here; it only shows the direction of calibration to follow whether the parameter estimates approach better or worse values. The second approach is to find the maximum value of the correlation coefficient, R, between the observed and the simulated runoff. The third approach is proposed by WMO (World Meteorological Organization) (WMO, 1986). In this approach, the best model parameters give the closest value to “1” for the F parameter defined in Equation 3:

$$F = 1 - \frac{\sum_{t=1}^{N}\left(Q_{m,\,t} - Q_{o,\,t}\right)^2}{\sum_{t=1}^{N}\left(Q_{m,\,t} - \overline{Q_m}\right)^2} \tag{3}$$

where, F is the calibration parameter to be close to 1; $Q_{m,t}$, the simulated runoff on the t^{th} day; $Q_{o,t}$, the observed runoff on the t^{th} day; $\overline{Qm}$, the mean value of the simulated runoff; and N, the number of days used in calibration. F is a parameter that was derived similarly to the Nash-Sutcliffe co-efficient of determination. There are 8 parameters to be calibrated in DRRSM: EC, DETCAP, α, β, γ, DEEP, SMIN and SMAX. These parameters are described in detail in Section 2.2.

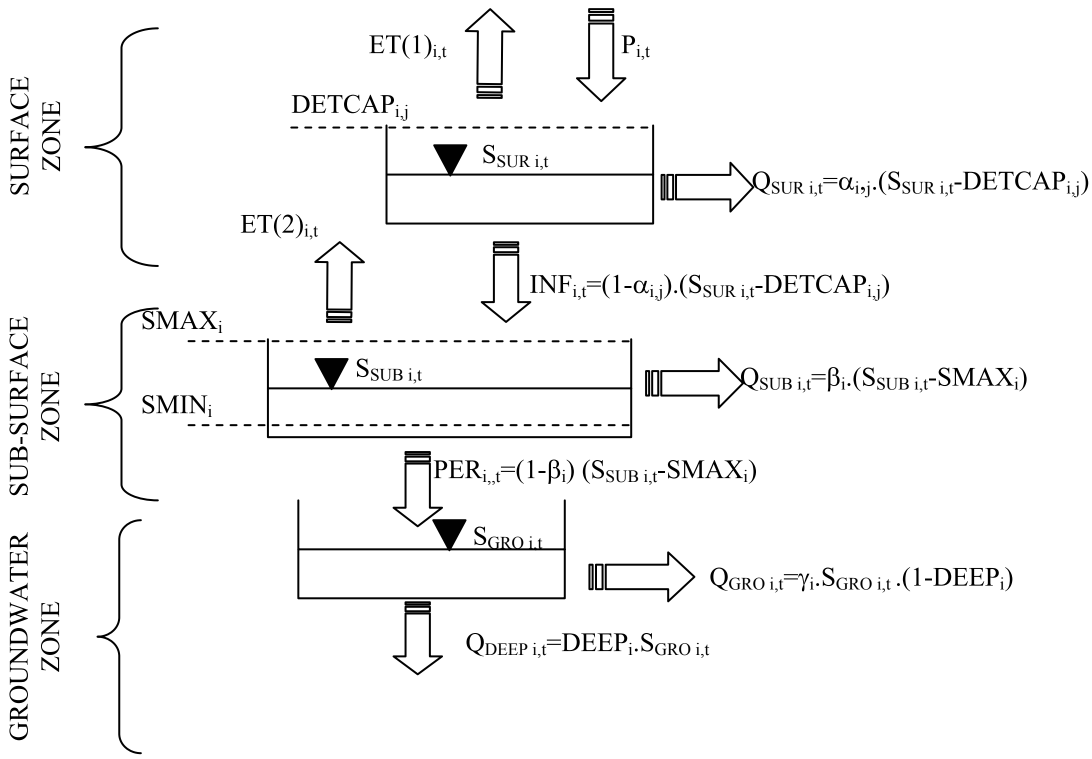

Figure 1. Vertical Water Budget Component of the DRRSM (ET: Evapotranspiration, P: Rainfall, DETCAP: Surface Detention Capacity, INF: Infiltration, PER: Percolation)

3. APPLICATION OF THE MODEL

DRRSM is applied to the Demirci watershed, which is a 818 km^2 subbasin of the Gediz River Basin along the Aegean coast of Turkey. Topography of the watershed along with land cover distribution are obtained from the USGS EROS Data Center at a resolution of 1 km x 1 km. The distribution of soil texture properties is obtained by digitizing traditional soil maps. Hydrometeorological data such as rainfall, temperature, evaporation and runoff at the outlet of the watershed) were available as time records at the relevant stations.

Manual calibration of the model is realized for the year 1995 and verification for 1996. At the end of 1995, there are big differences between the observed and the simulated runoff. These differences, which continue for 10 days, affect the calibration statistics. These differences are due to sampling errors in the observed runoff values at the end of year 1995, possibly caused by a fault in the water level meter. When the last 10 days of the year 1995 are eliminated from the calibration process, considerable changes occur in the calibration statistics (D, R, F) (Table 1) and in the mean value and standard deviations of both the observed and the simulated runoff (Table 2).

To test model performance, verification of the DRRSM for Demirci watershed is realized by using the calibrated parameters and the rainfall record of the year 1996. The simulation is carried out with the full record of the year 1996, and the results of the simulation are given in Fig. 2.

Table 1. Calibration statistics.

Calibration Year	*D*	*R*	*F*
1995 (N=365)	89	0.72	0.06
1995 (N=355)	19	0.88	0.80

Table 2. Means and standard deviations of the observed and simulated runoff values.

Year		*Number of days used for calibration*	*Means*	*Standard Deviations*
Full record of 1995	Observed Runoff	365	0.35	0.73
	Simulated Runoff	365	0.30	0.51
Last 10 days eliminated record of 1995	Observed Runoff	355	0.29	0.55
	Simulated Runoff	355	0.29	0.51

Tables 3 and 4 give the calibration statistics for the verification period 1996 and the mean and standard deviations of the observed and the simulated runoff values of the Demirci watershed, respectively. Figure 3 also shows all the runoff components which are surface, sub-surface and groundwater runoff during 1996.

Table 3. Calibration statistics (D, R, F) calculated for verification period 1996 of Demirci watershed

Verification Year	*D*	*R*	*F*
1996 (N=366)	39	0.87	0.66

4. CONCLUSION

DRRSM is a distributed hydrologic model with less data requirements and a more flexible structure compared to other complex models in the market. At its current state, the model uses simple linear equations for estimating runoff values from the storages on the land segments.

Table 4. Means and standard deviations of the observed and simulated runoff values for the verification period 1996.

Year		*Number of days used for verification*	*Means*	*Standard Deviations*
Full record of 1996	Observed Runoff	366	0.31	0.69
	Simulated Runoff	366	0.31	0.56

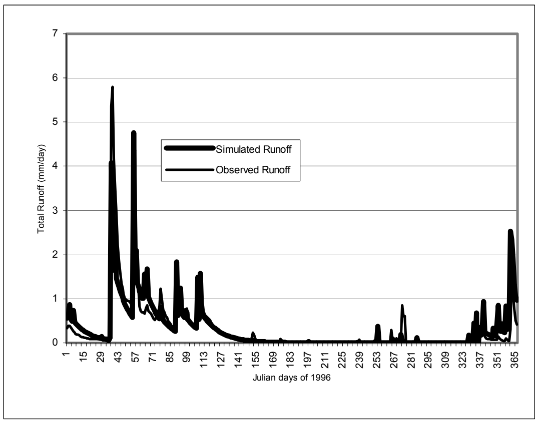

Figure 2. Simulated and observed runoff during the verification period 1996

Yet, the user can change the forms of equations to non-linear as the program is designed by means of independent components and object oriented programming techniques. However, changing the equations causes an increase in the number of parameters and makes the model structure divert from a parsimonious model type. The application of the DRRSM shows that the limited number of parameters and simple forms of equations are quite satisfactory since the spatial distribution of the parameters is considered. In addition, the results of the DRRSM in the Demirci watershed show that the GIS data obtained from the Internet sources such as topography and land cover with 1km x 1km resolution can be used for hydrologic analysis. Basically, the model has proved to be successful in cases where available data on catchment characteristics are insufficient to meet the calibration needs of over-parameterized complex models.

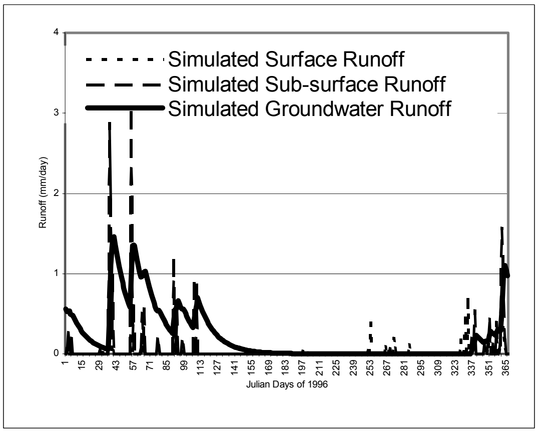

Figure 3. Simulated surface, sub-surface and groundwater runoff of Demirci watershed during 1996.

5. REFERENCES

Fistikoglu, O., Development of integrated tools for river basin management. Ph.D. Thesis DEU, Graduate School of Natural and Applied Sciences, Izmir, Turkey, 2002

Gan, T.Y., E.M. Dlamini and G.F.Biftu, Effects of model complexity and structure, data quality and objective function on hydrologic modeling. *Journal of Hydrology*. 32, 3513-3524, 1997

Perrin, C., C. Michel and V. Andreassian, Does a large number of parameters enhance model performance?. Comparative assessment of common catchment model structures on 429 catchments. *Journal of Hydrology*. 242, 275-301, 2001

WMO, Intercomparison of models of snowmelt runoff. Operational Hydrology Rep. 23, WMO-646, 1986.