

XIV SIMPÓSIO DE RECURSOS HÍDRICOS DO NORDESTE

VALIDAÇÃO DE DADOS DO GLOBAL PRECIPITATION MISSION (GPM) COM OBSERVAÇÕES IN SITU NA BACIA DO RIO SÃO FRANCISCO.

Autores: Erik Schmitt Quedi¹ ; Juliana Andrade Campos¹ ; Fernando Mainardi Fan¹

RESUMO – A utilização de dados de precipitação obtidos por meio do sensoriamento remoto fornece informações úteis para atividades econômicas e científicas, principalmente em regiões onde o monitoramento *in situ* é escasso. A validação de produtos de sensoriamento remoto constitui uma importante etapa prévia as aplicações operacionais, permitindo a mensuração e avaliação dos erros e incertezas associados a estes dados. Neste estudo, fez-se uma validação do produto acumulado diário IMERG-GPM *final run*, utilizando dados de 36 estações pluviométricas convencionais do INMET (Instituto Nacional de Meteorologia), na bacia hidrográfica do Rio São Francisco. Os resultados demonstraram que as métricas estatísticas apresentam desempenhos variáveis entre as regiões da bacia estudada, em geral, apresentando menor correlação linear e maiores erros nas regiões com maior precipitação média anual.

ABSTRACT– The precipitation data obtained through remote sensing is valuable for economic and science activities. The improved capability of both spatial and temporal representation is key factor for its use in many applications, especially when *in situ* monitoring is scarce. Through validation of remote sensing products, it is possible to issue, and evaluate the errors and uncertainties associated with the data. In this study the IMERG-GPM final run was validated using *in situ* data from 36 INMET (National Institute of Meteorology) rain gauges in the São Francisco river basin. The results showed that the statistical metrics presented different performances between the regions of the basin, in general, presenting smaller linear correlation and bigger errors in the regions with greater average annual precipitation.

Palavras-Chave – Bacia do Rio São Francisco, Sensoriamento Remoto, IMERG

1) Instituto de Pesquisas Hidráulicas, Universidade Federal do Rio Grande do Sul. Av. Bento Gonçalves, 950. erik.quedi@gmail.com

1. INTRODUÇÃO

A precipitação, em suas diversas formas de ocorrência, constitui-se como uma das principais variáveis do ciclo hidrológico. O conhecimento de sua ocorrência é fundamental para o entendimento dos processos climáticos, bem como seus impactos no ambiente e nas atividades econômicas, tais como agricultura, geração de energia, disponibilidade hídrica e respostas a desastres naturais.

A precipitação tem, como característica, uma grande variabilidade espacial e sazonal. Por este motivo, a utilização de instrumentos remotos, como os satélites, fornece informações úteis, principalmente em áreas com pouca cobertura de dispositivos *in-situ* (Arkin e Ardanuy, 1989; Kidd *et al.*, 2017).

As estimativas de precipitação por meio de instrumentos remotos vêm sendo desenvolvidas desde o início de missões espaciais, na década de 1970, embora maiores avanços científicos venham sendo obtidos nos últimos anos, a partir do aumento de missões espaciais que objetivam o monitoramento ambiental (Rast *et al.*, 2014).

A partir do desenvolvimento de dispositivos e métodos, atualmente existem diversos produtos oriundos de sensoriamento remoto que fornecem estimativas de precipitação, como o *Precipitation Estimation from Remotely Sensed Information using Artificial Neural Networks* (PERSIANN) (Sorooshian *et al.*, 2000); a técnica *MORPHing* (CMORPH) (Joyce *et al.*, 2004); o *TRMM Multi-satellite Precipitation Analysis* (TMPA) (Huffman *et al.*, 2007); o *Global Satellite Mapping of Precipitation* (GSMaP) (Kubota *et al.*, 2007); e o *Multi-Source Weighted-Ensemble Precipitation* (MSWEP) (Beck *et al.*, 2017). Neste estudo serão utilizados os dados do algoritmo *Integrated Multi-satellitE Retrievals for GPM* (IMERG) (Huffman *et al.*, 2018).

O *Global Precipitation Measurement* (GPM) constitui-se como uma missão espacial internacional, que objetiva o monitoramento climático (precipitação e neve) de todo o globo. O primeiro satélite da constelação GPM foi lançado em fevereiro de 2014, conjuntamente pelas agências espaciais americana (NASA) e japonesa (JAXA), demarcando a nova geração de estimativas espaciais de precipitação.

A missão GPM possui a premissa de melhorar a capacidade de estimativa de precipitação por meio de dispositivos sensores (*Dual-frequency Precipitation Radar*, *GPM Microwave Imager*, *GMI*), capazes de estimar precipitação (chuva e neve) em regiões e períodos que outros satélites não são capazes de mensurar (Huffman *et al.*, 2015; Huffman *et al.*, 2018). O algoritmo IMERG-GPM utiliza os dados proveniente da missão GPM, além de mesclar diversas fontes, como estimativas de precipitação de satélites micro-ondas, infravermelho calibrado com micro-ondas (IR), bem como

inclui informação de estações *in-situ*. Os produtos são disponibilizados em formato de grade, com resolução espacial de 0.1° e resolução temporal variável (de acordo com o produto).

Em vista da melhor capacidade de representação espacial e temporal do fenômeno da precipitação, sobretudo em regiões como as grandes bacias hidrográficas que compõem o território brasileiro, com baixo monitoramento *in situ*, as estimativas obtidas de produtos de sensoriamento remoto, como o IMERG-GPM, tornam-se alternativas importantes para aplicações científicas e econômicas.

Todavia, estes dados devem ser validados previamente a utilização operacional, de modo que se obtenha maior assertividade e conhecimento dos erros associados a estes (Bell e Kundu, 2003; Xie *et al.*, 2006; Renzullo *et al.*, 2011; Fleming *et al.*, 2011).

O estudo de Sharifi *et al.* (2016), que avaliou os produtos de precipitação por sensoriamento remoto IMERG-GPM (*Integrated Multi-Satellite Retrievals for GPM*), TMPA-3B42 (*TRMM Multi-Satellite Precipitation Analysis*) e o ERA-INTERIM (*European Centre for Medium-Range Weather Forecasts*), e validou com estações *in-situ* em regiões distintas do Iran. Os resultados apontam para uma maior capacidade qualitativa e quantitativa do produto IMERG-GPM em comparação aos outros analisados. No estudo de Prakash *et al.* (2016) foi realizada uma validação de produtos diários oriundos de sensoriamento remoto (IMERG-GPM, GSMaP e TMPA) em regiões dominadas pelo clima de monções, na Índia. Foi verificado que em regiões de ocorrências menores de precipitação, os produtos de sensoriamento possuem menor eficiência de detecção e estimativa, apresentando piores resultados nas métricas estatísticas aplicadas.

Neste estudo, objetivou-se avaliar os dados de precipitação do produto IMERG-GPM, utilizando como referência observações *in situ* de estações pluviométricas na bacia do Rio São Francisco.

2. METODOLOGIA

2.1. Área de estudo

A bacia do Rio São Francisco abrange os estados brasileiros de Minas Gerais, Bahia, Pernambuco, Alagoas e Sergipe. O principal curso d'água é o Rio São Francisco, e este atravessa regiões com condições naturais distintas. As regiões do Alto e Baixo São Francisco apresentam índices pluviométricos maiores que as regiões intermediárias (Médio e Submédio), conforme indica a Figura 01.

Os principais biomas da região correspondem a Mata Atlântica (região do Alto S.F., com maior densidade populacional), o Cerrado e a Caatinga.

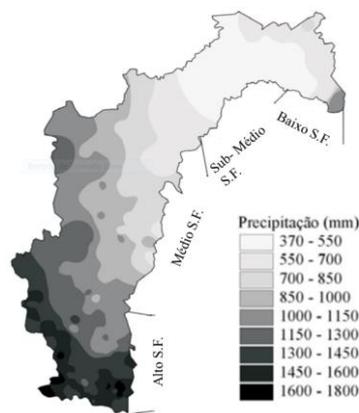


Figura 1 - Mapa de Precipitação - Médias pluviométricas anuais nas regiões da Bacia do Rio São Francisco. Adaptada de Pereira et al. (2007).

2.2. Dados de precipitação

Foram utilizados dados de precipitação oriundos de duas fontes distintas: o produto de precipitações IMERG (Integrated Multi-Satellite Retrievals for GPM), apresentado em Huffman *et al.*, (2018), e dados de precipitações *in situ* de 36 estações pluviométricas disponibilizados pelo Instituto Nacional de Meteorologia (INMET). Utilizou-se dados referentes ao período de 27/01/2014 até 28/02/2017, totalizando uma série de 1129 registros diários em cada ponto avaliado.

Inicialmente realizou-se uma correção do período de acumulo dos dados de precipitação do IMERG-GPM, uma vez que a estação pluviométrica mensura a chuva das 12:00 do dia anterior até 11:59 do dia seguinte (UTM-3), e os dados do produto GPM são acumulados para o período de 00:00 até 23:59 (UTM-3).

O valor da precipitação estimada pelo GPM foi considerado em relação ao centro do pixel da grade do produto GPM: se a estação convencional do INMET se encontra a uma distância menor que 1 km (0.01°) do centro de um pixel, o valor da série estimada pelo GPM nas estatísticas é aquela referente a este pixel; caso não se verifique esta primeira hipótese, é realizada uma interpolação pelo inverso da distância ao quadrado (IDQ) com os quatro pixels adjacentes.

Foram extraídos dados de precipitação do produto IMERG-GPM para as localidades das 36 estações de dados *in situ*. A séries de dados de chuva foram comparadas utilizando as métricas estatísticas descritas a seguir.

2.3. Métricas estatísticas

As estatísticas aplicadas foram baseadas no texto de Jolliffe (2003). Foram realizados cálculos de estatísticas que mensuram a performance e capacidade de detecção da ocorrência do evento de precipitação.

2.3.1. Estatísticas de performance

Como estatísticas de performance, foram utilizados o viés linear (BIAS), erro médio absoluto (MAE), a média do quadrado dos erros (MSE) e o coeficiente de correlação linear.

O Viés Linear é definido como a média da diferença entre as observações *in situ* e as estimadas pelo GPM, o viés pode ser positivo no caso de uma superestimava do GPM e negativo caso contrário. O viés linear não indica a magnitude do erro (valores positivos e negativos se cancelam).

$$BIAS = \frac{1}{n} \sum_{i=1}^n (f_i - o_i) = \bar{f} - \bar{o} \quad (1)$$

O erro médio absoluto indica a magnitude dos erros (medida linear, todos os erros têm o mesmo peso), e por ser um valor absoluto, não indica o sentido do erro (positivo ou negativo).

$$MAE = \frac{1}{n} \sum_{i=1}^n |f_i - o_i| \quad (2)$$

A medida do quadrado dos erros atribui mais peso aos erros maiores e é uma medida sensível a outliers e modelos com resolução grosseiras.

$$MSE = \frac{1}{n} \sum_{i=1}^n (f_i - o_i)^2 \quad (3)$$

O coeficiente de correlação linear indica a relação entre as duas séries de precipitação, isto é, se há algum indicio de tendências similares entre estas séries. O coeficiente de correlação linear pode assumir valores entre -1 e +1, o valor extremo negativo indica uma correlação linear negativa, do contrário uma correlação linear positiva, se não há correlação entre as séries o valor tenderá a zero.

$$CC = \frac{\sum_{i=1}^n (P_{S_i} - \bar{P}_S)(P_{O_i} - \bar{P}_O)}{\sqrt{\sum_{i=1}^n (P_{S_i} - \bar{P}_S)^2} \sqrt{\sum_{i=1}^n (P_{O_i} - \bar{P}_O)^2}} \quad (4)$$

2.3.2. Estatísticas categóricas

Este tipo de estatística constitui uma técnica para avaliação de estimativas em que se utiliza uma tabela de contingência, a qual indica a frequência de ocorrências ou não ocorrências do evento estimado, quando comparado com o observado *in situ*.

Fez-se uma análise da ocorrência entre as séries, e assim classificou-se as análises conforme a Tabela 01. Neste estudo, adotou-se um limiar de precipitação de 0.5 mm na série estimada por GPM, ou seja, valores abaixo destes foram considerados como não ocorrência (igual a zero).

Tabela 1- Tabela de Contingência

Estimativa (GPM)	Observação (INMET)		
	Sim	Não	Total
Sim	a	b	a + b
Não	c	d	c + d
Total	a + c	b + d	a + b + c + d = n

Probabilidade de detecção (PDD): Corresponde a parcela das observações in situ que foram estimadas corretamente pelo satélite. [0,1]

$$PDD = \frac{a}{a+c} \quad (5)$$

Taxa de alarme falso (TAF): Corresponde a fração de estimativas incorretas pelo satélite. [0,1]

$$TAF = \frac{b}{a+b} \quad (6)$$

Acurácia: Mensura a fração de estimativas corretas em relação ao total de dias comparados. [0,1]

$$ACC = \frac{a+d}{n} \quad (7)$$

Proporção de sucesso (PS): Corresponde a razão de eventos de precipitação corretamente estimados. [0,1]

$$PA = \frac{b}{a+b} \quad (8)$$

3. RESULTADOS E DISCUSSÕES

3.1. Estatísticas de performance

A região que apresentou maior magnitude de viés positivo foi a região do médio São Francisco, ou seja, os dados oriundos do GPM estão superestimando as chuvas nesta região, ao passo que a região do submédio São Francisco apresentou menor magnitude de viés, sendo ele negativo (subestimativa das chuvas).

O coeficiente de correlação linear demonstrou que há maior concordância entre as séries estimadas e observadas nas regiões com menores ocorrências de precipitação anual (Médio e

Submédio São Francisco. A Figura 2 ilustra a variabilidade do viés e do coeficiente de correlação linear ao longo da Bacia.

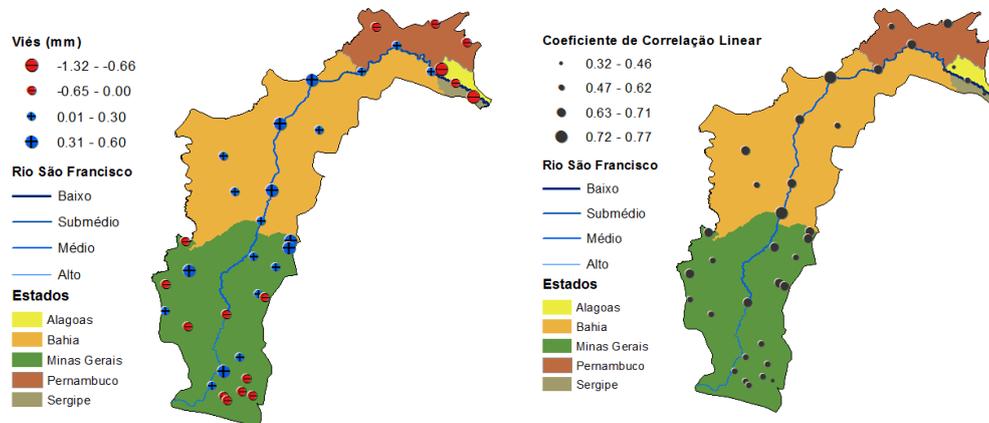


Figura 2 - Representação das métricas Viés e Coeficiente de correlação linear das estimativas do GPM em relação as estações INMET.

Os erros médios absolutos foram maiores nas regiões de cabeceiras (Alto e Médio São Francisco), sendo que estas as regiões apresentam maior volume de precipitação anual. Este comportamento também é observado por Aghakouchak *et al.* (2012), em que os autores também constataram que a magnitude dos eventos de precipitação é proporcional ao erro. A região próxima ao Submédio São Francisco (menor volume de precipitação anual) apresentou menores erros, e, nas proximidades da foz, o erro tende a aumentar.

Por sua vez, o erro médio quadrático evidencia as regiões onde ocorreram os maiores erros, sendo este verificado nas regiões de maiores volumes precipitados anualmente, isto é, àquelas próximas as nascentes, região compreendida pelo Alto e Médio São Francisco, conforme Figura 3.

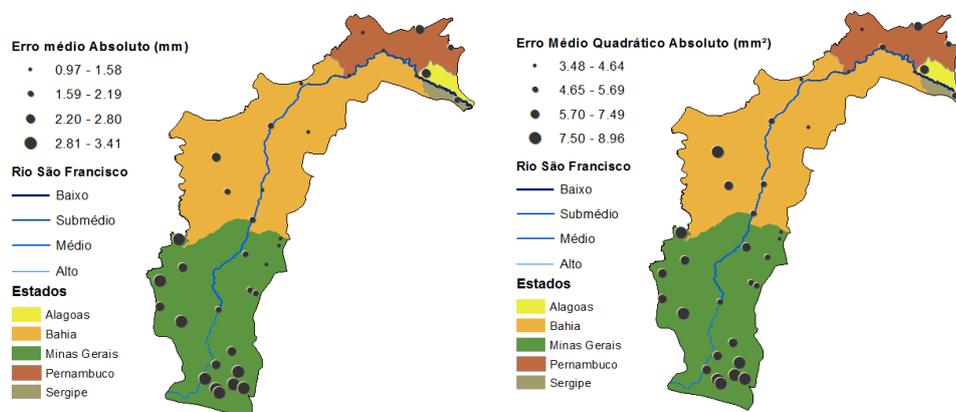


Figura 3 - Erro médio Absoluto e Erro Médio Quadrático das estimativas do GPM em relação as estações INMET.

3.2. Estatísticas categóricas

Os resultados da Probabilidade de Detecção (PDD) mostraram que os dados do GPM foram mais eficientes na detecção de ocorrência de precipitação na região do Alto e Médio São Francisco. Os piores valores de PDD foram valores médios (~50%), o que sugere uma razoável capacidade de estimativa da ocorrência de evento de precipitação em grande parcela da bacia hidrográfica, em especial nas regiões com maior densidade populacional.

A região entre o Submédio e Baixo São Francisco apresentou maiores taxas de alarme falso (Figura 4). Nota-se que estas regiões são aquelas que correspondem a um clima árido e baixa precipitação média anual.

Os dados de precipitação do IMERG-GPM apresentaram maior acurácia nas regiões com índices pluviométricos mais baixos, enquanto que no Alto e Baixo São Francisco, a acurácia se apresentou menor, conforme Figura 5.

As regiões próxima as cabeceiras e a foz da bacia (Alto e Baixo São Francisco, respectivamente) apresentaram menor proporção de estimativas corretas (acurácia), sugerindo que há maior chance de erro na estimativa do satélite devido ao maior número de observações de eventos de chuva. Todavia, a variação desta estatística foi relativamente baixa (~25%), apresentando razoável capacidade de estimativa correta da ocorrência de precipitação.

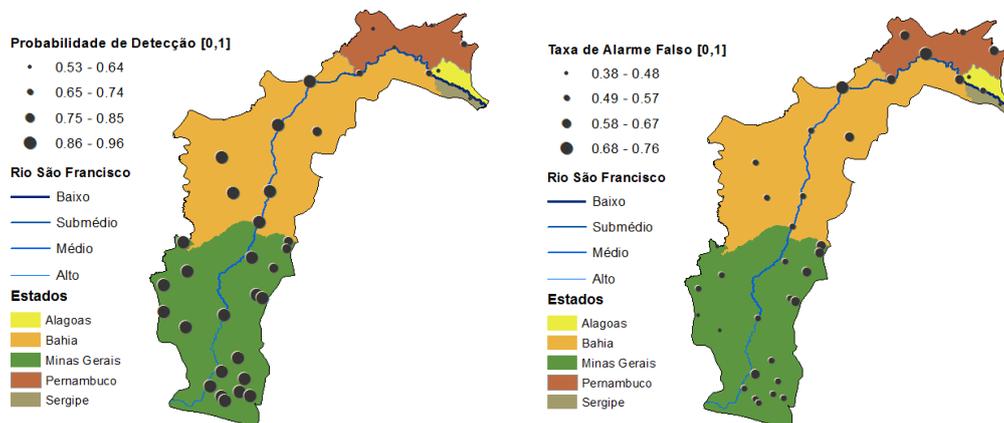


Figura 4 - Probabilidade de detecção e Taxa de alarme falso das estimativas do GPM em relação as estações INMET.

Quanto a proporção de Sucesso, a região do Submédio São Francisco apresentou a pior proporção de estimativa correta de ocorrência de precipitação. Isso pode ser explicado pela baixa ocorrência de precipitação nesta região, sugerindo uma deficiência na estimativa de precipitação pelo GPM nestas condições.

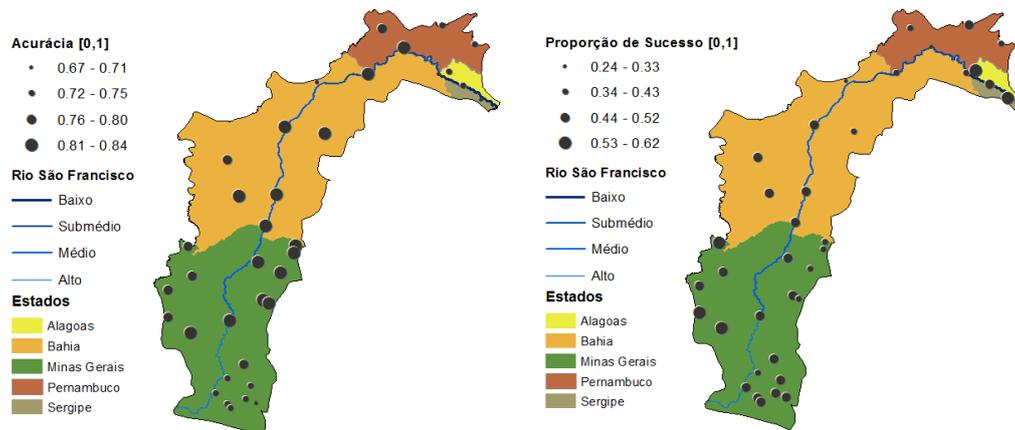


Figura 5 - Acurácia e Proporção de sucesso das estimativas do GPM em relação as estações INMET.

4. CONCLUSÕES

As regiões com maior índice pluviométrico (Alto e Baixo São Francisco) apresentaram os menores coeficientes de correlação linear, bem como a ocorrência de vieses negativos (estimativa do GPM subestimando a precipitação). Nestas regiões também se verificaram os maiores erros médios absolutos e erros médios quadráticos, o que implica em maiores magnitudes de erros.

A estatística de Probabilidade de Detecção mostrou boa capacidade do IMERG-GPM na detecção de ocorrência de precipitação em grande parte da bacia, embora tenha apresentado piores resultados na região com menor precipitação média anual (Submédio São Francisco).

Quanto a Acurácia, ou a estimativa correta de ocorrência ou não ocorrência de precipitação, os resultados foram melhores nas regiões com índices pluviométricos baixos (Submédio e Baixo São Francisco), enquanto que nas cabeceiras (Alto e Médio São Francisco) o resultado foi pior.

Também a Proporção de Sucesso atingiu melhores resultados nas regiões com maiores índices pluviométricos (Alto e Baixo São Francisco).

Sugere-se para estudos posteriores a inclusão de maior volume de dados observados, tais como estações telemétricas e outras bases de dados.

5. REFERÊNCIAS

- AGHAKOUCHAK, A.; MEHRAN, A.; NOROUZI, H.; BEHRANGI, A. (2012). "Systematic and random error components in satellite precipitation data sets". *Geophysical Research Letters*, Vol. 39. 2012.
- ARKIN, P.; ARDANUY, P. E. (1989). *Estimating Climatic-Scale Precipitation from Space: A Review*. *Journal of Climate*, v.2, nov. 1989.
- BECK, H. E.; VAN DIJK, A.I.J.M; LEVIZZANI, V.; SCHELLEKENS, J.; MIRALLES, D. G.; MARTENS, B.; DE ROO, A. (2017). *MSWEP: 3-hourly 0.25° global gridded precipitation (1979–2015) 10 by merging gauge, satellite, and reanalysis data*. *Hydrology And Earth System 11 Sciences*, v. 21, n. 1, p.589-615, 30 jan. 2017.

- BELL, T. L.; KUNDU, P. K. (2003). *Comparing satellite rainfall estimates with rain gauge data: Optimal strategies suggested by a spectral model*. J. Geophys. Res., 108(D3), 4121, doi: 10.1029/2002JD002641, 2003.
- FLEMING, K.; AWANGE, J. L.; KUHN, M.; FEATHERSTONE, W. E. (2011): *Evaluating the TRMM 3B43 monthly precipitation product using gridded raingauge data over Australia*. Australian Meteorological and Oceanographic Journal, 61, 3, pp. 171—184.
- HUFFMAN, G.J.; ADLER, R.F.; BOLVIN, D.T.; GU, G.; NELKIN, E.J.; BOWMAN, K.P.; HONG, Y.; STOCKER, E. F.; WOLFF, D. B. (2007). *The TRMM Multisatellite Precipitation Analysis (TMPA): 19 Quasi-Global, Multiyear, Combined-Sensor Precipitation Estimates at Fine Scales*. 20 Journal Of Hydrometeorology, v. 8, n. 1, p.38-55, fev. 2007.
- HUFFMAN, G.J.; BOLVIN, D.T.; BRAITHWAITE, D.; HSU, K.; JOYCE, R.; XIE, P. (2018). *GPM Integrated Multi-Satellite Retrievals for GPM (IMERG) Algorithm Theoretical Basis Document (ATBD) Version 5.2*. PPS, NASA/GSFC, 35 pp., 2018.
- HUFFMAN, G.; BOLVIN, D.; NELKIN, E. (2015). *Day 1 IMERG final run release notes*. Disponível em: https://pmm.nasa.gov/sites/default/files/document_files/IMERG_FinalRun_Day1_release_notes.pdf. Acesso em: 14 de Dez de 2017.
- JOLLIFFE, I. T.; STEPHENSON, D. B. (2003). *Forecast Verification: A Practitioner's Guide in Atmospheric Science*. ISBN 0-471-49759-2, 247p.
- JOYCE, R.J.; JANOWIAK, J. E.; ARKIN, P. A.; XIE, P. (2004). *CMORPH: A Method that Produces Global Precipitation 29 Estimates from Passive Microwave and Infrared Data at High Spatial and Temporal 30 Resolution*. Journal of Hydrometeorology, v.5, n. 3, p.487-503, jun. 2004.
- KIDD, C.; BECKER, A.; HUFFMAN, G. J.; MULLER, C. L.; JOE, P.; JACKSON, G. S.; KIRSCHBAUM, D. B. (2017). *So, How Much of the Earth's Surface Is Covered by Rain Gauges?* Bulletin of the American Meteorological Society, v. 98, n. 1, p.69-78, 2017.
- KUBOTA, T.; SHIGE, S.; HASHIZUME, H.; AONASHI, K.; TAKAHASHI, N.; SETO, S.; HIROSE, M.; TAKAYABU, Y.N.; USIO, T.; NAKAGAWA, K.; IWANAMI, K.; KACHI, OKAMOTO, K. (2007). *Global Precipitation Map Using Satellite-Borne Microwave Radiometers by the GSMaP Project: Production and Validation*. Geoscience and Remote Sensing, v. 45, n. 7, p.2259-2275, jul. 2007.
- PEREIRA, S. B.; PRUSKI, F. F.; DA SILVA, D. D.; RAMOS, M. M. (2007). *Estudo do comportamento hidrológico do Rio São Francisco e seus principais afluentes*. Rev. Brasileira de engenharia agrícola, Vol 11, n.6. 2007.
- PRAKASH, S.; MITRA, A. K.; AGHAKOUCHAK, A.; LIU, Z.; NOROUZI, H.; PAI, D. S. (2016). *A preliminary assessment of GPM-based multi-satellite precipitation estimates over a monsoon dominated region*. Journal of Hydrology. Vol 556, p.865-876. 2016.
- RAST, M.; JOHANNESSEN, J.; MAUSER, W. (2014). *Review of understanding of Earth's hydrological cycle: Observations, theory and modelling*. Surveys in Geophysics, v. 35, n.3, p. 491-513, 2014.
- RENZULLO, L.; CHAPPELL, A.; RAUPACH, T.; DYCE, P.; LI, M.; SHAO, Q. (2011). *An assessment of blended satellite-gauge precipitation products for Australia*. CSIRO: Water for a Healthy Country National Research Flagship
- SHARIFI, E.; STEINACKER, R.; SAGHAFIAN, B. (2016). *"Assesment of GPM-IMERG and Other Precipitation Products against Gauge Data under Different Topographic and Climatic Conditions in Iran: Preliminary Results"*. Remote Sensing. Vol 8(2), 135. 2016.
- SOROOSHIAN, S.; HSU, K.L.; GAO, X.; GUPTA, H.V.; IMAM, B.; BRAITHWAITE, D. (2000). *Evaluation of PERSIANN System Satellite-Based 30 Estimates of Tropical Rainfall*. Bulletin of the American Meteorological Society, v. 81, n. 9, p.2035-2046, set. 2000.
- XIE, P.; YATAGAI, A.; CHEN, M.; HAYASAKA, T.; FUKUSHIMA, Y.; LIU, C.; YANG, S. (2007). *A gauge-based analysis of daily precipitation over East Asia*. Journal of Hydrometeorology, 8, 607–626.