



available at www.sciencedirect.com



journal homepage: www.elsevier.com/locate/jhydrol



Predicting hourly-based flow discharge hydrographs from level data using genetic algorithms

Gokmen Tayfur ^{a,*}, Tommaso Moramarco ^b

^a Department of Civil Engineering, Izmir Institute of Technology, Urla, Izmir 35340, Turkey

^b Research Institute for Geo-Hydrological Protection, National Research Council, Via Madonna Alta, 126, 06128 Perugia, Italy

Received 11 May 2007; received in revised form 7 November 2007; accepted 18 December 2007

KEYWORDS

Genetic algorithm;
Flow hydrograph prediction;
Ungauged basins;
Elevation data;
Stage data;
Rating curve;
Modeling

Summary This study developed a genetic algorithm model to predict flow rates at sites receiving significant lateral inflow. It predicts flow rate at a downstream station from flow stage measured at upstream and downstream stations. For this purpose, it constructed two different models: First is analogous to the rating curve model (RCM) of Moramarco et al. [Moramarco, M., Barbetta, S., Melone, F., Singh, V.P., 2005. Relating local stage and remote discharge with significant lateral inflow. *J. Hydrologic Eng., ASCE*, 10(1)] and the second is based on summation of contributions from upstream station and lateral inflows using kinematic wave approximation. The model was applied to predict flow rates at three different gauging stations located on Tiber River, Upper Tiber River Basin, Italy. The model used average wave travel time for each river reach and obtained average set of parameter values for all the events observed in the same river reach. The GA model was calibrated, for each river reach and for each formulation, by three events and tested against three other events. The results showed that the GA model produced satisfactory results and it was superior over the most recently developed rating curve method. This study further analyzed the case where only water surface elevation data were used in the input vector to predict flow rates. The results showed that using elevation data produces satisfactory results. This has an implication for predicting flow rates at ungauged river sites since the surface elevation data can be obtained without needing the detailed geometry of river section which could change significantly during a flood.

© 2008 Elsevier B.V. All rights reserved.

Introduction

The correct computation of flow discharge is crucial for water resources engineering projects. It can be computed from available flow stage, channel width, and flow velocity

* Corresponding author. Fax: +90 232 750 6196.
E-mail addresses: gokmentayfur@iyte.edu.tr (G. Tayfur), T.Moramarco@irpi.cnr.it (T. Moramarco).

data. Measuring channel width requires topographic surveying which is expensive and sometimes impossible for inaccessible sections. Flow velocity can be measured using cable way and a current meter which becomes difficult, especially during severe floods. On the other hand, measuring flow stage is comparatively easy and cheap. Hence, researchers have tended to develop models relating flow stage to flow rate. For this purpose hydrologic and hydraulic methods have been developed. Hydrologic methods are based on only conservation of mass principle and they are limited for the cases where linearity is preserved. Hydraulic methods are based on conservation of mass and momentum equations that are often expressed by highly nonlinear partial differential equations whose solutions require highly sophisticated numerical techniques having numerical convergence and instability problems (Caleffi et al., 2003; Ying et al., 2004; Singh et al., 2004). Furthermore, for complex watersheds, hydraulic models require intensive topographic information and parameter estimation (Kouwen et al., 2005; Muleta et al., 2007).

Hence, researchers have employed simpler approaches (Franchini and Lamberti, 1994; Moramarco and Singh, 2001). Although the models of Franchini and Lamberti (1994) and Moramarco and Singh (2001) are simple they fail when applied to river reaches having lateral inflows. Franchini et al. (1999) developed model for estimating rating curve requiring flow stage and discharge data. However, their model is complex requiring estimation of nine parameters. Recently, Moramarco et al. (2005) developed a rating curve model (RCM) which is physically-based and applicable to cases where lateral inflows are significant. The RCM relates flow rate at a downstream site to flow rate at an upstream site and flow cross-sectional areas at the both sites with two parameters that are obtained using information on downstream base flow and downstream peak discharge. As such, it requires measurements of flow cross-sections through topographic surveying and flow depth measurements. It also requires flow velocity measurements at upstream end. Furthermore, the RCM model, for each event that occurs in the same river reach, has to determine the wave travel time and the model parameters. Most recently, Tayfur et al. (2007) applied artificial neural network (ANN) to predict flow rates using flow stage data in the input vector. Although ANN produced satisfactory results it does not reveal insight into understanding the physics of the process. ANN and GA models perform the same task that is the minimization of the error-based objective function. With regard to the methodologies and algorithms that both the models employ, they do not reveal insight into the understanding of the basic processes of the physical event. In that sense, both the models are black box models. However, ANN model is a further black box model since it does not yield an empirical equation. GA model, on the other hand, finds optimal values of existing empirical equations that can be readily used for predictive purposes. These existing empirical equations may shed a light onto the understanding of the physics of the processes. Furthermore ANN lacks extrapolation capability such that they cannot be applied to ungauged basins (Tayfur et al., 2007). As it is presented in a later section, GA model does not have such a shortcoming.

This study proposes to employ genetic algorithm (GA) to predict flow rates from flow stage data only. For this purpose, two formulations are proposed. One is analogous to the formulation of RCM model of Moramarco et al. (2005) and the other one is based on kinematic wave approximation. As pointed out earlier the RCM model relates flow rate at the downstream station to flow rate at upstream station and flow cross sectional-areas at upstream and downstream stations with parameters. This study however proposes to relate flow rate at a downstream station to flow stages at the upstream and downstream stations only. This study furthermore proposes to predict flow rate using only water surface elevation data without needing to know the detailed geometry of the river cross section which could change significantly during a flood. This enhances the rating curve assessment and it has an important implication for predicting flow rates at ungauged river sites.

Genetic algorithm (GA)

Genetic algorithm (GA) is a nonlinear search and optimization method inspired by biological processes of natural selection and the survival of the fittest. They make relatively few assumptions and do not rely on any mathematical properties of the functions. As such, GA has found wide application in water resources engineering (Liong et al., 1995; Guan and Aral, 1998, 2005; Sen and Oztopal, 2001; Jain et al., 2004; Singh and Datta, 2006; Cheng et al., 2005, 2006; Wu and Chau, 2006).

Basic units of GA consist of 'bit', 'gene', 'chromosome' and 'gene pool'. In GA, *gene* consists of *bits* (i.e. 0 and 1) and represents a model parameter (or a decision variable) to be optimized. A *chromosome* is the combined set of all the genes. Note that each variable is the gene and combination of genes forms the chromosome each of which is a possible solution for the variables. Finally, set of chromosomes form the *gene pool*. In the gene pool, each chromosome is a candidate for a solution.

The main GA operations basically consist of 'initial gene pool generation', 'evaluation of fitness for each chromosome', 'selection', 'cross-over', and 'mutation'. Initial population of chromosomes can be randomly generated by, for example, a uniform distribution or a normal distribution. Fitness of each chromosome can be obtained by the employment of the following equation:

$$F(C_i) = \frac{f(C_i)}{\sum f(C_i)} \quad (1)$$

where C_i is the chromosome i ; $F(C_i)$ is the fitness value of chromosome that is the percentage of variable in the pool; $f(C_i)$ is the value of objective function evaluated for chromosome i .

Selection can be performed randomly by, for example, a roulette wheel (Sen, 2004) or by ranking where the chromosomes are ranked according to their fitness from the fittest to weakest. Then the fittest ones are copied on the weakest ones.

By *cross-over*, new individuals are produced by changing the genes of the chromosomes. The chromosomes from the current generation are selected for the recombination process based on their relative fitness. Fig. 1 is an example for a single cut cross-over operation where the first two

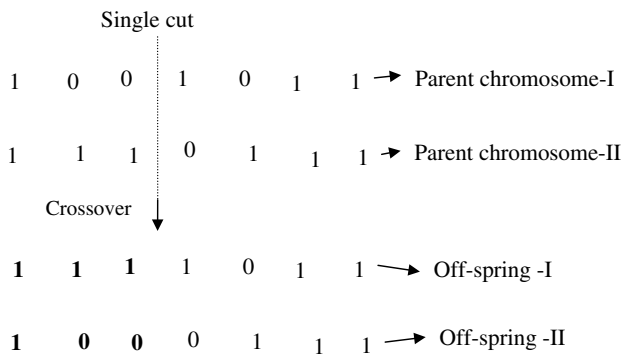


Figure 1 Example for a single cut cross-over operation.

chromosomes are cut from the 3rd digit from left, yielding new chromosomes at the bottom.

The last operation in GA is the *mutation* where particular bit (bits) is reversed (i.e. 1 to 0 or 0 to 1). In GA search, this is the perturbation that allows the GA to seek out new and novel solutions. By this process, the next trend would either faster converge to or diverge from the solution. In general, about 5% of the bits are subjected to mutation. Fig. 2 is an example presenting that the value of 153 goes to 57 after cross-over and then to 249 after mutation, scanning a large area of the solution domain. The details of GA can be obtained from Goldberg (1999) and Sen (2004), among others.

Rating curve method (RCM)

The RCM model of Moramarco et al. (2005) is:

$$Q_d(t) = \alpha \frac{A_d(t)}{A_u(t - T_L)} Q_u(t - T_L) + \beta \quad (2)$$

where Q_u is the upstream discharge; Q_d is the downstream discharge; A_d and A_u are the effective downstream and upstream cross sectional flow areas, respectively; T_L is the wave travel time; and α and β are the model parameters estimated by the following equations (Moramarco and Singh, 2001):

$$Q_d(t_b) = \alpha \frac{A_d(t_b)}{A_u(t_b - T_L)} Q_u(t_b - T_L) + \beta \quad (3a)$$

$$Q_d(t_p) = \alpha \frac{A_d(t_p)}{A_u(t_p - T_L)} Q_u(t_p - T_L) + \beta \quad (3b)$$

where $Q_d(t_b)$ is the base flow rate at the downstream section; $Q_d(t_p)$ is the peak discharge at the downstream section; t_p and t_b are the times when the peak stage and baseflow occur at the downstream section, respectively.

Base flow rate $Q_d(t_b)$ can be computed from the velocity measurements during low flows. The peak discharge $Q_d(t_p)$ is computed as

$$Q_d(t_p) = [Q_u(t_p - T_L) - Q^*] + q_p L \quad (4)$$

where $Q_u(t_p - T_L)$ represents the upstream discharge delayed for the wave travel time T_L ; Q^* is the discharge attenuation computed from Price formula (Price, 1973). $q_p L$ is the lateral flow estimated by (Moramarco and Singh, 2002):

$$\frac{A_d(t_p) - A_u(t_p - T_L)}{T_L} = q_p \quad (5)$$

Parameters α and β are obtained from the solution of Eqs. (3a) and (3b). Note that, for each event observed even in the same river reach, one needs to obtain different set of α and β values to apply the RCM model. Also, for each event, one needs to estimate the wave travel time. The estimation of wave travel time is given in Moramarco et al. (2005).

Proposed formulations

This study proposes two formulations for predicting flow rates from stage data. First one is analogous to the RCM model and expressed as

$$Q_d(t) = \alpha \frac{h_d^\beta(t)}{h_u^\gamma(t - T_L)} + \eta \quad (6)$$

where h_d and h_u is the flow stage at downstream and upstream station, respectively; α , β , γ , and η are the model parameters whose optimal values are found by the GA model. We call this model as ‘‘GA-Stage-I’’.

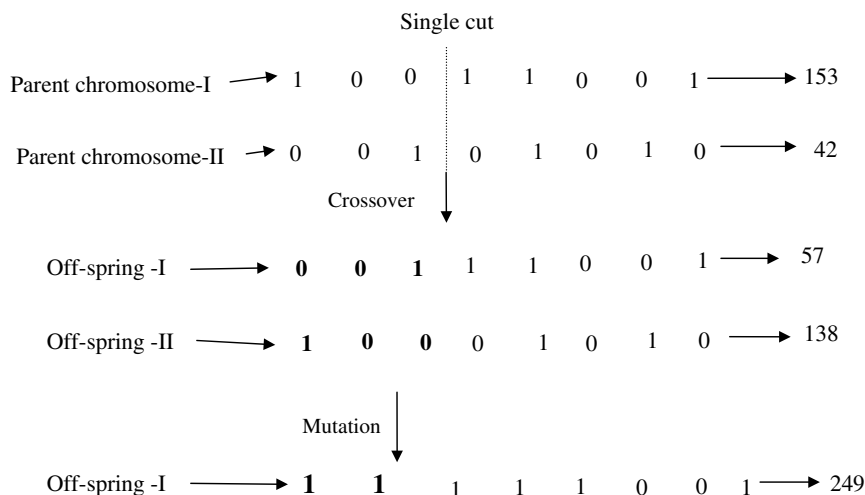


Figure 2 Example for cross-over and mutation operations.

According to Eq. (6), this model predicts flow rate at a downstream station from flow stage measured at upstream station a wave travel time earlier and the flow stage measured at downstream station with parameters whose optimal values are found by the GA model.

Second proposed model is based on kinematic wave approximation. That is flow rates at upstream and downstream stations are related to flow stages using kinematic wave approximation. The downstream flow is assumed as the contribution of upstream discharge at wave travel time earlier, at which the lateral inflow of the intermediate basin is added by surmising it as a power function of h_d . This model can be expressed as

$$Q_d(t) = \alpha_1 h_u^{\beta_1}(t - T_L) + \alpha_2 h_d^{\beta_2}(t) + \eta \quad (7)$$

where parameters α_1 , α_2 , β_1 , β_2 and η are found by the GA model. We call this model as "GA-Stage-II".

This study is also proposing two further formulations that predict flow rate at a downstream station from water surface elevation (elevation of a station from a reference datum plus flow depth, see Fig. 3) at upstream station a wave travel time earlier and water surface elevation at downstream station. This aspect is fundamental for the rating curve assessment because by, this, models would not depend on changes in the river section geometry that often occur during floods. This means that, unlike the RCM approach, the two formulations would not require the topographical surveys of gauged sections. Setting the hydrometric zero of downstream station as the reference datum, analogously to the RCM, the first formulation can be stated as

$$Q_d(t) = \alpha \frac{h_d^\beta(t)}{E^\gamma(t - T_L)} + \eta \quad (8)$$

where E is the upstream water surface elevation above the hydrometric zero of downstream gauged section (see Fig. 3). α , β , γ , and η are the model parameters whose optimal values are found by the GA model. We call this model as "GA-Elevation-I". It is a reasonably common practice in hydrological analysis to assume that the reference datum is the hydrometric zero at downstream station of a river reach, because this results in always obtaining positive values for water surface slope.

Also, this study is proposing the following formulation which is analogous to Eq. (7):

$$Q_d(t) = \alpha_1 E^{\beta_1}(t - T_L) + \alpha_2 h_d^{\beta_2}(t) + \eta \quad (9)$$

where optimal values of the parameters α_1 , α_2 , β_1 , β_2 and η are found by the GA model. We call this model as "GA-Elevation-II".

It has to be pointed out that the proposed formulations attempt to do reliable predictions using easily available

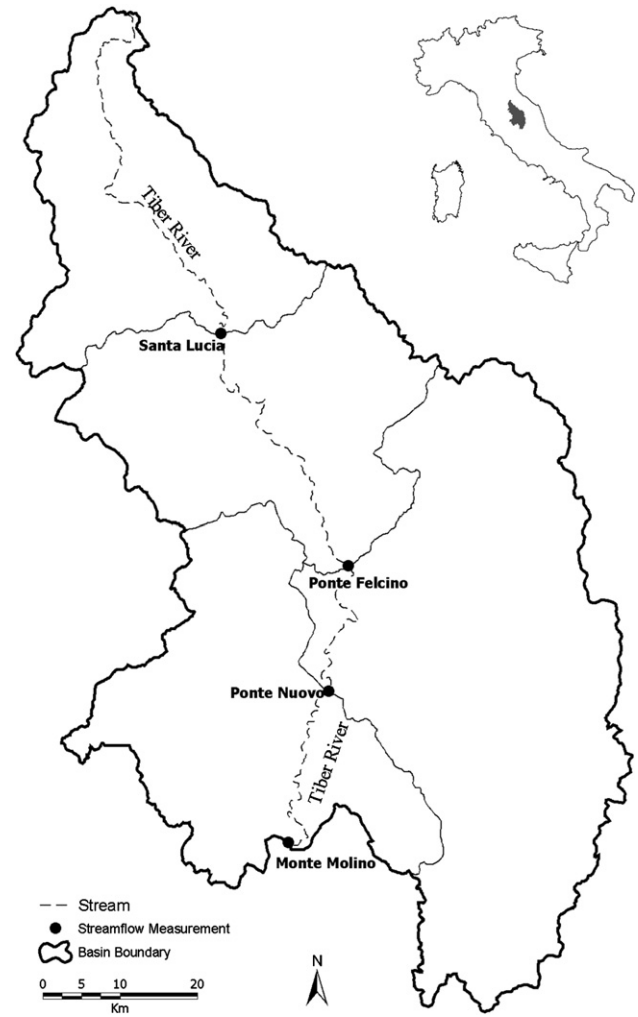


Figure 4 Upper Tiber river basin with the gauged sites and the respective drainage basins.

Table 1 Main geomorphological characteristics of Tiber River reaches

River	Bounded sections	Drainage area (km ²)	Reach length (km)	Mean slope	Mean width (m)
Tiber	Santa Lucia	935	44.6	0.0016	35
	Ponte Felcino	2035			
Tiber	Santa Lucia	935	70	0.0014	39
	Ponte Nuovo	4145			
Tiber	Santa Lucia	935	100.8	0.0012	44
	Monte Molino	5279			

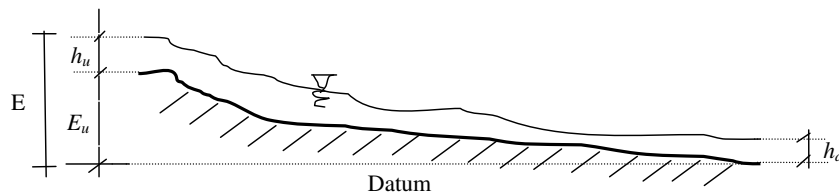


Figure 3 Schematic representation of water surface elevation from a reference datum set at downstream hydrometric zero.

data of flow stage and surface elevation. In Eqs. (6)–(9), the exponents are retained in order to account for the nonlinearity inherited in the flow processes. In hydraulic engineering studies, it is common to relate discharge to flow cross-section as $Q = \alpha A^\beta$; and flow velocity to flow depth $V = \alpha h^\beta$ (Singh, 1996). Moramarco et al. (2006) also employed $Q = \alpha h^\beta$. The η term in Eqs. (6)–(9), was also employed by Moramarco and Singh (2001) and Moramarco et al. (2005). It might stand for noise as it is commonly considered in stochastic models in hydrology.

GA model application

Watershed and hydrologic data

The Upper Tiber River basin in Central Italy, which has equipped river reaches, was selected for testing the GA model. Fig. 4 shows the location of the selected hydrometric sections along with subtended drainage areas. Table 1 summarizes the main characteristics of the selected river reaches. Each gauged section is equipped with a remote

ultrasonic water level gauge, and velocity measurements are carried out by current meter. The elevations above sea level of the hydrometric zero at stations are as follows: 261.10 m (Santa Lucia); 188.77 m (Ponte Felcino); 162.04 m (Ponte Nuovo); and 135.73 m (Monte Molino).

Six storm events for each river reach were considered for GA model calibration and application. The main properties of the selected flood events are summarized in Tables 2–4. It is seen that the lateral inflow contribution was significant in most of the events. Table 2 presents the storm events observed at Santa Lucia and Ponte Felcino river reach; Table 3 presents the storm events observed at Santa Lucia and Ponte Nuovo reach; and Table 4 shows events occurred at Santa Lucia and Monte Molino reach. Also, shown in these tables are the wave travel time and α and β parameter values used by the RCM model for each event. The events that are used for model calibration by the GA are marked by * in these tables. For each reach, 3 events were used for GA model calibration and 3 were used for model testing. For each river reach, the events were randomly grouped for calibration and application so as to avoid the bias in model performance. The wave travel time, on the

Table 2 Main characteristics of flood events observed at Santa Lucia and Ponte Felcino stations

Date	Santa Lucia station			Ponte Felcino station			RCM		
	Q_b (m ³ /s)	Q_p (m ³ /s)	V (10 ⁶ m ³)	Q_b (m ³ /s)	Q_p (m ³ /s)	V (10 ⁶ m ³)	T_L (h)	α	β
December 1990	9	418.6	47.8	5	404.2	56.8	2	1.15	-9
January 1994	35.6	107.9	8.3	50.8	240.7	17.8	2.5	1.7	-35
May 1995 ^a	4.2	71	10.3	8.8	138.7	19.1	4	1.24	-2.94
January 1997 ^a	18.2	120.2	24.3	36.2	225	51.8	3.5	1.28	-3.19
		146.2							
June 1997 ^a	5	345.6	27.6	10.8	449.6	49.1	5	1	0.25
January 2003	23.6	39.7	12.5	49.3	113.4	38.2	3.5	1.68	-24
		57.5			223.8				
		47.2			153.2				
Mean values							3.4	1.36	-12.41

For symbols see text. Note that there are double peak rates in January 1997 and triple peak rates in January 2003.

^a Used for GA model calibration.

Table 3 Main characteristics of flood events observed at Santa Lucia and Ponte Nuovo stations

Date	Santa Lucia			Ponte Nuovo			RCM		
	Q_b (m ³ /s)	Q_p (m ³ /s)	V (10 ⁶ m ³)	Q_b (m ³ /s)	Q_p (m ³ /s)	V (10 ⁶ m ³)	T_L (h)	α	β
December 1996	14	282.7	18.8	64.6	728.4	58	8.5	0.98	-12.8
April 1997 ^a	3.7	366	32.8	19.2	498.5	75.2	8.5	0.99	-11
December 1998 ^a	24	48.3	7.2	77.7	714.8	52.3	10	1.11	-30
February 1999	19	229	15.2	42.7	761.8	68.4	8	1.12	-28
December 2000 ^a	5	211	52.3	24.8	449.5	205.9	7.5	1.09	-14
		349.2			877.2				
January 2001	21	63.5	13.8	106	404.5	67.4	7.5	1.66	-90
		64.1			312.1				
Mean values							8.5	1.20	-30.8

For symbols see text. Note that there are double peak rates in December 2000 and January 2001.

^a Used for GA model calibration.

Table 4 Main characteristics of flood events observed at Santa Lucia and Monte Molino stations

Date	Santa Lucia			Monte Molino			RCM		
	Q_b (m ³ /s)	Q^p (m ³ /s)	V (10 ⁶ m ³)	Q_b (m ³ /s)	Q_p (m ³ /s)	V (10 ⁶ m ³)	T_L (h)	α	β
April 1997	3.7	366	32.8	27	572	93.7	14	0.87	2.3
December 1998 ^a	24	48.3	7.2	85	770	68.5	11	1.45	-55.8
February 1999 ^a	19	229	15.2	36	754	80.7	12.5	1.02	-7.6
December 2000	5	211	52.3	26	506	224.5	11.5	1.1	-4.9
		349.2			850				
January 2001	21	63.5	13.8	119	397.3	84.8	10.5	1.4	-63
		64.1			364				
May 2004 ^a	10	120.2	38.1	82	544.4	160.2	10.5	1.24	-25.7
		250.5			536.5				
Mean values							11.6	1.14	-14.61

For symbols see text. Note that there are double peak rates in December 2000, January 2001 and May 2004.

^a used for GA model calibration.

Table 5 Optimal parameter values for each proposed model for the Santa Lucia and Ponte Felcino reach

Parameters	GA-Stage-I Model	GA-Stage-II Model	GA-Elevation-I Model	GA-Elevation-II Model
α	2.508		9.089	
β	4.325		2.214	
γ	1.137		-0.114	
α_1		9.985		2.024
β_1		1.397		0.498
α_2		9.992		6.705
β_2		2.369		2.705
η	21.42	-1.236	-1.582	-14.56

Table 6 Optimal parameter values for each proposed model for the Santa Lucia and Ponte Nuovo reach

Parameters	GA-Stage-I Model	GA-Stage-II Model	GA-Elevation-I Model	GA-Elevation-II Model
α	6.520		0.769	
β	2.812		1.416	
γ	0.418		-0.919	
α_1		10.00		9.109
β_1		0.740		0.491
α_2		8.230		9.491
β_2		2.550		2.325
η	50.00	10.00	-49.96	-16.72

Table 7 Optimal parameter values for each proposed model for the Santa Lucia and Monte Molino reach

Parameters	GA-Stage-I Model	GA-Stage-II Model	GA-Elevation-I Model	GA-Elevation-II Model
α	10.00		0.930	
β	2.475		1.060	
γ	0.152		-0.960	
α_1		9.998		9.928
β_1		0.949		0.336
α_2		8.952		6.029
β_2		2.490		2.632
η	49.98	9.990	-49.95	49.91

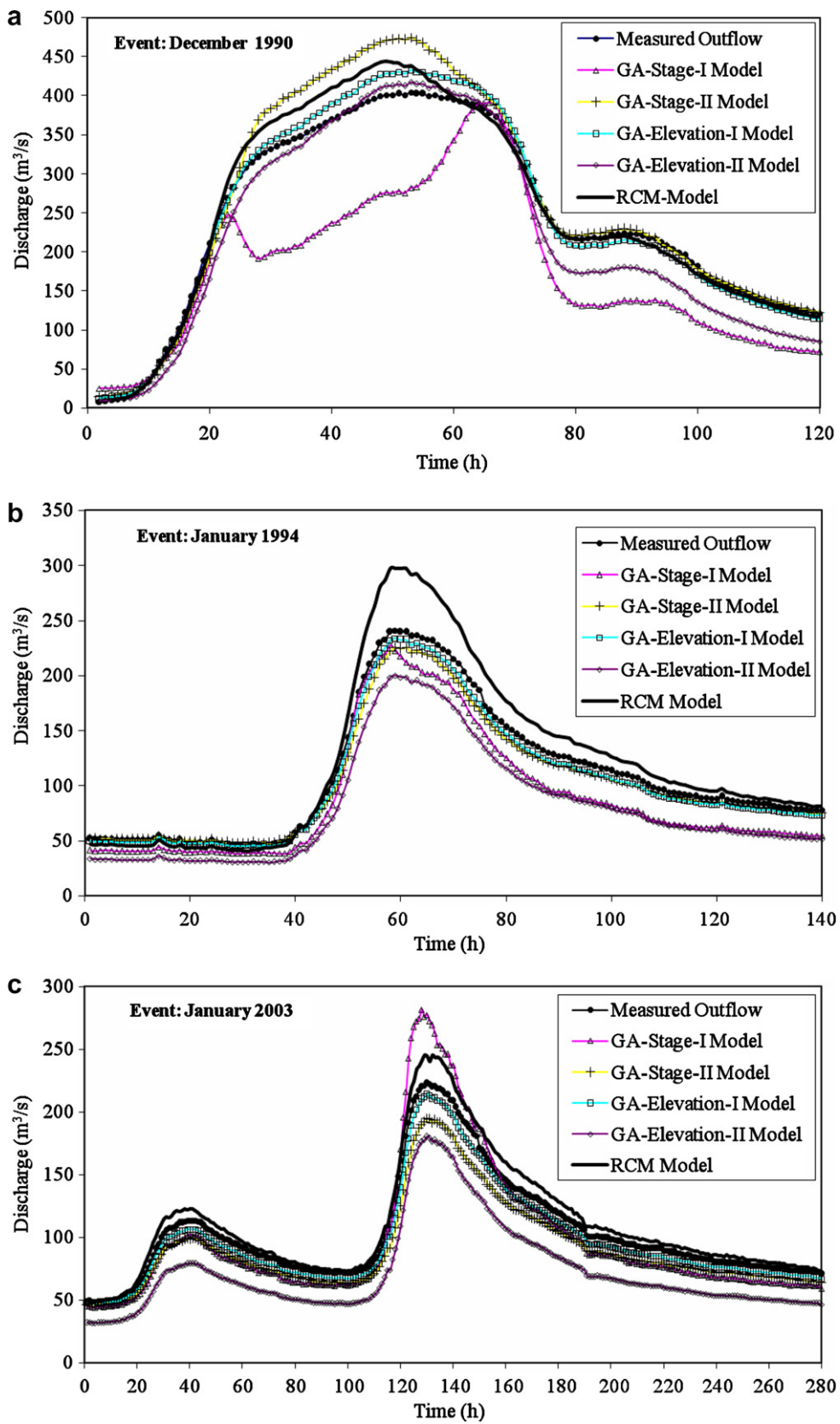


Figure 5 Hydrograph predictions by several models at downstream end of Santa Lucia–Ponte Felcino River Reach in (a) December 1990; (b) January 1994; and (c) January 2003.

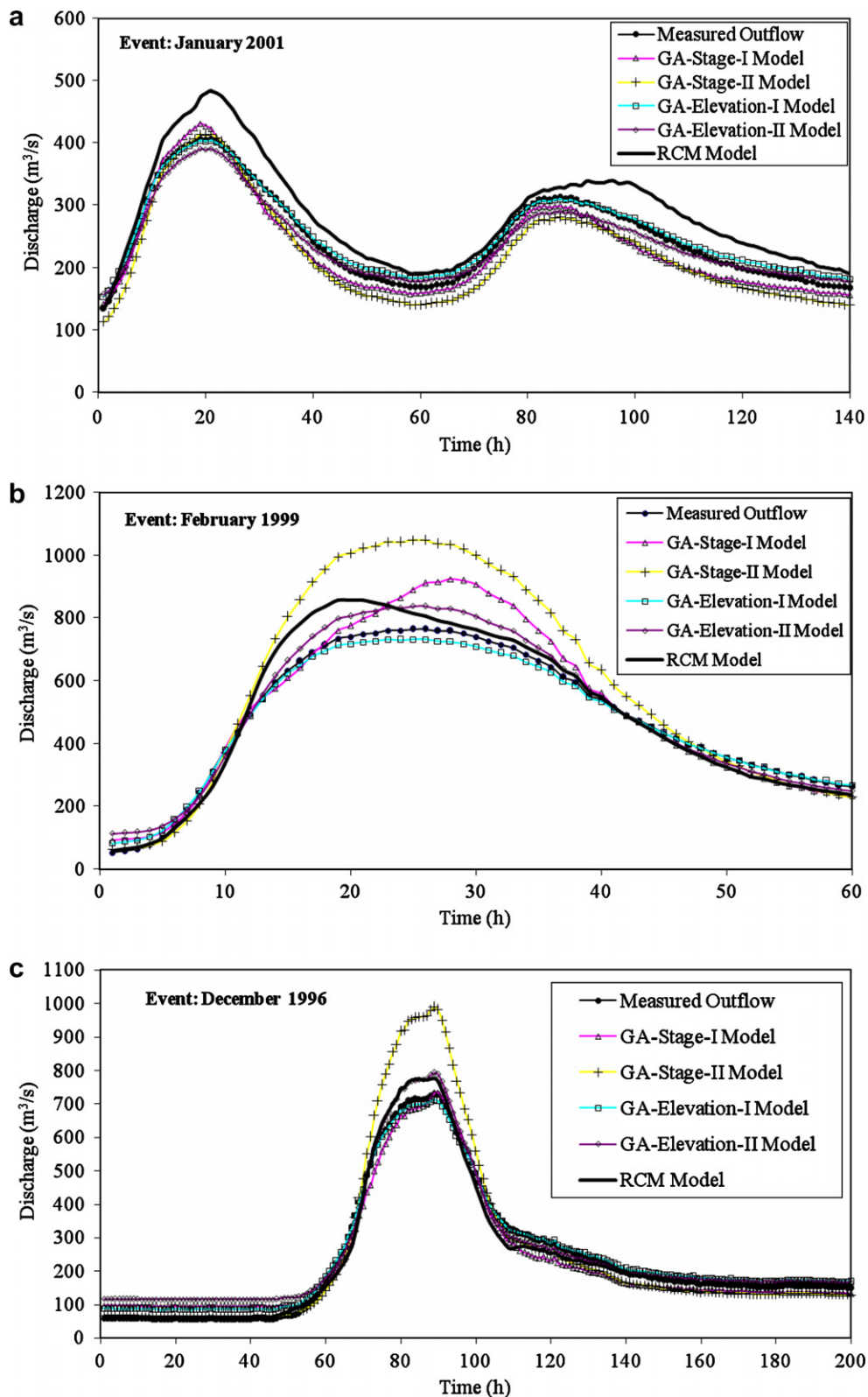


Figure 6 Hydrograph predictions by several models at downstream end of Santa Lucia–Ponte Nuovo River Reach in (a) January 2001; (b) February 1999; and (c) December 1996.

observed that all the runs, in fact, produced comparable results. Nevertheless, since each chromosome is a candidate for a solution, the GA model employed 500 chromosomes

in the gene pool. It also employed 80% cross-over rate, 4% mutation rate and 30,000 iterations. Note that it just takes couple of minutes of CPU time for the model to find a

solution in 30,000 iterations with 500 chromosomes in the gene pool.

Three events [marked by *] from Table 2 were used for calibrating the model parameters for the Santa Lucia–Ponte Felcino reach by the GA model. The calibrated values for each proposed model were summarized in Table 5. Note that, as pointed out earlier, unlike the GA model, the RCM finds different sets of values for the parameters and uses different wave travel time for each event observed in the same river reach. GA, on the other hand, for each river reach employs an average wave travel time and finds one set of average optimal model parameter values. Thus, it requires less information.

In a similar fashion, 3 events [marked by *] from Table 3 and 3 events [marked by *] from Table 4 were employed to calibrate the parameters for the Santa Lucia–Ponte Nuovo reach and Santa Lucia–Monte Molino reach, respectively. The calibrated parameter values for each proposed model were summarized in Table 6 for the Santa Lucia and Ponte Nuovo reach. For the Santa Lucia–Monte Molino reach, the optimal parameter values for each proposed model were summarized in Table 7.

Hydrograph predictions

Santa Lucia–Ponte Felcino reach

Fig. 5 presents simulations of 3 different hydrographs measured at Ponte Felcino station by the RCM model and the proposed equations [Eqs. (6)–(9)] whose parameters were obtained by the GA algorithm. As seen in Fig. 5a, GA-Stage-I model could not capture early stage of the measured hydrograph of December 1990. For this event, GA-Stage-II model showed an overestimation of the peak (Fig. 5a). Fig. 5b shows that although all the models captured the trend of the event January 1994, RCM shows significant overestimation of the peak while GA-Elevation-II model underpredicts it (Fig. 5b). The peak of January 2003 hydrograph was overestimated by the GA-Stage-I model while it was underestimated by the GA-Elevation-II model (Fig. 5c). It can be seen in Fig. 5 that, in all the 3 events, GA-Elevation-I model showed satisfactory performance in capturing the trend, time to peak and as well as the peak rates. Table 8 summarizes the computed relative errors with respect to peak discharge and time to peak for each event for all the models. Note that, in the case of peak rate, a negative error value indicates underestimation whereas a positive value indicates overestimation. In the case of time to peak, negative error value indicates early rise in reaching the peak rate while positive value indicates delay. According to Table 8, all the models showed comparable performances in terms of capturing the peak rate with, on the average, 12% error. However, GA-Elevation-I model had less than 5%, outperforming the other models. With respect to time to peak, except RCM and GA-Stage-I models, others reached the peak on time (Table 8).

Santa Lucia–Ponte Nuovo reach

Fig. 6 presents simulations of 3 different hydrographs measured at Ponte Nuovo station (Fig. 4) by the RCM model and

Table 9 Percentage errors in peak discharge [E_{Qp}] and time to peak [E_{Tp}] for the events in Fig. 6 [Santa Lucia–Ponte Nuovo reach]

Event	E_{Tp} (%)									
	GA-Stage-I	GA-Stage-II	GA-Elevation-I	GA-Elevation-II	RCM	GA-Stage-I	GA-Stage-II	GA-Elevation-I	GA-Elevation-II	RCM
January 2001	6.0	1.9	-0.56	3.9	18.9	-9.5	0	0	0	0
February 1999	21.0	37.2	-4.1	9.7	12.1	12.0	0	0	0	-20.0
December 1996	0.42	35.4	-2.6	8.6	5.0	1.1	0	0	0	1.1
Average	9.14	24.93	2.42	7.4	12.0	7.5	0	0	0	7.0

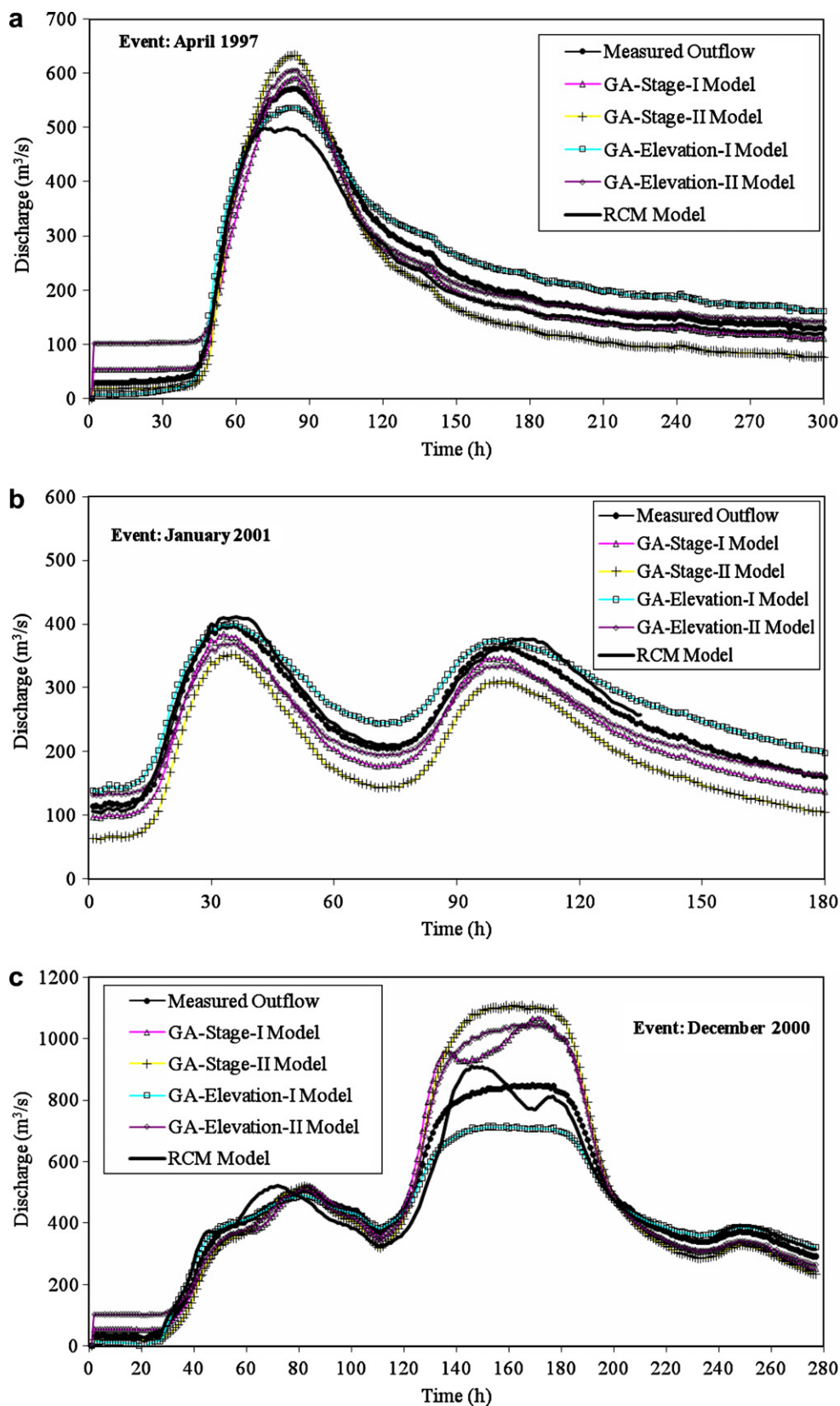


Figure 7 Hydrograph predictions by several models at downstream end of Santa Lucia–Monte Molino River Reach in (a) April 1997; (b) January 2001; and (c) December 2000.

the proposed equations [Eqs. (6)–(9)] whose parameters were obtained by the GA algorithm. As seen in Fig. 6a, RCM overestimated both peaks of the measured outflow. GA-Elevation-I model performed quite satisfactorily in simulating January 2001 (Fig. 6a) and as well as February 1999 (Fig. 6b) events. February 1999 measured outflow was significantly overestimated by GA-Stage-II model. All the models, except GA-Stage-II model, performed satisfactorily in predicting December 1996 measured outflow (Fig. 6c). Table 9 summarizes the computed relative errors with respect to peak discharge and time to peak for each event for all the models. According to Table 9, GA-Elevation-I model outperformed the other models with less than 3% error in predicting peak rates of the measured hydrographs. GA-Stage-II model showed poor performance with more than 20% error. With respect to the time to peak error, all reached the peak on time except GA-Stage-I and RCM models producing on the average 7% error (see Table 9).

Santa Lucia–Monte Molino reach

Fig. 7 presents simulations of 3 different hydrographs measured at Monte Molino station (Fig. 4) by the RCM model and the proposed equations [Eqs. (6)–(9)] whose parameters were obtained by the GA algorithm. As seen in Fig. 7a, RCM reached peak earlier and it underestimated the peak while GA-Stage-II model overestimated it. Other models showed comparable satisfactory performance for this hydrograph. According to Fig. 7b, although all the models captured the trend of the hydrograph closely, the RCM and GA-Elevation-I models predicted both the peaks satisfactorily while GA-Stage-II model underpredicted them. According to Fig. 7c, the prediction of December 2000 hydrograph was satisfactory by the RCM and GA-Elevation-I models. Although RCM captures the peak rate, it has a fluctuating trend around the peak (see Fig. 7c). Although GA-Elevation-I model slightly underpredicts the peak, it follows the trend of the hydrograph. GA-Stage-II model significantly overpredicted the peak. Table 10 summarizes the computed relative errors with respect to peak discharge and time to peak for each event for all the models. According to Table 10, with respect to the peak rate percent error, all the models produced comparable errors around 10% while GA-Stage-II model produced about 20% error. With respect to the time to peak percent error, all produced less than, on the average, 5% error.

Looking at the performances of the models based on the kinematic wave theory (GA-Stage II and GA-Elevation II), one could see that they provided results less accurate than those obtained by the other two models, and the one based on RCM approach. This was expected because of explicit representation of lateral inflows formulated as a function of stages at downstream end. This representation is, of course, an empirical approximation that does not guarantee the real contribution of the intermediate basin. On the other hand, the GA model based on RCM approach incorporating implicitly the lateral inflows, as proposed by Moramarco et al. (2005), has better performance for the selected flood events.

Above results imply that GA-Elevation-I model overall performed more satisfactorily among the proposed GA models for predicting hydrographs in the three river reaches.

Table 10 Percentage errors in peak discharge [E_{Op}] and time to peak [E_{Tp}] for the events in Fig. 7 [Santa Lucia–Monte Molino reach]

Event	E_{Tp} (%)									
	GA-Stage-I	GA-Stage-II	GA-Elevation-I	GA-Elevation-II	RCM	GA-Stage-I	GA-Stage-I	GA-Elevation-I	GA-Elevation-II	RCM
April 1997	3.4	10.6	-6.2	6.0	-12.8	1.2	0	0	0	-1.2
January 2001	-3.4	-11.5	0.45	-7.1	3.5	-8.3	0	0	0	0
December 2000	25.5	30.4	-17.0	23.3	7.1	1.8	-4.2	-8.9	0	-13.7
Average	10.8	17.5	7.9	12.1	7.8	3.8	1.4	3.0	0	5.0

Table 11 Mean absolute error (MAE) (m³/s) and Nash-Sutcliffe (NS) efficiency index values for GA-Elevation-I and RCM Models

	MAE (m ³ /s)		NS(%)	
	GA	RCM	GA	RCM
<i>S. Lucia–P. Felcino</i>				
December 1990	8.9	10.7	99	98
January 1994	4.3	10.2	99	90
January 2003	5.9	6.9	98	96
Average	6.4	9.3	99	95
<i>S. Lucia–P. Nuovo</i>				
January 2001	8.1	35.9	98	70
February 1999	13.1	29.4	100	97
December 1996	16.6	16.5	99	99
Average	12.3	27.3	99	89
<i>S. Lucia–M. Molino</i>				
April 1997	28.9	23.7	95	96
January 2001	24.3	12.9	90	97
December 2000	39.7	35.8	94	95
Average	30.9	24.1	93	96

Hence, we further analyse its performance against RCM model using additional error measures of MAE and Nash-Sutcliffe (NS) efficiency index. As pointed out earlier, MAE is one of the commonly employed error functions in the literature (Chang et al., 2005). According to Taji et al. (1999), it has the advantage of being less influenced by anomalous data than the square error functions. The NS index is also a commonly employed goodness-of-fit parameter (ASCE, 1993; Erpul et al., 2003; Kalin et al., 2003; Tayfur et al., 2007; Bardossy, 2007; among others) that can be applied to a variety of models (McCuen et al., 2006).

Table 11 presents the computed MAE and NS index values for the application events in Figs. 5–7 for the three reaches. As seen, for Santa Lucia–Ponte Felcino reach, GA model produced less MAE value and higher NS index (about 99%) than the RCM model for each event. Similar performance was observed for the Santa Lucia–Ponte Nuovo reach as well (Table 11). Especially for January 2001 event (Fig. 6a), GA produced 98% efficiency while RCM had 70%. For the Santa Lucia–Monte Molino reach, while RCM resulted in slightly less MAE value for each event, both the models had comparable NS efficiency index values (Table 11). GA produced on the average, for this reach, 92% NS while RCM had 96% efficiency. The reason that GA had slightly poor performance with respect to peak rate

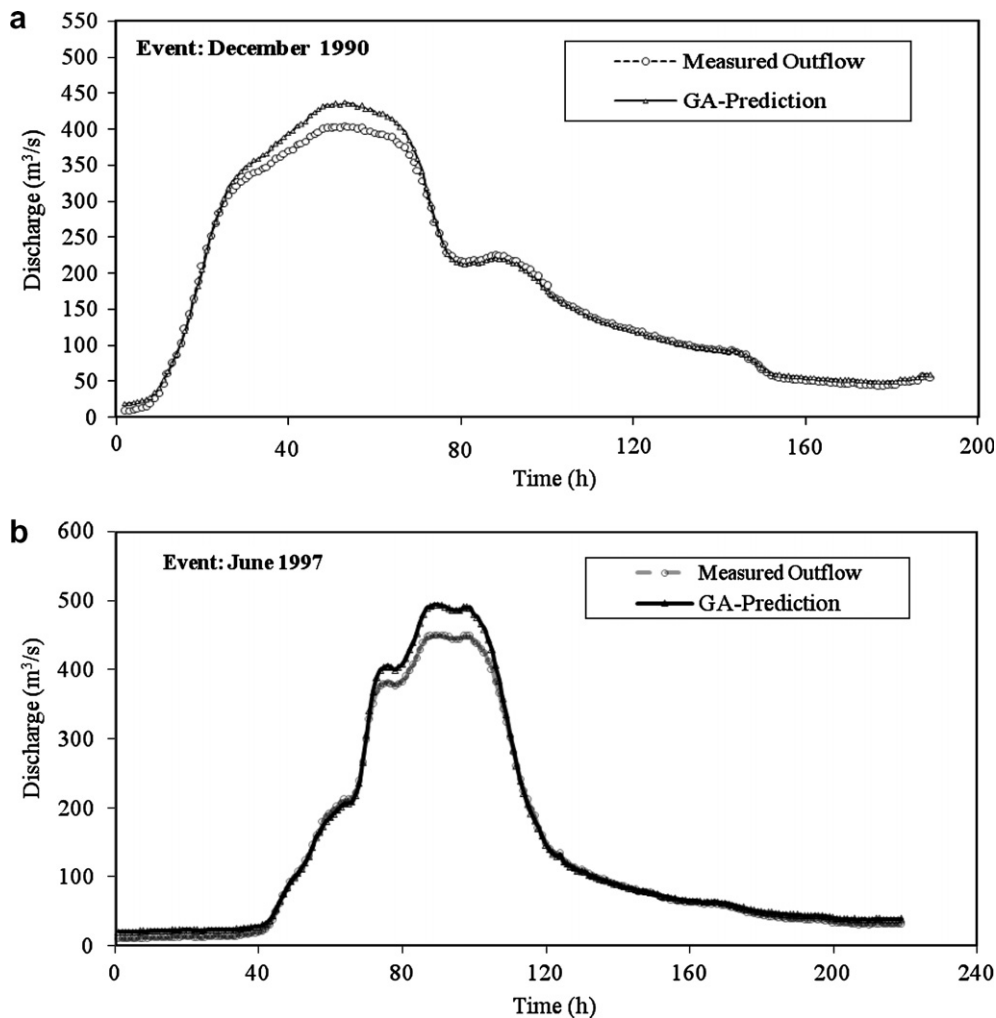


Figure 8 Simulation of outflow hydrographs observed at Ponte Felcino station in (a) December 1990; and (b) June 1997.

prediction than the RCM model for the Santa-Lucia–Monte Molino reach might be due to the fact that GA employs an average wave travel time for each event, as opposed to the exact wave travel time required by the RCM. As the distance between upstream and downstream stations increases, the assumed average wave travel between the two stations might not be very representative for each event. However, we have to also point out that for this branch the performance of the GA model can be considered satisfactory since the intermediate basin is about 4300 km² (85% of the whole basin) and water depth at section ends are referred to the same datum. As far as authors know, a so simple hydrologic approach which is only based on the two parameters assessment and able to simulate the discharge hydrograph at downstream end without having information on lateral inflows and flow areas at both ends is surely an enhancement in the practical hydrology. Nevertheless, when the computed MAE and NS values for all the events for the three river reaches in Table 11 are evaluated, the GA, overall, produced less MAE (16.5 m³/s) than the

RCM which had 20.2 m³/s of MAE. Also GA, overall, had a better efficiency with 97% than the RCM of 93%.

GA model extrapolation

This section investigates the extrapolation capability of GA model by calibrating the model parameters with lower peak hydrographs and then predicting the higher peak hydrographs. For this purpose, GA-Elevation I model was employed. As presented above, GA-Elevation I model, in general, showed satisfactory performance in predicting hydrographs observed at different river reaches. For the Santa Lucia–Ponte Felcino reach, January 1994 event (whose peak rate is 240.7 m³/s) was employed for calibration and December 1990 (whose peak rate is 404.2 m³/s) and June 1997 (whose peak rate is 449.6 m³/s) (Table 2) were chosen for the testing. It has to be pointed out that for both testing events the contribution of the intermediate basin is very significant. Fig. 8a and b show the simulations of December 1990 and June 1997 events, respectively. As

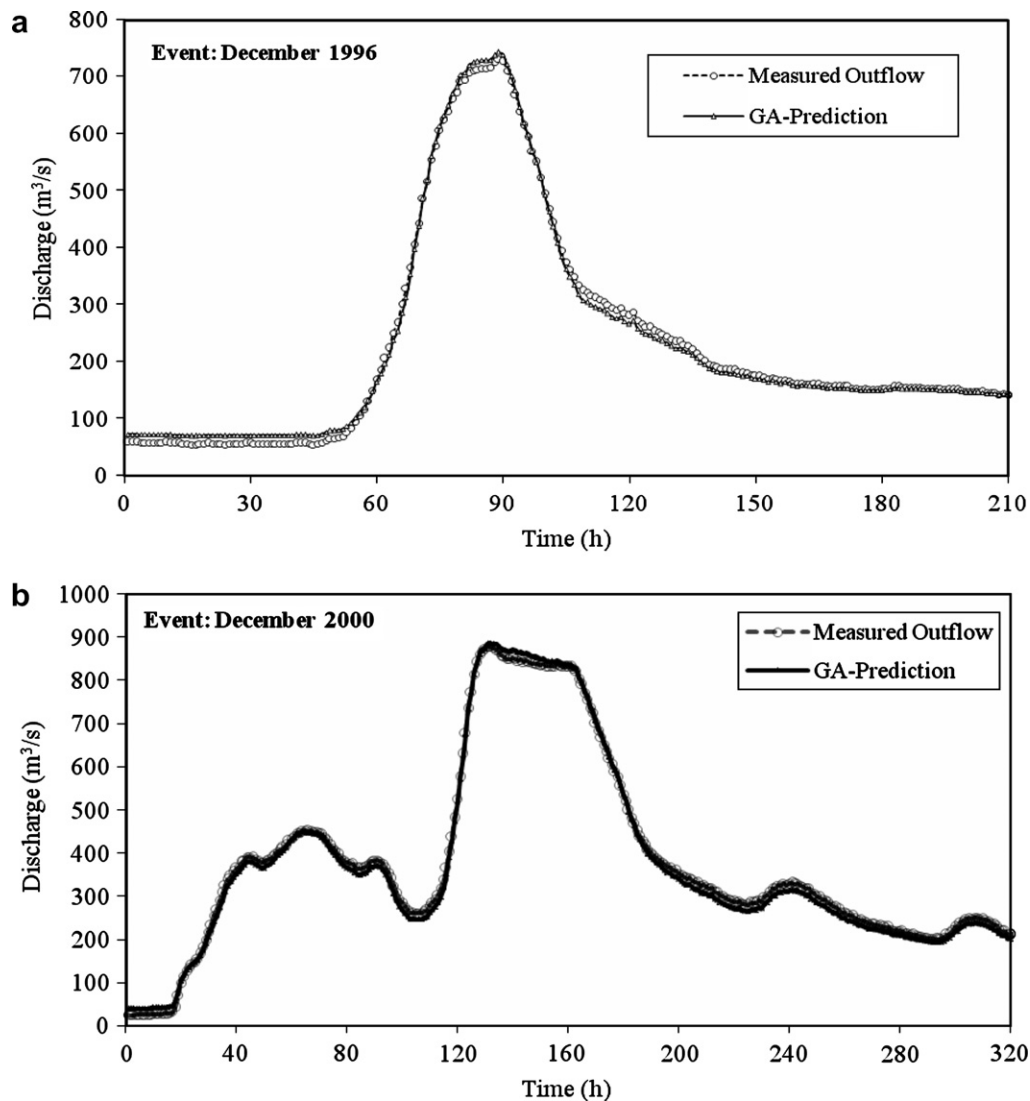


Figure 9 Simulation of outflow hydrographs observed at Ponte Nuovo Station in (a) December 1996; and (b) December 2000.

seen, the model showed a good performance in capturing the trends, time to peak (0% error), and peak rates (with relative errors of 8.2% and 9.6% for December 1990 and June 1997, respectively). The computed MAE values are 8.45 m³/s and 9.75 m³/s for December 1990 and June 1997 events, respectively. Although the peak rates predicted is almost 70% (in the case of December 1990) and 90% (in the case of June 1997) more than the peak of the event used in the calibration, the GA produced less than 10% error in predicting the peak rates. On the other hand, according to Fig. 5 in Tayfur et al. (2007), this error would have been almost 50% in the case of ANN extrapolation.

For Santa Lucia–Ponte Nuovo reach, April 1997 event (whose peak is 498.5 m³/s) was employed in calibration while December 1996 (whose peak is 728.4 m³/s) and December 2000 (whose peak is 877.2 m³/s) events (Table 3) were used for testing. Fig. 9a and b shows the simulations of December 1996 and December 2000 events, respectively. As seen, the model satisfactorily simulated both the hydrographs, capturing the time to peak with zero errors, and the

peak rates with 1.6% error for December 1996 and less than 1% error for December 2000. The computed MAE values are 8.71 and 11.4 m³/s for December 1996 and December 2000, respectively. Although the peak rates predicted is almost 50% (in the case of December 1996) and 80% (in the case of December 2000) more than the peak of the event used in the calibration, the GA produced less than 2% error in predicting the peak rates. This error, when employing ANN, according to Fig. 5 in Tayfur et al. (2007), would have been about 35% for December 1996 and 45% for December 2000.

For the Santa Lucia–Monte Molino reach, January 2001 event (whose peak is 397.3 m³/s) was used in the calibration and December 1998 (whose peak is 770 m³/s) and February 1999 (whose peak is 754 m³/s) (Table 4) were employed for testing. As seen the peaks of the tested hydrographs are almost 2 times the peak of the hydrograph used for calibrating the model parameters. According to Fig. 5 in Tayfur et al. (2007), the ANN would have made about 50% error in predicting these rates. Fig. 10a and b, on the other hand, show the simulation of the December 2000, and February 1999

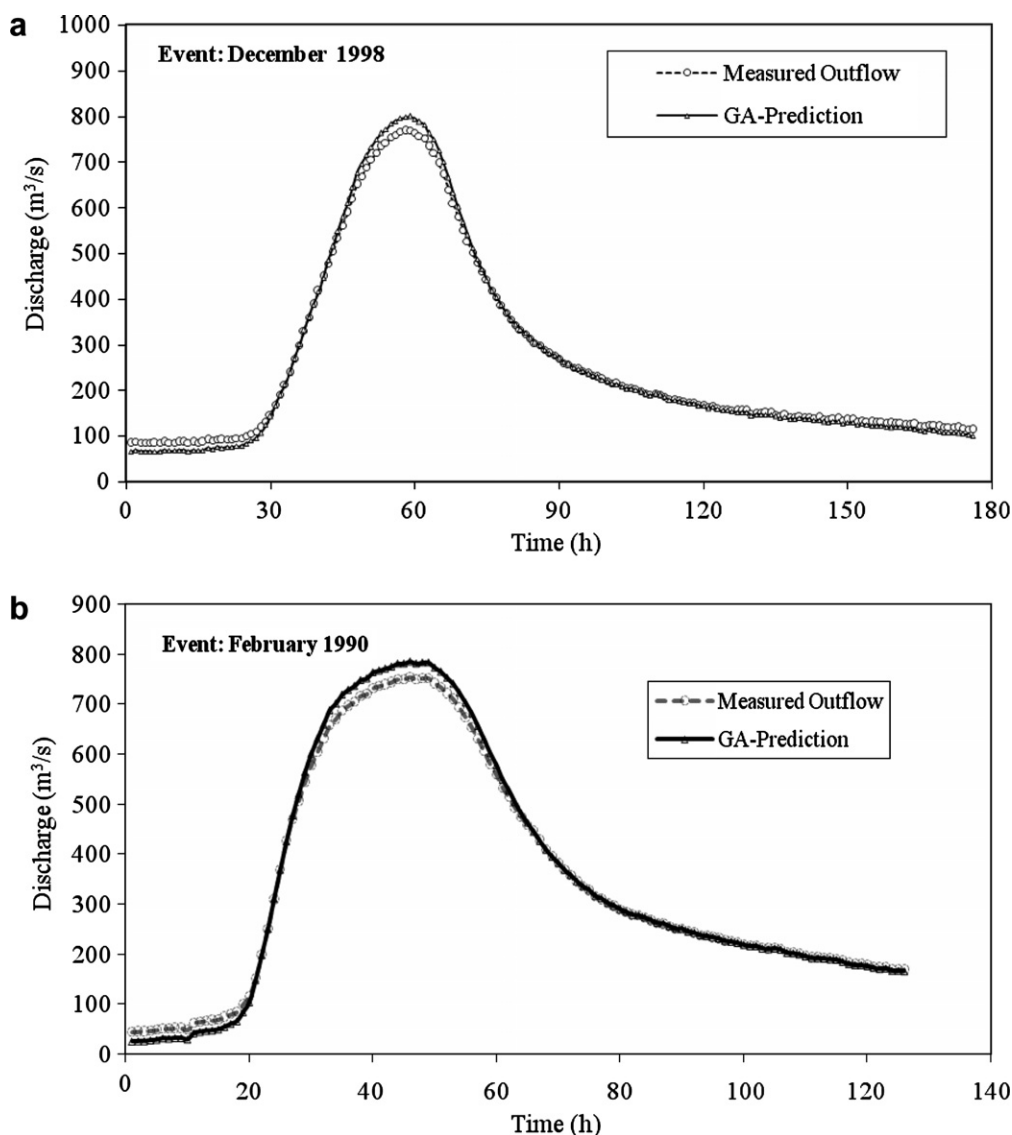


Figure 10 Simulation of outflow hydrographs observed at Monte-Molino Station in (a) December 1998; and (b) February 1990.

hydrographs, respectively. As seen, the model satisfactorily simulated the hydrographs, capturing the time to peak with zero errors, the peak rates with less than 4% error, and MAE values less than $12 \text{ m}^3/\text{s}$.

Above results imply that the GA model, unlike the ANNs, does not have extrapolation problem. It can be calibrated with lower peak events to predict higher peak hydrographs. This has an implication such that GAs could be employed for predicting hydrographs at ungauged basins.

Summary and concluding remarks

This study developed models to predict hourly-based hydrographs using level data only. It proposed two main formulations based on (1) RCM model of Moramarco et al. (2005) relating downstream discharge to upstream discharge and flow cross-sectional areas at both upstream and downstream stations through model parameters; and (2) kinematic wave model that expresses flow rate at each upstream and downstream station as a function of corresponding flow stage. In each approach, two different level data were used: (1) flow stage; and (2) water surface elevation (the elevation of a station from a reference datum plus the water level). As such, in total, four new formulations expressed by Eqs. (6)–(9), were proposed whose optimal parameter values were obtained by the genetic algorithms.

The developed models were calibrated and tested by measured hydrographs. The results indicated that GA-Elevation-I model, expressed by Eq. (8), showed satisfactory performance in capturing trend and peaks of the hydrographs observed at different river reaches whose wave travel time varied from 4 to 12 h. GA-Elevation-I model outperformed the RCM model whose parameters were obtained through classical method as outlined earlier. GA-Stage-II model (Eq. 7), in general, showed poor performance while other two models (Eqs. 6 and 9) showed comparable performances with RCM model.

The extrapolation capability of GA was also investigated in this study. For this purpose, GA-Elevation-I model was employed. It was calibrated with lower peak hydrographs and then applied to predict higher peak events for each river reach. The results showed that the GA model does not have an extrapolation problem.

Since GA-Elevation-I model turned out to be the best performing model, this model can be employed as a modeling tool for prediction purposes. This model, unlike of RCM approach, has the great benefit such that it requires elevation data without needing of cross sectional flow area and knowledge of inflows, widening its application also to ungauged river sites.

Acknowledgements

We thank the financial support (*International short-term mobility program for scientists/researchers from Italian and foreign institutions*) for the first author by the CNR (National Research Council) of Italian Government to carry out this research at the CNR-IRPI office of Perugia, Italy. The suggestions of reviewers for significantly improving the contents of the paper are gratefully acknowledged. Finally,

authors also thank the hydrology group of CNR-IRPI for their scientific and technical assistance.

References

- ASCE, 1993. Criteria for evaluation of watershed models. *J. Irrig. Drain. Eng.* 119 (3), 429–442.
- Bardossy, A., 2007. Calibration of hydrological model parameters for ungauged catchments. *Hydrol. Earth Syst. Sci.* 11, 703–710.
- Caleffi, V., Valiani, A., Zanni, A., 2003. Finite volume method for simulating extreme flood events in natural channels. *J. Hydraulic Res.* 41 (2), 167–177.
- Chang, C.L., Lo, S.L., Yu, S.L., 2005. Applying fuzzy theory and genetic algorithm to interpolate precipitation. *J. Hydrol.* 314 (1–4), 92–104.
- Cheng, C.T., Wu, X.Y., Chau, K.W., 2005. Multiple criteria rainfall-runoff model calibration using a parallel genetic algorithm in a cluster of computer. *Hydrol. Sci. J.* 50 (6), 1069–1087.
- Cheng, C.T., Zhao, M.Y., Chau, K.W., Wu, X.Y., 2006. Using genetic algorithm and TOPSIS for Xinanjiang model calibration with a single procedure. *J. Hydrol.* 316 (1–4), 129–140.
- Erpul, G., Norton, L.D., Gabriels, D., 2003. Sediment transport from interrill areas under wind-driven rain. *J. Hydrol.* 276, 184–197.
- Franchini, M., Lamberti, P., 1994. A flood routing Muskingum type simulation and forecasting model based on level data alone. *Water Resour. Res.* 30 (7), 2183–2196.
- Franchini, M., Lamberti, P., Di Giammarco, P., 1999. Rating curve estimation using local stages, upstream discharge data and a simplified hydraulic model. *Hydrol. Earth System Sci.* 3 (4), 541–548.
- Goldberg, D.E., 1999. *Genetic Algorithms*. Addison–Wesley, USA.
- Guan, J., Aral, M.M., 1998. Progressive genetic algorithm for solution of optimization problems with nonlinear equality and inequality constraints. *J. Appl. Math. Model.* 23, 329–343.
- Guan, J., Aral, M.M., 2005. Remediation system design with multiple uncertain parameters using fuzzy sets and genetic algorithm. *J. Hydrologic Eng., ASCE* 10 (5), 386–394.
- Jain, A., Bhattacharjya, R.K., Sanaga, S., 2004. Optimal design of composite channels using genetic algorithm. *J. Irrig. Drain. Eng., ASCE* 130 (4), 286–295.
- Kalin, L., Govindaraju, R.S., Hantush, M.M., 2003. Effect of geomorphological resolution on modeling of runoff hydrograph and sedimentograph over small watersheds. *J. Hydrol.* 276, 89–111.
- Kouwen, N., Danard, M., Bingeman, A., Luo, W., Senglenieks, F.R., Soulis, D.E., 2005. Case study: watershed modeling with distributed weather model data. *J. Hydrologic Eng.* 10 (1), 23–38.
- Liong, S.Y., Chan, W.T., ShreeRam, J., 1995. Peak flow forecasting with genetic algorithm and SWMM. *J. Hydraulic Eng., ASCE* 121 (8), 613–617.
- McCuen, H.R., Knight, Z., Cutter, A.G., 2006. Evaluation of the Nash–Sutcliffe efficiency index. *J. Hydrologic Eng.* 11 (6), 597–602.
- Moramarco, T., Singh, V.P., 2001. Simple method for relating local stage and remote discharge. *J. Hydrologic Eng., ASCE* 6 (1), 78–81.
- Moramarco, T., Singh, V.P., 2002. Accuracy of kinematic wave and diffusion wave for spatial-varying rainfall excess over a plane. *Hydrol. Processes* 16, 3419–3435.
- Moramarco, M., Barbetta, S., Melone, F., Singh, V.P., 2005. Relating local stage and remote discharge with significant lateral inflow. *J. Hydrologic Eng., ASCE* 10 (1).
- Moramraco, T., Barbetta, S., Melone, F., Singh, V.P., 2006. A real-time stage Muskingum forecasting model for a site without rating curve. *Hydrol. Sci. J.* 51 (1), 66–82.

- Muleta, M.K., Nicklow, J.W., Bekele, E.G., 2007. Sensitivity of a distributed watershed simulation model to spatial scale. *J. Hydrologic Eng.* 12 (2), 163–172.
- Palisade Corporation. Evolver, The Genetic Algorithm Solver for Microsoft Excel, September, 2001, Newfield, New York, USA.
- Price, R.K., 1973. Flood routing methods for British rivers. *Proc. Inst. Civ. Eng. Waters, Maritime Eng.* 55 (12), 913–930.
- Sen, Z., Oztopal, A., 2001. Genetic algorithms for the classification and prediction of precipitation occurrence. *Hydrol. Sci. J.* 46 (2), 255–267.
- Sen, Z. Genetic algorithm and optimization methods. Su Vakfi Yayinlari, Istanbul. (Turkish), 2004. ISBN: 975-6455-12-8.
- Singh, V.P., 1996. Kinematic Wave Modeling in Water Resources: Surface Water Hydrology. John Wiley, Hoboken, N.J.
- Singh, A.K., Kothyari, U.C., Raju, K.G.R., 2004. Rapidly varying transient flows in alluvial channels. *J. Hydraulic Res.* 42 (5), 473–486.
- Singh, R.M., Datta, B., 2006. Identification of groundwater pollution sources using GA-based linked simulation optimization model. *J. Hydrologic Eng., ASCE* 11 (2), 101–109.
- Taji, K., Miyake, T., and Tamura, H. 1999. On error back propagation algorithm using absolute error function. In: *International Conference on Systems, Man, and Cybernetics, IEEE SMC'99 Conference Proceedings, V. 5*, pp. 401–406.
- Tayfur, G., Moramarco, T., Singh, V.P., 2007. Predicting and forecasting flow discharge at sites receiving significant lateral inflow. *Hydrol. Processes* 21, 1848–1859.
- Wu, C.L., Chau, K.W., 2006. A flood forecasting neural network model with genetic algorithm. *Int. J. Environ. Poll.* 28 (3–4), 261–273.
- Ying, X., Khan, A.A., Wang, S.Y., 2004. Upwind conservative scheme for the Saint Venant equations. *J. Hydraulic Eng.* 130 (10), 977–987.