Revista Brasileira de Recursos Hídricos
Brazilian Journal of Water Resources
Versão On-line ISSN 2318-0331
RBRH, Porto Alegre, v. 28, e47, 2023
Special Session

https://doi.org/10.1590/2318-0331.282320230116

GHCen: a stochastic-conceptual approach for generating synthetic streamflow scenarios

GHCen: uma abordagem estocástica-conceitual para geração de cenários sintéticos de vazões

Felipe Treistman[1], Lucas de Souza Khenayfis[1] & Débora Dias Jardim Penna[1]

[1]Operador Nacional do Sistema Elétrico, Rio de Janeiro, RJ, Brasil
E-mails: felipe.treistman@ons.org.br (FT); lkhenayfis@ons.org.br (LSK); debora.jardim@ons.org.br (DDJP)

Received: October 06, 2023 - Revised: October 13, 2023 - Accepted: October 14, 2023

ABSTRACT

In Brazil, energy production predominantly relies on hydropower generation, necessitating precise hydrological planning tools to manage the uncertainty inherent in river flows. While traditional hydrological models provide valuable deterministic forecasts, addressing the need for probabilistic information remains crucial. This paper introduces a novel approach, the Hybrid Generator of Synthetic Streamflow Scenarios (GHCen), which combines a conceptual SMAP/ONS model with stochastic simulation techniques to generate synthetic streamflow scenarios. The stochastic methodology employed in GHCen effectively reproduces the key characteristics of precipitation processes on daily to annual scales. Through a comprehensive case study, conducted for 2021, GHCen demonstrates its capability to accurately replicate the hydrological behaviors from historical data. The analysis reveals a strong alignment between the synthetic scenarios and observed Natural Energy Inflow for the National Interconnected System, both monthly and in accumulated terms.

Keywords: Synthetic streamflow scenario generation; Hybrid model; Conceptual rainfall-runoff model.

RESUMO

No Brasil, a produção de energia depende predominantemente da geração hidrelétrica, necessitando de ferramentas precisas de planejamento hidrológico para gerenciar a incerteza inerente às vazões afluentes. Embora modelos hidrológicos tradicionais forneçam previsões determinísticas valiosas, endereçar a necessidade de informações probabilística continua a ser crucial. Este artigo apresenta uma nova abordagem, o Gerador Híbrido de Cenários Sintéticos de Afluências (GHCen), que combina um modelo conceitual SMAP/ONS com técnicas de simulação estocástica para gerar cenários sintéticos de vazões. A metodologia estocástica empregada no GHCen reproduz efetivamente as principais características dos processos de precipitação em escalas diárias a anuais. Através de um estudo de caso abrangente, realizado para 2021, o GHCen demonstra a sua capacidade de replicar com precisão os comportamentos hidrológicos a partir de dados históricos. A análise revela um forte alinhamento entre os cenários sintéticos e a Energia Natural Afluente observada para o Sistema Interligado Nacional, tanto mensalmente quanto em termos acumulados.

Palavras-chave: Geração de cenários sintéticos de vazões; Modelo híbrido; Modelo conceitual chuva-vazão.

INTRODUCTION

Despite the growing penetration of new renewable energy sources such as solar and wind, most of the Brazilian energy production still comes from hydropower plants. For this reason, the Brazilian ISO (called ONS) has a significant need for accurate representations of the uncertainty regarding river flows in its hydro plants for medium- and long-term planning operations.

Accurate runoff forecasts are achievable by conceptual hydrological models, which mimic the real soil and streamflow dynamics and tend to perform well (Ávila et al., 2023). Although effective, such forecasts only provide a single estimate value. Probabilistic forecasts fill this gap by providing additional information about the uncertainty of some pointwise forecasts. Some methods extend hydrological models directly to output probabilistic results, such as Tyralis & Papacharalampous (2021) but a more frequent approach is the employment of post-processing. This post-processing can be done either by modeling the prediction errors as in Sikorska-Senoner & Quilty (2021) and Wani et al. (2017) or by feeding the actual pointwise forecasts into another model as in Zhou et al. (2022) or by Humphrey et al. (2016) that produces either confidence intervals or complete predictive probability density functions.

An alternative approach is ensemble forecasts, where multiple possible paths of the forecast variable are simulated. The authors of Troin et al. (2021) group ensemble forecast methods into three main categories, based on the forcing type: exclusively historical streamflow data; weather forecasts from Numerical Weather Prediction (NWP) models, and historical and pseudo-historical weather data (satellite images, reanalysis data, reforecasts, etc.). The first kind encompasses all methods discussed in the previous paragraph, and indeed most post-processing approaches.

Ensembles of the second category, based on NWP forecasts, can be achieved by using ensemble weather forecasts, feeding deterministic forecasts into multiple hydrological models, or both. Ensemble forecasts, hydrological or weather, often demand post-processing to improve their skill. To this end, quantile mapping has been applied to hydrological ensembles by Wood & Schaake (2008) and to weather ensembles by Hamill & Scheuerer (2018). Also, the authors in Ye et al. (2015) propose a new method based on Generalized Linear Modelling which performs well in their case study. When combining multiple models with different levels of accuracy, the final predictive PDF is obtained most often by Bayesian Model Averaging (BMA), as in Liang et al. (2013). Another technique is Quantile Model Averaging (QMA), which is shown by Schepen & Wang (2015) to perform better when ensemble members are spread out amongst themselves.

The final category assumes past weather conditions are suitable proxies for future ones, which is mostly centered on the assumption that rainfall processes have less memory than runoff ones. In this paper, we propose a hybrid model for multivariate streamflow scenario generation based on the stochastic simulation of daily rainfall (Zhou et al., 2019), which is subsequently fed into a conceptual rainfall-runoff SMAP/ONS model (Operador Nacional do Sistema Elétrico, 2017). The model is shown to precisely capture key features of the historical data in a case study considering most of the Brazilian hydropower plants.

METHODOLOGY

The proposed methodology merges the conceptual rainfall-runoff modeling of the SMAP/ONS model with a stochastic methodology for the simulation of synthetic daily precipitation scenarios, resulting in a Hybrid Generator of Synthetic Streamflow Scenarios (GHCen). Based on the daily precipitation historical data, a stochastic model is estimated to generate daily precipitation synthetic scenarios, which are used as input to the SMAP/ONS for each subbasin, producing daily streamflow synthetic scenarios. Figure 1 shows the flowchart of GHCen.

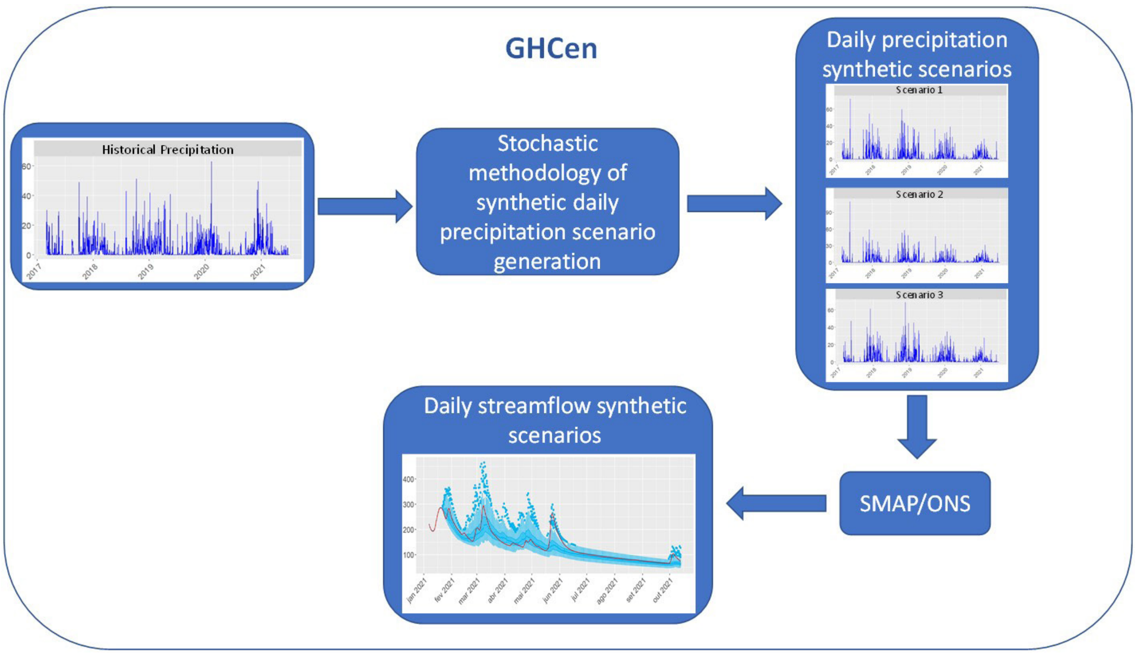

Figure 1. Flowchart of GHCen.

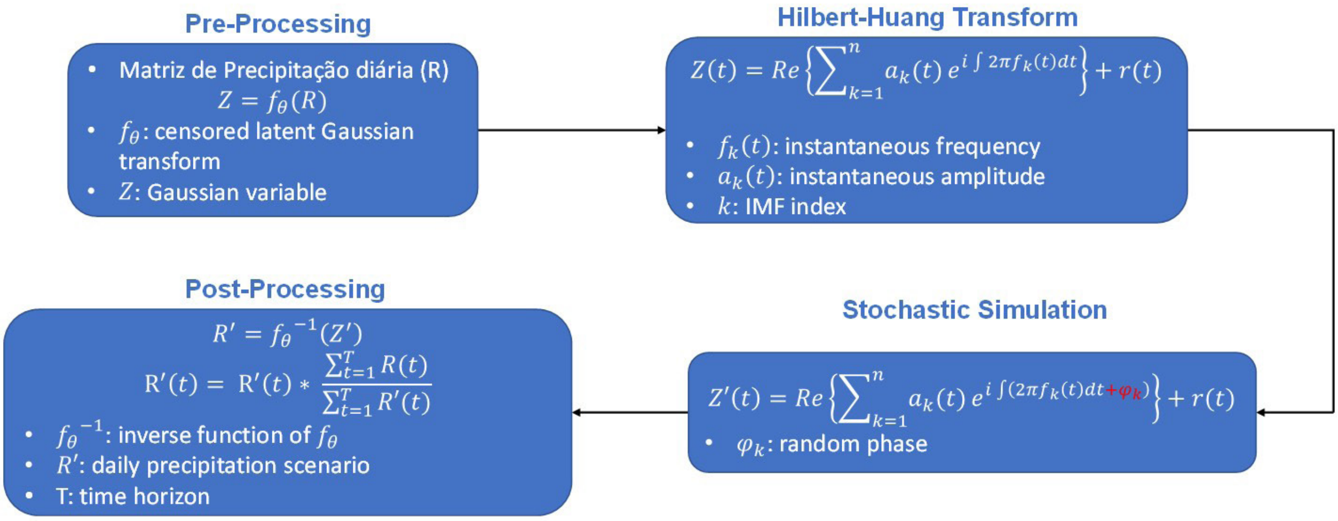

Figure 2. Flowchart from stochastic modeling of daily precipitation scenarios.

Methodology for generating daily precipitation scenarios

Stochastic modeling of daily precipitation scenarios is based on the model presented by Zhou et al. (2019). After initial tests, the empirical orthogonal function analysis was removed and a statistical downscaling was included as the last step, culminating in the version presented in this paper. The process of stochastic simulation of daily precipitation can be divided into four main steps, as illustrated in Figure 2.

Pre-processing

Considering the high asymmetry and rate of zeros (days without rain) in precipitation time series, it is necessary to apply pre-processing to the data before proceeding. A censored latent Gaussian transformation (Allard & Bourotte, 2015) is used to normalize the series, converting the daily precipitation R into a latent Gaussian variable Z. Equation 1 defines this transformation. Null values of precipitation correspond to values from the latent Normal below a determined quantile Z_0, while positive values are a non-linear function above Z_0. This limit is defined as the quantile of a normal distribution such that the cumulative probability up to it is equal to the unconditional probability of zeros in the time series. The estimation of the set of parameters θ is done through the maximum likelihood defined by Equation 2. Once the θ parameters have been estimated, Equation 3 is used to transform each R value into Z.

$$R = \begin{cases} R_m + b\left(e^{a(Z-Z_0)^c} - 1\right) & Z > Z_0 \\ 0 & Z \le Z_0 \end{cases} \quad (1)$$

Where R_m is the precision of the precipitation variable (0.1mm for instance); $\theta = (a, b, c)$ are the transformation parameters; Z is the latent variable; Z_0 is the limit of the latent variable that reproduces the percentage of dry days.

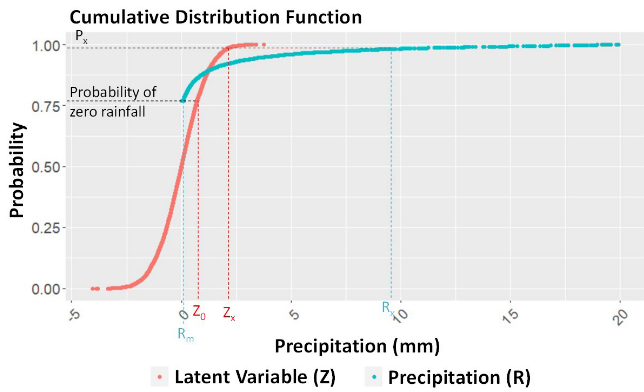

Figure 3. Cumulative distribution function from a generic daily rainfall time series R (blue line) and its transformation in a latent Gaussian variable Z (red line).

$$l(\ ) = \sum_{R_i \ R_m} \left\{ \begin{array}{c} \frac{f\ (R_i)^2}{2} - \frac{1}{c}\log a - \log c - \log(R_i - R_m + b) + \\ \left(\frac{1}{c} - 1\right)\log\left[\log\left(\frac{R_i - R_m}{b} + 1\right)\right] \end{array} \right\} \quad (2)$$

$$f_\theta(R) = \left\{a^{-1}\log\left[b^{-1}(R - R_m)\right] + 1\right\}^{\frac{1}{c}} + Z_0 \quad (3)$$

Figure 3 presents an example of the cumulative distribution function of a daily precipitation time series R (blue line) and its transformation into a normal variable Z (red line). In this example, the original series has approximately 75% of values equal to zero.

Hilbert-Huang transform

The second stage of the stochastic modeling of the daily rainfall consists of applying the Hilbert-Huang Transform (HHT)

(Huang et al., 1998; Huang & Wu, 2008). Unlike other available techniques (such as Fourier transform or Wavelet), HHT can be used in non-linear and non-stationary time series. HHT consists of the application of two procedures:

1. Empirical Mode Decomposition (EMD), which disaggregates the original series into a finite number of orthogonal functions, called IMFs (Intrinsic Mode Function);
2. Hilbert Spectral Analysis (HSA), applied to each IMF, whose product is a time-frequency-energy distribution, is called the Hilbert spectrum.

A function is called an IMF if it satisfies two conditions:

1. The number of extremes (peaks and valleys) and the number of times the x-axis is crossed must be the same as in the original series (it can differ by a maximum of one unit);
2. The local mean equals zero (the mean between the upper and lower envelopes equals zero).

By default, the number of extremums decreases as the procedure is performed, and the original signal $Z(t)$ is decomposed into high-frequency to low-frequency components, until it is no longer possible to perform the procedure. Once the sifting process is over, when it is no longer possible to obtain a new IMF, the original time series can be reconstructed according to Equation 4, adding the extracted IMFs, and a residue $r(t)$.

$$Z(t) = \sum_{k=1}^{n} IMF_k(t) + r(t) \tag{4}$$

In GHCen it is used an improved version of the EMD algorithm called Complete Ensemble Empirical Mode Decomposition with Adaptive Noise – CEEMDAN (Torres et al., 2011; Colominas et al., 2014). Figure 4 presents an example of applying the CEEMDAN to a normalized time series, obtaining 12 IMFs and a residual. Each IMF can be interpreted as one of the cycles that compose the stochastic process of precipitation, from the movement of cold fronts (high frequency) to inter-annual and multidecadal cycles (low frequency).

For each $IMF_k(t)$ obtained in the first step, its Hilbert transform, $H[IMF(t)]$, is calculated according to Equation 5:

$$H[IMF(t)] = \frac{1}{\pi}\int \frac{IMF(\tau)}{t-\tau} d\tau \tag{5}$$

The $IMF(t)$ (real signal) and its Hilbert transform $H[IMF(t)]$ form a complex conjugate pair of the analytical signal $S(t)$, Equation 6, which can be represented in the form of polar coordinates, Equation 7:

$$S(t) = IMF(t) + iH[IMF(t)] \tag{6}$$

$$S(t) = a(t)e^{i\theta(t)} \tag{7}$$

where $a(t)$ is the instantaneous amplitude obtained by Equation 8, and $\theta(t)$ is the instantaneous phase calculated by Equation 9:

$$a(t) = \sqrt{IMF^2(t) + H^2[IMF(t)]} \tag{8}$$

$$\theta(t) = \arctan\frac{H[IMF(t)]}{IMF(t)} \tag{9}$$

Given the instantaneous phase, the instantaneous frequency $f(t)$ is defined by Equation 10:

$$f(t) = \frac{1}{2\pi}\frac{d\theta(t)}{dt} \tag{10}$$

After applying the Hilbert transform to all IMFs obtained from the Gaussian time series $Z(t)$, the latter can be reconstructed

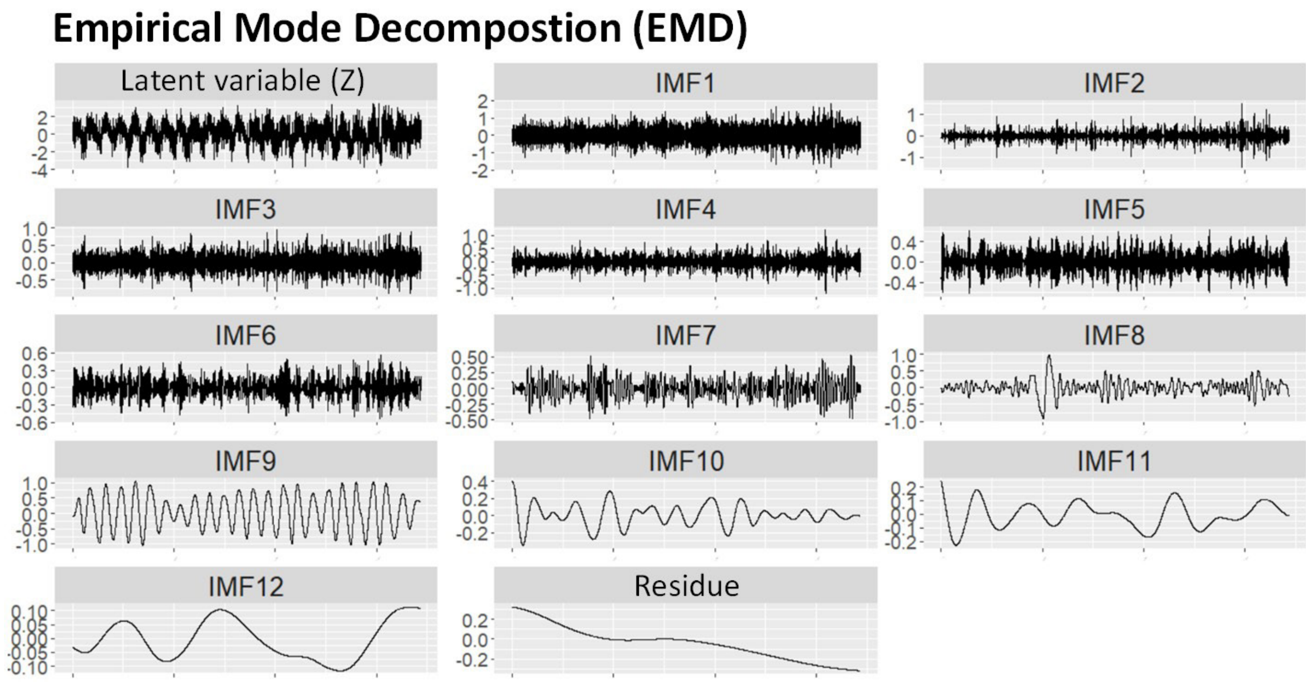

Figure 4. Result from the application of CEEMD into a latent Gaussian variable (Z) obtaining 12 IMFs and a residue r(t).

using Equation 12. This equation represents the end of the HHT, highlighting the amplitude and frequency of each IMF as a function of time.

$$Z(t) = Re\left\{\sum_{k=1}^{n} a_k(t) e^{i\int 2\pi f_k(t)dt}\right\} + r(t) \tag{11}$$

where *Re* represents the real part of the complex signal.

Stochastic simulation

In stochastic modeling, the time series under consideration can be perceived as a potential manifestation of the underlying stochastic phenomenon. To produce alternative conceivable manifestations that retain the fundamental characteristics of the historical data while accommodating the inherent unpredictability of the phenomenon, a random component is introduced as described in Equation 11 (Wen & Gu, 2009), resulting in Equation 12:

$$Z'(t) = Re\left\{\sum_{k=1}^{n} a_k(t) e^{i\int\left(2\pi f_k(t)dt + \varphi_k\right)}\right\} + r(t) \tag{12}$$

where $Z'(t)$ is the synthetic scenario of the same size as the original historical series; φ_k is a uniformly distributed phase noise between 0 and 2π.

To properly represent the most relevant characteristics of the historical series, the IMFs can be divided into significant and non-significant, applying a statistical significance (Wu & Huang, 2004), which compares the average energy of each IMF to the energy of white noise of the same average frequency. If an IMF is classified as not significant, this IMF is simulated by assigning phase noise φ_k. The IMFs identified as significant remain unchanged in the stochastic simulation phase, thus preserving the main characteristics of the historical series. A synthetic scenario of the latent Gaussian variable can then be obtained by summing all the significant (preserved) IMFs, the non-significant (simulated) IMFs, and the residual (preserved).

The processes described by Equations 1-12 are performed univariately for each sub-basin modeled. To preserve the spatial correlation between sub-basins, generating multivariate precipitation scenarios, the random phase noise φ_k, applied to non-significant IMFs is the same for IMFs of the same order in all sub-basins (Wen & Gu, 2009).

Post-processing

After the stochastic simulation stage and obtaining the synthetic scenarios $Z'(t)$ (which have normal distribution), the inverse transformation described in Equation 1 is performed to obtain the precipitation scenarios $R'(t)$. It should be noted that the censored latent Gaussian transformation also guarantee that all scenarios are positive. For the synthetic average of precipitation to be as close as possible to the historical average, a statistical downscaling is performed in the synthetic scenarios, given by Equation 14. Finally, the synthetic precipitation scenarios are limited to twice the maximum daily precipitation observed in each month, so that unrealistic scenarios are not generated.

It is important to note that this type of stochastic simulation can generate as many synthetic scenarios as necessary, with the same length as the original time series.

$$R'(t) = R'(t) * \frac{\sum_{t=1}^{T} R(t)}{\sum_{t=1}^{T} R'(t)} \tag{14}$$

Transformation of synthetic scenarios of daily precipitation in daily streamflow

After doing the processes described above, the daily precipitation scenarios can be used as input to the SMAP/ONS model (Operador Nacional do Sistema Elétrico, 2017), thus generating synthetic scenarios of daily streamflow. SMAP/ONS is a rainfall-runoff model, on a daily scale, based on SMAP (Lopes et al., 1982), with modifications aiming to better represent the specific characteristics of some basins. Figure 5 shows a representation of the SMAP/ONS model. The changes made to the SMAP/ONS model include:

- Addition of a fourth reservoir called plain reservoir (Rsup2);
- Possibility of using up to two recession coefficients in the surface runoff reservoir (K2t and K2t2);
- Use of precipitation adjustment coefficients for temporal representation;
- Use of adjustment coefficients for precipitation and potential evapotranspiration;
- Data assimilation and optimization process, so that, in its operational phase, the model can correct its state variables to reduce the deviation between the simulated and observed flows in a period before the day of the forecast.

The parameters for all calibrated sub-basins, which represent 100% of the HPPs in NIS, are published in Operador Nacional do Sistema Elétrico (2023). Given that the synthetic series of precipitation have the same size as the historical series, specific periods of these synthetic series can be selected so that they reproduce the hydrological behavior of years of greater interest for the simulation with the SMAP/ONS model. The selection of a specific period within the set of scenarios generated by the GHCen can consider the similarity with the recent hydrological situation or with predicted climatic variables.

Later, if it is necessary to reduce the size of the generated sample, clustering techniques can be applied to the synthetic scenarios of daily streamflow obtained at the end of the processing of the GHCen model (Operador Nacional do Sistema Elétrico, 2022). Depending on the desired application, synthetic streamflow scenarios can be calculated in terms of weekly or monthly averages.

DATA

Daily precipitation data

The daily precipitation data ranges from 1998 to 2021 and is a merging of three datasets: a combination of satellite and surface observation from Rozante et al. (2010); surface observations

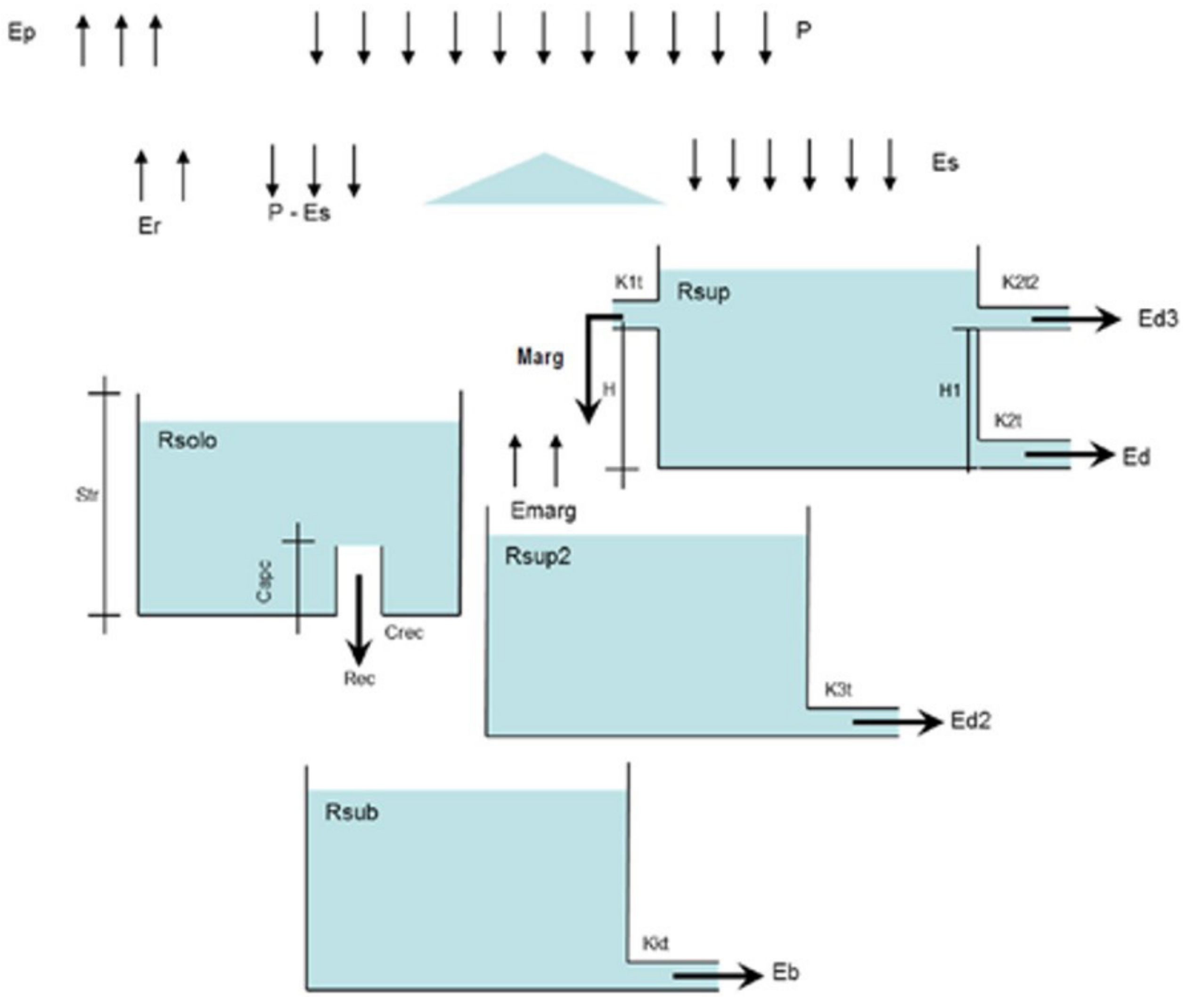

Figure 5. Representation of SMAP/ONS model.

from the National Center for Natural Disaster Monitoring and Alerts (CEMADEN) and Center for Weather Forecasting and Climate Studies/National Institute for Space Research (CPTEC/INPE); and a combination of satellite and surface observation from Operador Nacional do Sistema Elétrico (2020). For all sub-basins considered in the case study is calculated a mean daily precipitation over its area.

Streamflow data

The dataset used in this study includes natural daily and monthly streamflow readings from 146 hydropower plants across all major Brazilian river basins, dating from 1931 to 2021. The data was provided by ONS and can be accessed on their website (2023). To ensure that the readings only reflected natural streamflow, the anthropogenic effects of regulation, diversions, and reservoir evaporation were removed. For this study, only incremental natural streamflow was considered, which is the difference between the total natural streamflow readings from two adjacent HPPs.

CASE STUDY

Evaluation of the precipitation scenarios

To evaluate the potential of the methodology for generating daily precipitation scenarios, 200 synthetic scenarios of equal size to the evaluated historical record (1998 to 2021) were simulated for the 104 sub-basins currently calibrated with the SMAP/ONS model. This analysis will focus on the reproduction of several statistics relevant to the stochastic process of precipitation, namely:

- Mean, standard deviation, and skewness, on the daily (for each month), monthly and annual scales;
- Autocorrelations (lag1), on daily and annual scales;
- Spatial continuity ratio CR (Wilks, 1998);
- Sequences of days with and without rainfall and monthly percentage of days without rain;
- 95th percentile and daily maximum.

For all analyzed variables, it will be displayed a scatter plot comparing synthetic and historical values, and the correlation between both.

Evaluation of monthly streamflow and energy scenarios

To analyze the potential of GHCen for monthly streamflow scenarios a generation of 200 synthetic scenarios was made, starting in January 2021 until December 2021, using as base years 1998 until 2020. The synthetic scenarios generated by the GHCen model will be compared to the synthetic scenarios obtained with the historical rainfall as input to the SMAP/ONS model. This comparison aims to verify whether the synthetic scenarios generated with the GHCen model effectively capture the behavior simulated with the historical years.

Additionally, a comparison between the Natural Energy Inflow (NEI) scenarios for the NIS generated in the same case will be compared to the observed NEI and long-term mean (LTM), calculated from 1931 to 2021.

RESULTS AND DISCUSSIONS

Daily precipitation scenarios analysis

Figure 6 presents a comparison of historical and synthetic statistics of mean, standard deviation, and skewness on daily, monthly, and annual scales. There is a great adherence of synthetic statistics to historical ones. Even with the generation of synthetic scenarios being carried out daily, it is possible to notice that in the monthly and annual time scales, there is an adequate reproduction of the analyzed variables, including the skewness, which is usually a difficult variable to reproduce. Notably, in the stochastic model used, there is no parameterization regarding these statistics.

Figure 7 compares historical and synthetic autocorrelation (lag 1) and spatial correlation statistics, on daily and annual scales. Again, there is a good reproduction of the statistics evaluated in practically all locations, with a high correlation between synthetic and historical data. Figure 8 presents the spatial continuity ratio, which is a metric that measures the spatial intermittency of precipitation. It is noted that the synthetic series manages to reproduce this important behavior of the historical series, demonstrating that spatial coherence is maintained.

Figure 9 presents the comparison between the sequences of days with or without precipitation, synthetic and historical. It is observed that even in longer sequences, lasting 8 days or more than 28 days, the GHCen model manages to adequately reproduce this characteristic.

Figure 10 displays the monthly percentage of days without precipitation, indicating that the synthetic scenarios accurately replicate historical values. Furthermore, Figure 11 and 12 present the boxplot

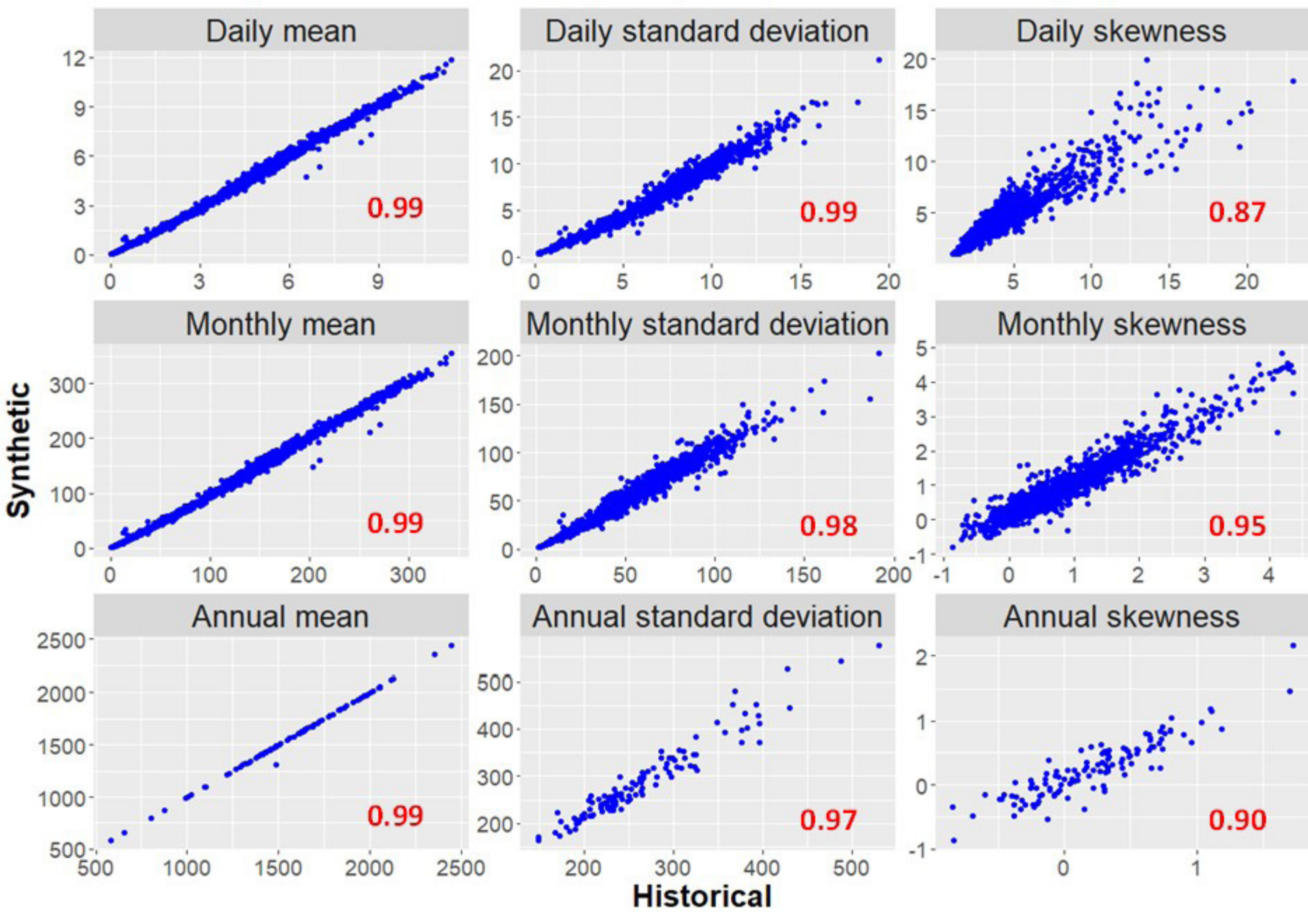

Figure 6. Comparison between historical and synthetic statistics on daily, monthly, and annual scales. In red, the correlation between historical and synthetic values is highlighted.

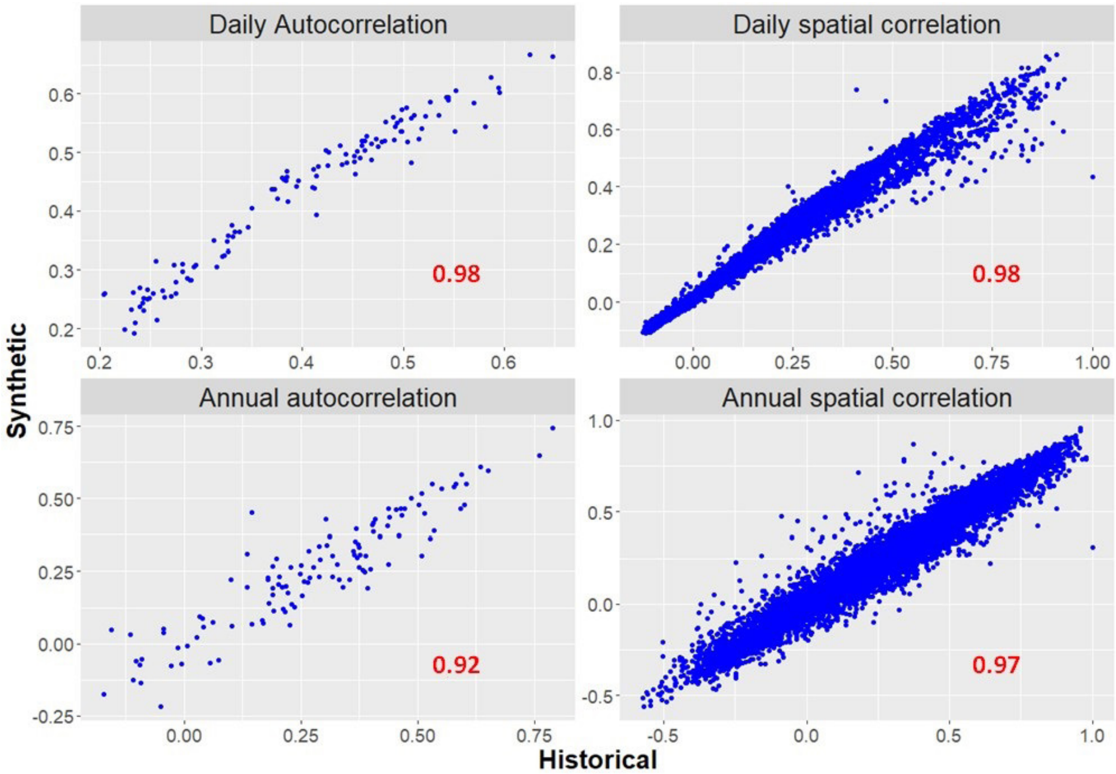

Figure 7. Comparison between autocorrelation and historical and synthetic spatial correlation on daily and annual scales.

of the maximums and synthetic 95th percentile for scenarios from all sub-basins, compared to the historical values shown as red dots, to examine the tail of the distributions of the historical record. It is noticeable that the GHCen model can generate greater maximums than those observed in the historical series, which is a crucial attribute for a model to produce synthetic scenarios.

Monthly streamflow scenarios analysis

This section presents an initial analysis of the GHCen model's streamflow scenario generation capability, monthly. To evaluate this, a total of 200 synthetic scenarios were generated, starting from January 2021 until December 2022, using the years from 1998 until 2020 as the base. The main aim of this comparison is to verify whether the synthetic scenarios generated by the GHCen model accurately capture the behavior observed in the selected hydrological years, in terms of incremental streamflow.

Figures 13 to 16 show boxplots of the scenarios generated for the Furnas, Itá, Tucuruí, and Capivara HPPs, based on the years 2001, 2006, 2011, and 2020. These scenarios are then compared to the scenario obtained with the SMAP/ONS model simulated with historical rainfall (represented by red lines). The comparison reveals significant adherence between the simulated scenarios with historical rainfall and the scenarios obtained with the GHCen model. Furthermore, the GHCen model manages to generate scenarios that mimic the hydrological behavior of each selected hydrological year, in addition to being able to generate both drier scenarios and scenarios with higher flows than those obtained with the simulation with historical precipitation.

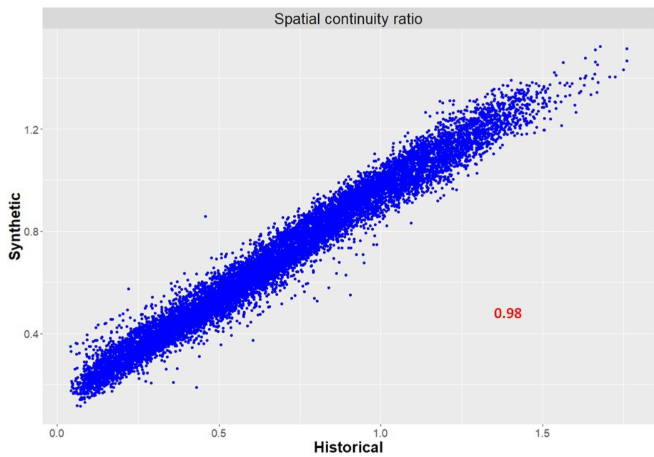

Figure 8. Comparison between historical and synthetic statistics of spatial continuity ratio.

Figure 17 shows the boxplot of the transformation of the streamflow scenarios to NEI scenarios for the NIS compared to the values observed in 2021 (black line) and LTM (red line). There is a good adherence to the NEI observed about the generated scenarios, mainly from April onwards. It should be noted that the year 2021 was one of the worst hydrological years in the historical record of Brazil, since 1931. Finally, Figure 18 presents the boxplot of the synthetic mean NEI between 2021, compared with the observed values (black dots) and the MLT (red dots). It can be noted that the synthetic scenarios generated were able to effectively capture the accumulated ENA that occurred throughout 2021.

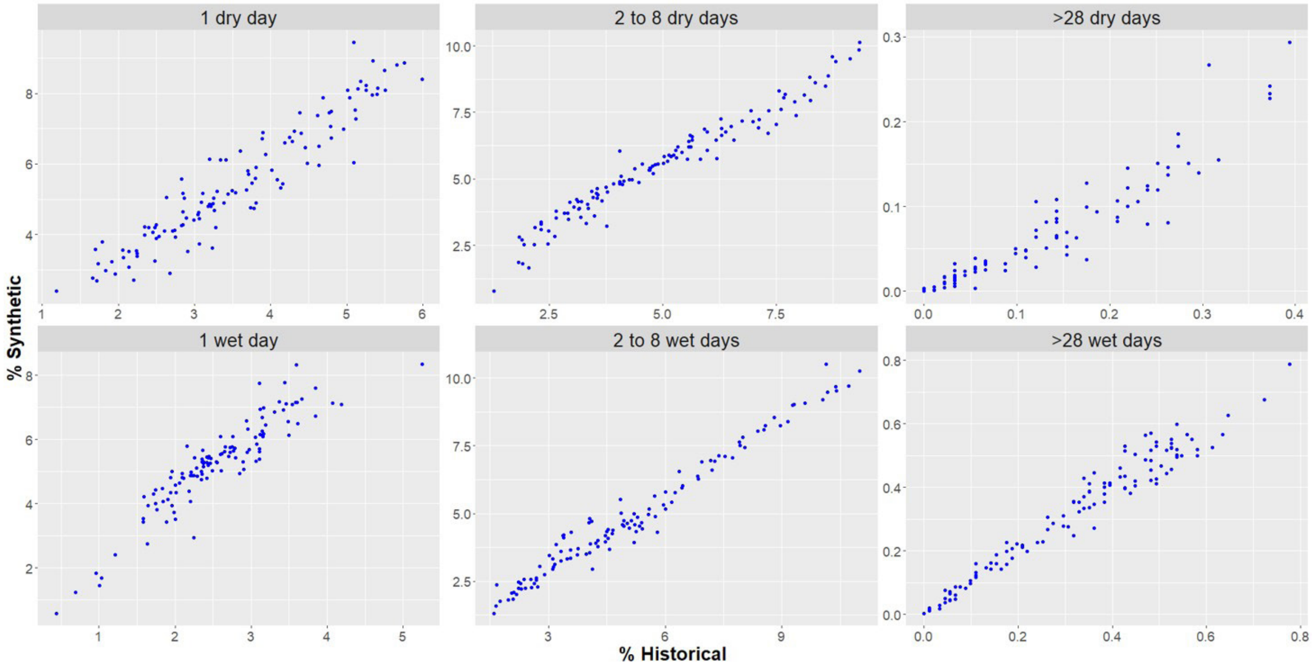

Figure 9. Comparison between historical and synthetic statistics of sequences of days with and without precipitation.

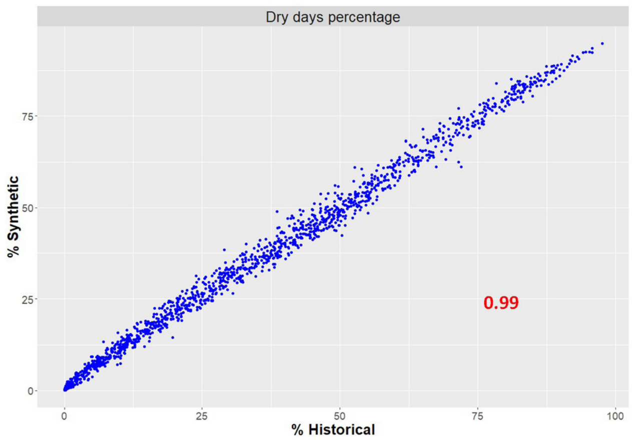

Figure 10. Comparison between the monthly percentage of days without synthetic and historical precipitation.

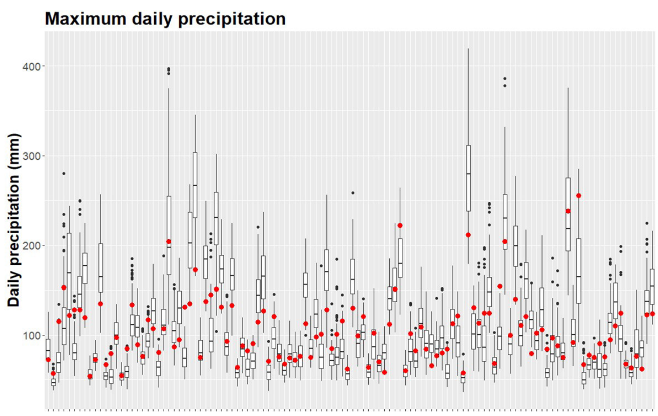

Figure 11. Boxplot of maximums obtained in each synthetic series, compared to historical values (red dots), by sub-basin.

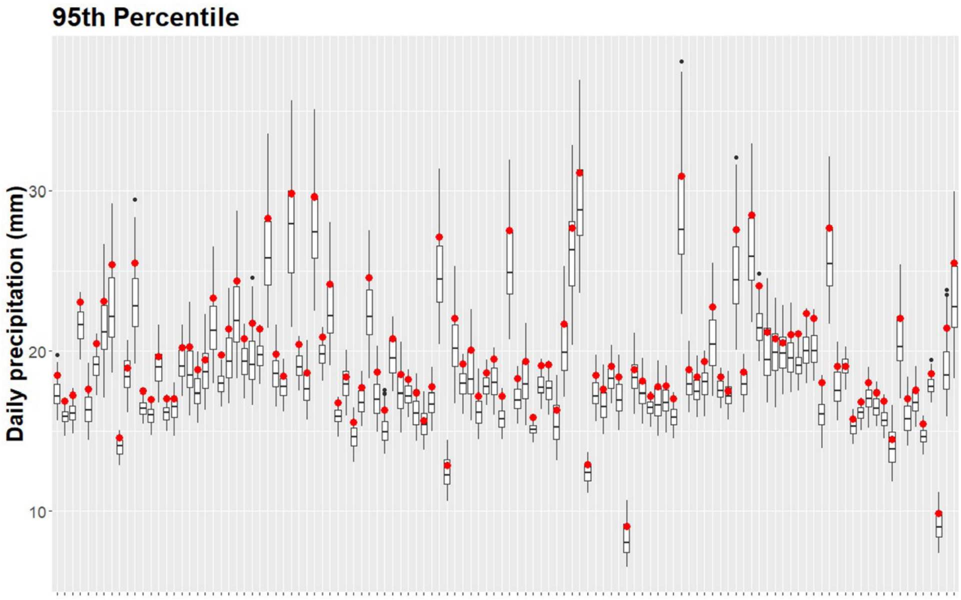

Figure 12. Boxplot of 95th percentile obtained in each synthetic series, compared to historical values (red dots), by sub-basin. Monthly streamflow scenarios analysis.

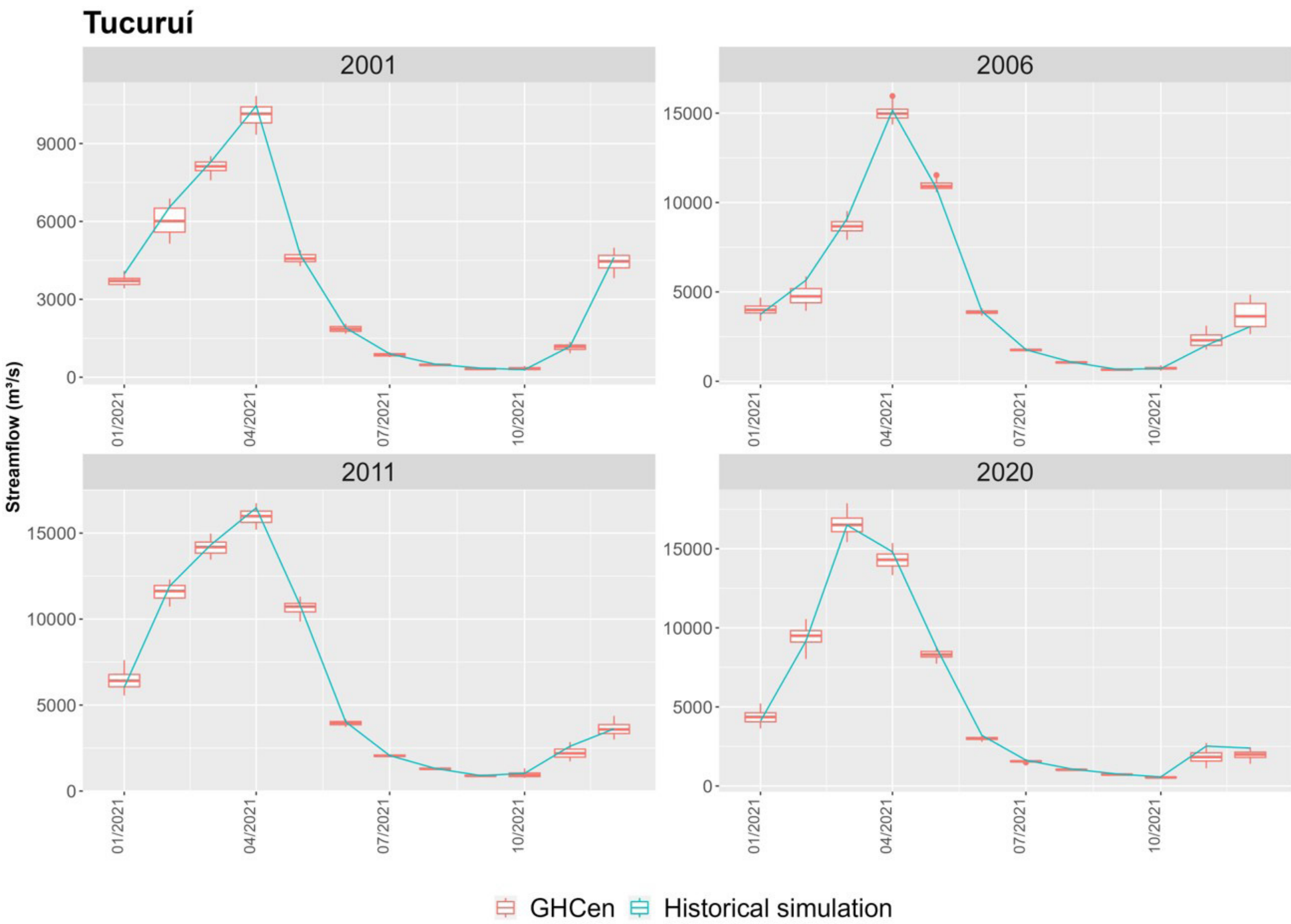

Figure 13. Boxplot of synthetic scenarios obtained with the GHCen for Tucuruí HPP based on the years 2001, 2006, 2011, and 2020 in comparison with the scenario obtained with the historical rainfall as input for the SMAP/ONS model (blue line).

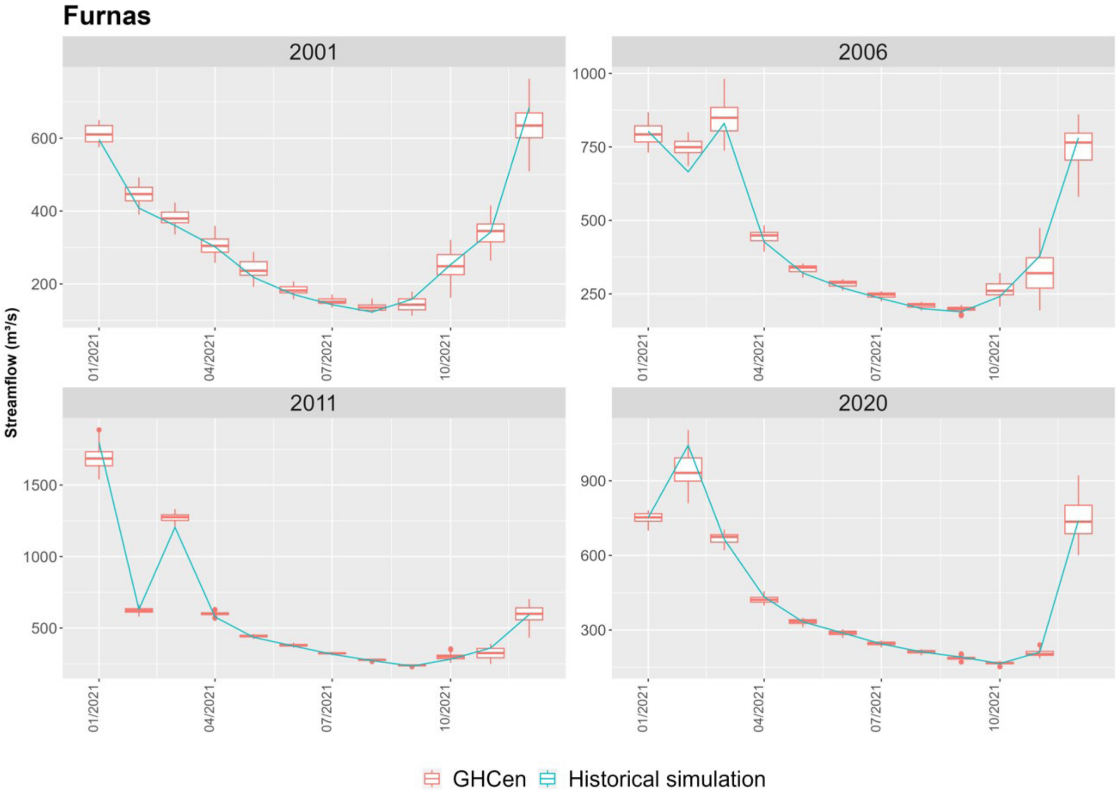

Figure 14. Boxplot of synthetic scenarios obtained with the GHCen for Furnas HPP based on the years 2001, 2006, 2011, and 2020 in comparison with the scenario obtained with the historical rainfall as input for the SMAP/ONS model (blue line).

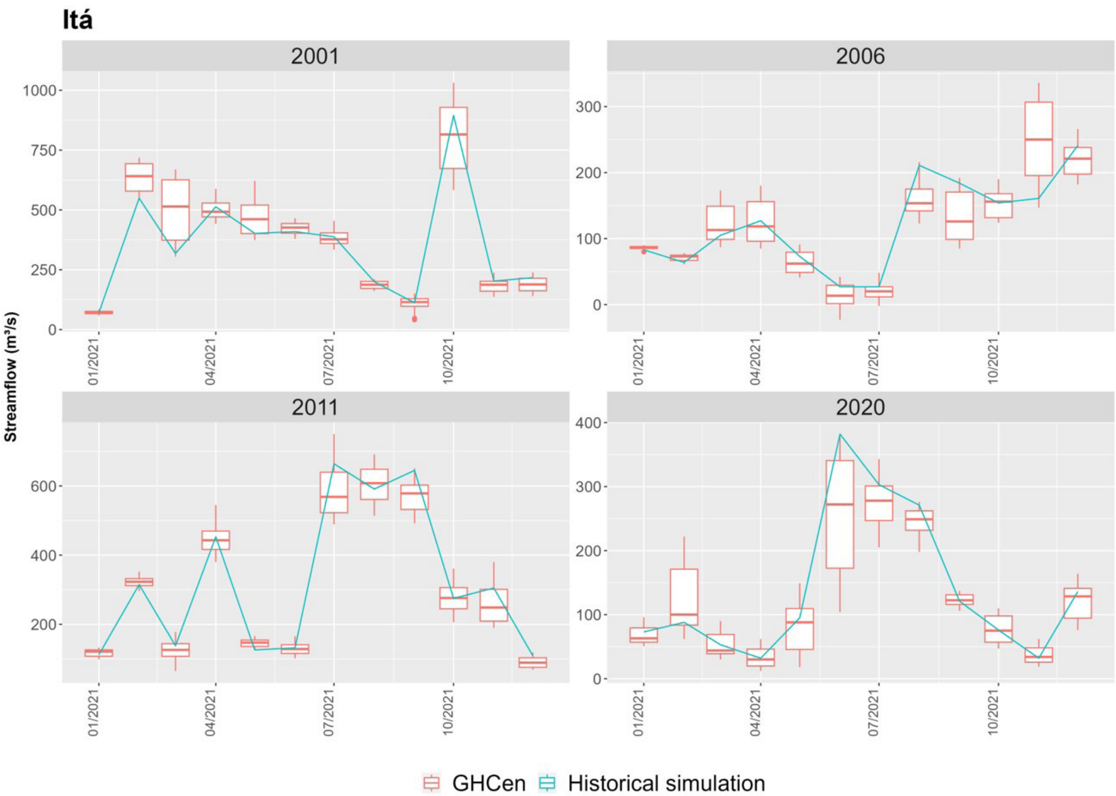

Figure 15. Boxplot of synthetic scenarios obtained with the GHCen for Itá HPP based on the years 2001, 2006, 2011, and 2020 in comparison with the scenario obtained with the historical rainfall as input for the SMAP/ONS model (blue line).

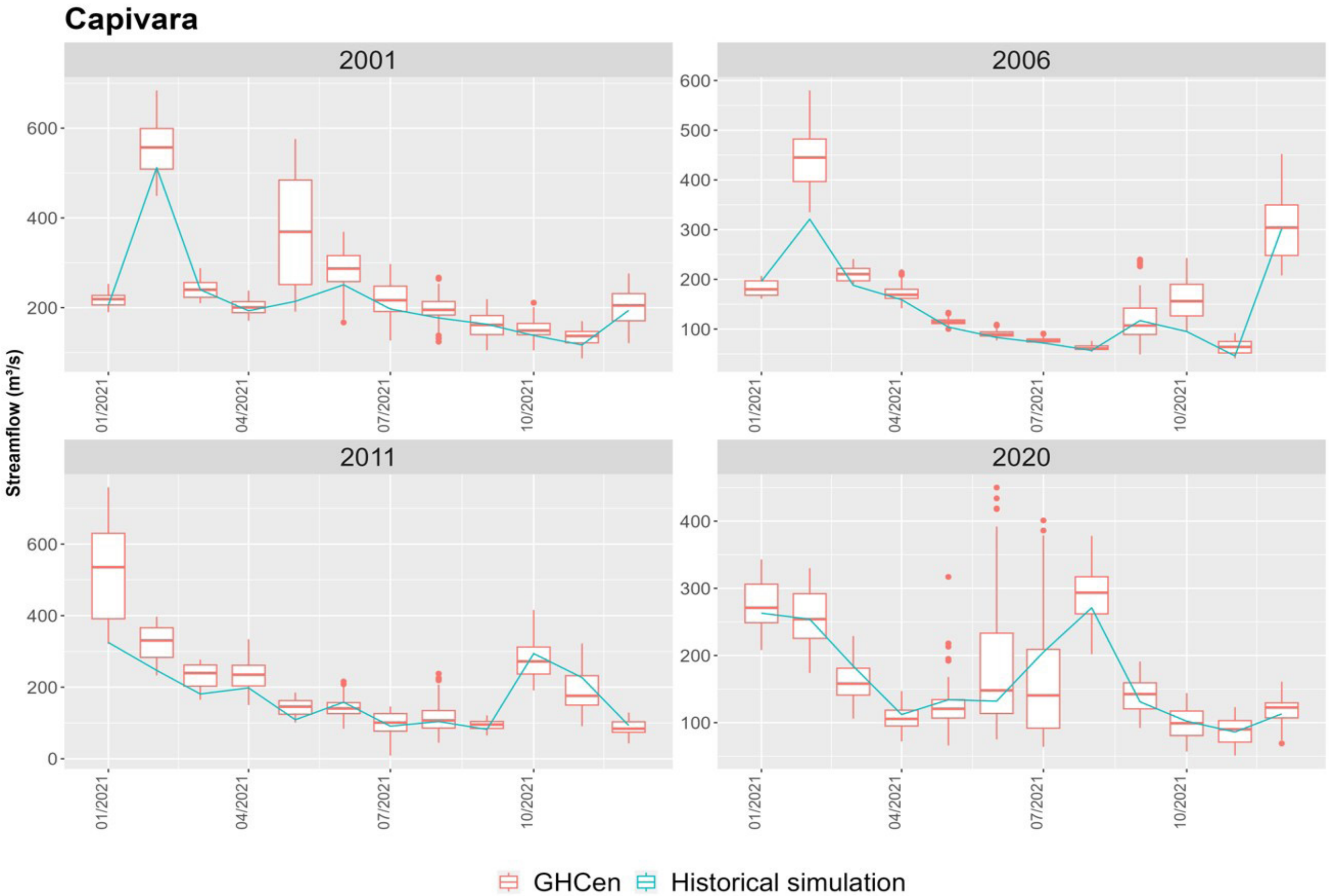

Figure 16. Boxplot of synthetic scenarios obtained with the GHCen for Capivara HPP based on the years 2001, 2006, 2011, and 2020 in comparison with the scenario obtained with the historical rainfall as input for the SMAP/ONS model (blue line).

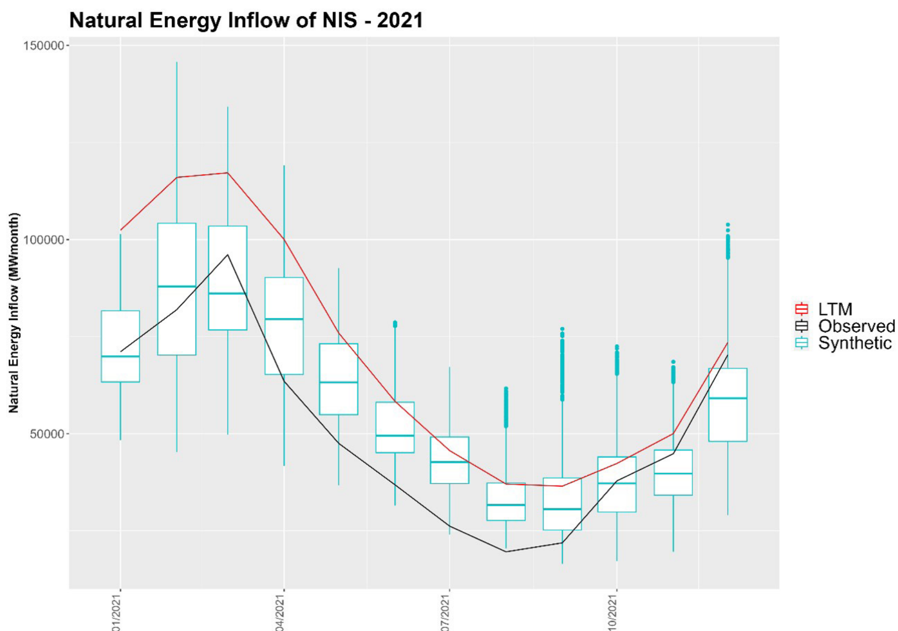

Figure 17. Boxplot of synthetic scenarios obtained with the GHCen model for the NIS compared to the values observed in 2021 (black line) and LTM (red line).

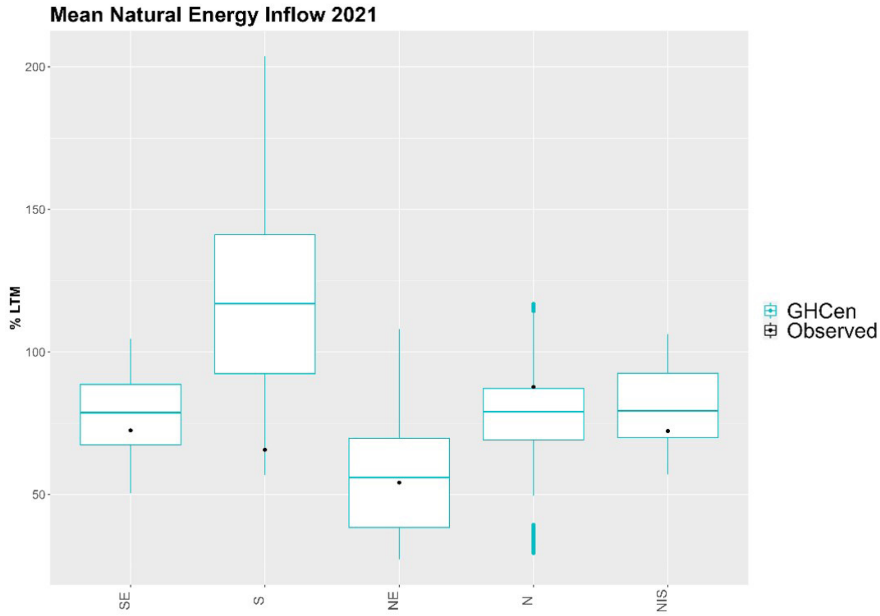

Figure 18. Boxplot of synthetic scenarios mean NEI of 2021 generated by GHCen (blue boxplot), compared to the observed values (blue dots) for each region and NIS.

CONCLUSIONS

This paper aimed to present a new model for generating synthetic streamflow scenarios. The proposed approach merges the conceptual SMAP/ONS model with a stochastic methodology for the simulation of synthetic scenarios of daily precipitation, resulting in a Hybrid Generator of Synthetic Streamflow Scenarios (GHCen). As can be seen throughout the evaluations presented, the stochastic model used to generate synthetic daily precipitation scenarios can reproduce the main characteristics of this stochastic process, in distinct time scales.

Furthermore, a case study from January 2021 was presented, demonstrating that the GHCen model can accurately replicate the hydrological behavior of the years from which the scenarios are based. When compared in terms of NEI, there is a close relation between the synthetic scenarios generated by the GHCen model and the observed series for the NIS. This holds for both monthly terms and the accumulated NEI during this period, which shows the potential of using GHCen as a synthetic scenario generator of monthly inflows.

Some improvements can be made to GHCen, such as selecting simulated years based on climate similarity to the current year. As the next steps, an exhaustive evaluation of the GHCen model should be carried out, to quantitatively verify its potential capacity to generate synthetic scenarios of monthly inflows, in comparison with the methodologies currently used.

REFERENCES

Allard, D., & Bourotte, M. (2015). Disaggregating daily precipitations into hourly values with a transformed censored latent Gaussian process. *Stochastic Environmental Research and Risk Assessment*, *29*(2), 453-462. http://dx.doi.org/10.1007/s00477-014-0913-4.

Ávila, L., Silveira, R., Campos, A., Rogiski, N., Freitas, C., Aver, C., & Fan, F. (2023). Seasonal streamflow forecast in the Tocantins River Basin, Brazil: an evaluation of ECMWF-SEAS5 with multiple conceptual hydrological models. *Water*, *15*(9), 1695. http://dx.doi.org/10.3390/w15091695.

Colominas, M. A., Schlotthauer, G., & Torres, M. E. (2014). Improved complete ensemble EMD: a suitable tool for biomedical signal processing. *Biomedical Signal Processing and Control*, *14*(1), 19-29. http://dx.doi.org/10.1016/j.bspc.2014.06.009.

Hamill, T. M., & Scheuerer, M. (2018). Probabilistic precipitation forecast postprocessing using quantile mapping and rank-weighted best-member dressing. *Monthly Weather Review*, *146*(12), 4079-4098. http://dx.doi.org/10.1175/MWR-D-18-0147.1.

Huang, N. E., & Wu, Z. (2008). A review on Hilbert-Huang transform: method and its applications to geophysical studies. *Reviews of Geophysics*, *46*(2), RG2006. http://dx.doi.org/10.1029/2007RG000228.

Huang, N. E., Shen, Z., Long, S. R., Wu, M. C., Shih, H. H., Zheng, Q., Yen, N. C., Tung, C. C., & Liu, H. H. (1998). The empirical mode decomposition and the Hubert spectrum for nonlinear and non-stationary time series analysis. *Proceedings - Royal Society. Mathematical, Physical and Engineering Sciences*, *454*(1971), 903-995. http://dx.doi.org/10.1098/rspa.1998.0193.

Humphrey, G. B., Gibbs, M. S., Dandy, G. C., & Maier, H. R. (2016). A hybrid approach to monthly streamflow forecasting: integrating hydrological model outputs into a Bayesian artificial neural network. *Journal of Hydrology*, *540*, 623-640. http://dx.doi.org/10.1016/j.jhydrol.2016.06.026.

Liang, Z., Wang, D., Guo, Y., Zhang, Y., & Dai, R. (2013). Application of Bayesian model averaging approach to multimodel ensemble hydrologic forecasting. *Journal of Hydrologic Engineering*, *18*(11), 1426-1436. http://dx.doi.org/10.1061/(ASCE)HE.1943-5584.0000493.

Lopes, J. E. G., Braga, B. P. F., & Conejo, J. G. L. (1982). SMAP: a simplified hydrological model. In: Singh VP, editor. *Applied modelling in catchment hydrology.* Littleton, CO: Water Resources Publications.

Operador Nacional do Sistema Elétrico – ONS. (2017). *Manual de metodologia - aplicativo SMAP/ONS*. Rio de Janeiro.

Operador Nacional do Sistema Elétrico – ONS. (2020). *NT 0076-2020: precipitação obtida pela combinação entre dados estimados por satélite e dados observados de estações meteorológicas.* Rio de Janeiro. Retrieved in 2023, October 6, from https://sintegre.ons.org.br/sites/10/45/paginas/servicos/produtos-pasta.aspx?RootFolder=/sites/10/45/Produtos/554/31-07-2020_145623/NTs

Operador Nacional do Sistema Elétrico – ONS. (2022). *NT 0098-2022: geração de cenários de afluências para a avaliação conjuntural do PEN 2022 utilizando o modelo GHCen.* Rio de Janeiro.

Operador Nacional do Sistema Elétrico – ONS. (2023). *NT 0097 2018: aplicação do modelo SMAP-ONS para previsão de vazões no âmbito do SIN - REV13.* Rio de Janeiro.

Rozante, J. R., Moreira, D. S., Gonçalves, L. G. G., & Vila, D. A. (2010). Combining TRMM and surface observations of precipitation: technique and validation over South America. *Weather and Forecasting*, *25*(3), 885-894. http://dx.doi.org/10.1175/2010WAF2222325.1.

Schepen, A., & Wang, Q. J. (2015). Model averaging methods to merge operational statistical and dynamic seasonal streamflow forecasts in Australia. *Water Resources Research*, *51*(3), 1797-1812. http://dx.doi.org/10.1002/2014WR016163.

Sikorska-Senoner, A. E., & Quilty, J. M. (2021). A novel ensemble-based conceptual-data-driven approach for improved streamflow simulations. *Environmental Modelling & Software*, *143*, 105094. http://dx.doi.org/10.1016/j.envsoft.2021.105094.

Torres, M. E., Colominas, M. A., Schlotthauer, G., & Flandrin, P. (2011). A complete ensemble empirical mode decomposition with adaptive noise. In *2011 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)* (pp. 4144-4147), Prague, Czech Republic. New York: IEEE. http://dx.doi.org/10.1109/ICASSP.2011.5947265.

Troin, M., Arsenault, R., Wood, A. W., Brissette, F., & Martel, J. L. (2021). Generating ensemble streamflow forecasts: a review of methods and approaches over the past 40 years. *Water Resources Research*, *57*(7), e2020WR028392. http://dx.doi.org/10.1029/2020WR028392.

Tyralis, H., & Papacharalampous, G. (2021). Quantile-based hydrological modeling. *Water*, *13*(23), 3420. http://dx.doi.org/10.3390/w13233420.

Wani, O., Beckers, J. V., Weerts, A. H., & Solomatine, D. P. (2017). Residual uncertainty estimation using instance-based learning with applications to hydrologic forecasting. *Hydrology and Earth System Sciences*, *21*(8), 4021-4036. http://dx.doi.org/10.5194/hess-21-4021-2017.

Wen, Y. K., & Gu, P. (2009). HHT-based simulation of uniform hazard ground motions. *Advances in Adaptive Data Analysis*, *01*(01), 71-87. http://dx.doi.org/10.1142/S1793536909000072.

Wilks, D. S. (1998). Multisite generalization of a daily stochastic precipitation generation model. *Journal of Hydrology*, *210*(1-4), 178-191. http://dx.doi.org/10.1016/S0022-1694(98)00186-3.

Wood, A. W., & Schaake, J. C. (2008). Correcting errors in streamflow forecast ensemble mean and spread. *Journal of Hydrometeorology*, *9*(1), 132-148. http://dx.doi.org/10.1175/2007JHM862.1.

Wu, Z., & Huang, N. E. (2004). A study of the characteristics of white noise using the empirical mode decomposition method. *Proceedings - Royal Society. Mathematical, Physical and Engineering Sciences*, *460*(2046), 1597-1611. http://dx.doi.org/10.1098/rspa.2003.1221.

Ye, A., Duan, Q., Schaake, J., Xu, J., Deng, X., Di, Z., Miao, C., & Gong, W. (2015). Post-processing of ensemble forecasts in low-flow period. *Hydrological Processes*, *29*(10), 2438-2453. http://dx.doi.org/10.1002/hyp.10374.

Zhou, L., Meng, Y., & Abbaspour, K. C. (2019). A new framework for multi-site stochastic rainfall generator based on empirical orthogonal function analysis and Hilbert-Huang transform. *Journal of Hydrology*, *575*, 730-742. http://dx.doi.org/10.1016/j.jhydrol.2019.05.047.

Zhou, Y., Cui, Z., Lin, K., Sheng, S., Chen, H., Guo, S., & Xu, C. Y. (2022). Short-term flood probability density forecasting using a conceptual hydrological model with machine learning techniques. *Journal of Hydrology*, *604*, 127255. http://dx.doi.org/10.1016/j.jhydrol.2021.127255.

Authors contributions

Felipe Treistman: GHCen conceptualization, software development, results generation and analysis, text writing, and final review.

Lucas de Souza Khenayfis: Methodology supporting, software supporting, results analysis, text writing, and review.

Débora Dias Jardim Penna: Methodology supporting, results analysis, text review.

Editor-in-Chief: Adilson Pinheiro

Associated Editor: Adilson Pinheiro