

XVII SIMPÓSIO DE RECURSOS HÍDRICOS DO NORDESTE

DOWNLOAD E PRÉ-PROCESSAMENTO DE DADOS HIDROCLIMÁTICOS DO HIDROWEB/ANA COM O PACOTE HYDROBR

Arthur Telles Calegario^[1]; Calebe Moraes de Araújo Assis^[1]; Clara Lopes Faria Brumano^[1];
Carlos Roberto Assis Damasceno^[1]; Jéssica Aragão^[1]; Ricardo Santos Silva Amorim^[1]; Michel
Castro Moreira^[1]; Demetrius David da Silva^[1]; Gilbert Vaz Junior^[1] & Artur Moisés Gonçalves
Lourenço^[2]

RESUMO – A aquisição e o pré-processamento de dados hidrológicos são essenciais para estudos nesta área. Este trabalho apresenta as funcionalidades do pacote *hydrobr*, desenvolvido no ambiente R, para download e pré-processamento de dados hidrológicos do portal Hidroweb da Agência Nacional de Águas e Saneamento (ANA). Foi realizado um estudo de caso nas bacias dos rios Piranhas, Paraíba e na zona hidrológica do litoral do Ceará e Paraíba. As estações fluviométricas (FLU) e pluviométricas (PLU) na área de estudo foram inventariadas, e seus dados foram adquiridos e organizados. Foram selecionadas estações com, no mínimo, 30 anos de dados, consistidos ou não, com falhas anuais inferiores a 5% no período de 1980 a 2024, considerando janeiro como início do ano hidrológico. O tempo de processamento das estações FLU e PLU foi de aproximadamente 54,3 e 13,6 minutos, respectivamente. Foram inventariadas 596 estações FLU e 798 PLU, das quais apenas 21 FLU e 58 PLU atenderam aos critérios estabelecidos. O baixo tempo de processamento permite o estudo de diferentes cenários com critérios variados de seleção. O *hydrobr* é um pacote gratuito e de código aberto que automatiza as análises, contribuindo para a reprodutibilidade e robustez dos estudos hidrológicos.

Palavras-Chave – *hydrobr*, *hydroversebr*, hidroweb, ANA

ABSTRACT - The acquisition and preprocessing of hydrological data are essential for studies in this field. This work presents the functionalities of the *hydrobr* package, developed in the R environment, for downloading and preprocessing hydrological data from the Hidroweb portal of the National Water and Sanitation Agency (ANA). A case study was conducted in the river basins of Piranhas, Paraíba, and in the hydrological zone of the coast of Ceará and Paraíba. The streamflow (FLU) and rainfall (PLU) stations in the study area were inventoried, and their data were acquired and organized. Stations with at least 30 years of data, whether consistent or not, with annual gaps of less than 5% during the period from 1980 to 2024 were selected, considering January as the start of the hydrological year. The processing time for FLU and PLU stations was approximately 54.3 and 13.6 minutes, respectively. A total of 596 FLU stations and 798 PLU stations were inventoried, of which only 21 FLU and 58 PLU met the established criteria. The low processing time allows for the study of different scenarios with varied selection criteria. *hydrobr* is a free, open-source package that automates analyses, contributing to the reproducibility and robustness of hydrological studies.

Key words – *hydrobr*, *hydroversebr*, hidroweb, ANA

¹ Universidade Federal de Viçosa, Departamento de Engenharia Agrícola, (31)3612-4004, tcalegario@gmail.com, calebe.assis@ufv.br, clara.brumano@ufv.br, carlos.damasceno@ufv.br, rsamorim@ufv.br, michelcm@ufv.br, demetrius@ufv.br; gilbert.junior@ufv.br

² Instituto Federal da Paraíba, Campus Princesa Isabel, (83) 99836-3759, artur.lourenco@ifpb.edu.br

1 INTRODUÇÃO

Os dados fluviométricos e pluviométricos são fundamentais na hidrologia pois é a ciência que estuda a distribuição, a dinâmica e a qualidade do ciclo da água na Terra. Estes dados integrados são cruciais para auxiliar os estudos de previsão de inundações e secas, gestão e planejamento de recursos hídricos, impactos ambientais dentre outros (Tucci, 2013).

O download e o pré-processamento de dados na hidrologia é de grande importância em decorrência da grande quantidade de dados de entrada para modelos e estudos hidrológicos. Segundo Magacho (2022), a execução de modelos hidrológicos é computacionalmente intensiva devido à quantidade de dados de entrada e aos múltiplos passos de processamento subsequentes.

De acordo com Melo (2008), a necessidade de se converter a base de dados disponível para uma forma que se possa trabalhar no software desejado faz com que o eventual uso de programas de conversão de formatos de dados gere elevada demanda de tempo nos estudos hidrológicos. Desta forma, o uso do pacote *hydrobr* (Calegario, Althoff e Lourenço, 2024), em ambiente R, pode auxiliar o processo de obtenção dos dados, baixando-os do portal Hidroweb da ANA (Agência Nacional de Águas e Saneamento Básico) e realizando a função de filtragem dos dados, podendo fazer a checagem de qualidade dos mesmos, selecionando dados consistidos e que não apresentem falhas.

Neste trabalho são apresentadas as funcionalidades do pacote *hydrobr* como ferramenta para realização de um trabalho colaborativo em software R (R Core Team, 2024). Para fins de análise das funcionalidades do pacote, realizou-se todo o processo de identificação, download e organização dos dados assim como seleção das estações fluviométricas (FLU) e pluviométricas (PLU) presentes nas áreas das bacias do rio Piranhas, Paraíba e da zona hidrológica do litoral do Ceará e Paraíba. Além disso, estimou-se o tempo necessária para tal processamento.

2 MATERIAL E MÉTODOS

2.1 Desenvolvimento metodológico

Na Figura 1 é apresentado um diagrama contendo as etapas (funções) necessárias para o download dos dados das estações hidrometeorológicas (precipitação e vazão) e pré-processamento dos dados hidrológicos, assim como os produtos gerados em cada etapa.

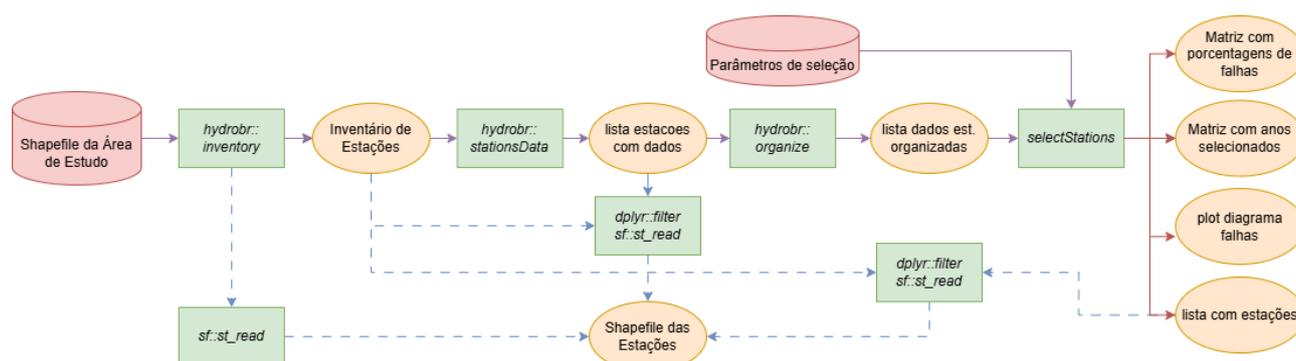


Figura 1 – Diagrama das etapas a serem executadas para o download e pré-processamento dos dados hidrológicos do Hidroweb/ANA

O pré-processamento consistiu no inventário, organização da base de dados e filtragem das estações com base no período-base adotado, máxima porcentagem de falhas admitida e o número de anos mínimos. Nos tópicos seguintes, são apresentadas as funções do *hydrobr* utilizadas para realização do estudo assim como os parâmetros de entrada. Para realização do processamento também foram utilizados os pacotes “*sf*” e “*dplyr*”.

hydrobr::inventory

O primeiro passo no processamento é fazer o download do inventário de estações da ANA. Para isso, utilizou-se a função *hydrobr::inventory*. Os parâmetros de entrada necessários são: (1) o *shapefile* da área de estudo ou a definição do estado desejado; e (2) a definição do tipo de estação de interesse (FLU ou PLU). O resultado obtido é um objeto "simple feature" contendo o inventário de estações na área de estudo, o qual é exportado como *shapefile* utilizando a função *sf::st_write*.

hydrobr::stationsData

O próximo passo consistiu no download do banco de dados de cada estação a partir do inventário obtido. Para isso, utilizou-se a função *hydrobr::stationsData*. Os parâmetros de entrada necessários são: (1) o objeto "simple feature" obtido a partir da função *hydrobr::inventory*; e (2) a definição da necessidade, ou não, de manter estações sem dados nos resultados.

O resultado obtido é uma lista em que cada elemento contém os dados de uma estação individualizada. A partir dessa lista, foram filtradas (utilizando *dplyr::filter*) do inventário de estações aquelas que continham dados, sendo essas, por fim, exportadas como *shapefile* utilizando a função *sf::st_write*.

hydrobr::organize

A função *hydrobr::organize* é utilizada para organizar os dados em séries temporais. Além disso, ela seleciona as colunas de data, vazão (ou precipitação) e nível de consistência. O resultado final é uma lista em que cada elemento contém os dados da estação individualizada organizados.

hydrobr::selectStations

O último passo no processo consiste na seleção das estações com base nos critérios de porcentagem de falha, nível de consistência, período-base (anos analisados), número mínimo de anos e mês de início do ano hidrológico.

Nesta etapa foi utilizado a função *hydrobr::selectStations*. Como parâmetros de entrada são necessários: (1) lista de estações obtidas na função *hydrobr::organize*; (2) mês de início do ano hidrológico; (3) anos de início e fim do período-base considerado; (4) modo no qual as porcentagens de falhas serão analisadas (mensal ou anual); (5) porcentagem de falhas mínimas aceitas; (6) número mínimo de anos completos na série.

O resultado consta de uma lista com quatro elementos, sendo esses: (1) listas individuais das estações selecionadas com respectiva série temporal; (2) matriz de falhas indicando anos (ou meses) selecionados nas estações que atenderam aos critérios de seleção; (3) matriz de porcentagem indicando a porcentagem de falhas anual (ou mensal) das estações selecionadas; e (4) bloco diagrama com estações selecionadas e anos selecionados, além da classificação da série histórica (anual ou mensal) segundo nível de consistência.

2.2 Estudo de caso

Como estudo de caso demonstrativo do algoritmo de download e pré-processamento dos dados do HidroWeb/ANA foram selecionadas as bacias que integram o estado da Paraíba. Utilizou-se o shapefile das sub-bacias nível 1 do Plano Nacional de Recursos Hídricos (ANA, 2007) para identificar as bacias do Litoral CE/PB, do rio Piranhas e do rio Paraíba (Figura 2). A área das bacias são, respectivamente, 30.286, 59.701 e 19.679 km².

Para o pré-processamento e filtragem dos dados foram considerados os seguintes critérios: (1) início do ano hidrológico no mês de janeiro; (2) anos com no máximo 5% de falhas; (3) período base de 1980 a 2024; e (4) estações que apresentavam 30 anos de dados. Cabe salientar que esses parâmetros foram adotados apenas para efeito demonstrativo do pacote. Para um estudo hidrológico mais robusto, é necessário realizar análises que otimizem a relação entre número de estações, anos de

dados e período-base. Além disso, a escolha do mês de início do ano hidrológico deve ser analisada conforme a finalidade do estudo a ser realizado.

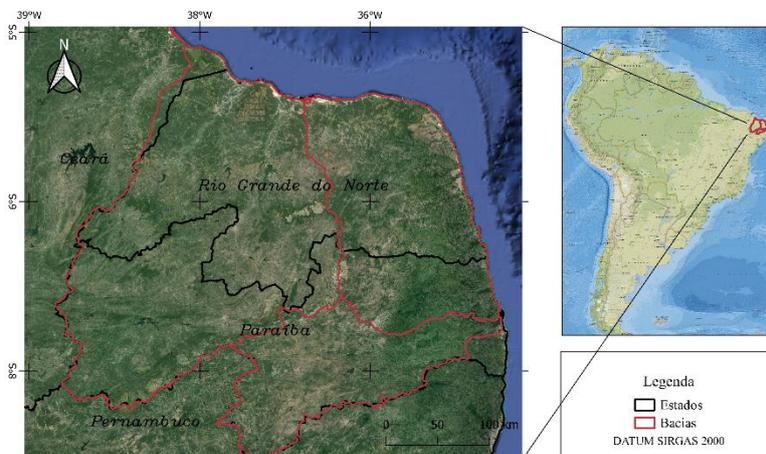


Figura 2 – Localização da área de estudo.

O computador utilizado para o processamento possui um processador “13th Gen Intel(R) Core(TM) i7-13700H 2.40 GHz”, 64 GB de RAM e “Windows 11 Home Single Language” como sistema operacional. A velocidade da rede constatada no momento do processamento foi de 300 Mbps, medida pelo site MINHA CONEXÃO (2024).

3. RESULTADOS E DISCUSSÃO

hydrobr::Inventory

O número de estações FLU e PLU inventariadas na área de estudo foram de 596 e 798, respectivamente. O tempo de processamento dessa etapa foi de 10 segundos para ambas. Os mapas com inventário de estações FLU e PLU são apresentados nas Figuras 3A1 e 3B1, respectivamente.

hydrobr::stationsData

O total de estações FLU com dados foi de 86 (Figura 3A2), enquanto para as estações PLU foi de 499 (Figura 3B2). O tempo de processamento para esta etapa foi de 10,9 minutos para as estações FLU e 33,6 minutos para as estações PLU. A título de exemplo, é apresentada na Tabela 1 uma parte dos dados disponibilizados no HidroWeb/ANA para a estação fluviométrica 37080000.

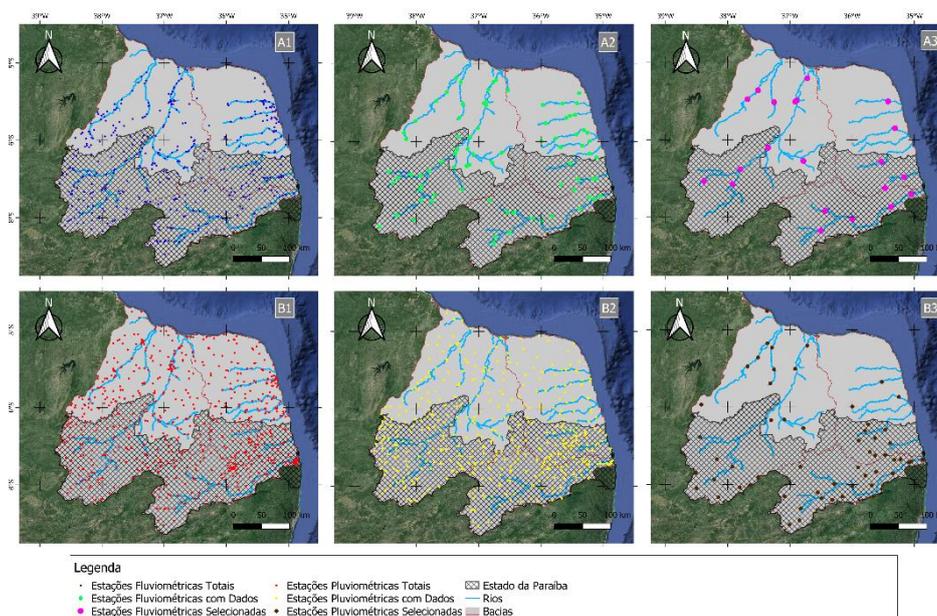


Figura 3 - Distribuição e seleção de estações fluviométricas (FLU) e pluviométricas (PLU): (A1) inventário de estações FLU e (B1) PLU; (A2) estações FLU e (B2) PLU com dados disponíveis; (A3) estações FLU e (B3) PLU selecionadas conforme os critérios adotados.

Tabela 1 – Amostra dos dados disponibilizados no hidroweb/ANA para a estação fluviométrica 37080000

Estação código	Nível consistência	Data	Média diária	Método obtenção vazões	Máxima	Mínima	Média
37080000	2	01/01/1973	1	1	1,6993*	1,6993*	1,6993*
37080000	2	02/01/1973	1	1	1,6993*	1,6993*	1,6993*
37080000	2	03/01/1973	1	1	19,4302	1,6993	3,5274

*os valores de vazões repetidos constam na base de dados da estação 37080000 da base de dados da ANA

hydrobr::organize

Posteriormente, os dados foram reorganizados utilizando a função *hydrobr::organize*. O tempo de processamento da função *hydrobr::organize* foi de 2,1 minutos para as estações FLU e 17,4 minutos para as estações PLU. A título de exemplo, é apresentada na Tabela 2 uma parte dos dados reorganizados disponibilizados no HidroWeb/ANA para a estação 37080000.

Tabela 2 – Amostra dos dados disponibilizados no hidroweb/ANA para estação fluviométrica 37080000 reorganizados pela função *hydrobr::organizes*

Código da estação	Nível de consistência	Data	Vazão (m ³ s ⁻¹)
37080000	2	01/01/1973	1,6993*
37080000	2	01/02/1973	1,6993*
37080000	2	01/03/1973	1,6993*

*os valores de vazões repetidos constam na base de dados da estação 37080000 da base de dados da ANA

hydrobr::selectStations

A partir dos parâmetros adotados para a seleção das estações, foram obtidas 21 estações FLU e 58 estações PLU. O tempo de processamento dessa etapa foi de 30 segundos para as estações FLU e 3,2 minutos para as estações PLU. Os mapas com o inventário das estações PLU e FLU selecionadas são apresentados na Figura 3A3 e 3B3, respectivamente.

O resultado da função *hydrobr::selectStations* é uma lista contendo 4 elementos. O primeiro deles é uma lista cujos elementos são as séries temporais de vazões individualizadas (tibble). O resultado para a estação 37080000 é apresentado na Tabela 3.

Tabela 3 – Amostra dos dados disponibilizados no hidroweb/ANA para estação 37080000 reorganizados pela função *hydrobr::selectStations*

Código da estação	Nível de consistência	Data	Vazão (m ³ s ⁻¹)	Ano civil	Mês do ano civil	Ano hidrológico	Mês do ano hidrológico	Máximo de falhas
37080000	2	01/01/1980	0*	1980	01/01/1980	1980	01/01/1980	<= 5%
37080000	2	01/02/1980	0*	1980	01/01/1980	1980	01/01/1980	<= 5%
37080000	2	01/03/1980	0*	1980	01/01/1980	1980	01/01/1980	<= 5%

*os valores de vazões repetidos constam na base de dados da estação 37080000 da base de dados da ANA

Na Tabela 3 constam o código da estação, o nível de consistência, a data, a vazão, o ano civil, o mês do ano civil, o ano hidrológico, o mês do ano hidrológico, assim como o máximo de falhas permitido considerando o ano hidrológico.

O segundo resultado da função *hydrobr::selectStations* é uma matriz que indica quais anos foram filtrados nas respectivas estações. Ou seja, anos com porcentagem maior que 5%, não são apresentados na série temporal resultante da função *hydrobr::selectStations*. Na Tabela 4 estão apresentados os resultados para algumas estações e anos hidrológicos, na qual os anos desconsiderados são representados por *FALSE*, enquanto os anos selecionados são representados por *TRUE*.

Tabela 4 – Amostra da matriz de falhas indicando anos hidrológicos selecionados (TRUE) nas estações que atenderam aos critérios de seleção pela função *hydrobr::selectstations*

Ano hidrológico	37080000	37084000	37185000	37220000	37380000	37410000	37470000
1980	TRUE	-	TRUE	FALSE	TRUE	TRUE	TRUE
1981	TRUE	-	TRUE	FALSE	TRUE	TRUE	TRUE
1982	TRUE	-	TRUE	FALSE	TRUE	TRUE	TRUE

O terceiro resultado da função `hydrobr::selectStations` é uma matriz que indica a porcentagem de falhas nos anos hidrológicos das respectivas estações selecionadas. Uma amostra dos resultados para esta função é apresentada na Tabela 5.

Tabela 5 – Amostra da matriz de porcentagem de falhas nos anos hidrológicos nas estações que atenderam aos critérios de seleção pela função `hydrobr::selectstations`

Ano hidrológico	37080000	37084000	37185000	37220000	37380000	37410000	37470000
1980	0	-	0	100	0	0	0
1981	0	-	0	100	0	0	0
1982	0	-	0	100	0	0	0

O quarto resultado é o diagrama de falhas e nível de consistência das respectivas estações e anos hidrológicos para o período base considerado (Figura 4). Os anos que apresentam cor azul escuro são aqueles cujos dados foram consistidos e apresentam menos de 5% de falha. Da mesma maneira, as células em vermelho representam dados que não foram consistidos, sendo que as células com cor escura indicam menos de 5% de falhas.

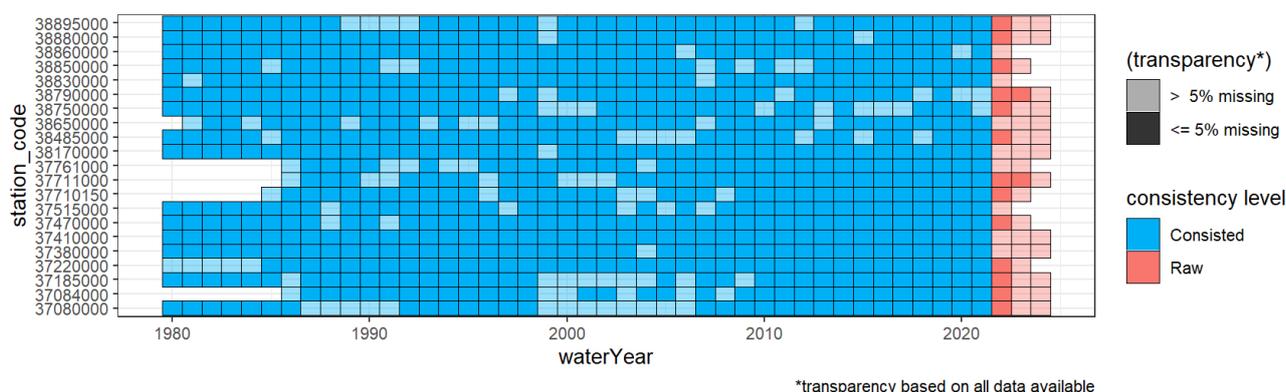


Figura 4 - Diagrama de falhas e consistência de dados das estações hidrológicas.

Na Tabela 6 tem-se o tempo necessário para realizar o todo o processamento hidrológico dos dados da ANA para área de estudo, assim como o número de estações obtidas em cada etapa do processamento.

Com base nos resultados apresentados podemos observar que o número de estações com dados que satisfazem aos critérios de menos de 5% de falhas nos anos hidrológicos e 30 anos de dados no período-base (1980-2024), sendo eles consistidos ou não, são reduzidos drasticamente. Para estações FLU e PLU a densidade da rede de estações (km²/estação) foi de 5.222 e 1.891, respectivamente, enquanto a média das densidades mínimas para estações FLU e PLU recomendadas pela Organização

Meteorológica Mundial (WMO, 2008), nas diferentes unidades fisográficas (Costal, Montanhoso, planícies interiores e ondulado), são 1.650 e 575, respectivamente.

Tabela 6. Tempo de processamento e número de estações conforme diferentes etapas do processamento

Função	Fluviométricas		Pluviométricas	
	Tempo de processamento (min)	Nº estações	Tempo de processamento (min)	Nº estações
<i>hydrobr::inventory</i>	0,1	596	0,1	798
<i>hydrobr::stationData</i>	10,9	86	33,6	499
<i>hydrobr::organize</i>	2,1	86	17,4	499
<i>hydrobr::selectStations</i>	0,5	21	3,2	58
Tempo total	13,6	-	54,3	-

O tempo de processamento das estações FLU e PLU (Tabela 6) foi de aproximadamente 1 hora e 20 minutos, sendo 13,6 min para o processamento das estações FLU e 54,3 min para as estações PLU. Tal fato demonstra a potencialidade do pacote *hydrobr*, visto que o baixo tempo de processamento permite o estudo de diferentes cenários dos critérios de seleção.

Outras potencialidades do pacote devem-se ao fator do mesmo ser gratuito, com código fonte aberto e ter como base um software amplamente utilizado por hidrólogos (Slater et al. 2019). Tais características permitem a verificação e correção de possíveis erros no código fonte e, dessa maneira, torna as análises realizadas pelo *hydrobr* mais robustas.

Além disso, a concentração dessas análises em um ambiente automatizado permite a reprodutibilidade das pesquisas por outros usuários, a partir da disponibilização de *scripts*, e reduz erros humanos associados ao processamento de grande bases de dados com softwares que exigem mais interações repetidas, como por exemplo o Sistema Computacional para Análise Hidrológica (SisCaH) (Souza et al. 2009).

Não menos importante, o software R possui diversos outros pacotes desenvolvidos para análises e modelagens hidrológicas (Slater et al. 2019), permitindo maior interconexão entre diversos pacotes e o *hydrobr*. O script gerado para o processamento realizado nesse trabalho é disponibilização em HYDROVERSEBR (2024a), enquanto exemplos de utilização do pacote são apresentando em HYDROVERSEBR (2024b)

AGRADECIMENTOS ao Centro de Referências em Recursos Hídricos (CRRH) da Universidade Federal de Viçosa (UFV) pelo suporte na realização da pesquisa e na análise dos dados, às agências de fomento: Fundação de Amparo à Pesquisa do Estado de Minas Gerais (Fapemig), Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) e Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) pelo financiamento da pesquisa.

4. REFERÊNCIAS

ANA (Agência Nacional de Águas e Saneamento Básico). (2007). Nível 1 de Sub-bacias do Plano Nacional de Recursos Hídricos. Brasília: ANA. Catálogo de Meta Dados da ANA. Disponível em: <https://metadados.snirh.gov.br/geonetwork/srv/eng/catalog.search#/metadata/f50527b9-24ed-41d5-b063-b5acfb25e10d>.

Calegario A.; Althoff D.; Lourenço A. (2024). hydrobr: Access to the Brazilian National Water Agency (Agência Nacional de Águas) hydrological data. R package version 0.0.0.9000.

DUBREUIL, V. et al. (2017). Les types de climats annuels au Brésil: une application de la classification de Köppen de 1961 à 2015. *EchoGéo*, v. 3, n. 41, p. 1-27.

HYDROVERSEBR. 2024A. 01_script_SNERH_2024_hydrobr. GitHub. <https://github.com/hydroversebr/miscellaneous>. Acesso em: 26 jun 2024

HYDROVERSEBR. 2024B. Disponível em: <https://www.youtube.com/@hydroversebr>. Acesso em: 06/07/2024.

MAGACHO, H. G. B.; TELLES, W. R.; BEDO, M. (2022). Pré-processamento de dados para Modelos Hidrológicos como algoritmo k-Medoids: O caso do Rio Pomba. *Proceedings of the 37th Brazilian Symposium on Data Bases*. Búzios.

MINHA CONEXÃO. (2024). Teste de velocidade. Disponível em: <https://www.minhaconexao.com.br/>. Acesso em: 26 jun 2024

MELO, H. de A. et al. (2008). Modelos hidrológicos e sistemas de informação geográfica (SIG): integração possível. *Anais do IX Simpósio de Recursos Hídricos do Nordeste*. Vol.1. Campina Grande.

R CORE TEAM. (2024) R. A language and environment for statistical computing. Vienna: R Foundation for Statistical Computing, Disponível em: <https://www.r-project.org/> Acesso em 22 de junho de 2024,

SLATER, L. J. et al. (2019). Using R in hydrology: a review of recent developments and future directions, *Hydrol. Earth Syst. Sci.*, 23, 2939–2963, <https://doi.org/10.5194/hess-23-2939-2019>.

TUCCI, C. E. M. (2013) *Hidrologia: Ciência e Aplicação*. 4a edição. Porto Alegre: Editora UFRGS, ABRH, 944 p.

SOUSA, H. T. et al. (2008). *SisCAH 1.0 - Sistema Computacional para Análises Hidrológicas*. Agência Nacional das Águas, Brasília, DF.

WMO. (2008). **Guide to hydrological practices**. World Meteorological Organization.