

PRECIPITAÇÃO ESTIMADA PARA O ESTADO DO PARÁ – BRASIL: VALIDAÇÃO DA MALHA DE DADOS DO *CHIRPS* v2.0

Lucas Pereira Soares ¹
Marcelo de Oliveira Moura ²

RESUMO

A partir deste estudo, foi desenvolvida a validação da base de dados de precipitação estimada do *Climate Hazards Group InfraRed Precipitation with Station data (CHIRPS) version 2.0 (v2.0)* ao estado do Pará. Considerou-se a aplicação da estatística mais comum a este processo de validação, conforme indicado por Pereira et al. (2013), Paredes-Trejo, Barbosa e Lakshmi Kumar (2017), Costa et al. (2019) e Silva et al. (2020), sendo utilizados os seguintes parâmetros: Erro Médio Absoluto (*Mean Absolute Error ou MAE*), Raiz do Erro Quadrático Médio (*Root-Mean-Square Error ou RMSE*), Viés do estimador (ou Bias), Coeficiente Determinação (R²) e, por fim, verificou-se a porcentagem para super/subestimação dos valores em relação as médias analisadas. Validados, estes dados tornam-se aptos a utilização em estudos climáticos com informações pluviométricas.

Palavras-chave: Precipitação, Validação, Climatologia Geográfica.

ABSTRACT

From this study, validation of the *Climate Hazards Group InfraRed Precipitation with Station data (CHIRPS) version 2.0 (v2.0)* estimated forecast database was developed for the state of Pará. Consider the application of the most common statistic to this validation process, as indicated by Pereira et al. (2013), Paredes-Trejo, Barbosa and Lakshmi Kumar (2017), Costa et al. (2019) and Silva et al. (2020), using the following parameters: Mean Absolute Error (MAE), Root Mean Square Error (Root-Mean-Square Error or RMSE), Estimator Bias (or Bias), Determination Coefficient (R²) and, finally, the percentage for over/underestimation of values in relation to proven averages is obtained. Once validated, these data become suitable for use in climate studies with rainfall information.

Keywords: Precipitation, Validation, Geographic Climatology.

INTRODUÇÃO

Devido à dificuldade em se obter uma malha de informações confiáveis e de boa resolução espaço-temporal, principalmente nas regiões que só recentemente tiveram uma ocupação territorial mais ampla, como é o caso da região Norte do Brasil, e em particular o estado do Pará, tornou-se comum, no desenvolvimento de estudos climáticos sob a alça geográfica, o uso de dados provenientes de precipitação estimada.

¹ Doutorando do PPGG-UFPB, contato: lucas.soares@ifpa.edu.br;

² Docente PPGG-UFPB, contato: geommoura@yahoo.com.br



No entanto, tais informações, provenientes, principalmente, de radares e satélites, ou de uma composição híbrida entre estes e dados observacionais, carecem de validação estatística para o seu pleno uso. Por essa necessidade, recorreu-se a literatura já edificada na comunidade científica, com destaque aos trabalhos de Pereira et al. (2013), Paredes-Trejo, Barbosa e Lakshmi Kumar (2017), Costa et al. (2019) e Silva et al. (2020), para fundamentar em princípio uma proposta de validação estatística dos dados do *Climate Hazards Group InfraRed Precipitation with Station data (CHIRPS) v2.0* (FUNK, C. et al. 2015) ao estado do Pará.

Partiu-se do entendimento que os dados CHIRPS v2.0 fornecem uma base de informações pluviométricas com boa resolução espacial, ~5 km para tamanho de grade, compondo um amplo rol, tanto sob o ponto de vista espacial, visto que tais informações são dispostas para todo o território paraense, como temporal, em uma disponibilidade de dados diários, mensais e anuais, dentro do intervalo temporal datado à pesquisa, entre 1990 e 2019.

Assim, nos limites do estado do Pará (Figura 1), procurou-se, sob o referencial teórico citado, definir os parâmetros estatísticos a serem utilizados na avaliação da similaridade entre os dados observacionais com aqueles dados gerados pelo *CHIRPS v2.0*, desenvolvendo a partir disto a validação da malha de dados pertinente a pesquisa.

Figura 1 – Localização do estado do Pará no território brasileiro.





METODOLOGIA

O roteiro metodológico é composto pela montagem do banco de dados observacionais e de modelagem pluviométrica que, organizados em torno dos parâmetros estatísticos, fomentam a validação dos dados *CHIRPS v2.0* conforme os procedimentos esboçados a seguir:

Espelhamento entre a malha montada pelos dados de superfície, coletada nos sítios *on line* da Agência Nacional de Águas (ANA) e do Instituto Nacional de Meteorologia (INMET), com a base do *CHIRPS v2.0*. Desta base superficial (Figura 2), foram organizadas as informações de 32 postos pluviométricos oriundos da ANA e um total de 9 estações meteorológicas, advindas do INMET.

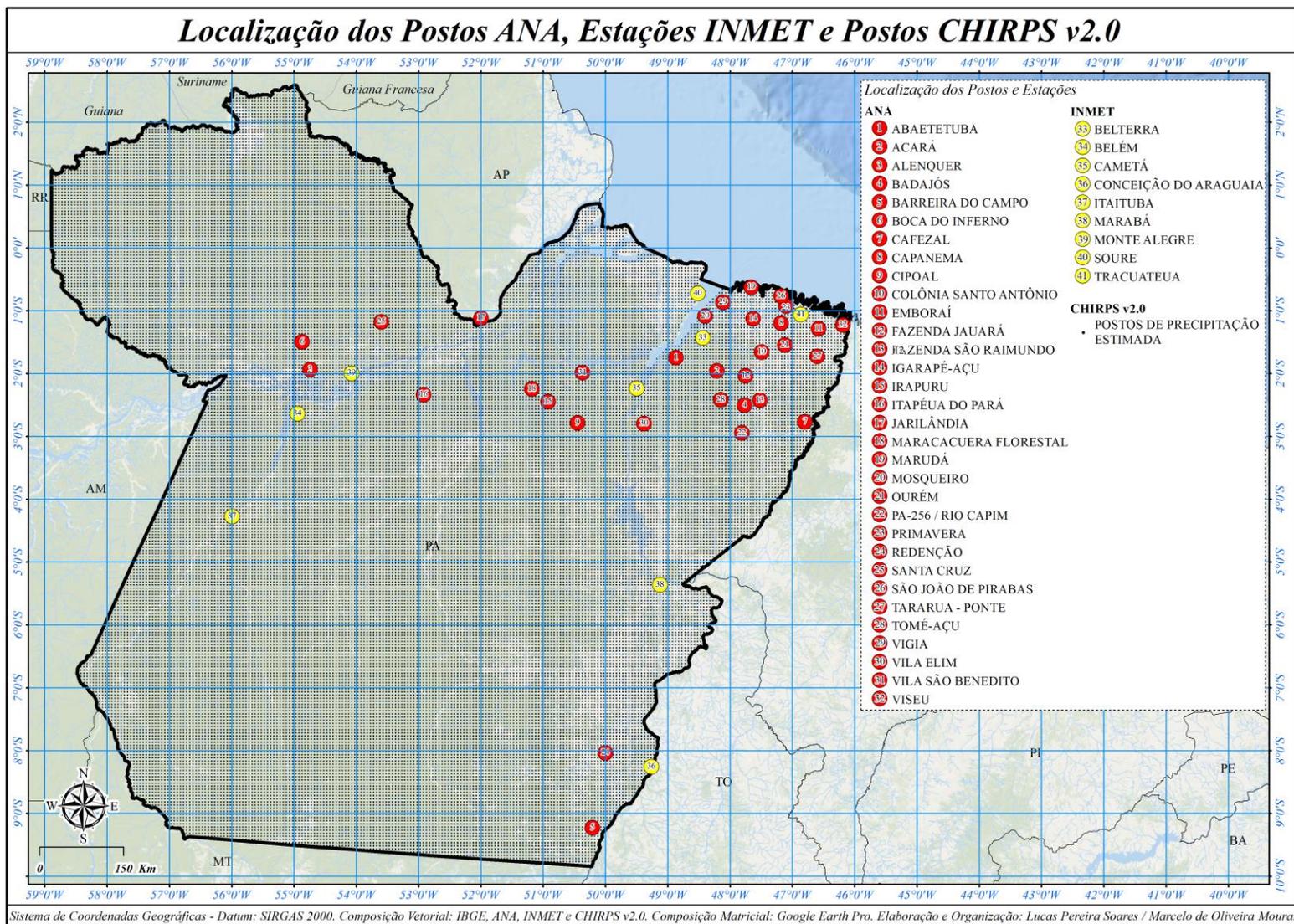
A escala espacial destes dados constitui-se sob os limites territoriais paraenses abrangendo um período de 30 anos, entre 1990 e 2019, sendo, predominantemente, formada por uma série temporal completa ou com ausências de dados pontuais. Observa-se uma distribuição não-homogeneizada destes postos e estações, que justifica a necessidade de compor uma base de dados mais robusta, como aquela disponibilizada pelo *CHIRPS v2.0*.

Para a validação dos dados *CHIRPS v2.0*, utilizou-se a localização dos 32 postos da ANA e das 9 estações do INMET, importando-os em formato *.xls* do *excel* para o *ArcGIS Pro*. A espacialização destes postos e estações deu-se por meio de sua localização de grade, considerando as coordenadas X (Longitude) e Y (Latitude) para a criação de pontos de grade vetoriais, posteriormente exportados ao formato vetorial. Após, procedeu-se a extração dos dados de precipitação estimada através do espelhamento, entre a localização geográfica da base de dados observacionais e base matricial do *CHIRPS v2.0*, pelo uso da ferramenta *extract values to points*, disponível no *ArcGIS Pro*.

Pela ferramenta *extract values to points* foram criados 30 arquivos *shapefiles* do tipo ponto, representativos de cada ano, aglutinando, dessa forma, em cada composição vetorial, os dados mensais extraídos dos arquivos matriciais. Após isso, os vetores, organizados por ano, com os dados mensais de precipitação estimada, foram convertidos em arquivos de texto e importados como arquivos do *excel*. Os dados de precipitação oriundo dos postos da ANA, das estações do INMET e da precipitação estimada do *CHIRPS v2.0* foram organizados em médias mensais para o período entre 1990 e 2019, a fim de verificar a similaridade entre estes.



Figura 2 - Localização dos postos ANA, estações INMET e postos CHIRPS v2.0.



Considerou-se a estatística mais comum a este processo de validação, conforme indicado por Pereira et al. (2013), Paredes-Trejo, Barbosa e Lakshmi Kumar (2017), Costa et al. (2019) e Silva et al. (2020), sendo utilizados os seguintes parâmetros: Erro Médio Absoluto (*Mean Absolute Error* ou *MAE*), Raiz do Erro Quadrático Médio (*Root-Mean-Square Error* ou *RMSE*), Viés do estimador (ou *Bias*), Coeficiente Determinação (R^2) e, por fim, verificou-se a porcentagem para super/subestimação dos valores em relação as médias analisadas.

Os resultados para o *MAE* e o *RMSE* são apresentados em milímetros (mm), quanto menor o valor desses parâmetros, melhor o modelo de previsão, neste caso, melhor é a previsão do *CHIRPS v.2.0* em relação a precipitação observada. Conforme BRUCE e BRUCE (2019, p. 185), na avaliação de modelos de previsão, o *RMSE* “é a métrica de desempenho mais importante da perspectiva da ciência de dados”, pois mede a precisão geral do modelo, e quando analisado em conjunto ao resultado do coeficiente de determinação (R^2), determina se o modelo ajusta bem os dados (BRUCE e BRUCE, 2019).

Conforme Miguel e Rennó (2020, p. 330), o R^2 é “utilizado para medir a concordância entre a chuva observada pelas estações e a chuva estimada [...], sem levar em conta a magnitude do erro”. Já as métricas *MAE* e *RMSE* representam essa magnitude, no entanto, sem indicar a “direção” deste erro, que é apontada na avaliação de Viés do estimador, ao informar se o modelo de previsão, sendo o *CHIRPS v.2.0*, está a subestimar a chuva, para o caso de Viés negativo, ou superestimar, se o Viés for positivo, em ambas as situações, valores próximos a zero indicam que o modelo é adequado à previsão (RABIEI e HABERLANDT, 2015).

RESULTADOS E DISCUSSÃO

Os parâmetros estatísticos utilizados na avaliação da similaridade entre os dados observacionais, provenientes dos 32 postos pluviométricos da ANA e das 9 estações meteorológicas convencionais do INMET, com aqueles dados gerados pelos *CHIRPS v.2.0*, encontram-se apresentados no Quadro 1 e correspondem as métricas: Erro Médio Absoluto (*Mean Absolute Error* ou *MAE*), Raiz do Erro Quadrático Médio (*Root-Mean-Square Error* ou *RMSE*) e o Viés do estimador (ou *Bias*), para o período entre 1990 e 2019.



Quadro 1. Valores das médias mensais, pela verificação conjunta entre a base de dados observacional e modelada para o estado do Pará entre 1990 e 2019, considerando os parâmetros de Viés do estimador (ou *Bias*), Erro Médio Absoluto (*MAE*) e Raiz do Erro Quadrático Médio (*RMSE*), os dois últimos na escala de milímetros (mm).

<i>Métricas</i>	<i>Jan</i>	<i>Fev</i>	<i>Mar</i>	<i>Abr</i>	<i>Mai</i>	<i>Jun</i>	<i>Jul</i>	<i>Ago</i>	<i>Set</i>	<i>Out</i>	<i>Nov</i>	<i>Dez</i>	<i>Média Anual</i>
<i>Erro Médio Absoluto (MAE) em mm</i>	28,6	32,2	28,8	24,7	27,1	16,7	16,0	13,0	12,9	10,2	11,8	27,3	20,8
<i>Raiz do Erro Quadrático Médio (RMSE) em mm</i>	34,2	40,3	38,4	32,5	34,7	21,7	21,0	16,5	17,8	13,3	15,3	33,9	26,6
<i>Viés (ou Bias)</i>	17,1	21,4	7,2	5,6	8,3	1,7	4,0	3,7	1,1	4,6	7,2	19,4	8,5

Fonte: elaborado pelos autores.

Revelados, estes dados do Quadro 1 indicam que as métricas estatísticas de erro concentram maiores valores durante o período chuvoso, principalmente, entre os meses de janeiro e maio, com redução no semestre de menor precipitação, a partir de junho e mantendo-se assim até meados de novembro, com a transição entre estes períodos, e durante o mês de dezembro, havendo, novamente, um aumento considerável dos parâmetros de erro. Essa dinâmica revelada nas métricas *MAE* e *RMSE* é a mesma observada nos trabalhos de Pereira et al. (2013) e Costa et al. (2019), indicando certa normalidade para a dinâmica entre os dados estimados e os observados durante os períodos de maior, menor e transição da precipitação.

O coeficiente de determinação (r^2) medido da relação entre os postos e estações com os dados *CHIRPS v.2.0*, tem um valor médio anual de 97,8%, com mínimo anual de 90,1% e máxima anual de 99,6%, indicando excelente proximidade entre as medidas do modelo com aquelas provenientes das matrizes de observação. Em complemento a estes valores, ressalta-se, novamente, o conjunto de parâmetros da Quadro 1 que, pelas métricas *MAE* e *RMSE*, ressalta-se as máximas de 32,2 mm e 40,3 mm, respectivamente, atingidas para ambos os parâmetros durante o mês de fevereiro, bem como, médias anuais de 20,8 mm para o *MAE* e 26,6 mm para o *RMSE*.

Considerando o universo dos dados observado no Quadro 2, tem-se, no geral, a média de valores precipitados para o estado do Pará superior a 2.000 mm/ano, indicando que as métricas *MAE* e *RMSE* apresentam-se aceitáveis, principalmente quando comparados as altas taxas do coeficiente de determinação para a associação entre os conjuntos de dados observados e estimados, indicando forte relação. Esses parâmetros mostram-se ainda mais promissores, quando observados junto com o Viés do estimador, que sob valores positivos e média anual de 8,5, indicam baixa tendenciosidade e baixa superestimação dos dados *CHIRPS v.2.0* quando comparados as informações coletadas junto a ANA ao INMET.

Quadro 2 - Precipitação média anual no estado do Pará para o período 1990-2019, considerando dados observado da ANA+INMET e dados estimados provenientes do CHIRPS v.2.0.

<i>Meses</i>	<i>Prec. Média ANA+INMET (mm)</i>	<i>Prec. Média CHIRPS (mm)</i>	<i>Superestimação CHIRPS (mm)</i>
<i>Jan</i>	272,9	292,6	19,7
<i>Fev</i>	333,6	356,2	22,6
<i>Mar</i>	410,8	417,0	6,2
<i>Abr</i>	371,7	378,5	6,8
<i>Mai</i>	277,3	285,3	8,1
<i>Jun</i>	152,7	155,5	2,8
<i>Jul</i>	118,6	121,8	3,2
<i>Ago</i>	68,9	71,4	2,6
<i>Set</i>	49,7	49,8	0,1
<i>Out</i>	55,7	59,7	4,1
<i>Nov</i>	71,9	79,9	8,0
<i>Dez</i>	141,3	158,4	17,1
<i>Total</i>	2324,9	2426,1	101,2

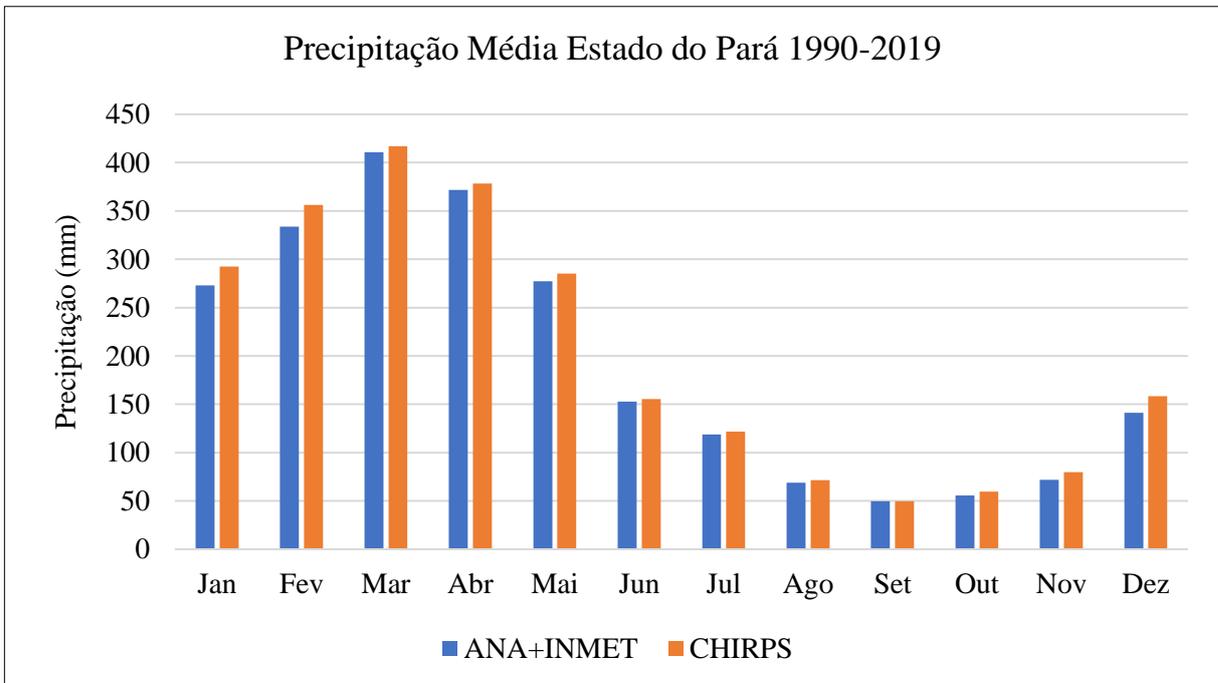
Fonte: elaborado pelos autores.

Destaca-se, também, a partir da Figura 3, a comparação entre as médias mensais de precipitação para a série estabelecida de 1990 a 2019, distribuídas no decorrer do ano, em uma relação que comporta os dados observados da ANA e do INMET, com aqueles modulados a



a partir do *CHIRPS v.2.0*, procurando desta associação, indicar a quantidade média de chuva superestimada ao estado do Pará.

Figura 3 - Distribuição da Precipitação média anual no estado do Pará para o período 1990-2019, considerando dados observado da ANA+INMET e dados estimados provenientes do *CHIRPS v.2.0*.



Fonte: elaborado pelos autores.

Pelo Quadro 2 e a Figura 3, nota-se que os dados *CHIRPS v.2.0* acompanham a distribuição anual das chuvas observadas nos postos e estações da ANA e do INMET. As quantidades conferem boa aproximação entre os valores estimados e a observação, representando anualmente, um acréscimo de 101,2 mm, consistindo em 4% de todo o valor precipitado, dado que as médias para o *CHIRPS v.2.0* referem-se ao valor de 2426,1 mm e para a composição da ANA e do INMET tem-se 2324,9 mm de média, no período 1990-2019. O mês com maior superestimação é fevereiro, com 22,6 mm, indicando uma pequena discrepância, como já apresentado, também, nas métricas *MAE* e *RMSE*, e no Viés do estimador.

Assim, através deste conjunto de parâmetros, em acordo com as métricas de erro e Viés, o coeficiente de determinação e os valores médios, destaca-se uma representação adequada dos dados *CHIRPS v.2.0* em relação aqueles valores observados nos postos e estações da ANA e do INMET para o estado do Pará, durante o período de 1990 a 2019, permitindo, a partir disso, a utilização dos dados do *CHIRPS v.2.0* em estudos climáticos com informações pluviométricas.

A validação de dados pluviométricos realizada neste estudo, alavanca ao estado do Pará uma nova proposta para o manuseio de uma base de dados que permite o desenvolvimento de estudos em um nível de detalhe anteriormente inviável, visto que as condições atuais da malha de dados observacionais pertinentes ao estado são defasadas.

Conceber o uso de dados provenientes da base *CHIRPS v2.0* reitera o papel de destaque aos avanços tecnológicos para suprir as necessidades comumente associadas aos trabalhos climáticos que requerem um alcance maior ao nível de detalhe, fato apresentado no contexto de integrações provenientes da abordagem geográfica do clima. Neste entendimento, a Climatologia Geográfica, fundada pela noção de ritmo, carece de análises constituídas sob uma robusta base de dados que, sob o âmbito pluviométrico, torna-se possível ao estado do Pará, após a realização deste trabalho, considerando a validação dos dados *CHIRPS v.2.0*.

REFERÊNCIAS

ANA. **Portal HidroWeb**: séries históricas. Disponível em:

<http://www.snirh.gov.br/hidroweb/serieshistoricas>. Acesso em: 4 jul. 2020.

BRUCE, P.; BRUCE, A. **Estatística prática para cientistas de dados**: 50 conceitos essenciais. 1ª edição. Rio de Janeiro — RJ: Alta Books, 2019.

COSTA, J. C. et al. VALIDAÇÃO DOS DADOS DE PRECIPITAÇÃO ESTIMADOS PELO CHIRPS PARA O BRASIL. **Revista Brasileira de Climatologia**, [S.l.], v. 24, jun. 2019. ISSN 2237-8642. Disponível em: <<https://revistas.ufpr.br/revistaabclima/article/view/60237>>. Acesso em: 10 abr. 2023.

FUNK, C. et al. The climate hazards infrared precipitation with stations - a new environmental record for monitoring extremes. **Sci Data** 2, 150066, 2015. Disponível em: <https://www.nature.com/articles/sdata201566>. Acesso em: 2 jul. 2020.

INMET. **Banco de Dados Meteorológicos do INMET**. Disponível em: <https://bdmep.inmet.gov.br/>. Acesso em: 2 jul 2020.

MIGUEL, B. H.; RENNÓ, C. D. Validação de estimativas de precipitação por radar meteorológico em uma bacia hidrológica na região central do estado de São Paulo, Brasil. **Anuário do Instituto de Geociências**, v. 43, n. 2, p. 325-339, 2020. DOI: <10.11137/2020_2_325_339>. Disponível em: <[doi:10.11137/2020_2_325_339](https://doi.org/10.11137/2020_2_325_339)>.

PAREDES-TREJO, F. J.; BARBOSA, H. A.; LAKSHMI KUMAR, T. V. Validating CHIRPS-based satellite precipitation estimates in Northeast Brazil. **Journal of Arid**



Environments, vol. 139, pp. 26–40, 2017. Disponível em:

<https://www.sciencedirect.com/science/article/abs/pii/S0140195X>. Acesso em: 10 jul. 2020.

PEREIRA, G. et. al. Avaliação dos dados de precipitação estimados pelo satélite TRMM para o Brasil. **Revista Brasileira de Recursos Hídricos**. v. 18, n. 3, p. 139-148, jul./set. 2013.

Disponível em: <<https://www.abrh.org.br/SGCv3/index.php?PUB=1&ID=109&SUMARIO>>. Acesso em: 14 abr. 2023.

SILVA, E. R. M. da et al. Análise do Desempenho da Estimativa de Precipitação do Produto CHIRPS para Sub-Bacia do Rio Apeú, Castanhal-PA. **Revista Brasileira de Geografia Física**, [S.l.], v. 13, n. 3, p. 1094-1105, jun. 2020. ISSN 1984-2295. Disponível em:

<<https://periodicos.ufpe.br/revistas/rbgfe/article/view/243392>>. Acesso em: 10 abr. 2023.

RABIEI, E.; HABERLANDT, U. 2015. Applying bias correction for merging rain gauge and radar data. **Journal of Hydrology**, 522: 544-557. Disponível em:

<<https://www.sciencedirect.com/science/article/abs/pii/S002272>>. Acesso em: 10 abr. 2023.