

Received October 25, 2020, accepted November 3, 2020, date of publication November 6, 2020, date of current version November 20, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.3036487

Daily Water Flow Forecasting via Coupling Between SMAP and Deep Learning

GUILHERME M. MACIEL, VINICIUS ALBUQUERQUE CABRAL, ANDRÉ LUÍS MARQUES MARCATO^{1b}, (Senior Member, IEEE), IVO C. SILVA JÚNIOR, AND LEONARDO DE M. HONÓRIO^{1b}

Electrical Engineering Department, Federal University of Juiz de Fora, Juiz de Fora 36036-900, Brazil

Corresponding author: André Luís Marques Marcato (andre.marcato@ufjf.edu.br)

This work was supported by Federal University of Juiz de Fora (UFJF).

ABSTRACT Hydrological models are essential tools to forecast daily water resources' availability, which are used to plan the short-term electrical systems' operation. However, there is a trade-off when choosing a given model. Complex models may provide good results depending on very complicated analytical and optimization procedures beyond sophisticated data, whereas simpler models offer reasonable results with much more amenable tuning approaches. To improve the quality of simpler models this article proposes the coupling of the Soil Moisture Accounting Procedure (SMAP) hydrological model with a Deep Learning architecture based on Conv3D-LSTM. In the proposed methodology, the SMAP is first optimized to obtain general parameters of the hydrographic basin. This optimized model's output is used as input to the Conv3D-LSTM estimator to provide the final results. This gray estimator model can generate fast and accurate results. Studies with the goal of forecast the natural flow seven days ahead are carried out for two large Brazilian hydroelectric plants to validate the method. The results obtained by the architecture are better than those obtained with decoupled techniques.

INDEX TERMS Deep learning, hydrological model, runoff forecasting, soil moisture accounting procedure.

I. INTRODUCTION

The generation of electric energy through hydropower plants is one of the main forms of generation applied in the world. This source is the most used in countries that have considerable water potential, due to its low operating cost, when compared to other sources [1].

Brazil is one of the countries that has an electrical matrix composed predominantly of hydropower plants. In this case, the thermoelectric plants operate in order to complement the generation deficit, when the hydropower generation and alternative sources do not supply the demand of the electric system. Thus, it is necessary to carry out studies to ensure a better use of this source and to reduce the dependence on generation by thermoelectric plants, which have higher operating costs.

In the planning stages of the operation of electrical systems composed of hydroelectric plants, it is necessary to know the

The associate editor coordinating the review of this manuscript and approving it for publication was F. R. Islam^{1b}.

availability of water resources for energy production days ahead. Thus, it is important to predict the amount of water that will be available and, consequently, the energy that can be extracted.

Based on climate forecast data, forecasts of affluent flow in rivers can be obtained through mathematical models that characterize the drainage basin in which the plant is inserted or through application of machine learning techniques.

From the point of view of hydroelectric plants, a good forecast of affluence and good planning can provide: greater generation at times when energy is more expensive; less spillage; optimization of the dispatch of the plant's machines; safety of the dam structure.

In Brazil, the electrical system is interconnected and managed by the National System Operator (ONS - *Operador Nacional do Sistema*), whose objective is the optimization and planning of energy generation and transmission, considering different time horizons.

Therefore, the definitions of the operation are developed by ONS and one of the challenges is the determination of

future flows according to the hydrological characteristics that influence the dynamics of the water along a basin: soil, relief, temperature, evapotranspiration, rain and moisture.

One of the main aspects that must be considered in the planning is the precipitation that increases the water level of the reservoirs, whereas the drought leads to an increase in generation by the thermoelectric plants due to a lower availability of generation by the hydropower plants.

Mathematical models have been studied and developed in order to improve the way a watershed is represented with the objective that the model presents a behavior close to the real one. Thus, knowing future climate forecast data and the characteristics of the basin, it is possible to obtain an approximation of the volume of water that will be available for the dispatch of the plant days ahead.

The objective and contribution of this work is the application of the Soil Moisture Accounting Procedure (SMAP) hydrological model with DNNs. The developed architecture produces a forecast of the natural inflow up to 7 days ahead and is based on a coupling between CNN in 3-Dimensions (Conv3D) and LSTM.

SMAP will be applied for processing soil data and generating input data for DNN. The SMAP model was chosen, since it is the model used by ONS for planning the entire Brazilian electrical system and is easy to be implemented. In addition, it is applied to several watersheds, which directly influences the planning, operation and price of energy throughout the national territory. The proposed methodology will be validated for two Brazilian regions where the HPP Peixe Angical (498 MW) and HPP Mascarenhas (198 MW) are located.

In addition to the main contribution of this work, which is an efficient architecture for predicting water flow, some important contributions are presented below, which were also developed throughout this article and which should be highlighted:

- application of precipitation data for the MERGE/CPTEC product, instead of the application of data observed by pluviometric stations;
- calibration structure of the SMAP model considering historical data for the representation of watersheds;
- local SMAP calibration at each forecast stage with a refinement of the parameters that provide a better characterization of the watershed in moments close to the forecast. This calibration is performed by the adapted Twiddle algorithm;
- precipitation data processing in grids through convolutional layers in three dimensions, aiming at a compaction of information and a space-time compensation of each measurement.

Finally, this article contributes, not only to studies that aim to make better use of HPP, but also to studies that aim to predict the availability of water resources in river basins.

The remainder of the paper is divided as follows. Section II shows related works and the state of the art regarding hydrological predictions. Section III describes the proposed

methodology and its respective steps. Section IV presents the data considered for the simulations and the respective results and analyzes. Finally, Section V presents a summary of the research conclusions and possible improvements for future validations.

II. RELATED WORKS

In recent years, several studies have been developed in order to predict hydrological behavior. In these studies, the learning machine techniques are the most used, mainly in works that have the objective of predicting river flow [2], [3]

In [4], water availability was predicted by forecasting the water level with the application of the Variable Infiltration Capacity (VIC) hydrological model for the Mekong River in Asia. In [5], a Feed Forward Neural Network (FFNN) was applied and compared with Long Short-Term Memory (LSTM). Its aim was also to forecast water levels, but applied to Polish lakes. In [6], Multilayer Perceptron (MLP) was applied for the same purpose, but for a navigable river and, in [7], the authors apply the level forecast for a HPP and point out that this study also contributes to ensuring the safety of the dam structures.

In addition to the water level, studies to predict precipitation are also very common. In [8] artificial neural networks (ANNs) and the Support Vector Machine (SVM) were applied, together with wavelet transformation for preprocessing of the input data.

Rasouli *et al.* [9] also apply Deep Learning (DL) techniques in order to predict river flow up to 7 days ahead. The techniques used are SVR, Bayesian Neural Network (BNN), Gaussian Process (GP) and Multiple Linear Regression (MLR) and are compared with respect to efficiency in forecasting.

Yaseen *et al.* [3] present an extensive review of the literature and demonstrate the applicability of DL techniques for predicting water flow in rivers and other hydrological aspects. In addition, they apply the SVR to compare to an extreme learning machine (ELM) model in predicting water flow in a Malaysian river.

The available rivers water flow is highly dependent on climatic aspects and must be considered during the forecasting process. In [10], Rokaya *et al.* show that the freezing and thawing of rivers impact the observed inflows in cold regions. For the forecast considering these aspects, they applied artificial neural networks (ANN) and showed that this tool is also widely used when considering the freezing effects of rivers.

In [11], water flow forecasting was also applied considering low temperature regions. However, the authors' approach considers snow melting from winter season and applies a sensitivity analysis of the initialization of soil moisture parameters.

The application of a Deep Belief Network (DBN) with variational mode decomposition (VMD) was the strategy applied in [12] in order to predict the flow of the Han River basin, in China. For the DBN calibration, the improved particle swarm optimization (IPSO) was used and the forecast was

made considering 1, 3, 5 and 7 days ahead. The same region was the subject of a study in [13], in which the LSTM was applied with the VMD.

In [14], an eigenmodel was used to model the behavior of flows in a basin as a function of precipitation and evapotranspiration. The model parameters were estimated based on probability density functions using a Markov chain Monte Carlo method.

Belvederesi *et al.* [15] apply the single-input sequential adaptive neuro-fuzzy inference system (ANFIS) to predict flow in a Canadian river. In [16], the authors apply a Convolution Regression based on Machine Learning (CRML) to predict the water flow of a river located in a Chinese hydrological basin.

Different mathematical models that represent a watershed have been applied in some studies. In [17] the following models are compared for river flow forecasting: Sacramento Soil Moisture Accounting (SAC-SMA), *modèle du Génie Rural à 4 paramètres Journalier (GR4J)*, McMaster University Hydrologiska Byråns Vattenbalansavdelning MAC-HBV, the Hydrologic Engineering Center Hydrologic Modeling System (HEC-HMS) and the University of Waterloo Flood Forecasting System (WAT-FLOOD). The Dynamically Dimensioned Search (DDS) method was used for the calibration of all models.

In [18], the Soil Moisture Accounting and Routine (SMAR) was compared to the following DL strategies in relation to the flow forecasting accuracy of the Heihe River in China: back-propagation neural network (BPNN), general regression neural network (GRNN) and a rotated GRNN (RGRNN) model.

Other approaches have also been the subject of studies for the prediction of flows. The authors of [19], [20] apply a rain-runoff model to determine soil status for short-term water flow forecasting with strategies that apply the ensemble Kalman filter (EnKF) to assimilate soil data. Reference [21] is also an example of the application of this method in problems of forecasting hydrological aspects, but with a focus on water quality. The authors of [22] apply EnKF to assimilate data for river flow forecasting using the Soil and Water Assessment Tool (SWAT) hydrological model.

In [23], the Mallow's coefficient was used with the M5Tree and MARS preprocessing data models for river flow forecasting. In [24], a technique based on chaos theory was applied in conjunction with a variation of genetic programming, with the same objective.

The application of river flow forecasting for regions where there are HPP can lead to a better use of the energy that can be generated. These studies also allow planning for the dispatch of machines, in addition to guaranteeing a future perspective regarding the price of energy for commercialization [25].

In [26], ANNs were applied with the objective of predicting the water flow 3 days ahead for a river in Slovenia, where a small capacity HPP is installed. In [27], an approach was proposed in which a probability density function was obtained from the Folker-Planck-Kolmogorov (FPK)

equation resolution. The main objective was to obtain the forecast of the monthly inflows of the Betania Hydropower reservoir, in a river in Colombia.

The coupling of techniques has been used in several studies in order to develop models for predicting hydrological aspects. The idea of this coupling is to obtain a greater forecasting capacity instead of applying the techniques separately. In [28], the predictability of droughts in different regions of China was investigated using a series of statistical, dynamic and hybrid models. Statistical and Dynamic models were coupled, through Bayesian model averaging that achieved the best results.

Zhou *et al.* [29] developed a forecast 4 days ahead of the inflow of the Three Gorges HPP reservoir. In this study, the unscented Kalman Filter (UKF) was applied with two DL techniques: BPNN and a non-linear auto-regressive with exogenous inputs recurrent neural network (NARX).

Machine learning techniques are widely used in isolation and in these models where there are couplings of different tools. In [30], the authors applied machine-learning quantile regression algorithms for probabilistic hydrological post-processing of flow calculated by the GR4J model.

A Brazilian hydrological basin was the subject of a study in [31], in which water flow was forecast in a river in which a 396 MW HPP is installed. The strategy used by the authors consisted of a forecast through stochastic optimization.

The Soil Moisture Accounting Procedure (SMAP) rain-flow model is one of the tools used by ONS to create flow scenarios for Brazilian basins. This deterministic and centralized model directly influences the planning of the operation of the electrical system and the dynamics of the Brazilian energy market.

In [32] the SMAP model was applied for the Paraopeba River region and the calibration was performed using the Dynamically Dimensioned Search (DDS) and Shuffled Complex Evolution (SCE) algorithms.

In [33] the Genetic Algorithm - AG - was applied to calibrate the parameters of the São Francisco River basin. The GA was identified as a robust tool with the capacity to obtain calibrations with a high degree of adherence between the observed and calculated flow curves.

In [34], the SMAP model was applied in a region of the Brazilian Northeast semiarid. The model calibration was performed using versions of AG, PSO and hybrid models with the Nelder Mead algorithm. In [35], the authors applied the model for the monthly forecast of water flows for HPP Água Vermelha. The model is applied in [36], to make monthly rain forecasts in the Três Marias basin (Brazil). The SMAP model can be used in short-term forecasts in its daily version. In [37] SMAP was used with the Weather Research and Forecasting (WRF) model to forecast floods in Rio de Janeiro.

The authors in [38] coupled a conceptual hydrological model in series with a machine learning tool. The research proposes a hybrid model through a post-processing of the state variables of the hydrological model named Tank Model through the SVM.

As can be seen in the state-of-the-art related papers, most of the work regarding hydrological prediction focus on hydrological models or learning strategies. However, as hydrological models are mathematical representation of a real and complex systems they provide just a good approximation of the problem.

On the other hand, learning strategies tries to map the entire system by co-relating the input/output. In this last approach, the size and complexity of the problem generates a huge solution space to be mapped and, as consequence, it is very difficult to any approach to have the necessary amount of data to proceed with a correct estimation.

Nevertheless, the idea presented by this article is to use the generalizability of deep learning approaches in smaller solution spaces. The hypothesis is that simpler hydrological models can have decent representation of a given region. This representation can explain and respond for several situations reducing the non mapped solution space. This new reduced space may be mapped by powerful deep learning approaches. This coupled strategy provides fast training models with excellent results.

III. PROJECT DEVELOPMENT

This work proposes the application of the SMAP model to assess the state of the soil reservoirs, which will be considered as input to the DNN. Based on a trained DNN and a well-calibrated SMAP model, the flow in the basin can be calculated as a function of precipitation and evapotranspiration. In the following sections, the methods applied and the structure of the developed tool will be described.

A. THE SMAP MODEL

The SMAP (Soil Moisture Accounting Procedure) model was originally developed by Lopes (1982) [39] which describes mathematically, in a simplified way, the behavior of watersheds. In this model, the water flow is calculated through flows from reservoirs that characterize different elements of the basin.

As the model was applied in studies, modifications were proposed in order to make it more suitable for a better characterization of the behavior of the water. In this work, the version used by ONS for planning studies of the Brazilian electrical system will be applied [40].

One of the input data for the SMAP model is the average rainfall observed in the basin for the period under analysis. The total rainfall in the basin is calculated as the average of the measurements of pluviometric stations located in the region. as follow:

$$Pb(t) = P1(t)ke_1 + P2(t)ke_2 + \dots + Pn(t)ke_n \quad (1)$$

in which $Pb(t)$ is the average precipitation observed in the basin, in the instant t (mm/day); $P1(t)$, $P2(t)$, ..., $Pn(t)$ are the precipitation observed at the stations pluviometric considered in the basin; and ke_1 ; ke_2 ; ...; ke_n are the spatial representation coefficients of each pluviometric station and

respect the following condition:

$$ke_1 + ke_2 + \dots + ke_n = 1 \quad (2)$$

Then, the model calculates the average precipitation of the day t ($Pd(t)$), which is defined by a weighted average of the observed precipitation of the days close to t , as shown in the following equation:

$$Pd(t) = [Pb(t-n) \dots Pb(t) Pb(t+1) Pb(t+2)] \cdot \begin{bmatrix} kt_{-n} \\ kt_{-n+1} \\ \vdots \\ kt_0 \\ kt_{t+1} \\ kt_{t+2} \end{bmatrix} \quad (3)$$

in which kt_{-n} , kt_{-n+1} , kt_0 , kt_{t+1} and kt_{t+2} are the time representation coefficients and n is the number of days in the past that will be considered in this time weighting.

Finally, the precipitation considered $P(t)$ is obtained by multiplying $Pd(t)$ by the factor $Pcof$, as shown in (4). $Pcof$ must be calibrated along with the other parameters, in order to guarantee the water balance.

$$Precp(t) = Pd(t) \cdot Pcof \quad (4)$$

However, rainfall data from rainfall stations are susceptible to measurement errors and failure to acquire and record data. In addition, their positions may not be the most appropriate from the point of view of the basin. Therefore, in this study, the precipitation data considered will be obtained by the product MERGE/CPTEC [41].

The precipitation provided by the MERGE product is freely accessible and is made available by the Brazilian National Institute for Space Research (INPE) [43]. The data are based on observations from GPM satellite and rainfall stations. In this product, the world is divided into small 10 km grid cells and daily average precipitation data is provided for the grid center. Through geographic coordinates, data related to regions of interest can be selected. This observed precipitation presents one more advantage in relation to the data obtained through stations, since the meteorological forecast products also present data in this format. Thus, the training and forecasting stages will present the same input data structure. As the MERGE product grids occupy the entire basin area in a uniform manner, spatial weighting is not necessary as is the case when data from sparsely located rainfall stations are used. Thus, this study uses the average of precipitation data from the grid located in the region of interest as input to the SMAP model, as it is a more robust approach and less susceptible to errors.

In addition to precipitation data, SMAP also requires daily evapotranspiration information to characterize the basin and flow calculations. In this work, the Hargreaves and Samani [42] method will be applied. In this methodology, only the maximum and minimum values of daily temperatures, which

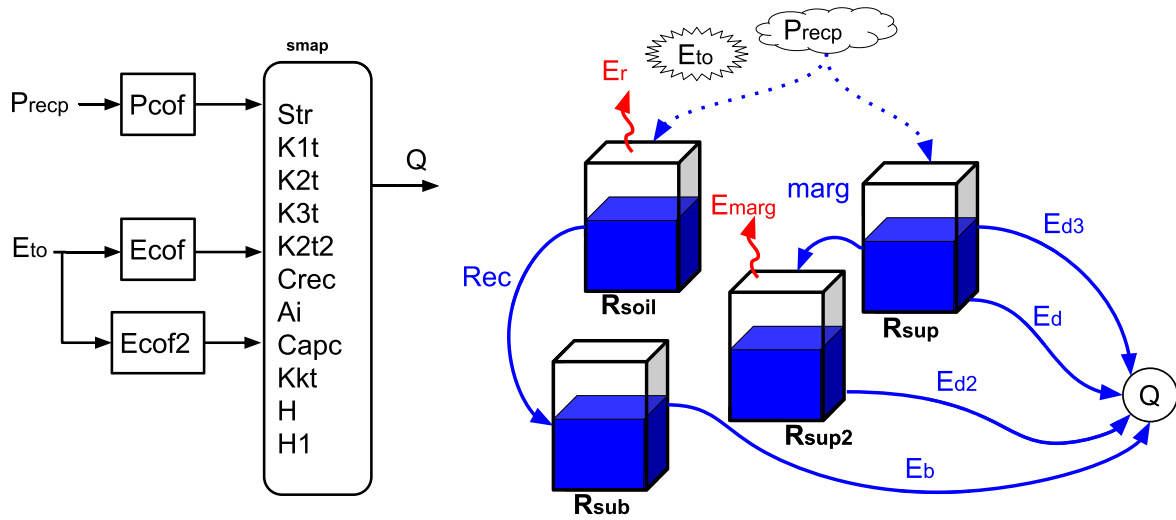


FIGURE 1. Parameters and schematic of the deterministic model Soil Moisture Accounting Procedure (SMAP).

were obtained from the National Institute of Meteorology (INMET) [44], are needed.

The daily SMAP model applied by ONS considers the representation of the basin by four reservoirs (R_{soil} , R_{sup} , R_{sup2} e R_{sub}). The model for this version is shown in Figure 1, The model of this version is shown in figure 1, in which the necessary parameters for the representation of the basin are also identified.

Potential evapotranspiration is responsible for reducing the R_{soil} and R_{sup2} reservoirs. After calculating the E_{to} evapotranspiration based on the described methodology, the values of E_{pr} and E_{pmarg} for the respective reservoirs are obtained as shown in the following expressions:

$$E_{pr} = E_{to} \cdot Ecof \tag{5}$$

$$E_{pmarg} = E_{to} \cdot Ecof2. \tag{6}$$

where $Ecof$ e $Ecof2$ are adjustment coefficients that must be calibrated.

At the beginning of the simulation, it is necessary to define the initial states of the reservoir. Thus, the values of the basic flow $E_b(0)$, the moisture content $Tu(0)$ and the superficial flow $Sup(0)$ must be defined, to obtain the initial states of the reservoirs R_{soil} , R_{sub} and R_{sup} , as shown in the following equations.

$$R_{soil}(0) = \frac{Tu(0) \cdot 100}{Str} \tag{7}$$

$$R_{sub}(0) = \frac{Eb(0)}{(1 - kk) \cdot AD} \cdot 86.4 \tag{8}$$

$$R_{sup}(0) = \frac{Sup(0)}{(1 - k2) \cdot AD} \cdot 86.4 \tag{9}$$

$$R_{sup2}(0) = 0 \tag{10}$$

in which Str is a constant that must be calibrated and AD is the drainage area of the watershed. The constants kk e $k2$

are dependent on parameters KKt and $K2t$, as shown in the following equations:

$$kk = \left(\frac{1}{2}\right)^{1/KKt} \tag{11}$$

$$k2 = \left(\frac{1}{2}\right)^{1/K2t} \tag{12}$$

Once all the parameters described have been defined, the iterative process of defining daily reservoir flows and states begins. The first parameter to be defined is the moisture content $Tu(t)$ as a function of the state of the R_{soil} reservoir, as shown in the following equation:

$$Tu(t) = \frac{R_{soil}(t)}{Str}. \tag{13}$$

Then, the runoff on the day t ($Es(t)$) is obtained, which will be different from zero, if the precipitation observed on the day t is greater than the soil abstraction Ai . The following expression shows the calculation of $Es(t)$ in this case:

$$Es(t) = \frac{(Precp(t) - Ai)^2}{(Precp(t) - Ai + Str - R_{soil}(t))} \tag{14}$$

The actual evapotranspiration $Er(t)$ will be equal to the potential $E_{pr}(t)$, if the difference $Precp(t) - Es(t)$ is greater than the potential evapotranspiration $E_{pr}(t)$. Otherwise, $Er(t)$ will be calculated as shown in the following expression:

$$Er(t) = Precp(t) - Es(t) + [E_{pr}(t) - Precp(t) + Es(t)] \cdot Tu(t) \tag{15}$$

The underground recharge $Rec(t)$ is a flow that models the water being absorbed by the soil and displacing to underground regions, represented by the underground reservoir R_{sub} . $Rec(t)$ is zero if the state of the $R_{soil}(t)$ soil reservoir is less than the product between field capacity $Capc$ and soil

saturation Str . If $Rec(t)$ is higher than the product, $Rec(t)$ is obtained as shown in the following equation:

$$Rec(t) = Crec \cdot Tu(t) \cdot (Rsoil(t) - Capc \cdot Str) \quad (16)$$

in which $Crec$ is an underground recharge constant and must also be calibrated.

The excess water that is stored on the banks of the river is represented by the $Rsup2$ reservoir and is characterized as an overflow from the $Rsup$ reservoir. To calculate this $Marg(t)$ overflow, it must be checked whether $Rsup(t)$ is greater than the $H1$ height of $Rsup$. If so, $Marg(t)$ is calculated as shown in the following equation, otherwise $Marg(t)$ is set to zero:

$$Marg(t) = (Rsub(t) - H1) \cdot (1 - k1) \quad (17)$$

in which $H1$ is included among the variables that must be calibrated and $k1$ is a dependent constant on $K1t$, as shown in the following expression:

$$k1 = \left(\frac{1}{2}\right)^{1/K1t} \quad (18)$$

Once the states of the reservoirs and the characteristic parameters of the basin are defined at the time t , the runoff from each reservoir that compose the water flow calculated by the model are calculated. The flow $Ed(t)$ is related to the reservoir $Rsup$, as shown in Figure 1, and is obtained through the following expression:

$$Ed(t) = \min [Rsup(t) - Marg(t), H] \cdot (1 - k2) \quad (19)$$

where H must be calibrated and represents a minimum height of the $Rsup$ reservoir for the runoff $Ed3(t)$ to be different from zero. $k2$ is a constant dependent on the $K2t$ parameter, as shown in the following equation:

$$k2 = \left(\frac{1}{2}\right)^{1/K2t} \quad (20)$$

The flow $Ed3(t)$ represents an existing flow when the surface reservoir $Rsup(t)$ is greater than H , which represents a high storage, characteristic of large water flows. Thus, if $Rsup(t)$ is greater than H , the flow $Ed3(t)$ is obtained from the following expression and, otherwise, $Ed3(t)$ is set to zero:

$$Ed3(t) = [Rsup(t) - H - Marg(t)] \cdot (1 - k2t2) \quad (21)$$

in which $k2t2$ is a dependent constant on $K2t2$, as shown in the following expression:

$$k2t2 = \left(\frac{1}{2}\right)^{1/K2t2} \quad (22)$$

The flows $Ed2(t)$ and $Eb(t)$, from the reservoirs $Rsup2$ and $Rsub$, respectively, can be obtained by the following expressions:

$$Ed2(t) = Rsup2(t) \cdot (1 - k3) \quad (23)$$

$$Eb(t) = Rsub(t) \cdot (1 - kk) \quad (24)$$

where $k3$ is a constant obtained by the following expression:

$$k3 = \left(\frac{1}{2}\right)^{1/K3t} \quad (25)$$

Thus, the water flow $Q(t)$ is calculated at time t using the following expression:

$$Q(t) = \frac{Ed(t) + Ed2(t) + Ed3(t) + Eb(t)}{86.4} \cdot AD \quad (26)$$

in which 86.4 is applied for unit conversion. Finally, the reservoir states for the next instant are calculated as shown in the following expressions:

$$Rsoil(t+1) = \min [Rsoil(t) + Precp(t) - Es(t) - Er(t) \\ Rec(t), Str] \quad (27)$$

$$Rsub(t+1) = Rsub(t) + Rec(t) - Eb(t) \quad (28)$$

$$Rsup(t+1) = Rsup(t) + Es(t) - Marg(t) - Ed(t) \\ - Ed3(t) + \max(0, Rsoil(t) + Precp(t) \\ - Es(t) - Er(t) - Rec(t) - Str) \quad (29)$$

$$Rsup2(t+1) = \max(Rsup2(t) + Marg(t) \\ - Ed2(t) - Emarg(t), 0) \quad (30)$$

B. HYBRID SMAP - DEEP LEARNING MODEL

Deep neural networks (DNNs) have a great capacity for modeling and pattern recognition. The main applications of DNNs in this area are for short-term water flow predictions and have observed inflows and rain predictions as inputs. In this study, the integrative hydrological model SMAP was used in parallel to provide an estimate of the distribution of water in the hydrographic basin, in order to discriminate possible redundancies related to the observed inflow values.

From the point of view of architecture training, the input variables obtained by SMAP contribute substantially. The training process of a rain-flow model based on Neural Networks consists of an optimization algorithm covering the solution region with several local minimums. The insertion of a mathematical model that partially characterizes the problem reduces the solution region and improves the training stage.

The dynamic characteristics of the hydrographic basin directly impact the observed flow and its behavior due to the precipitation in the region. Using these state variables of the SMAP model as input to the DNN gives the network better information about the basin and improves the water flow forecasting process.

This study does not use the calculated flow of the SMAP model as input data for DNN. The state variables of the SMAP model, which characterize the basin at each time t , are sent to DNN. Hence, the post-processing will be done in relation to the basin conditions and not to the observed flow, as it is commonly applied in coupling studies between hydrological models and Machine Learning techniques.

The first step of the proposed methodology is the calibration of the SMAP model based on a long history of inflow and precipitation. At the end of this process, the parameters obtained provide a good modeling of the watershed and a

good approximation between the calculated and the observed flow. It is a step that must be performed only once.

The second step is the local calibration of the SMAP, which corrects integrative errors and adjusts the parameters obtained in the previous step in order to improve the behavior of the basin over a period of 14 days and define the states of the reservoirs at the end of this period. In this step, the observed rainfall is also adjusted throughout the optimization process. Performing an iterative process, the reservoir states are obtained through an optimization for each day in the history. These states will be used as a training set for DNN.

The third stage is DNN training, which will be held for each day in the history. The observed rainfall data from 7 days ahead, the soil reservoir states and the observed flow are the input data. Water flows of 7 days ahead are the output data.

The SMAP model dynamically models the distribution of water in different soil layers, whereas the DNN has great potential for forecasting time series. The research applies the two models in order to take advantage of their intrinsic characteristics in a complementary way. Figure 2 presents an overview of the methodology, and the details of each step will be described in the following sections.

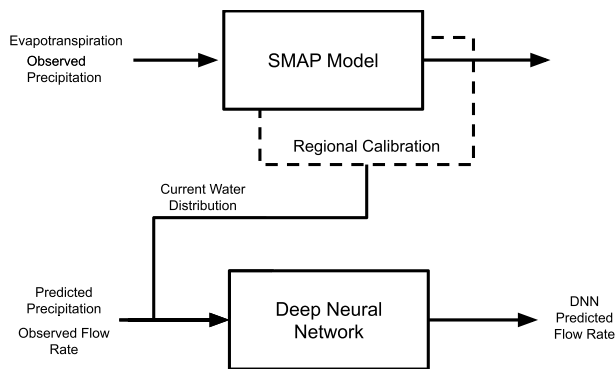


FIGURE 2. Overview of the coupling structure between the SMAP model and the deep neural networks.

1) SMAP CALIBRATION

The first step is to obtain the parameters of the SMAP model from a set of consolidated historical data of rain, flow and evapotranspiration. Some parameters of the SMAP model are dependent on geomorphological characteristics, which define specific values or intervals for each region of study. In this study, all model parameters are calibrated through the optimization process and the geomorphological characteristics are considered when defining their limits.

For this calibration, the differential evolution algorithm was used, a global search solver available in the SciPy library [46]. The default configuration of the solver was applied for this step in order to minimize the Mean Absolute Error (MAE) between the observed and calculated flow.

The behavior of the basin is dynamic and this characteristic can lead to errors when considering a model with fixed parameters. In order to adapt the method to this characteristic,

small adjustments are applied to the previously calibrated parameters for a specific period. These adjustments are made by the Twiddle algorithm [47], [48], a simple implementation method, based on local calibration and that does not require extra information about the function that must be minimized.

In addition to adjusting the model parameters, adjustments are also made to the observed precipitation data. Thus, the calculated flows will be closer to the observed flows and, therefore, the reservoir states will be better defined. Os integrative errors of the SMAP model With this local calibration, the integrative errors of the SMAP model are also mitigated. Thus, the application of simplified techniques for the estimation of daily evapotranspiration and precipitation data is sufficient for a good performance in the forecast and does not represent a limitation of the developed tool. This step is illustrated by Figure 3.

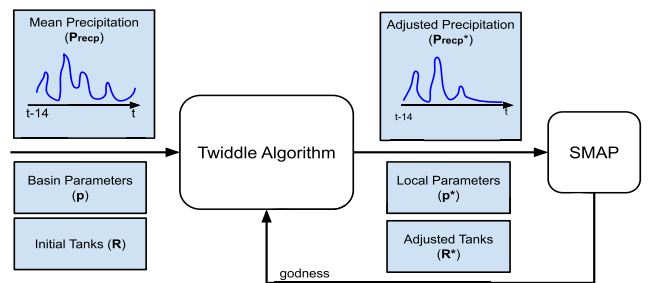


FIGURE 3. Scheme of the local SMAP calibration step via Twiddle algorithm.

Thus, considering the optimization applied for a day t , the goal of Twiddle is to improve the assertiveness between the flows observed and calculated for the period of 14 days before. As mentioned, the applied variations are smooth, so that the parameters obtained for the period are still close to those obtained in the previous step, due to the multimodal characteristic of the SMAP model. The variations allowed in this step are as follows:

- precipitation data *Precp*: $\pm 20\%$ of the observed values. This modification aims to correct possible measurement errors and ensure water balance;
- watershed parameters: $\pm 5\%$ of the values obtained in the calibration of the first stage.
- initial states of the reservoirs: $\pm 20\%$ of the values obtained for this day in the last iteration.

To apply Twiddle in the calibration of the SMAP model, some modifications were implemented in the original algorithm [47], [48], as shown below:

- 1) A scaling was performed based on the values of lower and upper bounds (lb, ub) of each variable. Thus, the variables will be limited between 0 and 1.
- 2) The update rate for each variable is restricted between $dxmin$ and $dxmax$. Thus, there is no dissipation or explosion of any rate.

The pseudo-code is shown in Algorithm 1. x_0 is the initial solution, x is the current scaled solution, dx is the array of

Algorithm 1: Bounded Twiddle Optimization

```

 $x = \text{scaling}(x_0; (lb, ub); (0, 1));$ 
 $best\_err = \text{fob}(\text{inv\_scaling}(x));$ 
while  $\text{sum}(dx) > \text{threshold}$  do
  for  $i$  in  $\text{range}(\text{len}(x))$  do
     $x_{best} = x[i];$ 
     $x[i] = x[i] + dx[i];$ 
     $x[i] = \text{constrain}(x[i]; 0; 1);$ 
     $err = \text{fob}(\text{inv\_scaling}(x));$ 
    if  $err < best\_err$  then
       $best\_err = err;$ 
       $dx[i] = dx[i] * (1 + \text{rate});$ 
       $dx[i] = \text{constrain}(dx[i]; dxmin; dxmax);$ 
    ;
    else
       $x[i] = x_{best} - dx[i];$ 
       $x[i] = \text{constrain}(x[i]; 0; 1);$ 
       $err = \text{fob}(\text{inv\_scaling}(x));$ 
      if  $err < best\_err$  then
         $best\_err = err;$ 
         $dx[i] = dx[i] * (1 + \text{rate});$ 
         $dx[i] = \text{constrain}(dx[i]; dxmin; dxmax);$ 
      else
         $x[i] = x_{best};$ 
         $dx[i] = dx[i] * (1 - \text{rate});$ 
         $dx[i] = \text{constrain}(dx[i]; dxmin; dxmax);$ 
      end
    end
  end
end
return  $\text{inv\_scaling}(x)$ 

```

x variations, $dxmin$ and $dxmax$ are set to 10^{-4} and 10^{-2} , respectively, and $\text{fob}()$ is the function that calculates the MAE between the observed and calculated flow through the parameters in x .

2) DNN INPUT DATA

After the calibration step of the SMAP model, the states of the reservoirs $R_{soil(t)}$, $R_{sub(t)}$, $R_{sup2(t)}$ and $R_{sup(t)}$ obtained for each instant t will be used as input data for a DNN. The forecasting step requires that the DNN input data be the observed flow rate, as they are normally applied in post-processing studies using Deep Learning techniques. In addition, the purpose of this work is that other dynamic characteristics of the basin are also used as input of the DNN, in order to better define the conditions in which the observed water flow occurred.

Therefore, at each forecast step, a local calibration of the SMAP model must occur in order to obtain the state variables $R_{soil(t)}$, $R_{sub(t)}$, $R_{sup2(t)}$ and $R_{sup(t)}$ for time t . This calibration aims to guarantee a better water balance of the model, a smaller error of these dynamic characteristics of the basin and an attenuation of integrative errors of the model.

In addition to the data obtained by SMAP, the DNN also has input data: the observed flow, which will be the starting point for the forecast curve; the daily precipitation data, in matrix format, with the concatenation of the data from the previous day $t - 1$ to the data of seven days ahead $t + 7$. Precipitation data from the previous day is considered, since the dynamics of the basin at a given time may be associated with the precipitation observed at previous times.

The MERGE product data provides daily rainfall in a matrix format, where each value is associated with rain in the center of each cell. For the treatment of these data, the first step is to identify the cells that are in the region of interest through the latitude and longitude coordinates. Thus, a cut is made in the MERGE data based on the minimum and maximum coordinates of the basin. Then, the squares that are outside the contour of the basin are identified. These boxes are not of interest to the study and, therefore, their values will be changed to 0 so that DNN ignores these points in the input matrix. Figure 4 shows the precipitation data entry of the model, forming a rectangular arrangement in three dimensions.

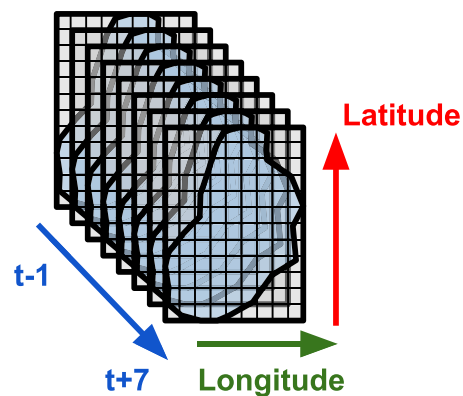


FIGURE 4. Structure of the matrix data provided by the MERGE/CPTEC product.

In the methodology validation stage, data of observed precipitation will also be considered for the calculation of the flow 7 days ahead and will be compared to the flow observed in the basin in the same period. However, the application of the method as a forecasting tool must be based on forecasting data. Using a network that processes data in the format provided by MERGE represents an advantage over the application of telemetric station data, since the forecast data is also provided in the same format.

3) DNN ARCHITECTURE

As previously described, after the treatment of precipitation data through geographic coordinates, a set of matrix data is generated and will be applied as input data for the developed architecture.

Due to the large size of the precipitation data, three 3D convolutional layers (Conv3D) were used interspersed with

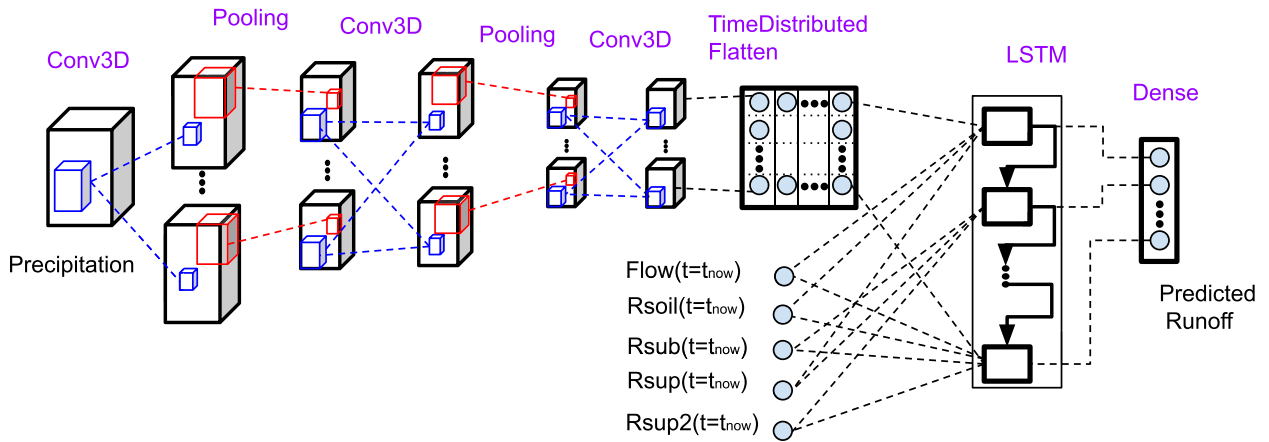


FIGURE 5. Scheme of the developed architecture. Input data: rainfall data and soil reservoir states. The architecture has 3 Conv3D + Pooling layers and a TimeDistributed Flatten layer for processing precipitation data. The LSTM layer processes the data so that a DNN can forecast the runoff data 7 days ahead.

pooling layers. In this step, the objective is to gradually reduce the size of the data and filter the relevant information. In addition, the spatial and temporal disposition of each grid is also considered in this processing, as there are clusters of data located nearby.

A TimeDistributed Flatten layer is applied to the output of the last Conv3D, in order to temporally separate the results obtained from the previous layers. It is important to highlight that this processing on the precipitation data is equivalent to the procedure applied by the SMAP model described in Equations 1 to 3. That is, the layers described process the temporal and spatial differences in the model’s response.

The output from the TimeDistributed Flatten layer will be applied as input data to an LSTM layer, for temporal processing. In addition, the reservoir states at the time t_{now} of execution are also input data for this layer. The applied LSTM layer is the current state-of-the-art. It is a very applied tool for forecasting time series and widely described and applied in other works [4], [5], [13].

The processing of data by the LSTM layer obtains results that will be applied as input to a DNN layer, the results of which are the expected flows 7 days ahead. Figure 5 illustrates the architecture described.

The Relu activation function was applied in all layers due to its simplicity and speed in training. The parameters used for each layer were defined based on empirical and conceptual studies, being dependent on the size of the basin. Thus, the neuron and filter data will be presented in Section III.

4) TRAINING

The cross-validation method was used with the split series series [45] for training and validation of the developed architecture. The training, validation and test sets are divided into chronological order. If the sets were randomly divided, the integrative characteristics and the time dependence of the system could reduce the tool’s forecasting potential.

In the Time Series Split method, the training step starts with a portion of the data. The data is added with more recent samples at each run and the neural network is retrained. At the end of the process, the average performance of the test sets for all iterations is obtained.

From the data history, the network is always validated with the last 14 samples, whereas the rest of the history is used for training. The number of samples in the validation was defined empirically and with the purpose of reducing the distance between the moment of the forecast and the training set.

At the end of each training and validation, a test is applied at the instant t in order to predict flows up to 7 days ahead. In this study, the simulations presented aim at a comparison between the flow predicted by the developed architecture and that observed for the same period. Thus, the execution of the tool will be carried out in such a way as to foresee intervals of 1 week, until data for comparison are obtained over the entire specified period.

The size of the training history will be increased by 7 dice, each run, whereas the validation and test intervals will be shifted. Figure 6 illustrates the behavior of the training, validation and test intervals throughout the runs.

The Adaptive Moment Estimation (Adam) method was the optimizer used for training the architecture. The metric used

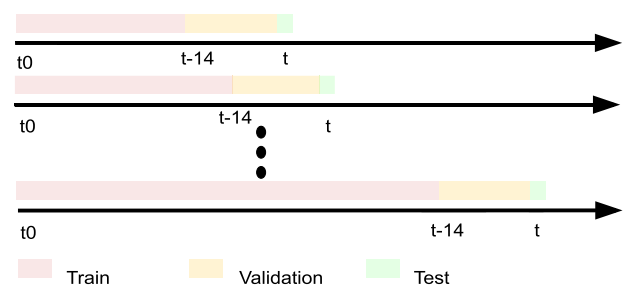


FIGURE 6. Cross-Validation: behavior of training, validation and testing sets throughout executions.

for validation was also the Mean Absolute Error (MAE), as in the SMAP calibration.

In each iteration, the DNN obtained in the previous iteration is retrained with a patience of 10 epochs, that is, the process is interrupted after 10 epochs without improvements on the validation set.

C. MODELS VALIDATION

The comparison between the results obtained by the proposed architecture and the flows observed in the region will be made using statistical metrics. The metrics considered are as follows: Mean Absolute Percentage Error (MAPE), Mean Absolute Error (MAE), Nash-Sutcliffe coefficient of efficiency (NSE), Root Mean Square Error (RMSE) and Pearson's correlation coefficient (r).

The coefficient NSE is the most used to evaluate the performance of hydrological models. The r coefficient indicates the correlation between the observed and predicted flow data, the MAE represents the average of the absolute error whereas the MAPE represents the average of the absolute error relative to the observed, in percentage. Finally, the RMSE represents an average quadratic error, which weighs the errors in order to intensify those that are higher. The statistical metrics described are calculated as shown in the following expressions:

$$MAPE = \frac{100}{N} \times \sum_{i=1}^N \left| \frac{Q(i) - Q_{obs}(i)}{Q_{obs}(i)} \right| \quad (31)$$

$$MAE = \frac{1}{N} \sum_{i=1}^N |Q(i) - Q_{obs}(i)| \quad (32)$$

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^N [Q(i) - Q_{obs}(i)]^2} \quad (33)$$

$$r = \frac{\sum_{i=1}^N (Q(i) - \bar{Q})(Q_{obs}(i) - \bar{Q}_{obs})}{\sqrt{\sum_{i=1}^N [Q(i) - \bar{Q}]^2} \sqrt{\sum_{i=1}^N [Q_{obs}(i) - \bar{Q}_{obs}]^2}} \quad (34)$$

$$NSE = 1 - \frac{\sum_{i=1}^N [Q(i) - Q_{obs}(i)]^2}{\sum_{i=1}^N [Q_{obs}(i) - \bar{Q}_{obs}]^2} \quad (35)$$

where Q represents the flow forecast, Q_{obs} the observed flow, \bar{Q} and \bar{Q}_{obs} the respective averages and N the amount of data. The metrics considered have different limits and optimal values as shown in Table 1.

TABLE 1. Comparison between the limits of statistical metrics.

Métrica	Minimum	Maximum	Optimal Value
MAPE	0	$+\infty$	0
MAE	0	$+\infty$	0
RMSE	0	$+\infty$	0
r	0	1	1
NSE	$-\infty$	1	1

For the calibration of the SMAP model and training of the architecture, the MAE was applied, as this metric is not influenced by drought periods such as MAPE nor by flood periods such as RMSE. However, all metrics are used to compare the performance of the forecast and calibration.

IV. RESULTS AND DISCUSSIONS

A. CASE STUDIES DESCRIPTION

To validate the developed architecture, two Brazilian hydrographic basins were considered in which HPP are installed: HPP Peixe Angical and HPP Mascarenhas. The objective is to apply the methodology for predicting daily unimpaired inflow over its reservoirs. HPP Mascarenhas is located in the state of Espírito Santo in a region with a tropical coastal climate. It is part of the Rio Doce basin, and its drainage area has approximately 75,500 km^2 . Its operation started in 1974 and has an installed capacity of 198 MW. HPP Peixe Angical is located in the state of Tocantins, in a predominantly semi-arid region. It is part of the Tocantins - Araguaia basin, with a drainage area of approximately 121,500 km^2 . Its operation started in 2002 and has an installed capacity of 498.8 MW. Figure 7 shows the drainage areas of each basin and their respective locations. The inflow of each HPP is obtained from the National Water Agency (ANA - Agência Nacional de Águas) [49].

Each grid of precipitation data provided by the MERGE product is a square of approximately 10 km from the side. Thus, a large amount of rainfall data is identified within each basin, due to the territorial extent of the drainage area. 1,116 daily precipitation data are identified for the region of HPP Mascarenhas and 2,420 for HPP Peixe Angical.

The large amount of daily data for each region justifies the application of Conv3D and Pooling layers for treatment and processing. The definition of the characteristics of each layer is dependent on the data from each basin. Thus, the architectural parameters for each basin were obtained empirically and are shown in the Table 2.

B. SMAP CALIBRATION

The first step in the application of the developed architecture is the calibration of the SMAP model considering a long history of flows, precipitations and evapotranspiration. The objective of this stage is the definition of parameters that model the watershed by comparing the observed and the calculated water flow.

The period from 09/2012 to 11/2016 was considered for this stage. Table 3 shows the minimum and maximum limits of each parameter considered in the calibration and the values obtained in this step for each hydrographic basin.

Figure 8 shows the behavior of the calibrated and observed flow over time and the data dispersion graphs for each watershed. The statistical metrics of the data for each region are presented in Table 4.

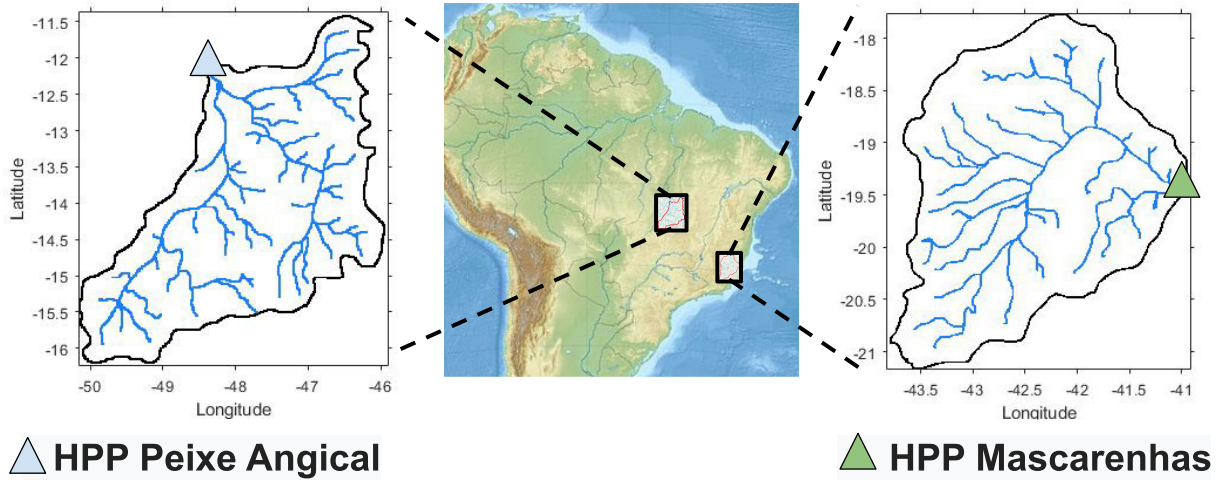


FIGURE 7. Location of the study regions and drainage area of the hydrographic basins: HPP Peixe Angical and HPP Mascarenhas.

TABLE 2. Architectural parameters for each hydrographic basin.

	Conv3D		Pool	Conv3D		Pool	Conv3D		LSTM	Dense
	Kernel	Filters	size	Kernel	Filters	size	Kernel	Filters	Units	Neurons
Mascarenhas	(2,4,3)	1	(1,2,2)	(1,4,4)	1	(1,2,2)	(1,3,3)	3	12	7
Peixe	(2,5,5)	1	(1,2,2)	(1,4,3)	1	(1,2,2)	(1,4,4)	3	12	7

TABLE 3. SMAP model parameters calibrated for each watershed.

Parameter	HPP Mascarenhas	HPP Peixe Angical	Lower Bound	Upper Bound
$Eb(0)$	290.1	361.1	0	1000
$Tu(0)$	50.7	11.31	0	100
$Sup(0)$	236.1	61.3	0	1000
$k1t$	5.32	2.72	0.1	10
$k2t$	5.01	9.37	0.1	10
$k3t$	29.3	10.6	0.1	60
$k2t2$	27.7	15.1	0.1	30
kkt	181.5	88.7	0.1	250
$Capc$	46	21.1	20	60
Ai	0.01	0.01	0.01	10
Str	157.7	234.5	0	300
$Crec$	8.56	0.29	0.1	100
$Ecof$	0.74	0.701	0.7	1.3
$Ecof2$	1.02	1.23	0.7	1.3
$H1$	88.2	53.61	0	200
H	3.4	13.9	0	200
$Pcof$	0.816	0.991	0.7	1.3

The results demonstrate that the SMAP model is a tool with excellent capacity to model watersheds. The next step is to adjust the parameters so that the representation of the basin in a shorter period is the closest to that observed. Thus, states of reservoirs that are less susceptible to errors are obtained and can be applied as input data for the proposed architecture. The next subsection presents the results obtained for the water flow forecasting step.

TABLE 4. Evaluation of water flows calculated by the calibrated SMAP model.

HPP	MAPE	MAE	RMSE	NSE	r
Mascarenhas	15.8	69.7	150.6	0.898	0.965
Peixe Angical	17.3	167.1	247.4	0.935	0.966

C. SMAP-DNN RESULTS

Initially, the entire history is calibrated by the Twiddle algorithm to obtain the reservoir states at each time point t . These data are used as the initial state of the reservoirs in the forecast period by the architecture.

For the cross-validation stage, the initial history from 09/2012 to 11/2016 was considered. The forecast was performed at intervals of 7 days. Thus, in the first run, the forecast is made from time t to time $t + 7$. In the next run, the forecast is made for time $t + 8$ to $t + 15$ and, thus, successively until 05/2020. Therefore, with each new execution, the training set increases by 7 samples and the validation and test sets are moved. For the validation of the proposed model, a long history is selected for the tool to perform the water flow forecast. This rainfall history, provided by MERGE, is an approximation of what happened in reality and is similar to prediction data in relation to the format in which they are made available and the accuracy of the data.

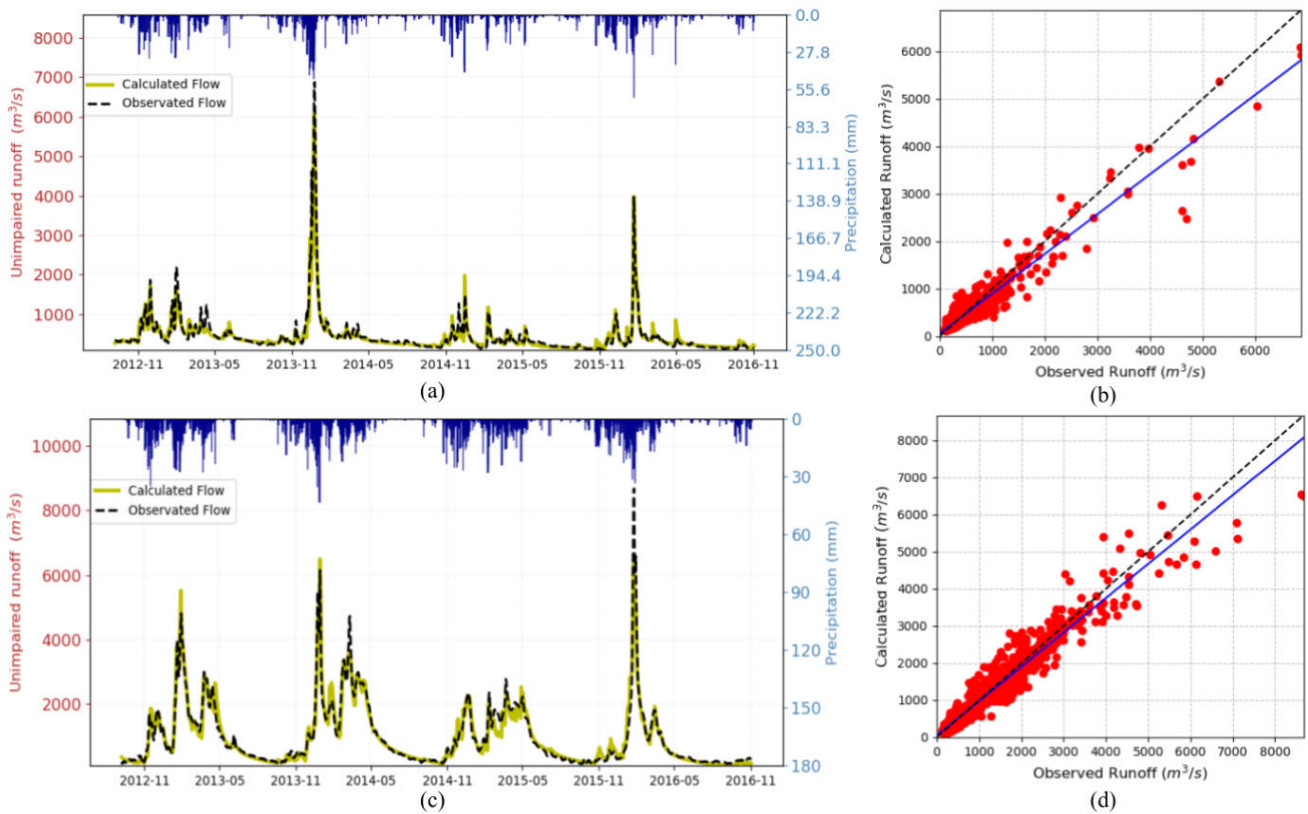


FIGURE 8. Comparison between the runoff observed and calculated by the SMAP model in the calibration step. (a) and (b) HPP Mascarenhas; (c) and (d) HPP Peixe Angical.

Hence, the MERGE data are used in this validation stage and play the role of the predicted precipitation in this period, since the validation must be done in a period in which observed flows are available for comparison with the DNN outputs.

To validate the methodology, the expected flows will be compared with the observed flow for the same period. In addition, the proposed architecture (SMAP-DNN) will be compared with two other forecasting strategies:

- the SMAP model, after local calibrations, will forecast based on future precipitation and evapotranspiration data;
- Another DNN model, with the same CONV3D-LSTM architecture, will make the forecast without applying the reservoir states as input data.

For the forecast, the SMAP model is fed with future data on rain and evapotranspiration. These data were not used for the calibration of the SMAP model or for training the network.

Due to the random nature of the training, strategies involving DL techniques were performed 20 times each. The Twiddle algorithm has a deterministic characteristic and, therefore, the SMAP model was calibrated only once at each forecast interval. The distribution of the results obtained by each technique are shown in Figure 9.

The results obtained demonstrate that the DNN has a better forecasting capacity than the SMAP model, when the two

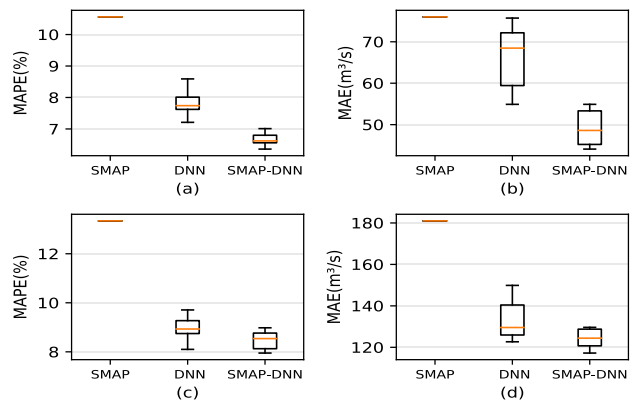


FIGURE 9. Comparison between MAPE and MAE distributions for prediction via the SMAP, DNN and DNN-SMAP model: (a) e (b) HPP Mascarenhas; (c) e (d) HPP Peixe Angical.

techniques are performed in an uncoupled manner. In addition, the graphs show that using the SMAP state variables as input data for the DNN improves the forecasting capacity, since these variables complement the flow observed with the dynamic characteristics of the basin at the time of the forecast.

The results of HPP Peixe Angical show that the application of the SMAP-DNN strategy reduces the mean and median

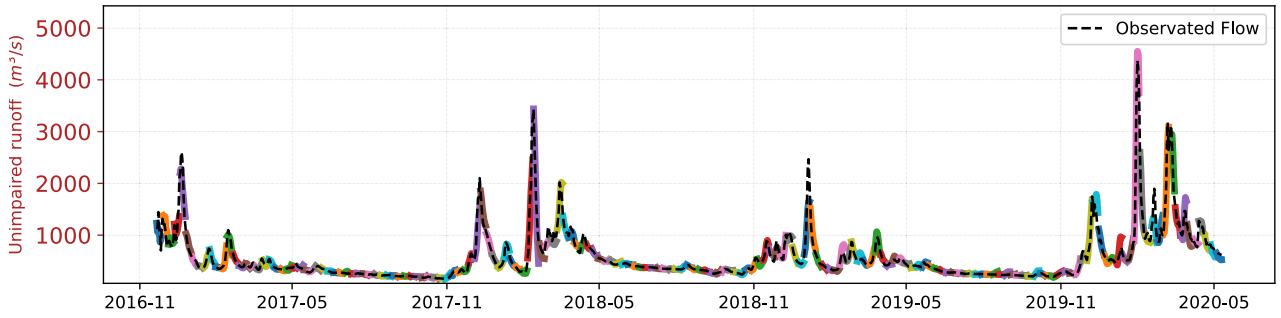


FIGURE 10. Simulation of Runoff Forecasting for HPP Mascarenhas.

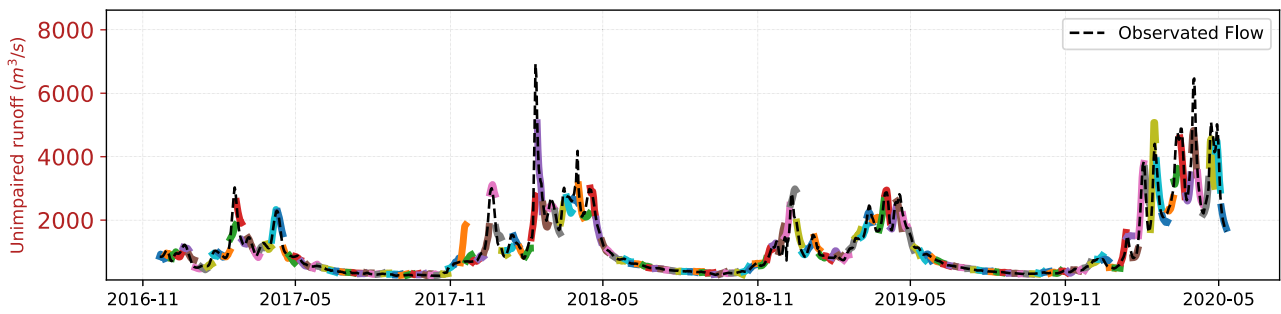


FIGURE 11. Simulation of Runoff Forecasting for HPP Peixe Angical.

of MAPE and MAE values. However, the results of HPP Mascarenhas showed more expressive reductions.

This difference between the results of the two regions can be explained by the following aspects:

- the SMAP model can better represent the region of HPP Mascarenhas, due to the intrinsic characteristics of the hydrographic basin;
- climate data for a region can be more accurate;
- characteristics of a watershed can make it more difficult to obtain optimal points through the calibration process;
- the hydrographic basin may have presented a different behavior throughout the forecast period than that observed in the calibration of the SMAP in the first step.

Despite the difference between the results of the two basins, both demonstrate that the proposed architecture is effective for the flow forecasting process in the short term. The flow data obtained by the proposed architecture are presented in Figure 10 and 11 for HPP Mascarenhas and HPP Peixe Angical, respectively. The black dotted line represents the observed flow and the colored lines represent the 7-day intervals that were generated in this forecasting step. The results show that there is great assertiveness in periods of drought and good forecast of the trend in periods of flood. This behavior can also be observed in the scatter plots of Figure 12.

Statistical metrics were also applied to assess the best forecast obtained for each region. The behavior of this forecast was analyzed according to the number of days ahead as shown in Table III-C. The results show that the correlation of HPP

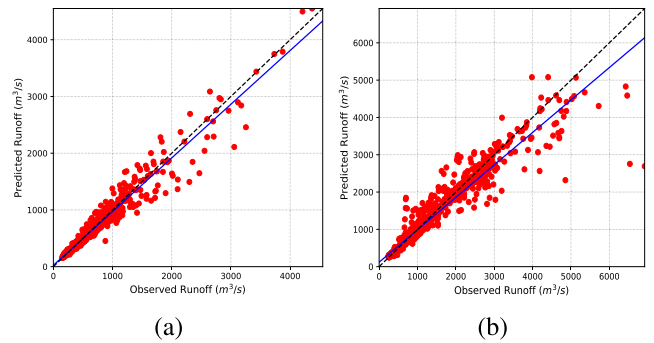


FIGURE 12. Scatter plot of observed and predicted runoff data. (a) HPP Mascarenhas; (b) HPP Peixe Angical.

Mascarenhas data is stronger than that of HPP Peixe Angical data. In addition, all metrics show that the Mascarenhas forecast performed better. However, the results obtained for both basins demonstrate that the developed architecture is an excellent tool for forecasting affluent flows. Table III-C also shows that the forecasts for the first day have a very low error and high correlation between the observed and calculated flow data. As expected, performance tends to decrease as it distances from the moment of execution. Even so, the latter also show a strong correlation between the data.

V. SUMMARY AND CONCLUSION

Forecasting the availability of water resources is an important study for several sectors, including hydroelectricity. A good estimate of future flows allows planning the operation of

TABLE 5. Statistical evaluation of the best forecast data obtained.

HPP Mascarenhas								
Metric	Total	Day1	Day2	Day3	Day4	Day5	Day6	Day7
MAPE	6.42	1.46	4.25	6.52	7.7	7.78	8.26	9.01
MAE	45.97	7.1	33.3	45.89	54.41	55.32	61.53	64.29
RMSE	93.09	11.49	69.93	86.81	105.55	113.91	131.91	132.07
r	0.979	0.999	0.987	0.983	0.978	0.976	0.968	0.964
NSE	0.958	0.999	0.974	0.964	0.954	0.951	0.937	0.929

HPP Peixe Angical								
Metric	Total	Day1	Day2	Day3	Day4	Day5	Day6	Day7
MAPE	7.82	1.61	4.01	6.22	8.39	9.95	11.78	12.82
MAE	113.12	14.66	52.56	93.21	128.42	151.62	173.76	177.62
RMSE	255.25	24.53	108.31	211.49	274.7	337.75	426.19	403.79
r	0.961	0.999	0.994	0.981	0.968	0.947	0.918	0.922
NSE	0.923	0.999	0.988	0.958	0.93	0.894	0.838	0.844

plants and the electrical system. This planning includes the forecast of generation and pouring in the short and long term. This work proposed an innovative tool to improve the assertiveness of water flow forecasts 7 days ahead in a watershed.

The methodology coupled the SMAP rain-flow model and a Deep Learning architecture Conv3D–LSTM. The Twiddle algorithm was applied for a smooth local adjustment of the SMAP model. Modifications were done in the algorithm to guarantee a gradual process and respecting the restrictions. Thus, the reservoir data was made available for architecture training in a more accurate way.

The 3D convolutional layers (Conv3D) showed great potential for obtaining spatio-temporal features of the rain. The developed tool was tested for reservoirs of two large plants in Brazilian basins with divergent characteristics. The results showed that the union of the two models showed better assertiveness in relation to the application of decoupled methods. The proposed architecture was built with the objective of defining only one scenario for each execution and this is the focus of this study. Due to the random nature of the DNN training process, stochastic analyzes can be developed, such as the use of quantile regression in training stage.

The comparison of water flows observed and predicted by the model demonstrated the existence of a strong correlation, validating the forecast stage. In addition, statistical analyzes were applied to the forecasts for each day of the forecast week. The results showed that the forecast assertiveness is greater at the beginning of the forecast. However, the past few days have presented acceptable results for planning and in accordance with the ideal limits of the statistical metrics considered.

For the simulations presented, rainfall data observed in a grid format provided by the MERGE/CPTEC product were used. For the application of the network as a forecasting tool, meteorological forecast data is required. The application of MERGE data represents an advantage from this point of view, since the forecast data has this same format.

ACKNOWLEDGMENT

The work reported in this paper was performed as part of an interdisciplinary research and development project undertaken by Federal University of Juiz de Fora (UFJF). The authors acknowledge the support of the following companies: Federal University of Juiz de Fora (UFJF), CAPES (Coordination for the Improvement of Higher Education Personnel), CNPq (National Council for Scientific and Technological Development), ANEEL (National Agency of Electrical Energy), and EDP (Energias de Portugal).

REFERENCES

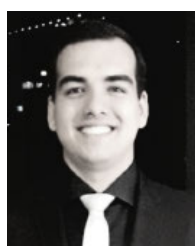
- [1] International Energy Agency (IEA). *Data & Statistics* Disponível em. Accessed: Sep. 1, 2020. [Online]. Available: <https://www.iea.org/>
- [2] S. W. Fleming and A. G. Goodbody, "A machine learning metasystem for robust probabilistic nonlinear regression-based forecasting of seasonal water availability in the US west," *IEEE Access*, vol. 7, pp. 119943–119964, 2019.
- [3] Z. M. Yaseen, S. O. Sulaiman, R. C. Deo, and K.-W. Chau, "An enhanced extreme learning machine model for river flow forecasting: State-of-the-art, practical applications in water resource engineering area and future research direction," *J. Hydrol.*, vol. 569, pp. 387–408, Feb. 2019.
- [4] C.-H. Chang, H. Lee, F. Hossain, S. Basnayake, S. Jayasinghe, F. Chishtie, D. Saah, H. Yu, K. Sothea, and D. Du Bui, "A model-aided satellite- altimetry-based flood forecasting system for the mekong river," *Environ. Model. Softw.*, vol. 112, pp. 112–127, Feb. 2019.
- [5] S. Zhu, B. Hrnjica, M. Ptak, A. Choiniński, and B. Sivakumar, "Forecasting of water level in multiple temperate lakes using machine learning models," *J. Hydrol.*, vol. 585, Jun. 2020, Art. no. 124819.
- [6] T. Zhou, Z. Jiang, X. Liu, and K. Tan, "Research on the long-term and short-term forecasts of navigable river's water-level fluctuation based on the adaptive multilayer perceptron," *J. Hydrol.*, vol. 591, Dec. 2020, Art. no. 125285.
- [7] Y. Shang, Y. Xu, L. Shang, Q. Fan, Y. Wang, and Z. Liu, "A method of direct, real-time forecasting of downstream water levels via hydropower station reregulation: A case study from Gezhouba hydropower plant, China," *J. Hydrol.*, vol. 573, pp. 895–907, Jun. 2019.
- [8] A. Belayneh, J. Adamowski, B. Khalil, and B. Ozga-Zielinski, "Long-term SPI drought forecasting in the awash river basin in ethiopia using wavelet neural network and wavelet support vector regression models," *J. Hydrol.*, vol. 508, pp. 418–429, Jan. 2014.
- [9] K. Rasouli, W. W. Hsieh, and A. J. Cannon, "Daily streamflow forecasting by machine learning methods with weather and climate inputs," *J. Hydrol.*, vols. 414–415, pp. 284–293, Jan. 2012.
- [10] P. Rokaya, L. Morales-Marin, and K.-E. Lindenschmidt, "A physically-based modelling framework for operational forecasting of river ice breakup," *Adv. Water Resour.*, vol. 139, May 2020, Art. no. 103554.
- [11] H. Bhuiyan, H. McNairn, J. Powers, and A. Merzouki, "Application of HEC-HMS in a cold region watershed and use of RADARSAT-2 soil moisture in initializing the model," *Hydrology*, vol. 4, no. 1, p. 9, Feb. 2017.
- [12] T. Xie, G. Zhang, J. Hou, J. Xie, M. Lv, and F. Liu, "Hybrid forecasting model for non-stationary daily runoff series: A case study in the han river basin, China," *J. Hydrol.*, vol. 577, Oct. 2019, Art. no. 123915.
- [13] G. Zuo, J. Luo, N. Wang, Y. Lian, and X. He, "Decomposition ensemble model based on variational mode decomposition and long short-term memory for streamflow forecasting," *J. Hydrol.*, vol. 585, Jun. 2020, Art. no. 124776.
- [14] T. Wöhling and L. Burberry, "Eigenmodels to forecast groundwater levels in unconfined river-fed aquifers during flow recession," *Sci. Total Environ.*, vol. 747, Dec. 2020, Art. no. 141220.
- [15] C. Belvederesi, J. A. Dominic, Q. K. Hassan, A. Gupta, and G. Achari, "Predicting river flow using an AI-based sequential adaptive neuro-fuzzy inference system," *Water*, vol. 12, no. 6, p. 1622, Jun. 2020.
- [16] C. Chen, Q. Hui, Q. Pei, Y. Zhou, B. Wang, N. Lv, and J. Li, "CRML: A convolution regression model with machine learning for hydrology forecasting," *IEEE Access*, vol. 7, pp. 133839–133849, 2019.
- [17] D. B. Wijayarathne and P. Coulibaly, "Identification of hydrological models for operational flood forecasting in St. John's, newfoundland, canada," *J. Hydrol., Regional Stud.*, vol. 27, Feb. 2020, Art. no. 100646.

- [18] S. Yin, D. Tang, X. Jin, W. Chen, and N. Pu, "A combined rotated general regression neural network method for river flow forecasting," *Hydrol. Sci. J.*, vol. 61, no. 4, pp. 669–682, Mar. 2016.
- [19] S. Meng, X. Xie, and S. Liang, "Assimilation of soil moisture and streamflow observations to improve flood forecasting with considering runoff routing lags," *J. Hydrol.*, vol. 550, pp. 568–579, Jul. 2017.
- [20] J. Samuel, A. N. Rousseau, K. Abbasnezhadi, and S. Savary, "Development and evaluation of a hydrologic data-assimilation scheme for short-range flow and inflow forecasts in a data-sparse high-latitude region using a distributed model and ensemble Kalman filtering," *Adv. Water Resour.*, vol. 130, pp. 198–220, Aug. 2019.
- [21] S. Loos, C. M. Shin, J. Sumihar, K. Kim, J. Cho, and A. H. Weerts, "Ensemble data assimilation methods for improving river water quality forecasting accuracy," *Water Res.*, vol. 171, Mar. 2020, Art. no. 115343.
- [22] A. Patil and R. Ramsankaran, "Improving streamflow simulations and forecasting performance of SWAT model by assimilating remotely sensed soil moisture observations," *J. Hydrol.*, vol. 555, pp. 683–696, Dec. 2017.
- [23] M. Rezaie-Balf, S. Kim, H. Fallah, and S. Alaghmand, "Daily river flow forecasting using ensemble empirical mode decomposition based heuristic regression models: Application on the perennial rivers in Iran and South Korea," *J. Hydrol.*, vol. 572, pp. 470–485, May 2019.
- [24] M. A. Ghorbani, R. Khatibi, A. Danandeh Mehr, and H. Asadi, "Chaos-based multigenetic programming: A new hybrid strategy for river flow forecasting," *J. Hydrol.*, vol. 562, pp. 455–467, Jul. 2018.
- [25] A. F. Hamlet, D. Huppert, and D. P. Lettenmaier, "Economic value of long-lead streamflow forecasts for Columbia river hydropower," *J. Water Resour. Planning Manage.*, vol. 128, no. 2, pp. 91–101, Mar. 2002.
- [26] T. Stokelj, D. Paravan, and R. Golob, "Enhanced artificial neural network inflow forecasting algorithm for Run-of-River hydropower plants," *J. Water Resour. Planning Manage.*, vol. 128, no. 6, pp. 415–423, Nov. 2002.
- [27] E. Domínguez and H. Rivera, "A fokker–planck–kolmogorov equation approach for the monthly affluence forecast of Betania hydropower reservoir," *J. Hydroinformatics*, vol. 12, no. 4, pp. 486–501, Oct. 2010.
- [28] L. Xu, N. Chen, X. Zhang, and Z. Chen, "An evaluation of statistical, NMME and hybrid models for drought prediction in China," *J. Hydrol.*, vol. 566, pp. 235–249, Nov. 2018.
- [29] Y. Zhou, S. Guo, C.-Y. Xu, F.-J. Chang, and J. Yin, "Improving the reliability of probabilistic multi-step-ahead flood forecasting by fusing unscented Kalman filter with recurrent neural network," *Water*, vol. 12, no. 2, p. 578, Feb. 2020.
- [30] G. Papacharalampous, H. Tyralis, A. Langousis, A. W. Jayawardena, B. Sivakumar, N. Mamassis, A. Montanari, and D. Koutsoyiannis, "Probabilistic hydrological post-processing at scale: Why and how to apply machine-learning quantile regression algorithms," *Water*, vol. 11, no. 10, p. 2126, Oct. 2019.
- [31] F. M. Fan, D. Schwanenberg, R. Alvarado, A. Assis dos Reis, W. Collischonn, and S. Naumann, "Performance of deterministic and probabilistic hydrological forecasts for the short-term optimization of a tropical hydropower reservoir," *Water Resour. Manage.*, vol. 30, no. 10, pp. 3609–3625, Aug. 2016.
- [32] I. S. Saraiva, "Regionalização de parâmetros de modelos de simulação hidrológica para emprego em bacias hidrográficas sem monitoramento fluviométrico," M.S. thesis, Dept. Water Resour., Federal Univ. Minas Gerais, Belo Horizonte, Belo Horizonte, Brazil, 2010.
- [33] G. R. Caponi, B. C. Silva, and A. H. M. Santos, "Calibração automática de um modelo chuva-vazão mensal por algoritmos genéticos," presented at the 21st Simpósio Brasileiro de Recursos Hídricos, 2011. [Online]. Available: <https://www.abrhydro.org.br/SGCv3/publicacao.php?PUB=3&ID=81&SUMARIO=1233>
- [34] C. Rampinelli and M. Costa, "Calibração do modelo chuva-vazão smap em uma estação do semiárido nordestino com os algoritmos pso, ag e nelder mead e seus efeitos na definição das séries de vazões médias mensais," presented at 21st Simpósio Brasileiro de Recursos Hídricos, 2015. [Online]. Available: <http://eventos.abrh.org.br/xxisbrh/>
- [35] F. D. N. R. da Silva, J. L. D. Alves, and M. Cataldi, "Climate downscaling over south america for 1971–2000: Application in SMAP rainfall-runoff model for grande river basin," *Climate Dyn.*, vol. 52, nos. 1–2, pp. 681–696, Jan. 2019.
- [36] L. F. G. De Paiva, S. M. Montenegro, and M. Cataldi, "Prediction of monthly flows for três marias reservoir (São Francisco river basin) using the CFS climate forecast model," *RBRH*, vol. 25, p. e16, Apr. 2020.
- [37] P. da Cunha Luz Barcellos and M. Cataldi, "Flash flood and extreme rainfall forecast through one-way coupling of WRF-SMAP models: Natural hazards in rio de janeiro state," *Atmosphere*, vol. 11, no. 8, p. 834, Aug. 2020.
- [38] M. Kwon, H.-H. Kwon, and D. Han, "A hybrid approach combining conceptual hydrological models, support vector machines and remote sensing data for rainfall-runoff modeling," *Remote Sens.*, vol. 12, no. 11, p. 1801, Jun. 2020.
- [39] J. E. G. Lopes, B. P. F. Braga, Jr., and J. G. L. Conejo, "SMAP—A simplified hydrologic model," in *Applied Modeling in Catchment Hydrology*. Littleton, CO, USA: Water Resources Publications, 1982, pp. 167–176.
- [40] Operador Nacional do Sistema Elétrico—ONS. *Manual do Usuário—Aplicativo SMAP*. Accessed: Jan. 2019. [Online]. Available: <http://www.ons.org.br/>
- [41] J. R. Rozante, D. S. Moreira, L. G. G. de Goncalves, and D. A. Vila, "Combining TRMM and surface observations of precipitation: Technique and validation over south america," *Weather Forecasting*, vol. 25, no. 3, pp. 885–894, Jun. 2010.
- [42] G. H. Hargreaves and Z. A. Samani, "Reference crop evapotranspiration from temperature," *Appl. Eng. Agricult.*, vol. 1, no. 2, pp. 96–99, 1985.
- [43] Instituto Nacional de Pesquisas Espaciais—INPE. *MERGE-GPM Daily Precipitation [Database]*. Accessed: Aug. 2020. [Online]. Available: <http://ftp.cptec.inpe.br/modelos/tempo/MERGE/GPM/DAILY/>
- [44] Instituto Nacional de Meteorologia—INMET. *Weather Database*. Accessed: Aug. 2020. [Online]. Available: <https://bdmep.inmet.gov.br/>
- [45] Scikit-Learning Modules. *Cross-Validation: Evaluating Estimator Performance*. Accessed: Aug. 2020. [Online]. Available: https://scikit-learn.org/stable/modules/cross_validation.html
- [46] SciPy. *SciPy Reference Guide, Differential Evolution*. Accessed: Aug. 24, 2020. [Online]. Available: https://docs.scipy.org/doc/scipy/reference/generated/scipy.optimize.differential_evolution.html
- [47] CS373: *Artificial Intelligence for Robotics*, Udacity, San Francisco, CA, USA, 2018.
- [48] Udacity. *Twiddle—Artificial Intelligence for Robotics [Streaming Video]*. Accessed: Aug. 2020. [Online]. Available: <https://www.youtube.com/watch?v=2uQ2BSZDvXs>
- [49] Agência Nacional de Águas. *Sistema Nacional de Informações Sobre Recursos Hídricos*. Accessed: Aug. 2020. [Online]. Available: <http://www.snirh.gov.br/hidroweb/apresentacao>



GUILHERME M. MACIEL received the degree in electrical engineering with emphasis on robotics and automation and the master's degree in electrical engineering in the field of electrical energy systems from the Federal University of Juiz de Fora, in 2016 and 2018, respectively, and the degree in mathematics from the Claretiano University Center, in 2020. He is currently pursuing the Ph.D. degree. He has worked as an Engineering Intern with ArcelorMittal Brazil, from 2015 to 2016.

He is currently a Professor of robotics with the Military College of Juiz de Fora, and a Researcher with INESC P&D Brazil. He has also been a collaborating member of the Local Robot Competition Team called Rinobot, since 2018. His research interests include optimization, machine learning, assistive robotics, and educational robotics.



VINICIUS ALBUQUERQUE CABRAL received the B.Sc. and M.Sc. degrees in electrical engineering from the Federal University of Juiz de Fora (UFJF), Juiz de Fora, Brazil, in 2017 and 2020, respectively, where he is currently pursuing the degree with the Postgraduation Program of Electrical Engineering (PPEE). He is currently a member of the Heuristic and Bioinspiration Optimization Group (GOHB), UFJF. His research interests include optimization, evolutionary algorithms, and power system analysis.



ANDRÉ LUÍS MARQUES MARCATO (Senior Member, IEEE) received the master's and Ph.D. degrees in electrical engineering from the Pontifical Catholic University of Rio de Janeiro. He is currently an Electrical Engineer with the Universidade Federal de Juiz de Fora (UFJF), Brazil, where he is also a Full Professor and the Head of the Electrical Energy Department. He has visited the Imperial College London and the Faculdade Engenharia da Universidade do Porto, as a Postdoctoral Student, in 2012. His research interest includes optimization techniques applied to hydrothermal coordination, operation, and expansion planning electrical systems. He coordinates or coordinated research projects with many companies in the Brazilian Electrical Sector, for example, Petrosbras, CEPEL, Light, CESP, Cemig, and Duke Energy.



LEONARDO DE M. HONÓRIO received the B.Sc. degree from the Federal University of Juiz de Fora (UFJF), in 1993, and the M.Sc. and Ph.D. degrees from EFEI, Brazil, in 1999 and 2002, respectively, all in electrical engineering. He was a Visiting Researcher with Porto and Irvine California University, in 2006 and 2012, respectively. He is currently an Associate Professor with UFJF. His current research interests include evolutionary algorithms, probabilistic methods, optimal power flow, robotics, autonomous vehicles, fuzzy logic, pattern recognition, and optimization.

• • •



IVO C. SILVA JÚNIOR received the B.Sc. and M.Sc. degrees from the Federal University of Juiz de Fora (UFJF), Brazil, in 2001 and 2003, respectively, and the D.Sc. degree from COPPE, Federal University of Rio de Janeiro, Rio de Janeiro, Brazil, in 2008. Since 2009, he has been a Professor with the Electrical Engineering Department, UFJF. His research interests include power economics, optimization, and power systems analysis.