EVALUATION OF UNCERTAINTIES OF RAINFALL-RUNOFF MODELS: TIME SCALE ISSUES

Gerald Souza da Silva; Alain Marie Bernard Passerat de Silans; Cristiano das Neves Almeida; Ana Cristina Souza da Silva

Federal University of Paraiba – Center of Technology – Water Resources and Environmental Engineering Laboratory

Author e-mail address: gerald-silva@gmx.de

ABSTRACT

The timescale of a rainfall-runoff-model is an important issue for the results that are achieved by the model. Rainfalls can be very intense and occur in a few hours or can last longer with the same overall quantity. Intense rainfalls are regularly flooding critical areas. This paper studies the relation between timescale and runoff simulated by lumped rainfall-runoff-models. The study is held in the semi-arid region in north-east Brazil. The water levels of monitored dams are used to forecast the runoff with observed rainfall data. Especially in the semi-arid region rainfalls that fill up these dams will occur concentrated in some days or weeks. Simulations on a monthly timescale are often used because of the absence of long-term continuous daily runoff data and safer results. However, how exact are these simulation to provide data for the several uses like water supply and the sizing of dams and the risks related.

Key Words: Rainfall-runoff model, Time scale

Introduction

For hydrological modeling purposes its necessary to decide in the beginning what kind of model will be applied for the simulation of hydrologic processes. Mostly common are deterministic and stochastic models, on a lumped or distributed special basis (CHOW, 1964). Also the temporal basis of the model has a significant impact for the type of model and the outcome of the simulation. In many cases it is not necessary to use very complex models to represent processes in nature, simple models mostly are sufficient for the simulation. However, for hydrological models are used timescales from seconds to month or years. Smaller time steps will probably give more exact results for processes in the nature on a smaller spatial scale, like runoff and sedimentation from experimental watersheds. But on the other hand smaller time steps will make the data collection more complex, more expensive and will require a more intensive work. Subsequently the processing time for analysis, for error correction and for post processing will be increased. Another circumstance is the availability of input data, for example, in northeast Brazil large distances and a lack of infrastructure make observation on a detailed spatial and temporal scale difficult. Most gauges are manually and provide only data on a larger timescale, normally 24 hour events. The semi-arid region in northeast Brazil is also defined by an extensive dry season and has only some months with rainfall which varies strongly each year. The rainfall events have special characteristics, it can rain very intense a few days which causes flood runoff (with the problems related, like pounding) or it can nearly not rain at all. This study aims to validate the temporal influence of hydrological modeling in the semi-arid region in northeast Brazil. The study is part of a research project named DISPAB (Metodologias para definição da disponibilidade hídrica em pequenos açudes e pequenas bacias hidrográficas da região semi-árida do Brasil), that is focused on the available water resources in small watersheds and small reservoirs in the semi-arid region in northeast Brazil.

Methods

To evaluate the differences caused by timescale a comparison of observed runoff data was studied and subsequently the influence to the lumped rainfall-runoff model SMAP (LOPES, BRAGA e CONEJO, 1981) was examined. The nonlinearity between rainfall and runoff is known and still aim of researches (WEN WANG, 2006). The effect of timescale can also be expected widely non-linear and depends upon various factors, like the spatial location and characteristics of the watersheds.

The inflow rates of monitored reservoirs in the semi-arid region were used to investigate time scale issues of the rainfall-runoff-model SMAP. There are two version of this model, a daily version and a monthly version. However, because of the different structures and the influence of the calibration process on the results a direct comparison is not possible. Therefore was used mainly the daily SMAP model with daily input data and output data on different timescales with the same model calibration. The main effect of using different timescales is the smoothing of mostly irregular time series data, therefore also different smoothing functions will be compared. To evaluate the influence basic statistics were used. Nonlinear statistics or explanatory statistics on time series weren't used, because non linear effects of the rainfall runoff model are not aim of this study.

Study Area

The study area is located in the semi-arid region in the state of Paraiba in northeast Brazil (Fig. 1). It covers an area of 56.439 km^2 and comprises a large number of reservoirs and dams, some of which are monitored by state water agency. The water levels from these reservoirs are used to calibrate a rainfall-runoff model. The used reservoirs have very different sizes and observations periods (see Table 1).

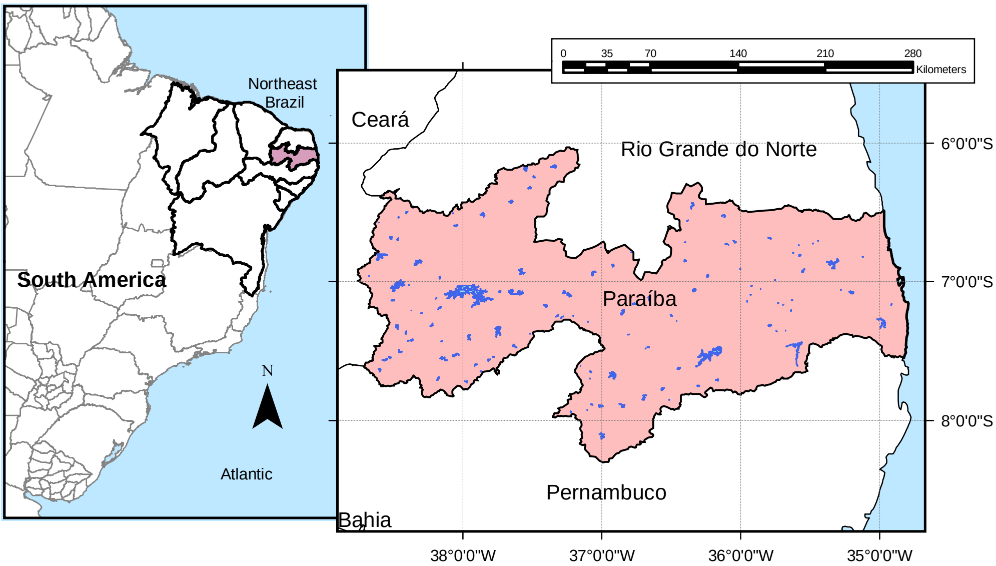

Fig. 1 - Study area: northeast Brazil – State Paraiba – Monitored Reservoirs

Capacity Classes(m³)	N° of reservoirs	Time Series
< 1.000.000	20	Varies from 1994 to 2009
de 1.000.000 a 5.000.000	36	Varies from 1994 to 2009
de 5.000.000 a 20.000.000	34	Varies from 1994 to 2009
de 20.000.000 a 100.000.000	26	Varies from 1994 to 2009
>> 100.000.000	5	Varies from 1994 to 2009

Table 1 – Capacity of the dams monitored by AESA

The hydrological model – SMAP daily

The chosen hydrological model was SMAP (Soil Moisture Accounting Procedure), the model was developed by Lopes *et al.* (1981). It is a deterministic, lumped, hydrological model used for rainfall–runoff transformation. The model schema is shown in Fig. 2. In the study region there is no base flow and the model had to been adapted for this study to run without underground reservoir (Rsub).

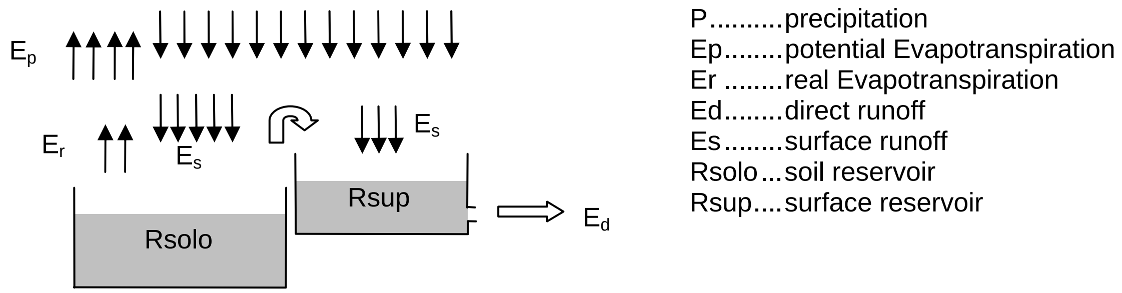

Fig. 2 Model SMAP modified.

The model tries to represent the rainfall runoff relation with only two tanks. One tank represents the capacity of the soil to store water. Only after his saturation additional rainfall will cause additional surface runoff. Soil depth and type will mostly control this tank. Another tank represents the recession in the watershed for the surface runoff to the outlet of the watershed, vegetation and watershed characteristics will influence this tank. Physical processes related to rainfall events can be straightforwardly explained by the model.

The input data are the average precipitation and potential evapotranspiration in the watershed. The model parameters are: STR, saturation capacity of the soil (Rsolo) 100–2000 (mm); K2, recession constant of runoff (Rsup) 0.2 to 0.5 (days); Ai, initial abstraction, from 2.5 to 5.1 (mm); CAPC, field capacity, 30–50 (%). The parameters “Ai” and “CAPC” represent the characteristics of the vegetation and the soil type. Typical values for “Ai” are 2.5 mm for fields, 3.7 mm for forests, and 5.0 mm for dense forests. Typical values for “CAPC” are: 30% for sandy soil, 40% for mixed soil, and 50% for clay soil.

Also noteworthy is that the monthly model would have only one reservoir (Rsolo) to simulate the rainfall runoff processes for the semi-arid region in northeast Brazil. This rainfall-runoff model is an even further abstraction for hydrological processes in nature and has no more a direct physical basis.

Objective Function (OF)

The objective function has the aim (objective) to minimize or maximize the relationship (function) between the observed and calculated data. The assessment, if the model is calibrated, is complex, the proper selection of an objective function has significant influence on the results, there are objective functions that prioritize low runoffs, high runoffs, or none of these. The used objective function for the analysis prioritized runoffs of greater magnitude (peak flow), but not too extreme far from the average, with 50% weight on the daily results and 50% on the monthly results. Equation 1 shows the selected objective function. All watersheds were calibrated automatically with genetic algorithms using this OF. The Java-package x2ga (SOARES JUNIOR, MOTTA, *et al.*, 2009)was developed and integrated to the rainfall-runoff model for this purpose.

$$OF = 50.\sum_{i=1}^{n}(Q_{obs/daily} - Q_{calc/daily})^{2} / \sum_{i=1}^{n}(Q_{obs/daily} - \overline{Q}_{obs/daily})^{2} + 50.\sum_{i=1}^{n}(Q_{obs/monthly} - Q_{calc/monthly})^{2} / \sum_{i=1}^{n}(Q_{obs/monthly} - \overline{Q}_{obs/monthly})^{2} \quad (1)$$

Results

In the semi-arid region the distribution and the intensity of rainfall events varies strongly every year. Fig. 3 shows for example the average observed runoffs to the Coremas Reservoir and the standard deviation (SD) from the last 16 years (1994-2010). It can be noticed that the SD is very high and varies with multiple factors from the monthly average.

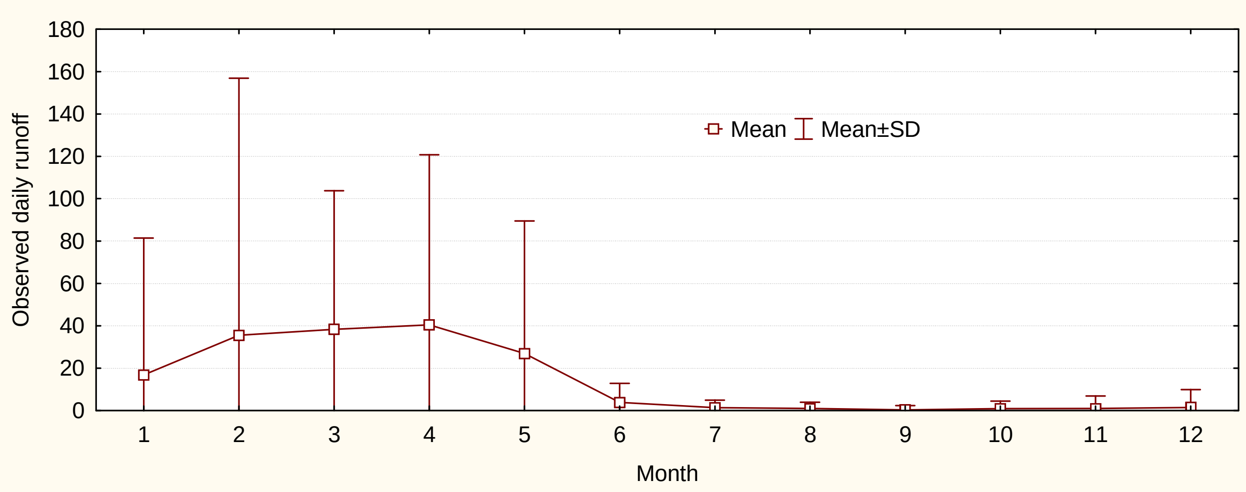

Fig. 3 - Mean Plot with Standard Deviation of observed daily runoff [m³/s] grouped by months

The advantage of a model which uses daily data for input instead of monthly is the higher information density for the calibration process. For example, a time series with a range from 20 years will have on monthly basis 240 registrations and on daily basis 7300 registrations. This issue plays an important role in the calibration process, a daily model has 30 times more values to calibrate the model. For example, in the semi arid-region the rain-season is only 3-4 months, which results small amount of rainfall-runoff data to calibrate. But within a month various rainfall-runoff events can be observed. Especially for watersheds with a shorter monitored time series data on daily scale can be essentially for an adequate calibration.

The next figures show a closer look on some events in the time series, which are interesting for time scale issues in the modeling process.

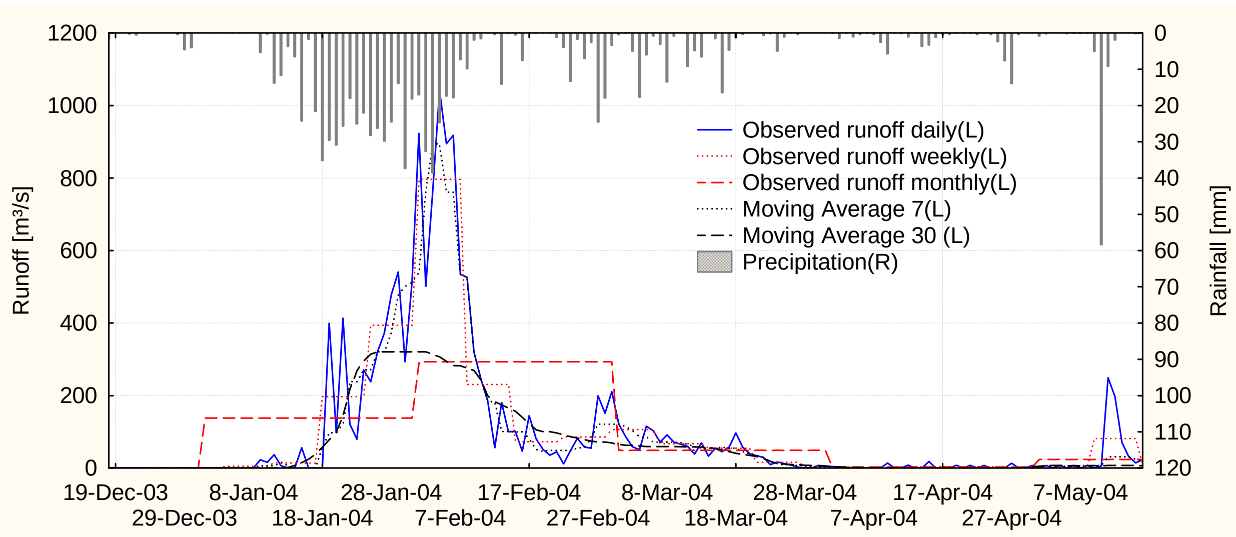

Fig. 4 – Hydrograph for the Coremas reservoir

Fig. 4 shows the hydrograph from the rain-season in 2004 with the daily observed inflow to the Coremas reservoir. The watershed has an drainage area of 6426 km². In this case the hydrograph illustrates the differences between the moving average using 30 days versus the monthly average from the Gregorian calendar and also the moving average using 7 days versus the weekly average of the Gregorian calendar. A significant difference can be seen in the monthly average, because of the rain period occurred in some weeks from the end of January to the beginning February. The rest of the months no intense rain or inflow was observed. In the semi-arid region concentrated rain periods (peak runoff) mostly happen in one or two weeks. If this peak runoff occurs between two months, many combinations between rainfall and runoff are possible. In this relative large watershed a rainfall runoff model on monthly basis could not be calibrated with this year, because the rainfall occurred mostly in January and the runoff in February, certainly other events in the time series will compensate such distributions, but more events like this will have significant influence in the calibration process. However, the average on a weekly basis seems always accompanying the 7 day moving average and can represent better (mathematically and graphically) the rainfall-runoff events in the semi-arid region.

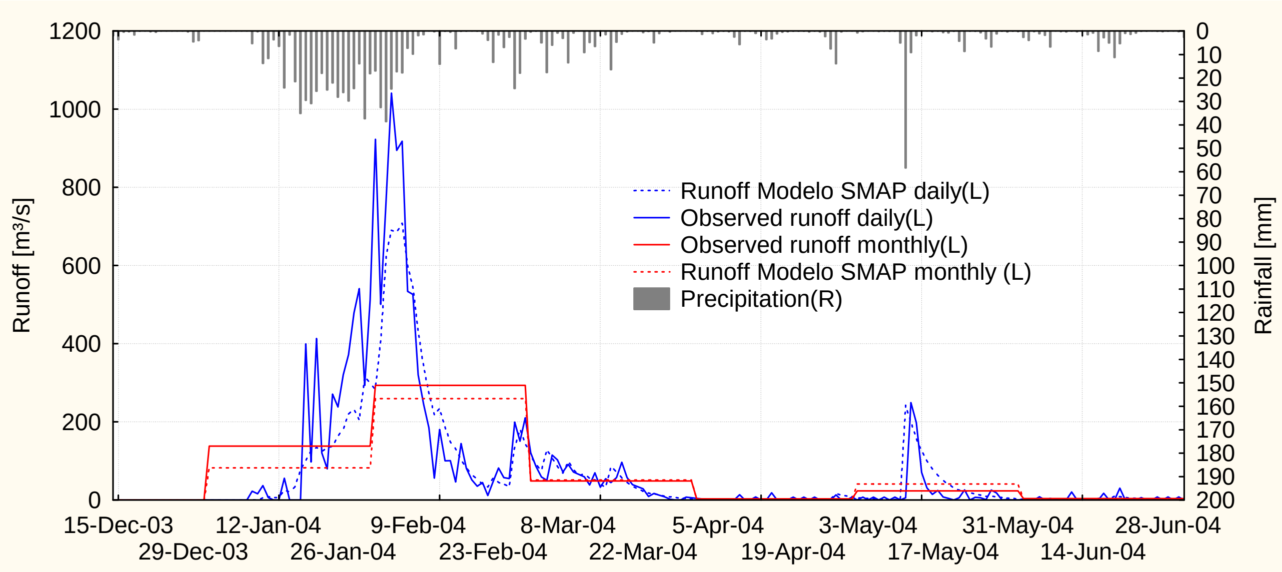

Fig. 5 – Hydrograph - Coremas reservoir – observed values vs. calculated values

The results for the calibrated model SMAP are shown in Fig. 5. The hydrograph shows the same rainfall event in year 2004 with the observed and calculated runoff on monthly and daily basis. The model was calibrated with a 16 year time series. In January, after the dry season, the monthly and daily results are showing significant differences. Mostly the monthly results differ, probably caused by different soil saturations in the watershed. The smoothing effect of the monthly average is very strong and doesn't seem related to rainfall events. Of course that the overall quantity of rainfall and runoff are the same, but the characteristics of runoff related to real processes in nature are lost. Also has to mentioned, that smaller differences on monthly scale, have a higher impact on the total outcome of the result than differences on daily basis, trough the factor 30. Besides on daily basis, the peak runoffs are mostly not simulated correctly. It seems the daily model don't respond to events which are highly concentrated in some days of the month, primarily because of the selected OF which has already a smoothing effect on the data. In average, the model performed well, with Nash–Sutcliffe model efficiency coefficients (NASH J. E., 1970) EC_{daily} = 0,8 and $EC_{monthly}$ = 0,94, but considerable differences between modeled and observed flows occur during periods of high peak flow.

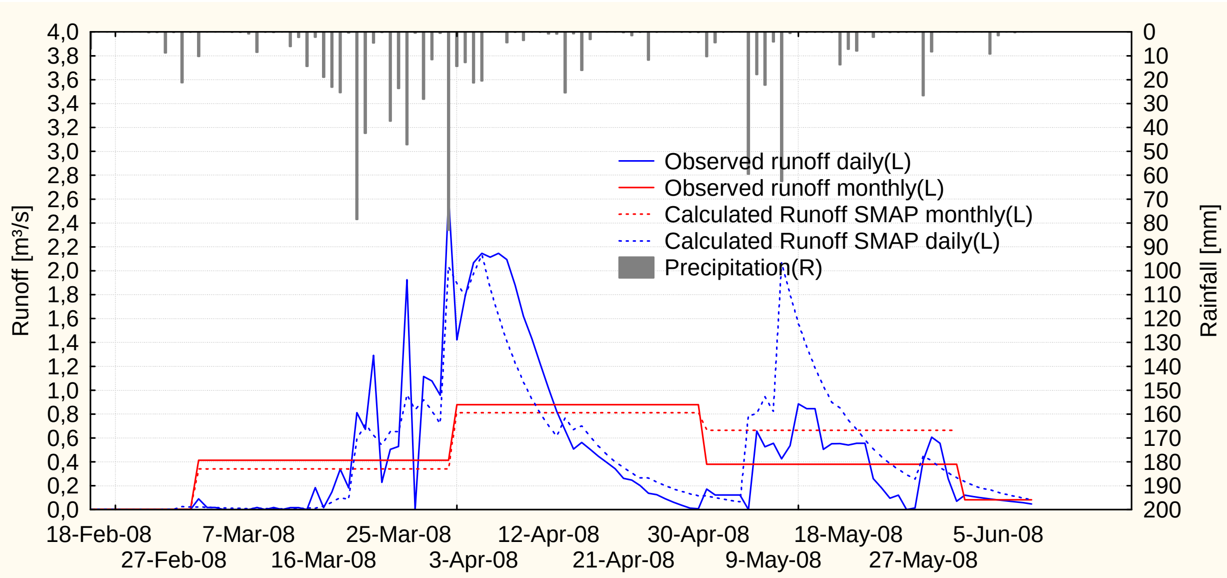

Fig. 6 – Hydrograph - Albino reservoir – observed values vs. calculated values

In comparison Fig. 6 shows a very small reservoir which captures a small sub watershed (17 km²) within the Coremas watershed. The hydrograph in year 2008 shows a similar behavior on smaller spatial scale. Rainfall runoff events in the semi-arid region are nearly always concentrated in a timescale inferior the monthly; this is in particular fact for small watersheds. Here too, differences on flood runoff are always present. This can be related to uncertainties in the input data, caused by incorrect rainfall distribution in the watershed, observation errors, dam infiltration or dam flooding. The highest uncertainties are associated with data from the most intense rain periods. It's nearly impossible to correctly simulate the peaks of the flood runoff, because of the lack of exact data. However overall, with Nash–Sutcliffe model efficiency coefficients EC_{daily} = 0,72 and $EC_{monthly}$ = 0,86 the model was able simulate the runoff satisfactorily.

For all simulated watersheds (69) the achieved daily Nash–Sutcliffe model efficiency coefficients were at maximum (Max) = 0.8 and at minimum (Min) = 0.16. On monthly basis the Nash–Sutcliffe model efficiency coefficients were Max = 0,95 and Min = 0,27. The average Nash–Sutcliffe model efficiency coefficient from 69 reservoirs were EC_{daily}=0.37 to $EC_{monthly}$=0.62. It can be seen that for a monthly basis, better results were found.

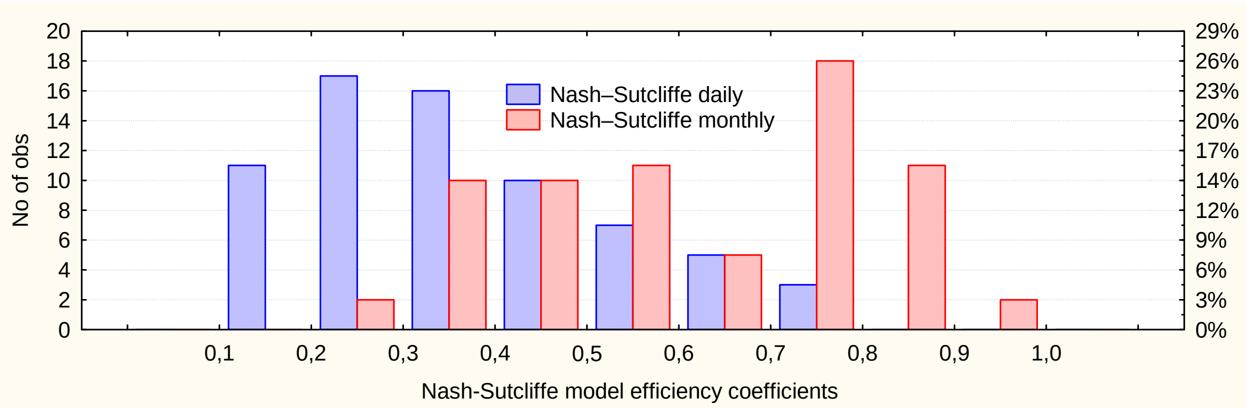

Fig. 7 - Histogram Nash–Sutcliffe model efficiency coefficients – daily and monthly

The model efficiency of all reservoirs shows nearly a Gaussian distribution related to data and model uncertainties (see Fig. 7). Almost all model efficiencies are improved, resulting in better “computational” results on the smoothed monthly data. The results showed the model efficiency is improving from daily to monthly results (see Fig. 8) with a constant value around 0.3.

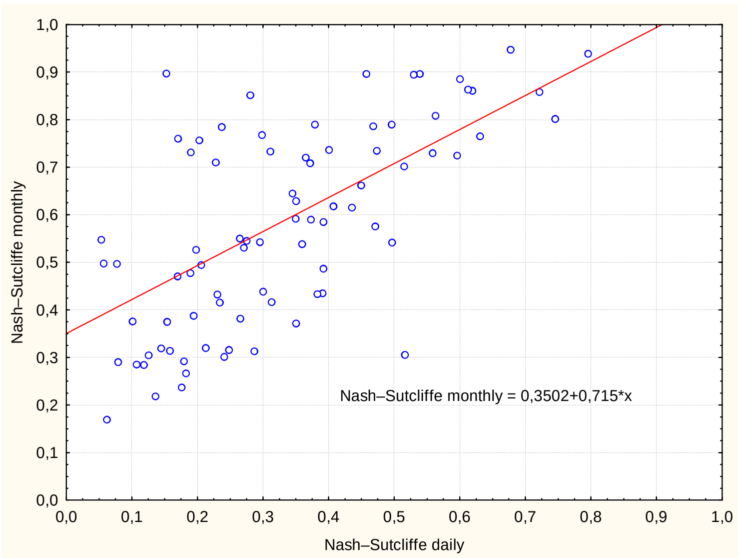

Fig. 8 - Scatterplot of Nash–Sutcliffe monthly against Nash–Sutcliffe daily

There is also a correlation between the watershed drainage area and the model performance, the average for smaller watersheds were CE_{daily}=0.32 and $CE_{monthly}$=0.57, the average for larger watersheds were CE_{daily}=0.45 and $CE_{monthly}$=0.67. Mostly very small watersheds performed not very well, because of the higher uncertainties in the input data related to the spatial scale, as the spatial distribution of rainfall.

Conclusion

Simulation showed that the model with input data on a daily base and output on a monthly base principally could not represent flood runoffs. This fact is derived from the circumstance that in the semi-arid region in north-east Brazil mostly highly concentrated rainfall events and related flood events occur in only few days in a month. This concentration is spatial as well as temporal. The use of monthly data will give a better estimation of the overall water availability in a watershed, for example, for water supply and management. However, flood runoffs can decrease the stored water availability, because these high peak runoffs mostly exceed the capacity of reservoirs. On monthly scale these peak runoffs are strongly smoothed and distributed within a month or two. Many observations showed continuous rainfall-runoff events passing two months. A model on monthly basis will have difficulties to represent such data because of the separation of rainfall and runoff. However monthly performances will be mostly acceptable, because relative differences are smaller, although summed up can result in significant differences. With many impropriate events (for example, events separated by months) on monthly timescale the overall model performance will suffer. Daily data has many errors and uncertainties. Smoothing the time series is recommendable, but the monthly timescale seems too high and inappropriate. Physically, the runoff depends largely on the rainfall intensity. The monthly model does not consider physical processes, only the total amount of rainfall and the “stored” water volume in the watershed are related to the watershed runoff. Normally models in civil engineering try to represent the nature; the monthly abstraction doesn’t seem adequate. However, for an estimation of the water availability for watershed the monthly model can be sufficient, but it can’t be used for other purposes, like sizing of reservoirs.

Time series uncertainties are also related to the spatial issues. There are countless types of watersheds and also many different rainfall distributions. Every watershed has its own characteristics, the time series are affected by observation errors, climatic data and uncertainties of the lumped model. These effects are in interaction with time scale related issues and are difficult to separate.

References

CHOW, V. T. **Handbook of applied hydrology**. New York: McGraw-Hill Book, 1964.

LOPES, J. E. G.; BRAGA, B. P. F.; CONEJO, J. G. L. **Hydrologic simulation:** application of a simplified model. Symposium of Water Resources. Fortaleza: Water Resources, Brazil. 1981. p. 42-62.

NASH J. E., S. J. V. River flow forecasting through conceptual models part I — A discussion of principles. **Journal of Hydrology**, v. 10 (3), p. 282–290, 1970.

SOARES JUNIOR, A. et al. **An XML-based Genetic Algorithm Tool:** System Features and Modeling. XXX Congresso Ibero-Latino-Americano de Métodos Computacionais em Engenharia. [S.l.]: [s.n.]. 2009.

WEN WANG, J. K. V. P. H. A. J. M. V. G. J. M. Testing for nonlinearity of streamflow processes. **Journal of Hydrology 322**, p. 247–268, 2006.