

<https://doi.org/10.1590/2318-0331.241920180031>

Data assimilation using the ensemble Kalman filter in a distributed hydrological model on the Tocantins River, Brasil

Assimilação de dados por filtro de Kalman por conjunto em um modelo hidrológico distribuído na bacia do rio Tocantins, Brasil

Karena Quiroz Jiménez¹ , Walter Collischonn¹ and Rodrigo Cauduro Dias de Paiva¹

¹Universidade Federal do Rio Grande do Sul, Porto Alegre, RS, Brasil

E-mails: karenaquiroz@gmail.com (KQJ), waltercollischonn@gmail.com (WS), rodrigo.paiva@ufrgs.br (RCDP)

Received: March 08, 2018 - Revised: November 02, 2018 - Accepted: January 22, 2019

ABSTRACT

In this work, the data assimilation method namely ensemble Kalman filter (EnKF) is applied to the Tocantins River basin. This method assimilates streamflow results by using a distributed hydrological model. The performance of the EnKF is also compared with an empirical assimilation method for hourly time intervals, in which two applications based on information transfer from gauged to ungauged sites and real time streamflow forecasting are assessed. In the first application, both assimilation methods are able to transfer streamflow to ungauged sites, obtaining better results when more than one station located upstream or downstream of the basin are gauged. In the second application, integration of a real time forecast model with EnKF is able to absorb errors at the beginning of the forecast. Therefore, a greater efficiency in the Nash-Sutcliffe index for the first 144 hours in advance in relation to its counterpart without assimilation is obtained. Finally, a comparison between both data assimilation methods shows a greater advantage for the EnKF in long lead times.

Keywords: Ensemble Kalman filter; Distributed hydrological model; Streamflow forecast; Information transfer.

RESUMO

Neste trabalho, o método de assimilação de dados por filtro de Kalman por conjunto (EnKF) é aplicado na bacia do rio Tocantins. Esse método atualiza as vazões do rio usando um modelo hidrológico distribuído. O desempenho de EnKF é também comparado com um método de assimilação empírico a intervalos de tempo horário, onde duas aplicações baseadas em transferência de informação de locais monitorados para não monitorados e previsão de vazão em tempo real são avaliadas. Na primeira aplicação, ambos os métodos de assimilação de dado conseguem transferir vazões a outros locais não monitorados, obtendo melhores resultados quando mais de uma estação localizada a montante ou a jusante da bacia são monitoradas. Na segunda aplicação, a integração de um modelo de previsão com EnKF consegue absorver os erros no início da previsão. Dessa forma, uma maior eficiência no índice de Nash-Sutcliffe para as primeiras 144 horas de antecedência é encontrada quando se compara com os resultados do modelo sem assimilação. Finalmente, a comparação entre os métodos de assimilação de dados no modelo de previsão mostra uma maior vantagem a favor de EnKF em maiores horizontes de previsão.

Palavras-chave: Filtro de Kalman por conjunto; Modelo hidrológico distribuído; Previsão de vazões; Transferência de informação.



INTRODUCTION

Advances in understanding the dynamics of the soil-water-air relationship encouraged the development of hydrological models of the earth surface, in order to perform a simple, even if realistic representation of water movement in the basin through a mass balance, differential equations of routing and energy transfer (BEVEN, 2012). Nevertheless, there are various sources of uncertainties involved, as in the structure of the hydrological model, in the estimation of the hydrological parameters and in the errors of the hydro meteorological forcing data to the model (VRUGT et al., 2005; LIU; GUPTA, 2007; GÖTZINGER; BÁRDOSSY, 2008; SALAMON; FEYEN, 2010). That is, the model parameters using an automatic or manual calibration process will always be associated with uncertainties. For cases of real time flood forecasting, it may be even more difficult due to errors in the rainfall forecast. Also, The sensitivity of the models to the initial conditions of short duration intensive rainfalls corresponding to the nonlinear effects and to the runoff responses may be another cause of uncertainties (CHEN et al., 2013). The quantification and reduction of these uncertainties is necessary to avoid greater errors in decision-making (LIU et al., 2012).

Advanced methods of data assimilation are an option for an optimal combination of information from models that are inherently imperfect, with also uncertain observations, thus obtaining estimates that are physically consistent with the reduction and quantification of the uncertainties (MCLAUGHLIN, 1995; WEERTS; EL SERAFY, 2006; LIU; GUPTA, 2007; CLARK et al., 2008; REICHLER, 2008; LIU et al., 2012; RIDLER et al., 2014; ERCOLANI; CASTELLI, 2017). Initially these methods were more popular in earth sciences for meteorological forecasting, characterization of atmospheric and oceanic conditions; however, in recent years they have been adapted for hydrological and hydraulic applications (LIU et al., 2012).

Data assimilation methods with applications of discharge data into distributed hydrological models are more abundant (WEERTS; EL SERAFY, 2006; CLARK et al., 2008; RAKOVEC et al., 2012; CHEN et al., 2012; DECHANT; MORADKHANI, 2012; CHEN et al., 2013; DUMEDAH; COULIBALY, 2013; ZHANG et al., 2014; MORADKHANI et al., 2005a; MORADKHANI et al., 2005b; SEO et al., 2009; XIE; ZHANG, 2010; PAUWELS; DE LANNOY, 2009; ERCOLANI; CASTELLI, 2017). However, recent advances in remote sensing encouraged the use of observations such as radar altimetry to estimate river levels (PAIVA et al., 2013b; ANDREADIS et al., 2007; BIANCAMARIA et al., 2011), microwave radiation to estimate soil humidity (NAGARAJAN et al., 2011; YAN; MORADKHANI, 2016), remote observation to infer the properties of snow as an extension of large scale coverage (ANDREADIS; LETTENMAIER, 2006), among others.

The first experiments of the advanced method of data assimilation in simulation and forecasting models were performed using lumped models with synthetic data (e.g. WEERTS; EL SERAFY, 2006; DUMEDAH; COULIBALY, 2013; DECHANT; MORADKHANI, 2012; MORADKHANI et al., 2005a; MORADKHANI et al., 2005b; NAGARAJAN et al., 2011; CHEN et al., 2013; SEO et al., 2009). The results of these studies were promising with a great potential for expansion into more complex systems and more realistic situations. Later, the applications

in semi-distributed and distributed hydrological models proved more interesting and appropriate. Chen et al. (2012) demonstrated that the data assimilation methods in semi-distributed models provided a better representation of the maximum streamflow peaks compared to the lumped models. Various methods of data assimilation in distributed hydrological models were applied in different basins worldwide, as in Andreadis and Lettenmaier (2006), Clark et al., (2008), Salamon and Feyen (2009), Mendoza (2010), Noh et al. (2011), Chen et al., (2012), Paiva et al. (2013b), Xie and Zhang (2010), Abaza et al. (2014), Ercolani and Castelli (2017), Xu et al. (2017), among others. Nevertheless, applications of data assimilation in South America basins are still insufficient. The advantage of the distributed models with data assimilation is forcing them with spatial data. Another advantage is the ability to simulate and predict hydrological variables at internal sites in the basin, and it should be underscored that the estimation of the variables at sites with poor or no monitoring in the basin means a new applications to the advanced methods of data assimilation (LIU et al., 2012).

The research cited demonstrated relative accurate when compared to a model that is not assimilated (PAIVA et al., 2013b; ABAZA et al., 2014; MENDOZA, 2010; ERCOLANI; CASTELLI, 2017) when they are applied in distributed hydrological models for real time streamflow forecasting. Additionally, these methods were useful to transfer information on streamflows, such as in Paiva et al. (2013b), Xie et al. (2014) and Zhang et al. (2014); Ercolani and Castelli (2017). The few works that exist on information transfer to ungauged sites showed reliable estimates for streamflow prediction in regions without much monitoring, but their results depend on the location of the monitored sites in relation to the ungauged sites (XIE et al., 2014)

The most disseminated advanced method of data assimilation is of the sequential or filtering methods. The most common methodologies are the ensemble Kalman filter (EVENSEN, 1994; EVENSEN, 2003) (EnKF) and the particle filter (ARULAMPALAM et al., 2002). The first is an extension of the classic Kalman filter method which was developed for linear systems. The EnKF is an approach of simulations by Monte Carlo to the traditional Kalman filter, where the error of matrix of covariance is done routing an ensemble of model states using the updated states of the previous period. The works of Evensen (1994) and Evensen (2003) introduced the ensemble Kalman Filter method presenting an algorithm of the formulation and numerical implementation. The results of multiple studies demonstrated that EnKF is robust and has computational efficiency, and is easy to implement compared to other advanced data assimilation methods (CHEN et al., 2013).

Another data assimilation method is the Particle Filter (PF), similar to EnKF, which utilizes the Monte Carlo method and the filtering theory based on the Bayesian theory. The central idea is to represent the function of posterior probability density of an ensemble of randomly chosen samples (particles), and each particle is associated with a weight (ARULAMPALAM et al., 2002; VRUGT et al., 2013). This method has the advantage of being applicable to models that present nonlinear functions with non-Gaussian density distributions. It updates the weights of the particles and not of the ensemble of state variables, which reduces

numerical instabilities, especially in distributed physical models (LIU et al., 2012). One of the first studies using PF in hydrological applications is that of Moradkhani et al. (2005a) developing a mathematical algorithm to analyze uncertainties in streamflow forecasting, considering uncertainties in the state variables together with the hydrological model parameters. Nevertheless, one may need a greater number of members compared to EnKF to provide more reliable forecasts.

Another group of advanced methods used in recent years is the variational method (COUSTAU et al., 2013), classified as a technique derived from the evolution of distribution of state variables over a time interval which involves the minimization of a cost function. The variational methods are computationally less expensive than EnKF (LIU; GUPTA, 2007). However, they work in a time window containing a sequence of observations, and are more appropriate for curve adjustment than for real time data (SALAMON; FEYEN, 2009).

Research presented by Paz et al. (2007), Collischonn et al. (2005), Collischonn et al. (2007b) and Fan et al. (2014) showed an empirical method of data assimilation for flood forecasting analyses in operational systems of inflows to the reservoirs. However, those techniques only allow assimilating the streamflow in rivers. In this sense, the advanced methods for data assimilation allow estimating the state variables together with the model parameters, considering uncertainties in the input and output data.

The objective of this study is to evaluate a data assimilation scheme using EnKF in a distributed hydrological model in the Tocantins River basin to test its usefulness for real time forecasting and transferring information. In this study, EnKF is coupled to the distributed hydrological model MGB-IPH to assimilate the river streamflows. A model for the generation of a synthetic series of rainfalls is used to alter the state variables of the hydrological model. The results are shown in terms of statistical performance among the simulations of the data assimilation model (EnKF), data assimilation by an empirical model (Empirical) described by Paz et al. (2007) and not using data assimilation (Open-Loop). This work is organized in 4 sections. The first is the Introduction containing the state of art of the application of data assimilation methods in the field of hydrology. The second is the Methodology containing a summary of the ensemble Kalman filter theory, the model of synthetic generation of rainfall and some considerations regarding system errors. The third is Discussion of Results, which describes the final results of the different simulations and, finally, the last section presents the conclusions.

METHODOLOGY

Distributed MGB-IPH Hydrological Model

MGB-IPH is a large-scale distributed hydrological model developed by Collischonn et al. (2007a). This model seeks out to represent processes of routing generation and streamflow propagation in the drainage system of the basin by using meteorological variables, and data derived from geographic information systems based on topography, land use and soil type. The drainage-basin area is divided into irregular cells namely “small basins” in which hydraulic and hydrological quantities are determined. Otherwise, the

Grouped Response Unit (GRU) approach is used for hydrological classification with similar combinations of soil type and land use without consideration of its exact location within the cell. It is important to mention that a cell contains a limited number of distinct GRUs.

MGB-IPH simulates the water balance based on physical relations divided into two distinct processes, vertical and horizontal. Among the vertical processes are canopy interception, evapotranspiration, infiltration, surface runoff, sub-surface and soil water budget, all of them simulated in each GRU. Horizontal hydrological processes include flow generated within each catchment is routed to the stream network using a linear reservoirs type model (surface, subsurface and base flow). Currently, MGB-IPH has three streamflow routing methods along the drainage network, besides the linear routing by Muskingum Cunge, the module of hydrodynamic routing with solution of the complete Saint-Venant equations (PAIVA; COLLISCHONN; BUARQUE, 2013a) and inertial flow routing method (PONTES et al., 2015).

The MGB-IPH model has an empirical data assimilation method coupled to updated streamflow by using routing module by Muskingum Cunge (PAZ et al., 2007; COLLISCHONN et al., 2005; COLLISCHONN et al., 2007b; TUCCI et al., 2006). The usefulness of this empirical model was demonstrated for reservoir inflow predictions based on quantitative precipitation forecast (COLLISCHONN et al., 2005; COLLISCHONN et al., 2007b) and flood forecasting using ensemble rainfall forecasts (FAN et al., 2014; FAN et al., 2016). A complete description of the empirical assimilation methodology can be found in the above mentioned articles.

Ensemble Kalman Filter (EnKF)

Sequential data assimilation consists of estimating the system states variables of the model recursively each time an observation becomes available (MORADKHANI et al., 2005a). When the systems follow a linear behavior of their equations, this problem may be solved by the Kalman filter as an optimal recursive algorithm. In the case of non-linear dynamic systems, the current state vector is linearized to use the extended Kalman Filter. However, this method presents many inconveniences, producing instabilities and divergences (EVENSEN, 2003). The EnKF presents as alternative to the traditional method of extended Kalman Filter (EVENSEN, 1994, 2003). In EnKF, the Monte Carlo method is used to generate an ensemble of trajectories of the model where the covariance matrix of error is done routing an ensemble of model states using the updated states of the previous time step. The model that represents the dynamic of a simulated system could be shown at discrete time intervals and can be represented by the following equation:

$$\mathbf{x}_{t+1} = f(\mathbf{x}_t, \boldsymbol{\theta}, \boldsymbol{\mu}_t) + \boldsymbol{\omega}_t \quad (1)$$

where f is a non-linear operator that represents the transition model from time t to time $t+1$; \mathbf{x} represents the model state variables; $\boldsymbol{\theta}$ represents the model parameters; $\boldsymbol{\mu}$ is the model input or forcing; and $\boldsymbol{\omega}$ represents the errors in the model structure, parameters, input data.

The most general equation that describes the relationship between the system observations and the states of the model is:

$$\mathbf{y}_t = h(\mathbf{x}_{t+1}, \boldsymbol{\theta}_t) + \mathbf{v}_{t+1} \quad (2)$$

where h is a non-linear operator that relates the state variables \mathbf{x} of the model to the observations \mathbf{y} and \mathbf{v} is the vector of the errors in observation.

It is considered that there is no bias in the errors of the model and observations, there is no correlation between the errors of the observations and of the model and the variance of errors is known. After updating the model in the time interval t , the model is used to calculate the states in the time interval $t+1$.

The next group of equations is the formulation of EnKF which consists of the stages of forecasting and updating at each time interval. The notation of time in the following formulas is eliminated in order to avoid confusion with the index that will identify each member of the ensemble. The original routines of the EnKF method are available at site <http://enkf.nersc.no/>. For further details and discussion the reader should look at Evensen (2003), Paiva et al. (2013b) and Clark et al. (2008).

Let \mathbf{X}^f be a $nstate \times N$ - size matrix, where $nstate$ is the number of state variables and N is the number of members of the ensemble, containing all the state vectors of the model in the forecasting stage. It can be represented as follows:

$$\mathbf{X}^f = (\mathbf{x}_1^f, \dots, \mathbf{x}_N^f) \quad (3)$$

where $\mathbf{x}_1^f, \dots, \mathbf{x}_N^f$ are the state vectors of the model for each member of the ensemble. According to Evensen (2003, 2004) in the formulation of EnKF, the mean of the ensemble states ($\bar{\mathbf{x}}^f$) is considered "true", given that the real states are unknown. For this purpose, the mean of the ensemble ($\bar{\mathbf{x}}^f$) is calculated to estimate the matrix of covariance of the errors of model (\mathbf{P}^f) based on the matrix of anomalies defined as $[\mathbf{X}'^f = (\mathbf{x}_1'^f, \dots, \mathbf{x}_N'^f)]$, with $\mathbf{x}_i'^f = \mathbf{x}_i^f - \bar{\mathbf{x}}^f$:

$$\mathbf{P}^f = \frac{1}{N-1} \mathbf{X}'^f \mathbf{X}'^{fT} \quad (4)$$

When an observation is available, the error of this forecast is calculated as $[y_i - h(\mathbf{x}_i^f)]$ known as innovation matrix, \mathbf{y}_i is a matrix of $nobs \times N$ size, where $nobs$ is an observations number and each element of these matrix is calculated for probability distribution model. Thus, the goal of data assimilation is to obtain an optimum estimate (\mathbf{x}^a) of the state variables, considering the errors of the model and of the observations. Consequently, the optimal, non-biased estimate and with minimum variance of the state variables is given by:

$$\mathbf{x}_i^a = \mathbf{x}_i^f + \mathbf{K}(\mathbf{y}_i - \mathbf{H}\mathbf{x}_i^f) \quad (5)$$

$$\mathbf{K} = \mathbf{P}^f \mathbf{H}^T (\mathbf{H}\mathbf{P}^f \mathbf{H}^T + \mathbf{R})^{-1} \quad (6)$$

where \mathbf{K} is the Kalman gain, \mathbf{P} is the matrix of variance-covariance of the errors of model $\boldsymbol{\omega}$, \mathbf{y}_i is the vector of the observations generated, \mathbf{R} is the matrix of variance-covariance of the errors of observations, and \mathbf{H} is the operator that takes the space of the model states to the observation space.

Uncertainty in rainfall forcing

In this study, the state variables of the hydrological model were generated from a model of generation of synthetic series by the distributive multiplicative error with a log-normal distribution in the rainfall observed. The reason for choosing rainfall as a forcing to be disturbed is the consideration that the greatest source of error in the simulation system and hydrological forecasting comes from precipitation. In this sense, the rainfall sequences are generated using Equation 7. This expression was already applied by Nijssen and Lettenmaier (2004) to evaluate the effects of the error on the sampling of precipitation estimated by satellite and in the generation of synthetic series for analysis of data assimilation in Paiva et al. (2013b). The rainfall values are disturbed using a log-normal distribution as shown below:

$$P_c = \left[\frac{1+\beta}{\sqrt{E^2+1}} \exp\left(\sqrt{\ln(E^2+1)} \epsilon\right) \right] P \quad (7)$$

Where P_c is the disturbed rainfall ($\text{mm} \cdot \Delta t^{-1}$), P is the observed rainfall ($\text{mm} \cdot \Delta t^{-1}$), E is the relative error of rainfall (%), β is the relative bias and ϵ is the random variable with the normal distribution correlated with mean zero and variance one. The value of the relative bias is considered zero as in Nijssen and Lettenmaier (2004). To calculate the random variable (ϵ), this is a function of the variable w of Equation 8 that represents a vector of random numbers generated with normal distribution (mean zero, variance equal to 1) and isotropic exponential correlation, in which the spatial correlation drops to e^{-1} in the distance τ_x called length of spatial decorrelation. At each spatial location, the temporal correlation was also considered using the following equation to simulate the temporal evolution of the errors. (EVENSEN, 2003).

$$\epsilon_{i,t} = \alpha \epsilon_{i,t-1} + \sqrt{(1-\alpha^2)} w_{i,t-1} \quad (8)$$

where t is the time interval, $\epsilon_{i,t}$ is a sequence of errors in time with a temporal correlation and after being calculated it is introduced into Equation 7; α estimates the temporal decorrelation according to the following relation:

$$\alpha = 1 - \frac{\Delta t}{\tau_t} \quad (9)$$

where τ_t is the length of temporal decorrelation.

Quantification of errors in the observed streamflow

The measurement errors of streamflow (Q) were modelled using the following relation:

$$Q_c = Q + \epsilon_Q; \quad \epsilon_Q \sim N\left(0, (\sigma_Q Q)^2\right) \quad (10)$$

where Q_c ($\text{m}^3 \cdot \text{s}^{-1}$) is the sequence of observations of Q ($\text{m}^3 \cdot \text{s}^{-1}$) accrue of ϵ_Q ($\text{m}^3 \cdot \text{s}^{-1}$), where ϵ_Q is the random error modelled by a normal distribution, where σ_Q is a parameters that must be specified.

Determination of the assimilation parameters of EnKF

A sensitivity analysis is performed to determine the parameters for the general procedure of EnKF. The first parameter is the number of members of the ensemble of the assimilation method EnKF (N) and the three others refer to the model of generation of synthetic rainfall series (Equation 7), the relative error of rainfall being (E ; %), the spatial decorrelation (τ_x , degrees) and the temporal decorrelation (τ_t , hours). This analysis of sensitivity is valid for hydrographic basins with more than one station to assimilate and where an ensemble of parameters valid for the entire basin is required.

The state variable to be assimilated is streamflow, therefore the sensitivity study is performed by two groups of river gauging stations considered, one for the data assimilation process and the other group for the verification process. For each of these groups a term called changes in root-mean-square error (Δrms), is calculated; greater negative values indicate a better performance of the model and it is calculated as follows:

$$\Delta rms = \frac{100(rms_2 - rms_1)}{rms_1} \quad (11)$$

where rms_1 e rms_2 are the root-mean-square error in observed streamflow with the Open-loop and EnKF streamflow, respectively.

Study area description

The Tocantins River basin is an area of study located in the central region of Brazil, with a drainage area of 310.000 km² up to the confluence with the Araguaia River (see Figure 1). The monthly mean temperature of the study area varies from 20 °C to 25 °C, approximately. The monthly mean maximums occur in the months of August and September while the monthly mean minimum occurs in July and August. The mean rainfall is 1480 mm.year⁻¹ and streamflow is 3300 m³.s⁻¹ at the Estreito station according to the estimated calculation for the 2008-2014. The basin topography elevations range from 83 to 1640 meters.

The Tocantins River basin was selected as area of study because it has a very important hydropower system formed by the Serra da Mesa, Cana Brava, São Salvador, Peixe Angical and Estreito hydroelectric plants. The Cana Brava and Estreito have an installed capacity for electricity generation of 1275 MW and 1087 MW, respectively. It was also chosen because it is a region that periodically suffers extreme events, typical situations in other regions of Brazil. Currently, the National Center of Monitoring and Alerts of Natural Disasters (*CEMADEN Centro Nacional de Monitoramento e Alertas de Desastres Naturais*) monitors the municipalities of Goiatins and Porto Nacional, located within the Tocantins River basin, classified as being vulnerable to hydrological risks. Likewise, the town of Imperatriz do Maranhão, downstream from Estreito suffers constant floods, with an impact on the populations living in the riparian areas.

The basin was discretized into 410 cells, 45 sub-basins and the integration of the use and soil type maps generated 6 different types of hydrological response units which are: forest in medium

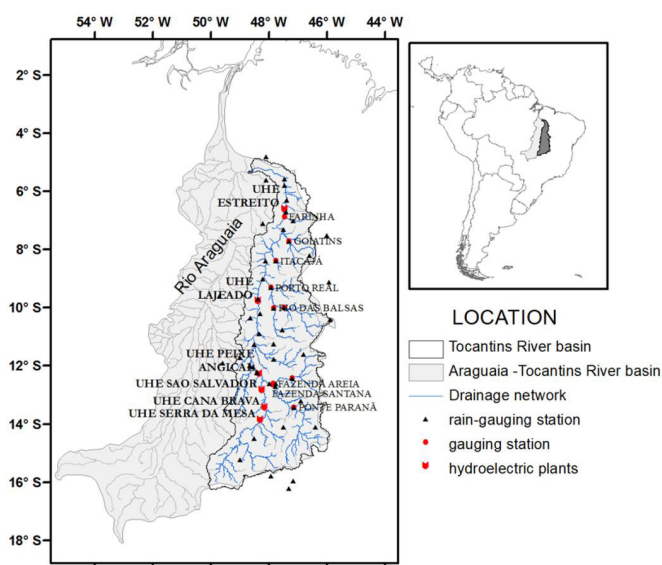


Figure 1. Location of the Tocantins river basin showing the 50 rain-gauging station and 6 gauging stations with naturalized streamflow.

soil (5%), forest in deep soil (9%), low vegetation in medium soil (19%), low vegetation in deep soil (36%), agriculture in deep soil (30%) and waterbody (1%).

Available data

The data on streamflow at hourly time intervals were extracted from 16 stations, 10 (ten) of them from hydroelectric power companies, as well as from the National Water Agency (ANA - *Agência Nacional de Águas*) and the other 6 (six) stations are from National Operator of the Electric System (ONS - *Operador Nacional de Sistema Elétrico*). The naturalized streamflow data of ONS is available at daily time intervals and for this work they were interpolated linearly to obtain hourly data. The drainage areas for the sites with a hydroelectric plant located along the mainstream of the river are more than 50.000 km² and less than 289.000 km². For the stream stations located in the Southeast and Northeast region of the basin the drainage areas range from 3.000 – 44.000 km².

The data on mean air temperature, relative humidity, wind velocity, atmospheric pressure and insolation were obtained from 15 climate gauging stations located around the basin, supplied by ANA and interpolated for hourly data. The precipitation was obtained from 50 rain-gauging station supplied by hydroelectric power companies and by the National Institute of Meteorology (*INMET Instituto Nacional de Meteorologia*). Considering the low density of the rain-gauging stations (1 station at every 6,200 km²) it was chosen to combine with TRMM satellite precipitation product. This option was performed to attempt to improve the response in the simulated streamflows at several basin stations using a methodology based on the work by Rozante et al. (2010). Quiroz (2017) used that methodology to determine temporal series of rainfall at hourly time intervals in the Tocantins River basin for the period of 1998-2014 called MergeHQ.

Uncertainties in the data assimilation system by EnKF

EnKF estimates the states in a nonlinear system, considering the uncertainty, both in the observation and in the structure of the hydrological model, parameters and input data. Thus, the choice of error values is important to obtain a better estimate of the state variables (LIU; GUPTA, 2007; NOH et al., 2011). In this study, the error in observations was defined based on other studies performed, that used the assimilation of data in different river basins. For instance, Paiva et al. (2013b) used 10% for daily streamflow in the Amazon region, Clark et al. (2008) used 10% for hourly streamflow in New Zealand, Noh et al. (2013) in Korea and Japan used 10% and Chen et al. (2013) used 20% in basins in America and in China. Based on this information, the value of 20% was considered as the percentage error of the observed streamflow for all stations.

Integration of the assimilation method using the ensemble Kalman filter with the hydrological model MGB-IPH

The rainfall observed was altered by means of a model for synthetic generation of rainfall. State variables were obtained for each member of the ensemble EnKF through the hydrological simulation process. Sensitivity analysis was performed to determine the parameters that involve the model of synthetic generation of rainfall and the number of members of the EnKF assimilation method. The state variables of MGB-IPH considered in assimilation are water storage in the soil layer, the volume in the reservoirs (surface, subsurface and groundwater), and routing streamflow. The state variables mentioned regarding the water bucket in the soil and the volumes of the three reservoirs were estimated in each GRU of the basin, others that govern the streamflow routing equations were estimated on the stream of the drainage network. Based on these considerations, a total of 5800 state variables for each member of the ensemble is constituted for matrix X^f . The state variables at the beginning of the data assimilation process were estimated based on the initial conditions of the hydrological model and repeated for each member of the ensemble. For all the results of the simulation with EnKF, it was chosen to consider the mean of the streamflows of all members of the ensemble at each time interval. The parameters of model MGB-IPH were considered invariant in time, i.e., model parameters after the calibration process are kept constant throughout the assimilation process by EnKF and streamflow forecasting.

DISCUSSION OF RESULTS

Calibration and verification of the model MGB-IPH parameters

An analysis of the calibration and verification of the model MGB-IPH parameters was performed independently for each type of rainfall data. The calibration period corresponded to

January/2008 - June/2012 and the validation period corresponded to July/2012 - June/2013.

Figure 2 shows the statistical coefficients of Nash-Sutcliffe (NS) for all gauging stations of the Tocantins River basin used in the study. The results of calibration and verification showed the NS efficiency greater than 0.60 with MergeHQ for locals with a drainage area greater than 20,000 km². It was also observed that the NS efficiency of basins with a drainage area close to 10.000 km² improved in the verification compared to the calibration period.

Figure 3 shows the streamflow hydrographs observed and simulated with MergeHQ for stages of calibration and verification of model MGB-IPH for four more representative stations. Model predictions with MergeHQ reproduced observed streamflow reasonably well to streamflow maximum and minimum events in the Tocantins River basin. Other studies such as those by Chiang et al. (2007) Jiang et al. (2012), Quiroz and Collischonn (2015) showed the efficiency of the methods of combining the rainfall to simulate the observed streamflow in a basin. In some cases, the combined rainfall can simulate adequately the peak of maximum observed streamflow. In other cases, a statistical analysis shows an improvement of least error in favor of the streamflows simulated with precipitation combined for a given time period. Quiroz (2017) and Fan et al. (2014) demonstrated the usefulness of these combined rainfall data in a forecast of

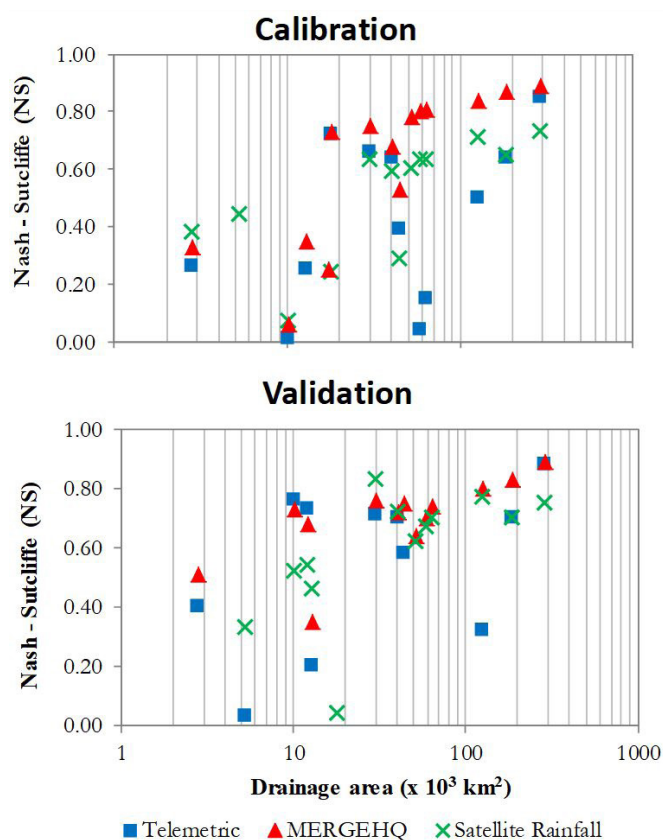


Figure 2. Statistical coefficients of Nash-Sutcliffe (NS) as a function of the accumulated drainage area corresponding to the gauging stations in the Tocantins river basin for the calibration (above) and validation periods (below).

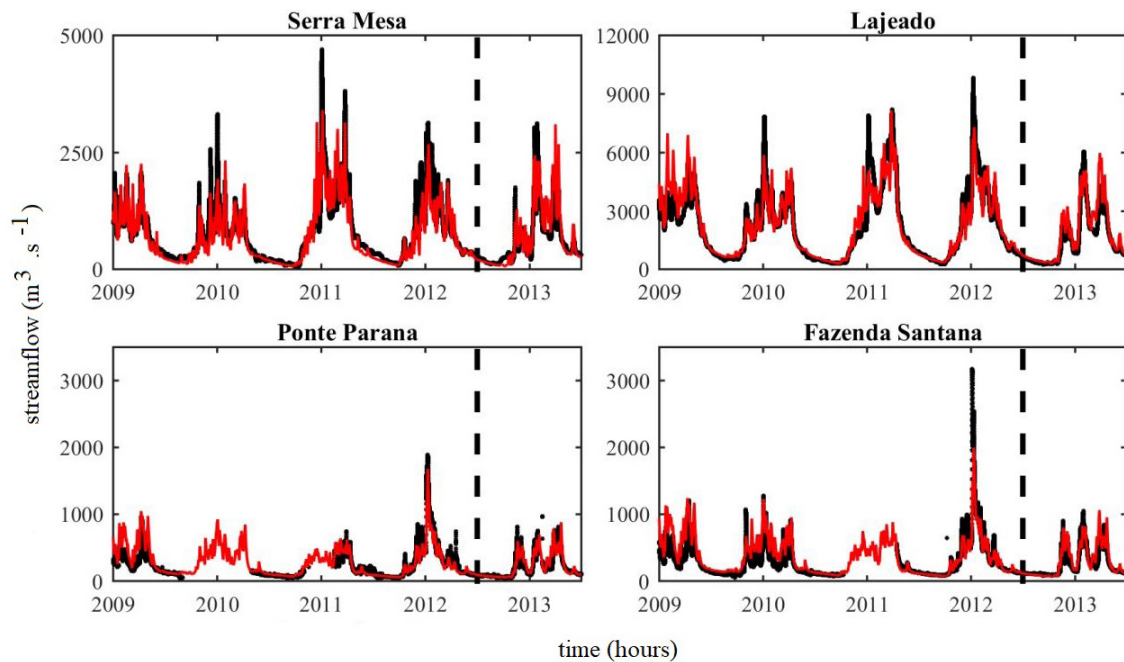


Figure 3. Hydrographs for the four most representative stations in the Tocantins river basin. The vertical dashed line corresponds to the division between the calibration (January, 2008 – June, 2012) and validation (July, 2012 – June, 2013) periods, respectively. Red line corresponds to the streamflows of simulation with MergeHQ and black line corresponds to the observed streamflows.

deterministic streamflows and probabilistic forecast by ensemble in the Tocantins river basin respectively. Thus, the combined rainfall given by MergeHQ was chosen as rainfall observed for sensitivity analysis, a process of assimilation to transfer information and analyze streamflow forecasts.

Sensitivity analysis

There is no norm to estimate the size ensemble. According to the literature consulted, in the case of basins with one control station it is usual to test several numbers of members and compare with a measure of error between the variables observed and assimilated for a given period as in Weerts and El Serafy (2006), Nagarajan et al. (2011), Chen et al. (2013), and Vrugt et al. (2013). Other studies that include the assimilation of several gauging stations merely test a given number of members, as in Mendoza (2010), Rakovec et al. (2012), Chen et al. (2012), and Ridler et al. (2014) varying from 30 to 200 members.

For this study, a procedure similar to that presented in Paiva et al. (2013b) was adopted to determine the number of members (N), the relative error precipitation (E ; %), spatial decorrelation (τ_x ; degrees) and temporal decorrelation (τ_t ; hours). Lajeado station was chosen to compose the temporal series of stations for the assimilation process due to its central location in the basin, whereas the Serra da Mesa, Cana Brava, São Salvador, Peixe Angical and Estreito stations were considered for the verification process. The time period to calculate the parameters was around 6 months from Jan/2012 to June/2012.

Figure 4 shows the results of the EnKF sensitivity analysis for the ensemble size, relative error precipitation, spatial and

temporal decorrelation. According to Equation 11, more negative values of Δrms produce improvement data assimilation model, interpreted as a better approximation of the hydrographs between the flows observed and simulated by EnKF. The values of Δrms for assimilation, composed by the Lajeado station, presented more negative values compared to the stations for verification composed by the Serra da Mesa, Cana Brava, São Salvador, Peixe Angical and Estreito stations.

According to the analysis showed in Figure 4, the data assimilation scheme by EnKF depends on the ensemble size. An improvement in the performance of the model is observed with the increase of this variable until $N = 100$, after this value, the Δrms value remains almost constant and larger in assimilation and validation sites, respectively. The data assimilation scheme is also sensitive to the precipitation relative error E , improving its performance in assimilation sites. However, when $E \geq 50\%$, the Δrms value is practically constant in validation sites. Also, a moderate and low dependence was observed between τ_x and σ_t , respectively. Finally, based on the sensitivity analysis it was decided to consider the following values: $N = 100$; $E = 50\%$; $\tau_x = 2^\circ$ e $\tau_t = 10$ hours.

Transfer of information from gauged sites to ungauged sites

Four scenarios are elaborated considering assimilate one, two and three from six gauging stations located in the mainstream course of the Tocantins River basin. The remaining stations are used to validate the performance of data assimilation model to transfer streamflow from gauge sites to ungauged sites. The first scenario (EST) considers Estreito gauge as monitored site and

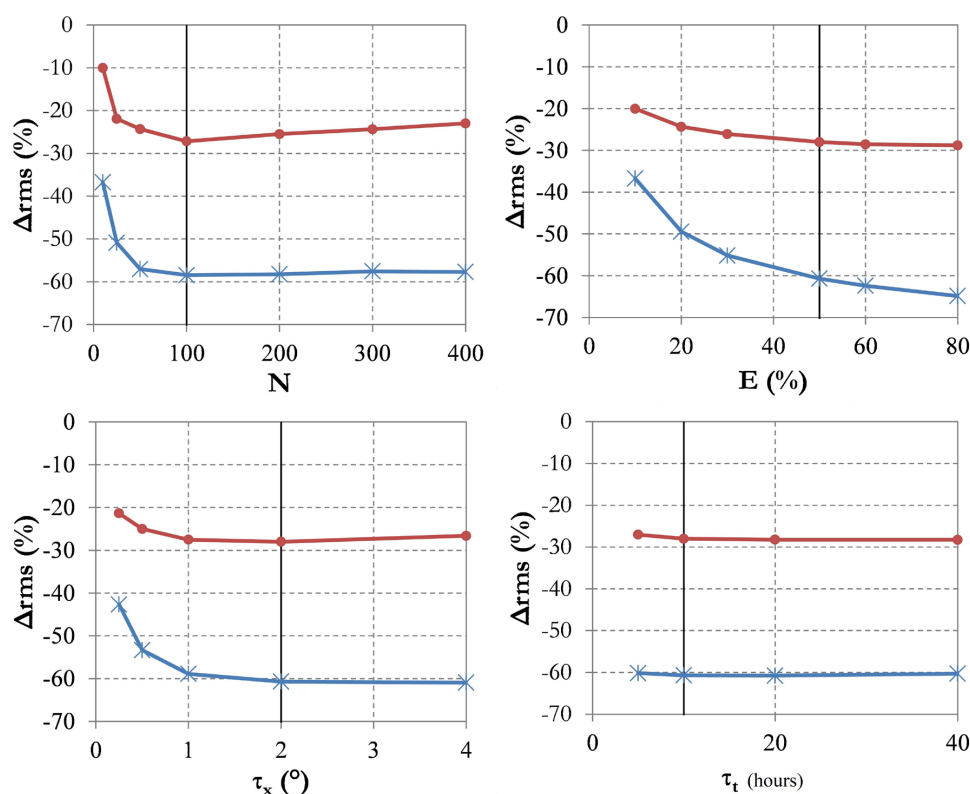


Figure 4. Ensemble size of EnKF and parameters of the precipitation synthetic model. Changes of the root mean square error for gauges with EnKF (star dots, blue color) and gauges with verification (circle dots, red color).

located further downstream in the basin. The second scenario (LAJ) considers Lajeado station as monitored site, located approximately in the central region of the basin. The third scenario (LAJ_EST) considers Lajeado station together with the Estreito station as monitored sites. Finally, the fourth scenario (SM_CB_SS) considers Serra da Mesa, Cana Brava and São Salvador stations as monitored sites located upstream in the basin.

As to the empirical method of data assimilation, the updated (assimilated) streamflows are based on a correction factor among the observed and calculated streamflows by the hydrological model and by the ratio of the accumulated drainage areas of the cell to be updated and the cell corresponding to the gauge with the observation (PAZ et al., 2007). This condition makes the statistical terms exact, presenting null error and perfect efficiency of Nash-Sutcliffe with values equal to 1 at sites with measured data. For the case of the statistical terms calculated with EnKF, the error will always exist because the covariance matrix is calculated based on the errors of the measurements of the gauge.

The data assimilation performance for transfer of information of streamflow is evaluated by comparison of observed and simulated series via EnKF and empirical method in ungauged sites. The Nash-Sutcliffe efficiency and the root mean square error (RMSE) are used as statistical terms for evaluations within the period from January/2009 to December/2011. Table 1 shows the results of RMSE and NS of the simulations of EnKF, empirical method and open-loop for four scenarios at four gauges located in the main network of the Tocantins river basin.

A comparison between both data assimilation methods showed a greater advantage for the empirical method with NS values greater than 0.70 for the last three scenarios. For instance, in scenario 2 (LAJ), NS for Peixe Angical is 0.92 and 0.82 for the empirical method and EnKF, respectively. On the other hand, already in the first scenario, EnKF presents a slight advantage over the empirical method.

The results in transfer of information to EnKF were able to absorb streamflow errors. However, the number of stations and their locations were important for evaluating model statistical performance. For the two first scenarios, where only one gauge was considered, a similar performance was observed between open-loop and EnKF in ungauged locations. For instance, Peixe Angical presents NS of 0.82 for both approaches. When the gauged locations conformed by more than one station, such as scenarios 3 and 4, located more downstream or upstream, the statistical performance favored Open-loop, with exception of the gauge Estreito in the four scenario, where Lajeado presents NS of 0.92 for EnKF and 0.87 for Open-loop, for instance.

Figure 5 shows the RMSE and NS for the gauges located in the southeast region of the basin (such as Fazenda Areia, Fazenda Santana, Rio da Palma and Ponte Paranã, see Figure 1) of scenario 3 (LAJ_EST). It is observed that in three of the four stations, monitoring at Lajeado and Estreito improved the statistical performance at these stations compared to the open-loop model.

The performance in monitored stations is also analyzed. The simulations by the empirical method are forced to the observed streamflow, while simulations by EnKF are corrected considering

errors in the observed streamflow and state variables. In the case of scenario 3 (LAJ_EST), the series of streamflow assimilated by EnKF at Lajeado and Estreito showed that NS of 0,96 and 0,97 compared to open-loop results with 0,87 and 0,56, respectively.

To illustrate the aforementioned results, the hydrographs of Peixe Angical and Estreito for scenario 3 (LAJ_EST) with EnKF, open-loop and observed streamflow simulations are shown in Figure 6. The streamflows of EnKF presented an adequate follow up of the observed maximum and minimum streamflows. In Estreito, the assimilated streamflows presented an adequate adjustment to the observed data, all this being more evident since January 2011. A statistical analysis shows that NS efficiency at Peixe Angical is 0.89 (EnKF) and 0.82 (open-Loop) and at Estreito the NS efficiency is 0.97 (EnKF) and 0.56 (open-Loop).

Streamflow forecasting

Streamflow forecasting was realized for two periods, the first one lasting from January 1, 2013 to May 15, 2013, and the second one from December 15, 2013 to March 15, 2014, obtaining in total 3575 and 1800 time intervals, respectively. In each of these

periods the forecast frequency was 1 hour with a lead time of 144 hours (6 days). Here, it was decided to consider the observed rainfall as the forecast rainfall that simulates a scenario of real time forecasting, in which the rainfall forecasts do not present errors. This consideration has already been made in other studies such as in Tucci et al. (2006) and Paiva et al. (2013b), where data assimilation methods were applied. The results of streamflow forecasting are shown at the gauges located in the main network of the Tocantins river, which are considered as those of the major importance for a real time forecasting system.

Figure 7 and Figure 8 show the efficiency of Nash-Sutcliffe (NS) based on the lead time for four representative gauges for each analyzed period. Vertical lines in gray color represented the superiority of performance, in terms of NS, of a data assimilation method in relation to another. Both figures show that the assimilation with both methods (EnKF and empirical method) was greater as compared to the efficiency of simulations without assimilation. In addition, the assimilation used in both methods proved to absorb the errors at the beginning of the forecast.

Also, Figure 7 shows greater efficiency of EnKF compared to the empirical method for lead times longer than

Table 1. Statistical results of four gauges most representatives for four scenarios of information transfer of streamflow in Tocantins river basin.

Scenario		Serra Mesa			Peixe Angical			Lajeado			Estreito		
		Empirical	EnKF	Open-Loop	Empirical	EnKF	Open-Loop	Empirical	EnKF	Open-Loop	Empirical	EnKF	Open-Loop
1	RMSE	334	301	302	659	508	506	815	624	615	0.00	1630	1575
	NS	0.73	0.78	0.78	0.70	0.82	0.82	0.77	0.86	0.87	1.00	0.53	0.56
2	RMSE	268	300	302	339	511	506	0.00	634	615	1549	1645	1575
	NS	0.82	0.78	0.78	0.92	0.82	0.82	1.00	0.86	0.87	0.57	0.52	0.56
3	RMSE	268	295	302	338	405	506	0.00	327	615	0.00	391	1575
	NS	0.82	0.79	0.78	0.92	0.89	0.82	1.00	0.96	0.87	1.00	0.97	0.56
4	RMSE	0.00	129	302	298	329	506	438	469	615	1585	1721	1575
	NS	1.00	0.96	0.78	0.94	0.93	0.82	0.93	0.92	0.87	0.55	0.47	0.56

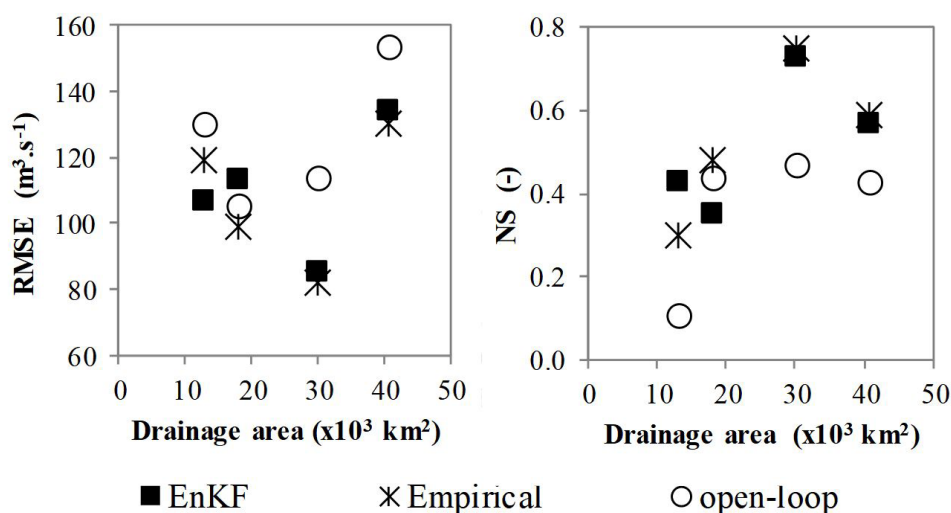


Figure 5. Root mean square error and Nash –Sutcliffe efficiency calculated for the Fazenda Areia, Fazenda Santana, Rio da Palma and Ponte Paraná corresponding to scenario 3 (LAJ_EST).

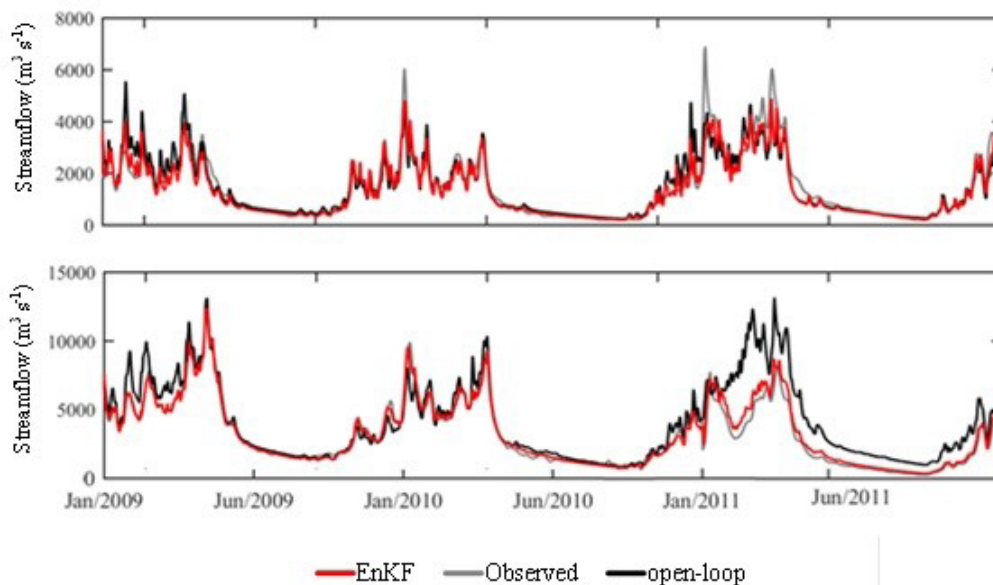


Figure 6. Hydrographs of the Peixe Angical (above) and Estreito (below) stations considering scenario 3 (LAJ_EST) with the simulations by EnKF, open-loop and observed streamflow for the period of January, 2009 – December, 2011.

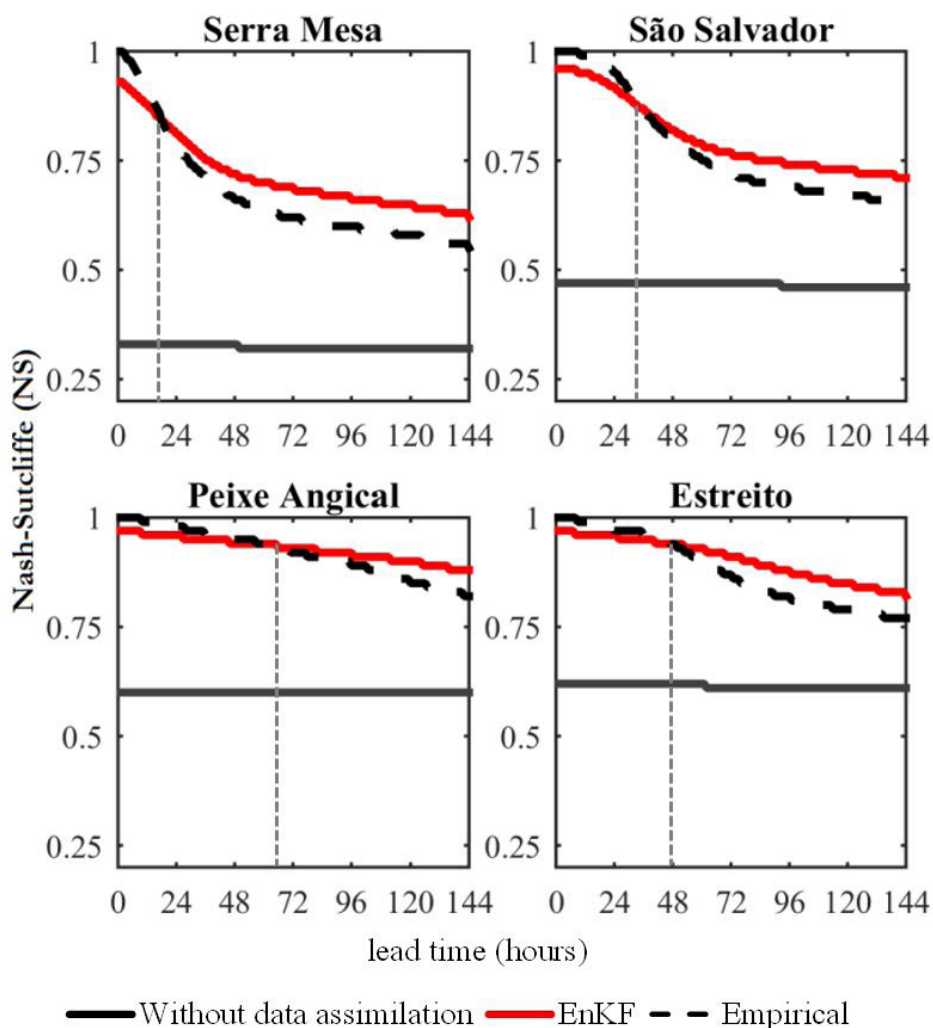


Figure 7. Nash-Sutcliffe efficiency coefficient (NS) as function as a function of lead time with a forecasting period from January 1, 2013 to May 15, 2013.

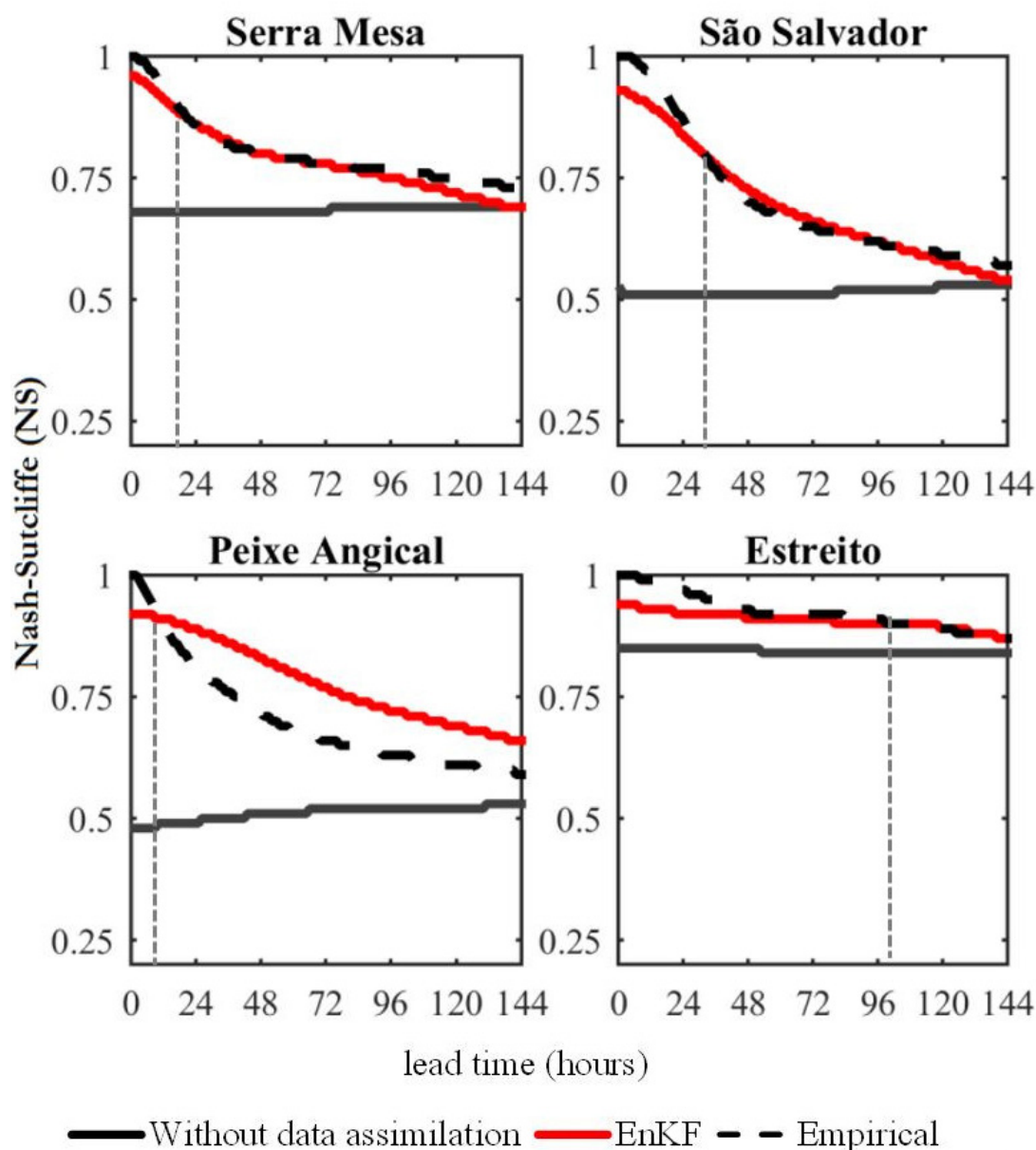


Figure 8. Nash-Sutcliffe efficiency coefficient (NS) as a function of lead time with a forecasting period from December 15, 2013 to March 15, 2014.

18, 37, 57 and 48 hours for Serra Mesa, São Salvador, Peixe Angical and Estreito, respectively, related to the first period. In Figure 8, the performance with EnKF expressed in terms of NS was superior to the empirical method for lead times longer than 10 hours for Peixe Angical and 100 hours for Estreito. However, in the Serra Mesa and São Salvador, the efficiency of NS calculated by the empirical method is nearly equal to that obtained with EnKF for lead times longer than 20 hours.

A visual analysis of the forecasts using the data assimilation methods is shown in Figure 9. This analysis was performed for maximum events during the first period of analysis at two gauges, Peixe Angical and Lajeado. It is observed that the forecasting hydrographs calculated with both data assimilation methods showed differences when compared to the maximum observed

streamflow. For instance, in Peixe Angical, the maximum streamflow obtained with the empirical method is overestimated, while this is underestimated with EnKF in relation to the observed maximum streamflow at January 1, 2013. In the same gauge, the streamflow forecast with both data assimilation methods were underestimated when compared to the maximum observed streamflow at January 26, 2013. In Lajeado, the maximum streamflow obtained with data assimilation methods were anticipated (18/1/2013) compared to the maximum observed streamflow. For the case of the second event (27/1/2013), the simulations with data assimilation were anticipated and underestimated in relation to the maximum observed streamflow. The forecasting analysis carried out at the beginning of this section has shown the usefulness of the data assimilation methods for several forecast horizons.

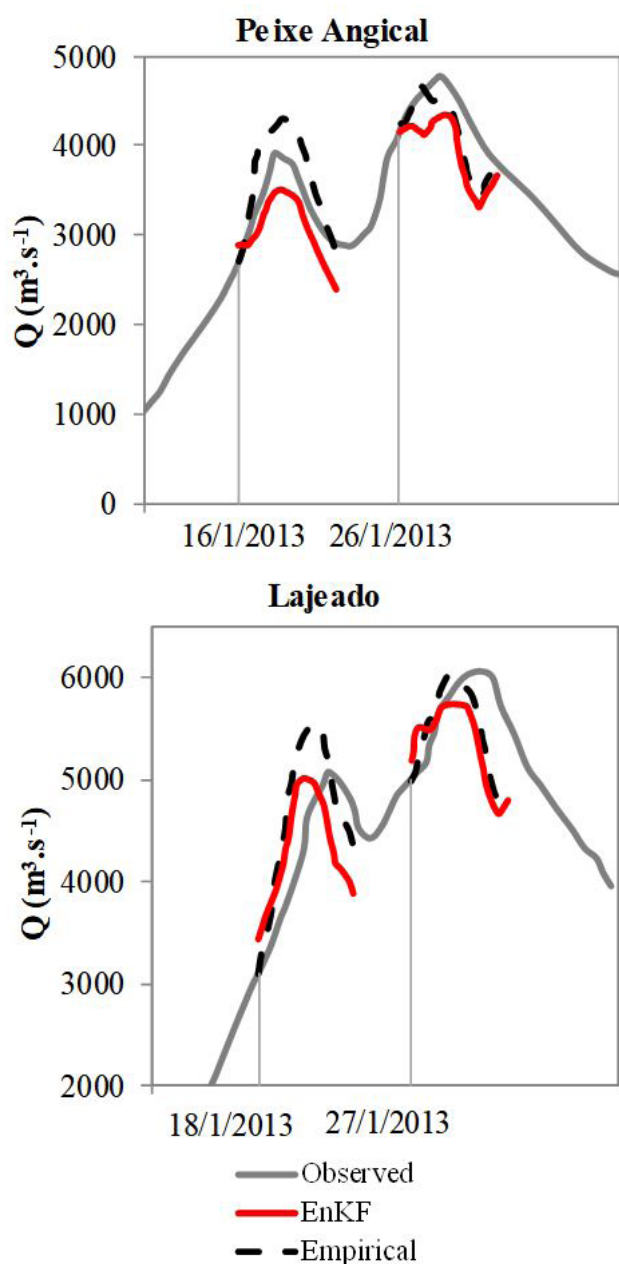


Figure 9. Streamflow forecasting hydrographs for Peixe Angical (above) and Lajeado (below) gauges with the beginning of forecasting at two instants in time corresponding to the first period of analysis.

CONCLUSIONS

In this study, a data assimilation method namely EnKF was tested to evaluate its usefulness for real time streamflow forecasting and the transfer of information of streamflow from gauge sites to ungauged sites. The results were presented in statistical terms using a distributed hydrological model applied to the Tocantins river basin in Brazil. A comparison to an empirical method of data assimilation was also studied. For assimilation with EnKF, the state variables were generated from a synthetic generation model of rainfall by means of the hydrological model. The parameters of the rainfall generation model, such as the relative error of

rainfall, as well as the ensemble size of EnKF were calculated using a sensitivity analysis.

According to the results presented herein, the data assimilation schemes (empirical and EnKF) proved to be reliable in the transferring information, although these results properly depend on the location of the gauge sites. Gauge sites located downstream and in the middle of the basin led to adequate results, even better than those obtained with the open-loop simulation. Also gauge sites upstream in the basin transferred streamflows at nearby gauges. The consideration of more than one gauge as the set of gauge sites led to the method of assimilation by EnKF to present better performance when compared to the performance of the open-loop simulations. Thus, the comparison of the two assimilation methods revealed a slightly better advantage, in statistical terms, in favor of the empirical method.

In the streamflow forecasting scheme, the simulations using EnKF proved to be promising, in which the NS efficiency was greater than that of the empirical method for greater lead time of forecasting (for instance, 10 hours for the Peixe Angical and 100 hours for the Estreito). Although the empirical method behave better in terms of computation time. However, the configuration for the management of uncertainties from various sources can be treated with the more advanced assimilation methods such as EnKF and not with the empirical method.

Finally, it is recommended that the results in transfer of information and real time streamflow should be discussed using other sources of uncertainties such as model parameters. The incorporation of uncertainties into the parameters of the distributed hydrological model is an analysis that has been discussed in recent years.

ACKNOWLEDGEMENTS

The author first is grateful for the financial support from PNP/DCAPES of Post-Graduate Program in Water Resources an environmental Sanitation (PPGRHSA) of Institute of hydraulic Research (IPH) of Federal University of Rio Grande do Sul (UFRGS).

REFERENCES

- ABAZA, M.; ANCTIL, F.; FORTIN, V.; TURCOTTE, R. Sequential streamflow assimilation for short-term hydrological ensemble forecasting. *Journal of Hydrology*, v. 519, p. 2692-2706, 2014. <http://dx.doi.org/10.1016/j.jhydrol.2014.08.038>.
- ANDREADIS, K. M.; CLARK, E. A.; LETTENMAIER, D. P.; ALSDORF, D. E. Prospects for river discharge and depth estimation through assimilation of swath-altimetry into a raster-based hydrodynamics model. *Geophysical Research Letters*, v. 34, n. 10, p. L10403, 2007. <http://dx.doi.org/10.1029/2007GL029721>.
- ANDREADIS, K. M.; LETTENMAIER, D. P. Assimilating remotely sensed snow observations into a macroscale hydrology model. *Advances in Water Resources*, v. 29, n. 6, p. 872-886, 2006. <http://dx.doi.org/10.1016/j.advwatres.2005.08.004>.

- ARULAMPALAM, M.; MASKELL, S.; GORDON, N.; CLAPP, T. A tutorial on particle filters for online nonlinear/non-Gaussian Bayesian tracking. *IEEE Transactions on Signal Processing*, v. 50, n. 2, p. 174-188, 2002. <http://dx.doi.org/10.1109/78.978374>.
- BEVEN, K. Down to Basics: Runoff processes and the modelling process. In: BEVEN, K. *Rainfall-Runoff Modelling: The Primer*. Hoboken: John Wiley & Sons, Ltd., 2012. cap. 1, p. 1-23.
- BIANCAMARIA, S.; DURAND, M.; ANDREADIS, K. M.; BATES, P. D.; BOONE, A.; MOGNARD, N. M.; RODRÍGUEZ, E.; ALSDORF, D. E.; LETTENMAIER, D. P.; CLARK, E. A. Assimilation of virtual wide swath altimetry to improve Arctic river modeling. *Remote Sensing of Environment*, v. 115, n. 2, p. 373-381, 2011. <http://dx.doi.org/10.1016/j.rse.2010.09.008>.
- CHEN, H.; YANG, D.; HONG, Y.; GOURLEY, J. J.; ZHANG, Y. Hydrological data assimilation with the Ensemble Square-Root-Filter: Use of streamflow observations to update model states for real-time flash flood forecasting. *Advances in Water Resources*, v. 59, p. 209-220, 2013. <http://dx.doi.org/10.1016/j.advwatres.2013.06.010>.
- CHEN, J.; ZHANG, W.; GAO, J.; CAO, K. Assimilating multi-site measurements for semi-distributed hydrological model Updating. *Quaternary International*, v. 282, p. 122-129, 2012. <http://dx.doi.org/10.1016/j.quaint.2012.01.030>.
- CHIANG, Y.; HSU, K.; CHANG, F.; HONG, Y.; SOROOSHIAN, S. Merging multiple precipitation sources for flash flood forecasting. *Journal of Hydrology*, v. 340, n. 3-4, p. 183-196, 2007. <http://dx.doi.org/10.1016/j.jhydrol.2007.04.007>.
- CLARK, M. P.; RUPP, D. E.; WOODS, R. A.; ZHENG, X.; IBBITT, R. P.; SLATER, A. G.; SCHMIDT, J.; UDDSTROM, M. J. Hydrological data assimilation with the ensemble Kalman filter: Use of streamflow observations to update states in a distributed hydrological model. *Advances in Water Resources*, v. 31, n. 10, p. 1309-1324, 2008. <http://dx.doi.org/10.1016/j.advwatres.2008.06.005>.
- COLLISCHONN, W.; ALLASIA, D.; SILVA, B. C.; TUCCI, C. M. The MGB-IPH model for large-scale rainfall - runoff modeling. *Hydrological Sciences Journal*, v. 52, n. 5, p. 878-895, 2007a. <http://dx.doi.org/10.1623/hysj.52.5.878>.
- COLLISCHONN, W.; HAAS, R.; ANDREOLLI, I.; TUCCI, C. E. Forecasting River Uruguay flow using rainfall forecasts from a regional weather-prediction model. *Journal of Hydrology*, v. 305, n. 1-4, p. 87-98, 2005. <http://dx.doi.org/10.1016/j.jhydrol.2004.08.028>.
- COLLISCHONN, W.; MORELLI TUCCI, C. E.; CLARKE, R. T.; CHOU, S. C.; GUILHON, L. G.; CATALDI, M.; ALLASIA, D. Medium-range reservoir inflow predictions based on quantitative precipitation forecasts. *Journal of Hydrology*, v. 344, n. 1-2, p. 112-122, 2007b. <http://dx.doi.org/10.1016/j.jhydrol.2007.06.025>.
- COUSTAU, M.; RICCI, S.; BORRELL-ESTUPINA, V.; BOUVIER, C.; THUAL, O. Benefits and limitations of data assimilation for discharge forecasting using and event-based rainfall-runoff model. *Natural Hazards and Earth System Sciences*, v. 13, n. 3, p. 583-596, 2013. <http://dx.doi.org/10.5194/nhess-13-583-2013>.
- DECHANT, C. M.; MORADKHANI, H. Examining the effectiveness and robustness of sequential data assimilation methods for quantification of uncertainty in hydrologic forecasting. *Water Resources Research*, v. 48, n. 4, p. W04518, 2012. <http://dx.doi.org/10.1029/2011WR011011>.
- DUMEDAH, G.; COULIBALY, P. Evaluating forecasting performance for data assimilation methods: The ensemble Kalman filter, the particle filter, and the evolutionary-based assimilation. *Advances in Water Resources*, v. 60, p. 47-63, 2013. <http://dx.doi.org/10.1016/j.advwatres.2013.07.007>.
- ERCOLANI, G.; CASTELLI, F. Variational assimilation of streamflow data in distributed flood forecasting. *Water Resources Research*, v. 53, n. 1, p. 158-183, 2017. <http://dx.doi.org/10.1002/2016WR019208>.
- EVENSEN, G. Sequential data assimilation with a non-linear quasi-geostrophic model using Monte Carlo methods to forecast error statistics. *Journal of Geophysical Research*, v. 99, n. C5, p. 10143-10162, 1994. <http://dx.doi.org/10.1029/94JC00572>.
- EVENSEN, G. The Ensemble Kalman Filter: theoretical formulation and practical implementation. *Ocean Dynamics*, v. 53, n. 4, p. 343-367, 2003. <http://dx.doi.org/10.1007/s10236-003-0036-9>.
- EVENSEN, G. Sampling strategies and square root analysis schemes for the EnKF. *Ocean Dynamics*, v. 54, n. 6, p. 539-560, 2004. <http://dx.doi.org/10.1007/s10236-004-0099-2>.
- FAN, F. M.; COLLISCHONN, W.; QUIROZ, K. J.; SORRIBAS, M.; BUARQUE, D. C.; SIQUEIRA, V. Flood forecasting on the Tocantins river using ensemble rainfall forecasts and real-time satellite rainfall estimates. In: INTERVATIONAL CONFERENCE ON FLOOD MANAGEMENT, 6., 2014, São Paulo. *Anais...* São Paulo: ICFM6, 2014.
- FAN, F. M.; PAIVA, R. C. D.; COLLISCHONN, W. Hydrological Forecasting Practices in Brazil. In: ADAMS III, T. E.; PAGANO, T. C. *Flood forecasting: a global perspective*. Amsterdam: Academic Press, 2016. p. 41-66. <http://dx.doi.org/10.1016/B978-0-12-801884-2.00002-5>.
- GÖTZINGER, J.; BÁRDOSSY, A. Generic error model for calibration and uncertainty estimation of hydrological models. *Water Resources Research*, v. 44, n. 12, p. W00B07, 2008. <http://dx.doi.org/10.1029/2007WR006691>.
- JIANG, S.; REN, L.; HONG, Y.; YONG, B.; YANG, X.; YUAN, F.; MA, M. Comprehensive evaluation of multi-satellite precipitation products with a dense rain gauge network and optimally merging their simulated hydrological flows using the Bayesian model averaging method. *Journal of Hydrology*, v. 452-453, p. 213-225, 2012. <http://dx.doi.org/10.1016/j.jhydrol.2012.05.055>.

- LIU, Y.; GUPTA, H. V. Uncertainty in hydrologic modeling: Toward an integrated data assimilation framework. *Water Resources Research*, v. 43, n. 7, p. W07401, 2007. <http://dx.doi.org/10.1029/2006WR005756>.
- LIU, Y.; WEERTS, A. H.; CLARK, M.; HENDRICKS FRANSSEN, H.-J.; KUMAR, S.; MORADKHANI, H.; SEO, D.-J.; SCHWANENBERG, D.; SMITH, P.; VAN DIJK, A. I. J. M.; VAN VELZEN, N.; HE, M.; LEE, H.; NOH, S. J.; RAKOVEC, O.; RESTREPO, P. Advancing data assimilation in operational hydrologic forecasting: progresses, challenges, and emerging opportunities. *Hydrology and Earth System Sciences*, v. 16, n. 10, p. 3863-3887, 2012. <http://dx.doi.org/10.5194/hess-16-3863-2012>.
- MCLAUGHLIN, D. Recent developments in hydrologic data assimilation. *Reviews of Geophysics*, v. 33, n. S2, p. 977-984, 1995. <http://dx.doi.org/10.1029/95RG00740>.
- MENDOZA, P. A. *Pronóstico de Caudales de Crecida Mediante un Modelo Hidrológico Distribuido y la Asimilación de Datos Observados*, 2010. 218 f. Tesis (Tesis para optar al grado de magister en ciencias de la ingeniería, mención recursos y medio ambiente hídrico) - Facultad de ciencias físicas y matemáticas-Departamento de Ingeniería civil, Universidad de Chile, Chile, 2010.
- MORADKHANI, H.; HSU, K. L.; GUPTA, H.; SOROOSHIAN, S. Uncertainty assessment of hydrologic model states and parameters: Sequential data assimilation using the particle filter. *Water Resources Research*, v. 41, n. 5, 2005a. <http://dx.doi.org/10.1029/2004WR003604>.
- MORADKHANI, H.; SOROOSHIAN, S.; GUPTA, H. V.; HOUSER, P. R. Dual state-parameter estimation of hydrological models using ensemble Kalman filter. *Advances in Water Resources*, v. 28, n. 2, p. 135-147, 2005b. <http://dx.doi.org/10.1016/j.advwatres.2004.09.002>.
- NAGARAJAN, K.; JUDGE, J.; GRAHAM, W. D.; MONSIVAIS-HUERTERO, A. Particle Filter-based assimilation algorithms for improved estimation of root-zone soil moisture under dynamic vegetation conditions. *Advances in Water Resources*, v. 34, n. 4, p. 433-447, 2011. <http://dx.doi.org/10.1016/j.advwatres.2010.09.019>.
- NIJSEN, B.; LETTENMAIER, D. P. Effect of precipitation sampling error on simulated hydrological fluxes and states: Anticipating the Global Precipitation Measurement satellites. *Journal of Geophysical Research, D, Atmospheres*, v. 109, n. D2, p. D02103, 2004. <http://dx.doi.org/10.1029/2003JD003497>.
- NOH, S. J.; TACHIKAWA, Y.; SHIIBA, M.; KIM, S. Applying sequential Monte Carlo methods into a distributed hydrologic model: lagged particle filtering approach with regularization. *Hydrology and Earth System Sciences*, v. 15, n. 10, p. 3237-3251, 2011. <http://dx.doi.org/10.5194/hess-15-3237-2011>.
- NOH, S. J.; TACHIKAWA, Y. O.; SHIIBA, M.; KIM, S. Ensemble kalman filtering and particle filtering in a lag-time window for short-term streamflow forecasting with a distributed hydrologic model. *Journal of Hydrologic Engineering*, v. 18, n. 12, p. 1684-1696, 2013. [http://dx.doi.org/10.1061/\(ASCE\)HE.1943-5584.0000751](http://dx.doi.org/10.1061/(ASCE)HE.1943-5584.0000751).
- PAIVA, R. C. D.; COLLISCHONN, W.; BUARQUE, D. C. Validation of a full hydrodynamic model for large-scale hydrologic modelling in the Amazon. *Hydrological Processes*, v. 27, n. 3, p. 333-346, 2013a. <http://dx.doi.org/10.1002/hyp.8425>.
- PAIVA, R. C. D.; COLLISCHONN, W.; BONNET, M.-P.; GONÇALVES, L. G. G.; CALMANT, S.; GETIRANA, A.; SANTOS DA SILVA, J. Assimilating in situ and radar altimetry data into a large-scale hydrologic-hydrodynamic model for streamflow forecast in the Amazon. *Hydrology and Earth System Sciences*, v. 17, n. 7, p. 2929-2946, 2013b. <http://dx.doi.org/10.5194/hess-17-2929-2013>.
- PAUWELS, V. R. N.; DE LANNOY, G. J. M. Ensemble-based assimilation of discharge into rainfall-runoff models: A comparison of approaches to mapping observational information to state space. *Water Resources Research*, v. 45, n. 8, p. W08428, 2009. <http://dx.doi.org/10.1029/2008WR007590>.
- PAZ, A. R.; COLLISCHONN, W.; TUCCI, C. E. M.; CLARKE, R. T.; ALLASIA, D. Data Assimilation in a large-scale distributed hydrological model for medium-range flow forecasts. In: PROCEEDINGS OF SYMPOSIUM HS2004 AT IUGG2007, 1., 2007, Perugia. Perugia: IAHS, 2007. p. 471-478.
- PONTES, P. M. R.; COLLISCHONN, W.; FAN, F. PAIVA, R. C. D.; BUARQUE, D. Modelagem hidrológica e hidráulica de grande escala com propagação inercial de vazões. *Revista Brasileira de Recursos Hídricos*, v. 20, n. 4, out./dez. 2015.
- QUIROZ, J.; COLLISCHONN, W. Método de combinação de dados de precipitação estimados por satélite e medidos em pluviômetros para a modelagem hidrológica, *Revista Brasileira de Recursos Hídricos*. v. 20, n. 1. 10.21168/rbrh.v20n1.p202-217, 2015.
- QUIROZ, K. J. *Previsão de vazão usando estimativas de precipitação por satélite e assimilação de dados*. 2017. 180 f. Tese (Doutorado em Recursos Hídricos e Saneamento Ambiental) - Instituto de Pesquisas Hidráulicas - IPH, Universidade Federal do Rio Grande do Sul, Porto Alegre, 2017. Disponível em: <http://hdl.handle.net/10183/170484>. Acesso em: 1 fev. 2018.
- RAKOVEC, O.; WEERTS, A. H.; HAZENBERG, P.; TORFS, P. J. J. F.; UIJLENHOET, R. State updating of a distributed hydrological model with Ensemble Kalman Filtering: Effects of updating frequency and observation network density on forecast accuracy. *Hydrology and Earth System Sciences*, v. 16, n. 9, p. 3435-3449, 2012. <http://dx.doi.org/10.5194/hess-16-3435-2012>.
- REICHLER, R. H. Data assimilation methods in the Earth sciences. *Advances in Water Resources*, v. 31, n. 11, p. 1411-1418, 2008. <http://dx.doi.org/10.1016/j.advwatres.2008.01.001>.
- RIDLER, M. E.; VAN VELZEN, N.; HUMMEL, S.; SANDHOLT, I.; FALK, A. K.; HEEMINK, A.; MADSEN, H. Data assimilation framework: Linking an open data assimilation library (OpenDA) to

- a widely adopted model interface (OpenMI). *Environmental Modelling & Software*, v. 57, p. 76-89, 2014. <http://dx.doi.org/10.1016/j.envsoft.2014.02.008>.
- ROZANTE, J. R.; MOREIRA, D. S.; GONCALVES, L. G. G.; VILA, D. A. Combining TRMM and Surface Observations of Precipitation: Technique and Validation over South America. *Weather and Forecasting*, v. 25, n. 3, p. 885-894, 2010. <http://dx.doi.org/10.1175/2010WAF2222325.1>.
- SALAMON, P.; FEYEN, L. Assessing parameter, precipitation, and predictive uncertainty in a distributed hydrological model using sequential data assimilation with the particle filter. *Journal of Hydrology*, v. 376, n. 3-4, p. 428-442, 2009. <http://dx.doi.org/10.1016/j.jhydrol.2009.07.051>.
- SALAMON, P.; FEYEN, L. Disentangling uncertainties in distributed hydrological modeling using multiplicative error models and sequential data assimilation. *Water Resources Research*, v. 46, n. 12, 2010. <http://dx.doi.org/10.1029/2009WR009022>.
- SEO, D.; CAJINA, L.; CORBY, R.; HOWIESON, T. Automatic state updating for operational streamflow forecasting via variational data assimilation. *Journal of Hydrology*, v. 367, n. 3-4, p. 255-275, 2009. <http://dx.doi.org/10.1016/j.jhydrol.2009.01.019>.
- TUCCI, C.; ANDREOLLI, I.; REGINA, J.; HAAS, R.; COLLISCHONN, W. Previsão de Vazão afluente a um reservatório utilizando previsão quantitativa de chuva. *Revista Brasileira de Recursos Hídricos*, v. 11, n. 3, p. 55-69, 2006. <http://dx.doi.org/10.21168/rbrh.v11n3.p55-69>.
- VRUGT, J. A.; DIKS, C. G. H.; GUPTA, H. V.; BOUTEN, W.; VERSTRATEN, J. M. Improved treatment of uncertainty in hydrologic modeling: Combining the strengths of global optimization and data assimilation. *Water Resources Research*, v. 41, n. 1, p. 1-17, 2005. <http://dx.doi.org/10.1029/2004WR003059>.
- VRUGT, J. A.; TER BRAAK, C. J. F.; DIKS, C. G. H.; SCHOUPS, G. Hydrologic data assimilation using particle Markov chain Monte Carlo simulation: Theory, concepts and applications. *Advances in Water Resources*, v. 51, p. 457-478, 2013. <http://dx.doi.org/10.1016/j.advwatres.2012.04.002>.
- WEERTS, A. H.; EL SERAFY, G. Y. H. Particle filtering and ensemble Kalman filtering for state updating with hydrological conceptual rainfall-runoff models. *Water Resources Research*, v. 42, n. 9, 2006. <http://dx.doi.org/10.1029/2005WR004093>.
- XIE, X.; MENG, S.; LIANG, S.; YAO, Y. Improving streamflow predictions at ungauged locations with real-time updating: application of an EnKF-based state-parameter estimation strategy. *Hydrology and Earth System Sciences*, v. 18, n. 10, p. 3923-3936, 2014. <http://dx.doi.org/10.5194/hess-18-3923-2014>.
- XIE, X.; ZHANG, D. Data assimilation for distributed hydrological catchment modeling via ensemble Kalman filter. *Advances in Water Resources*, v. 33, n. 6, p. 678-690, 2010. <http://dx.doi.org/10.1016/j.advwatres.2010.03.012>.
- XU, X.; ZHANG, X.; FANG, H.; LAI, R.; ZHANG, Y.; HUANG, L.; LIU, X. A real-time probabilistic channel flood-forecasting model based on the Bayesian particle filter approach. *Environmental Modelling & Software*, v. 88, p. 151-167, 2017. <http://dx.doi.org/10.1016/j.envsoft.2016.11.010>.
- YAN, H.; MORADKHANI, H. Combined assimilation of streamflow and satellite soil moisture with the particle filter and geostatistical modeling. *Advances in Water Resources*, v. 94, p. 364-378, 2016. <http://dx.doi.org/10.1016/j.advwatres.2016.06.002>.
- ZHANG, Y.; HONG, Y.; GOURLEY, J. J.; WANG, X.; BRAKENRIDGE, G. R.; DE GROEVE, T.; VERGARA, H. Impact of assimilating spaceborne microwave signals for improving hydrological prediction in ungauged basins. In: LAKSHMI, V.; ALSDORF, D.; ANDERSON, M.; BIANCAMARIA, S.; COSH, M.; ENTIN, J.; HUFFMAN, G.; KUSTAS, W.; VAN OEVELEN, P.; PAINTER, T.; PARAJKA, J.; RODELL, M.; RÜDIGER, C. *Remote Sensing of the Terrestrial Water Cycle*. Hoboken: John Wiley & Sons, Inc, 2014. chap. 27, p. 439-450. <http://dx.doi.org/10.1002/9781118872086.ch27>.

Authors contributions

Karena Quiroz Jiménez: The first author runs both codes: MGB-IPH hydrological model and EnKF data assimilation program. Also, she elaborated the tables, figures and content of the present paper.

Walter Collischonn: Author of the hydrological model MGB-IPH and review of the results.

Rodrigo Cauduro Dias de Paiva: Developer of updated version of EnKF for MGB-IPH model and review of the methodology and results.