

UNIVERSIDADE FEDERAL DE SANTA MARIA  
CENTRO DE TECNOLOGIA  
PROGRAMA DE PÓS-GRADUAÇÃO EM ENGENHARIA CIVIL

Renata Celante

**MSWEP COMO FONTE ALTERNATIVA DE DADOS DE  
PRECIPITAÇÃO PARA MODELAGEM HIDROLÓGICA COM O  
SWAT EM BACIAS HIDROGRÁFICAS COM ESCASSEZ DE DADOS**

Santa Maria, RS, Brasil  
2020

**Renata Celante**

**MSWEP COMO FONTE ALTERNATIVA DE DADOS DE  
PRECIPITAÇÃO PARA MODELAGEM HIDROLÓGICA COM O  
SWAT EM BACIAS HIDROGRÁFICAS COM ESCASSEZ DE DADOS**

Dissertação de mestrado apresentada ao curso de Pós-graduação em Engenharia Civil, Área de Concentração em Recursos Hídricos e Saneamento Ambiental, da Universidade Federal de Santa Maria (UFSM, RS), como requisito parcial para obtenção do título de **Mestre em Engenharia Civil**.

Orientadora: Prof<sup>a</sup>. Dr<sup>a</sup>. Rutineia Tassi  
Coorientadora: Dr. Miriam Fernanda Rodrigues

Santa Maria, RS  
2020

Celante, Renata

MSWEP COMO FONTE ALTERNATIVA DE DADOS DE PRECIPITAÇÃO  
PARA MODELAGEM HIDROLÓGICA COM O SWAT EM BACIAS  
HIDROGRÁFICAS COM ESCASSEZ DE DADOS / Renata Celante.-  
2020.

82 p.; 30 cm

Orientadora: Rutineia Tassi

Coorientadora: Miriam Fernanda Rodrigues

Dissertação (mestrado) - Universidade Federal de Santa  
Maria, Centro de Tecnologia, Programa de Pós-Graduação em  
Engenharia Civil, RS, 2020

1. Espacialização da precipitação 2. Preenchimento de  
falhas 3. Produto de satélite I. Tassi, Rutineia II.  
Rodrigues, Miriam Fernanda III. Título.

Sistema de geração automática de ficha catalográfica da UFSM. Dados fornecidos pelo autor(a). Sob supervisão da Direção da Divisão de Processos Técnicos da Biblioteca Central. Bibliotecária responsável Paula Schoenfeldt Patta CRB 10/1728.

© 2020

Todos os direitos autorais reservados a Renata Celante. A reprodução de partes ou do todo deste trabalho só poderá ser feita mediante a citação da fonte

E-mail: renata.celante@gmail.com

**Renata Celante**

**MSWEP COMO FONTE ALTERNATIVA DE DADOS DE  
PRECIPITAÇÃO PARA MODELAGEM HIDROLÓGICA COM O  
SWAT EM BACIAS HIDROGRÁFICAS COM ESCASSEZ DE DADOS**

Dissertação de mestrado apresentada ao curso de Pós-graduação em Engenharia Civil, Área de Concentração em Recursos Hídricos e Saneamento Ambiental, da Universidade Federal de Santa Maria (UFSM, RS), como requisito parcial para obtenção do título de **Mestre em Engenharia Civil**.

**Aprovado em 28 de Maio, 2020:**

---

**Rutineia Tassi, Dra. (UFSM)**

(Presidente/Orientadora)

---

**Miriam Fernanda Rodrigues, Dr. (UFSM)**

(Coorientador)

---

**Raviel Eurico Basso, Dr. (UFG) - Videoconferência**

---

**Vanessa Sari, Dra. (UFSM)- Videoconferência**

**Santa Maria, RS**

**2020**

## AGRADECIMENTOS

A realização dessa pesquisa só foi possível devido ao auxílio, compreensão e dedicação de várias pessoas, dessa maneira, gostaria de agradecer a todos que de certa forma contribuíram para a conclusão desse estudo, em especial:

À minha orientadora Professora Rutineia Tassi que desde a graduação me acompanhou, me auxiliou e me inspirou nessa jornada acadêmica. Obrigada pela confiança depositada em mim, e por sempre acreditar no meu trabalho. Você é incrível!

À minha coorientadora Professora Miriam Fernanda Rodrigues pela paciência e por todo auxílio prestado durante o desenvolvimento do estudo, suas considerações foram muito importantes no processo.

A todos os membros do grupo de pesquisa em Modelagem Hidroambiental e Ecotecnologias e principalmente a todos os colegas da salinha 1109B, vida longa ao café forte!

Ao Professor Raviel Eurico Basso que contribuiu imensuravelmente para o desenvolvimento deste trabalho, me auxiliando em cada passo e me encorajando a buscar sempre as melhores soluções.

Aos amigos queridos que fiz em Santa Maria que sempre me apoiaram em absolutamente tudo, em especial: Rafaela, Robson, Stephen, Ingrid, Jota, Brunna, Deborah e Juliana. Amo vocês, contem comigo sempre!

Aos meus pais, Renato e Rosmari, e ao meu irmão Dian, que nunca mediram esforços para garantir que não me faltasse nada. Muito obrigada por todo o suporte, auxílio e incentivo.

Ao PPGEC e á Lu.

Ao CNPQ, pelo auxílio financeiro concedido.

A todas as pessoas amigas não nomeadas, mas que de alguma forma auxiliaram nessa jornada, o meu muito obrigada.

## RESUMO

# MSWEP COMO FONTE ALTERNATIVA DE DADOS DE PRECIPITAÇÃO PARA MODELAGEM HIDROLÓGICA COM O SWAT EM BACIAS HIDROGRÁFICAS COM ESCASSEZ DE DADOS

AUTOR: Renata Celante  
ORIENTADORA: Rutineia Tassi  
COORDINADOR: Miriam Fernanda Rodrigues

Sendo a precipitação um dos elementos mais importantes no ciclo hidrológico, a obtenção de dados que representem com acurácia as variabilidades espaciais e temporais das precipitações em uma determinada área se torna, também, uma etapa relevante. Entretanto, em países subdesenvolvidos é comum encontrar uma rede pluviométrica com baixa densidade, com muitas falhas, com séries curtas e com qualidade questionável, o que torna a utilização desses dados um fator limitante para muitos estudos. Dessa forma, é necessário em muitos casos recorrer a fontes alternativas aos dados de precipitação medidos em solo, como aqueles oriundos de satélites. Nesse sentido, o produto Multi-Source Weighted-Ensemble Precipitation (MSWEP v.2) vem se destacando, em razão de sua cobertura global, resolução espacial e temporal refinada, demonstrando bons resultados também na modelagem hidrológica. Nesse trabalho, dados de precipitação do MSWEP v.2 foram utilizados como fonte alternativa de dados de precipitação na bacia do Rio Guaporé-RS, uma bacia caracterizada por uma rede de monitoramento pluviométrico bastante deficiente. Foi avaliada a qualidade da precipitação estimada pelo MSWEP v.2, em comparação com os dados de precipitação medidos em solo, e alternativas para aumentar a representatividade da séries de precipitação na bacia hidrográfica, como a extensão das séries, a redução do número de falhas, a ampliação da densidade de postos, e melhorias no processo de espacialização da precipitação. Para isso, cenários alternativos que incorporam informação do produto MSWEP v.2, ou utilizam totalmente a informação desse produto, foram comparados com um cenário de referência que utilizou a rede pluviométrica existente. As alternativas avaliadas mostraram-se como boa estratégia, uma vez que os problemas encontrados na rede pluviométrica existente foram efetivamente mitigados, obtendo-se séries longas, sem falhas e com alta densidade espacial. Após comprovada a viabilidade de utilização dos dados do MSWEP, as estratégias de montagem dos cenários alternativos de precipitação foram utilizadas como base de dados de entrada no modelo SWAT (Soil & Water Assessment Tool) e aplicados na bacia do Rio Guaporé-RS, para avaliar como essas informações melhoram a qualidade da modelagem hidrológica. Verificou-se que com dados alternativos de precipitação baseados no MSWEP foram obtidos resultados de modelagem muito similares aqueles resultantes das precipitações com medições em solo, no entanto, com destacada melhoria na simulação de vazões mínimas. A partir dos estudos realizados, constatou-se que o produto de precipitação do MSWEP v.2 pôde ser utilizado como alternativa para substituir, ou complementar, a rede pluviométrica da bacia hidrográfica com condições de baixa qualidade ou inexistência, pois permitiu aumentar a densidade e a qualidade da rede pluviométrica, diminuindo as incertezas na modelagem hidrológica.

**Palavras-chave:** Espacialização da precipitação. Preenchimento de falhas. Produto de satélite.

## ABSTRACT

### MSWEP AS AN ALTERNATIVE SOURCE FOR RAINFALL DATA FOR HYDROLOGICAL MODELING WITH SWAT IN DATA-SCARCE WATERSHEDS

AUTHOR: Renata Celante

ADVISOR: Rutineia Tassi

CO-ADVIDOR: Miriam Fernanda Rodrigues

As precipitation is one of the most important elements in the hydrological cycle, obtaining data that accurately represents the spatial and temporal variability of precipitation in a given area, becomes a relevant step. However, in developing countries it is common to find a rainfall network with low density, many gaps, short time series and with questionable quality, which makes the use of these data a limiting factor for many studies. Thus, it is necessary in many cases to use alternative sources for rainfall data measurement, such as those from satellites. In this sense, the Multi-Source Weighted-Esemble Precipitation product (MSWEP v.2) has stood out, due to its global coverage, refined spatial and temporal resolution, showing good results also in hydrological modeling. In this work, MSWEP v.2 precipitation product was used as an alternative source of precipitation data in the Guaporé-RS River watershed, a watershed characterized by a scarce rainfall network. The quality of precipitation estimated by MSWEP v.2, was evaluated in comparison with the ground-based rainfall data, and with alternatives to increase the representativeness of the precipitation series in the watershed, such as the extension of the series, reduction of the number of gaps, increase in rainfall network density, and improvements in the precipitation spatialization process. For this, alternative scenarios that incorporate measures from the MSWEP v.2 product, or use the product measured entirely, were compared with a reference scenario that used the available ground-based rainfall network. The evaluated alternatives proved to be a good strategy, since the problems found in the existing rainfall network were effectively mitigated, obtaining long series, without gaps and with high spatial density. After the viability of using MSWEP data was proven, the strategies for assembling alternative precipitation scenarios were used as an input database in the SWAT model (Soil & Water Assessment Tool) and applied in the Guaporé-RS River watershed, to assess how this measurement improves the quality of hydrological modeling. It was found that with alternative precipitation data based on MSWEP, modeling results obtained were very similar to those resulting from ground-based rainfall, however, with an improvement in the simulation low flow periods. From the studies carried out, it was found that the precipitation product of MSWEP v.2 could be used as an alternative to replace, or complement, the rainfall network of the watershed with conditions of low quality or nonexistence, as it allowed to increase the density and the quality of the ground-based rainfall network, reducing the uncertainties in hydrological modeling.

**Keywords:** Precipitation spatialization. Infilling gaps. Satellite product.

## LISTA DE FIGURAS

Figure 1 - Guaporé Watershed location and spatial distribution of ANA rain gauges and MSWEP v.2 grid cells. ANA rain gauges without rainfall recordings are represented by letter “X”, while the triangle symbol represents those station which provides rainfall data. ....	21
Figure 2- Incremental areas for daily statistical indicators agreement evaluation between ANA and MSWEP dataset.....	25
Figure 3- Daily statistical indicators ( $\rho$ , $r^2$ , RMSE and PBIAS) results for each ground-based ANA rain gauge against MSWEP v.2, and its correspondent spatial location. ....	27
Figure 4- Monthly mean rainfall and $\rho$ . ....	29
Figure 5- Example of rainfall spatial subtraction between scenarios for a random day (10/30/1983).Figure (a) for SC1, (b) for SC2, and (c) for SC3 spatialization, and the correspondent modular spatial difference between SC1 and SC2 (d) and SC1 and SC3 (e). ....	32
Figure 6- Temporal irregularity in the number of rain gauges used for interpolation in SC1..	33
Figure 7- Long-term (1979-2016) mean spatial rainfall and error (MEA). Figure (a),(b) and(c) shows the long-term mean areal rainfall for SC1, SC2 and SC3, respectively, followed by the spatial mean absolute error between the reference scenario SC1 and SC2 (d), and SC3(e). ....	34
Figure 8- Location and elevation of Guaporé watershed along with hydrographic network. ..	45
Figure 9- Location of hydrological, meteorological and precipitation stations, MSWEP v.2 grid and sub-basin division. ANA rain gauges without recordings data are represented by the letter “X”, while the filled circle symbol represents stations with data. ....	48
Figure 10- Guaporé Watershed distribution maps for: (a) Land use/Land cover (LULC) and (b) soil type. ....	49
Figure 11- Temporal analysis of ANA's rain gauge daily precipitation on the timeframe of this study and its code. ....	50
Figure 12- Spatial distribution of the rain gauges for scenario: a) RG only ANA gauges records, with no infilling gaps, with a total of 7 rain gauges used. b) MSWEP dataset accounting with 38 grid cells, and c) RG_MSWE, all ANA rain gauges from Guaporé watershed and its surroundings were used, along with MSWEP dataset to infill the gaps, with a total of 29 merged rain gauges. ....	52
Figure 13- Guapore’s streamflow time series (station code: 86580000).....	53
Figure 14- Guaporé's rating curve .....	54
Figure 15- Flowchart of how the methodology of this study was conducted, with model sensitivity analysis and calibration routines.....	55
Figure 16- Daily time series of the error components for Scenario RG and Scenario MSWEP. ....	59
Figure 17- Daily time series of error components for RG and RG_MSWE. ....	60
Figure 18- Daily precipitation frequency distribution in different ranges (0~0.1, 0.1~1, 1~5, 5~10, 10~20 and >20 mm/day) for MAP after IDW. ....	61
Figure 19- Simulated (Sim.) versus observed daily streamflow for $Q_{tRG}$ , $Q_{tMSWEP}$ and $Q_{tRG\_MSWEP}$ at gauging station. Pearson correlation ( $r$ ), interception (a) and regression gradient (b). ....	63
Figure 20- Observed and simulated daily streamflow for calibration period of $Q_t$ condition and two zoomed areas. ....	65
Figure 21- Normalized parameters with regards to the final minimum and maximum range and fitted value for the best simulation of each precipitation scenario in $Q_t$ condition. ....	66



Figure 22- Comparison of simulated daily streamflow for RG, MSWEP and RG_MSWEF and daily observed streamflow at gauging station for Q<500 approach. Pearson coefficient (r), intercept (a) and gradient (b).....	68
Figure 23- Observed and simulated daily flow hydrographs for calibration period of Q<500	70
Figure 24- Normalized parameters for Q<500. ....	71

## LISTA DE TABELAS

Table 1 - Scenarios description of infilling rainfall data. ....	24
Table 2- Rain gauge station code along with its snapping period, number of missing daily data, its percentage relative to the snapping period, and MSWEP daily validation metric results against ANA ground-based data. ....	26
Table 3- MSWEP monthly validation results for Pearson correlation coefficient ( $\rho$ ), coefficient of determination ( $r^2$ ), root mean square error (RMSE) and percentage bias (PBIAS). ....	28
Table 4 – Statistical indicators ( $\rho$ , $r^2$ , RMSE and PBIAS) results for daily average rainfall according to the incremental watershed area. ....	31
Table 5- Final set of parameters selected to perform calibration for $Q_t$ and $Q < 500$ , where sensitive parameters in each approach ( $Q_t$ or $Q < 500$ ) are highlighted with an “X”. ....	56
Table 6- Streamflow model calibration and validation statistical indicators performance for $Q_t$ ....	62
Table 7- Model uncertainties under $Q_t$ calibration and validation ....	67
Table 8- Metrics result between observed streamflow and precipitation scenarios for $Q < 500$ ....	68
Table 9- Model uncertainty prediction in simulated streamflow calibration and validation for $Q < 500$ . ....	72

## SUMÁRIO

<b>1</b>	<b>INTRODUÇÃO .....</b>	<b>12</b>
1.1	OBJETIVOS .....	15
1.1.1	Objetivo geral.....	15
1.1.2	Objetivos específicos.....	15
<b>2</b>	<b>CAPÍTULO I- USING MSWEP AS A DIRECT METHOD FOR FILLING MISSING RAINFALL DATA IN RAIN GAUGE SCARCITY WATERSHED .....</b>	<b>17</b>
2.1	INTRODUCTION .....	17
2.2	MATERIALS AND METHODS.....	20
2.2.1	Study area.....	20
2.2.2	Data.....	20
2.2.2.1	<i>Rainfall gauge data .....</i>	<i>20</i>
2.2.2.2	<i>MSWEP V2 database.....</i>	<i>22</i>
2.2.3	MSWEP V2 dataset quality control.....	22
2.2.4	Schemes for infilling data.....	23
2.3	RESULTS ANS DISCUSSION.....	25
2.3.1	MSWEP dataset validation.....	25
2.3.2	Infilling missing data and spatialization.....	30
2.3.2.1	<i>Statistical indicators result after infilling and spatialization.....</i>	<i>30</i>
2.3.2.2	<i>Visual spatial analyze.....</i>	<i>31</i>
2.4	CONCLUSION .....	35
2.5	REFERENCES .....	36
<b>3</b>	<b>CAPÍTULO II- ASSESSMENT OF MSWEP V.2 PRECIPITATION PRODUCT IN A DATA-SCARCE WATERSHED AND ITS UNCERTAINTY IN STREAMFLOW SIMULATION ON SWAT .....</b>	<b>41</b>
3.1	INTRODUCTION .....	41
3.2	MATERIALS AND METHODOLOGY .....	44
3.2.1	Study area.....	44
3.2.2	SWAT model.....	45
3.2.2.1	<i>Model setup and Model inputs.....</i>	<i>46</i>
3.2.2.2	<i>Precipitation datasets.....</i>	<i>49</i>
3.2.2.3	<i>Mean areal precipitation error.....</i>	<i>52</i>
3.2.2.4	<i>Streamflow data.....</i>	<i>53</i>
3.2.2.5	<i>Sensitivity analysis, model calibration and validation.....</i>	<i>55</i>
3.3	RESULTS AND DISCUSSIONS .....	58
3.3.1	Mean areal precipitation error.....	58
3.3.2	Model calibration and validation .....	61
3.3.2.1	<i>Model performance under <math>Q_t</math> condition .....</i>	<i>62</i>
3.3.2.2	<i>Model performance under <math>Q &lt; 500</math> condition.....</i>	<i>67</i>
3.4	CONCLUSIONS.....	72
3.5	REFERENCES .....	74
<b>4</b>	<b>CONCLUSÕES GERAIS .....</b>	<b>78</b>
<b>5</b>	<b>REFERÊNCIAS GERAIS .....</b>	<b>80</b>

## 1 INTRODUÇÃO

Por se tratar de um sistema fechado com entradas e saídas definidas, a bacia hidrográfica é o espaço ideal para avaliar o comportamento hídrico (TUCCI, 2012) e, justamente por isso, comumente compõem os elementos representados em modelos hidrológicos. De fato, o estudo das bacias hidrográficas através dos modelos hidrológicos possibilita a análise de mudanças climáticas, manejo dos recursos hídricos, inundações, e impactos causados pela mudança no uso e manejo do solo (ARNOLD *et al.*, 1998).

Entre os modelos hidrológicos atualmente disponíveis para tais análises estão os do tipo precipitação-vazão (P-V), que surgiram para suprir a necessidade de se obter séries hidrológicas mais longas e representativas da vazão, visto que geralmente as séries de precipitações disponíveis são mais longas, se comparadas com as séries de vazão isso, especialmente em decorrência da dificuldade de monitorar as variáveis hidrológicas (BEVEN, 2012; TUCCI, 2005). Soma-se a isso, o fato de que muitas séries de vazão não são homogêneas ou estacionárias no tempo, em consequência das modificações de uso do solo, e das variáveis climáticas, que podem ocorrer nas bacias hidrográficas ao longo do tempo (BRESSIANI *et al.*, 2015a; STRAUCH *et al.*, 2012; TUCCI, 2005). Essas características conferem aos modelos P-V a possibilidade de gerar séries de vazão, preencher falhas porventura existentes em uma série de vazão (BEVEN, 2012; MAIDMENT, 1993), avaliar cenários que incluam mudanças no uso do solo, mudanças climáticas, e práticas de conservação que influenciam os recursos hídricos (ABBASPOUR *et al.*, 2015; MORIASI *et al.*, 2015), e são especialmente importantes para lugares onde a rede de monitoramento hidrológico é escassa, falha ou inexistente (HUGHES, 2006).

No entanto, há que se considerar que a precisão da resposta do modelo P-V escolhido é diretamente influenciada pela qualidade dos dados de entrada utilizados, pela própria estrutura interna do modelo e pela sua capacidade de representação dos processos hidrológicos. Portanto, modelos mais complexos, como os de base física, necessitam de dados de entrada igualmente mais complexos (DEVRIES; HROMADKA, 1993). Dentre este tipo de modelo hidrológico com base física, está o Soil and Water Assessment Tool (SWAT), que tem se consolidado mundialmente (ABBASPOUR *et al.*, 2015; ARNOLD *et al.*, 2012; GASSMAN *et al.*, 2007) e no Brasil, com aplicações em estudos para diferentes finalidades (BRESSIANI *et al.*, 2015b). O SWAT possui rotinas e algoritmos que são capazes de simular o crescimento e a rotação de culturas, os impactos de práticas de uso e do manejo do solo na água, na produção de sedimento e na descarga e transporte de pesticidas provenientes da agricultura em bacias

hidrográficas (ARNOLD *et al.*, 2012). Em contrapartida à sua ampla aplicabilidade, o SWAT exige uma grande base de dados a serem fornecidos e/ou calibrados, para uma adequada representatividade dos processos simulados.

É preciso destacar ainda, que as condições de entrada de dados do modelo SWAT são condizentes com a realidade dos Estados Unidos, país onde o modelo foi desenvolvido e que, mesmo com grandes dimensões, possui uma ampla rede de monitoramento de suas bacias hidrográficas. Além disso, o banco de dados do SWAT possui os parâmetros e configurações originais referentes aos solos, clima, usos da terra, crescimento vegetal, operação de manejo, dentre outros, compilados a partir de estudos realizados nos Estados Unidos e que estão inseridos como padrão na base de dados do modelo.

Assim, gestores dos recursos hídricos e responsáveis pelo planejamento do uso e manejo do solo de uma bacia hidrográfica podem enfrentar certa dificuldade na obtenção dos dados necessários para o uso do SWAT em bacias hidrográficas brasileiras, dada a indisponibilidade e a baixa qualidade dos dados, a dificuldade de acesso a acervos de informações, em especial junto à instituições privadas e, ainda, problemas com o processamento e preparação do banco de dados para posterior utilização no modelo (BRESSIANI *et al.*, 2015b). Adicionalmente, grande parte das informações disponíveis se encontram dispersas em diferentes bases de dados, e em formatos que não são facilmente utilizáveis, tomando tempo e esforço por parte do usuário na sua preparação e conversão para o formato exigido pelo modelo (BRESSIANI *et al.*, 2015b).

Entretanto, independentemente do estudo a ser realizado com o SWAT (vazão, sedimento, parâmetros da qualidade da água, agroquímicos, etc.), o balanço hídrico é a força motora de todos os processos a serem simulados na bacia hidrográfica. Portanto, para quantificar precisamente o transporte de pesticidas, de sedimentos e de nutrientes, é fundamental que a simulação do ciclo hidrológico pelo modelo seja coerente com os processos que ocorrem na bacia (NEITSCH *et al.*, 2011), sendo a precipitação um elemento de extrema importância. A precipitação é o elemento que dá início ao ciclo hidrológico, sendo responsável por todos os outros processos subsequentes; assim, erros na sua estimativa podem produzir erros sérios na resposta do modelo hidrológico (LI *et al.*, 2018; STRAUCH *et al.*, 2012). Embora o Brasil conte com uma rede hidrográfica densa e vasta, com grande variabilidade espacial de características como clima, solo, vegetação e topografia, observa-se que a rede de monitoramento pluviométrica é bastante deficitária, o que pode comprometer a acurácia e precisão do modelo SWAT, uma vez que uma das grandes causas de resultados ruins na modelagem hidrológica do SWAT refere-se a baixa qualidade espacial da rede de monitoramento pluviométrico usada como dado de entrada (ARNOLD *et al.*, 2012).

A Agência Nacional das Águas (ANA), gestora do sistema de monitoramento hidrológico no Brasil conta com 2.722 estações pluviométricas, sendo que destas 60% possuem séries com mais de 30 anos de registros (ANA, 2017). Aparentemente esse indicador refletiria uma boa cobertura dessa rede de monitoramento de precipitação. No entanto, a utilização dos dados de precipitação é muitas vezes comprometida por sua qualidade, e pela má distribuição espacial desses postos. Muitas vezes, as séries de precipitação possuem falhas e descontinuidades, os postos pluviométricos são mal distribuídos espacialmente, e a densidade de estações pluviométricas é baixa, o que dificulta a utilização dos dados de precipitação no processo de modelagem. Além disso, observando a densidade de estações pluviométricas da ANA por km<sup>2</sup>, verifica-se um valor muito abaixo do recomendado pela Organização Mundial de Meteorologia (OMM). Este cenário é uma realidade também em outros países da América do Sul, onde a rede pluviométrica é insuficiente, com má distribuição espacial, e com baixa resolução temporal (SALIO *et al.*, 2015), o que origina dificuldades para a caracterização da precipitação, e mesmo para sua aplicação na modelagem hidrológica. Além disso, a aquisição de dados e de informações hidrológicas em campo pode ser onerosa, pois envolve infraestrutura adequada, mão de obra e tempo (FENSTERSEIFER; ALLASIA; PAZ, 2016); havendo ainda a possibilidade de risco para a equipe envolvida, nas situações de monitoramento de eventos de cheia.

Como forma de viabilizar que um maior número de trabalhos na área da modelagem hidrológica possa ser realizado em bacias hidrográficas com restrições em relação à qualidade e à disponibilidade dos dados de precipitação de postos pluviométricos, uma alternativa é uso de produtos de precipitação oriundos de satélites (STRAUCH *et al.*, 2012; TUO *et al.*, 2016). Esses produtos ser usados como fonte única de dados de precipitação na modelagem (THIEMIG *et al.*, 2012; YUAN *et al.*, 2017), como também, serem usados em conjunto com os dados já disponíveis na rede de monitoramento pluviométrico, em processos de preenchimento direto de falha das precipitações (DINKU *et al.*, 2014; GITHUNGO *et al.*, 2016; HUGHES, 2006; MISHRA, 2013).

Dentre as várias opções de produtos de precipitação disponíveis, o Multi-Source Weighted-Ensemble Precipitation- versão 2- (MSWEP v.2) se destaca por ser o primeiro produto de abrangência global, com resolução espacial de 0.1° e com dados de precipitação em escala temporal de três horas dos anos de 1979 a 2016 (BECK *et al.*, 2018). Alguns estudos já apresentaram resultados sobre uso de produtos de precipitação no Brasil como fonte de entrada do modelo SWAT, principalmente o produto TRMM 3B42 (CREMONINI; BRIGHENTI; BONUMÁ, 2014; STRAUCH *et al.*, 2012; TOBIN; BENNETT, 2014), que possui uma

resolução espacial de 0.25°. Todavia, trabalhos que usufruíram do produto MSWEP v.2 no modelo SWAT, até o momento, só são encontrados em outros países (SIRISENA *et al.*, 2018; TANG *et al.*, 2019); havendo, portanto, a necessidade de se investigar a viabilidade do uso dessas fontes de dados em simulações com o modelo SWAT para a realidade das bacias hidrográficas brasileiras.

Neste contexto, o presente trabalho é fundamentado na premissa de que devido à precária rede de monitoramento hidrológico e pluviométrico das bacias hidrográficas brasileiras, os usuários que desejam utilizar o modelo SWAT encontram inúmeras dificuldades na obtenção de dados representativos da variabilidade espacial da precipitação, e, dessa forma, fontes alternativas de informações podem ser necessárias. Neste sentido, este estudo investigou a qualidade do produto MSWEP v.2 em uma bacia hidrográfica de médio porte no Brasil, e utilizou esse tipo de informação como fonte alternativa aos dados precipitação de pluviômetros, para posterior aplicação no modelo hidrológico SWAT para a simulação dos processos hidrológicos.

## 1.1 OBJETIVOS

### 1.1.1 Objetivo geral

O estudo teve como objetivo avaliar a qualidade do produto de dados de precipitação estimados pelo MSWEP v.2 na bacia hidrográfica do Rio Guaporé-RS, e avaliar a qualidade da simulação hidrológica com o modelo SWAT a partir de cenários alternativos de fontes de precipitação.

### 1.1.2 Objetivos específicos

- Verificar a possibilidade de utilização da precipitação prevista pelo MSWEP v.2 como fonte alternativa aos dados de precipitação de postos pluviométricos, por meio de estratégias para preenchimento de falhas nas séries de dados pluviométricos e análise da precipitação média espacial da bacia hidrográfica.
- Avaliar o desempenho da simulação hidrológica da bacia do Rio Guaporé-RS com o modelo SWAT, a partir de diferentes cenários de dados de precipitação que incorporam informações do MSWEP v.2.

## 1.2 ESTRUTURA DA DISSERTAÇÃO

Esta dissertação foi dividida em quatro itens, que correspondem à Introdução, os Capítulos I e II e as Considerações Finais.

A Introdução, já apresentada, descreve os problemas que motivaram a realização deste trabalho, bem como os objetivos a serem desenvolvidos. Os dois itens seguintes, Capítulos I e II são apresentados na forma de artigo, e abordam as questões levantadas nos objetivos específicos, sendo que cada um destes capítulos corresponde a um objetivo específico, no contexto do objetivo geral. Por fim, o último item apresenta as conclusões finais a respeito dos resultados obtidos nos Capítulos I e II.



## 2 CAPÍTULO I- USING MSWEP AS A DIRECT METHOD FOR FILLING MISSING RAINFALL DATA IN RAIN GAUGE SCARCITY WATERSHED

### ABSTRACT

Long-term time series of rainfall accurate and reliable are essential for climate and hydrological studies. However, uninterrupted and complete rainfall datasets are often not available, especially in developing countries. Short time series and missing data are additional challenges for water resource planning, which compromise results leading to wrong and misleading conclusions. This paper proposes the use of Multi-Source Weighted-Ensemble Precipitation v.2 (MSWEP v.2) data as an alternative source for infilling missing daily rainfall data, or even to fully replace rain gauge records, at Guaporé Watershed (2430 km<sup>2</sup>), a typical mid-size Brazilian watershed characterized by a scarce rain gauge network density, time series interrupted and great number of missing data. The ground-based rainfall network at Guaporé watershed have 29 rain gauges, where only 14 have recorded data in the time frame of this study (1979-2016). MSWEP v.2 dataset was first validated against observed data in a daily and monthly time step, with four statistical indicators ( $\rho$ ,  $r^2$ , PBIAS, and RMSE). Then the methodology results of the infilling procedure were evaluated on the final product of the rainfall spatialization and estimation of the mean rainfall over the watershed with the statistical indicators, and with long-term mean spatial rainfall and error. The results indicated good agreement between ground-based rain gauges and MSWEP v.2 dataset for daily and monthly time step in validation procedure, presenting remarkable statistical indicators results. The use of MSWP v.2 as alternative for infilling gaps presented an important improvement on rainfall spatialization and variability, suggesting that its estimates can be used as a straightforward source for rainfall series in places where the ground-based rain gauges suffer from shortcomings and low network density.

### 2.1 INTRODUCTION

Rainfall is one of the most important parameters that drives the hydrologic cycle (KIDD; HUFFMAN, 2011), hence its data are essential for climate and hydrological studies, including water resource management strategies (HOU *et al.*, 2014). Long-term time series with high temporal and spatial variability are needed to allow reliable and accurate processes description and predictions in a watershed scale (ESLAMIAN, 2014). Even though is expected that the rain

gauge network fulfills these needs, rainfall data with high temporal and spatial variability is especially hard to find in developing countries where rain gauges are unevenly and sparsely distributed. Furthermore, even when the rain gauge network is fairly dense, the gauges may have several missing gaps and interrupted periods of monitoring (HUGHES, 2006; MAIDMENT *et al.*, 2017; VILLAZÓN; WILLEMS, 2010).

The Brazilian National Water Agency (Agência Nacional de Águas-ANA), for instance, had only 2,722 rain gauges managed by ANA in operation in the national hydrological network in 2016 (ANA, 2017), with daily data freely available. This amount of rain gauges implies a national average density of one rain gauge every 3,128 km<sup>2</sup>, while the minimum density recommended by World Meteorological Organization is one rain gauge every 900 km<sup>2</sup> (WMO, 2008), for the less restrictive physiographic unit considered. By comparison, the United States has an average national density of one rain gauge every 384 km<sup>2</sup> (USDC-NOAA, 2013). Besides the low number of rain gauges, the spatial distribution is uneven among the Brazilian states. The North (Amazon biome) and Midwest (Cerrado biome) regions have the lowest rain gauge density, while Southeast (Atlantic Forest biome) has the largest one. This disparity may occur due to accessibility issues and high cost of installation and maintenance (ANA, 2017). This disparity has higher effects on rainfall quality evaluation as the watershed area decrease, once the rain gauge density decreases as well. Furthermore, is common to find long-term time series with interruptions and missing values.

In this light, the estimative of missing rainfall values is the main strategy used to fill and obtain long-term rainfall time-series. Several techniques can be used to reconstruct and fill missing monthly and seasonal rainfall data, as the normal-ratio method, inverse distance method, and arithmetic mean method (CHOW *et al.*, 1988), artificial neural network (MWALE; ADELOYE; RUSTUM, 2012) or even the construction of a new rainfall database derived from satellite-based products (DINKU *et al.*, 2014; GITHUNGO *et al.*, 2016). However, missing daily rainfall data is a complex information to reconstruct due to the high spatial and temporal variability of each single-day event (SIMOLO *et al.*, 2010), which difficult the daily rainfall infilling procedure.

Despite this difficulty, good effort has been made, and a range of different methodologies regarding the infilling of missing daily rainfall data has been developed including artificial neural network (ANN) (COULIBALY; EVORA, 2007); ANN and regression technique (KIM; AHN, 2009); regressive methods (LO PRESTI; BARCA; PASSARELLA, 2010); objective automated procedure preserving both, probability distribution and long-term statistics (SIMOLO *et al.*, 2010); interpolation schemes (WAGNER *et al.*,

2012); probabilistic approach using Poisson-gamma distribution (HASAN; CROKE, 2013); multiple enchained imputation (DE CARVALHO *et al.*, 2017), and imputation algorithm based on the optimization of some regression methods (AIEB *et al.*, 2019). However, these approaches demand time and careful handling, in order to have a properly time series reconstruction.

Satellite-based precipitation products (SPPs) can offer a dataset that spans long-term periods (temporal coverage), with a time resolution ranging from hours to months (temporal resolution), which allows the estimation of spatial variability (spatial resolution) reaching remote areas (coverage). In addition, the SPP can be used as a simple and direct way of infilling missing daily rainfall data gathering in rain gauges (DINKU *et al.*, 2014; GITHUNGO *et al.*, 2016; HUGHES, 2006; MISHRA, 2013) or as the primary source of rainfall data in regions with sparse rain gauge network (THIEMIG *et al.*, 2012; YUAN *et al.*, 2017). Products from different satellites, as Global Precipitation Measurement (GPM), Terrain Rainfall Measurement Mission (TRMM), NOAA CPC Morphing Technique (CMORPH), Multi-Source Weighted-Ensemble Precipitation (MSWEP), have been used in order to fill missing daily rainfall data or as the main rainfall data. Each SPP launched can have improvements over one, two, three, or all four of the desired characteristics for rainfall estimation: temporal coverage, temporal resolution, spatial resolution, and coverage, providing some strong suit and weak points (KIDD, 2001).

The MSWEP is a global rainfall-product that merges satellite, rain-gauge and re-analysis data, and provides a long-term rainfall dataset (1979-2016) with a global 3-hourly 0.1° resolution. MSWEP Version 2 (v.2) had some unique aspects over the aforementioned characteristics comparing with other satellite products available, which includes reliable rainfall estimates over the entire globe, and high temporal and spatial resolution. Due to these characteristics, the best spatial and temporal resolution SPP in South Brazil has been considered those obtained from the MSWEP v.2 (BECK *et al.*, 2018).

Once MSWEP v.2 have been recently released, few papers evaluated its efficiency and accuracy when used in hydrological studies. The consistence of 12 SPP evaluation for hydrological modeling indicates that MSWEP v.2 was the most efficient to represent space-time accuracy and consistency in reproducing rain gauge rainfall estimates in the Lake Titicaca, South America Andean plateau (SATGÉ *et al.*, 2019). The MSWEP v.2 had better performance on hydrological modeling in China, when compared against different satellite products (WU *et al.*, 2018). Furthermore, the MSWEP v.2 had the best overall performance for 11 gauge-

corrected daily rainfall in the United States, comparatively with 10 other satellite products (Beck et al., 2019).

Whilst remarking the potential of MSWEP v.2, none of these manuscripts proposed its usage as a source for infilling missing daily rainfall data or replacing long gaps in the monitoring period. In this way, this study shows different analyses accomplished by using MSWEP v.2 dataset, aiming to improve the quality of the estimative of the mean areal precipitation of a typical mid-size watershed in South Brazil with scarce rain gauge data.

## 2.2 MATERIALS AND METHODS

### 2.2.1 Study area

The study was performed in the Guaporé watershed (2430 km<sup>2</sup>), located in the Rio Grande do Sul State (latitude 28°14'26.9" N, 28°58'02" S and longitude 51°54'59"E, 52°22'55"W ), Southern Brazil. The watershed elevation ranges from 14 m up to 847 m a.s.l., with an average elevation of 550 m a.s.l. (Figure 1). Regional climate is humid subtropical (Cfa), according to the Koppen's classification, with hot and humid summers, and cold to mild winters (ALVARES *et al.*, 2013). The mean temperature during the coldest month is around 12 °C, and 22 °C during the warmest month (WOLLMANN; GALVANI, 2012). The rainfall in this watershed is evenly distributed throughout the year, with average annual rainfall ranging from 1400-2000 mm (TIECHER, 2015).

Land use in this watershed is dominated by agriculture (55%) typically with three harvests per year with crop rotation. The main crops are soybeans, corn, wheat, ryegrass and tobacco. The vegetation comprises 44% of the watershed, in which 31% is covered by mixed forests and 13% by range grasses, and only about 1% by urban area. These percentages were acquired from a supervised classification of a 30m resolution satellite image from October 2014, provided by the United States Geological Survey website (USGC), Landsat 4-5 Thematic mapper. This study was carried out from the timeframe of 1979 to 2016, which is the range of the MSWEP v.2 dataset available.

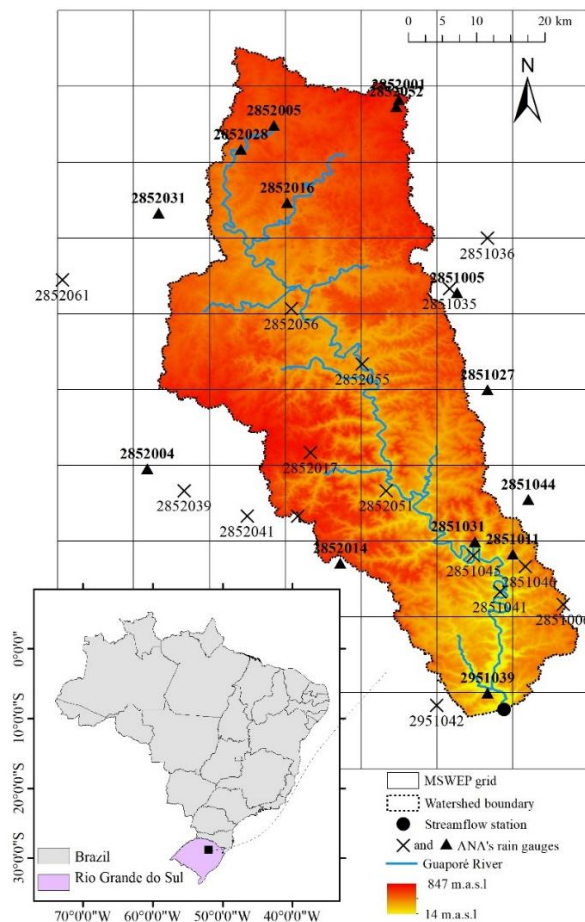
### 2.2.2 Data

#### 2.2.2.1 Rainfall gauge data

Guaporé Watershed has 16 ANA rainfall gauges available within its area, and several surrounding stations. Surrounding rainfall gauges were considered through previous evaluation in the spatialization process by using Thiessen Polygons, in order to ensure a dense and most well spatially distributed rainfall network in the whole watershed. Thus, a total of 29 rain gauges were selected, 16 rain gauges within the watershed boundaries and 13 rain gauges on its surroundings (Figure 1).

At the first sight, the rain gauge network density is suitable for the size of the study watershed. However, 8 rain gauge of the 16 rain gauges within the watershed have no recorded data, 6 have time series with 5 years or shorter, and only 2 remaining stations have longer time series, one with recordings from 2002 to 2016, and the other with the longest time series from 1979 to 2012. A total of 6 from the 13 surrounding watershed rain gauges were suitable for the studied period, with recordings to be analyzed.

Figure 1 - Guaporé Watershed location and spatial distribution of ANA rain gauges and MSWEP v.2 grid cells. ANA rain gauges without rainfall recordings are represented by letter “X”, while the triangle symbol represents those station which provides rainfall data.



Source: the author.

### 2.2.2.2 MSWEP V2 database

The satellite-based dataset used in this study is the Multi-Source Weighted-Ensemble Precipitation Version 2, or MSWEP v.2, released in 2018. MSWEP v.2 provides a gridded dataset for the period 1979-2016, and is the first fully global rainfall dataset with a spatial resolution of  $0.1^\circ$  (BECK *et al.*, 2018). The rainfall available in MSWEP v.2 is estimated by the integrated information including 76747 rain gauges, four satellite, and two reanalysis datasets. MSWEP v.2 has a high temporal resolution of 3 hours, which can be accumulated to daily or monthly record. The MSWEP v.2 has an upgrade, in comparison with MSWEP v.1, that uses daily (rather than monthly) gauge correction scheme accounting for regional differences in reporting times. This procedure allows minimize the timing mismatches when applying daily gauge corrections (BECK *et al.*, 2018). The MSWEP v.2 grid covers Guaporé watershed area with 38 cells (Figure 1), and a daily temporal resolution was used, fitting with the ANA database.

### 2.2.3 MSWEP v.2 dataset quality control

MSWEP v.2 quality control procedure was done against ANA ground-based rain gauge, or also called observed data, for the entire study time frame. Daily rainfall values of observed data (ground-based was considered reference) were evaluated against the observation in the nearest MSWEP grid cell in the same temporal resolution. As ANA dataset has some rain gauges with measurements available only above 1 mm of rainfall, to reduce the chances of mismatches and possible errors in rainfall estimation, a daily 1 mm cumulative rainfall threshold was taken as a pattern for considering rainfall occurrence, while values below this threshold were replaced by 0 in the whole ANA dataset. The MSWEP v.2 dataset was also modified, and all the daily accumulated rainfall below 1 mm were replaced to 0 in the grid cell.

The quality of daily MSWEP v.2 data was evaluated by the Pearson correlation coefficient ( $\rho$ ), the coefficient of determination ( $r^2$ ), the percent bias (PBIAS) and the root mean square error (RMSE). The  $\rho$  indicates the degree of statistical relationship between two variables (in this case, two rainfall estimates from different source), ranging from -1 to 1, being 1 a perfect positive correlation and -1 the perfect negative correlation. The results ( $\rho$ ) were discussed considering the absolute value, and the correlation quality between ground-based rainfall observation and MSWEP v.2 estimation following the references: i) very strong

correlation  $\rho \geq |0,9|$ , ii) strong correlation  $|0,7| \leq \rho < |0,9|$ , iii) moderate correlation  $|0,5| \leq \rho < |0,7|$ , iv) weak correlation  $|0,3| \leq \rho < |0,5|$ , and v) very weak correlation  $|0,0| \leq \rho < |0,3|$ . The  $r^2$  was used to measure the degree of linear association between ground-based rainfall and estimated by MSWEP v2, being that values close to 1 stands for a better match between these data.

PBIAS was used to measure the average tendency (percentage) of the MSWEP v.2 rainfall values to be larger or smaller than the ground-based data. The RMSE was used to evaluate the standard deviation of the residuals (prediction errors) and to represent how concentrated the data is around the line of best fit, in other words, measures how much error in millimeters exists between the two datasets, being 0 the optimal value.

Besides the daily evaluation, a monthly evaluation was done as well by using the same statistical indicators ( $\rho$ ,  $r^2$ , PBIAS, and RMSE), in order to check for possible seasonal differences on MSWEP v.2 efficiency to estimate rainfall. When monthly data were analyzed, 11 rain gauge stations were considered in order to calculate monthly statistical indicators between ANA and MSWEP data - 3 rain gauges (#2851011, #2852001 and #2951039) were excluded due to the short time-series. For the selected rain gauges and the nearest MSWEP grid cell, monthly rainfall average was calculated. Furthermore, for monthly data, it was considered a thorough analysis exclusively with ANA rain gauges in good conditions (GGC), with a threshold of at least 5 year of data, and less than 10% of missing values.

For both, daily and monthly analyses, only ANA rain gauges with recorded data was used, including those inside the watershed (8 stations) and on its surroundings (6 stations). Days or months with missing data were not considered.


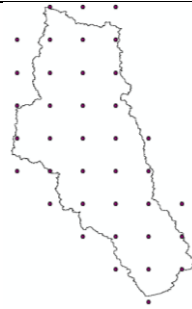

#### **2.2.4 Schemes for infilling data**

Three different strategies were used for infilling daily missing data (Table 1), were elaborated. Daily rainfall data was interpolated using the Inverse Distance Weighting (IDW) method in ArcGIS (ESRI, 2017) for the whole period (1979-2016), for each strategy. The interpolation procedure allows to evaluate the resulting differences of spatial variability of the rainfall strategies, or, hereafter called scenarios, over the watershed. The spatial resolution of the output IDW raster was approximately 238 m x 238 m.

The IDW methodology presume that information that are close to each other tend to be more similar than those are farther apart, and was chosen for its simplicity, wide use for geometric interpolation studies (LY; DEGRÉ; CHARLES, 2013; TUO *et al.*, 2016) accounting

for spatial heterogeneity in rainfall patterns (TUO *et al.*, 2016; WAGNER *et al.*, 2012), and due to good performance in hydrological modeling studies (RUELLAND *et al.*, 2008; TUO *et al.*, 2016; YANG *et al.*, 2015; YANTO; LIVNEH; RAJAGOPALAN, 2017).

Table 1 - Scenarios description of infilling rainfall data.

Name	Description	Spatial distribution of rainfall network
Scenario 1 (SC1)	It is the reference scenario, which uses only the rainfall records available on ANA dataset. The gaps were unfilled, and the data used for interpolation were those obtained purely from ANA. The rain gauges are heterogeneously distributed throughout the watershed. A total of 14 ANA rain gauges were used in this strategy.	
Scenario 2 (SC2)	The grid cell centroid value obtained from MSWEP v.2 dataset was used for daily rainfall IDW interpolation. The MSWEP rainfall product has high spatial homogeneity, accounting for 38 MSWEP grid cells. Total of 38 MSWEP grid cells were used in this strategy.	
Scenario 3 (SC3)	Measures from both, MSWEP V2 rainfall estimates and ground-based ANA rain gauge were merged. During the spatialization, when an ANA rain gauge had a missing value, the nearest MSWEP grid value was used for infilling. Furthermore, ANA rain gauges were fully replaced by MSWEP data for a long-term data with no records (e.g., those stations represented by X at Figure 1). Total of 29 rain gauges were used in this strategy.	

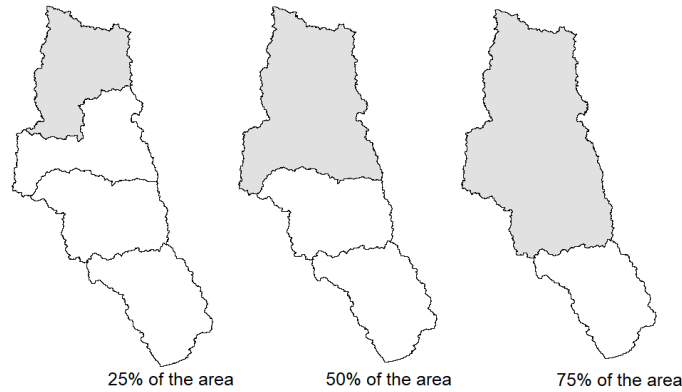
Source: the author.

After IDW spatialization procedure, the long-term average daily rainfall for Guaporé Watershed was extracted using zonal statistics in ArcGIS (ESRI, 2017) for all scenarios (1,2,3), in order to evaluate how the spatiotemporal irregularity of ground-based may affects the results of mean areal precipitation agreement between ANA and MSWEP dataset. This analysis was accomplished by dividing the watershed into four incremental areas of +25% (25%, 50%, 75% and 100% of the watershed area) from upstream to downstream (Figure 2). The  $\rho$ ,  $r^2$ , PBIAS



and RMSE were used to evaluate the agreement results between ANA and MSWEP rainfall data of 13880 days after spatialization, where Scenario 1 was fixed as the reference.

Figure 2- Incremental areas for daily statistical indicators agreement evaluation between ANA and MSWEP dataset.



Source: the author.

Additionally, the differences in the spatialized rainfall over the watershed, between SC2 and SC3 were measured considering SC1 as the reference. This analysis was performed pixel by pixel, over the 13880 days (1979-2016) according the Mean Absolute Error (MAE) statistical indicator (EQ. 1).

$$MAE = \frac{1}{13880} \sum_{d=1}^{13880} |Pr_d - Pe_d| \quad (1)$$

Where:  $d$  is the day,  $Pr$  is the reference rainfall (SC1),  $Pe$  is the estimated rainfall by the alternative scenarios (SC2, SC3).

## 2.3 RESULTS AND DISCUSSION

### 2.3.1 MSWEP dataset validation

Data for the 14 ANA rain gauges considered for validation check were processed in terms of daily rainfall against MSWEP v.2 data to perform the statistical indicators (Table 2). The correlation ( $\rho$ ) between MSWEP v.2 and the observed daily data was strong, ranging from 0.68 to 0.87, and the average was 0.78. The  $r^2$  average was 0.6, ranging from 0.43 to 0.75, which implies a good agreement between MSWEP v.2 and observed data. Quantitatively, the error (RMSE) ranged from 6.33 mm to 9.65 mm, with 7.56 mm average, in comparison to ANA rain

gauge data. The PBIAS indicates that MSWEP v.2 data underestimates rainfall data in average -15.6%, ranging from 0.40% (short-term time-series) to -29.9% (short-term time-series and high missing data).

Table 2- Rain gauge station code along with its snapping period, number of missing daily data, its percentage relative to the snapping period, and MSWEP daily validation metric results against ANA ground-based data.

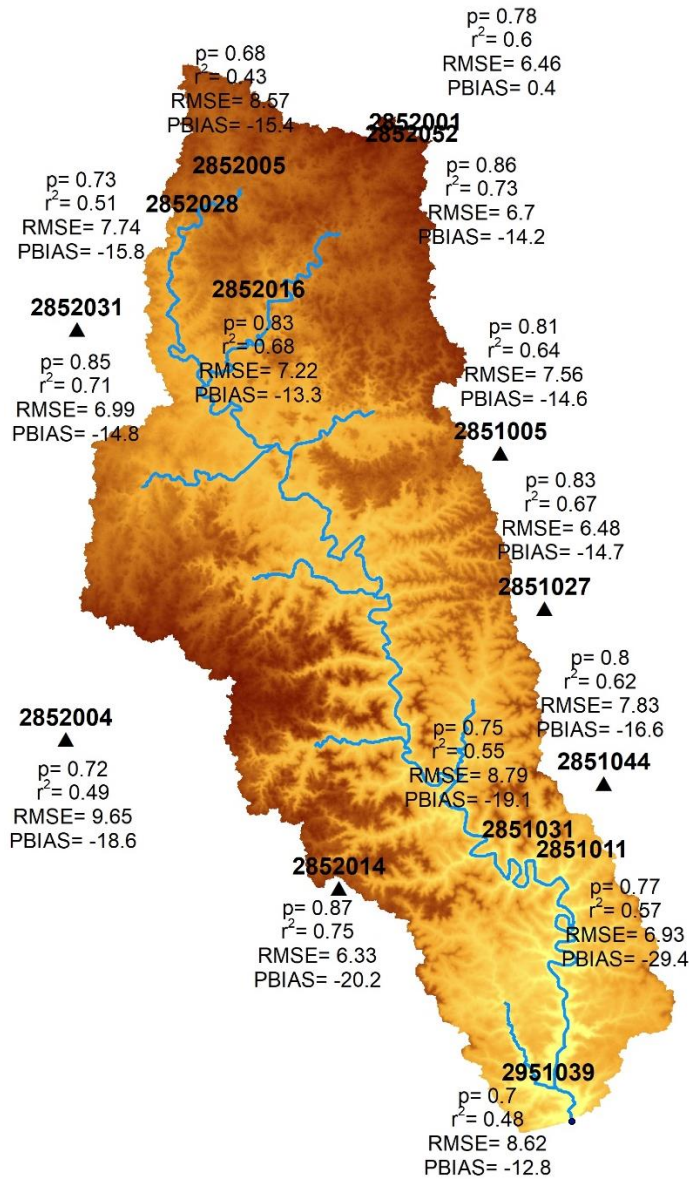
Location of the rain gauge	ANA rain gauge code	Monitoring period	Missing data (days)	Missing data (%)	$\rho$	$r^2$	RMSE (mm)	PBIAS (%)
Within the watershed	2851011	01/01/1979-11/26/1980	301	43	0.77	0.57	6.93	-29.4
	2851031	01/01/1979-12/31/1982	0	-	0.75	0.55	8.79	-19.1
	2852001	01/01/1979-12/31/1979	0	-	0.78	0.60	6.46	0.40
	2852005	01/02/1979-12/31/1982	0	-	0.68	0.43	8.57	-15.4
	2852016	01/01/1979-10/31/2012	0	-	0.83	0.68	7.22	-13.3
	2852028	01/02/1979-12/21/1982	0	-	0.73	0.51	7.74	-15.8
	2852052	12/20/2002-12/31/2016	521	10.16	0.86	0.73	6.70	-14.2
	2951039	01/01/1996-12/31/1997	30	4.10	0.70	0.48	8.62	-12.8
Watershed surroundings	2851005	01/01/1979- 11/30/2012	0	-	0.81	0.64	7.56	-14.6
	2851027	01/01/1979-11/30/2000	5113	63.90	0.83	0.67	6.48	-14.7
	2851044	05/01/1985-2/28/2015	1867	17.13	0.80	0.62	7.83	-16.6
	2852004	01/01/1979-12/31/2016	1155	8.32	0.72	0.49	9.65	-18.6
	2852014	01/01/1979-12/31/2012	0	-	0.87	0.75	6.33	-20.2
	2852031	01/03/1979-10/31/2012	0	-	0.85	0.71	6.99	-14.8
Average					0.78	0.60	7.56	-15.6

Source: the author.

The rainfall underestimation, probably, occurs mainly due to the convective rainfall occurrence during the summers at Guaporé watershed region. The intense heat and high humidity levels during the summers favors the convective rainfall activity, characterized by high intensity in a short-term duration, usually an hour or few hours. Consequently, MSWEP v.2 is unable to predict convective rainfall events, since its temporal resolution is longer than the convective rainfall event duration. The underestimation rainfall intensity trend has been reported by SPPs (HEROLD *et al.*, 2016), and by MSWEP product (AWANGE; HU; KHAKI, 2019; XU *et al.*, 2019).

Statistical indicators result according to the rain gauge location in the basin context, shows neither a pattern to a better or worst result related to elevation, for instance (Figure 3). However, the best  $\rho$  values were generally associated with rain gauges with long-term rainfall monitoring. In terms of RMSE and PBIAS, similar results were found, indicating a good overall prediction of MSWEP v.2 dataset.

Figure 3- Daily statistical indicators ( $\rho$ ,  $r^2$ , RMSE and PBIAS) results for each ground-based ANA rain gauge against MSWEP v.2, and its correspondent spatial location.



Source: the author.

For monthly evaluation, six ANA stations, among all evaluated ANA rain gauges met the requirement and are considered as gauge in good conditions (GGC), the six stations are highlighted in bold in Table 3. In general, the statistical indicators had good agreement between ANA rain gauge data and MSWEP v.2, with  $\rho$  ranging from -0.44 to 1.00, outstanding strong correlations with values above 0.8 in most cases. Furthermore,  $\rho > 0.75$  was observed for monthly rainfall evaluation when considered only ANA rain gauges in GGC.

Table 3- MSWEP monthly validation results for Pearson correlation coefficient ( $\rho$ ), coefficient of determination ( $r^2$ ), root mean square error (RMSE) and percentage bias (PBIAS).

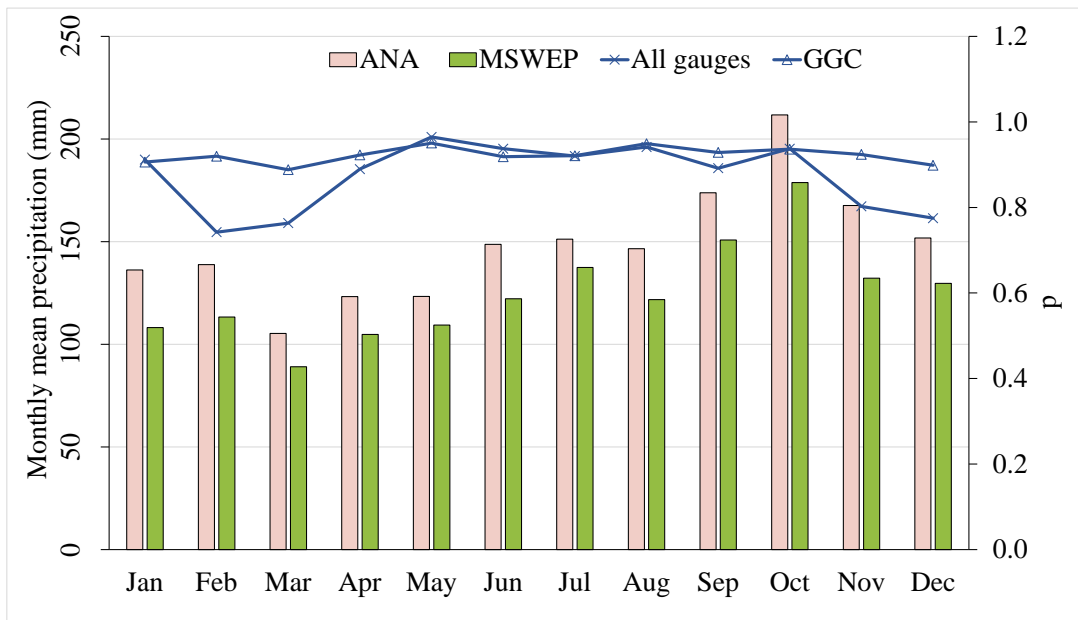
		Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
$\rho$													
Within the watershed	2851031	0.83	0.43	0.14	0.76	1.00	0.99	0.86	0.94	0.86	0.9	0.96	0.43
	2852005	0.98	0.36	0.48	0.82	1.00	1.00	0.95	0.96	0.73	0.99	-0.44	0.31
	<b>2852016*</b>	<b>0.89</b>	<b>0.91</b>	<b>0.92</b>	<b>0.96</b>	<b>0.95</b>	<b>0.95</b>	<b>0.94</b>	<b>0.92</b>	<b>0.94</b>	<b>0.9</b>	<b>0.93</b>	<b>0.96</b>
	2852028	0.98	0.04	0.64	0.98	0.99	1.00	0.97	0.84	0.79	0.98	0.91	0.62
	<b>2852052</b>	<b>0.95</b>	<b>0.92</b>	<b>0.94</b>	<b>0.95</b>	<b>0.97</b>	<b>0.97</b>	<b>0.98</b>	<b>0.99</b>	<b>0.96</b>	<b>0.99</b>	<b>0.97</b>	<b>0.97</b>
Watershed's surroundings	<b>2851005</b>	<b>0.95</b>	<b>0.95</b>	<b>0.87</b>	<b>0.92</b>	<b>0.91</b>	<b>0.88</b>	<b>0.91</b>	<b>0.96</b>	<b>0.93</b>	<b>0.94</b>	<b>0.88</b>	<b>0.89</b>
	2851027	0.99	0.91	0.97	0.98	0.96	0.94	0.95	0.97	0.9	0.96	0.94	0.92
	2851044	0.82	0.92	0.84	0.7	0.96	0.87	0.87	0.95	0.96	0.87	0.91	0.84
	<b>2852004</b>	<b>0.79</b>	<b>0.86</b>	<b>0.85</b>	<b>0.81</b>	<b>0.93</b>	<b>0.83</b>	<b>0.81</b>	<b>0.87</b>	<b>0.88</b>	<b>0.89</b>	<b>0.91</b>	<b>0.77</b>
	<b>2852014</b>	<b>0.98</b>	<b>0.94</b>	<b>0.94</b>	<b>0.95</b>	<b>0.98</b>	<b>0.95</b>	<b>0.95</b>	<b>0.98</b>	<b>0.96</b>	<b>0.97</b>	<b>0.97</b>	<b>0.94</b>
<b>2852031</b>	<b>0.88</b>	<b>0.94</b>	<b>0.8</b>	<b>0.94</b>	<b>0.96</b>	<b>0.93</b>	<b>0.93</b>	<b>0.97</b>	<b>0.9</b>	<b>0.92</b>	<b>0.88</b>	<b>0.86</b>	
Average $\rho$	0.91	0.74	0.76	0.89	0.96	0.94	0.92	0.94	0.89	0.94	0.8	0.77	
$r^2$													
Within the watershed	2851031	0.10	-0.76	-15.02	0.38	0.94	0.66	0.72	0.52	0.58	0.38	-1.28	-0.25
	2852005	0.95	-0.49	-1.26	0.62	0.99	0.83	0.88	0.06	0.18	0.64	-2.19	-0.06
	<b>2852016</b>	<b>0.66</b>	<b>0.75</b>	<b>0.80</b>	<b>0.89</b>	<b>0.85</b>	<b>0.83</b>	<b>0.83</b>	<b>0.74</b>	<b>0.83</b>	<b>0.78</b>	<b>0.78</b>	<b>0.86</b>
	2852028	0.75	-0.49	-0.30	0.88	0.95	0.86	0.90	-0.19	0.38	0.58	-0.69	0.38
	<b>2852052</b>	<b>0.76</b>	<b>0.68</b>	<b>0.63</b>	<b>0.88</b>	<b>0.86</b>	<b>0.86</b>	<b>0.91</b>	<b>0.85</b>	<b>0.83</b>	<b>0.94</b>	<b>0.86</b>	<b>0.91</b>
Watershed's surroundings	<b>2851005</b>	<b>0.80</b>	<b>0.81</b>	<b>0.66</b>	<b>0.76</b>	<b>0.72</b>	<b>0.64</b>	<b>0.77</b>	<b>0.88</b>	<b>0.81</b>	<b>0.83</b>	<b>0.71</b>	<b>0.69</b>
	2851027	0.82	0.74	0.78	0.82	0.82	0.65	0.74	0.86	0.67	0.77	0.57	0.81
	2851044	0.35	0.72	0.55	0.26	0.89	0.54	0.63	0.85	0.85	0.61	0.62	0.52
	<b>2852004</b>	<b>0.54</b>	<b>0.54</b>	<b>0.71</b>	<b>0.60</b>	<b>0.76</b>	<b>0.43</b>	<b>0.31</b>	<b>0.72</b>	<b>0.67</b>	<b>0.62</b>	<b>0.65</b>	<b>0.56</b>
	<b>2852014</b>	<b>0.68</b>	<b>0.62</b>	<b>0.68</b>	<b>0.72</b>	<b>0.90</b>	<b>0.69</b>	<b>0.67</b>	<b>0.84</b>	<b>0.80</b>	<b>0.86</b>	<b>0.72</b>	<b>0.64</b>
<b>2852031</b>	<b>0.64</b>	<b>0.81</b>	<b>0.50</b>	<b>0.86</b>	<b>0.91</b>	<b>0.79</b>	<b>0.78</b>	<b>0.88</b>	<b>0.72</b>	<b>0.77</b>	<b>0.68</b>	<b>0.62</b>	
Avg. $r^2$	0.64	0.36	-1.02	0.70	0.87	0.71	0.74	0.64	0.66	0.71	0.13	0.52	
RMSE (mm)													
Within the watershed	2851031	63.9	65.9	37.5	31.4	13.3	73.4	34.8	34.2	36.7	32.5	90.3	75.8
	2852005	10.6	38.9	34.1	21.2	5.8	35.9	22.9	32.6	33.1	60.5	69.6	61.2
	<b>2852016</b>	<b>44.6</b>	<b>46.8</b>	<b>23.3</b>	<b>28.8</b>	<b>35.0</b>	<b>25.2</b>	<b>37.1</b>	<b>43.0</b>	<b>44.0</b>	<b>49.1</b>	<b>42.5</b>	<b>25.9</b>
	2852028	33.5	41.4	17.7	11.3	16.0	29.1	21.5	33.6	25.6	58.3	66.7	46.9
	<b>2852052</b>	<b>30.5</b>	<b>38.5</b>	<b>40.8</b>	<b>24.0</b>	<b>26.3</b>	<b>27.9</b>	<b>28.3</b>	<b>38.9</b>	<b>32.8</b>	<b>28.0</b>	<b>27.5</b>	<b>27.9</b>
Watershed's surroundings	<b>2851005</b>	<b>40.1</b>	<b>38.3</b>	<b>35.6</b>	<b>41.8</b>	<b>47.9</b>	<b>35.8</b>	<b>41.0</b>	<b>26.3</b>	<b>49.5</b>	<b>43.0</b>	<b>37.2</b>	<b>33.4</b>
	2851027	52.4	34.7	22.6	17.5	17.3	32.0	24.3	32.9	26.6	40.8	46.2	27.7
	2851044	52.2	44.3	40.1	54.3	20.5	46.4	48.4	37.7	35.7	63.1	42.1	45.2
	<b>2852004</b>	<b>58.1</b>	<b>45.5</b>	<b>32.6</b>	<b>54.1</b>	<b>40.4</b>	<b>47.2</b>	<b>68.9</b>	<b>48.5</b>	<b>53.1</b>	<b>59.7</b>	<b>60.9</b>	<b>45.3</b>
	<b>2852014</b>	<b>51.2</b>	<b>43.2</b>	<b>40.4</b>	<b>32.8</b>	<b>27.1</b>	<b>41.0</b>	<b>47.1</b>	<b>33.9</b>	<b>42.6</b>	<b>37.2</b>	<b>51.6</b>	<b>41.1</b>
<b>2852031</b>	<b>45.4</b>	<b>37.2</b>	<b>35.3</b>	<b>27.6</b>	<b>29.9</b>	<b>29.7</b>	<b>42.9</b>	<b>30.6</b>	<b>48.2</b>	<b>51.7</b>	<b>48.4</b>	<b>43.7</b>	
Avg. RMSE	43.9	43.2	32.7	31.4	25.4	38.5	37.9	35.7	38.9	47.6	53.0	43.1	
PBIAS (%)													
Within the watershed	2851031	-42.9	-24.3	99.4	-22.4	-13.1	-28.6	6.5	-18.8	-15.0	-11.9	-35.7	-22.6
	2852005	-4.2	-18.9	-29.3	-11.5	-4.4	-19.0	-3.5	-19.7	-14.4	-22.3	-11.0	-9.5
	<b>2852016</b>	<b>-17.6</b>	<b>-16.7</b>	<b>-12.0</b>	<b>-9.2</b>	<b>-11.1</b>	<b>-11.4</b>	<b>-10.5</b>	<b>-20.0</b>	<b>-12.4</b>	<b>-11.2</b>	<b>-16.1</b>	<b>-11.0</b>
	2852028	-22.6	-16.3	-16.2	-13.0	-9.0	-18.3	-6.9	-17.2	-9.9	-22.7	-26.8	-1.6
	<b>2852052</b>	<b>-13.8</b>	<b>-18.0</b>	<b>-23.1</b>	<b>-7.7</b>	<b>-14.7</b>	<b>-12.8</b>	<b>-12.0</b>	<b>-20.6</b>	<b>-13.5</b>	<b>-8.7</b>	<b>-15.4</b>	<b>-12.7</b>
Watershed's surroundings	<b>2851005</b>	<b>-16.5</b>	<b>-17.4</b>	<b>-16.1</b>	<b>-16.8</b>	<b>-18.6</b>	<b>-15.7</b>	<b>-12.4</b>	<b>-12.5</b>	<b>-13.4</b>	<b>-11.3</b>	<b>-13.0</b>	<b>-12.6</b>
	2851027	-21.4	-12.9	-16.6	-12.9	-7.6	-18.3	-13.8	-15.1	-10.6	-12.5	-25.4	-7.3
	2851044	-23.7	-19.9	-18.8	-18.8	-8.2	-20.9	-15.9	-16.3	-9.3	-16.7	-19.6	-19.8
	<b>2852004</b>	<b>-18.7</b>	<b>-21.0</b>	<b>-11.5</b>	<b>-22.3</b>	<b>-13.5</b>	<b>-12.4</b>	<b>-26.2</b>	<b>-16.6</b>	<b>-17.0</b>	<b>-20.4</b>	<b>-22.9</b>	<b>-16.3</b>
	<b>2852014</b>	<b>-26.3</b>	<b>-23.8</b>	<b>-24.5</b>	<b>-17.7</b>	<b>-14.0</b>	<b>-20.1</b>	<b>-21.2</b>	<b>-18.9</b>	<b>-17.4</b>	<b>-13.4</b>	<b>-25.1</b>	<b>-23.0</b>
<b>2852031</b>	<b>-18.3</b>	<b>-15.6</b>	<b>-18.6</b>	<b>-9.3</b>	<b>-9.9</b>	<b>-13.2</b>	<b>-16.0</b>	<b>-16.1</b>	<b>-14.7</b>	<b>-14.4</b>	<b>-16.7</b>	<b>-15.5</b>	
Avg. PBIAS	-20.6	-18.6	-7.9	-14.7	-11.3	-17.3	-12.0	-17.4	-13.4	-15.1	-20.7	-13.8	

Source: the author.

\*The stations highlighted in bold are the ones which met the requirement and are considered as gauge in good conditions (GGC).

The monthly rainfall mean differences between ANA and MSWEP are not evenly distributed among months (Figure 4). However, the differences between monthly means are higher during September, October and November, which compose the main rainfall season of the Guaporé watershed. The driest months (March, April and May) had the lowest differences in monthly rainfall mean between both datasets. The overall  $\rho$  between ANA and MSWEP monthly mean combining all rainfall stations were lower for summer (January, February and March) and spring (October, November and December). However, considering only the GGC,  $\rho$  values for these seasons increase.

Figure 4- Monthly mean rainfall and  $\rho$ .



Source: the author.

The worst statistical indicators of  $r^2$ , RMSE and PBIAS were observed for summer and spring seasons, similarly to  $\rho$  results. The better performance of PBIAS in March is bias, once in this particular month one station (2851031) had MSWEP rainfall monthly mean 99.4% higher than the ground-based station (Table 3). This result directly affected the final rainfall monthly mean and PBIAS result for March, giving a wrong idea that even though it is summer and the pattern would be higher disagreement, MSWEP presented better accordance with observed data that was expected. This also explains the negative value of  $r^2$  for this month. Overall, MSWEP v.2 dataset had good performance for daily and monthly agreement with ANA gauges in the Guaporé watershed. MSWEP v.2 provided a good potential dataset for

watershed with sparse rain gauge network, and might, as well be used for infilling missing rainfall values.

### **2.3.2 Infilling missing data and spatialization**

Since the validation results between ANA and MSWEP rainfall dataset showed good agreement, based on the statistical indicators, the daily infilling procedure for SC3 (described on section 2.2.4) was possible. After infilling the gaps, daily interpolation for each scenario was accomplished for the entire period of study (1979-2016), and the following results are discussed in terms of long-term statistical indicators and spatial variability within the watershed, according incremental watershed area.

#### *2.3.2.1 Statistical indicators result after infilling and spatialization*

The scenario with only MSWEP data (SC2) produced an almost equivalent rainfall mean found in SC1, with strong correlation ( $\rho > 0.89$ ), good  $r^2 > 0.79$ , RMSE  $< 4.88$  mm and PBIAS  $-16.12\%$ . The RMSE reduced and PBIAS increased as the watershed area increased, possibly as a result of uncertainties related to the scarcity of rain gauge data in the downstream portion of the watershed, which outstand the underestimates related to MSWEP.

The strategy of merging MSWEP for infill ANA missing data in SC3 resulted in the best set of statistical indicators, with very strong correlation ( $\rho > 0.90$ ), very good  $r^2 > 0.80$ , low RMSE  $< 4.72$  mm and lower PBIAS  $< |14.63|$ . In this case, as the watershed area increased, statistical indicators results became worst, possibly as consequence of the higher number of rain gauges with short time series in the downstream portion of the watershed, which data were replaced by MSWEP data.

Table 4 – Statistical indicators ( $\rho$ ,  $r^2$ , RMSE and PBIAS) results for daily average rainfall according to the incremental watershed area

	<b>25% BA</b>	<b>50% BA</b>	<b>75% BA</b>	<b>100% BA</b>
<b><math>\rho</math></b>				
SC1-SC2	0.88	0.89	0.90	0.89
SC1-SC3	0.93	0.92	0.91	0.90
<b><math>r^2</math></b>				
SC1-SC2	0.77	0.79	0.80	0.79
SC1-SC3	0.86	0.85	0.83	0.80
<b>RMSE (mm)</b>				
SC1-SC2	5.61	5.18	4.91	4.88
SC1-SC3	4.37	4.42	4.53	4.72
<b>PBIAS (%)</b>				
SC1-SC2	-13.36	-14.03	-14.72	-16.12
SC1-SC3	-2.10	-11.43	-12.99	-14.63

Source: the author.

Thus, the simple replacement of rain gauge data by MSWEP v.2 dataset (SC2) could provide an acceptable result for estimate the mean rainfall in the watershed, although with uncertainties associated, mainly due to rainfall underestimates. However, when the MSWEP was used to infill the ANA rain gauge missing data (SC3), better results were observed after spatialization. The  $\rho$ ,  $r^2$  and RMSE had an improvement when compared with the direct analysis between rain gauge data and MSWEP dataset (Table 1). The PBIAS had a nearly constant 15% of underestimations.

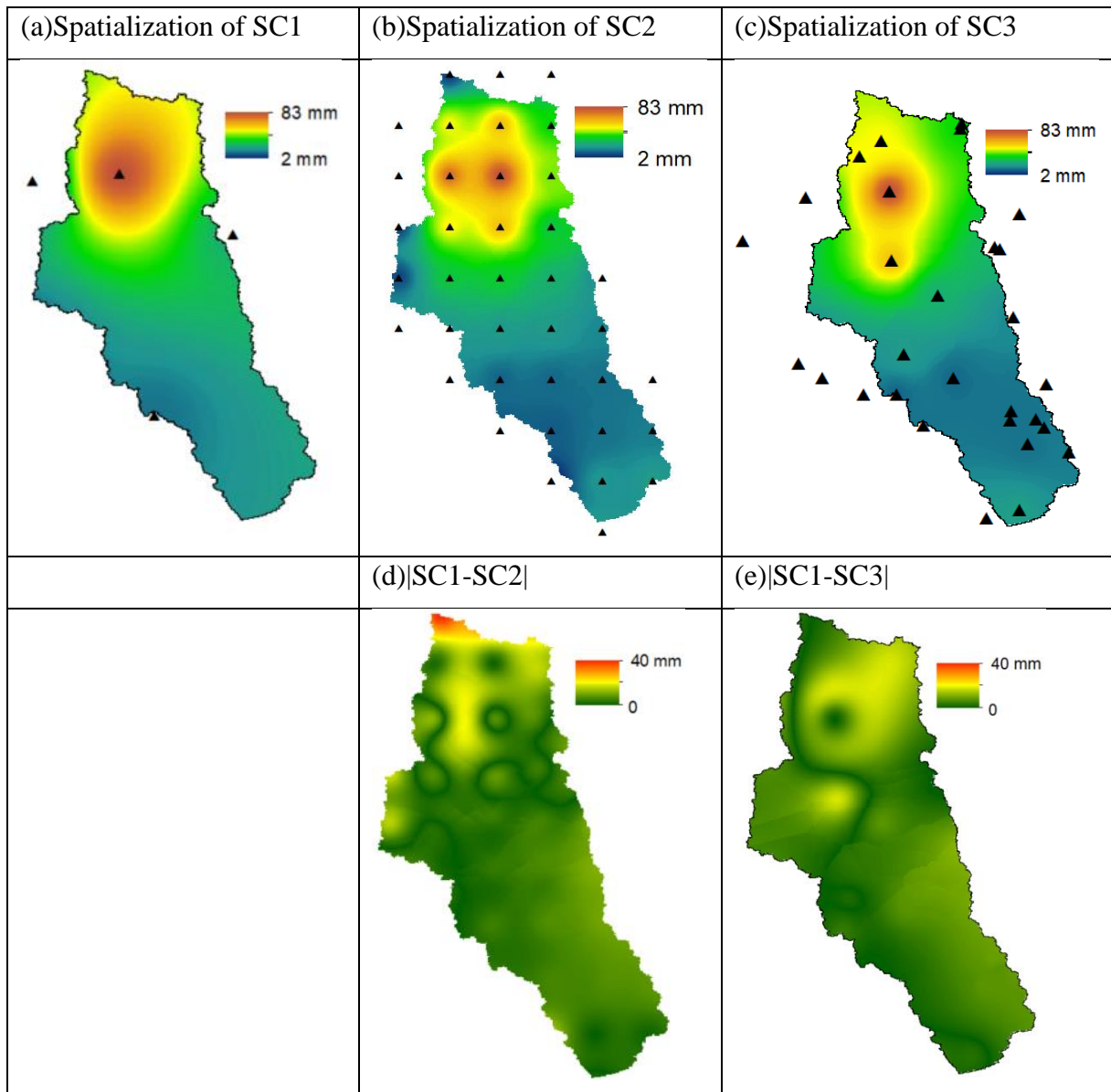
The infilling procedure, besides infilling the missing data, allows longer time series and a higher number of rain gauges in the spatialization process, which provide a more detailed mean areal rainfall for the watershed (VICENTE-SERRANO *et al.*, 2010). Furthermore, the incremental area analysis suggests that in areas where the ANA rain gauge network was denser, the infilling procedure using MSWEP data resulted in lower uncertainties in the spatialization. This finding is a plausible result once one of MSWEP v.2 primary resource for dataset construction is ground-based gauges (BECK *et al.*, 2018).

### 2.3.2.2 Visual spatial analyze

Firstly, it is presented a random day spatialization for an event with average rainfall over the watershed above 40 mm (Figure 5). On this specific day, SC1 had only one rain gauge within the watershed with rainfall data, and 3 closes to its borders. In addition, there are places

with a higher disagreement between SC1 and the other scenarios, especially in the north portion of the watershed. This coarse spatial distribution occurs throughout most part of the time series studied, which introduces possible interpolation errors due to high distance among ANA rain gauges.

Figure 5- Example of rainfall spatial subtraction between scenarios for a random day (10/30/1983). Figure (a) for SC1, (b) for SC2, and (c) for SC3 spatialization, and the correspondent modular spatial difference between SC1 and SC2 (d) and SC1 and SC3 (e).



Source: the author.

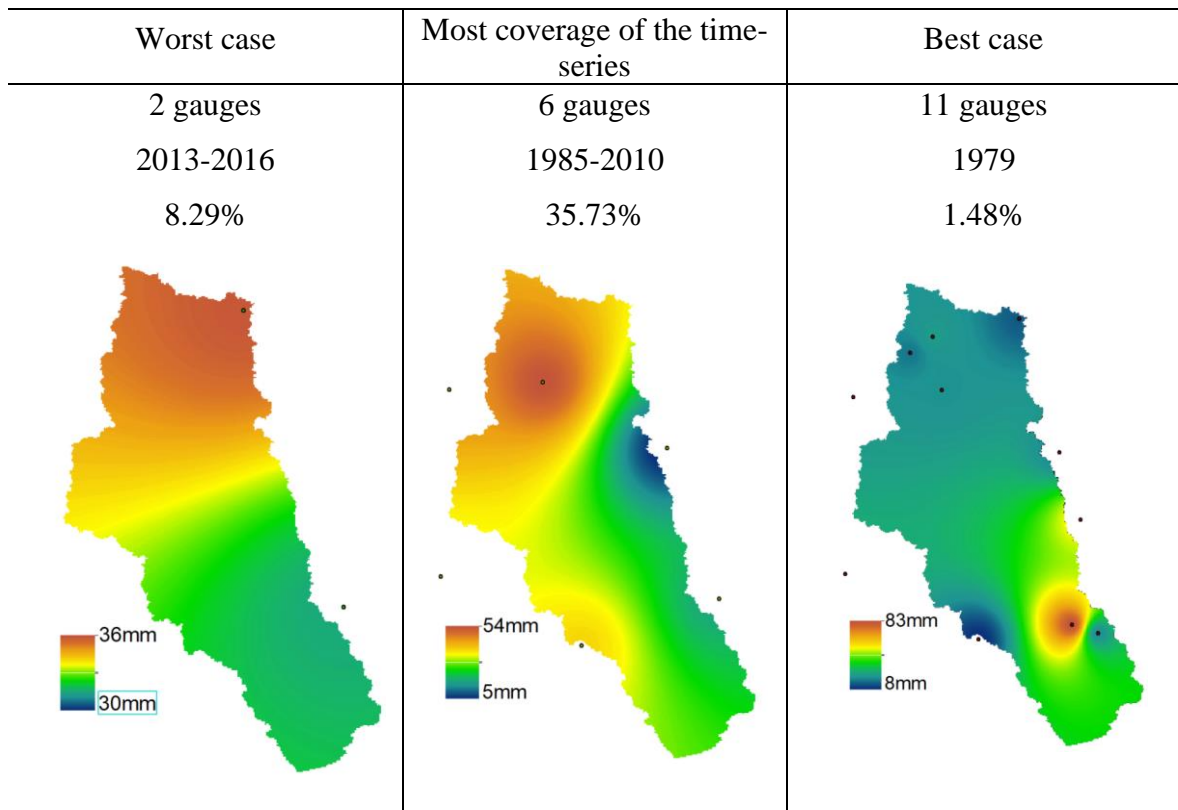
A geostatistical interpolation method, such as the simple kriging, the ordinary kriging or even the universal kriging, could be used to overcome the errors introduced by IDW interpolator. However, there is no interpolation method that stands out for every case of study,



since the best fit is more dependent on the conditions found in the specific area of study (LY; DEGRÉ; CHARLES, 2013) than the interpolation methodology. For instance, Zimmerman (1999) only found better interpolation results using ordinary kriging and universal kriging, when the rain gauge network was regular, the noise low and the spatial correlation strong, otherwise IDW performed better.

The inequality of ANA's rain gauges data throughout the years (Figure 6) stands the difficulties found to explore rainfall data in order to apply the most powerful interpolation methodologies. Along the most part of the time series (35.73%), exclusively six ANA rain gauges could provide daily rainfall data for interpolation for the reference scenario (SC1). The best situation was observed in less than 2% of the days (1.48%) in which 11 rain gauges were used in spatialization, and the worst condition was found along 8.29% of the days, where only 2 ANA rain gauges had available data. However, the percentage of good periods in the time-series is much smaller than the bad periods which only 2 ANA's stations were being used.

Figure 6- Temporal irregularity in the number of rain gauges used for interpolation in SC1.

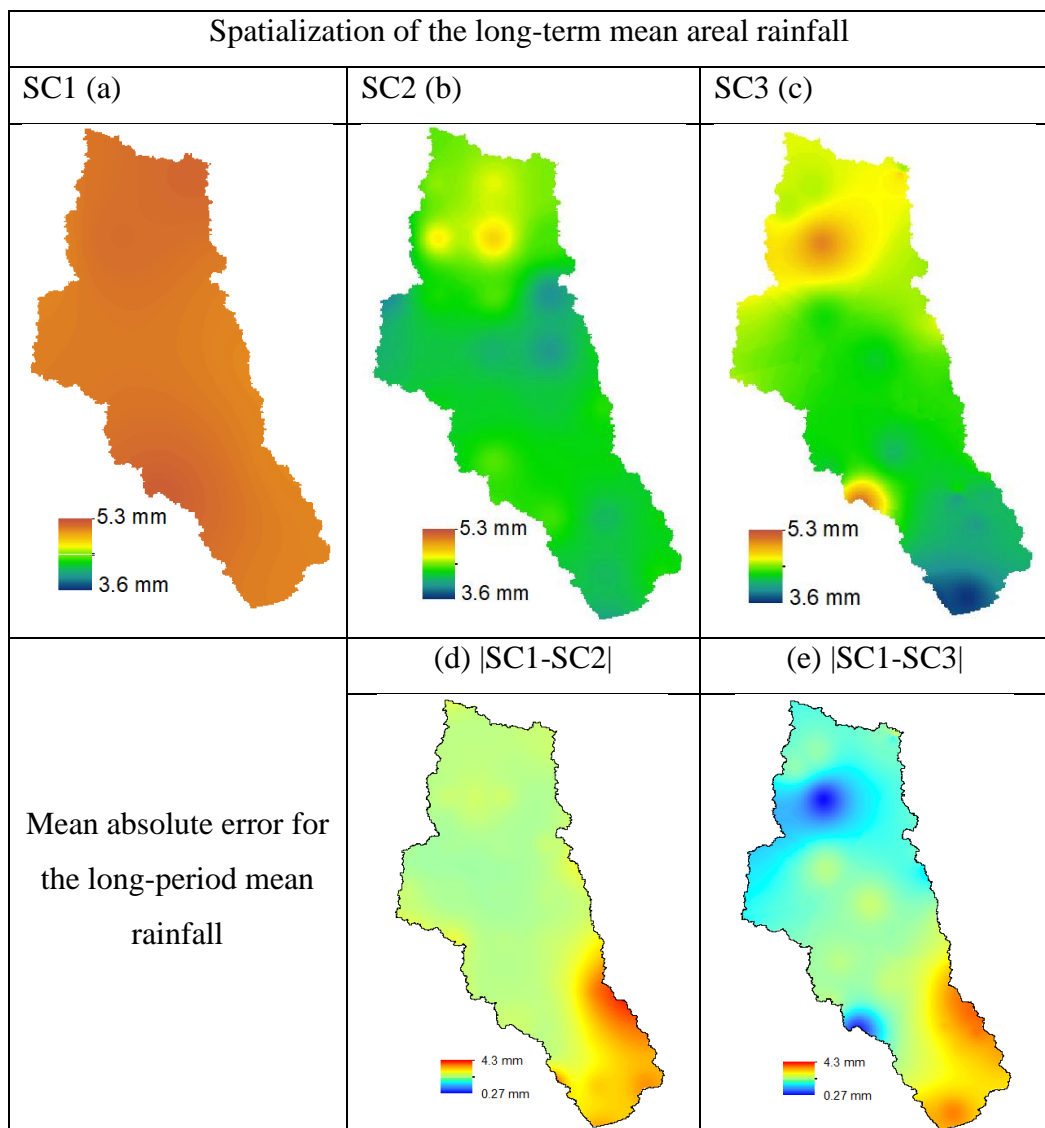


Source: the author.

The analysis of the long-term mean areal rainfall highlights the differences in daily rainfall due to the strategies of filling missing data, when compared with the reference scenario.

As a result, a set of two images were created, SC1-SC2 and SC1-SC3, providing a visual feature of the long-term spatial distribution of differences between the scenarios (Figure 7). This analysis is especially relevant towards hydrologic modeling, once it is possible to identify regions where the additional rainfall data influences the spatialization, which could represent a decrease in modeling uncertainties (STRAUCH *et al.*, 2012; TUO *et al.*, 2016).

Figure 7- Long-term (1979-2016) mean spatial rainfall and error (MEA). Figure (a),(b) and(c) shows the long-term mean areal rainfall for SC1, SC2 and SC3, respectively, followed by the spatial mean absolute error between the reference scenario SC1 and SC2 (d), and SC3(e).



Source: the author.

The mean areal rainfall is clearly different between SC1 and the two alternative scenarios (SC2 and SC3) (Figure 7). The SC1 had a nearly homogeneous rainfall spatialization, in contrast to SC2 and SC3 where high variability over the watershed was observed. This

finding endorses the potential advantage of using SC2 and SC3, once it provides a more realistic temporal/spatial rainfall variability over the watershed, being a valuable ally to use in hydrological models, which its good results rely fairly on the rain gauge network spatial coverage quality (ARNOLD *et al.*, 2012).

The mean spatial error presented a pattern of higher values in the south portion of the watershed, decreasing towards the north, which agree with the tendency of MSWEP to have worst performance in areas with poor ANA gauge network, as shown with the increment of areas analysis. Even though there are 3 rain gauges in the last quarter of the increment area, these gauges (2951039, 2851011, and 2851031) provide low influence over the long-term analysis (1979-2016), once their snap period are 1996-1997, 1979-1980, 1979-1982, with 4.10%, 43.00% and 0% of missing data, respectively.

The highest mean error is located on the south of the watershed, in general, for both cases (Figure 7 (d) and (e)), with the highest rainfall error of 4.3 mm. Two featured regions, one on the southwest and the other one in the north had the mean error lower than in the other regions (Figure 7 (e)). This result occurred due to the presence of two rain gauges in good conditions, in which long-term daily rainfall covers a period from 1979 until 2012 without any gaps. Thus, exclusively the period from 2012 to 2016 needed MSWEP v.2 data to infill missing rain gauge data, reducing the regional mean areal error. The average MAE for our base scenario and SC3 was 2.12 mm.

## 2.4 CONCLUSION

This study discusses the issues regarding poor rain gauge network density that is found in most developing countries, and presents ways of overcoming this problem by using satellite-based rainfall product to improve the spatial density of rain gauges, extend the time-series and specially infilling missing values in a daily time step. MSWEP v.2 product was tested in the replacement of rain gauge observation, as well as infilling missing rainfall data and extension of the time-series in a mid-size watershed in South Brazil.

The validation check procedure for the MSWEP v.2 dataset against the ground-based observation (rain gauge) showed good agreement, presenting remarkable statistical indicators values on a daily basis (0.78 for  $\rho=0.78$ ,  $r^2= 0.60$ , RMSE=7.56 mm, and PBIAS=-15.6%) . Following, the assessment of MSWEP v.2 as source for precipitation data, after IDW interpolation the scenario where MSWEP v.2 dataset fully replaced the ground-based rain gauges (SC2) succeeded in represent daily rainfall and its spatial variability in the watershed.

The average agreement in the whole watershed between MSWEP dataset and the ground-based reference (rain gauges) resulted in 0.89 for  $\rho$ , 0.79 for  $r^2$ , -16.12% for PBIAS, 4.88 mm for RMSE, and in a low MAE for the long-term around 2.53 mm.

When MSWEP v.2 data was merged with rain gauge data in order to infill rainfall gaps, or for replacing rain gauge data without any data throughout the analyzed period (SC3) the resulting statistical indicators were even better ( $\rho=0.90$ ,  $r^2=0.80$ , RMSE= 4.72 mm, PBIAS=-14.63% and MAE=2.12mm). The strategy of using the ground-based observation combined with MSWEP dataset produces a better result to estimate average daily rainfall over the watershed and longer time-series with high spatial variability.

The pronounced differences in the estimation of mean areal rainfall between MSWEP and the observed data were related to the lack of rain gauge representativeness (low density and shortcomings) in the watershed downstream region, which provided poor conditions for the spatialization process. In these cases, MSWEP dataset can be used in places where the rain gauge network is sparse and with shortcomings. The approach has the capability for extending daily rainfall time-series and improving the spatial variability of rainfall in the study area. A possible alternative to overcome MSWEP tendency of underestimation in further studies, could be the proposal of a monthly correction coefficient.

## 2.5 REFERENCES

AIEB, A. *et al.* A new approach for processing climate missing databases applied to daily rainfall data in Soummam watershed, Algeria. **Heliyon**, [S. l.], v. 5, n. 2, p. e01247, 2019. Disponível em: <https://doi.org/10.1016/j.heliyon.2019.e01247>

ALVARES, C. A. *et al.* Köppen's climate classification map for Brazil. **Meteorologische Zeitschrift**, [S. l.], v. 22, n. 6, p. 711–728, 2013. Disponível em: <https://doi.org/10.1127/0941-2948/2013/0507>

ANA. **Conjuntura dos recursos hídricos no Brasil**. [s. l.], 2017. Disponível em: <http://www3.snirh.gov.br/portal/snirh/centrais-de-conteudos/conjuntura-dos-recursos-hidricos>. Acesso em: 21 set. 2018.

ARNOLD, J. G. *et al.* SWAT: model use, calibration and validation. **American Society of Agricultural and Biological Engineers**, [S. l.], v. 55, n. 4, p. 1491–1508, 2012.

AWANGE, J. L.; HU, K. X.; KHAKI, M. The newly merged satellite remotely sensed, gauge and reanalysis-based Multi-Source Weighted-Ensemble Precipitation: Evaluation over Australia and Africa (1981–2016). **Science of the Total Environment**, [S. l.], v. 670, p. 448–465, 2019. Disponível em: <https://doi.org/10.1016/j.scitotenv.2019.03.148>

BECK, H. E. *et al.* MSWEP V2 global 3-hourly 0.1° precipitation: methodology and quantitative assessment. **Bulletin of the American Meteorological Society**, [S. l.], 2018. Disponível em: <https://doi.org/10.1175/BAMS-D-17-0138.1>

BECK, H. E. *et al.* Daily evaluation of 26 precipitation datasets using Stage-IV gauge-radar data for the CONUS. **Hydrology and Earth System Sciences**, [S. l.], v. 23, n. 1, p. 207–224, 2019. Disponível em: <https://doi.org/10.5194/hess-23-207-2019>

CHOW, V. T.; MAIDMENT, D. R.; MAYS, L. W. **Applied hydrology**. Water Resources Handbook, 1988.

COULIBALY, P.; EVORA, N. D. Comparison of neural network methods for infilling missing daily weather records. **Journal of Hydrology**, [S. l.], v. 341, n. 1–2, p. 27–41, 2007. Disponível em: <https://doi.org/10.1016/j.jhydrol.2007.04.020>

DE CARVALHO, J. R. P. *et al.* Modelo de imputação múltipla para estimar dados de precipitação diária e preenchimento de falhas. **Revista Brasileira de Meteorologia**, [S. l.], v. 32, n. 4, p. 575–583, 2017. Disponível em: <https://doi.org/10.1590/0102-7786324006>

DINKU, T. *et al.* Combined use of satellite estimates and rain gauge observations to generate high-quality historical rainfall time series over Ethiopia. **International Journal of Climatology**, [S. l.], v. 34, n. 7, p. 2489–2504, 2014. Disponível em: <https://doi.org/10.1002/joc.3855>

ESLAMIAN, S. **Handbook of engineering hydrology: modeling, climate change, and variability**. [S. l.]: CRC Press, 2014. *E-book*.

ESRI. **ArcGIS 10.4**. [S. l.: s. n.]

GITHUNGO, W. *et al.* Infilling monthly rain gauge data gaps with satellite estimates for ASAL of Kenya. **Hydrology**, [S. l.], v. 3, n. 4, 2016. Disponível em: <https://doi.org/10.3390/hydrology3040040>

HASAN, M. M.; CROKE, B. F. W. Filling gaps in daily rainfall data: a statistical approach. **20th International Congress on Modelling and Simulation, Adelaide, Australia**, [S. l.], n. December, p. 1–6, 2013. Disponível em: <http://www.mssanz.org.au/modsim2013/A9/hasan.pdf>

HEROLD, N. *et al.* How much does it rain over land? **Geophysical Research Letters**, [S. l.], v. 43, n. 1, p. 341–348, 2016. Disponível em: <https://doi.org/10.1002/2015GL066615>

HOU, A. Y. *et al.* The global precipitation measurement mission. **Bulletin of the American Meteorological Society**, [S. l.], v. 95, n. 5, p. 701–722, 2014. Disponível em: <https://doi.org/10.1175/BAMS-D-13-00164.1>

HUGHES, D. A. Comparison of satellite rainfall data with observations from gauging station networks. **Journal of Hydrology**, [S. l.], v. 327, n. 3–4, p. 399–410, 2006. Disponível em: <https://doi.org/10.1016/j.jhydrol.2005.11.041>

KIDD, C. Satellite rainfall climatology: A review. **International Journal of Climatology**, [S. l.], v. 21, n. 9, p. 1041–1066, 2001. Disponível em: <https://doi.org/10.1002/joc.635>

KIDD, C.; HUFFMAN, G. Global precipitation measurement. **Meteorological Applications**, [S. l.], v. 18, n. 3, p. 334–353, 2011. Disponível em: <https://doi.org/10.1002/met.284>

KIM, T. W.; AHN, H. Spatial rainfall model using a pattern classifier for estimating missing daily rainfall data. **Stochastic Environmental Research and Risk Assessment**, [S. l.], v. 23, n. 3, p. 367–376, 2009. Disponível em: <https://doi.org/10.1007/s00477-008-0223-9>

LO PRESTI, R.; BARCA, E.; PASSARELLA, G. A methodology for treating missing data applied to daily rainfall data in the Candelaro River Basin (Italy). **Environmental Monitoring and Assessment**, [S. l.], v. 160, n. 1–4, p. 1–22, 2010. Disponível em: <https://doi.org/10.1007/s10661-008-0653-3>

LY, S.; DEGRÉ, A.; CHARLES, C. Different methods for spatial interpolation of rainfall data for operational hydrology and hydrological modeling at watershed scale. A review. **Biotechnology, Agronomy and Society and Environment**, [S. l.], v. 17, n. 2, p. 392–406, 2013. Disponível em: <https://doi.org/10.6084/m9.figshare.1225842.v1>

MAIDMENT, R. I. *et al.* A new, long-term daily satellite-based rainfall dataset for operational monitoring in Africa. **Scientific Data**, [S. l.], v. 4, p. 1–17, 2017. Disponível em: <https://doi.org/10.1038/sdata.2017.63>

MISHRA, A. K. Effect of rain gauge density over the accuracy of rainfall: A case study over Bangalore, India. **SpringerPlus**, [S. l.], v. 2, n. 1, p. 1–7, 2013. Disponível em: <https://doi.org/10.1186/2193-1801-2-311>

MWALE, F. D.; ADELOYE, A. J.; RUSTUM, R. Infilling of missing rainfall and streamflow data in the Shire River basin, Malawi - A self organizing map approach. **Physics and Chemistry of the Earth**, [S. l.], v. 50–52, p. 34–43, 2012. Disponível em: <https://doi.org/10.1016/j.pce.2012.09.006>

RUELLAND, D. *et al.* Sensitivity of a lumped and semi-distributed hydrological model to several methods of rainfall interpolation on a large basin in West Africa. **Journal of Hydrology**, [S. l.], v. 361, n. 1–2, p. 96–117, 2008. Disponível em: <https://doi.org/10.1016/j.jhydrol.2008.07.049>

SATGÉ, F. *et al.* Consistency of satellite-based precipitation products in space and over time compared with gauge observations and snow- hydrological modelling in the Lake Titicaca region. **Hydrology and Earth System Sciences**, [S. l.], v. 23, n. 1, p. 595–619, 2019. Disponível em: <https://doi.org/10.5194/hess-23-595-2019>

SIMOLO, C. *et al.* Improving estimation of missing values in daily precipitation series by a probability density function-preserving approach. **International Journal of Climatology**, [S. l.], v. 30, n. 10, p. 1564–1576, 2010. Disponível em: <https://doi.org/10.1002/joc.1992>

STRAUCH, M. *et al.* Using precipitation data ensemble for uncertainty analysis in SWAT streamflow simulation. **Journal of Hydrology**, [S. l.], v. 414–415, p. 413–424, 2012. Disponível em: <https://doi.org/10.1016/j.jhydrol.2011.11.014>

THIEMIG, V. *et al.* Validation of satellite-based precipitation products over sparsely Gauged

African River basins. **Journal of Hydrometeorology**, [S. l.], v. 13, n. 6, p. 1760–1783, 2012. Disponível em: <https://doi.org/10.1175/JHM-D-12-032.1>

TIECHER, T. Fingerprinting Sediment Sources in Agricultural Catchments. [S. l.], p. 307, 2015.

TUO, Y. *et al.* Evaluation of precipitation input for SWAT modeling in Alpine catchment: A case study in the Adige river basin (Italy). **Science of the Total Environment**, [S. l.], v. 573, p. 66–82, 2016. Disponível em: <https://doi.org/10.1016/j.scitotenv.2016.08.034>

USDC-NOAA. **Precipitation-Frequency Atlas of the United States**. [S. l.]: U.S Department of Commerce, 2013. *E-book*.

VICENTE-SERRANO, S. M. *et al.* A complete daily precipitation database for northeast Spain: Reconstruction, quality control, and homogeneity. **International Journal of Climatology**, [S. l.], v. 30, n. 8, p. 1146–1163, 2010. Disponível em: <https://doi.org/10.1002/joc.1850>

VILLAZÓN, M. F.; WILLEMS, P. Filling gaps and Daily Disaccumulation of Precipitation Data for Rainfall-runoff model. **Area**, [S. l.], n. May, p. 1–9, 2010. Disponível em: [http://balwois.com/balwois/administration/full\\_paper/ffp-1381.pdf](http://balwois.com/balwois/administration/full_paper/ffp-1381.pdf)

WAGNER, P. D. *et al.* Comparison and evaluation of spatial interpolation schemes for daily rainfall in data scarce regions. **Journal of Hydrology**, [S. l.], v. 464–465, p. 388–400, 2012. Disponível em: <https://doi.org/10.1016/j.jhydrol.2012.07.026>

WMO. **Guide to Hydrological Practices**. Sixth ed. Geneva: World Meteorological Organization, 2008. v. *IE-book*. Disponível em: <https://doi.org/10.1080/02626667.2011.546602>

WOLLMANN, C. A.; GALVANI, E. Caracterização climática regional do Rio Grande do Sul: dos estudos estáticos ao entendimento da gênese. **Revista Brasileira de Climatologia**, [S. l.], v. 11, 2012.

WU, Z. *et al.* Hydrologic evaluation of Multi-Source satellite precipitation products for the Upper Huaihe River Basin, China. **Remote Sensing**, [S. l.], v. 10, n. 6, 2018. Disponível em: <https://doi.org/10.3390/rs10060840>

XU, Z. *et al.* Evaluating the accuracy of MSWEP V2.1 and its performance for drought monitoring over mainland China. **Atmospheric Research**, [S. l.], v. 226, p. 17–31, 2019. Disponível em: <https://doi.org/10.1016/j.atmosres.2019.04.008>

YANG, X. *et al.* Spatial Interpolation of Daily Rainfall Data for Local Climate Impact Assessment over Greater Sydney Region. **Advances in Meteorology**, [S. l.], v. 2015, 2015. Disponível em: <https://doi.org/10.1155/2015/563629>

YANTO; LIVNEH, B.; RAJAGOPALAN, B. Development of a gridded meteorological dataset over Java island, Indonesia 1985-2014. **Scientific Data**, [S. l.], v. 4, p. 1–10, 2017. Disponível em: <https://doi.org/10.1038/sdata.2017.72>

YUAN, F. *et al.* Assessment of GPM and TRMM multi-satellite precipitation products in streamflow simulations in a data sparse mountainous watershed in Myanmar. **Remote Sensing**, [S. l.], v. 9, n. 3, 2017. Disponível em: <https://doi.org/10.3390/rs9030302>



### **3 CAPÍTULO II- ASSESSMENT OF MSWEP V.2 PRECIPITATION PRODUCT IN A DATA-SCARCE WATERSHED AND ITS UNCERTAINTY IN STREAMFLOW SIMULATION ON SWAT**

#### **ABSTRACT**

In Brazil, as other developing countries, ground-based rainfall network is scarce, with low density, and with inherent problems of short time-series and gaps. These recurring problems effects directly the use of distributed hydrological models, once that a dense and well spatialized rainfall network is essential to diminish model uncertainties. Currently, satellite-based precipitation products provided an increasingly alternative source of precipitation data, some of them with global coverage, sub-daily time step, and with high spatial resolution. Accordingly, this paper focused on the use of MSWEP v.2 precipitation product as an alternative source of data to infill precipitation gaps in a mid-size Brazilian watershed, and assessed the outcomes of this approach when simulated in the hydrological model SWAT. Daily SWAT model performance for 2002 to 2016 timeframe was analyzed thought three different precipitation scenarios: 1) the available ground-based rainfall network, scenario called RG; 2) MSWEP v.2 dataset only, scenario called MSWEP and 3) a merging scenario where the gaps of scenario RG were infilled with scenario MSWEP, scenario called RG\_MSWEP. The alternative scenarios MSWEP, and RG\_MSWEP resulted in similar model performance, even presenting less uncertainties in the model prediction, in comparison to RG. Therefore, the MSWEP v.2 product can be a valuable data source for hydrological modeling, especially in developing countries or at remote locations because, allowing to reduce the uncertainties related to dubious rainfall data usage in water resources assessment.

#### **3.1 INTRODUCTION**

Ground-based rain gauges provide the most direct and reliable dataset to acquired precipitation estimates (AWANGE; HU; KHAKI, 2019; LIU *et al.*, 2019). The ability of a rain gauge to accurately represent the spatiotemporal variability, intensity, type, and occurrence of a precipitation event over a region is a challenging task (HOU *et al.*, 2014; STRAUCH *et al.*, 2012). These uncertainties effect the accuracy of a hydrological models prediction, once precipitation is one of the driven forces of the hydrological cycle and main input source in models (SIRISENA *et al.*, 2018; TANG *et al.*, 2019).

Problems related to precipitation dataset becomes even more pronounced in developing countries (HUGHES, 2006; WORQLUL *et al.*, 2018), where rain-gauge records commonly presenting some form of deficiency with discontinuities, inadequate length, dubious quality and presence of gaps are frequently (MWALE; ADELOYE; RUSTUM, 2012). In Brazil, for instance, the National Water Agency (ANA) managed 2,722 rain gauges in 2016 (ANA, 2017), leading to a national average network density of one rain gauge every 3,128 km<sup>2</sup>. This density is way below the minimum of one rain gauge every 900 km<sup>2</sup> recommended by World Meteorological Organization (WMO, 2008), regardless the local physiography. Furthermore, there is an uneven distribution of these rain gauges among the Brazilian states, which may be explained by accessibility issues and high costs involved (ANA, 2017). Additionally, several ANA rain gauges have rainfall discontinuities and short-term time series, with high numbers of gaps. Watershed modelers in developing countries are forced to cope with all these inherent problems when working with water resources management (STRAUCH *et al.*, 2012).

A worldwide agro-hydrological model used for hydrological simulations, land use and climate changes, contaminant transport, soil erosion and water resources management practices is the Soil and Water Assessment Tool (SWAT) (ARNOLD *et al.*, 1998). The SWAT model has an open source, and has been proven to be an effective tool for several applications and studies, ranging from catchment to continental scales (ABBASPOUR *et al.*, 2015). Even with an open source and several applications, SWAT model requires a more complex input dataset in comparison to a lumped model, once it is a semi-distributed model (TANG *et al.*, 2019), and this become an additional difficulty for its usage mainly in developing countries due to lack of adequate and detailed data required (BRESSIANI *et al.*, 2015a, 2015b; HUGHES, 2006; LI *et al.*, 2018; WORQLUL *et al.*, 2018).

In Brazil, modelers have to cope with the lack of available data, deficiency of easily accessible datasets and problems related to the processing of data required for SWAT application, contrasting with developed countries, where normally there is a wide monitoring network and easy access to the databases (BRESSIANI *et al.*, 2015b). In a survey for studies performed with SWAT in Brazil, only 66% of 102 SWAT studies had calibration results, and from these just 23% had validation results (BRESSIANI *et al.*, 2015b). Furthermore, 70% of the SWAT studies that reported calibration and validation results used a time series shorter than 5 years for each period (BRESSIANI *et al.*, 2015b). Thus, inherent problems of SWAT use in Brazil must be addressed to find possible alternatives to overcome the problems (BRESSIANI *et al.*, 2015b).

Satellite-based precipitation products are capable to overcome these limitations related to ground-based rainfall network being a good alternative as a data source, especially in data-scarce or ungauged regions, playing an important role in hydrological and meteorological studies, because of their large-scale coverage and high spatiotemporal resolution (MONDAL; LAKSHMI; HASHEMI, 2018; WU *et al.*, 2018). The multisource precipitation products has been used as a valuable source of rainfall measurements to calibrate and validate hydrological models against ground-based rain gauges (AJAJ; MISHRA; KHAN, 2019; LAKEW; MOGES; ASFAW, 2020; SENENT-APARICIO *et al.*, 2018; TANG *et al.*, 2019).

The Multi-Source Weighted-Ensemble Precipitation(MSWEP v.2; Beck *et al.*, 2018) has some unique aspects, among the satellite precipitation products available, as a long-term precipitation dataset (1979-2016) with a global coverage of 3-hours and 0.1° resolution. MSWEP v.2 takes advantage of satellite, rain gauges and re-analysis data, and to date is the satellite precipitation product with the best spatial and temporal resolution in South Brazil.

Although MSWEP v.2 has been widely used to validate and compare its dataset against others satellite products and ground-based observation (BECK *et al.*, 2019; LIU *et al.*, 2019; XU *et al.*, 2019), information about the use of MSWEP v.2 as an alternative source for precipitation dataset in hydrological models still incipient. MSWEP v.2 presented best precipitation volume estimation when compared to other three global precipitation datasets (TRMM 3B43, CFSR and PERSIANN) against ground-based gauge using multiple monthly water balance models (SENENT-APARICIO *et al.*, 2018). In the same way, MSWEP v.2 could capture better precipitation intensity variability in different climatic regions when was compared other three satellite precipitation products (AgMERRA, PERSIANN-CDR and TMPA), where the satellite products were evaluated against ground-based rainfall observation and analyzed its performance regarding daily streamflow prediction in SWAT (TANG *et al.*, 2019). Furthermore, MSWEP estimates allowed better daily streamflow prediction than the other three products (TANG *et al.*, 2019). To date, no work assessing MSWEP v.2 as primarily source of precipitation estimation in streamflow prediction using SWAT model in Brazil was found. Generally, works assessing streamflow prediction on SWAT by using satellite rainfall are based on TRMM 3B42 product (CREMONINI; BRIGHENTI; BONUMÁ, 2014; STRAUCH *et al.*, 2012; TOBIN; BENNETT, 2014).

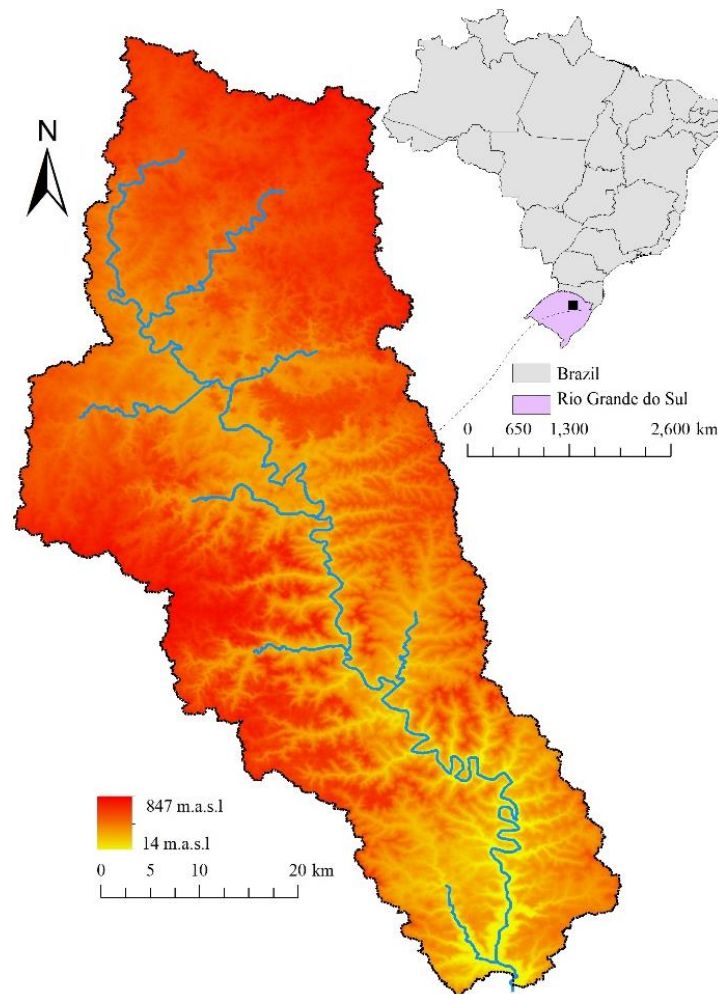
In view of the above issues addressed, this study aimed to assess different alternatives for improving precipitation estimates in a mid-size watershed in South Brazil by using MSWEP v.2 dataset, to overcome rain gauge dataset limitations, and how it performs in hydrological application in SWAT model.

## 3.2 MATERIALS AND METHODOLOGY

### 3.2.1 Study area

The Guaporé River Watershed is located in Southern Brazil at 28°14'26.9" S -28°58'02"S Latitude and 51°54'59" W -52°22'55.71" W Longitude (Figure 8). The study area covers about 2430 km<sup>2</sup> and the elevation of the catchment ranges from 14m up to 847m.a.s.l, with an average elevation of 550m.a.s.l. According to the Köppen's classification, the regional climate is humid subtropical, type Cfa, which is characterized by hot and humid summers, and cold to mild winters. The mean temperature during the coldest month is around 12°, and 22° during the warmest month (WOLLMANN; GALVANI, 2012). Annual rainfall is evenly distributed throughout the year, ranging from 1400 to 2000 mm (TIECHER, 2015). Agriculture is the dominant land use type in the catchment, and covers approximately 55% of the watershed area. Typically, three crops are yearly harvested with rotation, being soybeans, corn, wheat, ryegrass and tobacco, the main crops farmed. The vegetation comprises 44% of the catchment, in which 31% represents mixed forest and 13% represents range grasses, urban areas occupy only about 1% (these percentages were acquired from a supervised classification of the image used as input for land cover/use in SWAT, more details in section 3.2.2.1).

Figure 8- Location and elevation of Guaporé watershed along with hydrographic network.



Source: the author.

### 3.2.2 SWAT model

SWAT is a world-wide known hydrological model developed by the Agricultural Research Service (ARS) of the United States Department of Agriculture (USDA). SWAT is a semi-distributed, time-continuous, watershed scale simulation model, designed to assess and predict the impact of the alternative management practices regarding agriculture and water resource (ARNOLD *et al.*, 2012). The model can be used to simulate a variety of process, including hydrology, vegetation growth, erosion, nutrients, pesticides, non-point source of agricultural pollution and agricultural management. SWAT defines the watershed and its sub-basins based on topography, and subsequently, the sub-basins are divided into Hydrologic Response Units (HRU), which are unique combinations of the same land use, soil type, and slope.

The hydrological cycle is calculated based on the water balance equation (ARNOLD *et al.*, 1998), as given by Equation 1 .

$$SW_t = SW_0 + \sum_{i=1}^t (R - Q - ET - P - QR) \quad (1)$$

Where,  $SW_t$  is the soil water content at time  $t$ ,  $SW_0$  is the initial soil water content, and  $R$ ,  $Q$ ,  $ET$ ,  $P$  and  $QR$  are precipitation, runoff, evapotranspiration, percolation, and return flow respectively. All units are in mm and at the time  $t$ .

The main channel system receives water from the soil profile through lateral flow, return flow and surface runoff. All these components are calculated for each HRU (land phase), and then, the loading of water is transferred to the main channel and routed through the mainstream associated with each sub-basin. The shallow aquifer is recharged from soil profile percolation, while a small fraction of the total recharge of shallow aquifer can percolate to the deep aquifer (ARNOLD *et al.*, 1998).

To calculate surface runoff and infiltration the Soil Conservation Service (SCS) Curve Number (CN) or the Green & Ampt method are available. When the Curve Number method is being used, daily curve number calculation can be done as a function of soil moisture, or as function of plant evapotranspiration. SWAT model is able to calculate evapotranspiration with Hargreaves, Priestley-Taylor or Penman-Monteith methods (NEITSCH *et al.*, 2011). The complexity and required weather data vary with the chosen method.

Among the aforementioned methods, in this study based on the available climate data, Penman-Monteith was used to calculate potential evapotranspiration. The CN method was chosen to simulate surface runoff and infiltration because sub-daily precipitation data are not available for Guaporé watershed in order to use Green & Ampt method. The method of daily curve number calculation as function of plant evapotranspiration was chosen, once its use was documented to outcome the best performance in another study in a Brazilian watershed (BRESSIANI *et al.*, 2015a).

### 3.2.2.1 Model setup and Model inputs

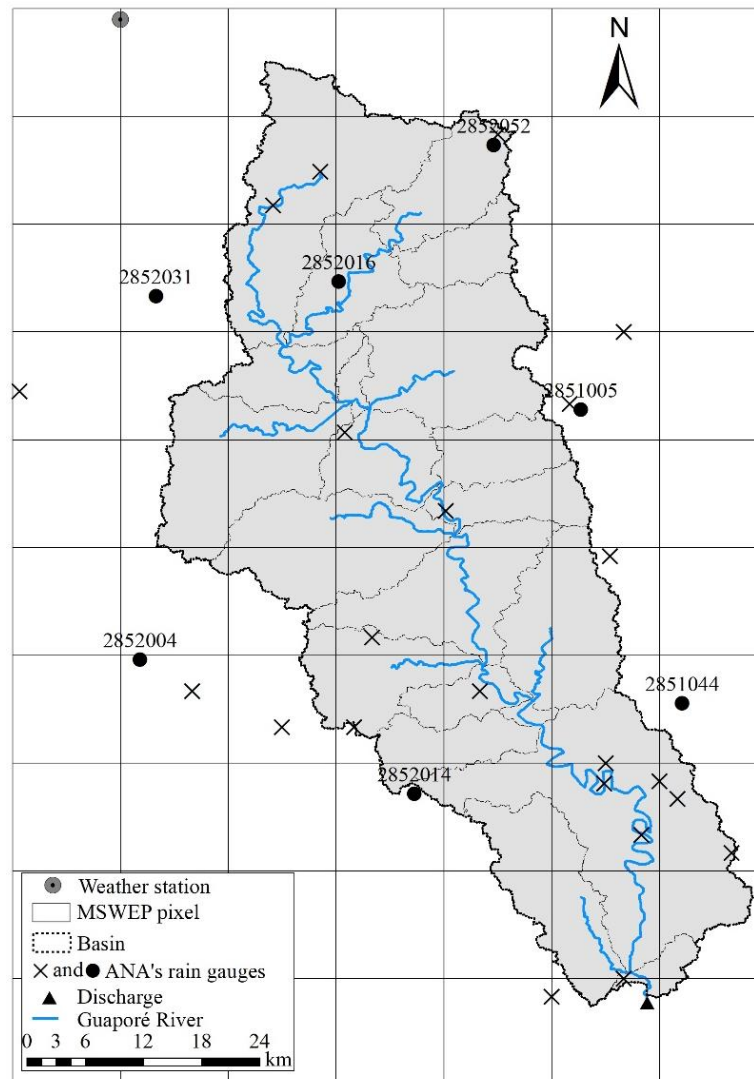
This study was carried out using the ArcSWAT 2012 version coupled with ArcGIS interface (ESRI, 2017). A digital elevation model (DEM) at 30 m resolution acquired from

United States Geological Survey (USGS) National Elevation Dataset (NED) was used to delineate Guaporé watershed. The watershed was divided into 21 sub-basins, according to the threshold drainage area of 6000 ha (Figure 9). All the meteorological inputs (solar radiation, air temperature, wind speed and relative humidity) were obtained from Meteorological Database for Teaching and Research-BDMEP (INMET, 2017) .

The land use/land cover (LULC) image, at 30 m resolution, was acquired from USGS earth explorer service (Satellite Landsat 4-5, 2015), and then classified into five classes of LULC (5 bands) with a supervised classification process , the image used is from October 2014 (Figure 10 a) . Two different management plans were designed to agriculture according to the crop rotation, harvest dates, winter or summer crops, and most farmed crop in each city/sub-basin. A data survey from Brazilian Institute for Geography and Statistics (IBGE, 2017) for the 22 cities, completely or partially within Guaporé watershed, was done to assess the most farmed crops. In the high lands, in the upstream areas, 14 sub-basins were defined to receive a crop rotation with wheat, soybeans and corn, while in the low lands, 7 downstream sub-basins were selected to have crop rotation with corn, tobacco and ryegrass.

The minimum base temperature for plant growth ( $T_{BASE}$ ) in farmed crops, initial leaf area index ( $LAI_{INIT}$ ), initial dry weight biomass ( $BIO_{INIT}$ ) and total number of heat units needed to the plants maturity ( $PHU_{PLT}$ ) for evergreen forest, were adjusted to the humid subtropical conditions of our study area. The potential heat units for crops were calculated and added to the management input file, once the plants grow is guided by the cumulative days exceeding the minimum (base) temperature. Several references from the study area, or areas with similar characteristics were used to adjust all these parameters.

Figure 9- Location of hydrological, meteorological and precipitation stations, MSWEP v.2 grid and sub-basin division. ANA rain gauges without recordings data are represented by the letter “X”, while the filled circle symbol represents stations with data.

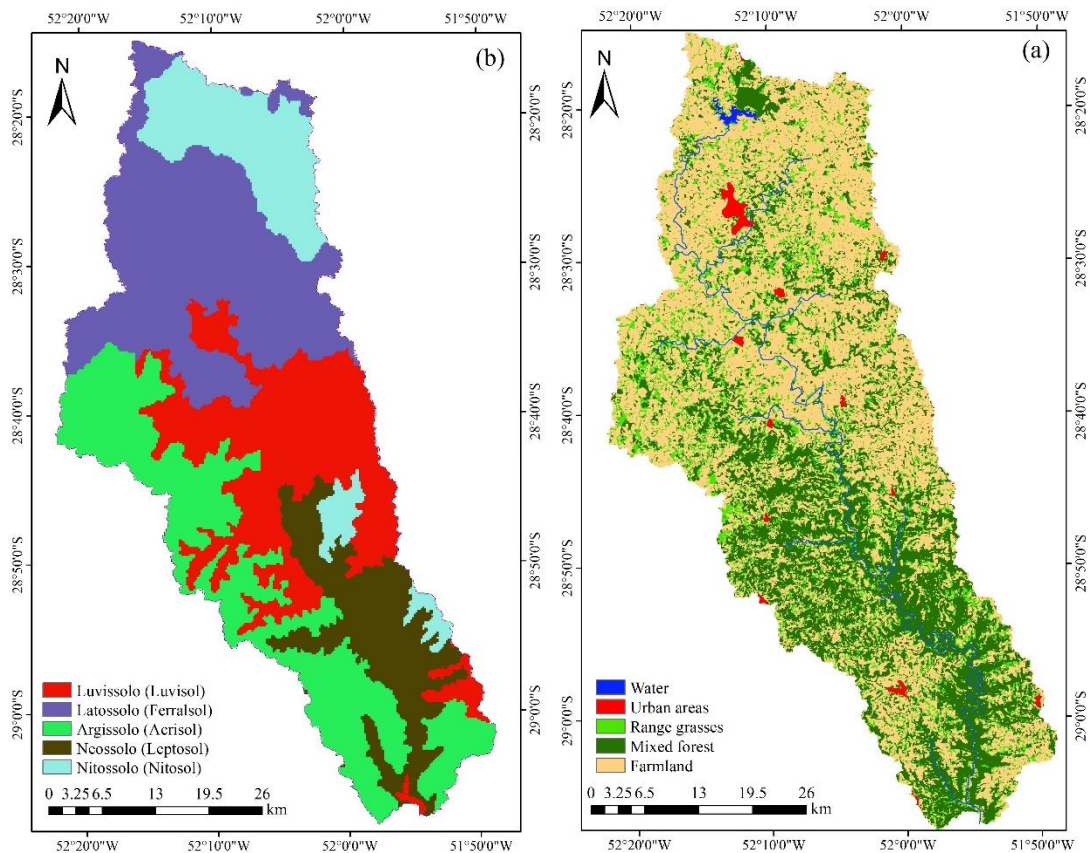


Source: the author.

The soil map was acquired from Santos et al. (2018), with a 1:750.000 scale (Figure 10 b), and the tabular properties (physical soil properties) were obtained according to the procedures described by Rodrigues (2015) and the classes updated according to the Brazilian Soil Classification System (SANTOS et al., 2018). Due to the focus of the analysis in this work, hydrological data as precipitation and streamflow are presented in the following sections.



Figure 10- Guaporé Watershed distribution maps for: (a) Land use/Land cover (LULC) and (b) soil type.



Source: the author.

### 3.2.2.2 Precipitation datasets

Two datasets covering 2002-2016 have been used to generate three alternative precipitation (scenarios) inputs for SWAT model, based on ground-based rain gauge (RG) observations and MSWEP v.2 (MSWEP) dataset, according to:

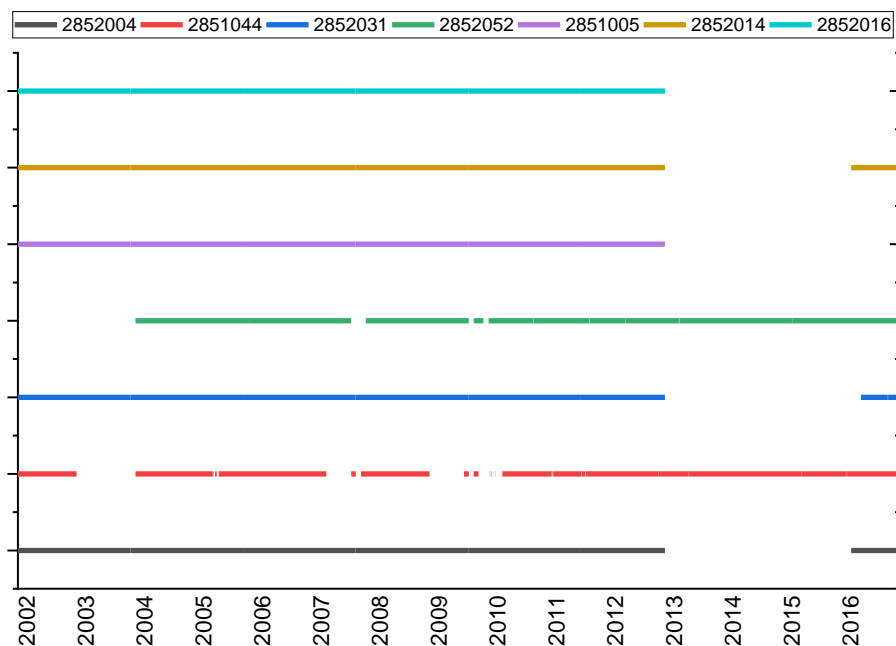
- i) Scenario RG: considers exclusively the ground-based precipitation data from ANA rain gauges;
- ii) Scenario MSWEP: considers exclusively satellite precipitation product from MSWP v.2 (BECK *et al.*, 2018);
- iii) Scenario RG\_MSWEP: corresponds to a merged scenario, where ground-based precipitation data from ANA rain gauges are preferentially used, and MSWEP v.2 is directly used as source for infilling missing daily rain gauge data.

#### i) Scenario RG

Daily precipitation within or nearby the watershed was selected from 29 rain gauges available in the ANA database (Figure 9). At first sight, when considering only the rain gauges within the watershed, the network obtained seems to be suitable, presenting a density of 1 rain gauge every 151 km<sup>2</sup>, which is an acceptable density according to the WMO recommendations. However, from the 16 rain gauges within the watershed, only 2 have recorded data in the study timeframe, leading to a rain gauge density of 1 station every 1215 km<sup>2</sup>, which represents a lower density than the less restrictive one recommended by WMO. In summary, considering the rain gauges within the watershed and on its surroundings, only 7 rain gauges have recorded data in the study timeframe (from 2002 up to 2016), which provided precipitation data for Scenario RG (Figure 12-a).

Besides the low rain gauge density found within the watershed and its surroundings, none of the seven rain gauges with data have an uninterrupted time series throughout the selected timeframe from 2002 up to 2016 (Figure 11). This analysis highlights the fragility of the ground-based monitoring network, common in developing countries (HUGHES, 2006; MAIDMENT *et al.*, 2017), and the need of alternative strategies to overcome these difficulties. Further, the low rain gauge density within the watershed reinforces the need of a previous interpolation procedure to cope with the limitations imposed by the centroid method used in SWAT.

Figure 11- Temporal analysis of ANA's rain gauge daily precipitation on the timeframe of this study and its code.



Source: the author.

The SWAT model, as default, only uses data from the rain gauge nearest to the centroid (centroid method) of each sub-basin as source for daily precipitation (NEITSCH *et al.*, 2011) during the simulations. However, the centroid method is insufficient to calculate a sub-basin's mean areal precipitation (MAP) (XUE *et al.*, 2019), especially in a daily time step. The use of interpolation methods to estimate the MAP, for instance, the direct weighted averaging, the surface-fitting, the inverse distance weighting (IDW) or the Thiessen method, has proved to decreased the uncertainties in streamflow simulation on SWAT, allowing properly representation of the rainfall spatial variability over the watershed (ANDERSSON *et al.*, 2012; CHO *et al.*, 2013; ZEIGER; HUBBART, 2017). In order to overcome this limitation, the inverse distance weighting (IDW) was used to interpolate each precipitation scenarios (RG, MWEP and RG\_MSWEF) in a daily time step, the mean areal precipitation (MAP) for each sub-basin was calculated and designated to its centroid. Hence each scenario has a total of 21 precipitation input in order to mislead SWAT model.

## **ii) Scenario MSWEP**

The MSWEP dataset is the first fully global precipitation dataset with a spatial resolution of 0.1°, with temporal resolution of 3 hours and available for the period of 1979-2016 (BECK *et al.*, 2018). MSWEP uses different types of data sources to provide reliable precipitation estimation, including 76747 rain gauges, four satellite (CMORPH, GridSat, GSMaP, and TMPA 3B43RT), and two reanalysis datasets (ERA-Interim and JRA-55), endorsing its strength as a satellite-based precipitation product. MSWEP v.2 grid have 38 cells over Guaporé watershed used as source for precipitation data (Figure 9), which means that Scenario MSWEP had the equivalent to 38 “remote” stations (Figure 12-b)

The suitability of MSWEP v.2 dataset over our study area was previously assessed on a prior work (CELANTE *et al.*, 2020<sup>1</sup>) considering the period of 1979-2016. The MSWEP v.2 data was capable to represent the temporal and spatial variability of daily and monthly rainfall in the watershed, proving to be an interesting data source for infill missing precipitation data, especially where rain gauges were scarce (CELANTE *et al.*, 2020).

## **iii) Scenario RG\_MSWEF**

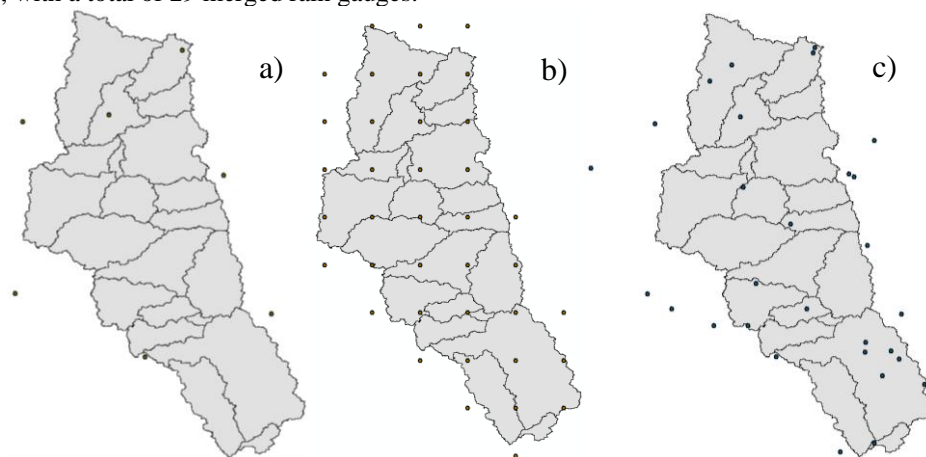
---

<sup>1</sup> CELANTE *et al.*, 2020 refere-se ao primeiro artigo apresentado na dissertação, que será submetido para publicação em revista a ser definida.

This last scenario corresponds to a merging scenario using the available ANA rain gauge dataset and MSWEP v.2 dataset. During<sup>2</sup> IDW interpolation process, when an ANA rain gauge presented a missing daily value, the nearest MSWEP v.2 grid cell value was used to fill the missing value. Also, ANA rain gauges without any recorded data in the timeframe of this study were fully replaced by the nearest MSWEP v.2 grid cell value.

Therefore, instead of a network with only 7 rain gauges (Scenario RG), this scenario has a rain gauge density of 29 merged stations (RG\_MSWEF) providing precipitation data (Figure 12-c). This scenario potentially has the advantage of providing a more realistic representation of rainfall temporal/spatial variability in the whole watershed.

Figure 12- Spatial distribution of the rain gauges for scenario: a) RG only ANA gauges records, with no infilling gaps, with a total of 7 rain gauges used. b) MSWEP dataset accounting with 38 grid cells, and c) RG\_MSWEF, all ANA rain gauges from Guaporé watershed and its surroundings were used, along with MSWEP dataset to infill the gaps, with a total of 29 merged rain gauges.



Source: the author.

### 3.2.2.3 Mean areal precipitation error

A temporal analysis of the watershed MAP error of the whole watershed was accomplished according to the methodology presented by Tian et al. (2009) for our calibration period (2002-2011), in order to have a better understanding of the uncertainties imposed by the spatialized precipitation datasets. This methodology assesses the error of satellite-based product against ground-based rain gauge data by using an error decomposition scheme.

This methodology is based on the decomposition of the error component in order to track the error source. The total error, also called total bias (E), is divided into three independent

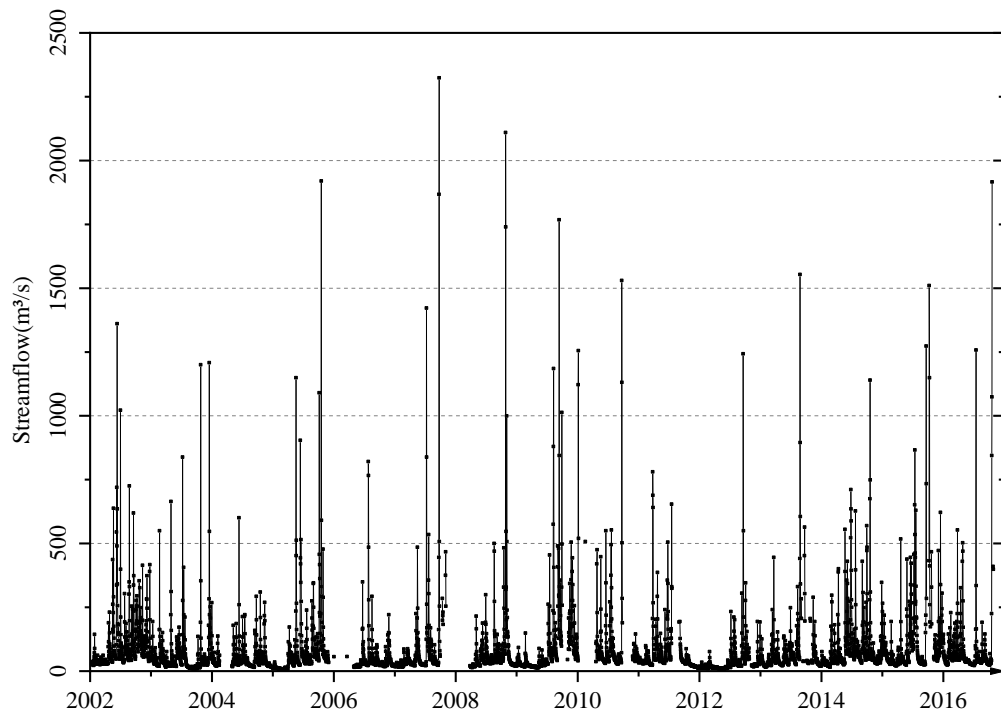
parts: i) the hit error, or hit bias (H), which corresponds to the difference of detected rain rate between the reference data (ANA rain gauges) and the evaluated data (MSWEP v.2); ii) the missed precipitation (-M) is an event where the reference data (RG) reported precipitation and the evaluated data (MSWEP and RG\_MSWEP) reported otherwise; and iii) the false precipitation (F) represents an precipitation event reported by the evaluated dataset (MSWEP and RG\_MSWEP) while the reference dataset (RG) reported no precipitation. Thus, the total bias may be written as  $E = H - M + F$  (TIAN *et al.*, 2009). The threshold used to determinate rain/no-rain was 1 mm/day, and a 31-day running average was applied to each error time series to avoid visual cluttering.

An intensity distribution of daily precipitation at different magnitudes was also assessed, in order to evaluate the ability of SPPs to detect low and high amounts of precipitation in comparison to ground-based rain gauge (ALAZZY *et al.*, 2017; TANG *et al.*, 2019).

#### 3.2.2.4 Streamflow data

The daily streamflow data was obtained from Guaporé watershed station (streamflow station # 86580000 of ANA). The streamflow data from 2002 to 2016 had 814 days (~15%) of missing data (Figure 13).

Figure 13- Guapore's streamflow time series (station code: 86580000).

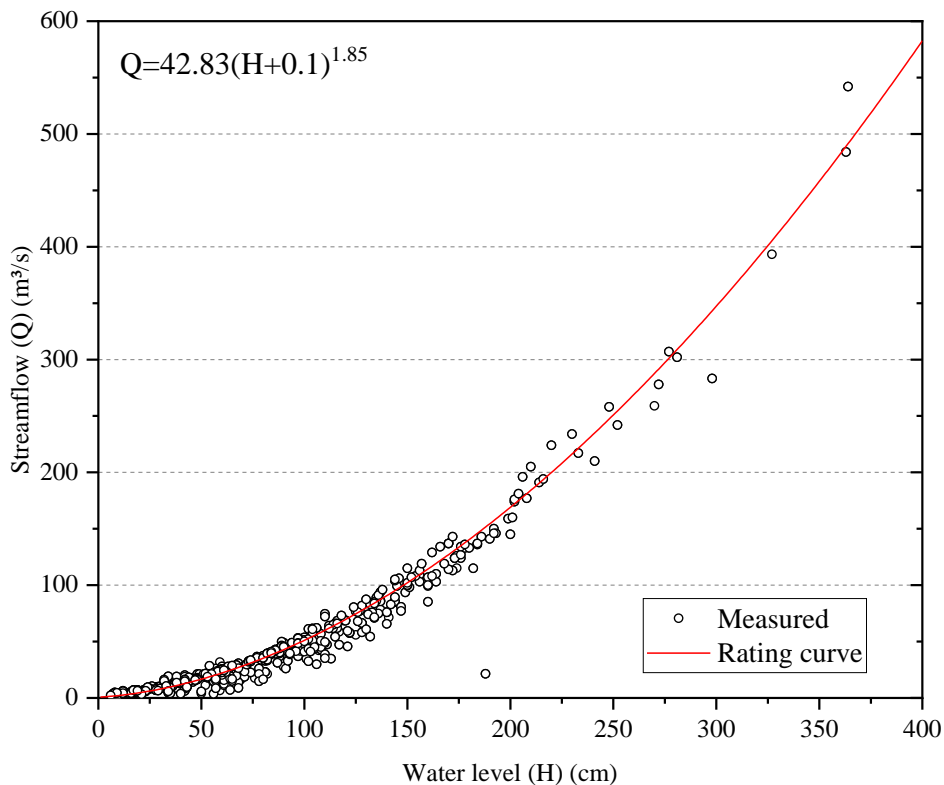


Source: the author.

In addition, when assessing streamflow data, it is important to evaluate the station rating curve used for the specific gauge. It is known that extreme floods are often post-event evaluated under indirect estimates, making the extrapolation of the rating curve a fundamental step in the process of streamflow estimation, being a substantial source of error (LANG *et al.*, 2010).

This streamflow time series had some high streamflow measures, which extrapolate those well-known limits in the upper end of the streamflow rating curve (Figure 14) . The streamflow measurement higher than 550 m<sup>3</sup>/s (Figure 13) is estimated by the extrapolation of the equation adjusted to the stage-discharge data.

Figure 14- Guaporé's rating curve



Source: the author.

Thus, the three precipitation scenarios were calibrated and validated, due to these uncertainties on high streamflow, according to the following streamflow data set: 1) full streamflow time series data ( $Q_t$ ), which includes high streamflow data comprised in the extrapolation of the rating curve; and 2) partial streamflow time series data, which contains exclusively streamflow observations below 500 m<sup>3</sup>/s ( $Q < 500$ ), leading to a streamflow time series 90 days shorter than  $Q_t$  (removed flows represent 1.93% of  $Q_t$ ). Therefore, six different

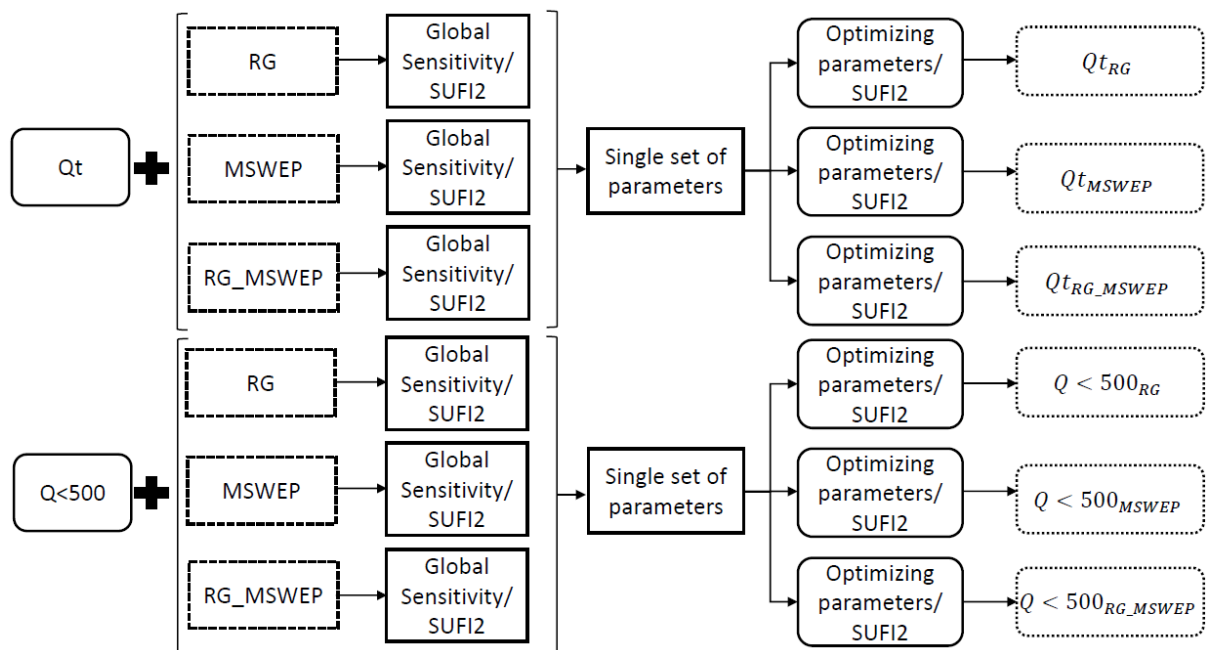
scenarios were assessed, and the approaches for the sensitivity analysis, model calibration and validation are present below.

### 3.2.2.5 Sensitivity analysis, model calibration and validation

Daily model calibration and validation was carried out from 2002-2011 and 2012-2016 time periods, respectively, and four years (1998-2001) were used as warm-up period in order to initialize the model state variables. With all scenarios set, the Sequential Uncertainty Fitting-2 (SUFI-2) algorithm, available in SWAT-CUP program (ABBASPOUR *et al.*, 2007a), was used to perform model sensitivity analysis, automatic model calibration and validation. The coefficient Nash-Sutcliffe (NSE) was used as the objective function to optimize model calibration.

After the construction of each set of precipitation input (RG, MSWEP, and RG\_MSWEP), 6 different SWAT-CUP projects were created in order to proceed with the assessment (Figure 15). Following suggestions from other studies (ABBASPOUR *et al.*, 2007a, 2007b), a set of 21 flow parameters were selected to perform sensitivity analysis, in an initial interaction of 1000 simulations.

Figure 15- Flowchart of how the methodology of this study was conducted, with model sensitivity analysis and calibration routines.



Source: the author.

The most sensitive parameters were identified by the global sensitivity analysis by using p-value, at a significance level of 5% (Table 5). Then, a set of merging parameters was defined according to each parameter that was sensitive, regarding the streamflow being calibrated ( $Q_t$  or  $Q < 500$ ). Thus, after the sensitivity analysis for RG, MSWEP, and RG\_MSWEP scenarios under  $Q_t$  conditions, a single set of the most sensitive parameters among the three precipitation scenarios was used to proceed with calibration.

A total of 10 parameters were sensitive to RG, MSWEP and RG\_MSWEP when calibrating with  $Q_t$ , and 11 parameters when calibrating with  $Q < 500$  (Table 5). The final range and fitted value of the calibrated parameters will be presented normalized, with regards to the final minimum and maximum range and fitted value for the best simulation of each precipitation scenario.

Table 5- Final set of parameters selected to perform calibration for  $Q_t$  and  $Q < 500$ , where sensitive parameters in each approach ( $Q_t$  or  $Q < 500$ ) are highlighted with an “X”.

Parameter	Description	Initial range	$Q_t$	$Q < 500$
r_CN2	SCS runoff curve number	-0.4 - 0.4	X	X
v_ALPHA_BF	Base-flow recession factor (days)	0 - 1	X	X
v_SHALLST	Initial depth of water in the shallow aquifer	0 - 4000	X	X
v_GW_REVAP	Groundwater revap coefficient	0.02 – 0.2		X
v_GWQMN	Threshold water depth in shallow aquifer for return flow to occur	0 - 4000		X
v_CH_K2	Effective soil hydraulic conductivity (mm/hr)	0.01- 400	X	X
v_CH_N2	Manning’s $n$ value for the main channel	0 - 0.3	X	X
v_ESCO	Soil evaporation compensation factor	0 - 1	X	X
r_SOL_AWC()	Average available soil water content	-0.5 – 0.5	X	X
r_SOL_K()	Saturated hydraulic conductivity (mm/day)	-0.5 – 0.5	X	X
v_CNCOEF	Plant ET curve number coefficient	0.5 – 2	X	
v_ALPHA_BNK	Baseflow alpha factor for bank storage	0 – 1	X	X

Source: the author.

Each SWAT-CUP project (RG $_{Q_t}$ , MSWEP $_{Q_t}$ , RG\_MSWEP $_{Q_t}$ , RG $_{Q < 500}$ , MSWEP $_{Q < 500}$  and RG\_MSWEP $_{Q < 500}$ ) and the corresponding set of sensitive parameters ( $Q_t$  or  $Q < 500$ ) was calibrated with several iterations, each with 600 simulations, after the parameter sensitivity analysis. After every interaction on the calibration process, the SUFI-2 algorithm suggests a new set of parameter range to be used in the next iteration. Before setting the suggested range to a new iteration, the upper and lower limits were verified, assuring that the new range was



within the physical reality of each parameter, and the SWAT model official documentation (NEITSCH *et al.*, 2011).

The calibration process was considered satisfactory when between two successive iterations no improvement over model performance was reached. Then, the ranges of the calibrated parameters were used in 600 simulations to validate the model.

Model performance for calibration and validation was also assessed by using the coefficient of determination ( $r^2$ ) and the percentage bias (PBIAS). The model performance was evaluated as recommended by Moriasi *et al.* (2015), which classify the performance as very good ( $NSE > 0.8$ ,  $r^2 > 0.85$ ,  $PBIAS < \pm 5$ ), good ( $0.7 < NSE \leq 0.80$ ,  $0.75 < r^2 \leq 0.85$ ,  $\pm 5 \leq PBIAS \leq \pm 10$ ), satisfactory ( $0.5 < NSE \leq 0.7$ ,  $0.60 < r^2 \leq 0.75$ ,  $\pm 10 \leq PBIAS \leq \pm 15$ ) and not satisfactory ( $NSE \leq 0.5$ ,  $r^2 \leq 0.60$ ,  $PBIAS \geq 15$ ). NSE (Equation 2),  $r^2$  (Equation 3) and PBIAS (Equation 4) are calculated as follows:

$$NSE = 1 - \frac{\sum_i (Q_m - Q_s)_i^2}{\sum_i (Q_{m,i} - Q_{m,avg})^2} \quad (2)$$

$$r^2 = \frac{[\sum_{i=1}^n (Q_{m,i} - Q_{m,avg})(Q_{s,i} - Q_{s,avg})]^2}{\sum_{i=1}^n (Q_{m,i} - Q_{m,avg})^2 \sum_{i=1}^n (Q_{s,i} - Q_{s,avg})^2} \quad (3)$$

$$PBIAS = 100 * \frac{\sum_{i=1}^n (Q_m - Q_s)_i}{\sum_{i=1}^n Q_{m,i}} \quad (4)$$

where  $Q_{m,i}$  and  $Q_{s,i}$  are measured and simulated streamflow at each time step  $i$ ;  $Q_{m,avg}$  and  $Q_{s,avg}$  are the mean measured and simulated streamflow, and  $n$  is the number of time steps.

Additionally, the error agreement between observed and simulated streamflow data was evaluated with Equation 5, which prioritizes low flow periods, once NSE and  $r^2$  are more adequate to evaluate high flows (MORIASI *et al.*, 2015). Low errors in the agreement between observed and simulated streamflow are found when  $E_{min}$  is closest to 0.

$$E_{min} = \left[ \frac{1}{n} \sum_{i=1}^n \left( \frac{1}{Q_{m,i}} - \frac{1}{Q_{s,i}} \right)^2 \right]^{1/2} \quad (5)$$

Lastly, the Person correlation coefficient ( $r$ ) between observed and simulated daily streamflow was evaluated, where  $r$  ranges from 0 to 1. Even though if the results for  $r$  are high

(close to 1), the regression gradient (b) and interception (a) needs to be considered to evaluate the correlation because an equation that systematically overestimates or underestimates streamflow along the entire simulation could still result in an  $r$  close to 1 (KRAUSE; BOYLE; BÄSE, 2005). Thus, a good agreement between simulated and observed streamflow is found when an intercept value is close to 0, and a gradient close to 1.

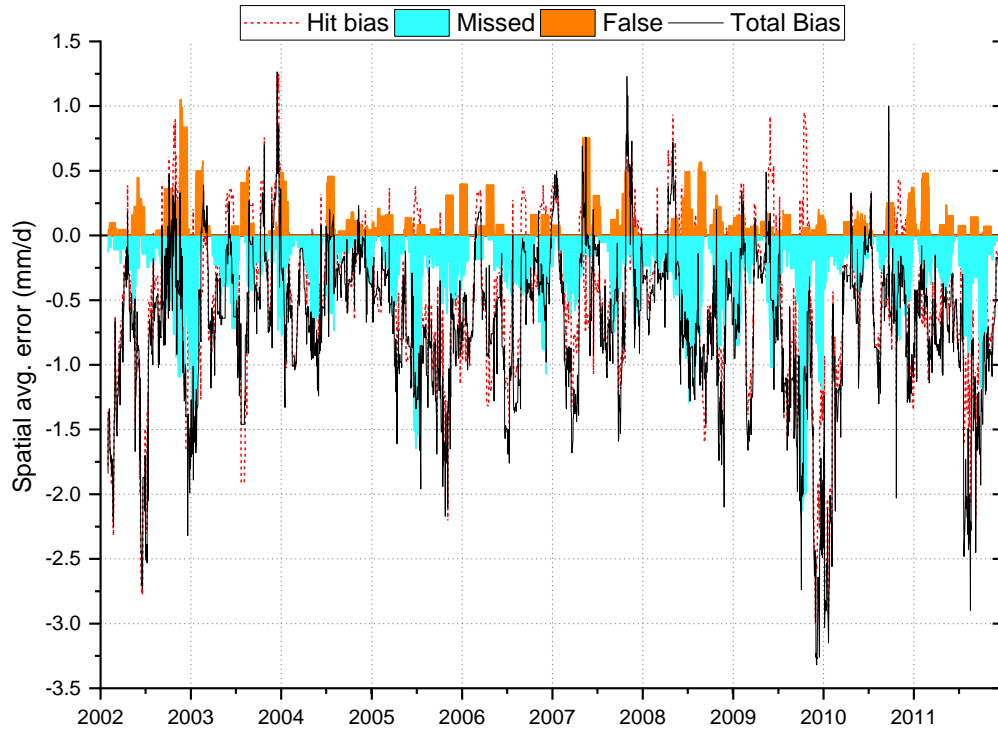
The assessment of the uncertainty in the streamflow model prediction on SWAT-CUP is the 95% probability distribution (95PPU). The 95PPU represents the propagation of the parameters' uncertainty, and results are the model outputs in a stochastic calibration approach, which is given in an envelope of good solutions (Abbaspour et al., 2007). The fitness between the model result (expressed as 95PPU) and observation data used to calibrate the model was quantified with the p-factor and the r-factor. The p-factor is the percentage of observation data bracketed by the 95PPU band, and r-factor is the average width of the 95PPU band. The p-factor ranges from 0 to 1 (1 represents a 100% of the measured data enveloped by the 95PPU band), for streamflow analysis it is recommended p-factor  $>70\%$  , and r-factor close to 1 (ABBASPOUR et al., 2007a).

### 3.3 RESULTS

#### 3.3.1 Mean areal precipitation error

The MSWEP v.2 product has a tendency of underestimate daily precipitation intensity, based on the systematic trend of negative values of total bias component. The total bias (E) is close to hit bias (H) throughout the time series, in general, and both have larger amplitude than missed (M) and false (F) precipitation, indicating that the main error comes from hit bias, thus underestimation of MSWEP v.2 dataset (Figure 16).

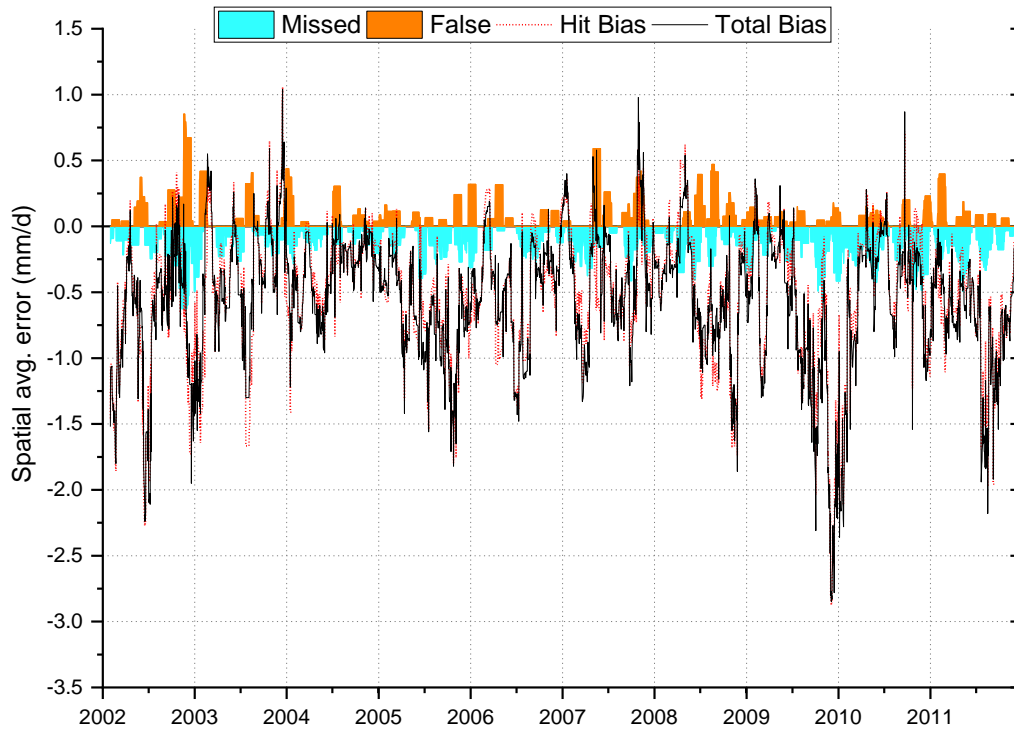
Figure 16- Daily time series of the error components for Scenario RG and Scenario MSWEP.



Source: the author.

The missed precipitation error is decreased when considering the infilling approach (RG\_MSWEF) (Figure 17). Although total bias still has a systematic underestimation pattern, hit bias and total bias are even more similar, and missed precipitation is smaller. Thus, the RG\_MSWEF scenario which combines both datasets resulted in an improvement in precipitation accuracy, with a narrower total bias variability (-2.85 to 1.04), when compared against MSWEF scenario (-3.33 to 1.26).

Figure 17- Daily time series of error components for RG and RG\_MSWEP.

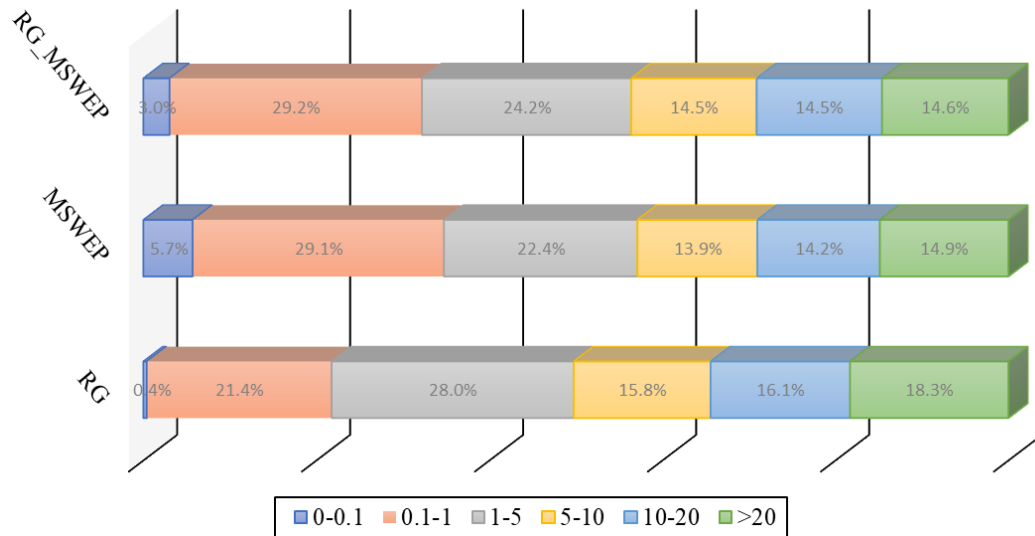


Source: the author.

The same underestimation pattern was found on a previous work, following a different approach (CELANTE *et al.*, 2020) and could be explained by the rainfall characteristics that occurs at Guaporé watershed region. During summer, the intense heat and high humidity levels result in convective rainfall events, commonly characterized by high rainfall intensity in short-term duration, usually an hour or few hours (PEREIRA BRITTO; BARLETTA; MENDONÇA, 2008). Possible, MSWEP v.2 is unable to predict convective rainfall events, once its temporal resolution is longer than the duration of a convective event. Additionally, the SPPs has trend of underestimate precipitation intensity (HEROLD *et al.*, 2016), and this pattern has been specially found for MSWEP products (AWANGE; HU; KHAKI, 2019; XUE *et al.*, 2019).

The underestimation of MSWEP v.2 dataset is highlighted for a frequency distribution of daily precipitation at different magnitudes (Figure 18). The highest precipitation frequency rainfall range for MSWEP and RG\_MSWEP scenarios is 0.1~1 mm/d, while in the reference scenario RG the highest frequency ranges from 1~5 mm/d.

Figure 18- Daily precipitation frequency distribution in different ranges (0~0.1, 0.1~1, 1~5, 5~10, 10~20 and >20 mm/day) for MAP after IDW.



Source: the author.

It is worth remembering that in order to apply the error components methodology (TIAN *et al.*, 2009), the threshold of 1 mm was chosen in order to identify rain/no rain day. The threshold of 1 mm used to identify rain/no rain day is also the up limit of the highest frequency range for MSWEP and RG\_MSWEF (Figure 18). The -M must have been influenced negatively by this threshold, once -M represents a rain event which the ground-based station (RG) reported rain, and the satellite (MSWEP and RG\_MSWEF) missed. This behavior occurred because only 21.8% of precipitation events of RG were below the threshold of 1 mm/d, while 34.8% and 32.2% of rainfall events were below 1 mm/d for MSWEP and RG\_MSWEF, respectively.

### 3.3.2 Model calibration and validation

Model calibration and validation results are following presented in separated topics. The first topic (3.3.2.1) is destined to presents the results related to the model performance during calibration and validation using all streamflow data ( $Q_t$ ) provided by ANA, assessed for RG, MSWEP and RG\_MSWEF precipitation scenarios. Thereafter, calibration and validation results are presented for the same precipitation scenarios by using only the streamflow data with higher reliability ( $Q > 500$ ) (3.3.2.2).

### 3.3.2.1 Model performance under $Q_t$ condition

Model simulation based on ground-based observation (RG) outperformed the scenarios that used MSWEP v.2 dataset (MSWEP and RG\_MSWEP), in both, calibration (2002-2011) and validation period (2012-2016) (Table 6) in regards to model performance. A ‘very good’ fit (Moriassi et al., 2015) based on NSE, between observed daily streamflow and  $Q_{tRG}$  simulation was found, for both calibration and validation periods, followed by  $Q_{tMSWEP}$  and  $Q_{tRG\_MSWEP}$  simulations that despite smaller NSE still are classified as a ‘good’ model performance.

Table 6- Streamflow model calibration and validation statistical indicators performance for  $Q_t$

	NSE	PBIAS (%)	$r^2$	E <sub>min</sub>	Mean streamflow Simulation-Observation (m <sup>3</sup> /s)	Standard deviation Simulation-Observation
Calibration						
RG	0.81	-4.5	0.84	0.02	86.58-82.88	115.07-152.43
MSWEP	0.70	7.8	0.75	0.03	76.42-82.88	100.87-152.43
RG_MSWEP	0.73	4.6	0.77	0.03	79.07-82.88	104.66-152.43
Validation						
RG	0.83	-4.8	0.84	0.68	88.19-84.17	116.12-138.60
MSWEP	0.75	3.4	0.77	0.37	81.34-84.17	101.95-138.60
RG_MSWEP	0.76	2.0	0.78	0.05	82.49-84.17	101.96-138.60

Source: the author.

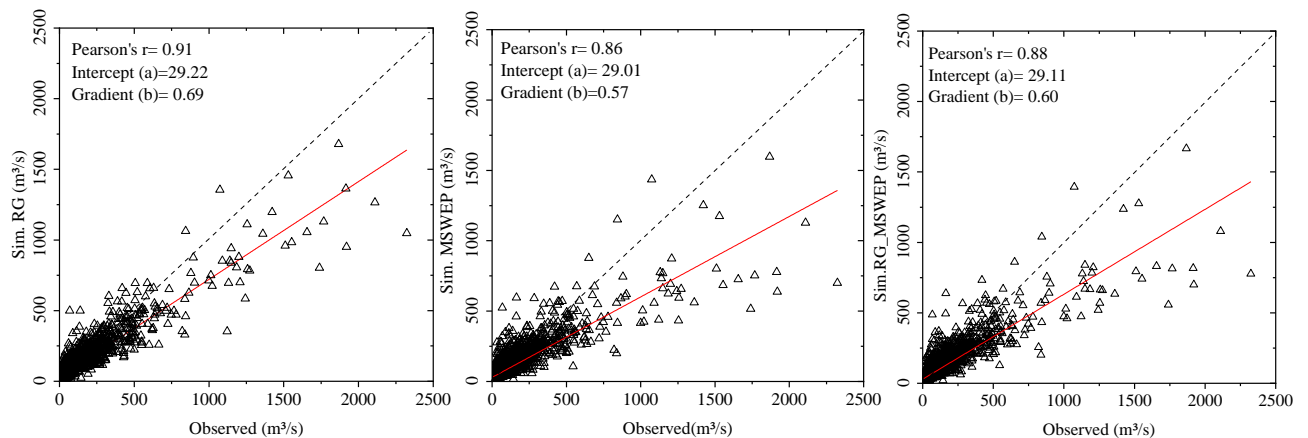
The model based on gauge observation ( $Q_{tRG}$ ) had a better overall performance to represents streamflow based on NSE and  $r^2$ , in comparison  $Q_{tMSWEP}$  and  $Q_{tRG\_MSWEP}$ , for both calibration and validation. It is interesting to notice that in validation period all scenarios presented better or equal results for NSE and  $r^2$ , if compared to calibration one. This pattern may be explained because of the higher numbers of streamflow peaks (outliers) in calibration period, which were not properly represented by SWAT and affected the indicators efficiency. The PBIAS for scenarios  $Q_{tMSWEP}$  and  $Q_{tRG\_MSWEP}$  which are 7.8% and 4.6% in calibration, and 3.4% and 2.0% in validation, respectively, highlights the influence of rainfall underestimation by MSWEP v.2 dataset on streamflow modeling resulting in lower streamflow results than gauge observations, while  $Q_{tRG}$  produced a continuous overestimation of -4.5% and -4.8%. The results for model performance found by Strauch et al. (2012) also achieved higher NSE value with a rain gauge driven model. This work used precipitation inputs based on ground-based rain gauges, with variations of the gauge data using a moving average and Thiessen polygons, and also by using precipitation data from TRMM product, in a 215 km<sup>2</sup> Brazilian watershed.

All scenarios had similarly low error, close to 0, between observed and simulated streamflow, for low streamflow occurrence during calibration. On the other hand, in validation, RG had the highest error of 0.68, followed by MSWEP with 0.37, and RG\_MSWEP with 0.05.

Even though the use of MSWEP v.2 resulted in the weaker overall model performance during the calibration (scenario MSWEP), the statistical indicators ( $r^2$ , PBIAS and NSE) from the validation are good (MORIASI *et al.*, 2015) and close to RG\_MSWEP. This result was possible affected by the lower density of ANA rain gauges with available data for the validation period (Figure 11), which led to poor quality of rainfall spatialization used as input in SWAT sub-basins. Therefore, MSWEP v.2 was a suitable alternative to overcome the spatialization quality for hydrological modeling in regions with data scarcity.

The correlation ( $r$ ) between observed and simulated daily streamflow for  $Q_{tRG}$ ,  $Q_{tMSWEP}$ , and  $Q_{tRG\_MSWEP}$  ranged from positive 0.91, 0.86 and 0.88, respectively (Figure 19). The intercepts values were 29.22 (RG), 29.01 (MSWEP) and 29.11 (RG\_MSWEP), and gradients were 0.69, 0.57 and 0.60, respectively. The results for  $r$  suggest that the simulated and the observed have good correlation, however, the values of gradient and interception reveals that maybe a curvilinear fitting would provide a best fit, and that the variance of the simulated increase with the increase in observed.

Figure 19- Simulated (Sim.) versus observed daily streamflow for  $Q_{tRG}$ ,  $Q_{tMSWEP}$  and  $Q_{tRG\_MSWEP}$  at gauging station. Pearson correlation ( $r$ ), interception ( $a$ ) and regression gradient ( $b$ ).



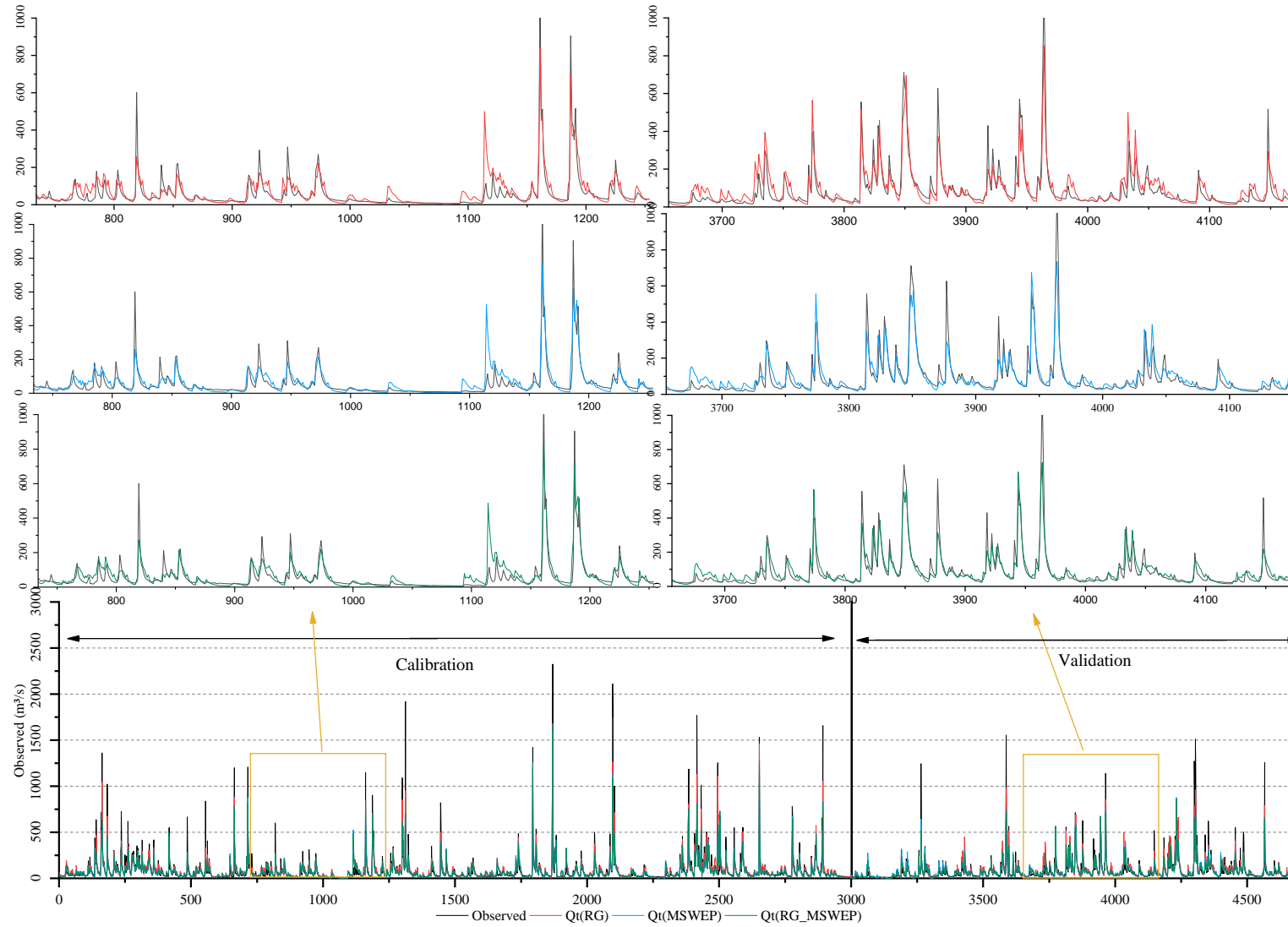
Source: the author.

All precipitation scenarios provided similar results for low streamflow in calibration period, while the best fit for validation was achieved with RG\_MSWEP (Figure 20- Observed and simulated daily streamflow for calibration period of  $Q_t$  condition and two zoomed areas. Figure 20).  $Q_{tRG}$  provided better streamflow prediction during high streamflow, in

comparison to  $Q_{t_{MSWEP}}$  and  $Q_{t_{RG\_MSWEP}}$ , which explains higher NSE result, although with a systematic underestimation of high peaks.



Figure 20- Observed and simulated daily streamflow for calibration period of Qt condition and two zoomed areas.

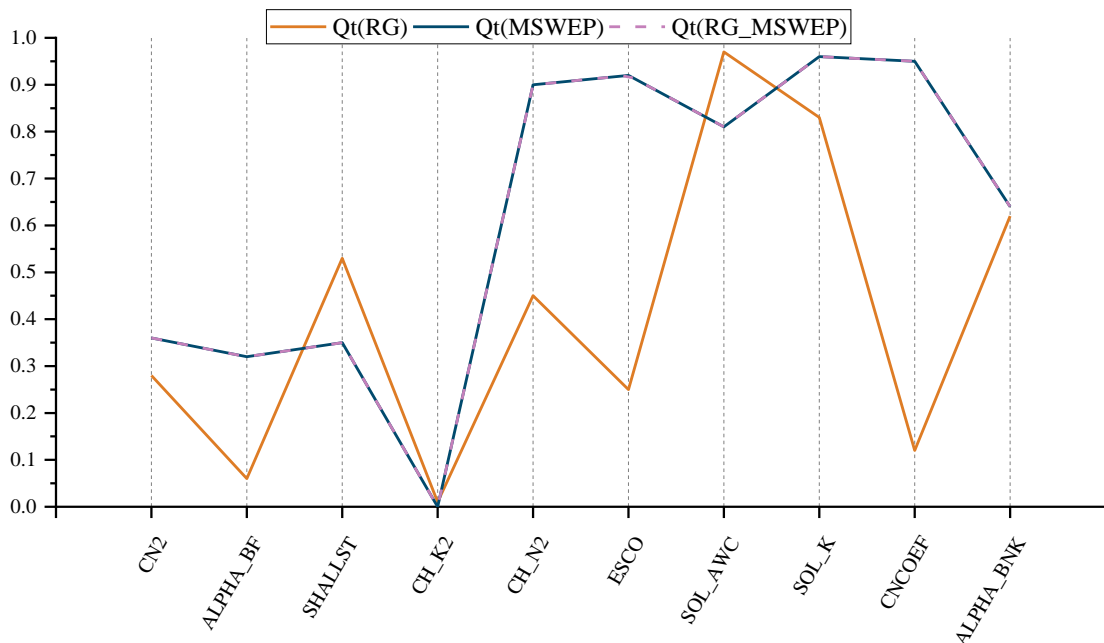


Source: the author.

The final range and fitted value of the 10 parameters, those sensible during Qt sensitive analysis, for calibration were the same for  $Q_{tMSWEP}$  and  $Q_{tRG\_MSWEP}$  (Figure 21). The CH\_K2, ALPHA\_BNK and CH\_N2, parameters responsible for describe the physical process affecting flow routing at watershed scale were the ones with less variation among the precipitation scenarios, had similar final minimum and maximum range and fitted parameters values.

Parameters related to soil evaporation (ESCO) and retention for daily CN depending on plant evapotranspiration (CNCOEF) had the highest deviation among the scenarios. ESCO fitted value ranged from 0.19 ( $Q_{tRG}$ ) to 0.62 ( $Q_{tMSWEP}$  and  $Q_{tRG\_MSWEP}$ ), and CNCOEF from 0.97 to 1.58. The CN2 represents the land use/land cover curve number and has large effects on surface runoff generation. Hence,  $Q_{tRG}$  with a smaller CN2 fitted value, in comparison to  $Q_{tMSWEP}$  and  $Q_{tRG\_MSWEP}$ , generates less runoff, and conversely, and has a higher available water content in the soil (SOL\_AWC) coping with a smaller hydraulic conductivity (SOL\_K). For  $Q_{tMSWEP}$  and  $Q_{tRG\_MSWEP}$  the opposite occurs, where CN and SOL\_K fitted values are higher and SOL\_AWC is lower, in order to compensate the precipitation underestimates due to MSWEP v.2 product.

Figure 21- Normalized parameters with regards to the final minimum and maximum range and fitted value for the best simulation of each precipitation scenario in Qt condition.



Source: the author.

The ground-based precipitation scenario (RG) captured 73% (calibration) and 81% (validation) of the streamflow observations within the 99PPU band with an r-factor of 0.6 and

0.7, respectively (Table 7).  $Q_{t_{MSWEP}}$  and  $Q_{t_{RG\_MSWEP}}$  had more consistent uncertainties prediction for the calibration period capturing 85% and 84% of the observed streamflow within its 95PPU band, with an r-factor of 0.49 and 0.5, respectively.

Table 7- Model uncertainties under  $Q_t$  calibration and validation

	Calibration		Validation	
	p-factor	r-factor	p-factor	r-factor
RG	0.73	0.6	0.81	0.70
MSWEP	0.85	0.49	0.81	0.53
RG_MSWEPE	0.84	0.5	0.81	0.53

Source: the author.

### 3.3.2.2 Model performance under $Q < 500$ condition

A total of 90 days (1.93%) of streamflow time series (2002-2016) was suppressed in order to obtain only streamflow data below the threshold of 500 m<sup>3</sup>/s and to meet a more reliable time series. Calibration statistical indicators had a worst overall result for  $Q < 500_{RG}$  approach in comparison to  $Q_{t_{RG}}$  approach, although the model performance was ‘good’ (MORIASI *et al.*, 2015) to simulate streamflow. The same behavior was found for comparison between  $Q < 500_{MSWEP}$  and  $Q < 500_{RG\_MSWEP}$  (Table 8). The NSE and PBIAS on calibration period, when compared to  $Q_t$  evaluation, provided a slight improvement for  $Q < 500_{MSWEP}$  and  $Q < 500_{RG\_MSWEP}$  simulations. The NSE improved from 0.7 ( $Q_{t_{MSWEP}}$ ) to 0.73 ( $Q < 500_{MSWEP}$ ), and 0.73 ( $Q_{t_{RG\_MSWEP}}$ ) to 0.75 ( $Q < 500_{RG\_MSWEP}$ ).

The  $Q < 500_{MSWEP}$  and  $Q < 500_{RG\_MSWEP}$  provided an overestimation in comparison to observed streamflow data in the validation process, where PBIAS ranged from 3.4% and 2.0% (validation in  $Q_t$ ) to -3.4% and -5%, respectively. This behavior is unlike the systematic underestimation of streamflow resulting from the use of MSWEP v.2 data ( $Q_t$  calibration and validation, and  $Q < 500$  calibration).

Moreover,  $Q_{t_{RG}}$  presented a streamflow underestimation of model simulation for validation period in this assessment, while in the former approach the  $Q_{t_{RG}}$  provided a consistent streamflow overestimation for validation and calibration

The Emin had a significant improvement over the validation period for RG and MSWEP, while RG\_MSWEPE the error increased with  $Q < 500$  approach in comparison to  $Q_t$  validation, but it is still the lower error among the scenarios. This result reinforces the need for precaution when using streamflow data in the extrapolation area of the rating curve, because it

could force the model to cope with unrealistic peak flows. In the calibration period a low error of 0.01 is found for RG, MSWEP, and RG\_MSWEF.

Table 8- Metrics result between observed streamflow and precipitation scenarios for  $Q < 500$

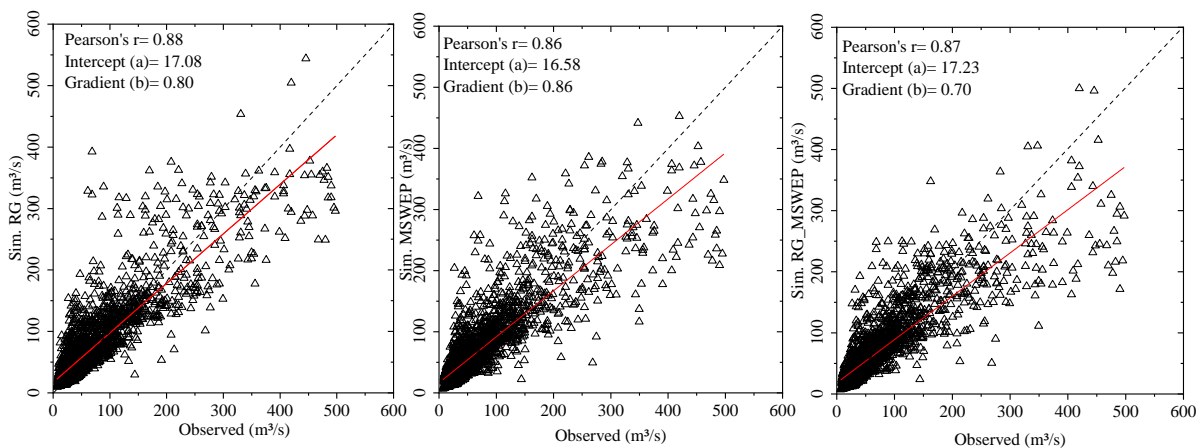
	NSE	PBIAS (%)	$r^2$	Emin	Streamflow mean Simulation- Observation( $m^3/s$ )	Standard deviation Simulation- Observation
Calibration						
RG	0.76	-6.4	0.77	0.01	70.63-66.37	68.10-73.66
MSWEP	0.73	0.4	0.74	0.01	66.11-66.37	65.20-73.66
RG_MSWEF	0.75	2.2	0.75	0.01	64.92-66.37	60.82-73.66
Validation						
RG	0.75	4	0.76	0.24	66.63-69.37	63.81-75.54
MSWEP	0.76	-3.4	0.76	0.29	71.71-69.37	63.72-75.54
RG_MSWEF	0.77	-5	0.77	0.12	72-83-69.37	63.99-75.54

Source: the author.

Very similar results of the statistical indicators were achieved for all precipitation when restricting the analysis of model simulation to discharges below  $500 m^3/s$ , unlike the results found for calibration and validation on Qt approach, where RG stand out over MSWEP and RG\_MSWEF performance.

Pearson correlation coefficient results are 0.88, 0.86 and 0.87, interceptions are 17.08, 16.58 and 17.23, and lastly, gradients are 0.80, 0.75 and 0.70. Even though Pearson's coefficient itself are lower, in the former approach interception and gradients were worst, suggesting that with  $Q < 500$  approach the results are more cohesive with less variance (Figure 22).

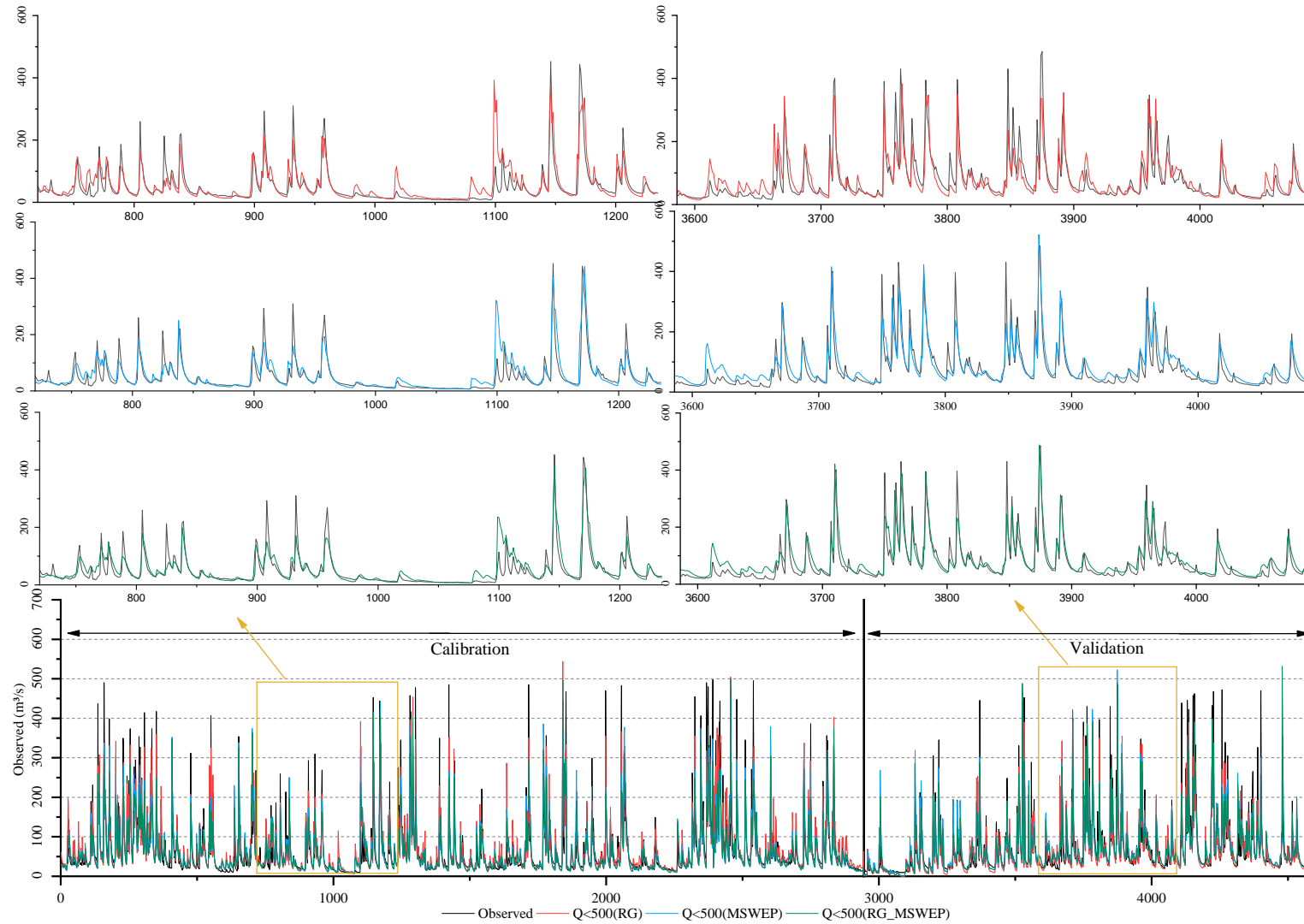
Figure 22- Comparison of simulated daily streamflow for RG, MSWEP and RG\_MSWEF and daily observed streamflow at gauging station for  $Q < 500$  approach. Pearson coefficient ( $r$ ), intercept ( $a$ ) and gradient ( $b$ ).



Source: the author.

The streamflow simulation results to all precipitation scenarios are very similar and accurately represent the streamflow observations (Figure 23).

Figure 23- Observed and simulated daily flow hydrographs for calibration period of  $Q < 500$

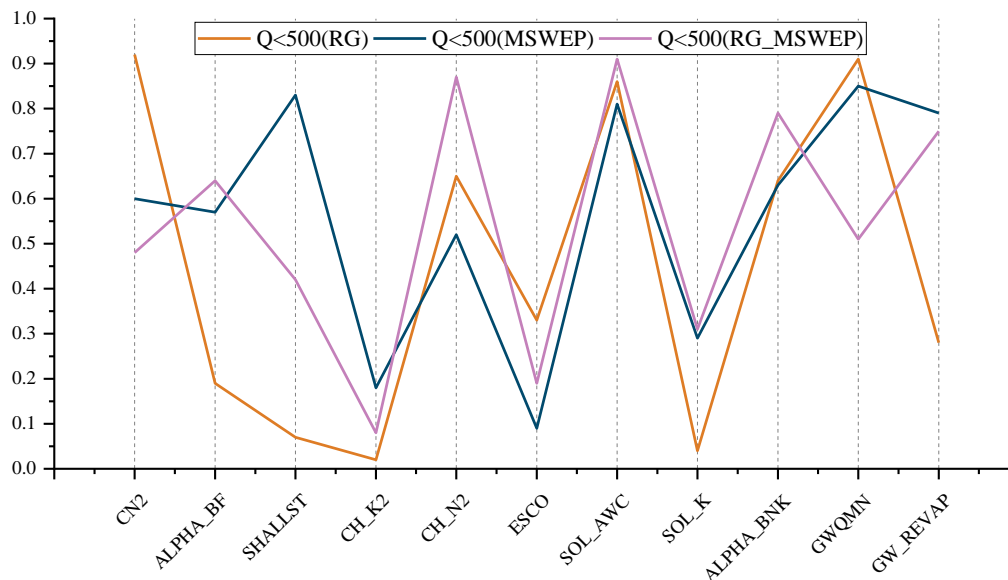


Source: the author.

The sensitive analysis considering  $Q < 500$  resulted in a set of 11 sensitive parameters (Figure 24). For the final calibrated parameters, in this streamflow approach, unlike  $Q_t$  approach,  $Q < 500_{MSWEP}$  and  $Q < 500_{RG\_MSWEP}$  had different final calibration ranges and fitted value (Figure 24).

The parameters related to groundwater flow (GWQMN, ALPHA\_BF, SHALLST, and GW\_REVAP) appears to have a greater influence over model simulation in the  $Q < 500$  approach, which contrast with the total flow ( $Q_t$ ) approach, where the parameters related to evapotranspiration and soil properties were the most important to surface runoff. The improvement of low streamflow simulation for  $Q < 500$  approach (Figure 23) in comparison to  $Q_t$  approach (Figure 20) possible occurred due to the GWQMN parameter influence as sensitive and part of calibration process. During low streamflow periods, groundwater contribution to streamflow can be significant and GWQMN is an important parameter related the occurrence of return flow.

Figure 24- Normalized parameters for  $Q < 500$ .



Source: the author.

The  $Q < 500_{MSWEP}$  and  $Q < 500_{RG\_MSWEP}$  had the best model uncertainty prediction result reaching 79% and 78% of streamflow measurement data bracketed by the 95PPU band in the calibration period (Table 9). In the validation period, both  $Q < 500_{MSWEP}$  and  $Q < 500_{RG\_MSWEP}$  reached 87%, with high 95PPU bandwidth in all simulations. The worst model uncertainty

prediction for calibration ( $\rho$ -factor=0.65; r-factor=0.8) and validation ( $\rho$ -factor=0.77; r-factor=0.74) periods was  $Q < 500_{RG}$ .

Table 9- Model uncertainty prediction in simulated streamflow calibration and validation for  $Q < 500$ .

	Calibration		Validation	
	p-factor	r-factor	p-factor	r-factor
RG	0.65	0.80	0.77	0.74
MSWEP	0.79	0.84	0.87	0.82
RG_MSWEPE	0.78	0.86	0.87	0.83

Source: the author.

### 3.4 CONCLUSIONS

This study aimed to evaluate a simple approach to overcome difficulties found in scarce or/and poor rain gauge regions to assess hydrologic modeling using SWAT model, where the rainfall network commonly suffers with low density, gaps, and series with short timeframe. After identifying the main drawbacks from the ground-based rainfall gauges at the study area, it was proposed the use of a satellite-based product (MSWEP v.2) in order to overcome the inherent missing precipitation and gaps problems throughout the timeframe of the study. Therefore, 3 precipitation scenarios were evaluated: 1) ground-based precipitation from rain gauges (RG); 2) MSWEP v.2 dataset only (MSWEP), and 3) precipitation from the ground-based gauges with infilling of missing precipitation data with MSWEP v.2 dataset (RG\_MSWEPE).

Firstly, we evaluated the error components of our three precipitation scenarios with regards to the mean areal precipitation (MAP). MSWEP v.2 dataset provided a systematic underestimation of MAP, being the hit bias the main source of error. This error component was also present in the merging scenario (RG\_MSWEPE) although in small amplitude. Furthermore, the intensity distribution of the MAP in the 3 scenarios showed that the most frequent precipitation ranged between 0-1 mm for MSWEP and RG\_MSWEPE, which represents events with none or small changes on the hydrological cycle. The most frequent precipitation in RG ranged from 1 to 5 mm. These results highlight that MSWEP v.2 product underestimated daily rainfall.

The assessment of SWAT model performance for the three precipitation scenarios was done under two different streamflow approaches,  $Q_t$  with the full streamflow time series, and  $Q < 500$  which excluded high streamflow data in the extrapolation of the rating curve. These



approaches highlighted the importance of a previous analysis of the streamflow time series. SWAT model had the best metrics performance result when simulated with RG scenario (NSE=0.81 for  $Q_{tRG}$ ). However, the algorithm used in SWAT-CUP (SUFI-2) is a stochastic procedure, which provides a set of several good simulations and its uncertainties prediction. For this reason, that  $Q_{tRG}$  cannot be fully considered the best model.

MSWEP and RG\_MSWEF (in both  $Q_t$  and  $Q<500$  approaches) provided better model uncertainty prediction than RG, with a higher percentage of streamflow measurement bracketed by the 95PPU band, when evaluating the result as stochastic. The bandwidth had an improvement primarily due to the removal of the high streamflow data (rating curve extrapolation), which resulted in a smaller standard deviation, comparing  $Q_t$  results against  $Q<500$ . The results for model uncertainties (p-factor and r-factor) also implies that uncertainties estimated based on parameter uncertainties (parameter range) were unable to capture all source of uncertainty, suggesting that a good amount of uncertainties comes notably from precipitation.

Better agreements between observed and simulated streamflow were found for low streamflow periods, especially with the merging scenario in the validation period ( $Q<500$  RG\_MSWEF), when restricting the process of calibration and validation to streamflow data under a reliable threshold of the rating curve. It is not possible to infer which evaluated scenario was the best for representing rainfall of Guaporé watershed. The results provided only the suitability of the SWAT model in the adjustment of parameters in order to transform different precipitation input from different sources to streamflow time series. Nevertheless, the merging scenario, which used ground-based rainfall and satellite precipitation data, illustrated a potential and reliable alternative to assess hydrologic models in poorly rain gauge watershed.

Although the merging scenario provided additional rainfall information including spatial variability, it was not possible to conclude its benefit on the modeling results, once only one streamflow gauge was available. However, these rainfall improvements possibly provided a positive effect over model performance at the sub-basin scale, due to the unavailability of rain gauges or even complete time-series for long periods along the timeframe evaluated.

The MSWEP v.2 precipitation dataset allowed to extend the ground-based precipitation time series, at the same time the rain gauge network spatial density was virtually increased, and missing precipitation records were infilled, diminishing the problems that drove the aims of this work. The findings of this work can also encourage and support hydrologists from other developing countries with the same inherent problems to follow similar methodology.

## 3.5 REFERENCES

- ABBASPOUR, K. C. **SWAT - CUP SWAT: Calibration and Uncertainty Programs**. [S. l.: s. n.].
- ABBASPOUR, K. C. *et al.* A continental-scale hydrology and water quality model for Europe: Calibration and uncertainty of a high-resolution large-scale SWAT model. **Journal of Hydrology**, [S. l.], v. 524, p. 733–752, 2015. Disponível em: <https://doi.org/10.1016/j.jhydrol.2015.03.027>
- ABBASPOUR, K. C. *et al.* **SWAT - CUP: SWAT calibration and uncertainty programs**. Zurich: [s. n.], 2007 a.
- ABBASPOUR, K.C. *et al.* Modelling hydrology and water quality in the pre-alpine/alpine Thur watershed using SWAT. **Journal of Hydrology**, [S. l.], v. 333, n. 2–4, p. 413–430, 2007 b. Disponível em: <https://doi.org/10.1016/j.jhydrol.2006.09.014>
- AJAAJ, A. A.; MISHRA, A. K.; KHAN, A. A. Evaluation of satellite and gauge-based precipitation products through hydrologic simulation in Tigris River basin under data-scarce environment. **Journal of Hydrologic Engineering**, [S. l.], v. 24, n. 3, p. 1–18, 2019. Disponível em: [https://doi.org/10.1061/\(ASCE\)HE.1943-5584.0001737](https://doi.org/10.1061/(ASCE)HE.1943-5584.0001737)
- ALAZZY, A. A. *et al.* Evaluation of Satellite Precipitation Products and Their Potential Influence on Hydrological Modeling over the Ganzi River Basin of the Tibetan Plateau. **Advances in Meteorology**, [S. l.], v. 2017, 2017. Disponível em: <https://doi.org/10.1155/2017/3695285>
- ANA. **Conjuntura dos recursos hídricos no Brasil**. [s. l.], 2017. Disponível em: <http://www3.snirh.gov.br/portal/snirh/centrais-de-conteudos/conjuntura-dos-recursos-hidricos>. Acesso em: 21 set. 2018.
- ANDERSSON, J. C. M. *et al.* Improved SWAT Model Performance With Time-Dynamic Voronoi Tessellation of Climatic Input Data in Southern Africa. **Journal of the American Water Resources Association**, [S. l.], v. 48, n. 3, p. 480–493, 2012. Disponível em: <https://doi.org/10.1111/j.1752-1688.2011.00627.x>
- ARNOLD, J. G. *et al.* LARGE AREA HYDROLOGIC MODELING AND ASSESSMENT PART I: MODEL DEVELOPMENT 1. **Journal of the American Water Resources Association**, [S. l.], v. 34, n. 1, p. 73–89, 1998.
- ARNOLD, J. G. *et al.* SWAT: model use, calibration and validation. **American Society of Agricultural and Biological Engineers**, [S. l.], v. 55, n. 4, p. 1491–1508, 2012.
- AWANGE, J. L.; HU, K. X.; KHAKI, M. The newly merged satellite remotely sensed, gauge and reanalysis-based Multi-Source Weighted-Ensemble Precipitation: Evaluation over Australia and Africa (1981–2016). **Science of the Total Environment**, [S. l.], v. 670, p. 448–465, 2019. Disponível em: <https://doi.org/10.1016/j.scitotenv.2019.03.148>
- BECK, H. E. *et al.* MSWEP V2 global 3-hourly 0.1° precipitation: methodology and quantitative assessment. **Bulletin of the American Meteorological Society**, [S. l.], 2018. Disponível em: <https://doi.org/10.1175/BAMS-D-17-0138.1>

BECK, H. E. *et al.* Daily evaluation of 26 precipitation datasets using Stage-IV gauge-radar data for the CONUS. **Hydrology and Earth System Sciences**, [S. l.], v. 23, n. 1, p. 207–224, 2019. Disponível em: <https://doi.org/10.5194/hess-23-207-2019>

BRESSIANI, D. A.; *et al.* Effects of different spatial and temporal weather data resolutions on the stream flow modeling of a semi-arid basin, Northeast Brazil. **International Journal of Agricultural and Biological Engineering**, [S. l.], v. 8, n. 3, p. 1–16, 2015 a. Disponível em: <https://doi.org/10.3965/j.ijabe.20150803.970>

BRESSIANI, D. A.; *et al.* Review of Soil and Water Assessment Tool (SWAT) applications in Brazil: Challenges and prospects. **International Journal of Agricultural and Biological Engineering**, [S. l.], v. 8, 2015 b.

CHO, J. *et al.* Effect of Spatial Distribution of Rainfall on Temporal and Spatial Uncertainty of SWAT Output. [S. l.], n. September 2009, 2013. Disponível em: <https://doi.org/10.13031/2013.29143>

CREMONINI, J. BRIGHENTI, Tássia Mattos; BONUMÁ, Nádia Bernardi. Avaliação Do Uso Dados De Satélite Para O Cálculo Da Vazão Em Uma Pequena Bacia Hidrográfica Por Meio Do Modelo Swat Evaluation of the Use of Satellite Data for Discharge Calculation in a Small Hydrographic Basin By Means of. [S. l.], 2014.

ESRI. **ArcGIS 10.4**. [S. l.: s. n.]

HEROLD, N. *et al.* How much does it rain over land? **Geophysical Research Letters**, [S. l.], v. 43, n. 1, p. 341–348, 2016. Disponível em: <https://doi.org/10.1002/2015GL066615>

HOU, A. Y. *et al.* The global precipitation measurement mission. **Bulletin of the American Meteorological Society**, [S. l.], v. 95, n. 5, p. 701–722, 2014. Disponível em: <https://doi.org/10.1175/BAMS-D-13-00164.1>

HUGHES, D. A. Comparison of satellite rainfall data with observations from gauging station networks. **Journal of Hydrology**, [S. l.], v. 327, n. 3–4, p. 399–410, 2006. Disponível em: <https://doi.org/10.1016/j.jhydrol.2005.11.041>

IBGE. **Produção agrícola municipal**. [s. l.], 2017. Disponível em: <https://sidra.ibge.gov.br/pesquisa/pam/tabelas>. Acesso em: 17 out. 2019.

INMET. **Meteorological Database for Teaching and Research**. [S. l.: s. n.]

KRAUSE, P.; BOYLE, D. P.; BÄSE, F. Comparison of different efficiency criteria for hydrological model assessment. **Advances in Geosciences**, [S. l.], v. 5, p. 89–97, 2005. Disponível em: <https://doi.org/10.5194/adgeo-5-89-2005>

LAKIEW, H. B.; MOGES, S. A.; ASFAW, Dereje Hailu. Hydrological performance evaluation of multiple satellite precipitation products in the upper Blue Nile basin, Ethiopia. **Journal of Hydrology: Regional Studies**, [S. l.], v. 27, n. January, p. 100664, 2020. Disponível em: <https://doi.org/10.1016/j.ejrh.2020.100664>

LANG, M. *et al.* Extrapolação des courbes de tarage par modélisation hydraulique, avec

application à l'analyse fréquentielle des crues. **Hydrological Sciences Journal**, [S. l.], v. 55, n. 6, p. 883–898, 2010. Disponível em: <https://doi.org/10.1080/02626667.2010.504186>

LIU, J. *et al.* Evaluation and comparison of CHIRPS and MSWEP daily-precipitation products in the Qinghai-Tibet Plateau during the period of 1981–2015. **Atmospheric Research**, [S. l.], v. 230, n. July, p. 104634, 2019. Disponível em: <https://doi.org/10.1016/j.atmosres.2019.104634>

MAIDMENT, R. I. *et al.* A new, long-term daily satellite-based rainfall dataset for operational monitoring in Africa. **Scientific Data**, [S. l.], v. 4, p. 1–17, 2017. Disponível em: <https://doi.org/10.1038/sdata.2017.63>

MONDAL, A.; LAKSHMI, V.; HASHEMI, H. Intercomparison of trend analysis of Multisatellite Monthly Precipitation Products and Gauge Measurements for River Basins of India. **Journal of Hydrology**, [S. l.], v. 565, n. April, p. 779–790, 2018. Disponível em: <https://doi.org/10.1016/j.jhydrol.2018.08.083>

MORIASI, D. N. *et al.* Hydrologic and Water Quality Models: Performance Measures and Evaluation Criteria. **Transactions of the ASABE**, [S. l.], v. 58, n. 6, p. 1763–1785, 2015. Disponível em: <https://doi.org/10.13031/trans.58.10715>

MWALE, F. D.; ADELOYE, A. J.; RUSTUM, R. Infilling of missing rainfall and streamflow data in the Shire River basin, Malawi - A self organizing map approach. **Physics and Chemistry of the Earth**, [S. l.], v. 50–52, p. 34–43, 2012. Disponível em: <https://doi.org/10.1016/j.pce.2012.09.006>

NEITSCH, S. L. *et al.* Soil & Water Assessment Tool Theoretical Documentation Version 2009. **Grassland, Soil, and Water Research Laboratory**, [S. l.], p. 647, 2011.

PEREIRA BRITTO, F.; BARLETTA, R.; MENDONÇA, M. Regionalização Sazonal E Mensal Da Precipitação Pluvial Máxima No Estado Do Rio Grande Do Sul. **Revista Brasileira de Climatologia**, [S. l.], v. 3, p. 35–51, 2008. Disponível em: <https://doi.org/10.5380/abclima.v3i0.25425>

RODRIGUES, M. F. **Dinâmica hidrossedimentológica de pequenas bacias hidrográficas florestais**. 2015. - Universidade Federal de Santa Maria, [s. l.], 2015.

SANTOS, H. G. dos *et al.* **Sistema brasileiro de classificação de solos**. 5. ed. [S. l.]: EMBRAPA, 2018. *E-book*.

SENET-APARICIO, J. *et al.* Using Multiple Monthly Water Balance Models to Evaluate Gridded Precipitation Products over Peninsular Spain. **Remote Sensing**, [S. l.], v. 10, n. 6, 2018. Disponível em: <https://doi.org/10.3390/rs10060922>

SIRISENA, T. A. J. G. *et al.* Effects of different precipitation inputs on streamflow simulation in the Irrawaddy River Basin, Myanmar. **Journal of Hydrology: Regional Studies**, [S. l.], v. 19, n. October, p. 265–278, 2018. Disponível em: <https://doi.org/10.1016/j.ejrh.2018.10.005>

STRAUCH, M. *et al.* Using precipitation data ensemble for uncertainty analysis in SWAT

streamflow simulation. **Journal of Hydrology**, [S. l.], v. 414–415, p. 413–424, 2012. Disponível em: <https://doi.org/10.1016/j.jhydrol.2011.11.014>

TANG, X. *et al.* Assessing the uncertainties of four precipitation products for SWAT modeling in Mekong River Basin. **Remote Sensing**, [S. l.], v. 11, n. 3, p. 1–24, 2019. Disponível em: <https://doi.org/10.3390/rs11030304>

TIAN, Y. *et al.* Component analysis of errors in Satellite-based precipitation estimates. **Journal of Geophysical Research Atmospheres**, [S. l.], v. 114, n. 24, p. 1–15, 2009. Disponível em: <https://doi.org/10.1029/2009JD011949>

TIECHER, T. Fingerprinting Sediment Sources in Agricultural Catchments. [S. l.], p. 307, 2015.

TOBIN, K. J.; BENNETT, Marvin E. Satellite precipitation products and hydrologic applications. **Water International**, [S. l.], v. 39, n. 3, p. 360–380, 2014. Disponível em: <https://doi.org/10.1080/02508060.2013.870423>

WMO. **Guide to Hydrological Practices**. Sixth ed. Geneva: World Meteorological Organization, 2008. v. *IE-book*. Disponível em: <https://doi.org/10.1080/02626667.2011.546602>

WOLLMANN, C. A. ; GALVANI, E. Caracterização climática regional do Rio Grande do Sul: dos estudos estáticos ao entendimento da gênese. **Revista Brasileira de Climatologia**, [S. l.], v. 11, 2012.

WORQLUL, A. W. *et al.* Performance of bias corrected MPEG rainfall estimate for rainfall-runoff simulation in the upper Blue Nile Basin, Ethiopia. **Journal of Hydrology**, [S. l.], v. 556, p. 1182–1191, 2018. Disponível em: <https://doi.org/10.1016/j.jhydrol.2017.01.058>

WU, Z. *et al.* Hydrologic evaluation of Multi-Source satellite precipitation products for the Upper Huaihe River Basin, China. **Remote Sensing**, [S. l.], v. 10, n. 6, 2018. Disponível em: <https://doi.org/10.3390/rs10060840>

XU, Z. *et al.* Evaluating the accuracy of MSWEP V2.1 and its performance for drought monitoring over mainland China. **Atmospheric Research**, [S. l.], v. 226, p. 17–31, 2019. Disponível em: <https://doi.org/10.1016/j.atmosres.2019.04.008>

XUE, F. *et al.* Evaluating the impact of spatial variability of precipitation on streamflow simulation using a SWAT model. **Water Policy**, [S. l.], v. 21, n. 1, p. 178–196, 2019. Disponível em: <https://doi.org/10.2166/wp.2018.118>

ZEIGER, S.; HUBBART, J. An assessment of mean areal precipitation methods on simulated stream flow: A SWAT model performance assessment. **Water (Switzerland)**, [S. l.], v. 9, n. 7, 2017. Disponível em: <https://doi.org/10.3390/w9070459>

## 4 CONCLUSÕES GERAIS

Devido aos problemas relacionados à baixa densidade da rede de monitoramento pluviométrico no Brasil, somados à curta extensão das séries de dados e recorrentes falhas, é necessário que, em muitos estudos, sejam encontrados procedimentos alternativos para a aquisição de dados de precipitação em diversos locais.

Neste trabalho, o produto de precipitação MSWEP v.2 mostrou-se uma importante fonte alternativa de dados de precipitação para ampliar os dados de precipitação disponíveis em uma bacia hidrográfica brasileira de médio porte com escassez de dados pluviométricos.

Com uma metodologia simples, o produto do MSWEP v.2 foi utilizado para realizar o preenchimento diário de falhas existentes nas séries de precipitação, torná-las mais extensas e, adicionalmente, melhorar o processo de espacialização da precipitação. Isso foi possível em razão da boa correlação existente entre a série diária de precipitação monitorada em postos pluviométricos e o produto MSWEP v.2, embora este tenha apresentado uma leve tendência sistemática na subestimativa das precipitações, se comparado com os dados observados.

O produto do MSWEP v.2 foi utilizado na proposição de cenários alternativos de precipitação para a modelagem hidrológica com o modelo SWAT, de forma a avaliar o potencial de melhoria na qualidade da simulação hidrológica, em razão de uma série de precipitação mais consistente. Além do banco de dados de precipitação baseado na rede pluviométrica, foram utilizadas precipitações oriundas exclusivamente do MSWEP v.2 e, ainda, a combinação destes dados com dados observados em campo.

Como resultado, o cenário que combinou as diferentes fontes de informação de precipitações produziu vazões simuladas muito similares àquelas observadas. Assim, a inclusão de informações de precipitação oriunda do MSWEP v.2 aos dados de precipitação monitorados em redes convencionais permitiu a utilização de séries de precipitação com boa qualidade, e mostrou-se uma fonte alternativa de dados bastante interessante, principalmente para regiões onde a rede pluviométrica é deficitária ou inexistente.

## FUTUROS TRABALHOS

No desenvolvimento do trabalho, foi possível identificar alguns pontos importantes que podem vir a serem estudados em futuros trabalhos. Primeiro, verifica-se a oportunidade de se obter um fator de correção dos dados de precipitação do MSWEP v.2, uma vez que foi identificada uma subestimativa sistemática dos dados de precipitação.

Ainda, em relação ao uso do SWAT no Brasil, verifica-se a necessidade de criação de um banco de dados unificado, pois os dados necessários para a utilização do modelo estão muitas vezes em posse de empresas privadas, ou grupos de pesquisas com baixo alcance de divulgação. Outro ponto interessante, seria a criação de um “manual” de uso do SWAT no Brasil, uma vez que por ter sido criado nos EUA, sua base de dados é adaptada para as condições climáticas e vegetais de lá. O que se verifica é que em muitos casos, poucos modeladores no Brasil estão atentos para as adaptações que devem ser feitas para que o modelo possa ser usado de forma adequada nas bacias hidrográficas brasileiras.

## 5 REFERÊNCIAS GERAIS

ABBASPOUR, K. C. *et al.* A continental-scale hydrology and water quality model for Europe: Calibration and uncertainty of a high-resolution large-scale SWAT model. **Journal of Hydrology**, [S. l.], v. 524, p. 733–752, 2015. Disponível em: <https://doi.org/10.1016/j.jhydrol.2015.03.027>

ANA. **Conjuntura dos recursos hídricos no Brasil**. [s. l.], 2017. Disponível em: <http://www3.snirh.gov.br/portal/snirh/centrais-de-conteudos/conjuntura-dos-recursos-hidricos>. Acesso em: 21 set. 2018.

ARNOLD, J. G. *et al.* Large area hydrologic modeling and assessment part i: model development 1. **Journal of the American Water Resources Association** [S. l.], v. 34, n. 1, p. 73–89, 1998.

ARNOLD, J. G. *et al.* SWAT: model use, calibration and validation. **American Society of Agricultural and Biological Engineers**, [S. l.], v. 55, n. 4, p. 1491–1508, 2012.

BECK, Hylke E. *et al.* MSWEP V2 global 3-hourly 0.1° precipitation: methodology and quantitative assessment. **Bulletin of the American Meteorological Society**, [S. l.], 2018. Disponível em: <https://doi.org/10.1175/BAMS-D-17-0138.1>

BEVEN, K. **Rainfall-Runoff Modelling: the primer**. [S. l.: s. n.]. 2. ed. 2012. Disponível em: <https://doi.org/10.1002/9781119951001>

BRESSIANI, D. De A.; *et al.* Effects of different spatial and temporal weather data resolutions on the stream flow modeling of a semi-arid basin, Northeast Brazil. **International Journal of Agricultural and Biological Engineering**, [S. l.], v. 8, n. 3, p. 1–16, 2015 a. Disponível em: <https://doi.org/10.3965/j.ijabe.20150803.970>

BRESSIANI, D. De A.; *et al.* Review of Soil and Water Assessment Tool (SWAT) applications in Brazil: Challenges and prospects. **International Journal of Agricultural and Biological Engineering**, [S. l.], v. 8, 2015 b.

CREMONINI, J.; BRIGHENTI, T. M.; BONUMÁ, N. B.. Avaliação Do Uso Dados De Satélite Para O Cálculo Da Vazão Em Uma Pequena Bacia Hidrográfica Por Meio Do Modelo Swat. Anais do XXII SIMPÓSIO BRASILEIRO DE RECURSOS HÍDRICOS, 2014, Florianópolis/SC.

DEVRIES, J. J.; HROMADKA, T. V. Chapter 21 Computer Models for Surface Water. In: MAIDMENT, D. R. (org.). **Handbook of Hydrology**. New York: McGraw-Hill, 1993.

DINKU, T. *et al.* Combined use of satellite estimates and rain gauge observations to generate high-quality historical rainfall time series over Ethiopia. **International Journal of Climatology**, [S. l.], v. 34, n. 7, p. 2489–2504, 2014. Disponível em: <https://doi.org/10.1002/joc.3855>

FENSTERSEIFER, C.; ALLASIA, D. G.; PAZ, A. R. Assessment of the TRMM 3B42 Precipitation Product in Southern Brazil. **Journal of the American Water Resources Association**, [S. l.], v. 52, n. 2, p. 367–375, 2016. Disponível em: <https://doi.org/10.1111/1752->



1688.12398

GASSMAN, P. W. *et al.* The Soil and Water Assessment Tool: historical development, applications, and future research directions. **American Society of Agricultural and Biological Engineers**, [*S. l.*], v. 50, n. 4, p. 1211–1250, 2007.

GITHUNGO, W. *et al.* Infilling monthly rain gauge data gaps with satellite estimates for ASAL of Kenya. **Hydrology**, [*S. l.*], v. 3, n. 4, 2016. Disponível em: <https://doi.org/10.3390/hydrology3040040>

HUGHES, D. A. Comparison of satellite rainfall data with observations from gauging station networks. **Journal of Hydrology**, [*S. l.*], v. 327, n. 3–4, p. 399–410, 2006. Disponível em: <https://doi.org/10.1016/j.jhydrol.2005.11.041>

LI, D. *et al.* Adequacy of TRMM satellite rainfall data in driving the SWAT modeling of Tiaoxi catchment (Taihu lake basin, China). **Journal of Hydrology**, [*S. l.*], v. 556, p. 1139–1152, 2018. Disponível em: <https://doi.org/10.1016/j.jhydrol.2017.01.006>

MAIDMENT, D.; **Handbook of Hydrology**. [*S. l.*]: McGraw-Hill Professional, 1993.

MISHRA, A. K.. Effect of rain gauge density over the accuracy of rainfall: A case study over Bangalore, India. **SpringerPlus**, [*S. l.*], v. 2, n. 1, p. 1–7, 2013. Disponível em: <https://doi.org/10.1186/2193-1801-2-311>

MORIASI, D. N. *et al.* Hydrologic and Water Quality Models: Performance Measures and Evaluation Criteria. **Transactions of the ASABE**, [*S. l.*], v. 58, n. 6, p. 1763–1785, 2015. Disponível em: <https://doi.org/10.13031/trans.58.10715>

NEITSCH, S. L. *et al.* Soil & Water Assessment Tool Theoretical Documentation Version 2009. **Grassland, Soil, and Water Research Laboratory**, [*S. l.*], p. 647, 2011.

SALIO, P. *et al.* Evaluation of high-resolution satellite precipitation estimates over southern South America using a dense rain gauge network. **Atmospheric Research**, [*S. l.*], v. 163, p. 146–161, 2015. Disponível em: <https://doi.org/10.1016/j.atmosres.2014.11.017>

SIRISENA, T. A. J. G. *et al.* Effects of different precipitation inputs on streamflow simulation in the Irrawaddy River Basin, Myanmar. **Journal of Hydrology: Regional Studies**, [*S. l.*], v. 19, n. October, p. 265–278, 2018. Disponível em: <https://doi.org/10.1016/j.ejrh.2018.10.005>

STRAUCH, M. *et al.* Using precipitation data ensemble for uncertainty analysis in SWAT streamflow simulation. **Journal of Hydrology**, [*S. l.*], v. 414–415, p. 413–424, 2012. Disponível em: <https://doi.org/10.1016/j.jhydrol.2011.11.014>

TANG, X. *et al.* Assessing the uncertainties of four precipitation products for SWAT modeling in Mekong River Basin. **Remote Sensing**, [*S. l.*], v. 11, n. 3, p. 1–24, 2019. Disponível em: <https://doi.org/10.3390/rs11030304>

THIEMIG, V. *et al.* Validation of satellite-based precipitation products over sparsely Gauged African River basins. **Journal of Hydrometeorology**, [*S. l.*], v. 13, n. 6, p. 1760–1783, 2012. Disponível em: <https://doi.org/10.1175/JHM-D-12-032.1>

TOBIN, K. J.; BENNETT, M. E. Satellite precipitation products and hydrologic applications. **Water International**, [S. l.], v. 39, n. 3, p. 360–380, 2014. Disponível em: <https://doi.org/10.1080/02508060.2013.870423>

TUCCI, C. E. M. **Modelos hidrológicos**. 2. ed. [S. l.]: Editora da UFRGS, 2005.

TUCCI, C. E. M. **Hidrologia: ciência e aplicação**. 4. ed. [S. l.]: UFRGS/ABRH, 2012.

TUO, Y. *et al.* Evaluation of precipitation input for SWAT modeling in Alpine catchment: A case study in the Adige river basin (Italy). **Science of the Total Environment**, [S. l.], v. 573, p. 66–82, 2016. Disponível em: <https://doi.org/10.1016/j.scitotenv.2016.08.034>

YUAN, F. *et al.* Assessment of GPM and TRMM multi-satellite precipitation products in streamflow simulations in a data sparse mountainous watershed in Myanmar. **Remote Sensing**, [S. l.], v. 9, n. 3, 2017. Disponível em: <https://doi.org/10.3390/rs9030302>