Analysis of the worth of rainfall-runoff information using statistical entropy concepts

Aymen BEN JAAFAR [1], Zoubeida BARGAOUI [1]

[1] *Université Tunis El Manar, ENIT, Laboratoire de Modélisation en Hydraulique et Environnement ; BP 37, 1002, Tunis. Tunisia*

Abstract

Rainfall and runoff (R-R) databases are among the most commonly used by hydrologists for assessing the rainfall runoff transformation and estimating evapotranspiration and principal water flows (percolation, infiltration) at the watershed scale and daily to annual time scales. Recently, the concept of statistical entropy was proposed for hydrologic data compression. Following this idea, R-R database of two moderate size watersheds from Northern Tunisia (Sejnane 376 km², Joumine 418 Km^2 nowadays controlled both by dams) is analyzed using some entropy related concepts. The general objective is to integrate the transfer of information between both series in rainfall runoff model calibration and to perceive its impact on model calibration and validation performances. The methodology is based on estimating probability density of daily rainfall and mean daily discharge using the univariate and bivariate non parametric kernel method to quantify entropy and information gain. The integration is performed at the yearly scale. The interpretation of results aims to detect periods (a number of consecutive years) of observation characterized by significant information gain. The worth of choosing such periods as calibration periods is examined. For these basins, it is found that the period with the most information gain results in the best Nash coefficients in calibrating runoff at monthly and decadal scales. With respect to model validation and when calibrating the model using other periods of the same length, model performance in validation remains good for the periods of maximum information gain.

Keywords : rainfall, runoff, statistical entropy, information gain, Nash coefficients, model calibration, model validation, BBH.

1. Introduction

The problem of hydrological models abilities to be transferred from period to period has recently gained interest, as demonstrated by the new IAHS Scientific Decade: “Panta Rhei” (Montanari et al., 2013). There is a kind of agreement that a rainfall-runoff (RR) model calibrated on a given period will generally not be able to simulate flows with a similar efficiency on another period (Merz et al., 2011; Coron et al., 2012). This lack robustness represents a serious problem for model application in practical conditions.

Model efficiency is often measured through root-mean-square error and NS efficiency (Nash and Sutcliffe, 1970). These criteria have the advantage of reflecting the model efficiency on all simulated time steps and can even be used to build model robustness criteria, as discussed by Coron et al. (2012).

The objective of this study is to integrate the transfer of information between series of rainfall and runoff in model calibration in order to analyze model robustness hydrologic conditions. The methodology is presented in the next section. There, the entropy concept of information gain is briefly presented as well as calibration criteria and the rainfall runoff model adopted in the study. The subsequent section presents rainfall and runoff daily series for the study basins. Then it is focused on the results in terms of information gain findings, as well as model calibration and validation performances. Finally the conclusion is drawn about the worth of using maximum information gain periods for model calibration and validation.

2. Methodology

A water balance model operating at daily resolution is considered to infer the rainfall-runoff relationships.

Presentation of the BBH model

The conceptual water balance model BBH (Bucket Bottom Hole) is a lumped model, with a daily time scale, developed by Kobayashi et al. in 2001. The BBH is a water balance model which involves seven parameters: a (parameter related to field capacity), b (humidity drying soil parameter), c (the potential capillary rise), D (root soil depth), p (soil porosity), η (moisture retention capacity) and σ (the resistance of the vegetation cover to evapotranspiration). A reparametrization of this model has been undertaken by (Bargaoui and Houcine (2010)) so as to transform parameters a, b, and c which were initially estimated using rainfall-runoff data during model calibration process, in order to adopt pedo transfer parameters: Ks (the conductivity at saturation of the ground), Sfc (the capacity of the ground field) and B (a shape parameter representative of soil retention curve) instead of calibrated parameters. Thus, it is proposed to estimate model parameters Ks, Sfc, p and B using basin soil texture information. To that purpose, an average spatial value is adopted and weights are assumed as the percent of the basin surface occupied by a given soil texture class. Pedo transfer data are from Abid (2015). The parameters (D, η, σ) are estimated by model calibration using calibration criteria.

Calibration model criteria

In the next, calibration of (D, η, σ) is achieved by considering three hypothetic and potential values for D generally assumed in literature (D=300 mm, D=500 mm; D=1000 mm) and by screening the interval of variation of ($0< \eta <1$; $0< \sigma <1$) with an increment $\Delta\eta=\Delta\sigma=0.01$. The mean absolute relative error (AARE) which is a criterion associated to the water balance is considered to solve the parameters selection. Solutions that verify AARE less than a fixed threshold equal to 0.2 are selected. Other solutions are eliminated.

Then, the Nash-Sutcliffe coefficient for two time scale resolutions is considered: Monthly ($Nash_1$) and by decade ($Nash_2$) are adopted to rank solutions obtained using AARE. The set presenting the best values for $Nash_1$ and $Nash_2$ is retained. To select the calibration period the concept of mutual information is then introduced.

Mutual information or information gain

The mutual information is "a measure of the variables' mutual dependence" and is defined as follows (Krstanovic and Singh, 1992):

$$I_m(X,Y) = H(X) + H(Y) - H(X,Y) \tag{1}$$

with:

I_m (X, Y): the mutual information associated with two variables X and Y

H (X, Y) the joint entropy associated with two variables X and Y

H (X): the entropy associated with variable X

H (Y): the entropy associated with variable Y

In the present, X is daily rainfall, Y is daily runoff and mutual information $I_{m,k}$ (X, Y) is estimated for each year *k*.

Figure 1 illustrates the link between the various concepts related to the entropy.

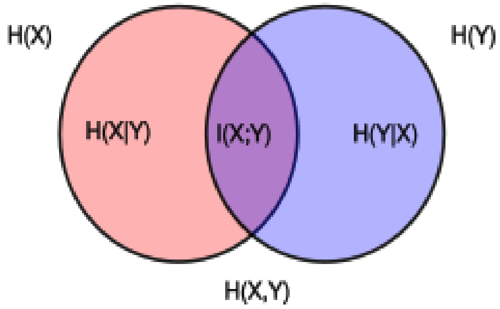

Figure 1. Link between the various concepts related to the entropy

(http://fr.wikipedia.org/wiki/Information mutuelle)

We identify at a yearly basis years where the information that X and Y share is high are selected. Also, the analysis seeks to identify the sub period (a succession of years) where the gain is at its maximum in comparison to other sub periods and to adopt them as calibration periods. Such sub periods will be assigned as “Type A periods”. Non parametric marginal density distributions (Silverman, 1986) are estimated for both daily rainfall series and for the series of daily mean flow as well as non parametric joint probability distribution as presented in Yang and Burn (1994). Then, mutual information (information gain) is investigated

Model performance assessment

A comparison between calibration using type A period to the other sub-periods (assigned as type B) is achieved; To that end, continuous periods having the same size (length) as the highest gain period are selected. Using both the monthly Nash criteria ($Nash_1$) and the decadal Nash ($Nash_2$), this comparison is used to examine whether the period of type A (with strongest gain) gives up to the best performance levels in model calibration. For model validation, best solutions obtained from the various calibration periods (either of type A or B) are applied to the remaining sub periods in order to assess the worth of using the strongest gain of information period as calibration period.

3. **Results and conclusions**

The evaluation of information content for series of mean daily rainfall and runoff was performed for two catchments of 376 Km^2 (Sejnane Déversoir) and 418 Km^2 (Joumine Arima) situated Northern of Tunisia. Sejnane Dévesoir has a mean annual rainfall equal to 2.36 mm/day and Joumine Arima 1.82 mm/day.

Using an observation period of 15 years (1961-1975) for the watershed of Sejnane Déversoir and also 15 years (1967-1981) for the watershed of Joumine Arima, the evaluation of information gain is achieved for an annual time scale, adopting hydrological years (first day of each year: September the first). It results that the periods between the hydrological year 1972-1973 and the hydrological year 1975-1976 for the Sejnane watershed and from the hydrological year 1967-1968 to the hydrological year 1971-1972 for the Joumine watershed bring the most information gain between rainfall and runoff data (period Type A). While the basins are neighbors, the periods of maximum information gain are not the same.

Table 1 and 2 illustrate the best Nash obtained from calibration respectively for Sejnane and Joumine basins.

Table 1. The best $Nash_1$ and $Nash_2$ obtained in calibration (Sejnane basin)

D (mm)	**300**				**500**				**1000**			
Type of period	A	B1S: 1961-1964	B2S: 1965-1968	B3S: 1969-1972	A	B1S	B2S	B3S	A	B1S	B2S	B3S
$Nash_1$	0.77	0.87	0.89	0.82	0.84	0.83	0.89	0.89	0.9	0.76	0.81	0.81
$Nash_2$	0.6	0.78	0.76	0.65	0.79	0.77	0.75	0.79	0.76	0.71	0.77	0.74

Table 2. The best $Nash_1$ and $Nash_2$ obtained in calibration (Joumine basin)

D (mm)	**300**			**500**			**1000**		
Type of period	A	B1J: 1972-1976	B2J: 1977-1981	A	B1J	B2J	A	B1J	B2J
$Nash_1$	0.62	0.64	0.28	0.66	0.4	0.66	0.8	0.44	0.78
$Nash_2$	0.52	0.25	0.32	0.56	0.37	0.28	0.72	0.41	0.51

From Table 1 and Table 2, it is seen that for both basins it is found that the best results with respect to Nash coefficients for monthly (Nash1) and decadal (Nash2) scales are obtained for the period of type A when compared to the other calibration periods with the same duration (type B).

Table 3 and 4 illustrate the best efficiency criteria $Nash_1$ and $Nash_2$ obtained from transferring the model (all the solutions fulfilling the AARE > 0.2 criterion are here tested) to other periods.

Table 3. The best $Nash_{1val}$ and $Nash_{2val}$ obtained from validating all the solutions calibrated from every sub-period on the other three periods for Sejnane basin

Type	A calib.		B1S calib.		B2S calib.		B3S calib.	
A valid.			0.9	0.78	0.9	0.78	0.9	0.78
B1S Valid.	0.71	0.55			0.76	0.67	0.76	0.66
B2S valid.	0.81	0.75	0.82	0.77			0.82	0.77
B3S valid.	0.77	0.64	0.82	0.77	0.82	0.75		

Table 4. The best $Nash_1$ and $Nash_2$ obtained from validating all the solutions calibrated from every sub-period on the other two periods for Joumine basin

Type	A calib.		B1J calib.		B2J calib.	
A valid.			0.77	0.65	0.79	0.7
B1J valid.	0.4	0.37			0.43	0.41
B2J valid.	0.66	0.26	0.61	0.1		

Thus, when using the set of parameters calibrated on the basis of type B period, it is found that model efficiency remains good for the period of maximum information gain (type A) for both studied basins while it deteriorates for the other periods. So, this experiment reflects the worth of using maximum information gain periods for water balance model calibration. When the calibration period is selected as the period presenting the maximum information gain, Nash model calibration criteria was found higher and model validation was ensured. While such a result might be intuitive, conclusions need to be checked using many other basins before becoming effective.

Acknowledgements: Part of this research was funded by the PEER project (Partnerships for Enhanced Engagement in Research): Contribution to drought identification and alert in Northern Tunisia, 2015.

References

[1] Abid,N., Estimation of actual and potential evapotranspiration using remote sensing in northern Tunisia: 26th IUGC Genearal Assembly 2015 International Union Geodesy and Geophysics, June 22-July 2, Prague, Czech Republic, 2015.

[2] Bargaoui, Z., Houcine. A., 2010. Sensitivity to calibration data of simulated soil moisture related drought indices. Revue Sécheresse, Vol. 21(4). 1-7.

[3] Coron, L., Andréassian, V., Perrin, C., Lerat, J., Vaze, J., Bourqui, M., and Hendrickx, F.: Crash testing hydrological models in contrasted climate conditions: An experiment on 216 Australian catchments, Water Resources Research, 48, W05552, doi:10.1029/2011WR011721, 2012.

[4] Kobayashi, T., Matsuda, S., Nagai, H. and Tesima, J., 2001: A bucket with a bottom hole (BBH) model of soil hydrology, 5 pages. IAHS Publ.no. 270.

[5] Krstanovic, P. F. and Singh, V. P., (1992). Evaluation of rainfall networks using entropy, Water Ressources Management, 6: 279-293.

[6] Merz, R., Parajka, J., and Blöschl, G.: Time stability of catchment model parame ters - implications for climate impact analyses, Water Resources Research, 47, W02531, doi:10.1029/2010WR009505, 2011.

[7] Montanari, A., Young, G., Savenije, H. H. G., Hughes, D., Wagener, T., Ren, L. L., Koutsoyiannis, D., Cudennec, C., Toth, E., Grimaldi, S., Bl¨oschl, G., Sivapalan, M., Beven, K., Gupta, H., Hipsey, M., Schaefli, B., Arheimer, B., Boegh, E., Schymanski, S. J., Di Baldassarre, G., Yu, B., Hubert, P., Huang, Y., Schumann, A., Post, D. A., Srinivasan, V., Harman, C., Thompson, S., Rogger, M., Viglione, A., McMillan, 5 H., Characklis, G., Pang, Z., and Belyaev, V.: “Panta Rhei—Everything Flows”: Change in hydrology and society—The IAHS Scientific Decade 2013–2022,

Hydrological Sciences Journal, 58, 1256–1275, doi:10.1080/02626667.2013.809088, 2013.

[8] Nash, J. and Sutcliffe, J.: River flow forecasting through conceptual models part I. A discussion of principles, Journal of Hydrology, 10, 282–290, doi:10.1016/0022 1694(70)90255-6, 1970.

[9] Silverman, B. W. (1986). Density Estimation for Statistics and Data Analysis, Chapman &Hall, London.

[10] Yang,Y. and Burn,D.H., 1994. An entropy approach to data collection network design, Journal of Hydrology 157,307-324.