XVIII SIMPÓSIO DE RECURSOS HÍDRICOS DO NORDESTE
XVIII ENCONTRO DE RECURSOS HÍDRICOS EM SERGIPE

Automação em Python do fluxo de geração e análise de séries sintéticas produzidas pelo ClimaBR

Abner da Silva Lins 1; Filipe de Carvalho Lemos 2; Germano Gondin Ribeiro Neto 3; Cristiano das Neves Almeida 4; Antônio Edmilson de Alcântara Dantas5; José Lindemberg Vidal Barbosa 6

RESUMO – A geração de séries sintéticas de precipitação é uma importante alternativa para aplicações hidrológicas e ambientais em regiões com baixa ou nenhuma disponibilidade de dados observados. Nesse contexto, o modelo ClimaBR (Oliveira, 2003) destaca-se como uma ferramenta desenvolvida para representar características estatísticas de eventos pluviométricos típicos das condições climáticas brasileiras. Entretanto, a utilização de suas saídas em estudos de erosividade e modelagem hidrológica requer etapas adicionais de processamento e análise. O presente trabalho teve como objetivo desenvolver um fluxo de trabalho computacional em Python para automatizar o processamento das séries sintéticas geradas pelo ClimaBR. A metodologia implementada consistiu em rotinas em Phyton para preparação dos dados de entrada, reconstrução dos hietogramas por meio da função dupla exponencial, discretização temporal dos eventos em intervalos de 10 minutos e geração de análises estatísticas e gráficas. A avaliação dos resultados foi realizada por meio da comparação entre variáveis observadas e sintéticas, incluindo precipitação total dos eventos, duração da chuva, intensidade máxima em 30 minutos e erosividade. Os resultados indicaram que o sistema desenvolvido foi capaz de reproduzir satisfatoriamente os padrões estatísticos gerais das séries analisadas. Em relação aos dados observados, verificou-se subestimação da duração dos eventos (viés de -44%) e da erosividade (viés de -21%), enquanto a intensidade máxima em 30 minutos (I30) foi superestimada (viés de +30%). Apesar dessas diferenças a representação geral das características dos eventos mostrou-se consistente. A metodologia proposta amplia as possibilidades de aplicação das séries sintéticas produzidas pelo ClimaBR, contribuindo para estudos relacionados à erosão hídrica, conservação do solo e modelagem hidrológica.

Palavras-Chave – ClimaBR; Python; Séries Sintéticas

ABSTRACT– The generation of synthetic rainfall series is an important alternative for hydrological and environmental applications in regions with limited or no availability of observed rainfall data. In this context, the ClimaBR model (Oliveira, 2003) stands out as a tool developed to represent the statistical characteristics of rainfall events typical of Brazilian climatic conditions. However, the use of its outputs in rainfall erosivity and hydrological modeling studies requires additional processing and analysis. This study aimed to develop a computational workflow in Python to automate the processing of synthetic rainfall series generated by ClimaBR. The implemented methodology consisted of Python routines for preparing the input data, reconstructing rainfall hyetographs using the double-exponential function, discretizing rainfall events into 10-minute intervals, and generating statistical and graphical analyses. The results were evaluated by comparing observed and synthetic rainfall event characteristics, including total event rainfall, rainfall duration, maximum 30-minute intensity (I30), and rainfall erosivity. The results showed that the developed system was able to satisfactorily reproduce the overall statistical patterns of the analyzed series. Compared with the observed data, rainfall duration was underestimated (bias of −44%) and rainfall erosivity showed a bias of −21%, whereas the maximum 30-minute intensity (I30) was overestimated (bias of +30%).

Despite these differences, the overall representation of rainfall event characteristics proved to be consistent. The proposed methodology expands the applicability of synthetic rainfall series generated by ClimaBR, contributing to studies related to water erosion, soil conservation, and hydrological modeling.

Keywords: ClimaBR; Python; Synthetic Rainfall Series.

INTRODUÇÃO

A precipitação pluvial constitui uma das variáveis hidrológicas mais importantes para estudos relacionados à conservação do solo, modelagem hidrológica, dimensionamento de obras hidráulicas e gestão de recursos hídricos. Além do volume precipitado, características como duração, intensidade e distribuição temporal da chuva exercem influência direta sobre processos como geração de escoamento superficial, transporte de sedimentos e erosão hídrica do solo. Oliveira (2003) destaca que a adequada representação da precipitação requer não apenas informações referentes à lâmina precipitada, mas também ao perfil temporal dos eventos pluviométricos e com essa prerrogativa foram desenvolvidos modelos de geração.

Os primeiros geradores estocásticos de precipitação surgiram no começo da década de 1960, com o trabalho de Gabriel e Neumann (1962) para ocorrência de chuva, e depois foram ampliados para quantidades de chuva em Todorovic e Woolhiser (1975). Com a finalidade de permitir simulações de cenários hidrológicos e climáticos sem depender apenas de séries observadas, que são frequentemente curtas e incompletas. Estes dados são importantes na hidrologia, agricultura e engenharia para dimensionamento de reservatórios, sistemas de drenagem, modelagem de erosão do solo e planejamento agrícola por exemplo.

O CLIMABR é um modelo estocástico desenvolvido para gerar séries sintéticas de precipitação representativas das condições climáticas brasileiras. Seu desenvolvimento foi baseado em procedimentos utilizados por outros modelos, principalmente o gerador climático CLIGEN (USDA-ARS, 1997), amplamente empregado em estudos de erosão do solo e hidrologia, capaz de simular não apenas a precipitação, mas também variáveis como temperatura máxima e mínima, radiação solar e vento (Min et al., 2011).

Incorporando adaptações para melhorar a representação das características das chuvas observadas no Brasil, O modelo simula, para cada dia chuvoso, a precipitação total, a duração do evento, a intensidade máxima instantânea e o tempo de ocorrência dessa intensidade máxima, permitindo a geração de eventos sintéticos estatisticamente equivalentes aos observados.

No entanto o modelo não faz todas as etapas sozinho. o cálculo dos parâmetros estatísticos de entrada necessários ao CLIMABR e a determinação dos coeficientes da função dupla exponencial não são

executados pelo modelo de forma independente, tornando necessário o desenvolvimento de rotinas adicionais para sua aplicação.

Diante dessa limitação, este estudo teve como objetivo reproduzir a metodologia proposta por Oliveira (2003) por meio de scripts desenvolvidos em Python. Inicialmente, foi implementada uma rotina para o processamento das séries históricas e cálculo automático dos parâmetros estatísticos de entrada do CLIMABR. Em seguida, foi desenvolvido um segundo algoritmo responsável pela determinação dos coeficientes da função dupla exponencial de área unitária, reconstrução do perfil temporal dos eventos sintéticos.

Para avaliar a capacidade do método em reproduzir o regime observado, foram selecionadas séries históricas de precipitação sub-horária das redes CEMADEN e ANA com, no mínimo, dez anos de registros contínuos. Essas séries foram agregadas para a escala diária e utilizadas como dados de entrada do CLIMABR, responsável pela geração das séries sintéticas.

Posteriormente, os eventos sintéticos reconstruídos por meio da função dupla exponencial foram comparados aos eventos observados, com o objetivo de avaliar a capacidade do gerador de séries sintéticas em reproduzir as principais características hidrológicas dos eventos de chuva e sua aplicação na estimativa da erosividade.

METODOLOGIA

Área de Estudo e Conjunto de Dados

Neste estudo, foram utilizados os eventos sintéticos gerados pelo modelo ClimaBR (Oliveira, 2003), desenvolvido para reproduzir as características estatísticas de eventos de precipitação representativos das diferentes condições climáticas brasileiras. Para a avaliação do desempenho do modelo, os eventos sintéticos foram comparados a séries observadas de alta resolução temporal provenientes de estações pluviográficas distribuídas pelo território nacional, considerando características como precipitação total do evento, duração, intensidade máxima em 30 minutos (I30) e índice de erosividade (EI30).

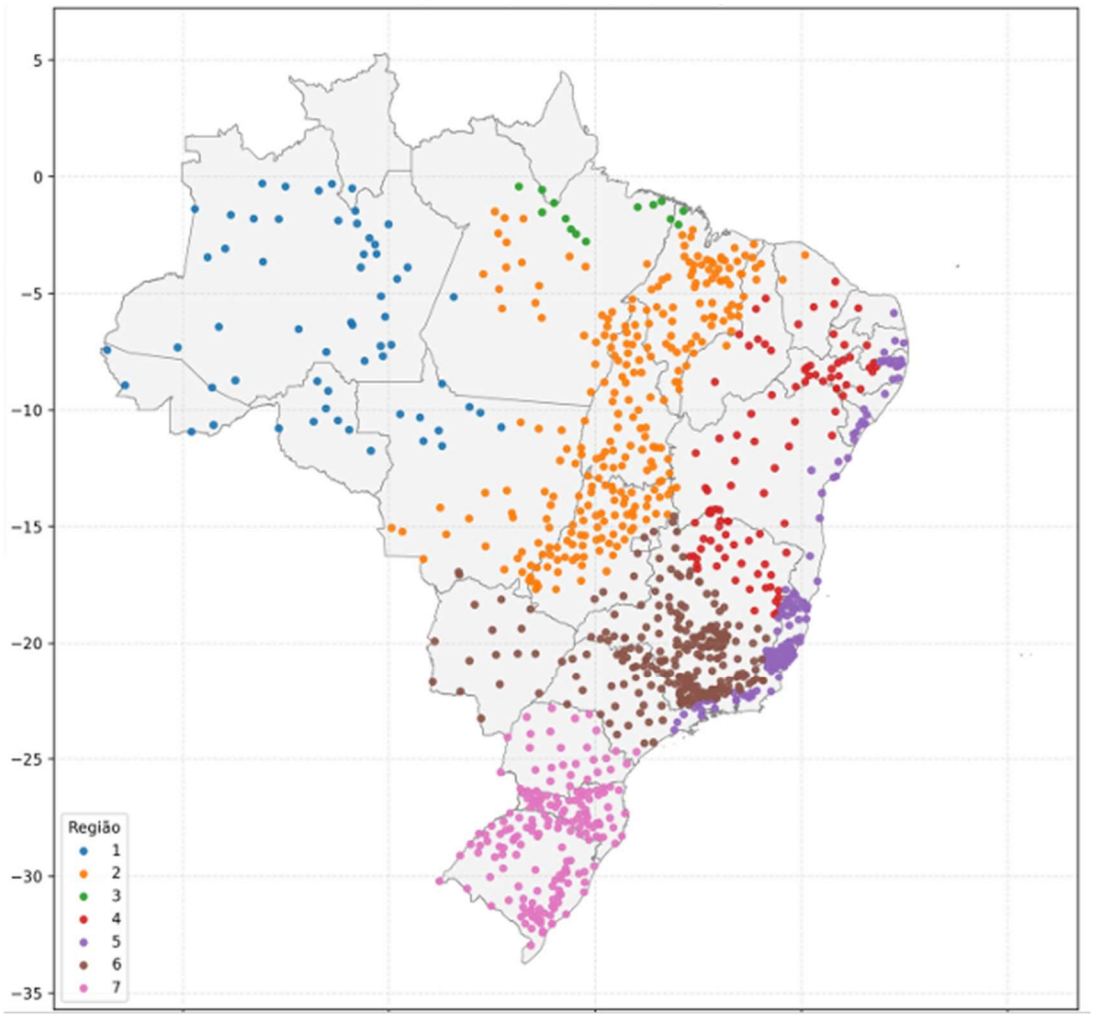


Figura 1 – Área de Estudo e Estações da Hidro Web com as regiões climáticas consideradas pelo ClimaBR

Cálculos

A Figura 2 apresenta o fluxograma da metodologia deste estudo. O processo inicia com a preparação dos dados de entrada do modelo ClimaBR, seguida pela geração das séries sintéticas de precipitação (ClimaBR), reconstrução dos hietogramas por meio da função dupla exponencial, discretização temporal dos eventos em intervalos de 10 minutos e, por fim, avaliação do desempenho das séries sintéticas em relação aos dados observados.

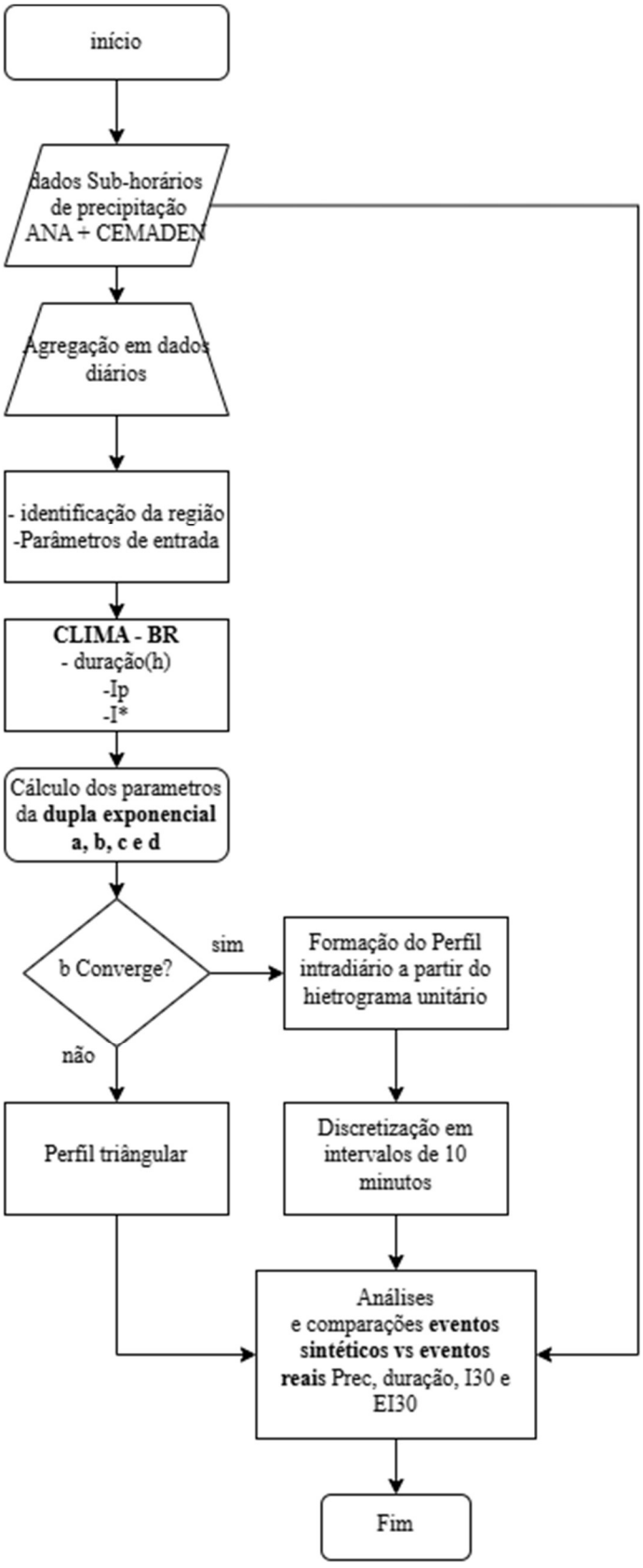


Figura 2 – Fluxograma

Entrada dos Dados

Inicialmente, os dados sub-horários de precipitação foram agregados em séries diárias para o cálculo das variáveis de entrada do ClimaBR. Para cada estação, foram estimadas as estatísticas mensais de precipitação (média, desvio padrão e coeficiente de assimetria) e as propriedades da cadeia de Markov de primeira ordem, representadas pelas probabilidades de ocorrência de dias chuvosos consecutivos (PCC) e da transição de dias secos para dias chuvosos (PCS), utilizadas como parâmetros de entrada do modelo.

O modelo ClimaBR considera a variabilidade climática brasileira por meio da divisão do território nacional em sete regiões homogêneas, cada uma caracterizada por um conjunto específico de parâmetros estatísticos calibrados a partir de séries históricas de precipitação. Dessa forma, a seleção da região climática correspondente constitui uma etapa fundamental da preparação dos dados de entrada, garantindo que os eventos sintéticos gerados representem adequadamente as características pluviométricas locais.

A precipitação média mensal ($\bar{P}$) foi calculada por:

$$\bar{P} = \frac{\sum_{i=1}^{n} P_i}{n}$$

em que $\bar{P}$é a precipitação média mensal (mm), $P_i$representa a precipitação diária observada no dia ie $n$corresponde ao número de dias do mês. em que $\bar{P}$é a precipitação média mensal (mm), $P_i$representa a precipitação diária observada no dia ie $n$corresponde ao número de dias do mês.

O desvio padrão mensal (S) foi obtido por:

$$S = \sqrt{\frac{\sum_{i=1}^{n}(P_i - \bar{P})^2}{n-1}}$$

em que Sé o desvio padrão da precipitação diária, $\bar{P}$é a precipitação média mensal, $P_i$representa a precipitação diária e né o número de observações.

O coeficiente de assimetria (Sk) foi calculado por:

$$Sk = \left(\frac{n}{(n-1)(n-2)}\right)\sum_{i=1}^{n}\left(\frac{P_i - \bar{P}}{S}\right)^3$$

em que Skrepresenta o coeficiente de assimetria da precipitação diária, utilizado para descrever o grau de assimetria da distribuição dos valores de precipitação em relação à média.

As probabilidades de transição da cadeia de Markov de primeira ordem foram estimadas por:

$$PCC = \frac{N_{CC}}{N_C}$$

em que N_{CC}representa o número de dias chuvosos precedidos por outro dia chuvoso e $N_C$corresponde ao número total de dias chuvosos.

A probabilidade de transição de um dia seco para um dia chuvoso foi calculada por:

$$PCS = \frac{N_{SC}}{N_S}$$

em que N_{SC}representa o número de transições de dias secos para dias chuvosos e $N_S$corresponde ao número total de dias secos.

Ocorrência dos eventos

A ocorrência de dias chuvosos é simulada por meio de uma cadeia de Markov de primeira ordem, utilizando as probabilidades de transição entre dias secos e chuvosos (PCC e PCS). Após a definição da ocorrência de precipitação, a lâmina total diária é gerada a partir da distribuição de Pearson Tipo III. Em seguida, a duração do evento é estimada por relações empíricas entre as intensidades média e máxima da chuva, enquanto a intensidade máxima é obtida por uma equação potencial ajustada a dados observados.

Por fim, o instante de ocorrência da intensidade máxima é definido por meio de números aleatórios, caracterizando a posição do pico de intensidade ao longo do evento. Vale ressaltar que todas essas etapas, desde a simulação da ocorrência dos eventos até a estimativa da duração, intensidade máxima e posição do pico, são realizadas pelo executável do ClimaBR. Assim, por não constituírem o foco deste estudo, sua descrição detalhada não é apresentada, podendo ser consultada em Oliveira (2003).

Dupla exponencial

Após a geração dos eventos sintéticos pelo ClimaBR, os perfis intradiários de precipitação foram reconstruídos utilizando a função dupla exponencial proposta por Oliveira (2003) de acordo com a figura 4.

$$i(t) = \begin{cases} ae^{bt}, 0 \leq t \leq t_p \\ ce^{-dt}, t_p < t \leq 1 \end{cases}$$

Onde $i(t)$= intensidade normalizada da precipitação; t= tempo normalizado do evento; t_p= tempo normalizado de ocorrência da intensidade máxima; i_p= intensidade máxima normalizada; a, b, ce d= parâmetros da função dupla exponencial.

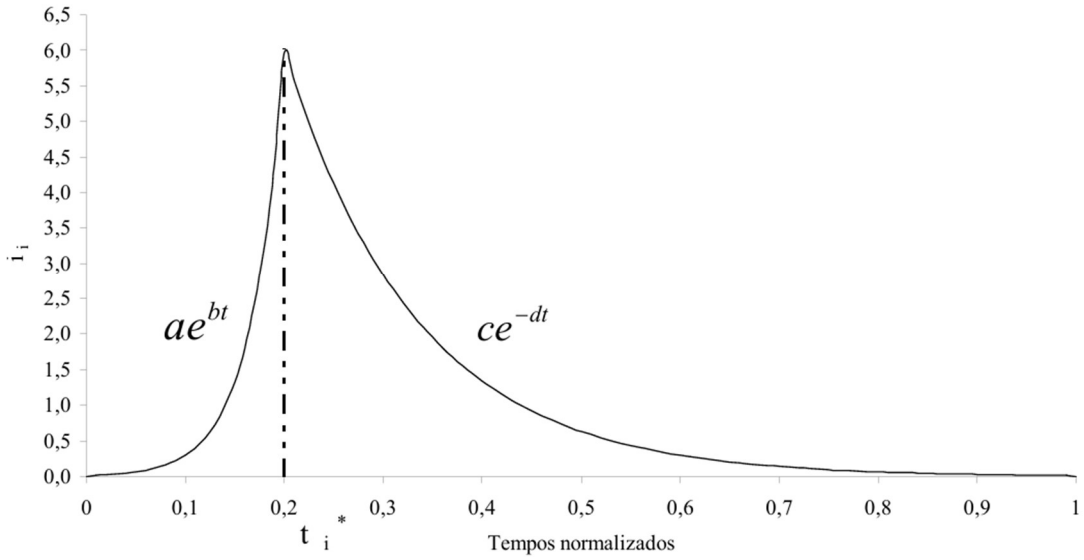


Figura 3 – Representação gráfica da função dupla exponencial e das variáveis associadas.

A função é definida a partir dos parâmetros a, b, ce d, estimados em função da duração do evento, da intensidade máxima instantânea e da posição temporal do pico de intensidade.

$$1 - e^{-bt_p} = \frac{bt_p}{i_p}$$

$$d = \frac{bt_p}{1 - t_p}$$

$$a = i_{pe}^{-bt_p}$$

$$c = i_{pe}^{dt_p}$$

Quando não foi possível obter convergência para o parâmetro b, adotou-se como alternativa um perfil triangular, conforme recomendado por Oliveira (2003). Os hietogramas reconstruídos foram posteriormente discretizados em intervalos de 10 minutos para o cálculo das propriedades dos eventos de precipitação.

Avaliação do método

A avaliação da capacidade do método em reproduzir as características dos eventos pluviométricos foi realizada por meio da comparação entre os dados observados sub horários e os eventos sintéticos gerados pelo CLIMABR. Para isso, foram analisadas as distribuições da precipitação total do evento (Prec), duração do evento (Dur), intensidade máxima em 30 minutos (I30) e erosividade da chuva (EI30).

A escolha dessas duas últimas grandezas se deve a sua importância em estudos hidrológicos e sua dependência de uma representação adequada do perfil de intensidade da chuva intra-evento.

Inicialmente, utilizaram-se histogramas para comparar a frequência de ocorrência dos valores observados e sintéticos, permitindo identificar diferenças na distribuição e dispersão das variáveis. Em seguida, foram empregadas curvas de densidade por estimativa de kernel (Kernel Density Estimation – KDE), que fornecem uma representação contínua da distribuição de probabilidade dos dados, possibilitando uma avaliação mais detalhada da forma das distribuições.

Por fim, foram construídos gráficos quantil-quantil (QQ-plots), nos quais os quantis das séries sintéticas foram comparados aos quantis das séries observadas. Nesse tipo de gráfico, quanto mais próximos os pontos estiverem da reta de referência 1:1, maior é a semelhança entre as distribuições analisadas, permitindo avaliar a capacidade do modelo em reproduzir não apenas os valores médios, mas também a variabilidade e os eventos extremos.

RESULTADOS

A capacidade do ClimaBR foi avaliada por meio da comparação entre as séries sintéticas geradas para um ano aleatório utilizando como entrada as propriedades de uma estação do CEMADEN e os eventos observados com resolução sub horária da mesma estação para um ano real.

Foram analisadas quatro propriedades fundamentais dos eventos de chuva: precipitação acumulada do evento (Prec), duração (Dur), intensidade máxima em 30 minutos (I30) e erosividade do evento (EI30) por ser uma variável que é muito influenciada pela estrutura intraevento da chuva. A comparação foi realizada utilizando Histogramas, distribuições de frequência Kernel Density Estimation (KDE) e gráficos Quantile–Quantile (Q–Q), permitindo avaliar tanto a forma das distribuições quanto a representação dos diferentes quantis das variáveis analisadas.

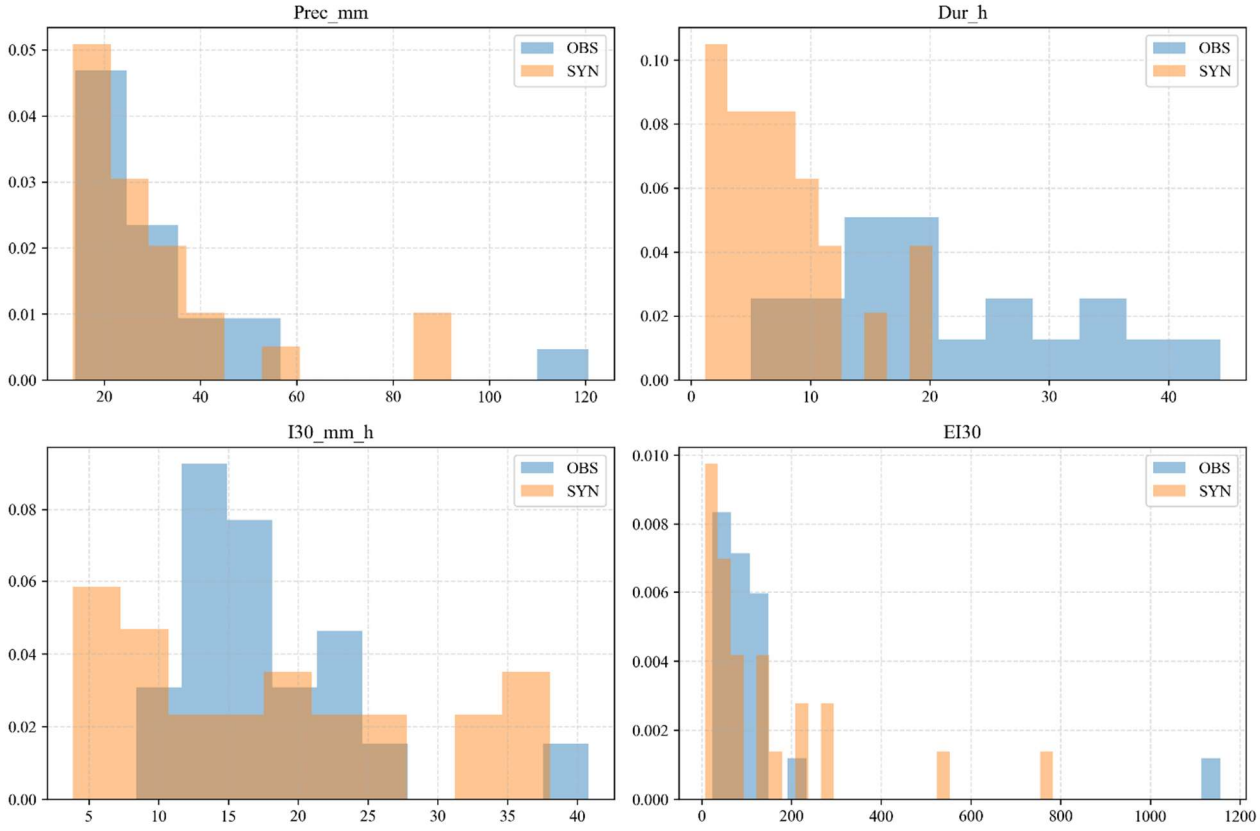


Figura 4 – Histogramas das frequências (Precipitação, Duração, I30 e EI30)

Na figura 4 a distribuição da precipitação total dos eventos sintéticos apresentou uma concordância com os dados observados sub horários para os eventos de pequena e média magnitude. Tanto os histogramas quanto as curvas de densidade mostram uma sobreposição entre as distribuições, significando que o CLIMABR reproduz bem a frequência dos eventos mais comuns. Entretanto,

observa-se uma leve superestimação dos eventos em torno de 80–90 mm e uma subestimação dos eventos extremos acima de aproximadamente 100 mm.

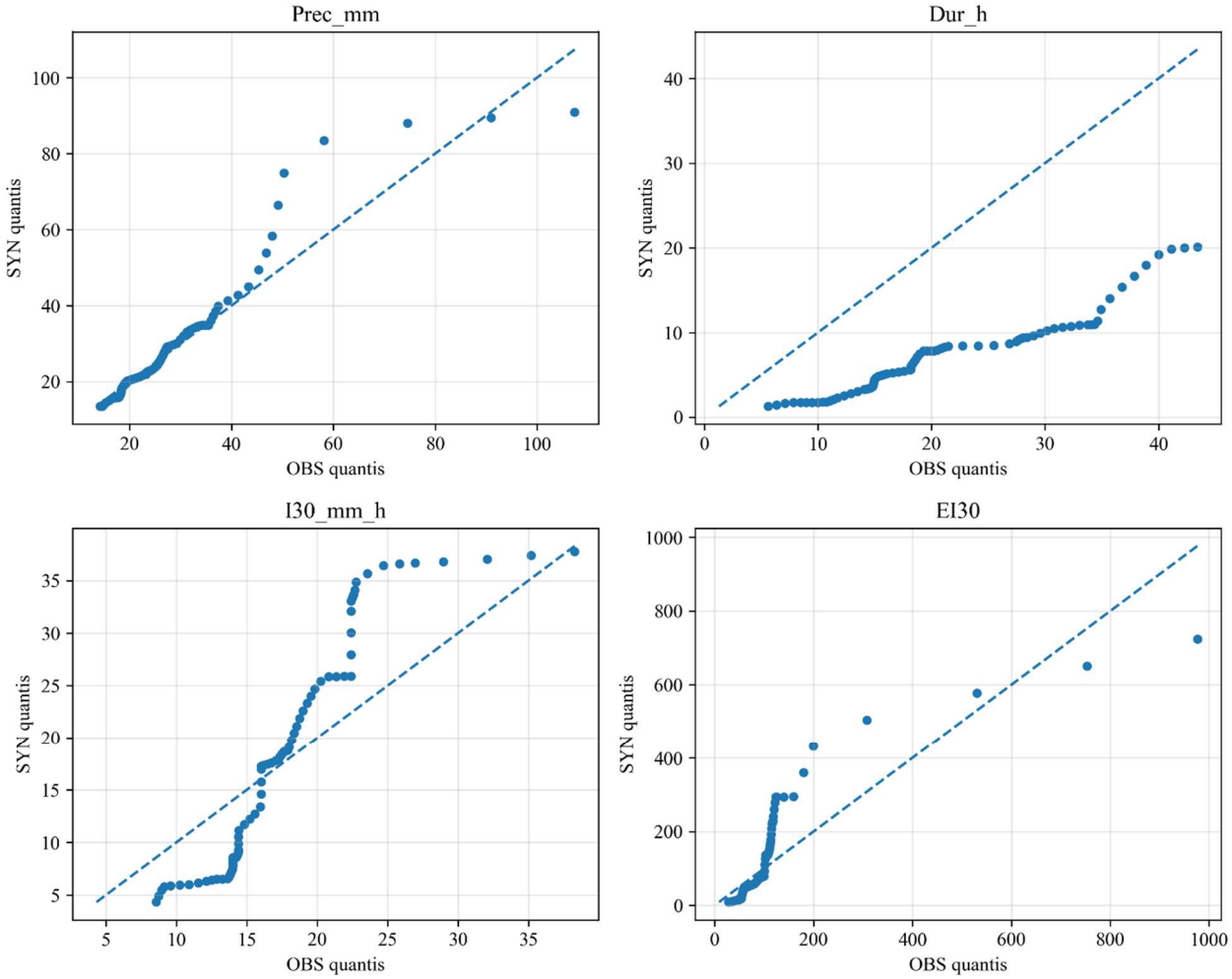


Figura 5 – Quantile – Quantile (Precipitação, Duração, I30 e EI30)

De acordo com a figura 5, precipitação acumulada apresentou boa concordância entre os quantis centrais, com desvios crescentes para os maiores eventos. A duração foi sistematicamente subestimada pelo CLIMABR, mostra a dificuldade do modelo em reproduzir eventos de longa duração. Em contrapartida, os quantis da intensidade máxima em 30 minutos foram superestimados para parte da distribuição, causando uma compensação entre menores durações e maiores intensidades. Como consequência, os quantis do EI30 apresentaram boa concordância para eventos de baixa e média erosividade, porém com tendência de subestimação nos eventos mais erosivos, indicando limitações do modelo na representação dos extremos.

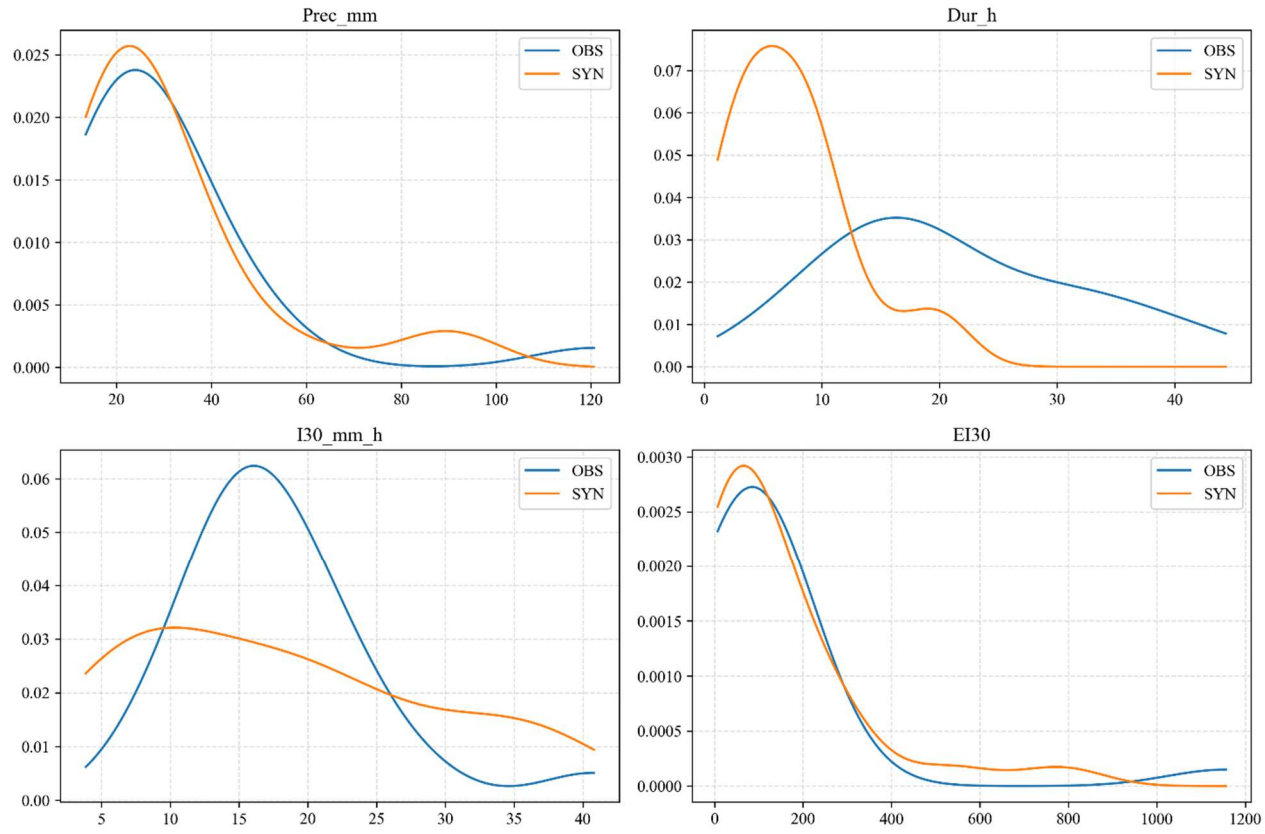


Figura 6 – Distribuição de frequencia KDE (Precipitação, Duração, I30 e EI30)

As distribuições de densidade mostram diferenças na capacidade do ClimaBR em reproduzir as propriedades dos eventos. A precipitação acumulada (Prec) apresentou boa concordância com os dados observados para os eventos de baixa e média magnitude. Para a duração (Dur), as séries sintéticas concentraram-se em eventos mais curtos, indicando uma sub-representação de eventos longos.

A intensidade máxima em 30 minutos (I30) apresentou distribuição mais dispersa que a observada, com maior frequência de eventos de intensidade intermediária. Apesar dessas diferenças nas variáveis individuais, a distribuição do EI30 foi reproduzida de forma satisfatória, especialmente para os eventos de menor e média erosividade, indicando que o modelo é capaz de representar adequadamente a distribuição geral da erosividade dos eventos.

CONCLUSÃO

O CLIMABR representa o perfil temporal de cada evento utilizando uma função dupla exponencial ajustada a partir da intensidade máxima instantânea padronizada e do tempo padronizado de ocorrência do pico. Dessa forma, o modelo fornece as informações necessárias para representar a distribuição temporal da precipitação em cada evento sintético.

Ademais, a metodologia proposta mostrou-se eficiente para o tratamento de grandes volumes de dados, proporcionando maior reprodutibilidade, padronização e rapidez na execução das análises. Os resultados obtidos indicaram desempenho satisfatório na reconstrução dos eventos sintéticos e na obtenção das variáveis derivadas avaliadas.

O sistema desenvolvido melhora o potencial de aplicação das séries sintéticas produzidas pelo ClimaBR, permitindo sua utilização em estudos de erosividade, conservação do solo, modelagem hidrológica e planejamento de recursos hídricos. Como trabalhos futuros, recomenda-se a implementação completa dos algoritmos originais do ClimaBR em ambiente Python, eliminando a dependência do executável proprietário e ampliando a flexibilidade do sistema.

AGRADECIMENTOS

Os autores agradecem à instituição organizadora do simpósio pela oportunidade de divulgação e discussão dos resultados desta pesquisa. Agradecem também à Fundação de Apoio à Pesquisa do Estado da Paraíba (FAPESQ) pela concessão da bolsa de pesquisa, fundamental para o desenvolvimento das atividades realizadas. Por fim, os autores expressam seu agradecimento ao Programa de Pós-Graduação em Engenharia Civil e Ambiental (PPGECAM) da Universidade Federal da Paraíba (UFPB) pelo apoio institucional, infraestrutura e ambiente acadêmico que possibilitaram a execução deste trabalho.

REFERÊNCIAS

OLIVEIRA, V. P. S. (2003). Modelo de geração de séries sintéticas de precipitação. Tese de Doutorado, Universidade Federal de Viçosa.

BARBOSA, L. R.; ALMEIDA, C. N.; COELHO, V. H. R.; FREITAS, E. S.; C. O. GALVÃO; ARAÚJO, J. C. Sub-hourly rainfall patterns by hyetograph type under distinct climate conditions in Northeast of Brazil: a comparative inference of their key properties. REVISTA BRASILEIRA DE RECURSOS HÍDRICOS, v. 23, p. 1-14, 2018

GABRIEL, K.R. AND NEUMANN, J. 1962: A Markov chain model for daily rainfall occurrence at Tel Aviv. Quarterly Journal of the Royal Meteorological Society 88, 90–95.

MIN, Y. M.; KRYJOV, V. N.; AN, K. H. et al. *Evaluation of the weather generator CLIGEN with daily precipitation characteristics in Korea*. *Asia-Pacific Journal of Atmospheric Sciences*, 47, 255–263, 2011. doi:10.1007/s13143-011-0014-y.

TODOROVIC, P. AND WOOLHISER, D.A. 1975: A stochastic model of n-day precipitation. Journal of Applied Meteorology 14, 17–24.

USDA-ARS. *General Description of the CLIGEN Model and its History*. Washington, DC: Agricultural Research Service, United States Department of Agriculture