A NON-LINEAR RAINFALL-RUNOFF MODEL WITH A SIGMOID GAIN FACTOR TO SIMULATE FLOW FREQUENCY DISTRIBUTION CURVES FOR AMAZON CATCHMENTS

MARCO VALÉRIO A. VINAGRE, CLAUDIO J.C. BLANCO, ANDRÉ L. AMARANTE MESQUITA

Doctoral Program in Natural Resources Engineering in the Amazon – PRODERNA/ITEC/UFPA – Federal University of Pará, Rua Augusto Corrêa, 01, 66075-110, Belém-Pará-Brazil; Mailto: vinagre@ufpa.br; blanco@ufpa.br; andream@ufpa.br

The objective of this paper is to simulate flow frequency distribution curves for Amazon catchments with the aim of scaling power generation from small hydroelectric power plants. Thus, a simple nonlinear rainfall-runoff model was developed with sigmoid-variable gain factor due to the moisture status of the catchment, which depends on infiltration, and is considered a factor responsible for the nonlinearity of the rainfall-runoff process. Data for a catchment in the Amazon was used to calibrate and validate the model. The performance criteria adopted were the Nash-Sutcliffe coefficient (R^2), the RMS, the $Q_{95\%}$ frequency flow percentage error, and the mean percentage errors ranging from $Q_{5\%}$ to $Q_{95\%}$. Calibration and validation showed that the model satisfactorily simulates the flow frequency distribution curves. In order to find the shortest period of rainfall-runoff data, which is required for applying the model, a sensitivity analysis was performed whereby rainfall and runoff data was successively reduced by 1 year until a 1.5-year model application minimum period was found. This corresponds to one hydrological year plus the 6-month long "memory". This analysis evaluates field work in the ungauged sites of the region.

KEY WORDS: Hydrological Modeling, Flow Frequency Distribution Curves, Amazon Small Catchments.

Marco Valério A. Vinagre, Claudio J. C. Blanco, André L. Amarante Mesquita: NELINEÁRNY ZRÁŽKOODTOKOVÝ MODEL SO SIGMOIDÁLNYM ZDROJOVÝM FAKTOROM NA SIMULÁCIU ČIAR ROZDELENIA PRIETOKOV PRE POVODIA AMAZONKY. J. Hydrol. Hydromech., 59, 2011, 3; 11 lit., 10 obr., 5 tab.

Cieľom tohto príspevku je simulácia čiar rozdelenia prietokov pre povodia rieky Amazonka pre potreby hodnotenia premeny energie v malých hydroelektrárňach. Preto bol vyvinutý jednoduchý nelineárny zrážko-odtokový model so sigmoidálne sa meniacim zdrojovým faktorom v závislosti od obsahu vody v povodí, ktorý závisí od infiltrácie a je považovaný za faktor, spôsobujúci nelinearitu zrážkoodtokových procesov. Pre kalibráciu a validizáciu modelu boli použité údaje z povodí rieky Amazonka. Použili sme tieto hodnotiace kritériá: Nashov-Sutcliffov koeficient (R²), RMS, $Q_{95\%}$, chyba určenia odtoku v percentách, a priemerná percentuálna chyba v rozsahu od $Q_{5\%}$ do $Q_{95\%}$. Kalibrácia a validizácia ukázala, že model simuluje čiary rozdelenia prietokov uspokojivo. Aby bolo možné nájsť najkratšie obdobie pre nájdenia závislosti zrážky – odtok, ktorá je potrebná pre aplikáciu v modeli, použili sme citlivostnú analýzu tak, že údaje zrážky – odtok boli postupne redukované o jeden rok, až kým nebolo nájdené minimálne obdobie pre aplikáciu vzťahu zrážky – odtok 1,5 roka. Toto obdobie zodpovedá jednému hydrologickému roku, plus 6 mesiacov dlhá "pamäť". Touto analýzou boli vyhodnotené výsledky terénnych meraní v oblastiach, kde neboli k dispozícii merania odtoku.

KĽÚČOVÉ SLOVÁ: hydrologické modelovanie, čiara rozdelenia prietokov, malé povodia riekyAmazonka.

1. Introduction

Rainfall-runoff modeling constitutes an important tool for projects in hydroelectric power plant reservoirs (*Kachroo*, 1992). However, one must consider the problem of over-configuration of

models, a topic that has occupied the attentions of many researchers, with cases in which thirteen parameters were effectively replaced by only four (*Jakeman* and *Hornberger*, 1993).

According to experience obtained from a number of seminars on river flow forecast, which were held

at University College Galway, Ireland (*Ahsan*, 1994) using linear components with constant gain factors, the efficiency of the model was shown to be more sensitive to the magnitude of the gain factor than to the actual shape of the impulse response function. Moreover, the non-linearity of the rainfall-runoff process has been known for so long due to infiltration, which is dependent upon soil moisture and evaporation, among other factors (*Jun*, 2001).

The Amazon region presents problems with respect to obtaining data for conducting proper hydrological modeling and for quantitative assessment of the main components of the hydrological cycle. In addition to data shortage, another complicating factor for hydrological modeling is the complexity of the watershed's hydrological processes such as evapotranspiration, soil moisture and runoff generation (*Ribeiro Neto*, 2008).

In this context, *Blanco* et al. (2005) found that the data commonly available for small catchments in the Amazon are those related to rainfall, and they thus investigated a rainfall-runoff cause-and-effect relationship. To do that, the simple linear model was applied, calibrated, and validated and had its sensitivity analyzed over a catchment in the Amazon. The model was used to simulate flow frequency distribution curves required for hydro-energy studies, and had satisfactory results, with errors of less than 9% in the range between $Q_{10\%}$ and $Q_{95\%}$.

Thus, the objective of this paper is to contribute to the simulation of flow frequency distribution curves of catchments in the Amazon through a nonlinear, calibrated, validated hydrological model of an Amazon catchment, with an aim of sizing the power generation of small hydroelectric power plants.

2. Test catchment

Fig. 1 shows the test catchment used to calibrate and validate the hydrological model that was analyzed.

The test catchment is located 160 km east of Belém, in Pará State, Brazil, and has a drainage area of 82 km². The fluviometric station 31600000F, known as Marambaia, in that catchment, is located at latitude $-01^{\circ}39'06''$ and longitude $-47^{\circ}07'03''$ with altitude equal to 50m, within the limits of the municipality of Capitão Poço. It is the only station in the region that has long series of flow data; however, it lacks rainfall data. For that reason, data was taken from the nearest rainfall station, 00147016P with altitude equal to 40m, in the city of Ourém (latitude $-1^{\circ}33'02''$, longitude $-47^{\circ}07'01''$), some 15 km distant from the fluviometric station of the catchment under study. The difference in altitude between the rainfall and fluviometric station is equal to 10m. That is the same catchment analyzed by *Blanco* et al. (2005). The authors analyzed rainfall correlation in the region and concluded that the data from the Ourém rainfall station (Fig. 1) can be used as data for the Marambaia fluviometric station (Fig. 1).

Fig. 2 shows the flow frequency distribution curves (raw and smooth) observed for the test catchment.



Fig. 1. Test catchment.

It can be seen that the smooth curve adequately represents the raw curve, and, henceforth, the smooth curve will be used in the comparison of the results of simulated flows for the test catchment.



Fig. 2. Raw and smooth flow frequency distribution curves.

3. Methodology

3.1 Simple Linear Rainfall-Runoff Model (SLM)

Since the introduction of the unit hydrograph theory by *Sherman* (1932), the non-parametric form of linear time-invariant model has occupied a singular niche in the history of the development of deterministic rainfall-runoff models. In time domain, the relationship between input x(t) and output y(t), of a linear time-invariant system in the continuous non-anticipative, initially relaxed, time is expressed by the convolution integral of the form,

$$y(t) = \int_{0}^{1} U(\tau)x(t-\tau)dt,$$
 (1)

where $U(\tau)$ is the unit impulse response function of the system. Since the input variable x(t) is the effective rainfall in the basin, with the dependent output variable y(t) being the effective flow resulting from the rainfall (that is, the stream discharge minus the base flow) at the mouth of the basin, the impulse response of the system is conceptually identical to the unit hydrograph of the basin, as introduced by *Clark* (1945), generalizing the theory of unit hydrograph. By definition, the total effective rainfall should equal the volume of runoff; therefore, the instantaneous unit hydrograph of a basin must meet the following equality:

$$\int_{0}^{\infty} U(\tau) \, \mathrm{d}\tau = 1.0 = \text{unit_volume.}$$
(2)

However, when a linear system is nonconservative, i.e., considering the loss in the input volume due to infiltration and evaporation, the above-stated equality does not exist, and the area between the time axis and the impulse response function is often called steady gain factor of system G, which can be represented by:

$$G = \int_{0}^{\infty} \mathbf{U}(\tau) \,\mathrm{d}\tau \tag{3}$$

i.e., the Gain Factor is smaller than the unit, as it incorporates the effects of infiltration and evaporation, which in the case of catchments in the Amazon are not known, allowing the model to be applicable to the available data.

Normalizing the unit impulse response function, using the gain factor, as follows:

$$h(\tau) = U(\tau) / G \tag{4}$$

from (1) above the following is obtained:

$$y(t) = G \int_{0}^{t} h(\tau) x(t-\tau) \mathrm{d}\tau, \qquad (5)$$

where the kernel is called the normalized unit impulse response function.

In rainfall-runoff modeling, the relationship explained by the Eq. (5) with $G \neq 1$ occurs when a linear and time-invariant relationship is established between the total rainfall and the total discharge of a basin, rather than between effective rainfall and effective runoff. The resulting model is the continuous form of the Simple Linear Model (SLM) (Blanco et al., 2005). The SLM, though simple, is an important starting point in modeling rainfallrunoff, as opposed to the application of the unit hydrograph theory, which requires considerable effort in determining the effective rainfall and separating the base flow from total flow. The SLM can be applied to raw data and as such, can serve as an expeditious verification of linearity existing in the rainfall-runoff relationship.

In the absence of a perfect linear relationship between rainfall (total) x(t) and runoff (total) y(t), it is more appropriate to rewrite Eq. (5) defining the SLM with the inclusion of a model output error e(t)term, as follows:

$$y(t) = G \int_{0}^{t} h(\tau) x(t-\tau) d\tau + e(t) .$$
 (6)

Assuming that the error term e(t) has an expected value of zero and that both the input x(t) and the output y(t) have non-zero expected values, it is verified that the gain factor *G* satisfies the relationship of the form:

$$G = \lim_{T \to \infty} \frac{\int_{0}^{T} y(t) dt}{\int_{0}^{T} x(t) dt},$$
(7)

where a mathematically relaxed system at the beginning and the end was assumed. Within the context of the SLM, the above equality implies that for a sufficiently long period of rainfall-runoff data, the gain factor G of the SLM is approximately equal to the relationship between the total runoff volume and the total rainfall accumulated in the period.

Throughout the SLM adjustment process using real rainfall-runoff data, the gain factor estimated by Eq. (3) is often verified to be approximately equal to the long-term runoff coefficient for the basin being considered, as estimated by Eq. (7) with T set as the calibration period instead of the infinite (Kachroo and Natale, 1992). This merely reflects the linearity and time invariance assumed and inherent to the formulation of the SLM. At any rate, it is not plausible that in nature the constant proportion of rainfall would always be transformed into runoff. Rather, it is known that the higher the moisture of the basin, the greater the proportion of rainfall that would become runoff. Since the formulation of the SLM does not accept this intuitive concept, the flow simulated in the SLM is expected to overestimate the dry season flows and to underestimate the rainy season flows, as found in Blanco et al. (2005) and Labat and Mangin (2000). The application of the SVM model aims to improve the simulation of the dry season flows that characterize high frequency flows in the flow frequency distribution curves. Such flows are used to size the power generation of small hydroelectric power plants.

3.2 Sigmoid-Variable Gain Factor Model (SVM)

For the proposed model, the following inputoutput relationship is assumed:

$$y(t) = \int_{0}^{t} G(\tau)h(\tau)x(t-\tau)\mathrm{d}\tau + e(t), \qquad (8)$$

where G is the sigmoid-variable gain factor.

3.2.1 Sigmoid function

An input-output relationship of the form given by Eq. (8) where G(t) presents the function in a sigmoid form will be hereinafter referred to as Sigmoid-Variable Gain Factor Model or simply SVM, the input of which is the total rainfall x(t) and whose output is the total flow y(t).

Although the rating for the variable gain factor G(t) in (8) and in subsequent equations contains the time t as an argument, it should be understood that G(t) cannot be generally expressed in the parametric form as a variable equation at time t. Instead, G(t) should be viewed as a time function whose instantaneous value depends on the moisture u(t) of the drainage basin soil. This assumption holds the concept of temporal SVM variability.

The following considerations are relevant to the selection of the appropriate function to represent the gain factor:

- The gain factor varies between zero and the unit. This condition reflects the equation of mass conservation, as reflected in the physical impossibility of surface runoff volume larger than the rainfall volume.
- The gain factor should increase monotonically with increasing moisture in the basin.

While many other functions may be considered for describing the variation of the gain factor with respect to the function of soil moisture in the basin u(t), this paper focuses on the sigmoid function, which is:

$$G(u(t)) = \frac{1}{(1+e^{-u(t)})}.$$
(9)

Fig. 3 shows the behavior of the sigmoid function adopted for the variable gain factor of the proposed model.



Fig. 3. Sigmoid function graph.

Adoption of this function takes into account the behavior of the rainfall-runoff phenomenon in nature, as the rainfall volume is divided and seeps into the soil, runs off on the surface and evaporates. If the soil has low moisture, infiltration is facilitated to the detriment of the runoff and evaporation. As moisture increases, infiltration decreases and the portion that runs on the surface increases, covering the natural terrain depressions, and exposing the water to the action of evaporation.

It can be observed that the gain factor G(u(t)) in Eq. (9) varies depending on the soil moisture u(t), which may not be a simple function of time and may depend upon the inputs and/or outputs of the system. As such, Eq. (8) can be considered as a nonlinear input-output ratio. Thus, the nonlinearity of the input-output ratio of (9) is simple in nature and there is the possibility of submitting the SVM to a linear treatment through transformation of the gain factor function as per below.

Eq. (9) states that:

$$(1+e^{-u(t)}) = \frac{1}{G(t)} \Rightarrow e^{-u(t)} = \frac{1}{G(t)} - 1 \Rightarrow$$
$$e^{-u(t)} = \frac{1-G(t)}{G(t)} e^{u(t)} = \frac{G(t)}{1-G(t)}$$
(10)

whence

$$u(t) = Ln \left[\frac{G(t)}{1 - G(t)} \right]. \tag{11}$$

This is equivalent to rendering of the sigmoid function linear by the 'logarithmization' of Eq. (10), formed from the sigmoid variable gain factor G(t), so as to simplify its use in the estimation of flows, as will be seen below.

3.2.2 Auxiliary function – weighted daily rainfall z(t)

For Kachroo and Liang (1992), the "system memory" is the time interval between the occurrence of precipitation and the moment when their influence on the flow ceases. In determining the non-linear gain factor, ranging between zero and the unit, it is necessary to determine the rainfall/runoff ratio and then its normalization, dividing it by its peak during the period. It turns out that for daily rainfall/runoff data, there may be days without precipitation, in which, the daily rainfall, x(t) == 0, brings about singularities that preclude the determination of the rainfall/runoff ratio and, therefore, the gain factor. To avoid that condition, this study employed the auxiliary function z(t), assumed as the weighted daily rainfall for the period comprising the past *m* days, which represent the "system memory". The function $z_i = z(t)$ (Eq. 12) is obtained from the convolution of the weights by the daily precipitations, as follows:

$$z_{i} = \frac{1}{d} \sum_{j=1}^{j=e} [w_{j} \sum_{k=1}^{k=d} x_{(n=i-j,k+1)}], \qquad (12)$$

where
$$\sum_{j=1}^{j=e} w_j = 100\%$$
 (13)

and the "memory" being m = e.d (14)

In Eq. (12), x_n is the total daily rainfall of day n. Therefore, the "memory" period with a duration of m days shown in the Eq. (14) is subdivided into e subintervals j of d days, the variance of k from 1 to d scans each subinterval, which receives the impact of the weights W_j . Such weights have amplitudes ranging from 0% to 100%, and their sum is equal to 100%, according to Eq. (13).

3.2.3 Auxiliary function of soil moisture u(t)

This function, which has a simple mathematical structure, is based on the weighted daily rainfall of day *i*, $z_i = z(x(t))$, from the assumption that the gain factor of day *i* depends on the function $u_i = u(x(t))$, which represents the state of soil moisture on the day. Thus, $G_i = G(u(x(t)))$, and, therefore, can be written as $G_i = G(u_i)$. Henceforth, for the sake of simplicity, $u_i = u(x(t))$ is called simply moisture.

Since the input data for the model are numerical rainfall-runoff pairs, the inputs of function u_i are the rainfall data of day *i*, x_i , ordered in time, so as to provide a function u_i , whose domain and codomain are consistent with the sigmoid function adopted for the gain factor *G* of SVM.

Thus, function $u(x(t)) = u_i$, adopted to represent the basin's soil moisture, must incorporate the "memory" of the effect of recent and past rainfalls, so as to provide an appropriate indicator of soil moisture on day *i*. For this, the moisture function u_i was assumed, given by

$$u_i = \frac{x_i - z_i}{x},\tag{15}$$

where u_i , dimensionless, is the soil moisture function for day *i*, x_i is the rainfall on day *i*, z_i is the weighted daily rainfall on day *i* and \bar{x} is the daily average rainfall for the entire period, either for calibration or validation.

Therefore, the procedure adopted for estimating u_i and G_i is parsimonious, these two functions being dependent only on rainfall events that make up the "memory" of the system, given that the problem under study is the development of a simple rainfall-runoff model whereby evaporation data and explicit components of water balance are not considered.

3.2.4 Impulse response

After adopting the auxiliary functions (Eqs. (12) and (15)) and through some algebric refinement, Eq. (8) becomes:

$$y(t) = G[u(x(t))]z(x(t)).$$
 (16)

Thus, from the daily rainfall impulse x(t), which determines the values of the auxiliary functions z(x(t)) and u(x(t)) and the impulse response G[u(x(t))], we have the daily runoff Q_{i} , which can simply be rewritten as

$$Q_i = G_{iz_i} \tag{17}$$

i.e., the daily runoff is equal to the product from the gain factor by the daily weighted rainfall.

3.2.5 SVM flowchart

Fig. 4 shows the SVM flowchart applied to rainfall and runoff data for the test catchment. The index *o* represents the observed values, index *s* is the simulated values, such as, in the runoff rate Q_o (observed) and Qs (simulated).

It is observed in Fig. 4 that from the data of observed daily rainfall-runoff, one may determine the values of the auxiliary functions z and u, as well as the Gain Factor G and Ln[G/(1-G)]. After that, using the non-linear/linear transformation, by logarithmization, u is estimated and then finally the values of Q_s , from which the flow frequency distribution curve is generated.



Fig. 4. Flowchart representing the application of SVM.

3.3 Performance criteria

Since the objective of this study is the simulation of flow frequency distribution curves, with Q_o and Q_s being the observed and simulated flows at time *i*, $\overline{Q_o}$ the average flow observed and *n* being the total number of observations for assessing SVM performance, the following performance criteria are suggested:

i) The Nash-Sutcliffe coefficient (R^2) calculated between the observed and simulated flows, defined by

$$R^{2} = 1 - \frac{\sum_{i=1}^{n} (Q_{oi} - Q_{si})^{2}}{\sum_{i=1}^{n} (Q_{oi} - \overline{Q}_{o})^{2}}.$$
(18)

The closer one is to the unit, the better the performance.

ii) The square root of mean square error, RMS, measured between the observed and simulated flows, given by

$$RMS = \sqrt{\frac{\sum_{i=1}^{n} (Q_{oi} - Q_{si})^2}{n}} .$$
(19)

The closer to one is to zero, the better the performance.

iii) The percentage errors for the flow rate of 95% frequency flow, given by

$$e_{Q95\%} = \left| \frac{Q_{s95\%} - Q_{o95\%}}{Q_{o95\%}} \right|.100\%.$$
(20)

The closer one is to zero, the better the performance.

iv) The average of the percentage errors in the interval between the flow rates from $Q_{95\%}$ to $Q_{5\%}$ frequency, given by

$$e_{mediumQ95\% toQ5\%} = \frac{\sum_{i=5\%}^{i=95\%} \left| \mathcal{Q}_{si} - \mathcal{Q}_{oi} \right|}{n},$$
(21)

where n is the number of Q_i values in the interval, and the closer to zero, the better the performance.

4. Results and discussion

4.1 Model calibration

In this case, the data period ranges from 06/05/1993 through 12/31/1999. The seven months of 1993 were used as system memory, simulating the flow frequency distribution curve for the six remaining years of the rainfall and runoff data samples, that is, between the years 1994 and 1999. One of the steps of the calibration process is the optimization of system memory *m*, whereas optimizing it also means optimizing *e* and *d* and weights *w* (Eqs. (12), (13) and (14)). The objective functions used were R^2 (Eq. (18) and *RMS* (Eq. (19)), which also were used as performance criteria. The method employed was that of trial and error. Fig. 5 shows the optimum value of *m* by maximizing R^2 and minimizing *RMS*.



Fig. 5. System memory optimized for six years.

T a b l e 1. w in function of n interval for the calibration model.

Interval for n	w [%]
i to i-14	2.40
i-15 to i-29	2.40
i-30 to i-44	4.00
i-45 to i-59	4.00
i-60 to i-74	17.5
i-75 to i-89	17.3
i-90 to i-104	2.40
i-105 to i-119	2.40
i-120 to i-134	175
i-135 to i-149	17.5
i-150 to i-164	6 20
i-165 to i-179	0.30

Thus, for m = 180, d = 15 and e = 12, the weights w are given in Tab. 1. Fig. 6 shows the flow frequency distribution curves for the observed and simulated flows, as well as the error between them for the calibration period of the test catchment.

In Fig. 6, it can be verified that for the frequency range between 5% and 95%, the percentage errors have a mean value of 6.9%. It can be observed that for $Q_{95\%}$, the simulated flows in relation to the observed ones differ by only 0.4%. It is also observed that at high frequencies in the range of $Q_{98\%}$ to $Q_{100\%}$ there is a distance of 40% between the simulated flows and the observed ones. However, for the purpose of this research, which is scaling of power generation, this does not represent an obstacle.

4.2 Model validation

In this case, the data period ranges from 07/05/2000 through 12/31/2006. It is observed that, after the model calibration and subsequent system memory optimization, the period of the rainfall and runoff data sample is exactly 6 years and 180 days,

and the 180 days of year 2000 were used as the optimized system memory. Fig. 7 shows the flow frequency distribution curves observed and simulated and the error between them for the validation period. As with calibration, the flow frequency distribution curves were simulated for the remaining six years of the validation period, that is, from 2001 to 2006.

Through Fig. 7, it can be verified that in the frequencies between 5% and 95%, the percentage errors are limited to 10%, having a mean value of 5.5%. As occurred with calibration, it is also observed that at high frequencies in the range of $Q_{98\%}$ to $Q_{100\%}$ there is a gap of 40% between the simulated and observed flows. Tab. 2 displays the results for the performance criteria of the application of SVM to the data for the small catchment.

According to Tab. 2, it can be observed that the estimates made by the SVM are significantly close to the observed values. The average errors are lower than 5.5% in the intervals between $Q_{5\%}$ and $Q_{95\%}$, both in the calibration and validation periods.



Fig. 6. Observed and simulated flow frequency distribution curves and the error in the calibration period.



Fig. 7. Observed and simulated flow frequency distribution curves and the error in the validation period.

Table 2.	Performance	Criteria of the	e SVM for	the test catchment.
----------	-------------	-----------------	-----------	---------------------

Period	Nash-Sutcliffe	RMS	e _{Q95%}	e _{medium Q95%} to Q5%
Calibration	0.97	0.08	1.7%	3.9%
Validation	0.99	0.10	7.7%	5.5%

4.3. Comparison of SVM results and SLM results

The SVM was applied to data for the period from January 1, 1993 and December 31, 1999 for the test catchment. Fig. 8 shows the flow frequency distribution curves for the observed and simulated flows and the errors for the SVM and SLM models of *Blanco* et al. 2005.

Fig. 8 shows the improvement of estimates in the range from $Q_{90\%}$ to $Q_{100\%}$ of the SVM in relation to the SLM. Since this range is the one containing the high frequency values, particularly $Q_{90\%}$ and $Q_{95\%}$ broadly used for design and scaling of hydropower, the contribution of this research to the hydrology of catchments in the Amazon is thus demonstrated. It is further verified that the SVM simulates droughts better than the SLM of *Blanco* et al. (2005). In the remaining frequency ranges, the performance of both models was very similar, with the SLM having a slight advantage.

Tab. 3 shows the results for the performance criteria for the application of SVM and SLM to the test catchment data.

T a ble 3. SVM and SLM performance – from 1993 to 1999.

Model	R²	RMS	e _{Q95%}	$e_{ m medium} Q_{95\%}$ to $Q_{5\%}$
SVM	0.97	0.08	1.7%	3.9%
SLM	0.90	0.14	9.4%	3.7%

It is observed again in Tab. 3 that the SVM simulates high frequency flows better than does the SLM, represented by $Q_{95\%}$. Regarding the mean error in the interval $Q_{95\%}$ to $Q_{5\%}$, this was smaller for the SLM than for SVM. This is explained by the fact that the SLM develops its projection around the mean value for the period, whereas the SVM operates considering the weighted daily rainfall function z(t) and soil moisture in the basin u(t) with memory, so that the simulated flows incorporate the effect of the rainfall that occurred during the memory of the previous 180 days.



Fig. 8. Observed and simulated flow frequency distribution curves by the SVM and SLM and errors for the calibration period.

4.4 Sensitivity analysis for the SVM

A sensitivity analysis for the model in relation to the size of the rainfall/runoff samples was carried out. The flow frequency distribution curves simulated with the truncated samples should be similar to the validation curves. The analysis allows for estimating the time required for applying the model to catchments having no flow data. The data used in model calibration were progressively reduced by 1 year, whereas at each reduction, model calibration was conducted and applied to the 6 years (2001-2006) of rainfall data used in the model validation for simulation of the flow frequency curve. Tab. 4 shows the calculated values for the performance criteria according to the extended calibration period. By analyzing the table, it was verified that the quality of the performance criteria that were considered decreases considerably when the data sets are reduced to 1 year. Thus, it was concluded that 1.5 years is the minimum data set needed for applying the model.

T a b l e 4. Performance criteria for sensitivity analysis.

Period	Nash-	RMS	$eQ_{95\%}$	$e_{medium}Q_{95\%}$
(years)	Sutcliffe			to $Q_{5\%}$
6.5	0.99	0.10	7.7%	5.5%
5.5	0.98	0.18	2.6%	5.2%
4.5	0.98	0.18	3.2%	5.0%
3.5	0.99	0.15	4.5%	5.7%
2.5	0.98	0.16	5.3%	5.5%
1.5	0.99	0.13	5.4%	6.8%
1.0	0.47	0.88	23.1%	68.5%
RMS (m ³ /s)				D ²



Fig. 9. System memory optimized for 1.5 years.

In the case of optimizing m for the truncated series, Fig. 9 shows the results for the optimization of m related to the minimum series, that is, 1.5 years, for the period ranging from July of 1998 to December of 1999. For the other truncated series of 1.5

years within the 6.5 years that were analyzed, the results were also equal to m = 180 days.

T a ble 5. w in function of n interval for the sensitivity analysis.

Interval for n	w [%]
i to i-14	8 00
i-15 to i-29	8.00
i-30 to i-44	14.0
i-45 to i-59	14.0
i-60 to i-74	11.0
i-75 to i-89	11.0
i-90 to i-104	8.00
i-105 to i-119	0.00
i-120 to i-134	6.00
i-135 to i-149	0.00
i-150 to i-164	3.00
i-165 to i-179	5.00

Thus, for m = 180, d = 15 and e = 12, the weights w are given in Tab. 5. Fig. 10 compares the simulated flow frequency distribution curves with calibration periods of 6.5 years (maximum) and 1.5 years (minimum). It shows that the maximum percentage errors are lower than 15% for flow rates between $Q_{5\%}$ and $Q_{95\%}$, the average errors being lower than 6.8%. This comparison demonstrates that the model satisfactorily simulates the flow frequency distribution curve for the test catchment except for the flood flows, whether with the 6.5year sample or with that of 1.5 years. This analysis also determined that the flow frequency distribution curves are comparable, whenever the month of July is taken as origin for the period between 1993 and 1999. However, only the results for the period between July 1998 and December 1999 are presented (Fig. 10). In that case, this paper established two periods of drought in order to position the floods in the region, which occur between the months of January and June. As for Blanco et al. (2005), they established two flood periods to position the droughts and thus establish the sample that simulates the entire hydrograph. The difference between these two results stems from the difference between the sources taken by the two models: that of Blanco et al. (2005) takes as its source the month of January; whereas the optimized model of this paper takes the month of July as the source. Nevertheless, from the hydrological viewpoint, the results are similar, since both models require a full hydrological year plus the repetition of a period of flood or drought to establish the minimum sample that is characteristic of the hydrology of the test catchment.



Fig. 10. Simulated flow frequency distribution curves and errors for calibration periods of 6.5 years and 1.5 years.

5. Conclusions

The concept of Sigmoid-Variable Gain Factor Model (SVM) adopts the physically realistic assumption of a gain factor varying in a sigmoid form with the prevailing moisture in the basin soil; it is assumed as the function of the weighted daily rainfalls for the previous 180 days. The result is a simple non-linear model, which satisfactorily represents the rainfall-runoff phenomenon, considering the nonlinearity caused by infiltration and evaporation, making it adequate in terms of the reality of the available data (mostly rainfall data) for catchments in the Amazon.

The results from the application of the model began with the optimization of the system memory, m, which had as objective functions R^2 and RMS, which were also used as performance criteria for the model, resulting in an optimized memory, m == 180 days. Optimization of memory m = 180 days, beginning in July, coincides with the dry season, from July to December in the region, thus strengthening the simulation of high frequency flows.

With the optimization of the system's memory, the data for the catchment could be applied in the SVM, resulting in a satisfactory simulation of the frequency distribuition curve, which showed mean errors of less than 3.9%, when compared with the frequency distribuition curve observed, in the intervals between $Q_{5\%}$ and $Q_{95\%}$ during the calibration period, and less than 5.5% in the validation period.

It can also be observed that the SVM has improved performance in relation to the SLM in the range of from $Q_{90\%}$ to $Q_{100\%}$. This range contains the high flow frequency values, particularly $Q_{90\%}$ to $Q_{95\%}$ widely used for sizing designs of hydroelectric projects. Thus, the contribution of this paper to the hydrology of catchments in the Amazon, as presented here, is demonstrated by the fact that the SVM simulates droughts more adequately than the SLM of *Blanco* et al. (2005), as mentioned herein. In the remaining flow frequency ranges, the performance of both models was very similar, with a slight advantage for the SLM.

In the sensitivity analysis, the performance criteria adopted were the Nash-Sutcliffe coefficient (R^2), the RMS, the errors of $Q_{95\%}$, and the mean errors between $Q_{5\%}$ and $Q_{95\%}$. R^2 was always observed as being greater than 98%, except during the calibration period of 1.0 year, which was considered unsatisfactory. The mean errors are all lower than 6.8%, except during the calibration period of 1.0 year, which was also considered unsatisfactory. Therefore, 1.5 years is the minimum time for applying the model to Amazon catchments lacking data, which are located in hard-to-access areas, making it important to estimate that time in planning the work site. The application of SVM in the test catchment under consideration shows that this approach is promising, as it presents results for the simulated flow frequency distribuition curve significantly close to the observed flow frequency distribuition curve, thus providing realistic results for the sizing designs of hydropower production using Small Hydroelectric Power Plants in the Amazon.

List of symbols

- dt time derived,
- $d\tau$ time derived,
- *e* exponential function,
- e subintervals j of d days,
- e(t) output error in function of time,
- e% percentage errors,
- $e_{\text{medium} Q5\% to Q95\%}$ average of percentage errors in the interval between the flow rates from $Q_{95\%}$ to $Q_{5\%}$ frequency,
- G gain factor,
- Go observed gain factor,
- Gs simulated gain factor,
- $h(\tau)$ normalized unit impulse response function,
- Fo observed effective flow rate/weighted rain factor,
- Fs estimated effective flow rate/weighted rain factor,
- m system memory [day],
- n total number of observations,
- Q_b basic flow rate [m³ s⁻¹],
- Q_o observed flow rate [m³ s⁻¹],
- Q_s simulated flow rate [m³ s⁻¹],
- Q_o average observed flow [m³ s⁻¹],
- $Q_{5\%} 5\%$ frequency flow [m³ s⁻¹],
- $Q_{10\%} 10\%$ frequency flow [m³ s⁻¹],
- $Q_{95\%} 95\%$ frequency flow [m³ s⁻¹],
- $Q_{98\%}$ 98% frequency flow [m³ s⁻¹],
- \widetilde{RMS} root mean square [m³ s⁻¹],
- R^2 Nash Sutcliffe coefficient,
- SLM Simpe Linear Model,
- SVM- Sigmoid-Variable Gain Factor Model,
- T calibration period [day],
- t = time [day],
- $U(\tau)$ unitary impulse response function,
- u(t) soil moisture function,
- x(t) basin daily rain [mm day⁻¹],

- y(t) effective flow rate due to precipitation at the mouth of the basin [m³ s⁻¹],
- w_j weight of weighted rainfall and soil moisture functions,
- z(t) weighted rain function [mm day⁻¹],

 τ – time [day].

REFERENCES

- AHSAN M., O'CONNOR K.M., 1994: A simple non-linear rainfall-runoff model with a variable gain factor. J. Hydrol., 155, 151–183.
- BLANCO C.J.C., SECRETAN Y., FAVRE A.C., SLIVITZKY M., 2005: Raifall-runoff model for simulation of the flow duration curves of small non gauged basins in Amazonia. (In French with abstract in English.) Canad. J. Civ. Eng., 32, 803–811.
- CLARK C.O., 1945: Storage and the unit hydrograph. Trans ASCE, 110, 1419–1446.
- JAKEMAN A.J., HORNBERGER G.M., 1993: How much complexity is warranted in a rainfall-runoff model. Wat. Resour. Res., 29, 8, 2637–2649.
- JUN X., 2001: A system approach of real-time hydrological forecasting applied to a catchment basin. International Workshop on non-structural measures for water management problems, London, Canada, October 2001.
- KACHROO R.K., 1992: River flow forecasting. Part 1. A discussion of the principles. J. Hydrol., 133, 1–15.
- KACHROO R.K., LIANG G.C., 1992: River flow forecasting. Part 2. A discussion of the principles. J. Hydrol., 133, 17–40.
- KACHROO R.K., NATALE L., 1992: Non-linear modelling of the rainfall-runoff transformation. J. Hydrol., 135, 341–369.
- LABAT R.A., MANGIN A., 2000: Rainfall-runoff relations for karstic springs. Part I: convolution and spectral analyses. J. Hydrol., 238, 123–148.
- RIBEIRO NETO A. et al., 2008: Simulation in the Amazon basin using limited data: the Madeira River. (In Portuguese with abstract in English.) Rev. Bras. Rec. Hídr., *13*, 3, 47–58.
- SHERMAN L.K., 1932: Stream flow from rainfall by the unit hydrograph method. Engng. News Rec., 108, 501–505.

Received 25 August 2010 Accepted 12 April 2011