Improving Satellite-based precipitation estimates

A spatiotemporal object-oriented approach to error analysis and correction

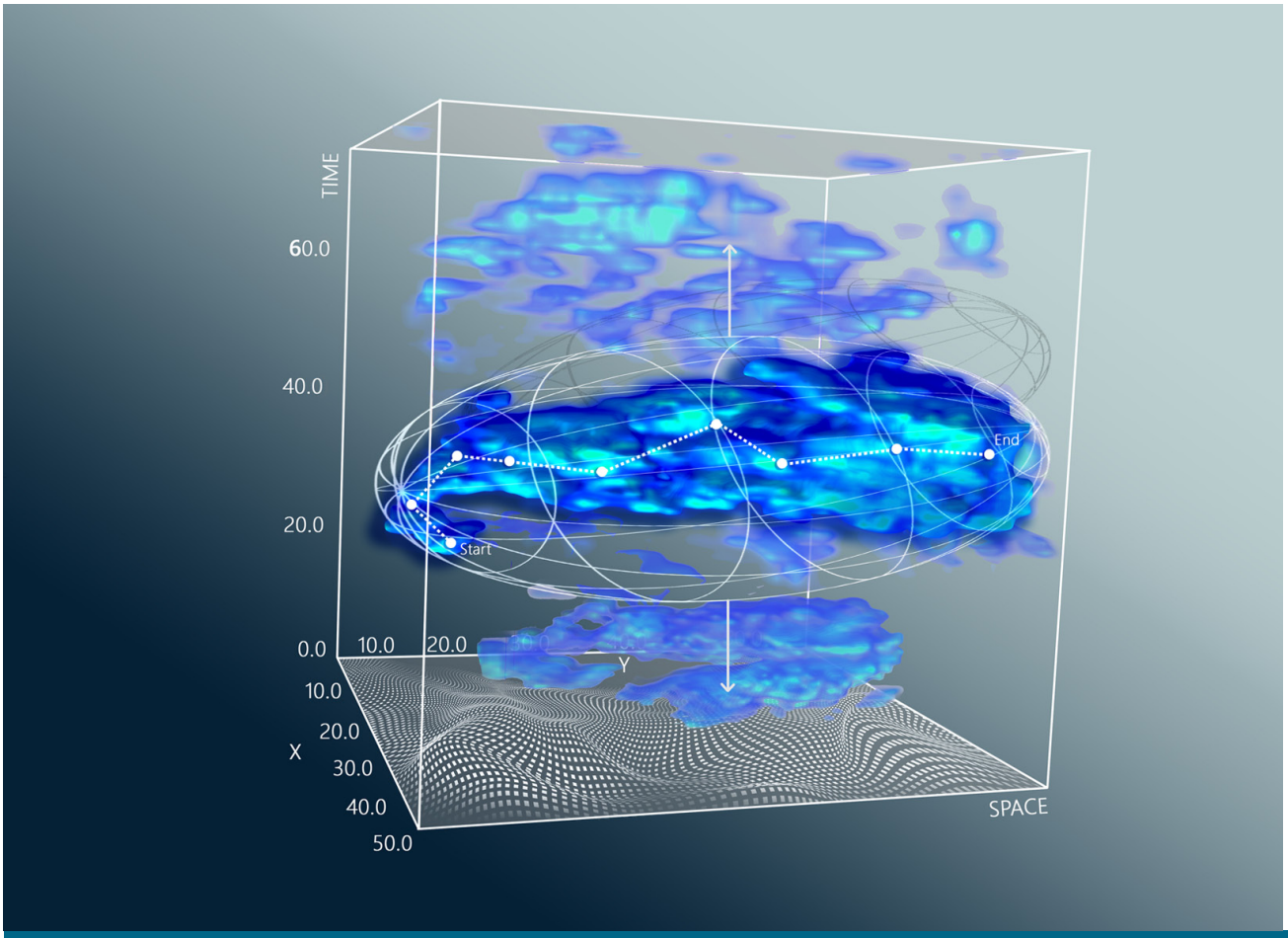
Laverde Barajas, M.A.

Publication date
2022

Document Version
Final published version

Citation (APA)
Laverde Barajas, M. A. (2022). *Improving Satellite-based precipitation estimates: A spatiotemporal object-oriented approach to error analysis and correction*. [Dissertation (TU Delft), Delft University of Technology].

Important note
To cite this publication, please use the final published version (if applicable).
Please check the document version above.



Improving Satellite-Based Precipitation Estimates

A Spatiotemporal Object-Oriented Approach to Error Analysis and Correction

Miguel Angel Laverde Barajas

IMPROVING SATELLITE-BASED PRECIPITATION ESTIMATES

A SPATIOTEMPORAL OBJECT-ORIENTED APPROACH TO ERROR ANALYSIS AND CORRECTION

Miguel Angel Laverde Barajas

IMPROVING SATELLITE-BASED PRECIPITATION ESTIMATES

A SPATIOTEMPORAL OBJECT-ORIENTED APPROACH TO ERROR ANALYSIS AND CORRECTION

DISSERTATION

Submitted in fulfillment of the requirements of
the Board for Doctorates of Delft University of Technology
and
of the Academic Board of the IHE Delft
Institute for Water Education
for
the Degree of DOCTOR
to be defended in public on
11, April 2022, at 15:00 hours
in Delft, the Netherlands

by

Miguel Angel LAVERDE BARAJAS
Master of Science in Civil Engineering,
National Autonomous University of Mexico, Mexico City, Mexico,

born in Bogota, Colombia.

This dissertation has been approved by the

promotor: Prof.dr. D.P. Solomatine

copromotor: Dr. G.A. Corzo Perez

Composition of the doctoral committee:

Rector Magnificus TU Delft	Chairman
Rector IHE Delft	Vice-Chairman
Prof.dr. D.P. Solomatine	TU Delft/ IHE Delft, promotor
Dr. G.A. Corzo Perez	IHE Delft, copromotor

Independent members:

Prof.dr.ir. N.C. van de Giesen	Technische Universiteit Delft
Prof.dr.rer.nat.-ing A. Bardossy	University of Stuttgart, Germany
Prof.dr.ir. A.H. Weerts	WUR/ Deltares
Dr. D.B. Kirschbaum	NASA Goddard Space Flight Center, USA
Prof.dr.ir. H.W.J. Russchenberg	TU Delft, reserve member

Prof.dr.ir. R. Uijlenhoet contributed significantly to the realization of the dissertation.

This research was conducted under the auspices of the Graduate School for Socio-Economic and Natural Sciences of the Environment (SENSE)

Published by IHE Delft, Institute for Water Education

`www.un-ihe.org`

ISBN 978-90-73445-40-6

To God
To my family
To Pin

Miguel Laverde-Barajas

Contents

Summary

During the past decades, events caused by extreme rainfall events have affected over a billion people and have contributed to 80% of the total global economic damages caused by all-natural hazards. Scientific evidence shows that changes in the rainfall patterns due to climate change affects the distribution, frequency and magnitude of flood events. Therefore, extreme rainfall's spatial and temporal distribution plays a vital role in the study of extremities. However, traditional rain-gauge measurements limit the analysis of extreme rainfall systems at a finer scale. Satellite Precipitation Products (SPPs) are an essential source of information for hydrological applications. Compared to in-situ measurements, satellite products provide spatial variability. However, these rainfall estimates from SPPs are subject to multiple systematic and aleatory errors that are difficult to address.

In many hydrological applications, errors in SPP estimates are compared with pixel values against the corresponding observed data. However, pixel-oriented approaches are unable to capture critical systematic errors due to displacement and timing. Object-based methods become an alternative way to improve SPP data based on the spatiotemporal structure of extreme rainstorm events. This research explores the analysis of the spatiotemporal structure of extreme rainfall events in monsoonal environments. The following objectives aim to understand SPP products' error characteristics in space and time for building a solid concept that contributes to hydrological extremes and their areas of application for error correction. This research is carried out in two monsoonal areas: the catchment of the Tiete river over the sub-catchments of Piracicaba, Capivari and Jundiai Rivers, Brazil, and the Lower Mekong basin over Thailand.

An initial exploration analyzed the importance of extreme events' temporal and spatial characteristics in the performance of SPPs using an analytical approach. This research evaluated the performance of four Near real-time Satellite precipitation products (NRT SPP) to estimate the spatial and temporal characteristics of several extreme rainfall events: local and short duration, long-lasting, short and spatially extended, and long-lasting and spatially extended extreme rainfall. The methodology used an analytical approach to independently define and characterize extreme events' spatial and temporal composition. Results obtained demonstrated the importance of the spatiotemporal component in the performance of NRT SPPs and the limitations of the current rainfall analysis approach to describe the structure of extreme rainfall events.

A new spatiotemporal object-based rainfall analysis method was developed to address the limitations of existing object-based approaches for describing the spatiotemporal structure of extreme rainfall events at the catchment scale. This method, Spa-

tiotemporal Contiguous Object-based Rainfall Analysis (ST-CORA), analyses the structure of rainstorm events at the catchment scale based on the physical features in space and time (volume, maximum rainfall, duration and spatial extension). This methodology is applied over the Tiete River in Brazil to classify different types of extreme events during monsoon seasons and evaluate the performance of SPPs for several rainstorm events. Several comparisons between rainstorm events obtained from an NRT SPP and gauge-corrected weather radar showed new insights to evaluate several sources of error based on the spatiotemporal features of rain objects. These results pointed to the need to evaluate how this satellite-based error for rainstorms could propagate in a hydrological basin.

The evaluation of the hydrological response of two systematic satellite error sources for rainstorm estimation was presented: location in space and time for displacement and magnitude for volume. Using ST-CORA, error sources extracted from the CMORPH SPP were propagated through a distributed hydrological model (Wflow) calibrated over the sub-catchment of the Capivari River in Brazil. Results from different rainstorm types showed that systematic errors sources due to location and magnitude from the SPP affect the shape, phase and amplitude of the streamflow in the catchment.

The development of a spatiotemporal object-based bias correction method for rainstorm detection is presented. This method, ST-CORAbico, compares the spatiotemporal features from SPP and observed datasets to correct errors due to displacement and volume. This methodology is applied over the Mekong basin in Thailand to correct the Multi-satellitE Retrievals for GPM (GPM-IMERG) during monsoon seasons. Results showed a considerable reduction in the bias and the total error of the GPM IMERG SPP data.

A hybrid machine learning/object-based bias correction method for operational purposes is proposed. This methodology combined the ST-CORAbico with a K-nearest neighbour Machine Learning classifier to bias correct rainstorm events in the absence of observed data. This hybrid approach called ST-kNNbico was tested over the Mekong basin in Thailand to correct the monsoon season of 2018 based on historical data from 2014 to 2017. The results presented that ST-kNNbico effectively corrected bias and the total error of the GPM IMERG SPP data without parallel observed data. These results demonstrated the applications of hybrid spatiotemporal bias correction methods for operational purposes.

Overall, this research provides a framework of spatiotemporal error analysis in satellite-based precipitation products. Furthermore, the spatiotemporal object-based rainfall analysis ST-CORA is currently incorporated into the operational rainstorm tracker system for the Lower Mekong Basin (LMB) called Rainstorm Tracker (rainstorms-servir.adpc.net). This system is designed to monitor and alert the severity of rainstorm events over the LMB in near real-time and real-time.

Samenvatting

In de afgelopen decennia hebben rampen die veroorzaakt zijn door regenval meer dan een miljard mensen getroffen en bijgedragen aan 80% van de totale wereldwijde economische schade van natuurrampen. Wetenschappelijk bewijs toont aan dat klimaatverandering de spreiding, frequentie en omvang van neerslagpatronen en daarmee de omvang van overstromingen kan beïnvloeden. De verdeling van regenval over de tijd en ruimte (hierna: spatio-temporal) spelen daarom een cruciale rol in de studie naar extremen. Echter geven traditionele neerslagmetingen een te beperkt beeld van extreme neerslag systemen op een precieze schaal. Op satelliet gebaseerde neerslag informatie, ookwel Satellite Precipitation Products (SPP's) zijn een essentiële informatiebron voor hydrologische toepassingen. Vergeleken met in-situ of lokale metingen bieden satelliet producten inzicht in de ruimtelijke variabiliteit. Echter, neerslag inschattingen van SPP's zijn onderhevig aan meerdere systematische en willekeurige fouten die moeilijk te verhelpen zijn.

Bij veel hydrologische toepassingen worden fouten in SPP-schattingen vergeleken met pixelwaarden tegen overeenkomstige waargenomen gegevens. Deze pixel gerichte benadering kan echter geen kritische systematische fouten vastleggen als gevolg van spatio-temporale veranderingen. Object-based methoden zijn een alternatieve manier om SPP-gegevens te verbeteren op basis van de spatio-temporele verdeling van neerslag. Dit onderzoek is gericht op een analyse van deze spatio-temporele structuur van extreme neerslag in moessongebieden. De doelstellingen zijn bedoeld om fouten van SPP-producten in ruimte en tijd te begrijpen en om een solide concept te bouwen dat bijdraagt aan hydrologische extremen en hun toepassing voor foutcorrectie. Dit onderzoek is uitgevoerd in twee moessongebieden: het stroomgebied van de Tiete-rivier over de deelstroomgebieden van de rivieren Piracicaba, Capivari van Jundiai in Brazilië en het deelstroomgebied in de lagere Mekong bekken in Thailand.

In een eerste verkenning is het belang aangetoond van het analyseren van het spatio-temporele domein bij extreme gebeurtenissen met SPP's met behulp van een analytische benadering. Dit onderzoek heeft de prestaties geëvalueerd van vier 'near real-time' Sattelite Precipitation Products (NRT SPP) om de verschillende spatio-temporele kenmerken van verschillende extreme neerslaggebeurtenissen in te schatten: lokaal en van korte duur, langdurig, kort en een groot oppervlak, langdurig en een groot oppervlak met extreme neerslag. Er is een analytische benadering gebruikt om de ruimtelijke en temporele samenstelling van extreme gebeurtenissen onafhankelijk te definiëren en karakteriseren. De resultaten toonden het belang aan van de ruimtelijke

en temporele componenten in de prestaties van NRT SPP's en de beperkingen van de huidige benadering van neerslag analyse om de structuur te beschrijven.

In dit onderzoek is een nieuwe spatio-temporal object-based neerslag analyse ontwikkeld om de limieten van bestaande object-based aanpakken om spatiotemporele structuren van extreme neerslag gebeurtenissen op stroomgebiedsniveau aan te tonen. De methode, Spatiotemporal Continguous Object-based Rainfall Analysis (ST-CORA) analyseert de structuur van een neerslaggebeurtenis op stroomgebiedsniveau gebaseerd op de fysieke eigenschappen in ruimte en tijd (zoals volume, maximale neerslag, duur van de neerslag en ruimtelijke extensie). De methode is toegepast op het Tiete stroomgebied in Brazilië om verschillende soorten extreme gebeurtenissen in moessonseizoen te classificeren en te evalueren en aan te tonen wat de prestatie van SSP's is voor verschillende neerslag gebeurtenissen. Verschillende vergelijkingen tussen neerslag gebeurtenissen uit een NRT-SSP methode en neerslagmeter-gecorrigeerde radarmetingen hebben nieuwe inzichten in beeld gebracht om verschillende bronnen van fouten in beeld te brengen van neerslaggebeurtenissen in het spatio-temporele domein. De resultaten tonen aan satelliet-gebaseerde fouten in neerslaggebeurtenissen door kunnen werken in een hydrologisch bekken.

Het evalueren van de hydrologische effecten van twee systematische satelliet foutbronnen voor het schatten van regenbuien zijn in dit onderzoek gepresenteerd: de locatie in ruimte en tijd voor verplaatsing van de gebeurtenis en het volume van de gebeurtenis. Met behulp van ST-CORA zijn foutbronnen uit de CMPORPH SPP gehaald door middel van een gedistribueerd hydrologisch model (Wflow) welke gekalibreerd is over het deelstroomgebied van de Capivari rivier in Brazilië. Het resultaat laat zien dat verschillende neerslaggebeurtenissen met systematische fouten in locatie en grootte van de SPP de vorm, fase en amplitude van de stroom in het stroomgebied beïnvloeden.

Een spatio-temporele object-based bias correctiemethode voor neerslag detectie is gepresenteerd. De methode, ST-CORAbico vergelijkt de spatiotemporal kenmerken van SPP en waargenomen datasets naar gecorrigeerde fouten als gevolg van locatie en volume van neerslag. Deze methode is toegepast op het Mekong-bekken in Thailand om de 'multi-satellite retrievals for GPM (GPM-IMERG) tijdens het regenseizoen te corrigeren. De resultaten laten een aanzienlijk verminderde vertekening zien van de totale fout van de GPM IMERG SPP-gegevens.

Een hybride machine-learning/object-based bias correctie methode voor operationeel gebruik is daarnaast voorgesteld. Deze methode combineerd de ST-CORAbico met de K-nearest neighbour Machine Learning classificatie methode om neerslaggebeurtenissen te corrigeren bij de afwezigheid van gemeten gegevens. Deze hybride benadering, ST-kNNbico genaamd, is getest in het Mekong-bekken in Thailand waarmee de effecten van het regenseizoen van 2018 gecorrigeerd zijn op basis van historische data van 2014-2017. De resultaten lieten zien dat ST-kNNbico de fouten van de GPM IMERG SPP effectief corrigeerde zonder parallel gemeten gegevens. Deze resultaten laten het

effect zien van het toepassen van hybride spatio-temporele bias-correctiemethoden voor operationele doeleinden.

Dit onderzoek biedt een raamwerk voor spatio-temporal fouten analyses in satelliet gebaseerde neerslag gebeurtenissen. Daarnaast is de spatio-temporale object-based neerslag analyse ST-CORA inmiddels opgenomen in het operationele neerslag volgsysteem voor de Lagere Mekong Delta (Lower Mekong Delta) genaamd Rainstorm Tracker (rainstorms-servir-adpc.net). Dit systeem is ontworpen om de ernst van neerslag over de delta in bijna real-time en real-time te bewaken en wanneer nodig te waarschuwen.

1

Introduction

1.1. Motivation

Extreme rainfall is one of the principal factors in triggering floods and landslides. During the last decade, events caused by heavy rainfall have affected more people than any other disaster. According to the EM-DAT (International Disaster Database), extreme rainfall events have impacted around 3 billion people, constituting more than 60% of the total global economic damages resulting from disasters over the past 20 years (CRED, 2015). Scientific evidence shows that one of the main consequences of climate change is the intensification of the hydrological cycle (Allen and Ingram, 2002; Held and Soden, 2006; Donat et al., 2016). Changes in the atmospheric temperature have altered precipitation dynamics, leading to greater intensity and frequency of extreme events. These changes are significant for monsoon climates, where extreme weather patterns will have severe consequences for the region's food security, infrastructure, and people's livelihoods.

Understanding the spatiotemporal features of extreme rainfall events are vitally important for effective flood disaster management. However, information on the dynamics of rainstorm events is scarce due to the limited capability of rain gauge measurements in accurately representing the fine-scale details of precipitation systems. Satellite-based precipitation products have been revolutionary in hydrological applications, providing uninterrupted quasi-global rainfall information at a high temporal and spatial resolution. Currently, there is a wide range of satellite-based precipitation products which have arisen from the combination of multiple passive microwaves (PMW) and infrared (IR) sensors carried by Geostationary-Earth-Orbiting (GEO) and Low-Earth-Orbiting (LEO) satellites, respectively (e.g. Huffman et al., 2007; Joyce et al., 2004; Sorooshian et al., 2000). IR sensors assess the rainfall via the cloud top temperature, while PMW sensors analyse the clouds' emission, absorption, and diffusion signals. The combina-

tion of those operational instruments has allowed for greater accuracy in precipitation estimation (Huffman et al., 2007).

Despite the availability of high-resolution rainfall data, numerous researchers have shown that characterising small-scale variability of rainfall patterns is still challenging (e.g. Grayson and Blöschl, 2001). Studies suggest that satellite errors tend to arise from multiple factors such as the measuring devices (e.g. Hu et al., 2016; Qiao et al., 2014), the size of the basin (e.g. Moazami et al., 2013), the climate regimes and seasons (e.g. Thiemig et al., 2012; Sapiano and Arkin, 2009; Mei et al., 2014), and the geographical conditions (e.g. Mei et al., 2016a; Dinku et al., 2007; Guo et al., 2015; AghaKouchak et al., 2011). In the case of rainstorm events, errors are associated with the capability of satellites in detecting heavy rainfall rates at low temporal resolutions (e.g. AghaKouchak et al., 2011; Marra et al., 2017). These uncertainties have limited the operational applications of satellite-based precipitation products at the catchment scale.

1.2. Pixel-based and Object-based methods for extreme rainfall error analysis

Satellite-based precipitation products are traditionally evaluated by comparing the individual pixels of the satellite with the corresponding observed data. Based on this approach, the error is calculated using several numerical and categorical metrics, which include the Root Mean Square Error (RMSE), Bias Correction Coefficient (r), Probability of Detection (POD), False Alarm Ratio (FAR), among others. While these measurements emphasise the performance in terms of accuracy, such metrics do not explicitly estimate errors associated with location, spatial representation and geometric patterns (Baldwin and Kain, 2006; Casati et al., 2008). These error types have serious implications in the accuracy of the hydrologic response with regards to floods (e.g Casse et al., 2015; Bitew and Gebremichael, 2011; Mei et al., 2016b; Vergara et al., 2014; Yilmaz et al., 2005; Li et al., 2009).

In response to the limitations of traditional pixel-based approaches, various researchers have proposed several spatial verification methods to characterise pattern errors in rainstorm prediction (Ebert and McBride, 2000; Davis et al., 2009b; Wernli et al., 2008). These 'non-traditional' methods, in the field of forecast verification, avoid the double penalties from point-to-point matches between the observed and estimated fields (e.g. rainfall estimated but not observed and vice-versa). Methods are broadly grouped into two categories: filtering and displacement Gilleland et al. (2009). Filtering methods evaluate the rainstorm event at a coarser resolution to provide information about the scale of the performance, e.g., (neighbour or fuzzy Roberts (2005); Marsigli et al. (2006) and scale separation, Casati et al. (2004); Casati (2010); Mittermaier (2006)). On other hand, displacement methods are able to evaluate the rainstorm

configuration as an interconnected mass figure (e.g., Object-based Ebert and McBride (2000); Davis et al. (2009b); Wernli et al. (2008) and field transformation methods Keil and Craig (2007, 2009a)).

Among the available spatial verification methods, object-based approaches have been used in hydrology to analyse and correct several systematic errors of rainstorms detected by satellite due to location, rotation, intensity, and shape (e.g. Skok et al., 2009; Li et al., 2015a, 2016). Feature or object-based methods base the rainstorm analysis on the structural properties of the rainstorm fields. Several studies have shown the capabilities of these methods in characterising and correcting displacement errors of high-resolution satellite-based products in areas with extreme conditions, such as monsoonal environments (e.g. Demaria et al., 2011; Li et al., 2014; Le Coz et al., 2019). Despite the applicability of object-based methods for error analysis of satellite-based precipitation products, these methods are constrained by the two-dimensional analysis of rainstorms. This approach presents challenges in describing the temporal and spatial evolution of the rainstorm events, however, extreme rainfall analysis for flood estimation requires a better understanding of the capabilities and limitations of satellites to estimate the temporal and spatial structure of rainstorm events.

1.3. Spatiotemporal perspective to analyse the error in extreme rainfall

Several efforts have been made for characterising the dynamic nature of precipitation systems, especially for Numerical Weather Prediction models events (e.g. Clark et al., 2014; Prein et al., 2017a; Li et al., 2020a; Sellars et al., 2013, 2015; Prein et al., 2017b). Examples in remote sensing data can be found in the PERSIANN CONNected precipitation objECT (CONNECT) Dataset Sellars et al. (2013). This global inventory uses a spatiotemporal object-based approach to describe the four-dimensional properties of Mesoscale Convective Systems (MCS) (e.g., average intensity, starting and ending location (latitude and longitude), duration (h) and event speed (km/h)). These spatiotemporal approaches are generally set up at a regional scale for analysing the climatology of large scale precipitation systems such as typhoons, hurricanes, cold fronts, among others. (e.g. Sellars et al., 2015; Prein et al., 2017b). At catchment scale, spatiotemporal methods have the disadvantage of false merging small precipitation systems during the segmentation process (Chang et al., 2016). These storm events are significant in tropical monsoonal climates, where cell convective storms are responsible for extreme flash flooding causing major human and economic damage.

In general, spatiotemporal object-based methods provide a unique perspective of the physically-based features of the storm events. Based on this analysis, errors due to displacement in space and time between the satellite and the observations can be identified while providing object properties (e.g. orientation, magnitude, duration,

extension) as additional attributes for evaluation. However, the evaluation of these systematic errors is limited by the use of subjective approaches in which the performance of precipitation datasets can be evaluated by comparing the storms features against reference data (e.g. Clark et al., 2014; Pinto et al., 2015). The lack of characterisation of satellites' error for storm detection brings several challenges in reducing systematic error in satellite-based products. Developing a method to efficiently reduce the error in satellite incorporating the elements of the spatiotemporal analysis is one of the significant challenges for an accurate storm prediction, especially for operational applications.

1.4. Research questions

This research aims to analyse the spatiotemporal error dynamics of storm events estimate by satellite. This objective is to integrate the development of a methodology to analyse the spatiotemporal structure of storm events at catchment scale and a method for reducing systematic error in satellites.

To achieve this aim, we formulate the following research questions:

1. How vital is the spatiotemporal dynamics of extreme rainfall events in satellite performance regarding the error?
2. How can spatiotemporal rainstorm dynamics be integrated into error estimation?
3. What is the effect of spatiotemporal errors on the hydrological response?
4. How can spatiotemporal error information be used to improve bias correction of satellite data?
5. How can artificial intelligence be used to improve spatiotemporal bias correction methods in operational applications?

1.5. Scientific relevance

The accuracy of satellite-based products for rainstorm prediction is crucial for many meteorological and hydrological applications. For instance, systematic errors in precipitation measurements can lead to overestimating or underestimating the flood prediction in the catchment. Similarly errors associated with intensity, satellite-based estimates also contain errors due to displacement and time. However, the analysis and reduction in the types of error are poorly studied. This dissertation bridges this gap by developing a methodology to evaluate satellites' spatiotemporal characteristics of rainstorm events. In this methodology, components of pattern recognition approaches are implemented in a multidimensional framework to identify the primary sources of systematic errors in satellites. Based on this physically-based approach, gaps in bias

correction schemes are also addressed by developing a spatiotemporal method to reduce several systematic errors in rainstorm prediction detected by satellites. This approach incorporates machine learning technologies for correcting satellite-based products used in operational applications.

1.6. Scientific contributions

The scientific contributions can be summarised as follows:

- A new method (M_1) is developed for analysing the **spatiotemporal structure of rainstorm events** at catchment scale. This method is implemented over several monsoonal environments to analyse the regional climatology of the rainstorm events and evaluate the accuracy of satellite-based precipitation products. Further developments address the limitations in object-based approaches by incorporating a probabilistic rainstorm segmentation algorithm. This approach overcomes the model sensitivity for segmenting sub-rainstorm events inside large convective systems.

- A new **bias correction method** (M_2) is developed to reduce several sources of systematic error in satellites for rainstorm prediction. This method uses the spatiotemporal features of the rainstorm from M_1 to correct two types of error: displacement in space and time and volume. The limitations for applications in disaster management are discussed.

- A **machine learning implementation** (M_3) of the bias correction method M_2 was developed to reduce error due to displacement and volume in the absence of measured data. This method addresses the use of object-based bias correction methods for operational applications.

- An in-depth **hydrological study** was conducted, aimed at evaluating the hydrological impact of several sources of error in satellites for rainstorm event estimation. Based on the spatiotemporal error analysis obtained by M_1, location and magnitude error sources are subtracted from a near-real-time satellite-based precipitation product and used as forcing for a distributed hydrological model. Results revealed that error sources in satellites due to location and magnitude affect different components of the catchment streamflow. The main source of this error is conditioned by different types of rainstorm events.

The methods developed in this research were applied to satellite-based precipitation products. However, multiple grid-based measurements and weather model outputs can also be incorporated to verify the accuracy of rainstorm prediction at the catchment level.

A better analysis of the physically-based features of rainstorm events detected by satellites creates new opportunities for supporting disaster management strategies associated with extreme rainfall. The further development of a system based on the spatiotemporal object-based rainfall analysis to improve the decision-making process of flash flood warning systems in the Lower Mekong Basin have been undertaken. This system, called 'Rainstorm Tracker', is being developed under the SERVIR-Mekong program to monitor and evaluate the severity of rainstorm events with the potential to trigger flash floods. Preliminary results and more information about the project is available at the following link: `http://rainstorms-servir.adpc.net`.

1.7. Outline

The outline of the dissertation is divided into nine chapters as follows:

Chapter 2 reviews the theoretical background of the research based on the analysis of three components. First, remote sensing technologies for measuring precipitation are analysed, describing the most common satellite products used in hydrology. The second component reviews existing methodologies for evaluating errors for rainstorm prediction and new approaches using object-based methods. The third component presents a brief description of the bias correction methods used to reduce systematic errors in rainfall.

Chapter 3 presents a general description of the two monsoonal study areas used for the development of the proposed methods: The tropical catchment of the Tiete river in Brazil and the Lower Mekong Basin in Thailand (Isan area). This chapter explores the physical, geographical and atmospheric characteristics of the study areas and discusses the factors that make the areas prone to extreme rainfall. In the Tiete river, the area is analysed according to three different domains: the first domain corresponds to the sub-catchments of the rivers, Piracicaba, Jundiai, and Capivari used in chapter 4; the second domain is the catchment of Tiete river used for the development of method M_1 in chapter 5, and finally the sub-catchment of the Capivari river used for the hydrological study presented in chapter 6.

Chapter 4 deals with an evaluation of different near-real-time satellite-based precipitation products for extreme rainfall prediction using an analytical approach. The advantages and disadvantages of current methods to analyse rainstorms are discussed, and the opportunities to develop new methodologies for spatiotemporal rainstorm analysis.

Chapter 5 describes the development of the method to analyse the spatiotemporal dynamics of rainstorm events M_1. Two applications of this method are shown to classify different types of rainstorms and validate a satellite-based product for rainstorm events.

Chapter 6 presents the use of M_1 to evaluate the hydrological impact of two systematic error sources: location and magnitude. For this aim, the study uses the

distributed hydrological model wflow developed by Deltares to represent the rainfall-runoff model of the Capivari river. The model calibration and validation are presented as well as the results obtained from the evaluation.

Based on M_1 and the error analysis previously described, **Chapter 7** introduces the development of the spatiotemporal object-based bias correction method, M_2, for the reduction of error due to displacement and volume. This method is applied over the Lower Mekong Basin in Thailand to correct the near real-time satellite-based product, GPM IMERG, from NASA.

Chapter 8 describes the use of M_2 for correcting rainstorm events in the absence of in-situ data. The method, M_3, uses a machine learning approach to identify matches in historical rainstorm events to correct the events detected by satellite-based GPM-IMERG. The limitations and prospects for operational purposes are discussed.

Chapter 9 summarises the main findings, existing challenges, limitations and further developments. In addition, chapter 9 details an additional implementation use-case of M_2 under the SERVIR-Mekong program, which supports decision-making regarding flash flood warning systems in the Lower Mekong Basin.

2

Theoretical background: Precipitation measured from satellite

This chapter presents the theoretical background of the remote sensing technologies used to measure rainfall and reviews existing methods to evaluate and correct systematic errors in satellites for rainstorm prediction. The first section evaluates the existing rainfall analysis methodologies, particularly satellite-based approaches for near-real-time (NRT) rainfall prediction. The most common NRT satellite-based datasets used in hydrology are also outlined in this section. The second section reviews existing methodologies for evaluating errors in rainstorm prediction. Spatial object-based verification methods are discussed as well as new approaches that incorporate the temporal dimension of rainstorm analysis. In the final section, methodologies for rainfall bias correction are presented. Limitations and opportunities for operational purposes are also discussed.

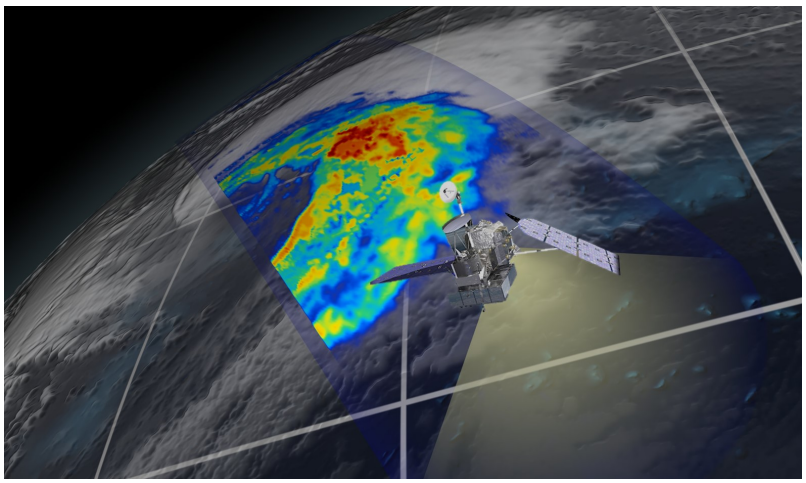


Figure 2.1: *GPM satellite mission architecture. NASA (2014)*

2.1. Introduction

An accurate observation of rainfall is a key factor to describe the water cycle process over land. A study suggests that the spatiotemporal dynamics of measured rainfall can explain almost 70-80% of the terrestrial water cycle variability (Syed et al., 2004). A wide range of instruments is available to measure rainfall with different spatial and temporal scales. Ground monitoring networks (e.g., rain gauges and weather radars) are traditionally used in many hydrological early warning systems. For instance, rain gauges are relatively cheap, easy to maintain, and offer a direct estimation of rainfall at a single point (Sinclair and Pegram, 2005). On the other hand, weather radars provide a continuous estimation of rainfall with high spatial and temporal resolution at the regional scale (Maggioni et al., 2016a). However, such data is affected by limitations in the precipitation measurement. Rain gauges are prone to human errors during the data acquisition process. In addition, the spatial representation depends on the network density and assumptions in the interpolation techniques. Limitations in the radar-based methods involve errors in the reflectivity rain rate (Z–R) relationship, problems with ground clutter, signal attenuation and beam blockage, among others (e.g. Hasan et al., 2014; Kirstetter et al., 2012; Liu et al., 2015; Uijlenhoet and Berne, 2008).

Over the past decades, earth observation systems have become crucial in estimating rainfall in catchments (e.g AghaKouchak and Nakhjiri, 2012; Azarderakhsh et al., 2011; Kidd et al., 2009; Pan et al., 2010). Satellite-based precipitation products (SPPs) provide uninterrupted quasi-global information with high spatial and temporal resolution. The use of these products for estimating rainfall in extreme conditions presents new opportunities to support disaster management on a global scale. This is especially the case in poorly gauged basins, where SPPs might be the only data source for predicting flow downstream with enough lead time to initiate a response Maggioni and Massari (2018).

Despite multiple advances in satellite rainfall monitoring, products are still subject to errors from multiple sources (e.g. Hu et al., 2016; Qiao et al., 2014; Thiemig et al., 2012; Mei et al., 2016a). With extreme rainfall estimations, SPPs still present numerous challenges resulting from errors in detecting the magnitude, the position and the timing of the rainstorm event. Errors due to displacement and timing are crucial in evaluating the impact of flow in a catchment. However, these errors are poorly evaluated. Spatial verification methods have recently been used in hydrology to analyse the structure of rainstorm events. Defining a rainstorm as an object presents new opportunities to evaluate the complete structure of rainstorm events in both space and time.

2.2. Satellite precipitation measurement

Satellite precipitation products are derived from a large constellation of operational sensors, including Low-Earth-Orbiting (LEO) and Geostationary-Earth-Orbiting (GEO) satellites, as well as from observing spectral ranges (visible (VIS), infrared (IR), Passive Microwave (PMW) and Active Microwave (AMW)). Retrieval techniques from satellite methods present several strengths and weaknesses for retrieving precipitation above the surface. For instance, VIRS and IR sensors carried by GEO satellites provide helpful information about the characteristics of the cloud brightness temperature observed from space with a rapid temporal update cycle (e.g. 30 mins or lower). However, cloud top temperature and height measurements are not always well correlated with surface rainfall. Unlike VIS and IR sensors, PMW and AMW sensors from LEO satellites can penetrate clouds, providing a more direct physical connection with the raindrops, snow and ice. Despite the low temporal and spatial resolution, this direct measurement of the hydrometeor profiles allows for quantitative precipitation detection in the lower atmosphere and the surface. Table 2.1 summarises the different methods to retrieve rainfall from satellites.

Table 2.1: *Summary of satellite methods for the retrieval of rainfall, modified from Ferraro and Smith (2013)*

Observation Spectrum	Satellite type	Sensor example	Advantages	Caveats
Visible (VIS)	GEO LEO	GOES, Imager AVHRR	Cloud type, Cloud evolution	Cloud tops, Indirect rain rate.
Infrared (IR)	GEO LEO	GOES, Imager AVHRR	Cloud temperature Cloud evolution	Cirrus contamination, Indirect rain rate.
Passive Microwave (PMW)	LEO	SSM/I. AMSR-R, TMI, GPM	Direct measure of rain, especially over oceans	limited temporal and spatial resolution, Indirect rain rate in land.
Active Microwave (AMW)	LEO	TRMM PR, CloudSat, CPR, GPM	Direct measure of vertical structure of rain	Narrow swath width, limited temporal resolution, Rain rate sensitivity/saturation.

2.2.1. VIS/IR based satellite precipitation products

Multiple precipitation products have been developed to observe cloud physics for measuring rainfall. SPPs based solely on VIS/IR sensors are characterised by a high temporal and spatial resolution with low latency. One of the most common IR-based

SPPs is the Precipitation Estimation from Remotely Sensed Information using Artificial Neural Networks (PERSIANN) (Sorooshian et al., 2000). This product uses a cloud classification procedure using an Artificial Neural Network (ANN) to derive rainfall rates at 0.04° each half-hour. Input data from PERSIANN are brightness temperature measurements obtained from IR sensors of the NOAA's Geostationary Operational Environmental Satellites (GOES), Meteosat 5 and 7 geostationary meteorological satellites (GMS). PERSIANN data includes three different versions, namely PERSIANN, PERSIANN-CDR and PERSIANN-CCS. PERSIANN and PERSIANN-CDR (Climate Data Record) incorporate quality controlled input data to derive rainfall available at two days and three months latency, respectively. Meanwhile, PERSIANN-CCS (Cloud Classification System) is available in near real-time with a latency of 1 hour. Recently Nguyen et al. (2020) presented the quasi-real-time PERSIAN Dynamic Infrared Rain Rate (PDIR-Now) dataset. This new dataset aims to improve the PERSIANN-CCS dataset providing hourly precipitation estimates at 0.04° × 0.04° spatial resolution with a short latency (15–60 min). This short latency makes it suitable for operational monitoring in many hydrometeorological applications.

Similar to PERSIANN-CCS, other near-real-time SPP approaches have been developed using VIS/IR sensors to support early warning systems and heavy forecast efforts. The Hydro-Estimator (Hydro) developed by Scofield and Kuligowski (2003) uses the infrared cloud top brightness temperature information from GEOS as primary information to estimate rainfall rates with 0.04° spatial resolution every 15 minutes. The methodology is based on the automatic precipitation estimation algorithm, NESDIS Vicente et al. (1998). Non-detected factors in the IR sensor such as moisture availability, evaporation, orographic modulation and thermodynamic profile effects are corrected using atmospheric information from the Global Forecast System (GFS). Due to the high resolution and low latency (10 min), the hydro-estimator has been widely used as the primary input in many flash flood early warning systems around the world, including the CAFFG System in Central America, the TerraMA2 in Brazil and The Flash Flood Guidance System (FFGS) implemented in more than 65 countries around the world.

2.2.2. IR/PMW based precipitation products

Multi-Sensor Global Precipitation products use the superior sampling obtained from GEO satellites combined with the vertical profile measurements of the hydrometeor from LEO satellites to create a more reliable rainfall dataset. The advantages and disadvantages of VIS/IR sensors discussed so far can be optimised by combining the GEO-IR estimates with LEO-PMW sensors. Several techniques have been developed to merge IR and PMW measurements optimally. For example, the Tropical Rainfall Measuring Mission (TRMM) Multi-satellite Precipitation Analysis (TMPA) developed by Huffman et al. (2007) uses PMW data from the TRMM mission and other sensors

(SSM/I, AMSU, MHS) to calibrate IR estimates from GEO satellites by extrapolating rainfall measurements when the PMW data is not available. This combination increases the temporal sampling up to 3 hours with 0.25° x 0.25° spatial resolution and improves the performance compared to IR-based measurements. The latency of TMPA near-real-time (TMPA-RT) version 7* ranges between 6 and 9 hours. After two months, a scientific version of the TMPA data is released, incorporating rain gauge data to correct the SPP data.

The Climate Prediction Centre Morphing Method (CMORPH) developed by Joyce et al. (2004) uses a different approach to combine PMW and IR data for estimating rainfall. This SPP data uses IR data only to generate the atmospheric motion vector of two successive IR images. Then, the generated motion vectors are used to propagate the rainfall fields from diverse PWM data such as the TMI, SSM/I, AMSU-B and AMSR-E. This technique, also known as the MORPHING technique, provides high-resolution data at 0.0727°(8km) spatial resolution generated every 30 minutes. Version 1.0 of CMORPH is available in three different versions: raw CMORPH (CMORPH-RAW); CMORPH bias-corrected (CMORPH-CRT); and a gauge-satellite blended product, CMORPH-BLD. CMORPH in near real-time corresponds to the raw version and is available up to 4 hours after the measurement.

The deployment of more robust satellite missions has presented an opportunity to develop multiple SPP datasets. This is the case of the NASA/JAXA Global Precipitation Mission (GPM). This mission was launched in 2014 upon the success of the TRMM mission after more than 20 years of operational service. The GPM mission is composed of an international network of satellites for the global observation of rain and snow (Figure 2.2). The Core Observatory of GPM consists of an advanced radar/radiometer system that extends the capability to measure light rain (< 0.5 mm hr-1), solid precipitation and the microphysical characteristics of precipitating particles (NASA, 2014).

Several techniques have been developed under the GPM mission to retrieve precipitation rates. The Integrated Multi-satellitE Retrievals for GPM (GPM-IMERG), developed by the US GPM Science Team (Huffman et al., 2015), combines information from multiple infrared, passive-microwave, and satellite-radar sensors to provide rainfall estimations at 0.1-degree spatial resolution every half-hour. GPM-IMERG computes Early, Late, and Final runs. The first two runs are near real-time versions of IMERG and are available at 4-6 hours and 18 hours latency, respectively. In the Early version, rainfall estimations are propagated forward while the Late has both forward and backward propagation allowing the incorporation of climatological gauge data. The final version is obtained three months after the measurements. In this run, GPM-IMERG ingests the monthly rainfall analysis from the Global Precipitation Climatology Centre (GPCC; Schneider et al. (2008)). This version is used for scientific purposes, as it is considered to be the most reliable version (Huffman et al., 2015).

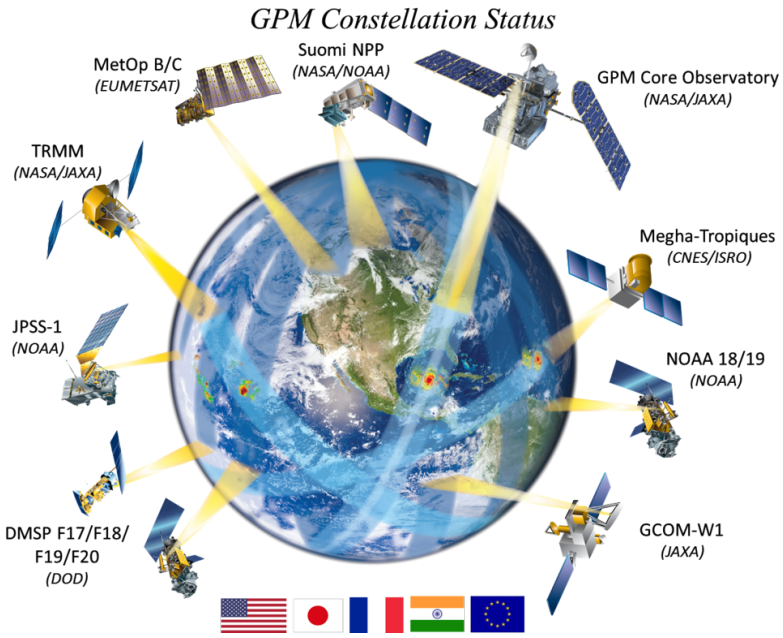


Figure 2.2: *GPM constellation program. Credits: NASA (2015)*

On the other hand, the GPM Science team from the Japan Aerospace Exploration Agency (JAXA) developed the Global Satellite Mapping of Precipitation (GSMaP; Aonashi et al. (2009)). This product was created under the GPM mission, combining multiple IR and PWM sensors to derive half-hour precipitation quasi global maps at 0.1 x 0.1 spatial resolution. The methodology uses a similar approach to the one developed in CMORPH, in which IR sequential images are used to derive the motion field. However, a Kalman filter approach is used as a blending methodology. GSMaP product is available as standard (GSMAp MVK) and gauge-adjusted product (GSMaP Gauge), near real-time (GSMaP NRT) and near real-time adjusted (GSMaP Gauge NRT), real-time product (GSMaP NOW) and reanalysis product (GSMaP Gauge RNL). The latency of the GSMaP products is three days for the standard product and up to half an hour for the real-time version. Using GsMAP, nowcasting approaches have been created at the quasi global scale. The GSMaP RIKEN (GSMaP RNC) developed by (Otsuka et al., 2019) is an extrapolation-based nowcasting system designed to ingest GSMaP for predicting rainfall up to 6 hours in advance. The methodology applies the motion vector technique to sequential GSMaP NRT images, and then an ensemble Kalman filter is used to blend the nowcast rainfall data in real-time. GSMaP RNC has been available since 2016 with 0.1 x 0.1 degrees spatial resolution, updated hourly.

2.2.3. Validation of SPP

Benchmarking SPPs against ground-based measurements are an important requirement for validating the rainfall retrieval in a particular area. A large number of studies have focused on comparing the performance of SPPs in different environments and climates

(e.g. Sapiano and Arkin, 2009; Ebert et al., 2007; Guo et al., 2015; Dinku et al., 2010). Maggioni et al. (2016b) presents a comprehensive review of SPP validation studies across different regions around the world. At the same time, Sun et al. (2018) analysed the performance of SPPs around the world compared with gauge and reanalysis rainfall datasets. Overall, SPPs have an important advantage in providing precipitation data with a high temporal and spatial resolution (almost) anytime and (almost) everywhere. This is especially important in more remote, inaccessible areas such as oceans, mountains, forests and deserts.

However, due to the indirect nature of measurements, SPPs are affected by several errors and uncertainties due to several factors, including the topography, seasonality, climatology, and sensor type. For example, several studies have shown that SPPs present several challenges when detecting rain in complex areas where the precipitation is dominated by the topography and the high variability of rainfall fields (Derin and Yilmaz, 2014). In terms of seasonality and climatology characteristics, several studies showed a good performance in warm conditions and over tropical and equatorial regions (e.g. Ebert et al., 2007; Dinku et al., 2010), contrasted by a weak detection in semiarid areas and winter conditions. (e.g. Kidd et al., 2012; Peña-Arancibia et al., 2013). The type of used sensor retrieval plays an important role in the performance of the SPPs. IR sensors generally miss light precipitation and orographic events, while PWM retrieval algorithms typically underestimate orographic precipitation, especially in the cold seasons (Derin and Yilmaz, 2014).

2.2.4. Extreme precipitation and hydrologic validation

The performance of SPPs in extreme conditions plays an important role in areas of hydrology such as flood monitoring and risk management. Capturing rainstorm events from extreme rainfall requires synchronicity between the event and the time that the satellite overpasses (Prat and Barros, 2010). Many studies have used the growing length of satellite records to analyse the statistical distribution of extreme long-term rainfall. Prat and Nelson (2020) described recent studies on the long-term evaluation of SPPs. 'Extreme' has several definitions, including being defined according to percentiles (75th, 90th, 95th and above) (e.g. AghaKouchak et al., 2011; Boers et al., 2013; Ringard et al., 2015), as well as thresholds for daily rainfall (above 20mm, 500mm, 100mm) (e.g. Tan and Santo, 2018; Gao and Liu, 2013; Miao et al., 2015). These studies conclude that there is no ideal SPPs for extreme prediction. SPPs tend to underestimate extreme precipitation values with a lower performance at the higher threshold or percentile values AghaKouchak et al. (2011); Gao and Liu (2013); Breña-Naranjo et al. (2015).

Short-term extreme event prediction, commonly associated with hydrometeorological extremes, represent a real challenge for SPP techniques. In operational applications, an accurate representation of those events demands the description of both location and spatial extension at high temporal and spatial resolutions. This is particularly

important in monsoonal environments, where convective rainstorm events are highly localised during a short period. Several studies have evaluated the ability of SPPs to represent the dynamics of extreme events at several temporal and spatial samplings. For example, Mehran and AghaKouchak (2014) evaluated the temporal capability of three NRT SPPs to estimate heavy precipitation rates over the United Stated at different temporal accumulations (3, 6, 12 and 24 h). This study showed that at high temporal resolutions (3h), none of the SPPs was able to represent heavy rainfall accurately. However, the performance increases at higher temporal accumulations. On the other hand, Habib et al. (2012) evaluated the spatial sampling of CMORPH for detecting more than 130 rainfall events in southern Louisiana, United States. The results showed that events estimated by satellites contain higher levels of random errors at finer resolutions, reducing over coarser resolutions. While temporal and spatial sampling can result in errors, the performance of the SPPs are directly related to the type of precipitation system.

In addition to the evaluation of SPPs from an atmospheric perspective, extreme validation includes evaluating the hydrological response in terms of the flood. Hydrological validation focuses on examining the ability of SPPs to predict floods in a particular catchment. This process requires a hydrological model that faithfully captures the runoff generation and the streamflow interaction. Several studies have evaluated the performance of SPPs used in hydrology at different scales: a global (Revilla-Romero et al., 2015; Dottori et al., 2016; Beck et al., 2017), mesoscale and large-scale river basins (Falck et al., 2015; Mazzoleni et al., 2019; Dembélé et al., 2020) or a medium and small catchment scale (Li et al., 2009; Casse et al., 2015; Bitew and Gebremichael, 2011; Mei et al., 2016b; Vergara et al., 2014). In a review presented by (Maggioni and Massari, 2018), it is argued that the bias in the forcing precipitation varies from minor to major issues in the flood prediction due to multiple factors including, the type of SRP sensor; the precipitation type; the geomorphological conditions; and the hydrological model formulation. The implementation of SPPs in operational hydrological applications requires quick observation updates that consider additional rainfall information to correct the rainfall data and reduce the inherent bias of SPP datasets (Serrat-Capdevila et al., 2014; Maggioni and Massari, 2018).

2.3. Error analysis in satellite precipitation data

The International Precipitation Working Group (IPWG)[1] is an international scientific community created to promote standards for satellite precipitation measurement and successive validation and verification of SPP datasets (Kidd et al., 2008). Based on an extensive set of surface reference datasets (radar and rain gauge) distributed in several

[1]The International Precipitation Working Group (IPWG) was endorsed during the 52nd session of the WMO Executive Council in 2000, which encouraged the Coordination Group for Meteorological Satellites (CGMS) to participate in the formation of the IPWG

areas worldwide, the IPWG validation and verification program provides a visual and statistical analysis of daily satellite/model precipitation data in near real-time. These statistics involve standard (bias, Root Mean Square Error (RMSE); correlation coefficient) and categorical (probability of Detection (POD); False Alarm Ratio (FAR); Critical Success Index (CSI)) error score metrics. While these measurements provide helpful information about the accuracy and skill of SPPs in a clear, simple way, metrics are limited due to the skewness of the precipitation intensities (Kidd et al., 2020). In extreme rainfall detection, the point-to-pixel comparison of standard and categorical scores do not consider important intrinsic structural features of rainstorm event data such as location, spatial representation and geometrical patterns (Baldwin and Kain, 2006; Wilks, 2011). For instance, SPP rainstorm events with corrected size and structure can be over-penalised by slight displacement in space. This is known as the "double penalty".

In response to the limitations of "traditional" verification methods, spatial verification metrics, emerging from weather forecasting, have become an alternative way to evaluate rainfall data in extreme conditions (e.g. Skok et al., 2009; Li et al., 2015a, 2016). This approach considers the spatial features of rainfall fields to evaluate the performance of satellites. Errors can be analysed in physical terms based on their nature (e.g. displacement, volume, and pattern). Spatial methods can be broadly grouped into neighbour or fuzzy (Roberts, 2005; Marsigli et al., 2006), scale separation (Casati et al., 2004; Casati, 2010; Mittermaier, 2006), object-based e.g., Ebert and McBride (2000); Davis et al. (2009b); Wernli et al. (2008), and field transformation (Keil and Craig, 2007, 2009a). The first two categories can be described as spatial filtering methods, in which the verification statistics are evaluated at coarser resolutions to provide information about the scale of the performance. Object-based and field transformation are considered displacement verification methods when estimated rainfall fields, defined as an object, are spatially manipulated (displacement, rotation, scaling, among others) to try and fit the observed value.

2.3.1. Object-Based Analysis

In nature, an object can be identified by summarising all intrinsic characteristics. A plant, for instance, could be identified in an image by observing characteristics such as colour, size, shape, among others. Similar to our eyes registering all this information to decide what type of plant we are observing, object-based methods analyse the information of a physical process to understand its characteristics and examine their relationships. Object-based or object-oriented analysis is defined as the study of the statistics of the population of objects. The analysis could be incorporated into a group of images or gridded data to describe an object with low computational complexity (Blaschke et al., 2004). In the literature, object-based methods have been used for many purposes, from investigating terrain morphology and detecting its changes

(Blaschke, 2010; Desclée et al., 2006; Yu et al., 2010; Amatya et al., 2021) to evaluating the performance of climatic models (Davis et al., 2009b; Grams et al., 2006). In atmospheric sciences, object-based models are widely used for monitoring and nowcasting precipitation systems (Dixon and Wiener, 1993; Han et al., 2009; Johnson et al., 1998) and the spatial verification of weather products (Ahijevych et al., 2009; Brown et al., 2004; Davis et al., 2009b; Ebert and McBride, 2000; Li et al., 2020b). In the latter case, multiple spatial verification methods have been proposed to evaluate model accuracy in terms of spatial pattern, intensity, and displacement, such as the contiguous rain area (CRA) (Fig. 2.3) (Ebert and McBride, 2000), object-based diagnostic evaluation (MODE) (Davis et al., 2009b), and structure, amplitude, and location (SAL) (Wernli et al., 2008). Gilleland et al. (2009) discussed the capabilities of spatial object-based verification methods in proving information on structure errors of rainfall estimation at multiple scales.

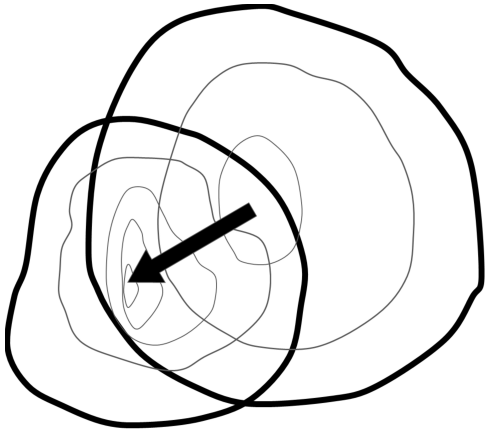


Figure 2.3: *Schematic diagram of a contiguous rain area. The arrow shows the level of displacement between the observed and predicted rainfall fields. Modified from Ebert et al. (2007)*

Advances in high-resolution satellite rainfall detection have led to deploying these methods for evaluating the accuracy of satellite rainfall products at large scales. For instance, Skok et al. (2009) implemented the MODE method to analyse the properties and spatial distribution of the rainfall systems from PERSIAN and TMPA version 7 over the equatorial Pacific Ocean. Additionally, Demaria et al. (2011) used the CRA method to identify systematic errors in South America using TMPA, PERSIANN, and CMORPH. Li et al. (2015a) developed an object-based approach to validate satellite-based rainfall products against ground observations. This method was later applied by Li et al. (2016) for evaluating three high-resolution satellite precipitation products (PERSIANN, CMORPH, and TMPA) in the United States. One of the main advantages of these methods compared to traditional gridded-based approaches was incorporating new information about errors in shape, orientation, and displacement into rainfall analysis.

2.3.2. Spatiotemporal Object-Based Methods

Object-based methods have in common a 2D approach to identify rainfall objects in space (latitude, longitude). The evolution of rainfall systems in time is analysed simply through a connectivity or "tracking" function (e.g., overlapping, centroid displacement). Once the attributes of two pairs of consecutive objects are defined in space, the function identifies the likelihood that both rainfall systems belong to each other (Fig. 2.4). However, rainfall systems are often characterised by evolution in time as well as spatial structure (Davis et al., 2009a). Integrating the temporal dimension into the object-based analysis marks a fundamental challenge to rainfall systems analysis.

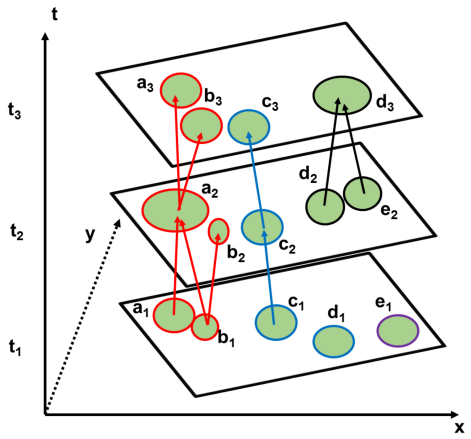


Figure 2.4: *Rainstorm event tracking based on centroids. The arrows represent vector motions of rainfall fields a, b, c, d, and e in time steps t1, t2, t3. Colours describe the rainstorm object geometry. Modified from Liu et al. (2016)*

Spatiotemporal object-based methods are currently developed to identify and compare the temporal and spatial evolution of rainfall systems. Advances in "big data" analysis have given a new perspective to analyse rainfall systems in space and time. This new approach transforms gridded rainfall values into a spatiotemporal dimensional grid (latitude, longitude, time) as volumetric pixels or "voxels." The algorithm identifies objects in space and time connected by a common attribute to obtain a 4D object (latitude, longitude, time, and intensity). Several features can be identified from the object type, such as volume area, duration, average speed, and centroid, among others.

In the literature, there are a couple of examples of this method. For example, Sellars et al. (2013) created PERSIANN-CONNECT, a 4D rainfall dataset created from the high-resolution PERSIANN satellite-based rainfall data to identify the characteristics of rainfall systems at a large scale. This dataset is stored in a PostgreSQL database and can be easily consulted at http://connect.eng.uci.edu/. Another example developed by this approach is the spatiotemporal verification method MODE Time-Domain (TD) (Davis et al., 2009a), an object-based verification method created to

evaluate the forecast from high-resolution numerical weather prediction (NWP) models. MODE-TD follows the same methodology of the spatial verification MODE but incorporates temporal evolution in the precipitation system. This method has been used in diverse atmospheric applications (Clark et al., 2014; Prein et al., 2017b; Ayat et al., 2021).Mittermaier and Bullock (2013) compared MODE and MODE-TD methods over the United Kingdom for evaluating the spatial and temporal characteristics of cloud cover forecast from km-scale NWP models. This comparison showed notable differences between the two methods. The inclusion of the time dimension provided a different perspective on the verification; however, the choice of method-specific parameters plays a critical role in identifying rainfall objects over time.

2.4. Error correction methods

The inherent bias of satellite products can be reduced by combining the SPP-based data with ground measurements. It is shown that bias correction methods significantly improve the SPP-forced hydrologic model performance in several regions around the world (Serrat-Capdevila et al., 2014). For instance, products such as the Climate Hazards Group Infrared Precipitation with Station data (CHIRPS, Funk et al. (2015)) or Multi-Source Weighted-Ensemble Precipitation (MSWEP, Beck et al. (2016)) databases demonstrate that merging multiple satellites, reanalysis, and gauge-based rainfall dataset strengthens the performance of each product (Beck et al., 2017).

Bias correction approaches are designed to use ground-based rainfall measurements to reduce the systematic bias of SPP data and other precipitation datasets, including Global Circulation Models (GCM) and Regional Climate Models (RCM). Multiple bias correction methods have been developed for precipitation data, including simple linear and scaling methods, probabilistic distribution maps and machine learning approaches. Other methods use physically-based approaches to correct rainfall data in extreme conditions. The characteristics of these methods are described below.

2.4.1. Linear and scaling methods

These methods correct the mean bias of SPPs using the ratio factor (linear or scaling) calculated from the relationship between the SPP and the ground-based reference data (Lenderink et al., 2007; Teutschbein and Seibert, 2012; Fang et al., 2015). This additive or multiplicative factor is applied uniformly or spatially distributed (using a spatial interpolator) over the rainfall field. The simplicity of these methods allows the incorporation of additional components (e.g. sequential time windows and topographic clusters) to analyse temporal and orographic effects in the error correction (Gumindoga et al., 2019). However, these methods do not correct the bias in the standard deviation.

2.4.2. Probabilistic methods

Probabilistic distribution mapping methods correct the mean and the standard deviation of the precipitation data by matching the distribution of the satellite for the ground data. Methods such as the Distribution Transformation (DT) (Bouwer et al., 2004) or the Gamma Quantile mapping (GQM) (Piani et al., 2010) have the flexibility to correct data without parallel reference data. This process facilitates the analysis in future rainfall simulations (e.g. for climates projections in GCM and RCM) or when ground data is not yet available (e.g.in real-time monitoring in SPPs), Yuan et al. (2019); Tong et al. (2020). While these methods have been successfully applied in multiple fields, errors are corrected just as a function of the rainfall intensity values. Other approaches use multivariate distribution techniques such as N-dimensional distribution methods (e.g. Dekens et al., 2017; Cannon, 2016, 2018) or Copulas Functions (e.g. Vrac and Friederichs, 2015; Kim et al., 2019) to correct the distribution and the spatial and temporal dependence of the rainfall data. However, the quality of the bias correction is limited by the referenced data as well as the physical process described by the SPP Hempel et al. (2013).

2.4.3. Machine Learning methods

Machine learning methods have been relatively recently explored in correcting the systematic error in SPPs. Machine learning technology has been revolutionary in many fields of remote sensing, including land-use change detection (Cao et al., 2019), flood monitoring (Corzo and Solomatine, 2007; Kişi, 2007; Zhao et al., 2019), precipitation nowcasting (Moon et al., 2019), among others. Methods such as Support Vector Regressions, K-near Neighbours, Decision Trees, Artificial Neural Networks have been implemented over several types of rainfall data sets against ground data to reduce the systematic errors in the measurement (e.g. Moghim and Bras, 2017; Kolluru et al., 2020; Chaudhary and Dhanya, 2020; Le et al., 2020). Based on their flexibility to ingest large amounts of data, ML-based bias correction approaches can handle multiple physiography and atmospheric variables to create SPP merged products. For example, the Machine Learning–based Precipitation Ensemble Technique (MLPET) proposed by Bhuiyan et al. (2019) combines daily soil moisture, terrain elevation, and atmospheric variables (temperature and humidity) with multiple global precipitation data sets.

Advanced deep neural network (DNN) approaches can directly analyse complex relationships of raw satellite data when correcting errors. For instance, Tao et al. (2016) use a Stacked Denoising Autoencoder (SDAE) to correct bias and false alarms of the PERSIANN-CCS data over the central United States. SDAE uses the cloud imagery of the IR satellites to reproduce the estimated difference between the SPP and the ground data and then reduce it from the original data. Recently (Le et al., 2020) applied a 2-dimensional convolutional neural network called the convolutional autoencoder (Con-

vAE) to correct pixel-by pixel satellite data in space and time. This method was applied to correct PERSIANN-CDR against grid-based rain gauged dataset APHRODITE [2] Results over the Mekong basin region showed that the ML method overcomes probabilistic methods, especially during extreme conditions. While these methods have performed well when describing distribution trends and spatial relationships, ML methods are limited by the physical constraints of the rainfall dynamics. ML models are only trained by the available data without any scientific principles or laws. This limitation may bring physically unrealistic predictions for several scientific problems (Corzo Perez, 2009; Wang et al., 2020).

2.4.4. Physically-based bias correction methods

Physically-based bias correction methods offer an integrated approach to describing extreme rainfall dynamics and correcting the main error components (volume displacement and timing). Available methodologies use spatial verification methods to match SPP rainfall events (defined as a object) with respect to gridded-based ground data (e.g. Skok et al., 2009; Li et al., 2015a, 2016). Several systematic errors can be corrected, including location, rotation, intensity, and shape. For instance, Demaria et al. (2011) used the object-based method, Contiguous Rainfall Analysis (CRA, Ebert and McBride (2000)), to correct the location error of CMORPH, PERSIANN, and the TMPA datasets over the Plata basin. Recently, Le Coz et al. (2019) used the field transformation method, called Feature Calibration and Alignment technique (FCA), to correct the error due to location in the GPM-IMERG late version over Sub-Saharan Africa. The application of these methods has been useful for correcting the displacement errors when the grid resolution is high, and the rainstorm event is small while preserving the higher spatial variability of SPP rainstorms. One important shortcoming of physical correction methods is the two-dimensional analysis of rainstorm events. In addition, the bias correction depends on a parallel grid-based data set to match the satellite event. As previously discussed, the inclusion of the temporal data into the spatial verification method will contribute to a complete diagnosis of the temporal and spatial evolution of the rainstorm events. Under this concept, other sources of error, such as displacement and rotation time, can be evaluated.

2.5. Bias correction for operational purposes

Methods for bias correction methods face several challenges in operational applications when observed measurements are absent or not yet available. A parallel observed

[2]APHRODITE (Asian Precipitation-Highly Resolved Observational Data Integration towards Evaluation of Water Resources) is a grid-based dataset created by the Research Institute for Humanity and Nature (RIHN) and the Meteorological Research Institute of Japan Meteorological Agency (MRI/JMA). APHRODITE is available at 0.25 x 0.25 from 1998 to 2015 over three domains: Monsoon Asia, the Middle East and Russia

data is essential in bias correction to estimate the discrepancies of the SPP data. However, based on the assumption that the error structure in the past determines its configuration in the future, probabilistic and Machine Learning approaches to overcome this limitation by using observed historical data. These methods are found not only in operational applications but also for the correction of forecast data from General Circulation Models. For instance, probabilistic mapping methods such as the Gamma Quantile Mapping uses observed historical data as reference to calculate the empirical cumulative distribution function (CDF) (e.g. Hashino et al., 2007; Piani et al., 2010; Zhao et al., 2017). Machine Learning methods include historical data in the training phase to correct SPP as well forecast data Gagne et al. (2014); Zarei et al. (2021); Tao et al. (2016); Le et al. (2020).

As opposed to probabilistic and machine learning approaches, the dependency of simultaneous observed data is the limiting factor of physically-based bias correction methods for operational applications. Available implementations using methods such as Contiguous Object-based Rainfall analysis methods or transformation methods relied on both SPP and observed data which is not always available, to correct errors due to displacement and volume (e.g. Demaria et al., 2011; Le Coz et al., 2019).

Combining the physical understanding offered by object-based bias correction methods with the operability of probabilistic or machine learning methods to work without parallel-ground data can increase the performance of independent bias correction methods. In remote sensing, Object-based Image Analysis methods (OBIA) are combined with ML models to improve classification accuracy in diverse areas, including urban planning (Wang et al., 2021), vegetation and agricultural mapping (Heumann, 2011; Duro et al., 2012; Zhang et al., 2018), natural hazards detection (Aksoy and Ercanoglu, 2012; Li et al., 2015b) among others. In comparison with pixel-based approaches, machine learning methods incorporated with OBIA reflect the human understanding of the real world based on the explicit knowledge of the object to classify in multiple dimensions. Further implementations of both physically-based and machine learning bias correction methods as a hybrid approach provide a promising alternative to reduce systematic errors in SPP data due to displacement and volume.

2.6. Conclusions

In this chapter, we introduced the main characteristics of the Satellite Precipitation Products (SPPs) and the methods used for analysing and correcting errors for rainstorm prediction. The algorithms and the temporal and spatial attributes of each SPP product were described based on the type of sensor retrieval used to measure rainfall. We only focused on the most common NRT satellite-based data sets used in hydrology.

The research presented in this dissertation will be based on a spatiotemporal method to analyse and correct rainfall events detected by satellites. The methods de-

veloped are conceptualised using rainstorm events as a multidimensional object. Such conceptualisation allows satellite errors in rainstorm prediction and their hydrological impact to be evaluated by analysing the complete structure of the event in space and time. The bias correction method developed using the proposed approach will consider the primary systematic sources of error to reduce the bias of SPPs. Real-time applications are also considered by combining the spatiotemporal bias correction method with a machine learning approach. Applications of this method are discussed when observed data is not readily available.

3

Study areas: Extreme rainfall in monsoon environments

This chapter describes the physical, geographical and atmospheric characteristics of the two monsoonal study areas selected to develop the proposed methodologies - the catchment of the Tiete River in Brazil (chapter 4-6) and the Lower Mekong Basin Thailand (chapter 7,8). Both topical catchments are characterised by extreme rainfall conditions triggering landslides and extreme flash flood events. In both areas, the rainstorm analysis is focused on the monsoon season. In the Tiete area, the period covers the South American monsoon season from December to February. While in the Lower Mekong Basin, the period spans the Asian summer monsoon season, which runs from June to October.

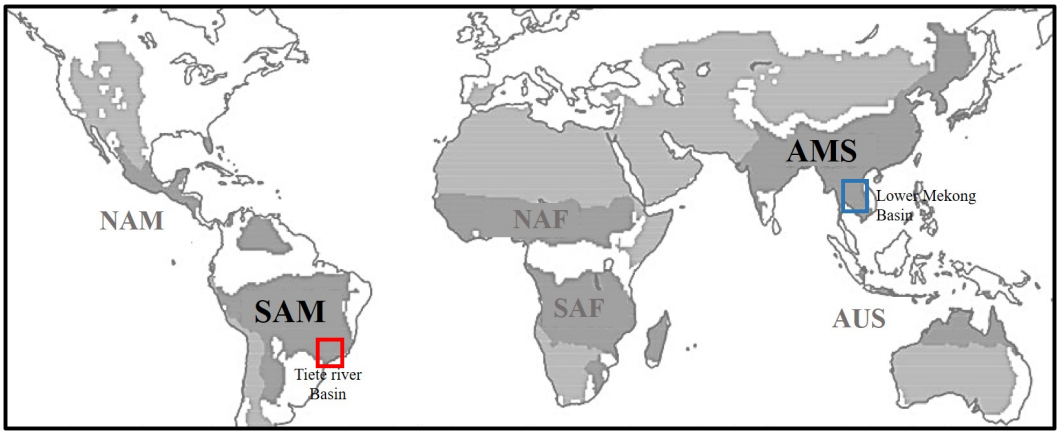


Figure 3.1: *Global monsoon regions and study areas. Monsoon land areas (dark grey); dry lands (light grey); Tiete river (red rectangle) and Lower Mekong basin (blue rectangle). Modified from Zeng and Zhang (2020)*

3.1. Introduction

According to the latest Intergovernmental Panel on Climate Change (IPCC) Report (Masson-Delmotte et al., 2021), the observed changes from the global surface temperature are leading to the increment in the frequency and intensity of rainfall, specially marked in monsoon environments. Nearly two-thirds of the world's population live in monsoon areas. Many of these societies living in developing countries rely on rain-fed agriculture for their subsistence. As a result, any disruption to the amount, timing and location of monsoon rain will dramatically impact the development of these societies.

Extreme rainfall arises from inter-and extra-tropical climatic conditions in monsoon regions triggering devastating landslides and floods. According to the Emergency Disasters Database (EM-DAT), from 1990 to 2019, Asia and America ranked first and second, respectively, as the two highest impacted regions in the world due to rainstorms and floods (ADRC, 2021). In 2019 alone, events during the Asian summer monsoon season affected almost 56 million people in Asia and cost around 51 thousand million dollars in economic damage. Meanwhile, extreme rainfall that occurred during the South American monsoon season, exacerbated by climatic oscillations such as the El Niño–Southern Oscillation (ENSO), triggers severe floods in the tropical west coast and south-eastern regions of South America with marked socioeconomic effects across countries like Brazil, Uruguay and northern Argentina (Cai et al., 2020). Due to this ongoing problem, the role of the scientific community is key to understanding the causes of error in SPPs and reducing the systematic error arising from the prediction of extreme rainfall.

This research focuses on two areas localised South America and the East Asian summer monsoon - the Tiete river in Brazil and the Lower Mekong Basin. The Tiete river basin has a high population density and is home to several of the region's most significant cities, such as the metropolitan Sao Paulo and Campinas. On the other hand, the Lower Mekong Basin, which spans Vietnam, Myanmar, Cambodia, Lao PDR and Thailand, is one of the richest and diverse areas in the region. Both areas are highly prone to natural hazards triggered by extreme rainfall events. This chapter describes the physical, geographical and atmospheric characteristics of the two monsoonal study areas and the domains used across this dissertation.

3.2. Tiete river system

This system is influenced by the convective precipitation band of the South Atlantic Convergence Zone, one of the most distinctive features of the South American Monsoon System Boers et al. (2013). During the Austrian summer season corresponding to December to February, the catchment experiences extreme climatological conditions due to low-altitude humid air fluxes that transport large amounts of rain from the tropical Atlantic Ocean to the continent. These conditions make the area prone to

landslides and flash floods (Sprissler, 2011). In this research, the analysis made over the Tiete river is divided into three different domains (Figure 3.2): the Tiete river; the sub-catchments of the Piracicaba, Jundiai and Capivari rivers; and finally, the catchment of the Capivari river used for the hydrological analysis. The physical and geographical characteristics for each domain will be briefly described below.

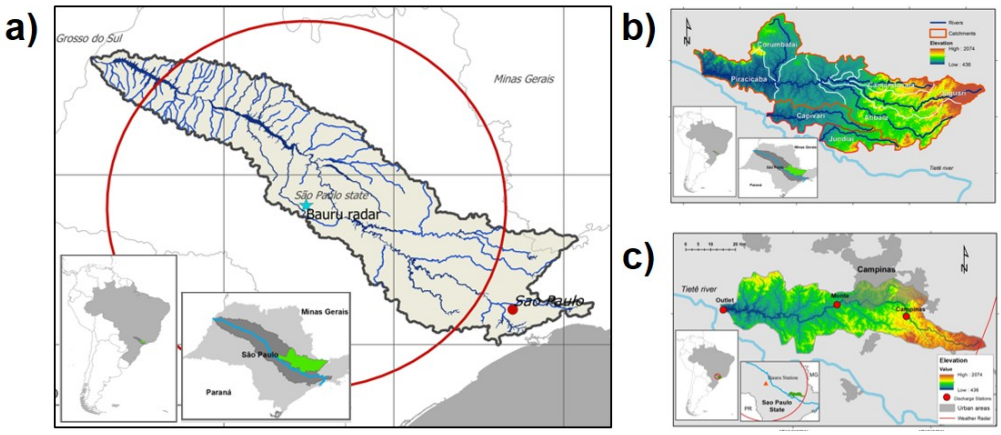


Figure 3.2: *Tiete river system: a) Tiete river domain; b) Piracicaba, Capivari and Jundiai rivers domain; c) Capivari river*

3.2.1. Tiete river

The Tiete river is part of the Paraná River basin, one of Brazil's central river systems and the second-longest river in South America, after the Amazon river. This river starts its journey at the eastern source in the São Paulo Metropolitan Region and flows 1,100 km backwards the Atlantic Ocean, where it joins the Paraná river. The total surface area of the basin is 72,000 km^2, in which 70% of its territory is affected by industries and agricultural activities Barrella and Petrere Jr (2003). The average rainfall ranges between 1,200 and 3,000 mm per year (Marcuzzo, 2020). The Tiete river is the primary source of energy for the state. The cascade reservoir system located across the river provides energy and water to the metropolitan area of Sao Paulo city, the highest urban centre in the country and one of the most populous cities in the world (Milano et al., 2018). The Tiete river is the domain for developing the spatiotemporal object-based rainfall method ST-CORA presented in chapter 5. The domain of this study comprises the area covered by the weather radar of the station Barau, located in the middle of the Tiete basin.

3.2.2. Piracicaba, Capivari and Jundiai rivers

The catchment areas of the Piracicaba, Capivari and Jundiai rivers, or PCJ, are among the Tiete river's administrative regions. This area is one of the most important industrial and economic centres accounting for 5.8% of Brazil's GNP. The PCJ covers a

drainage area of 14,138 km2 with an elevation ranging from 436 to 2,074 AMSL distributed across three zones: upper altitude zones, located on the east side and in a small area near to Corumbataí; intermediate or middle altitude zones, mainly located in the central part; and low altitude zones, located in the western part, where the rivers flow to the Tietê River. This area is particularly vulnerable to extreme rainfall events due to the higher population living in flood-prone areas Pompermayer et al. (2003); de Moraes Gonçalves et al. (2015). Between 2000 and 2011, natural hazards related to heavy rainfall affected more than 26,000 inhabitants. In this research, the PCJ is studied in chapter 4 to evaluate NRT SPP for extreme rainfall event prediction.

3.2.3. The subcatchment of Capivari river

The catchment of the Capivari river is used in Chapter 6 to evaluate the hydrological response of several systematic errors in SPPs for the prediction of rainstorm events. The drainage area is 1655 km^2, and the elevations range from 400 to 1000 m above sea level. This subcatchment can be subdivided into three geographic zones: upper, middle, and lower. Elevated areas characterise the upper area with a semi-urbanised distribution. These zones end at the Campinas streamflow station located on the east part of the city of Campinas. The middle zone is located at the stream gauge Monte station and is covered by the urban area of Campinas city, which is the main urban centre of the catchment. Finally, the lower zone covers the streamflow outlet to the Tiete River and corresponds to rural and semi-rural areas of lower elevations.

3.3. Lower Mekong basin system

The Mekong basin system is one of the largest rivers in the world and the longest river in Southeast Asia. In the Lower Mekong Basin, the river flows across Thailand, Vietnam, Lao PDR, Cambodia, and Myanmar, covering a total drainage area of 184.000 km^2. During the period between June and October, the hydroclimate of the basin is dominated by the covariability of the two subsystems of the Asian Summer Monsoon season - Indian summer monsoon (ISM) from the west, and the east, the Pacific East Asian summer monsoon (EASM) (Delgado et al., 2012; Yang et al., 2019). The confluence between the widespread and extended rains of ISM and the localised rainfalls generated by the EASM system trigger intense flash floods and landslides in the basin (Joy, 2012). In this research, we evaluated the Lower Mekong basin region located in Thailand. According to the Mekong River Commission (MRC), floods in this region have cost around 127 million US$ in damages and impacted more than 900 thousand people between 1980 and 2008 (Joy, 2012).

3.3.1. Lower Mekong basin in Thailand

This area is located in the northeastern region of Thailand corresponding to the Isan region (Figure 3.3). The geomorphological conditions are flat lowland and plateau areas (below 250 m) with soft and intermediate slopes. The primary land use of the area is predominantly rural, and the majority of its population belong to the agricultural sector. According to Sondergaard et al. (2016), the Isan is the region with the lowest income in the country. This socioeconomic situation incorporates additional challenges for dealing with floods. In this research, this area is used in Chapters 7 and 8 to develop the spatiotemporal bias correction method and the operational bias correction approach using machine learning.

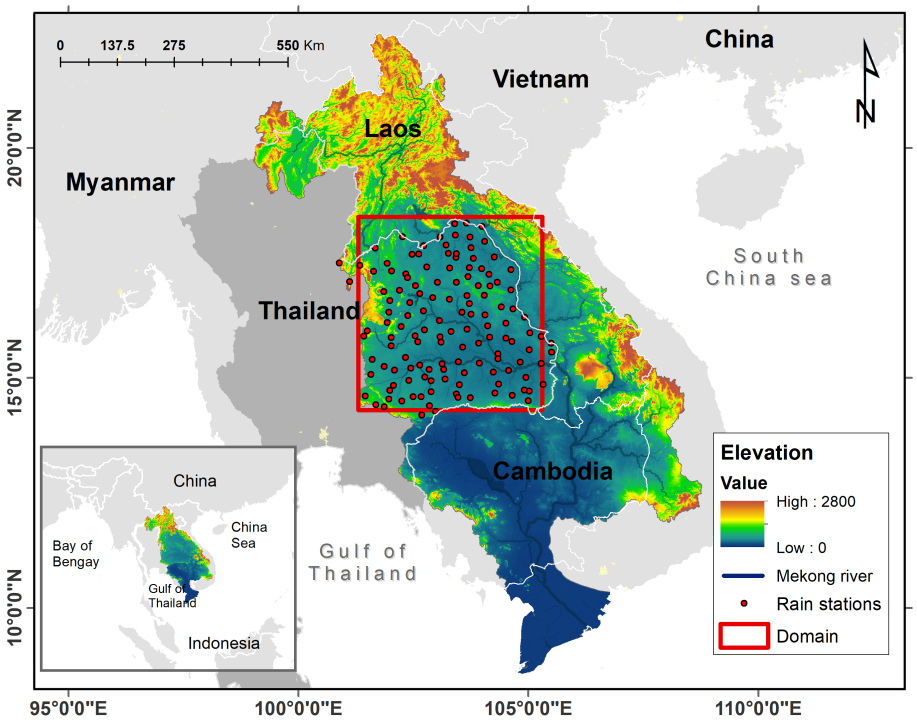


Figure 3.3: *Digital elevation of the Lower Mekong Basin in Thailand. Red rectangle corresponds to the study area in the Isan region. Red dots represent the hourly rain gauge stations from* `www.thaiwater.net`*. Blue line represents the Lower Mekong river network.*

3.3.2. Tiete river versus the Lower Mekong basin

The Tiete and the Lower Mekong Basin (LMB) present multiple similarities and essential differences in hydroclimatologic and socioeconomic factors. In terms of hydroclimatology, both regions present similar circulation patterns from the east of the boundary. This warm and moist environment coming from the Atlantic Ocean and the South China Sea to the Tiete river and the LMB, respectively, generates high convective rainstorm events during monsoon seasons (Zhou and Lau, 1998). The short

and localised life cycle of these events represents an important challenge for measuring rainfall from satellites.

Regarding socio-economic factors, both regions present remarkable differences in terms of economy and demographic distribution. Tiete river is mainly an industrial hub in the region with high urbanisation along the river. On the other hand, the LMB is primarily dominated by the agriculture sector and therefore, its population is mainly rural. As mentioned above, both areas share a common history associated with flood events due to extreme weather. Devastating flood events have impacted the lives of millions of people living in both areas.

4

Space-time error analysis for extreme rainfall events detected by satellites

This chapter explores the performance of the four Near real-time Satellite precipitation products (NRT SPP) to estimate different rainstorm events types in a subtropical catchment in south-eastern Brazil. (1) local and short duration, (2) long-lasting, (3) short and spatially extent, and (4) spatially extent and long-lasting extreme rainfall events are classified using an analytical approach based on the temporal and spatial characteristics of extreme values. This extreme value is evaluated at different threshold levels (75th, 90th, and 95th). NRT SPPs are evaluated against in-situ data from the thirteen hourly Automatic Weather Stations during the monsoon seasons from 2007 to 2014. Results show that the product performance highly depends on the spatiotemporal characteristics of rainfall events. All four NRT SPP tend to overestimate intense rainfall in the study area, especially in high altitude zones. CMORPH had the best overall performance to estimate different types of extreme spatiotemporal extreme events. The advantages and disadvantages of available rainstorm analysis methods are discussed as well the opportunities of new technologies to understanding the spatiotemporal features of rainstorm events estimated by SPP.

This chapter is partly based on the publication: Laverde-Barajas, M., Corzo Perez, G. A., Dalfré Filho, J. G., & Solomatine, D. P. (2018). Assessing the performance of near real-time rainfall products to represent spatiotemporal characteristics of extreme events: case study of a subtropical catchment in south-eastern Brazil. International journal of remote sensing, 39(21), 7568-7586.

4.1. Introduction

The distribution and intensity of extreme rainfall play an important role in the hydrological cycle. Characteristics of rainfall events such as magnitude, duration and spatial extent determine the level of damage associated with natural hazards. In this context, an accurate representation of the temporal and spatial components of extreme rainfall is crucial for the correct assessment of water resource availability and predict potential water-related risks (Thiemig et al., 2012).

During the last two-three decades, earth observation systems are providing an important input to the weather monitoring and forecasting systems (e.g. Kidd et al., 2009; Pan et al., 2010; Azarderakhsh et al., 2011). Satellite-based rainfall products provide uninterrupted global information with an up to 0.5 hours intervals and up to 8 km spatial resolution (e.g. Huffman et al., 2007; Joyce et al., 2004; Sorooshian et al., 2000). Despite the multiple advantages of satellite-based products, several studies have shown errors in satellite rainfall estimations related to the measuring devices (e.g. Hu et al., 2016; Qiao et al., 2014), the size of the basin (e.g. Moazami et al., 2013), the climate regimes and seasons (e.g. Thiemig et al., 2012; Sapiano and Arkin, 2009; Mei et al., 2014), and the geographical conditions (e.g. Mei et al., 2016a; Dinku et al., 2007; Guo et al., 2015; AghaKouchak et al., 2011). For extreme rainfall detection, uncertainties associated with the capacity of detecting heavy rainfall rate at short temporal resolutions (e.g. AghaKouchak et al., 2011; Marra et al., 2017) have limited their use in operational applications.

Several studies have evaluated the behaviour of satellite-based products to represent the spatial and temporal characteristics of extreme events. Temporal capabilities of different NRT satellite products have been investigated by Mehran and AghaKouchak (2014) across the United States. Analysing heavy rainfall at different temporal accumulations, the authors indicated that all high temporal resolution products (3-hourly) presented problems for estimating high rainfall rates.Gebregiorgis and Hossain (2015) analysed the spatial performance of different NRT products around the world. Based on error variance models, they showed how diverse geophysical settings impact the products' performance. In the case of South America, just a few studies have analysed the performance of satellite-based rainfall products in estimating the spatiotemporal characteristics of extreme rainfall. Ringard et al. (2015) evaluated four satellite-based rainfall products against in situ measurements over French Guiana and North Brazil. Dividing the study area into six climatic zones and analysing daily and monthly rainfall data, their analysis showed that estimates of low-intensity rainfall have relatively high accuracy, while convective-type rainfalls were poorly estimated by satellite products. Boers et al. (2015) analysed the spatial characteristics of extreme events estimated by two satellite-based products and one re-analysis dataset over the South American monsoon system. Using the complex networks theory, these researchers found substantial

differences in estimating the extreme rainfall patterns between the different rainfall products, especially in the South-East of Brazil.

For hydrological applications, rainfall analysis needs a better understanding of how the characteristics of an extreme event (intensity, duration and spatial distribution) are represented by different datasets. The main objective of this study is to evaluate the performance of the four near-real-time (NRT) satellite-based rainfall products to represent different extreme rainfall events (ERE). The raw version of Climate Prediction Centre (CPC) Morphing algorithm (CMORPH; Joyce et al. (2004)), the Tropical Rainfall Measuring Mission, Multisatellite Precipitation Analysis in real-time (TMPA-RT; Huffman et al. (2007, 2010)), the Precipitation Estimation from Remotely Sensed Information using Artificial Neural Networks - Global Cloud Classification System (PERSIANN-GCCS; Sorooshian et al. (2000)) version 7 and the Hydro-Estimator (Hydro; Scofield and Kuligowski (2003)) are compared against an Automatic Weather Station network (AWS; Brunini (2017)) during the core of monsoon seasons (December, January, February) from 2007 to 2014. The spatiotemporal analysis focuses on the four different extreme event types proposed by Boers et al. (2015).

4.2. Rainfall Products

The four NRT satellite-based rainfall products are evaluated on seven monsoon seasons from a common analysis period from December 2007 until February 2014. This selection is based on the following criteria: (1) good correlation in previous studies over the region; (2) good spatial and temporal resolution; (3) operation in NRT or "early run" satellite products. It is to highlight that the NRT are referred to non-gauge corrected rainfall estimations, which are available soon after their generation; (4) common covering period over the monsoon seasons; (5) their free access and publication. The latency of these rainfall products ranges from 1 to 18 hours. Within the products evaluated, two products combine passive microwave and infrared sensors (IR-PMW) and two use infrared sensors to estimate rainfall (IR). The use of IMERG SPP data which is the natural replacement of the TMPA, was not considered in this study because, by the time of the analysis, available records of this product started from April 2014. However, further studies can analyse this dataset SPP based on a reprocessed version of IMERG from 2000 to the present (released in 2019).

Considering each product has different spatial and temporal resolutions, rainfall products with spatial resolution finer than 0.25° were up-scaled by aggregation in which the products are averaged to larger scales for matching the spatial resolution of the referenced dataset. However, it is noted the data aggregation can introduce further uncertainties derived from the assumptions used in the aggregation algorithm (Gebere et al., 2015). Aggregated rainfall data in space and time has its own contribution to the total error since the lack of knowledge of the rainfall process at different spatial scales

will further propagate errors in the data. A comparison of SPP at several temporal and spatial scales is recommended in areas with the variability of rainfall is high, however, this study just focuses on the errors of SPP for extreme events predictions.

Table 4.1: *NRT satellite-based rainfall products used*

Product	Provider	Spatial coverage	Temporal coverage	Type	Spatial and temp res.
CMORPH V1.0 Raw	NOAA-CPC	60°N–60°S	Since 1 January 1998	IR-PMW	0.07°approx./3 h
TMPA-RT	NASA/JAXA	50°N–50°S	Since 1 January 1998	IR-PMW	0.25°/3 h
PERSIANN-GCCS	UC Irvine	60°N–60°S	Since 1 March 2000	IR	0.04° approx./0.5 h
Hydro-Estimator	NOAA/NESDIS	90°N–90°S	Since 1 January 2007	IR	0.04°/15 min

4.3. Study area and reference data

This study is applied in the tropical catchments of the Piracicaba, Capivari and Jundiai rivers (PCJ) in Brazil described in Chapter 3 (Figure 3.2b). We analysed climatic conditions during the core of monsoon in South America from December to February from 2007 to 2014. The reference data is based on 13 Automatic Weather Stations (AWS) observations provided by the Integrated Centre of Agrometeorological information CIIAGRO (Brunini, 2017). These stations are part of a dense network of hourly real-time information for agrometeorological monitoring in the PCJ catchment. The data were quality controlled to reduce possible errors and noise in the measurements. In the first part, we compared the product at the point-based location-scale to avoid taking interpolation errors into account. However, in the following parts, we used an interpolated gridded data to represent the spatial representation of the referenced data. Hourly AWS measurements were interpolated using the Inverse Distance Weighted method (IDW) (Wackernagel, 2013) set to 0.25° × 0.25° from 2007 to 2014.

4.4. Methodology

The methodology involves three parts: First, we analyse the spatial error distribution at different high intensities levels between satellite-based rainfall products and AWS gauges during monsoon seasons. In the second part, we evaluate the performance of the products to identify different extremes rainfall events (ERE) types. In the third part, we assess the sensitivity of the event-based performance to different intensity rainfall thresholds.

4.4.1. Spatial error distribution of intense rainfall

The spatial error of satellite products is analysed on a point-cell basis for the three intervals: above the 75th percentile (strong rainfall), above the 90th (extreme rainfall), and above the 95th (most extreme rainfall). This method compares the grid points of satellite products and the reference data with the nearest rain gauges values (Thiemig

et al., 2012; Dembélé and Zwart, 2016). However, in the areas where two or more stations lie in a cell, rainfall values are compared using the average-point measurements to the cell (Thiemig et al., 2012). In these cases, the spatial rainfall variability will be limited by the product resolution which may lead to an under or overestimation (Peleg et al., 2018). For this analysis we used three widely-used statistical measures to quantify the errors: the correlation coefficient (r) to analyse the linear correlation between the satellite products and AWS measurements, the Root Mean Square Error (RMSE) to evaluate the magnitude error, and the relative bias (Bias) to evaluate the systematic bias of the products (Li et al., 2014) (Eq.4.1 - 4.3).

$$r = \frac{\sum_{i=1}^{N}(P_{sat_i} - \overline{P_{sat}})(P_{ref_i} - \overline{P_{ref}})}{\sqrt{\sum_{i=1}^{N}(P_{sat_i} - \overline{P_{sat}})^2}\sqrt{\sum_{i=1}^{N}(P_{ref_i} - \overline{P_{ref}})^2}} \tag{4.1}$$

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(P_{sat_i} - P_{ref_i})^2} \tag{4.2}$$

$$Bias = \frac{\sum_{i=1}^{N} P_{sat_i} - P_{ref_i}}{\sum_{i=1}^{N} P_{ref_i}} * 100 \tag{4.3}$$

where n is the number of samples; P_{sat} are the satellite-based measurements and P_{ref} is the reference value, $\overline{P_{sat}}$ and $\overline{P_{ref}}$ are the mean of the satellite measurements and the reference values.

4.4.2. Performance in detecting different extreme rainfall events in spatiotemporal context

The performance of NRT products to represent rainfall events is evaluated considering temporal and spatial characteristics. For intense rainfall considered as extreme (above 90th percentile), we defined different ERE types according to the classification proposed by Boers et al. (2015). This methodology defines four types of rainfall events defined by their duration and spatial extension:

a. Local and short extreme events (LSE), which are only determined by their high magnitude.

b. Local and long-lasting extreme events (LLE) which are characterised by the long duration and high magnitude.

c. Spatially extensive extreme events (SEE), which are identified by their extension and high magnitude. In this study, unlike proposed by Boers et al. (2015), SEE

is composed of events (connected) with a high magnitude. This was done to analyse the spatiotemporal behaviour at the same threshold.

d. Spatially extensive and long-lasting extreme events (SLE) which are determined by their high magnitude, long duration and wide extension.

To identify the classes of EREs the following procedure is used:

1. Estimate the temporal duration of each event, by employing a running-mean filter to the gridded rainfall time series at time step i:

$$\tilde{P}_i = \frac{\sum_{s=-w}^{N} P_{i-s}}{2w+1} \tag{4.4}$$

where P_i is the filter input and $\tilde{P}_i$ is the filtered rainfall value. This method uses the moving average period defined as $2w+1$. w is the width of the running-mean filter given by table 2.

2. Considering P_i and $\tilde{P}_i$ to be samples of rainfall intensities, compute the pth percentile of these samples (for rainfall above 0.2 mm). Rainfall events are defined as time steps i for which P_i and $\tilde{P}_i$ are above a threshold T^p, depending on the type of event (strong, extreme, most extreme) (Table 2).

$$e_i := \begin{cases} 1, & \text{if } P_i | \tilde{P}_i > T^p. \\ 0, & \text{otherwise.} \end{cases} \tag{4.5}$$

where e_i is a binary event indicator associated with the time step i. T^p is the threshold corresponding to the pth percentile.

3. To identify the spatial extent of extreme rainfall, we employed the so-called connected component labelling method to group similar rain cells into homogeneous groups (Szeliski, 2010; He et al., 2017). This method connects cell values identified as an event in each time step as object pixels. This spatial type of event e_i^q is defined by the size of its connected label relative to the spatial threshold S_i^q (Boers et al., 2015).

$$e_i^q := \begin{cases} 1, & \text{if } C_i^q > S^q. \\ 0, & \text{otherwise.} \end{cases} \tag{4.6}$$

where C_i^q is the group of connected cells (considered as events) at each time step, is the spatial threshold in space. Considering the hydro-meteorological scales of rainfall systems observed in the region, we defined extensive events as events with an area bigger than 2° Œ 2° ($S_i^q = 8$ cells; each of 0.25° Œ 0.25°).

Taking into account the characteristics of ERE, we defined the spatial and temporal components of each ERE type. Table 4.2 shows the components of each type of ERE. It should be noted that these four classes are not mutually exclusive, so one event may be classified into two or more types of ERE.

Table 4.2: *Spatial and temporal characteristics of each ERE type*

ERE type	Magnitude	Temporal (w)	Spatial (q)
LSE	T^{75}, T^{90}, T^{95}	-	-
LLE	T^{75}, T^{90}, T^{95}	2	-
SEE	T^{75}, T^{90}, T^{95}	-	8 cells
SLE	T^{75}, T^{90}, T^{95}	2	8 cells

Considering the magnitude, duration and spatial extension of the reference dataset during monsoon seasons, we analysed the frequency and spatial patterns for all types of ERE. To avoid double-counting of EREs, we separated join extreme rainfall events selecting the inter-arrival time of each type of event (e.g. Dunkerley, 2008, 2010). By concept, LSE and SEE events usually range between 3 to 6 hours, while LLE last for 9-15 hours and SLE event ranges from 6 to 12 hours.

The capabilities of the NRT satellite-based rainfall products for each ERE type are evaluated based on four skill score metrics proposed by AghaKouchak and Mehran (2013); Wilks (2011). The Frequency Bias Index, Probability of Detection, False alarm ratio, and Critical Success Index is used to evaluate the performance concerning rainfall events, as presented in equations 4.7 - 4.10:

a. Event-based Frequency Bias Index (EFBI): indicates the level of underestimation or overestimation of an event "e". Its ranges from 0 to infinity with a perfect score of 1. I represent the indicator function of number exceedances.

$$EFBI = \frac{\sum_{i=1}^{N} I(P_{sat_i} | P_{sat_i} \in e \,\&\, P_{ref_i} \notin e) + \sum_{i=1}^{N} I(P_{sat_i} | P_{sat_i} \notin e \,\&\, P_{ref_i} \in e)}{\sum_{i=1}^{N} I(P_{sat_i} | P_{sat_i} \in e \,\&\, P_{ref_i} \in e) + \sum_{i=1}^{N} I(P_{ref_i} | P_{sat_i} \in e \,\&\, P_{ref_i} \notin e)} \tag{4.7}$$

b. Event-based Probability of Detection (EPOD): is defined as the ratio of the correct detections of "e". It ranges from 0 to 1, with a perfect score of 1.

$$EPOD = \frac{\sum_{i=1}^{N} I(P_{sat_i} | P_{sat_i} \in e \,\&\, P_{ref_i} \in e)}{\sum_{i=1}^{N} I(P_{sat_i} | P_{sat_i} \in e \,\&\, P_{ref_i} \in e) + \sum_{i=1}^{N} I(P_{sat_i} | P_{sat_i} \notin e \,\&\, P_{ref_i} \in e)} \tag{4.8}$$

c. Event-based False alarm ratio (EFAR): represents the ratio of the incorrect detections belonging to the event "e". Its ranges from 0 to 1 with a perfect score

of 0.

$$EFAR = \frac{\sum_{i=1}^{N} I(P_{sat_i} | P_{sat_i} \in e \,\&\, P_{ref_i} \notin e)}{\sum_{i=1}^{N} I(P_{sat_i} | P_{sat_i} \in e \,\&\, P_{ref_i} \in e) + \sum_{i=1}^{N} I(P_{ref_i} | P_{sat_i} \in e \,\&\, P_{ref_i} \notin e)} \quad (4.9)$$

d. Event-based Critical Success Index (ECSI): Corresponds to the combination of EPOD and EFAR to identify the overall performance skill of ERE. Its ranges from 0 to 1, with 0 as the perfect score

$$ECSI = \frac{\sum_{i=1}^{N} I(P_{sat_i} | P_{sat_i} \in e \,\&\, P_{ref_i} \in e)}{\sum_{i=1}^{N} I(P_{sat_i} | P_{sat_i} \in e \,\&\, P_{ref_i} \in e) + \sum_{i=1}^{N} I(P_{ref_i} | P_{sat_i} \notin e \,\&\, P_{ref_i} e) + \sum_{i=1}^{N} I(P_{sat_i} | P_{sat_i} \in e \,\&\, P_{ref_i} \notin e)} \quad (4.10)$$

4.4.3. Event-based performance for different rainfall intensities

The sensitivity of the event-based performance at different intensity rainfall thresholds is evaluated in the third part of the methodology. Following the process described in section 4.2, we compared the products' performances for rainfall events defined above-defined above T^{75}, T^{90}, T^{95} percentiles. For each satellite product, the performance is evaluated using EFBI, EPOD, EFAR, ECSI scores.

4.5. Results and discussion

4.5.1. Error distribution of SPP at high-intensity levels

Figure 4.1 shows the spatial location of errors of the satellite products, compared to the reference AWS data during monsoon seasons from 2007 to 2014. Figures 2a-c present the spatial distribution of r, RMSE and Bias, Figures 2d shows the cross-correlation between the satellite products and the reference data over the study area. According to the results, satellite products had problems for the high altitude zones. CMORPH was the product with the lowest quantitative errors, indicated by the highest correlation (mean r 0.65), and the lowest magnitude error and the systematic bias (mean RMSE 0.58 mm hour-1, Bias 4.5%). For TMPA-RT and PERSIANN-GCCS, errors were generally high for elevated altitude zones. TMPA-RT presented a better correlation and lower Bias than PERSIANN-GCCS, however, the error magnitude was higher. Hydro was the product with the highest error over the whole study area with r and RMSE values of around 0.33 and 1.16 mm hour-1 respectively, and a Bias higher than 40%.

Concerning the measurement of the error at different rainfall levels, Figure 4.2 shows the r, RMSE and Bias errors of NRT products above 75th, 90th and 95th

percentiles thresholds: error increases with an increase of rainfall intensity. CMORPH and TMPA-RT had the lowest errors at high intensities. CMORPH presented the highest correlation coefficient dropping from 0.4 at T^{75} to 0.2 at T^{95} while TMPA-RT had the lowest overestimation increasing from 25% at T^{75} to 38% at T^{95}. In contrast, PERSIANN-GCCS and Hydro have the highest errors (the lowest correlation and high overestimation).

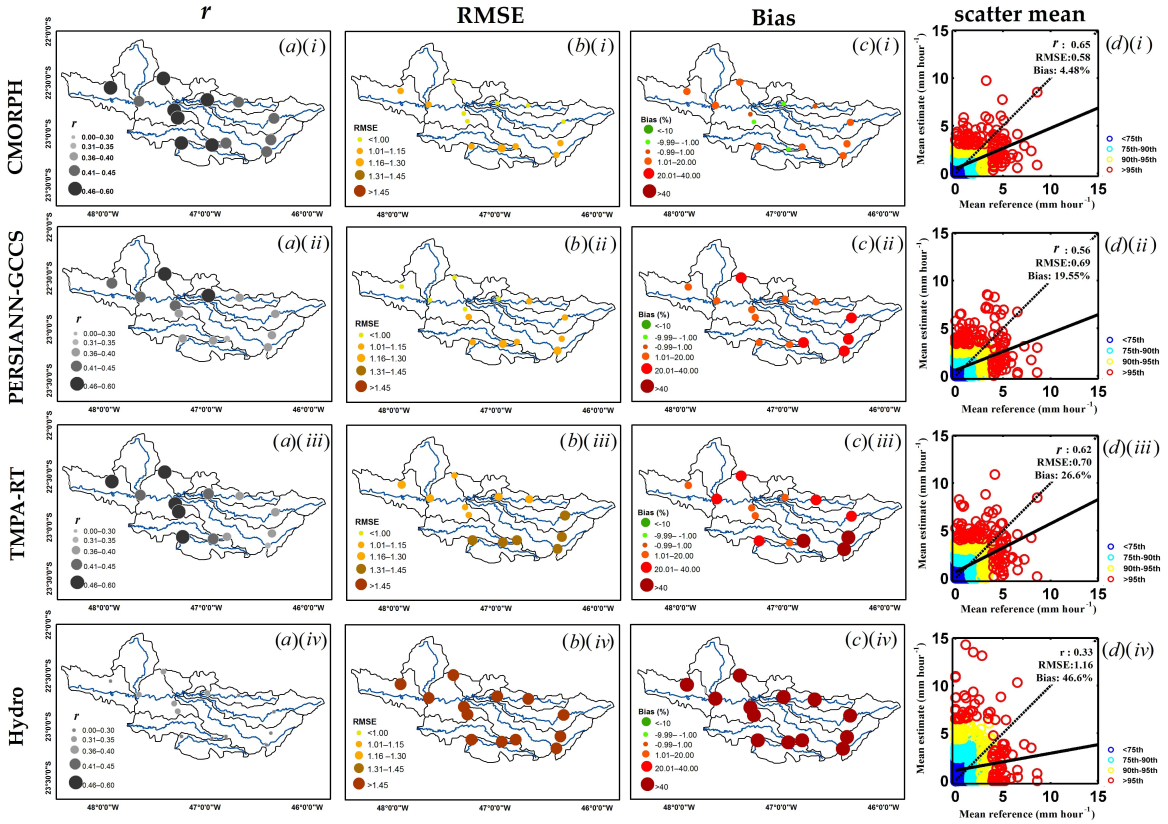


Figure 4.1: *Spatial distribution of (a) r; (b) RMSE; (c) Bias errors and (d) scatterplot of the mean rainfall of (i)CMORPH, (ii) PERSIANN-GCCS,(iii)TMPA-RT (iv) Hydro, during monsoon seasons from 2007 to 2014*

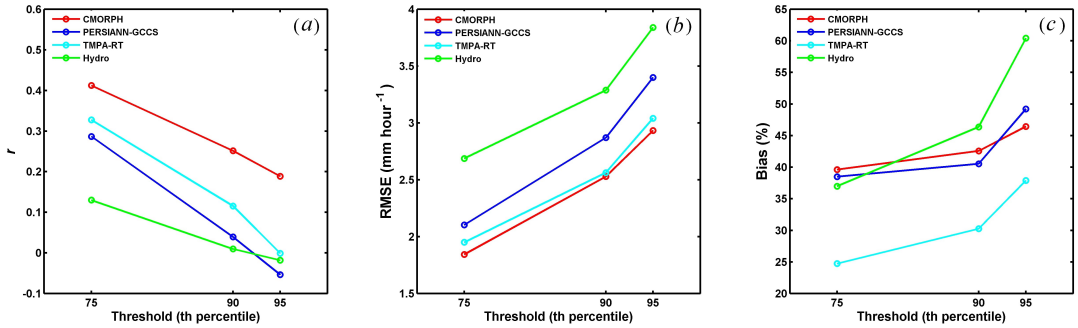


Figure 4.2: *Quantitative errors of satellite rainfall products versus AWS measurements above different rainfall thresholds.(a) r; (b) RMSE; (c) Bias*

4.5.2. Performance of satellite-based products to represent different ERE types

Local and short extreme events (LSE):

The performance of satellite-based products in terms of EFBI, EPOD, EFAR and ECSI for LSE events is shown in Figure 4.3. In general, performance is not very high, especially at high altitude zones. CMORPH showed better performance (low EFBI and EFAR). TMPA-RT showed the best EPOD and together with CMORPH had the better ESCI score distributed mainly in the middle and low altitude zones. PERSIANN-GCCS and Hydro had a poor score for LSE detection, being the products with the lowest performance for LSE.

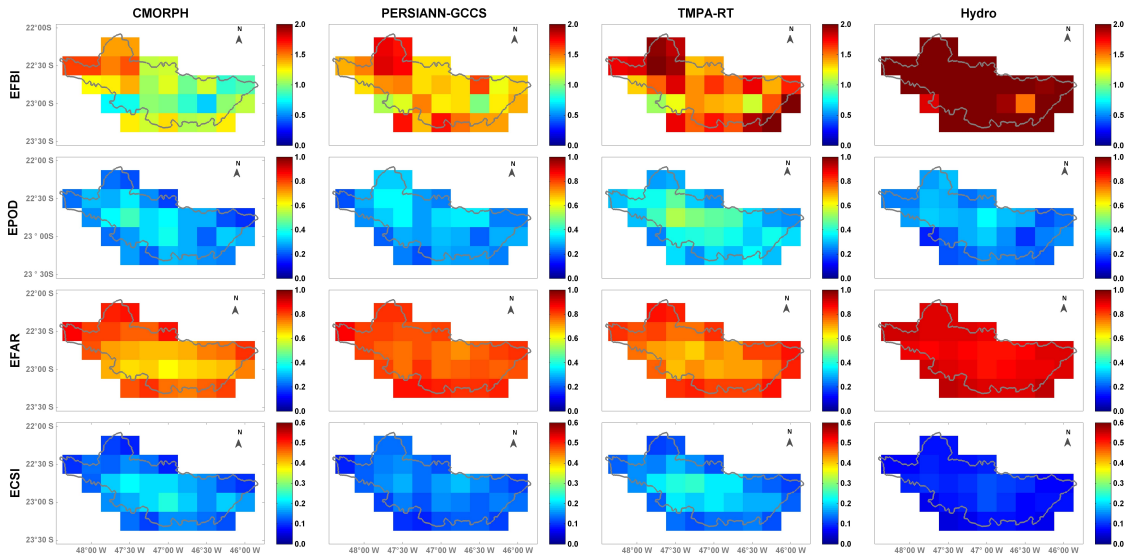


Figure 4.3: *Performance of NRT products in estimating LSE events*

Local and long-lasting extreme events (LLE):

Figure 4.4 presents the performance of satellite products for LLE events. When the extreme rainfall events are longer in time, TMPA-RT and CMORPH slightly underestimated LLE events as opposed to Hydro and PERSIANN-GCCS, which tend to overestimate them. Compared with the short duration events, satellite products performed better in detecting LLE events, with higher EPOD detection and fewer EFAR estimations. CMORPH had a better EPOD score while TMPA-RT had a better score in false alarm estimations. Overall, CMORPH presented the highest performance for LEE events with a mean ESCI of 0.23, followed by TMPA-RT and PERSIANN-GCCS. Hydro is the product with the lowest performance for LSE.

Spatially extensive extreme events (SEE):

Performances for SEE events is presented in Figure 4.5. The results show the limited capacity of satellite products to detect this type of extreme event. Hydro was highly

biased over the whole catchment while PERSIANN-GCCS and TMPA-RT were biased mainly over lower altitude areas. On the other hand, CMORPH was unbiased in lower zones but underestimates over the elevated areas. Regarding the capacity to detect SEE events, TMPA-RT is better, with a high EPOD score over lower zones. CMORPH had the lowest number of false alarm detections with an EFAR value around 0.6. In general, TMPA-RT had the best performance for SEE events with a mean ECSI of 0.2, followed by CMORPH with a mean ECSI of 0.15, PERSIANN-GCCS with mean ECSI of 0.14 and finally Hydro with ESCI of 0.1.

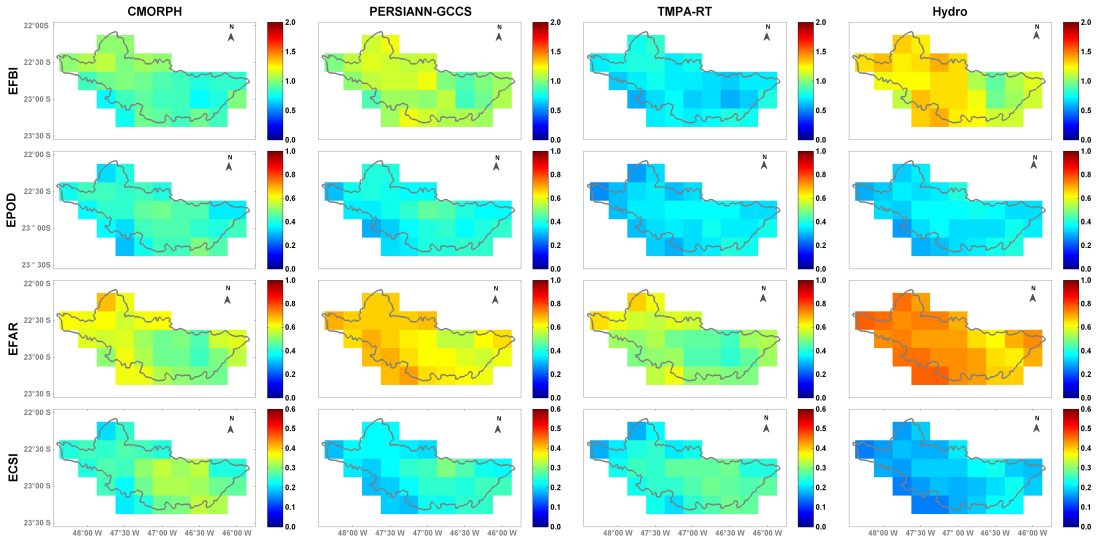


Figure 4.4: *Performance of NRT products in estimating LLE events*

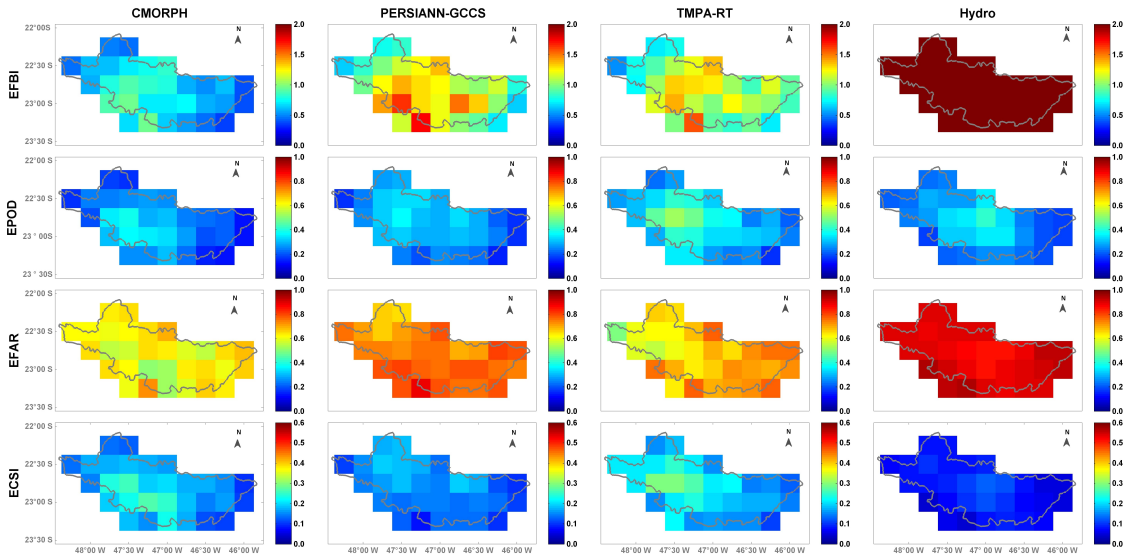


Figure 4.5: *Performance of NRT products in estimating SEE events*

Long-lasting and spatially extensive extreme events (SLE):

Figure 4.6 shows the performance of satellite products for the SLE events, and one can see the products' higher performance. CMORPH and TMPA-RT were slightly unbiased while Hydro and PERSIANN-GCCS were marginal biased, mainly in low and middle altitude zones. In comparison, CMORPH was slightly better in EPOD and together with TMPA-RT, showed the lowest false alarm score.

In terms of ECSI, CMORPH had the best performance for detecting SLE events with a mean ECSI value of 0.3 distributed over middle and high altitude zones. In contrast, Hydro was the product with the lowest score with a mean ECSI of 0.15.

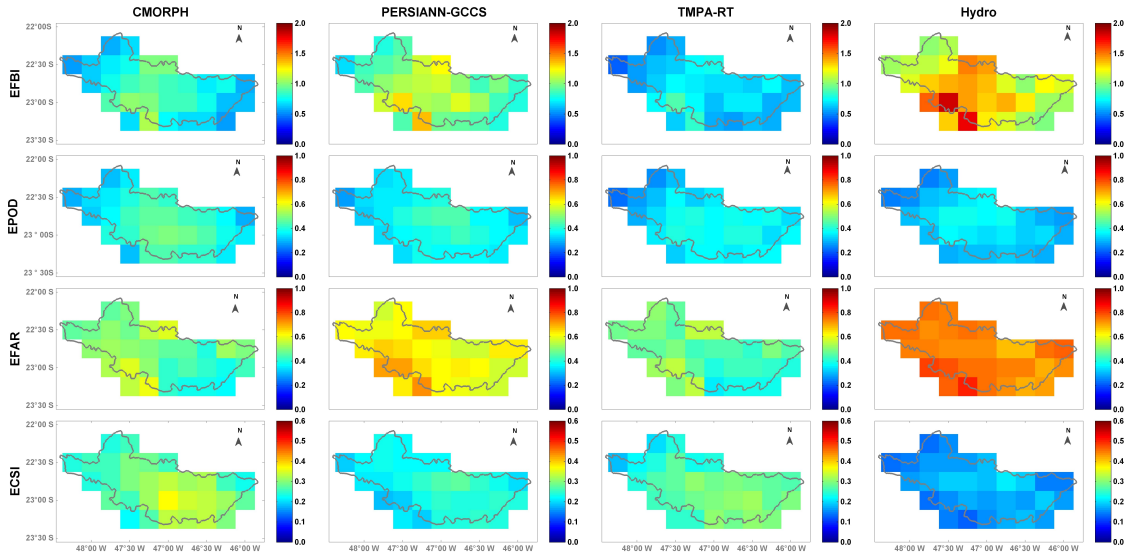


Figure 4.6: *Performance of NRT products in estimating SLE events*

Event-based performance at different rainfall intensity thresholds

To analyse the influence of the intensity threshold on the evaluation of satellite performance, Figure 4.7 shows the comparison between the skill scores for each type of event and the rainfall intensity threshold defined as strong, extreme and most extreme rainfall (above T^{75}, T^{90} and T^{95} respectively). Lines show the average values in the whole area, while the shaded regions enveloping them represent the dispersion between the 25 and 75% percentiles. Downward- and upward-pointing triangles are the minimum and maximum score levels for each product.

Results show that the performance of the evaluated SPP is reduced under high-intensity levels of rainfall. During the evaluated period, the bias of the SPP had a significant increment at higher thresholds. Hydro had the most considerable increment, especially LSE and SEE events. Short duration rainfall events were more difficult to estimate than extensive and long-duration events accurately. In the case of EPOD, the performance of SPP marginally decreased for high-intensity events.

EFAR results steadily increased as the event threshold increased, showing a consid-

erable deterioration EFAR scores at higher thresholds. Overall, the capacity of estimating different rainfall events in terms of ECSI showed that long temporal rainfall events score for all products had a steep drop (50 % approximately), except TMPA-RT, which score also dropped but just marginally. ECSI score for short duration events dropped around 30% for all products, demonstrating low sensitivity at higher intensities.

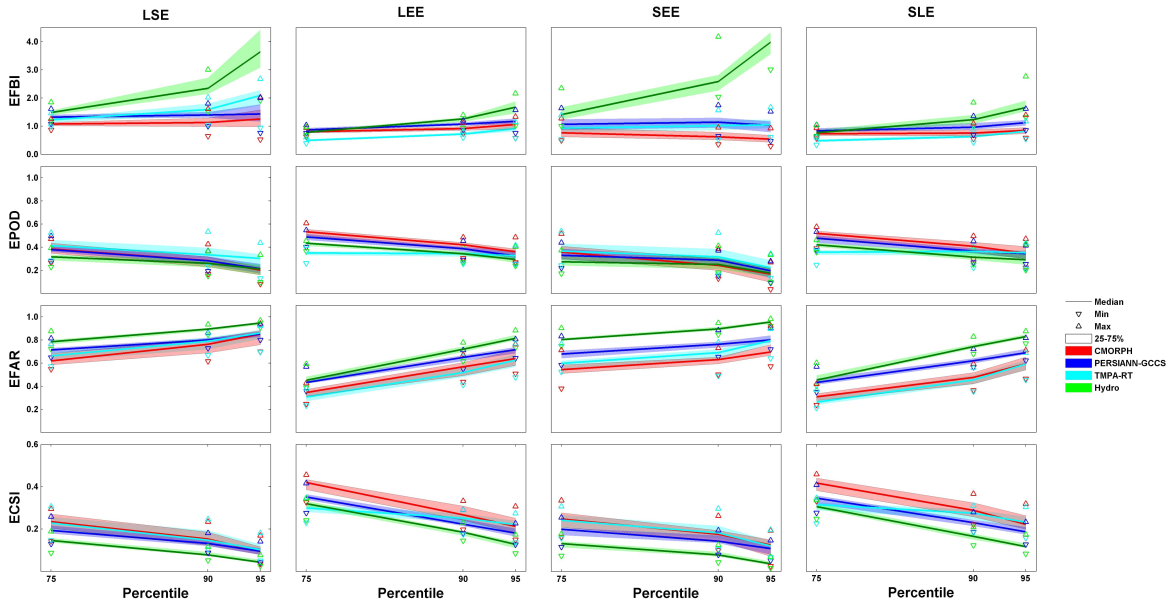


Figure 4.7: *The performance of the four products for EREs of different intensities. Lines are 50% percentiles; shaded ranges - the 25% and 75% percentiles. Downward- and upward-pointing triangles are the minimum and maximum score level for each product*

4.6. Conclusions

In this chapter, we evaluated the performance of four NRT satellite-based products for representing different types of EREs in the subtropical catchment of the Piracicaba, Capivari and Jundiai rivers in Brazil. CMORPH, PERSIANN-GCCS, TMPA-RT and Hydro were compared against hourly rain gauge information from AWS during monsoon seasons from 2007 to 2014. The applied methodology identified the spatiotemporal characteristics of extreme rainfall events, classifying them as four EREs according to their magnitude, duration and spatial extent. We analysed the errors at different rainfall intensities, the performance to detect different extreme events, and the sensitivity of the performance at different thresholds.

We can conclude that all products tend to overestimate rainfall over the study area. CMORPH had the lowest quantitative error in r, RMSE and Bias at the point-based location. However, TMPA-RT showed lower Bias at high intensities levels. On the other hand, Hydro was the product with the highest error over the whole study area.

The performance of the NRT rainfall products depends on the spatiotemporal characteristics of rainfall events. In general, the short duration events are more difficult

to predict than the long ones. For all ERE types, the study showed that CMORPH and TMPA-RT products exhibited the best performance while PERSIANN-GCCS and Hydro displayed the lowest. TMPA-RT had the best EPOD detections for short temporal events in the same way as CMORPH in spatially extensive events. CMORPH presented the lowest false alarm detections in short-duration events and together with TMPA-RT had the lowest EFAR for spatially extensive events. For ECSI, CMORPH showed the highest performance of all satellite products.

It has been found that the performance of the product is strongly affected by the intensity of rainfall events: the bias increases with intensity. Concerning the capacity to predict different types of events, in most of the rainfall products, the performance of correct estimations marginally decreased, while the frequency of incorrect estimations considerable increased for high-intensity rainfalls. In general, the performance of all products decreased at high rainfall for all types of events.

The results show the importance of taking into account the spatiotemporal characteristics for product verification. Even though the methodology analyses the characteristics of the extreme event using a pixel-based approach, the results show an interesting evaluation of the capabilities of NRT to estimate different EREs. Further research will incorporate new verification methods such as feature-based to analyse the spatiotemporal structure of extreme events. These methods can be easily applied to evaluate the capabilities of products at different resolutions without using up-scaling techniques, which may contribute to the product error. Another further approach can be the selection of an optimal combination of products. The technology of fuzzy committees models (e.g. Fenicia et al., 2007; Kayastha, 2014) could be a possible candidate for this.

5

Spatiotemporal Object-Based method for rainstorm analysis

This chapter proposes a spatiotemporal object-based method to analyse the structure of rainstorm events in space and time at a catchment scale. This method, called Spatiotemporal Object-based Contiguous Rainfall Analysis or ST-CORA, uses a multidimensional connected-component labelling algorithm to cluster convective rainfall regions as a 4D rainstorm event object (longitude, latitude, time, rainfall intensity). Several features can be extracted from this rainstorm object, such as volume, area, duration, orientation, speed, among others. The methodology is applied to the subtropical catchment of the Tiete River to identify and classify different types of extreme events during monsoon seasons and verify a Near-Real time satellite-based product. Results show the importance of spatial and temporal structures in the comparison of products for real-life events. This method also provides insights to understand better the rainfall concentration (location) of events and their dynamic over catchments.

This chapter is partly based on the publication: Laverde-Barajas, M., Corzo, G., Bhattacharya, B., Uijlenhoet, R., & Solomatine, DP (2019). Spatiotemporal analysis of extreme rainfall events using an object-based approach. In Spatiotemporal Analysis of Extreme Hydrological Events (pp. 95-112). Elsevier.

5.1. Introduction

Satellite-based precipitation information has become an important input in water management, flood monitoring, and forecast systems. Multiple products derived from Infrared and Passive Microwave satellite-sensors have provided high-resolution precipitation data in near and real-time (e.g. Huffman et al., 2007, 2015; Nguyen et al., 2020; Joyce et al., 2004; Nguyen et al., 2020). Despite recent advances in earth observation systems, numerous researchers have shown that characterising small-scale variability of rainfall patterns is still challenging Grayson and Blöschl (2001). Continuous verification statistics traditionally characterises Grid-based studies to analyse the rainfall dynamic and evaluate product quality. While these measurements provide useful information in terms of correlation, metrics do not consider important intrinsic features of rainfall data such as location, volume, and type of event.

Object-based methods are an alternative approach to analysing rainfall. This methodology evaluates rainfall estimation based on the structural properties of rainfall fields. Based on this method, a rainfall event can summarise all intrinsic attributes and statistics representing space and time. Different authors have demonstrated the use object-based methods to evaluate the performance of satellite-based products Davis et al. (2009a); Ebert and McBride (2000); Li et al. (2015a); Skok et al. (2009). Object-based methods in space and time are becoming a powerful tool to analyse the complete structure of large-scale rainfall systems such as events like typhoons, hurricanes, or cold fronts and evaluate their prediction (e.g. Davis et al., 2009b; Sellars et al., 2013; Mittermaier and Bullock, 2013). However, in the case of small-scale (catchment-scale) convective systems, this analysis is challenging (Sillmann et al., 2017). Floods caused by this type of event, e.g. mesoscale and rainstorm-scale rainfall events, produce extreme flash flooding, causing major human and economic damage. Understanding rainfall dynamics plays an important role in hydrological applications. The variation in shape and location of rainfall events affects the runoff volume over the catchment (e.g Arnaud et al., 2002; Foufoula-Georgiou and Vuruputur, 2001; Haile et al., 2011).

A new object-based method is proposed to analyse the spatiotemporal structure of extreme rainfall events at the catchment scale. The Spatiotemporal Object-based contiguous Rainfall Analysis method (ST-CORA) enables the feature extraction of different types of rainfall fields from satellite products through a multidimensional connected component labelling algorithm. This research describes the innovative application of a spatiotemporal object-based method to analyse extreme rainfall events at the catchment scale, using a subtropical catchment of the Tiete River, Brazil, to identify extreme rainfall events and the evaluation of near-real-time (NRT) satellite-based products.

5.1.1. ST-CORA description

ST-CORA is a spatiotemporal object-based method designed to analyse the spatiotemporal features of different rainstorm events at catchment scales (duration, spatial extent, magnitude, and centroid). This method uses a multidimensional connected-component labelling algorithm to detect regions with similar features in space and time. The error verification for SPP allows the error decomposition in displacement, volume and pattern error. The methodology for error verification is composed of three steps briefly described below.

5.1.2. Identification of spatiotemporal convective objects

In the first step, spatiotemporal rainfall objects are built using a multidimensional connected-component labelling algorithm (Sedgewick, 1998; Acharya and Ray, 2005). This algorithm groups connected voxels (used in 3D instead of pixels) into a disjoint object, assigning a unique identifier (label). This operation is realised for binary information of "effective rainfall" voxels $S[x, y, t]$ (1 = "true" or 0 = "false"), segmented by rainfall intensity threshold (Eq. 5.1). The choice of intensity threshold is defined by the user. By default this threshold is 1 mm/h (e.g. Ebert, 2005):

$$S_{[x,y,t]} := \begin{cases} 1, & \text{if } R_{x,y,t} \geq \text{IT}. \\ 0, & \text{otherwise}. \end{cases} \tag{5.1}$$

where $R_{x,y,t}$ is the rainfall voxel, and IT is the rainfall intensity threshold. Once all $S_{[x,y,t]}$ voxels have been determined, the connected-component labelling algorithm identifies spatiotemporal convective objects as follows:

1. Scan all voxels in a neighbour system (from top to bottom and left to right) assigning preliminary labels to $S_{[x,y,t]}$ as:

$$c(S_{[x,y,t]}) = \left\{ N_{[x,y,t]} \in \alpha s : S_{CR} = S_N \right\} \tag{5.2}$$

   where $c(S_{[x,y,t]})$ is the preliminary label, S_s,S_N are the properties of the voxel $S_{[x,y,t]}$ and its neighbours $N_{[x,y,t]}$, respectively, and vs is the neighbour system in space and time (Fig. 5.1).

2. If a neighbour has more than two $c(S_{[x,y,t]})$, it is assigned the lower label recording the label equivalences in a union-find table.

3. Resolve the table of equivalences classes using the union-find algorithm (Sedgewick, 1998).

4. Make a second iteration relabelling $c(S_{[x,y,t]})$ on the resolved equivalences classes.

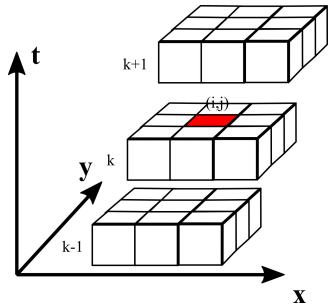


Figure 5.1: *Neighbour system in space and time (26 voxel neighbours)*

After the connected-component labelling algorithm is applied, ST-CORA applies two additional algorithms: a size filtering for noise removal and morphological closing to delineate the 4D object. A size-filtering algorithm removes objects lower than a size threshold T defined as noise. The selection of T is defined based on the spatial and temporal resolution of the rainfall product (default six voxels). The second algorithm solves false merging resulting from the labelling component algorithm. A morphological closing algorithm divides or merges convective objects with low or robust connectivity (Fig. 5.2). This algorithm uses a dilation erosion process similar to the one employed in object-based algorithms to separate zones with a weak connection in weather radar (e.g. Han et al., 2009). For dilation, boundaries of convective objects are expanded, while for erosion, those boundaries are removed. This procedure performs first a morphological dilation followed by erosion to delineate the convective object.

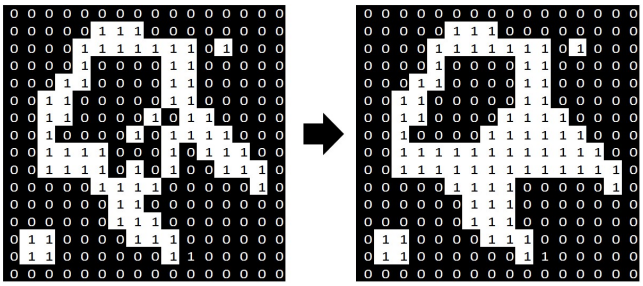


Figure 5.2: *Morphological closing*

5.1.3. Selection of extreme rainfall events

In the second step, extreme convective events are identified based on the critical mass threshold. This parameter corresponds to the minimum volume of rainfall (km^3) necessary to be considered as an extreme event Grams et al. (2006). The value of the critical mass threshold is defined by the user and is typically obtained depending on the maximum extension, and the convective object volume (e.g. Demaria et al., 2011). Once all rainfall events are identified, a descriptive statistical algorithm calculates diverse characteristics such as the total volume of the rainstorm event (m^3), maximum intensity (mm/h), maximum area (km^2), rainstorm duration (h), and weighted cen-

troid (latitude, longitude, time). This algorithm organises the event properties in a data structure (N x d DataFrame) for easy analysis and visualisation (Fig. 5.3).

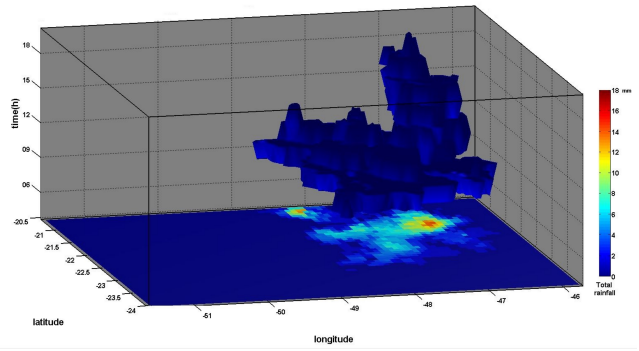


Figure 5.3: *Example of a rainfall object. Rainfall event over the subtropical catchment of the Tiete River, Brazil (24 Feb 2008)*

5.1.4. Classification of extreme event

The last step corresponds to the classification of extreme rainfall events by hydrometeorological criteria. Based on three main characteristics of extreme events (maximum area, duration, and volume of rainfall) (Fig. 5.4), four types of rainfall events are identified: local and short extreme events (LSE); local and long-duration extreme events (LLE); spatially extensive extreme events (SEE); and long-duration and spatially extensive extreme events (SLE). SEE, and LSE are events with a slow or fast motion, which are extended over large areas (long-lived) (e.g., mesoscale level). LSE and LLE are associated with small convective systems with a slow or fast movement, which contains a large amount of rainfall falling over a reduced area (short-lived) (e.g., a city, small draining catchment), for example, convective rainstorms and rainfall cells, among others.

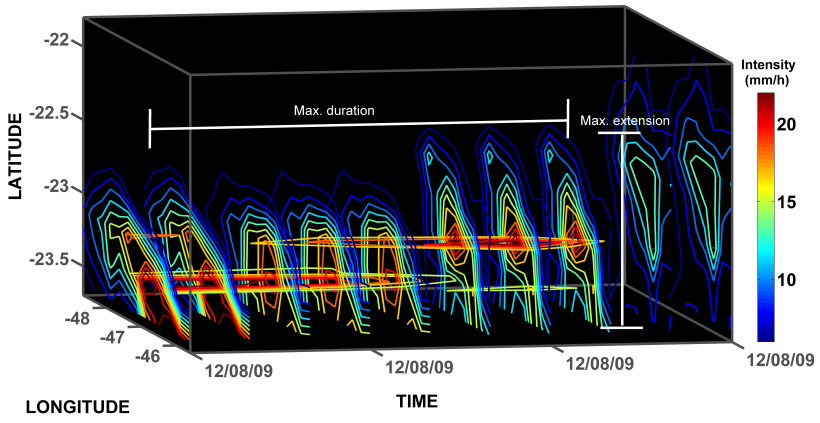


Figure 5.4: *Spatiotemporal characteristics of extreme rainfall*

5.1.5. Spatiotemporal verification

Several error metrics can be applied to rainfall objects to evaluate the performance of satellite products starting from standard verification methods such as continuous statistics (e.g., mean error, RMSE, correlation coefficient) and categorical verification metrics (e.g., BIAS, POD, FAR) to most sophisticated diagnostics verification methods such as error decomposition (Ebert and McBride, 2000), correspondence ratio (Stensrud and Wandishin, 2000), or displacement and amplitude score (Keil and Craig, 2009b). This method can be used for identification at different resolutions and scales.

5.2. Applications over the Tiete river, Brazil

The study site for this research is the subtropical catchment of the Tiete River (Fig. 5.8). The area is part of the Parana River basin, one of Brazil's central river systems. Due to the location, the area is strongly impacted by inter-and extratropical climatic conditions, e.g., the South American monsoon system (Boers et al., 2013). This factor makes the area prone to severe landslides and flash floods (Sprissler, 2011). This research analyses the climatic conditions during the monsoon core in South America from December to February from 2007 to 2017.

Near real-time satellite rainfall information from CMORPH (Joyce et al., 2004) is used for the study. The selection of this dataset responds to a previous study in the area made by Laverde-Barajas et al. (2017), where this dataset showed superior performance in estimating different types of rainfall events. CMORPH is used at 8 km (0.0727 degrees) with half-hour temporal resolution (1 h aggregated). Ground measurement is obtained by the weather radar station Bauru (CAPPI 3.5) from the Sao Paulo State University Faculty of Sciences. Ground measurements used have a spatial resolution of 1k (0.01 degrees) every 15 min (aggregated to 1 h).

5.2.1. Rainstorm identification

ST-CORA requires the definition of two parameters: the convective threshold and the critical mass threshold. The definition of convective threshold is critical in the definition of rainfall systems. Several studies in spatial verification have found that reduction of this threshold allows the incorporation of small systems with the deficiency of creating unrealistic rainfall objects (Demaria et al., 2011). In this study area, it was observed that the parameter of the best fit is 1 mm/h. The critical mass threshold value is defined by the relationship between rainfall systems' maximum area and volume. In the Tiete River catchment, it is found that convective systems larger than 5000 km^2 (8% of the population) contribute almost 90% of the rainfall during the monsoon season (Table 5.1). This area corresponds to a total volume of around 0.14 km^3 (Fig. 5.5). Based on this result, a critical mass threshold of 0.1 km^3 was selected. By applying this

method over hourly information from CMORPH in the catchment area, 694 extreme rainfall events were identified during the monsoon season from 2007 to 2017. For each rainfall event, several descriptive characteristics were extracted. These include volume, intensity, maximum area, duration, and weighted centroid.

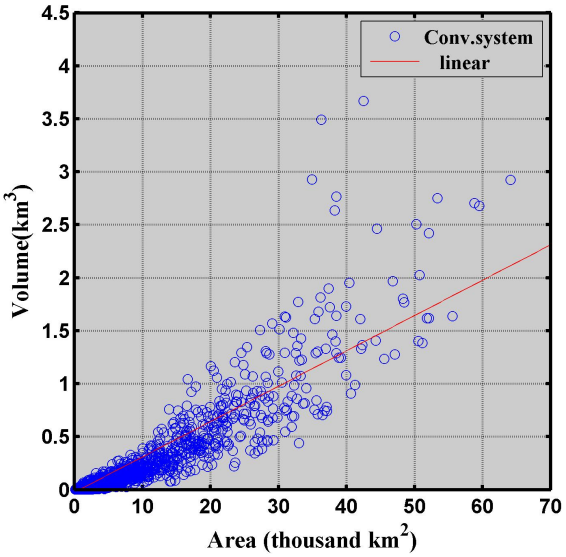


Figure 5.5: *Relationship between maximum event area and total rainfall volume.*

Table 5.1: *Contribution of the Rainfall Fields Into Total Amount of Water During the Monsoon Season*

Rainfall features	Rainfall contribution (%)	Population fraction (%)
Size range 75-500 km^2	1%	55%
Size range 500-2.000 km^2	5%	28%
Size range 2.000 - 5.000 km^2	8%	8%
Size range 5.000 - 10.000 km^3	11%	3%
more than 10.000 km^2	76%	5%

5.2.2. Spatiotemporal features of extreme rainfall events

Figure 5.6 presents the histograms of extreme rainfall systems in terms of total volume, maximum rainfall, duration, and maximum extension (area). According to the results, the total volume of the rainfall presented a positively skewed distribution commonly associated with these systems. On the other hand, maximum rainfall values of the events followed a normal distribution ranging from 5 to 50 mm/h with a median value of 30 mm/h. Concerning spatiotemporal characteristics, rainfall events are frequently longer in time (mean = 14 h STD = 6 h), typically varying between 8 and 18 h. In space, the maximum extension of the events ranges between 6000 and 12,000 km, covering between 4% and 8% of the total catchment area.

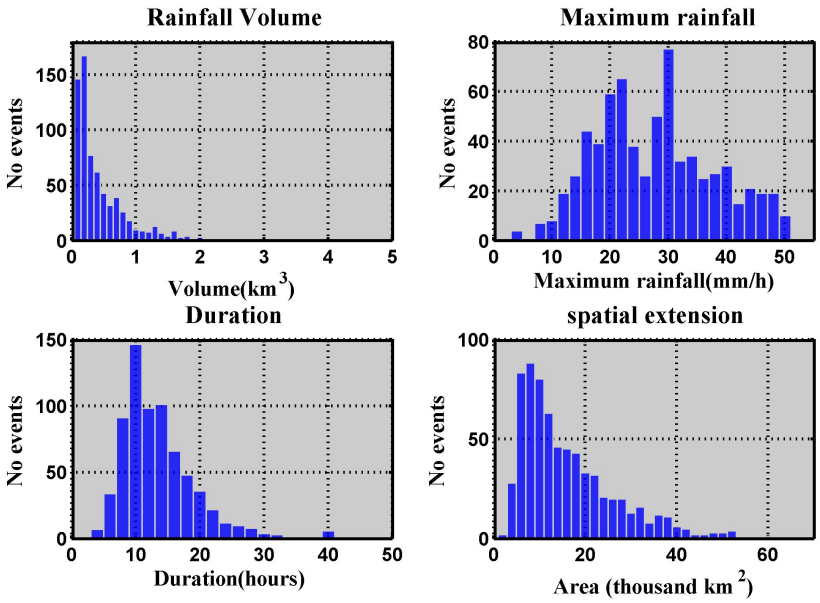


Figure 5.6: *Histograms of rainfall events characteristics. Rainfall Volume, Maximum rainfall, Duration and Spatial extension*

Based on the temporal and spatial characteristics of extreme rainfall systems, Fig. 5.7 presents the classification of extreme rainfall systems according to volume, duration, and maximum extension. In the study area, long-duration events are considered systems with a duration longer than 12 h, while spatially extensive events present a maximum extension of larger than 100 × 100 km^2. According to the results from 694 events, 314 (45.3%) are SLE events, 188 (27.1%) are LSE events, 133 (19%) are SEE events, and 59 (8.4%) are LLE events.

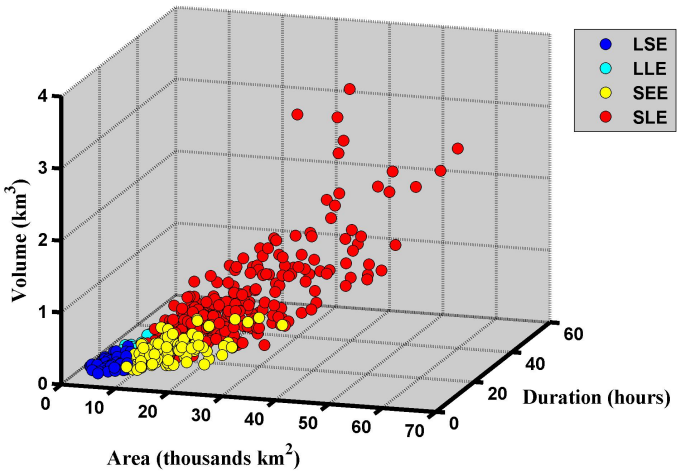


Figure 5.7: *Classification of extreme rainfall events.LLE, Local and long-duration extreme events; LSE, local and short extreme events;SEE, spatially extensive extreme events; SLE, long-duration and spatially extensive extreme events*

5.2.3. Application for satellite-based rainfall products verification.

The capacities of this methodology for product verification are evaluated for two types of extreme rainfall systems. The first event is an LSE event that occurred on February 12, 2011, between 16:00 and 21:00 h, causing several flash floods in the southeastern part of the Tiete River (IMPACT). The second event is an SLE event over the region between January 11 and 12, 2011. This rainfall event was one of the most critical extreme events in the area, where the extreme rainfall triggered massive floods and landslides over important cities of the state. Thirteen people lost their lives, and thousands lost their homes and other buildings (NOAA, 2011). NRT satellite-based information from CMORPH is compared against weather radar located in the study area. Both products are compared based on their main characteristics (maximum, volume, duration, and area), and their performance was evaluated grid to grid in terms of the level of displacement (concerning the weighted centroid), RMSE, and correlation coefficient. In this latter case, weather radar was upscaled to match the grid size of the satellite product.

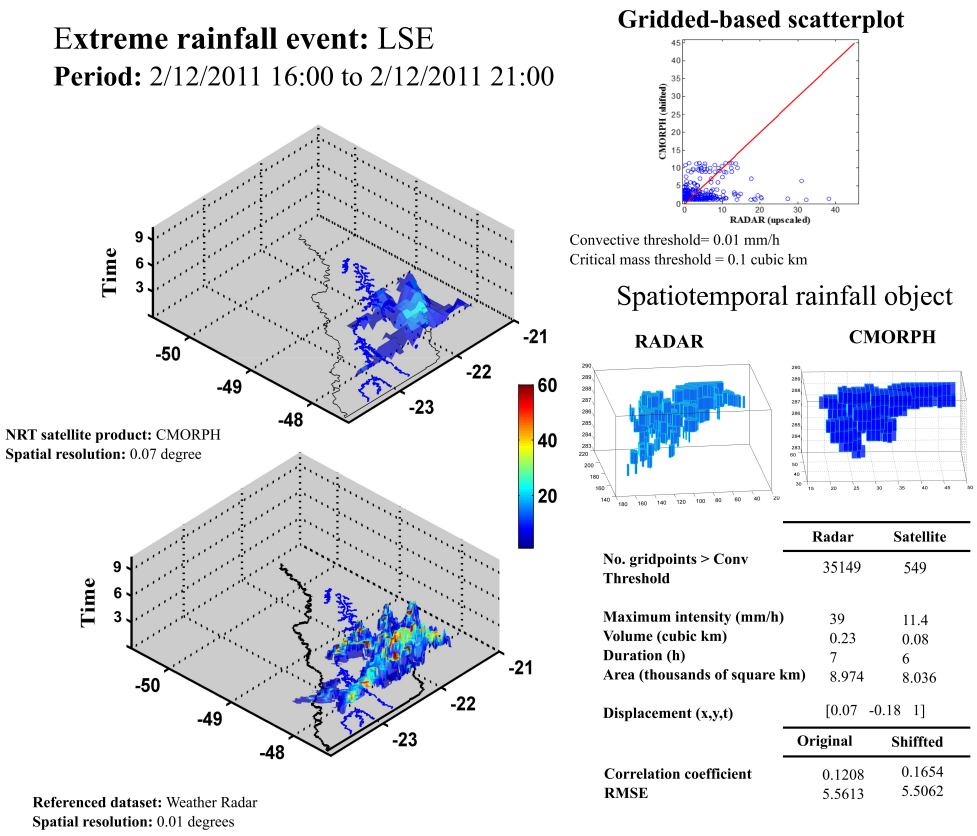


Figure 5.8: *Spatiotemporal verification for local and short extreme rainfall events. CMORPH, Climate prediction centre morphing technique; NRT, near-real time.*

Figure 5.8 displays the verification used for the LSE event of February 12, 2011.

The results showed a high rainfall concentration over the southeastern part of the region in both products. CMORPH product and the weather radar objects presented similar structures in terms of shape but had essential differences in terms of magnitude. CMORPH tended to underestimate the magnitude of the extreme event and the total volume of the system. The characteristics in space and time presented a slight displacement to the northwest in space and a displacement in time of 1 h. Radar and CMORPH products covered similar extensions (8974 and 8036 km2); however, CMORPH had a short duration (7 and 6 h). Based on two continuous verification metrics, CMORPH had a maximum correlation (shifted) of 0.16 and RMSE of 5.5 mm/h.

Verification of the SLE event on January 11 2011, is presented in Fig. 5.9. According to the results, CMORPH was able to capture important characteristics of extreme events. In the case of magnitude, both 4D objects presented similar characteristics. Maximum rainfall value was slightly overestimated (5 mm/h), as was volume (0.4 km^3). In the case of space and time characteristics, CMORPH presented a considerable shift to the northwestern part of the catchment. The maximum extension estimated covers 403,200 km^2 being greater than the radar estimate (323,500 km^2).

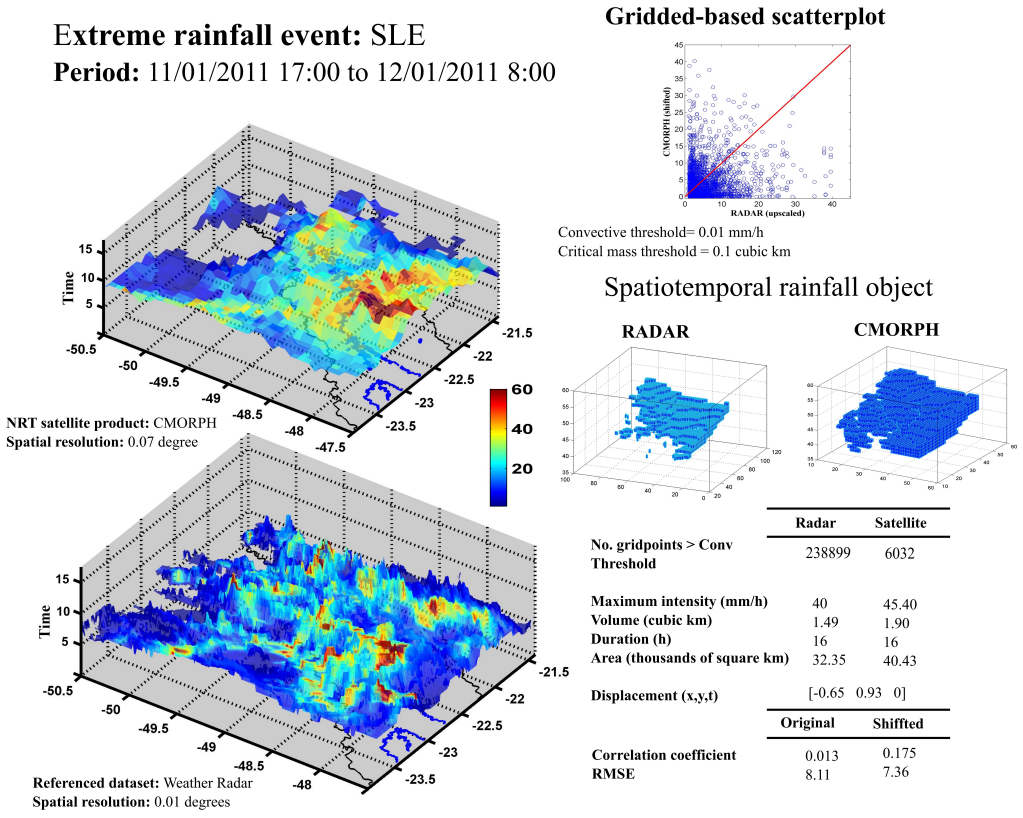


Figure 5.9: *Spatiotemporal verification for long-duration and spatially extensive extreme rainfall events. CMORPH, Climate prediction centre morphing technique; NRT, near-real time.*

On the other hand, the verification of the SLE event on January 11, 2011, is presented in Fig. 5.9. According to the results, CMORPH was able to capture essential characteristics of extreme events. In the case of magnitude, both 4D objects presented similar characteristics. Maximum rainfall value was slightly overestimated (5 mm/h), as was volume (0.4 km3). In the case of space and time characteristics, CMORPH presented a considerable shift to the northwestern part of the catchment. The maximum extension estimated covers 403,200 km2, being more significant than the radar estimate (323,500 km^2).

In both cases, the rainfall event had a duration of 16 h over the area. Evaluation in a gridded-based approach displayed a maximum correlation of 0.175 and an RMSE of 7.36 mm/h.

5.3. Conclusions

Object-based methods offer a unique perspective to analyse rainfall systems, providing a complete picture of the dynamics of rainfall fields. In this research, the analysis of rainfall events at the catchment scale further analyses extreme events in space and time. This method provides a complete diagnosis of extreme events identifying a large number of characteristics such as volume, area, duration, orientation, and speed, among others. The proposed methodology is very flexible and could be applied to different regions through the adjustment of two parameters: the convective threshold and the critical mass threshold. These parameters are susceptible, and they can be fundamental in the analysis. This method opens the door to discoveries in hydrological sciences toward understanding rainfall dynamics over catchments in extreme conditions. Possible new developments could lead to a better understanding of the flood generation process, evaluating the relationship between rainfall patterns and runoff. This methodology can induce new ways of describing flood events based on the spatiotemporal characteristics of rainfall events. Additionally, other remote sensing products may be evaluated using this approach, for instance, the Global Precipitation Measurement satellite product from NASA/JAXA (Huffman et al., 2015), among others. This method could be used to evaluate the capabilities of high-resolution rainfall products over a determined area.

6

Hydrological response of satellite-based error sources for extreme rainfall events

This chapter uses the ST-CORA method to evaluate the hydrological response of two systematic satellite error sources for rainstorm estimation: location in space and time for displacement and magnitude for volume. Both error sources from the Near Real-time (NRT) CMORPH product are subtracted based on gauged-adjusted weather radar and used as input forcing for hourly calibrated distributed hydrological model set up in the Capivari catchment, Brazil. Due to location and magnitude, synthetic rainstorm scenarios are created by adjusting the shift and intensity distribution of the SPP rainstorm object concerning the radar data. Two types of rainstorm events in the study area are evaluated: a short-lived and a long-lived rainstorm. The results indicate that the spatiotemporal characteristics obtained by ST-CORA reflect the primary source of errors of the CMORPH rainstorm detection. It is found that location is the primary source of error for the short-lived rainstorm event, while volume is the primary source in the long-lived rainstorm event. The subtraction of both errors leads to an essential reduction of the simulated streamflow in the catchment.

This chapter is partly based on the publication: Laverde-Barajas, M., Perez, GC, Chishtie, F., Poortinga, A., Uijlenhoet, R., & Solomatine, DP (2020). Decomposing satellite-based rainfall errors in flood estimation: Hydrological responses using a spatiotemporal object-based verification method. Journal of Hydrology, 591, 125554

6.1. Introduction

The spatiotemporal characteristics of rainstorm events, such as magnitude, duration and spatial extent, are some of the main triggering factors of flooding. In conjunction with intrinsic characteristics of the catchment, changes in the rainstorm structure can dramatically impact the severity of flood damage (e.g. Saulnier and Le Lay, 2009; Bui et al., 2014; Viglione et al., 2010). Therefore, an accurate representation of rainstorm dynamics is crucially vital in disaster preparedness and response (Phongsapan et al., 2019). However, the high spatial and temporal variability of rainstorm events in prominent monsoon regions still makes the estimation challenging.

A great deal of effort has been made to provide rainfall measurements with high spatial and temporal resolution. Recently, satellite-based precipitation products (SPP) have widely used in hydrological applications for overcoming the lack of spatial representation of rain-gauges (e.g. Artan et al., 2007; Su et al., 2008; Stisen and Sandholt, 2010; Nikolopoulos et al., 2013). Despite numerous advances, satellite products are subject to several systematic and random errors from multiple sources (e.g Hu et al., 2016; Qiao et al., 2014; Dinku et al., 2010; Guo et al., 2015; Mei et al., 2014; Sapiano and Arkin, 2009; Thiemig et al., 2012; Poortinga et al., 2017). These errors have several implications in relation to their input for modelled flood response (e.g Casse et al., 2015; Bitew and Gebremichael, 2011; Mei et al., 2016b; Vergara et al., 2014; Yilmaz et al., 2005; Li et al., 2009).Maggioni and Massari (2018), analysing the lessons from using satellite precipitation products for riverine flood modelling around the world, argued that the performance of SPP-forced hydrological model depends on several factors including: the type of SPP sensor; the precipitation type; the geomorphological conditions; and the hydrological model formulation.

Several analytical methods have been used to analyse the error decomposition in satellites and evaluate the spatiotemporal dynamics in terms of streamflow (e.g. Viglione et al., 2010; Mei et al., 2017a,b). However, the error decomposition is limited by the configuration of the hydrological model. In contrast to analytical methods, rainfall object-based methods represent the cells and integrate spatiotemporal information from an event as an interconnected mass figure characterising the rainstorm events. These object-based methods offer an alternative way to analyse the error in rainstorm events (e.g Skok et al., 2009; Li et al., 2015a, 2016). These methods, applied in hydrological modelling, can evaluate the impact of several systematic SPP errors in forecasted streamflow, independently of the model configuration. For example, Demaria et al. (2011) used the Contiguous Rainfall Analysis (CRA, Ebert and McBride (2000)) to evaluate the hydrological impact of location errors presented in TMPA,CMORPH and PERSIANN SPPs. The analysis showed that errors due to location represented more than 65% of the total errors affecting the streamflow peak, thus also its volume.

Important challenges in object-based methods arise from the effect of displacements calculated in space and time. The two-dimensional approach of existing methods limits the description of the rainstorm event in space and time. Several researchers have Incorporated the temporal domain into object-based spatial methods to identify the main feature of convective rainstorm events in space and time. (e.g. Davis et al., 2009a; Mittermaier and Bullock, 2013; Sellars et al., 2013; Laverde-Barajas et al., 2019). Recent methods such as the Spatiotemporal Contiguous Object-based Rainfall Analysis (ST-CORA, Laverde-Barajas et al. (2019)) have been applied to evaluate the errors from NRT SPP for predicting rainstorms at catchment scale.

In this study, we used the ST-CORA method to evaluate the hydrological impact of two systematic error sources in the near real-time CMORPH for rainstorm event estimation: the error due to location and magnitude. Synthetic rainstorm scenarios, created by the subtraction of location and magnitude error sources, are propagated through the distributed hydrological model wflow-sbm Schellekens (2018) in the subtropical catchment of the Capivari River, Brazil. The model is calibrated at an hourly scale during 2016 using a gauge-corrected weather radar as reference data. Two types of rainstorm events are analysed: local and short events (short-lived rainstorms) and long duration, spatially extensive, extreme events (long-lived rainstorms). This approach can better understand the hydrological response to systematic errors in SPPs, especially in extreme conditions.

6.2. Methodology

6.2.1. Rainfall object estimation, ST-CORA

ST-CORA is a spatiotemporal object-based method developed by Laverde-Barajas et al. (2019) to analyse the spatiotemporal characteristics of different rainstorm events at catchment scale (duration, spatial extent, magnitude, and centroid). This methodology uses a multidimensional connected-component labelling algorithm to detect regions with similar features in space and time. Spatiotemporal rainstorm event objects are defined based on two parameters: the rainfall intensity threshold (IT) and the Critical Mass Threshold (CMT). IT defines the common intensity threshold for associating neighbouring groups of voxels (volume representation of a pixel). Based on the sensitivity analysis made by Laverde-Barajas et al. (2019) in the region, we selected an IT of 1 mm/h and a CMT of 0.01 [3] to identify and segment extreme convective objects. After obtaining the convective rainfall object, ST-CORA applies two error filtering algorithms to remove small noisy objects and solves the false merging errors from the connected-component labelling algorithm (Ceperuelo et al., 2006; Han et al., 2009). In this study, we defined noise to objects lower than four connected voxels in space and time ($voxel = 0.1°/hour$).

Following hydrometeorological and dynamical criteria to classify severe rainfall events proposed by (Molini et al., 2009, 2011; Boers et al., 2015), we classified rainstorm events in two groups depending on the maximum area, the rainstorm duration and the volume of rainfall: small convective systems with a short duration (short-lived) and long duration systems extended over large areas (long-lived). Short-lived events as extreme convective objects with a duration of less than 12 hours and spatial distribution greater than 86 x 86 km^2 Laverde-Barajas et al. (2018, 2019)

Spatiotemporal error analysis using ST-CORA

The error analysis for rainstorm detection using ST-CORA compares the spatiotemporal features detected by SPP against grid-ground data used as reference. Several error metrics can be evaluated based on this comparison. In this study, we used the error decomposition from the Contiguous Rainfall Analysis method (CRA) developed by Ebert and McBride (2000). This method analyses the spatial elements of the SPP error, decomposing the total error into displacement, volume and pattern error as shown below.

$$MSE_{total} = MSE_{displacement} + MSE_{volume} + MSE_{pattern} \tag{6.1}$$

Where

$$MSE_{displacement} = MSE_{total} - MSE_{location} \tag{6.2}$$

$$MSE_{volume} = \left(\overline{Y} - \overline{O}\right)^2 \tag{6.3}$$

$$MSE_{pattern} = MSE_{location} - MSE_{volume} \tag{6.4}$$

Where $MSE_{location}$ is calculated as the mean square error of the spatial shifted SPP, and $\overline{Y}$ and $\overline{O}$ are the mean satellite and observed values after the shift. In the context of satellite-based rainfall data, $MSE_{displacement}$ and MSE_{volume} are considered as systematic errors, while $MSE_{pattern}$ corresponds to intrinsic random errors in the measurements, which are calculated as the residual error of systematic errors (Hoffman et al., 1995; Ebert and Gallus, 2009). It is important to highlight that systematic errors integrate random errors due to non-corrected values from location and magnitude subtraction. In this research, we considered non-corrected errors as part of pattern errors in the rainstorm event.

6.2.2. Systematic error source subtraction

The use of ST-CORA allows for the temporal and spatial description of rainstorm events and identifies the primary sources of error in the measurement. Once the main rainstorm characteristics are identified (volume, maximum intensity, max area and duration), we subtracted location error as the primary source in displacement and magnitude as the main volume source by the following approach.

Location subtraction

In this research, we used the Principal Component Analysis method (PCA) proposed by Johnson and Hebert (1999) to remove the shifting and orientation effects of the satellite rainstorm in space and time (Grams et al., 2006) (Fig.6.1). PCA is a statistical method that describes the pattern of multi-variable data based on orthogonal variables called principal components (Abdi and Williams, 2010). Based on an ellipsoid wrapped over the object with weighted-centroid (WC), the methodology calculates the eigenvectors of the object (Fig.6.1a). These vectors describe the orthogonal basis in space and time, in which the object position is described (Johnson and Hebert, 1999). Location is subtracted by finding the best match between the observed and the satellite rainstorm. Location error subtraction first removes the shifting effect by matching the weighted-centroids of both objects in space and time, and then the orientation is adjusted by rotating the temporal eigenvector in the direction of the observed vector (Fig. 6.1b, c).

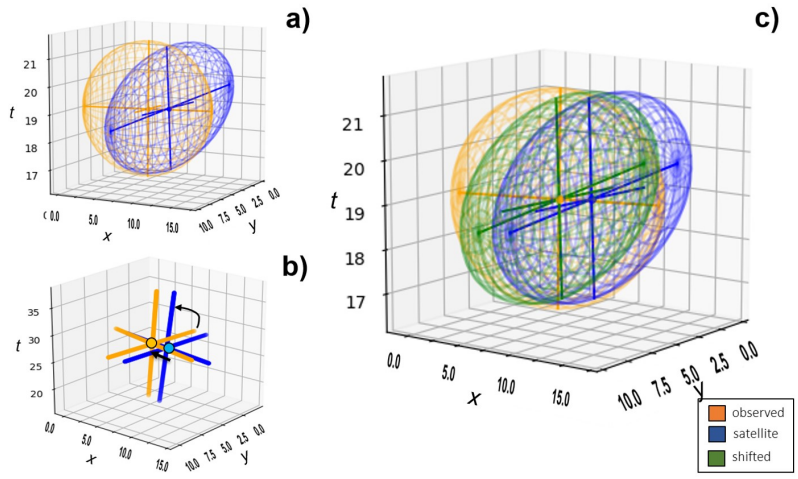


Figure 6.1: *Scheme of object matching for location subtraction. a) eigenvectors of a compressed the rainstorm object b) weighted centroid and eigenvector matching. c) shifted rainstorm object*

Magnitude subtraction

The magnitude subtraction statistically adjusts all moments (quantiles) of the intensity probability distribution function for the satellite rainstorm for the observed rainstorm. In this case, the adjusted satellite rainstorm intensity distribution should reflect the function of the observed rainstorm. Several methodologies have been developed to equalise the histograms and remove the bias from the estimated data (e.g Cannon et al., 2015a; Themeßl et al., 2012; Teng et al., 2015). For this research, we selected the extensively used Empirical Quantile Mapping method (EQM) (Themeßl et al., 2012). This method adjusts all moments of the cumulative probabilistic function ($ecdfs$) of the satellite's intensity for the observed data. The relationship between the $ecdf$ for observed ($ecdf_{obs}$) and the satellite ($ecdf_{sat}$) is based on the following equation:

$$EGM = ecdf_{obs}^{-1}(ecdf_{sat}^{-1}(I_s)) \tag{6.5}$$

where I_s is the intensity rainstorm distribution detected by the satellite.

Once the magnitude error source is subtracted, the volume equation, Eq. 6.3, can be reformulated to included the magnitude subtraction as follows:

$$MSE_{volume} = MSE_{total} - MSE_{magnitude} \tag{6.6}$$

6.2.3. Evaluation of the hydrological response

Synthetic scenarios, created by the individual subtraction of location and magnitude error sources from rainstorm events detected by satellites, are propagated through a hydrological model to evaluate the error response in terms of streamflow. The hydrological response is evaluated by comparing simulations from satellite scenarios against simulations from observed rainfall. We used several metrics to analyse the shape, phase and amplitude of the streamflow along the catchment. These metrics include the correlation coefficients (r), Root Mean Square Error ($RMSE$), the Mean Absolute Peak Time. Error ($MAPTE$) (Ehret and Zehe, 2011), Percentage Peak Effect index (PPE) and the Percentage Volume Effect (PVE) (Bennett et al., 2013).

$$r = \frac{\sum_{t=1}^{N}(Q_{rad,t} - \overline{Q_{rad}})(Q_{sat,t} - \overline{Q_{sat}})}{\sqrt{\sum_{t=1}^{N}(Q_{rad,t} - \overline{Q_{rad}})^2}\sqrt{\sum_{t=1}^{N}(Q_{sat,t} - \overline{Q_{sat}})^2}} \tag{6.7}$$

$$RMSE = \sqrt{\frac{1}{N}\sum_{t=1}^{N}(Q_{rad,t} - Q_{sat,t})^2} \tag{6.8}$$

$$MAPTE = \frac{1}{N}\sum_{t=1}^{N}|P_{rad,i} - P_{sat,i}| \tag{6.9}$$

$$PPE = \frac{Q_{rad_{max}} - Q_{sat_{max}}}{Q_{rad_{max}}} * 100 \tag{6.10}$$

$$PVE = \frac{V_{rad} - V_{sat}}{V_{rad}} * 100 \tag{6.11}$$

Here, $Q_{rad_{max}}$ and $Q_{sat_{max}}$ are the maximum discharge simulated by the radar (referenced value), and the satellite discharge values (original, location and magnitude) for each discharge station. $P_{rad,t}$ and $P_{sat,i}$ are the peak times and V_{rad} and V_{sat} are the simulated total volume based on the radar and the satellite, respectively.

The streamflow error variation between CMORPH and the location and magnitude scenarios for the short- and long-lived rainstorms is evaluated by the following equation:

$$Errorvariation_t = \frac{Q_{sat,t} - Q_{rad,t}}{Q_p} * 100 \tag{6.12}$$

where Q_p is the maximum discharge value reached in all rainstorm scenarios.

6.3. Study area and data available

The study area corresponds to the subtropical catchment of the Capivari River in south-eastern Brazil. In chapter 3 is presented a description of this area. We selected two rainstorm event types based on the Integrated Disaster Database system from the National Civil Protection Secretary (S2ID, SEDEC (2013)). The first event is a short-lived rainstorm that occurred on 1 January 2011. The second event is a long-lived rainstorm that took place between 3 and 5 January 2012. Both events drastically impacted the region, causing several flash floods and landslides in the catchment.

6.3.1. Rainfall data

In this study, we used the CPC MORPHING product (CMORPH) developed by the NOAA Climate Prediction Centre (Joyce et al., 2004). This data combines Passive Microwave and Infrared sensor information to produce high-resolution data every 30 minutes. We selected the Near-real-time version of CMORPH NRT due to superior performance for extreme rainfall estimation in the study area (Laverde-Barajas et al., 2018). We used CMORPH 1.0 raw version at an 8x8 km scale (0.727 degrees) and 30-minute resolution (1h aggregated).

Ground measurements were obtained from the Bauru weather radar station (Bauru CAPPI 3.5 km) from the Faculty of Science, Sao Paulo State University, located at 22° 21' 28" S latitude, 49° 01' 36" W longitude (Fig. 3.2c). This radar provides information at 1km^2 spatial resolution every 15 minutes. Weather radar rainfall data was gauged adjusted at hourly scale using eight automatic ground stations from the Integrated Centre of Agrometeorological information (CIIAGRO, Centro Integrado de Informações Agrometeorológicas) applying the reduction of the mean-field described in Amorati et al. (2012) (Eq. 6.13)

$$R_{merged} = RF, \tag{6.13}$$

where, R is hourly rainfall field and F is the weighted correction factor with temporal averaged windows, Δt, calculated as follows:

$$\overline{\log F_t} = \frac{\sum_{\tau=t-\Delta t}^{t} \sum_{i=1}^{N_\tau} \log_{10}\left(\frac{G_{i,t}}{R_{i,t}}\right)}{\sum_{t-\Delta t}^{t} N_\tau}, \tag{6.14}$$

where, G_i and R_i are the Automatic Weather Stations (AWS) and radar measurements at gauge i, respectively. A three-hour temporal window, Δt, is considered for the $\overline{\log F_t}$. In order to reduce the effect of differences in the spatial resolution (Peleg et al., 2018), we scaled up weather radar data by the mean aggregation to match the spatial resolution of the CMORPH dataset.

6.4. Capivari model

We used the distributed hydrological model wflow-sbm (version 2018.1) to represent the rainfall-runoff process in the Capivari River. This model was developed by project Deltares OpenStreams (Schellekens, 2018) based on the topog-sbm model (Vertessy and Elsenbeer, 1999). The wflow-sbm simulates the hydrological routing through a gridded mesh into a GIS environment called PCRaster-Python (Wesselung et al., 1996).

The representation of the hydrological cycle within wflow-sbm is divided into the three main routines: interception, soil, and water surface. Wflow-sbm simulates interception using the analytical Gash model (Gash et al., 1995). This model calculates the actual evapotranspiration of the canopy based on the Potential Evapotranspiration (PET), soil water content and land cover type. The soil routine runs per grid cell through the topog-sbm model. This model is a simple bucket scheme designed to transfer the water infiltration process into a saturated (S) and unsaturated store (US). The S and US interaction is based on the exponential decay of the saturated hydraulic conductivity with depth below the soil surface (Ksat) (Schellekens, 2018). The surface water routine uses the kinematic wave algorithm to simulate the river drainage and overland flow. The runoff is calculated per grid cell as the sum of residual rainfall water from the interception and soil routines and the horizontal and vertical interaction between cells estimated in the routing process.

6.4.1. Model setup

For each rainfall data product, we setup the wflow-sbm model in a 1km x 1km grid integrating several static and dynamic inputs. Static inputs are integrated by *i)* the Digital Elevation Model from the Shuttle Radar Topography Mission (SRTM) version 3.0 (Van Zyl, 2001); *ii)* land use map extracted from the Brazilian land use map of the National Institute of Geography and Statistics (IBGE, 2017) for 2014; and *iii)* the soil type map, extracted from the FAO Digital Soil Map of the World (DSMW-FAO, 1960).

In addition to the rainfall data, dynamic inputs are hourly air temperature (T) in degree Celsius and potential evapotranspiration (PET) maps based on the AWS from CIIAGRO. PET is calculated using the method of Abtew (1996) recommended by Tangune and Escobedo (2018) for the study area. This method calculates PET in mm/h using the air temperature and solar radiation as

$$PET = \frac{0.53}{2.501 - 0.002361T} Rs, \tag{6.15}$$

where Rs is the solar radiation data from AWS. Hourly T and PET data were interpolated using the Inverse Distance Weight method (IDW-2). This method is widely used in meteorology, and it was selected for providing reliable hourly information based on the total number of available stations Chung and Yun (2004); Chen et al. (2019). However, it is noted that interpolation methods are subject to several sources of uncertainty due to the spatial sampling and geomorphological conditions in the catchment. These uncertainties include the misrepresentation between meteorological variables and elevation DeGaetano and Belcher (2007).

6.4.2. Model calibration and validation

Wflow-sbm requires 15 input parameters to be calibrated in the study area (excluding the input snow parameters). We calibrated the Capivari hydrological model using hourly discharge data from the Early Warning Flood System of Sao Paulo state (SAISP) available in the area, with the upper zone calibrated based on the Campinas station between January and October 2016. In the middle zone, the model was calibrated based on the Monte station between July and October 2016. In the absence of hourly discharge data in the lower zone, the model was calibrated based on a virtual discharge point in the outlet area calculated using the spatiotemporal linear estimator proposed by Paiva et al. (2015):

$$\hat{Q}(S_o, t_o) = \sum_{i=1}^{N} \lambda_i Q(S_i, t_i) \tag{6.16}$$

Here, $\hat{Q}$ is the virtual station at location S_i, Q are Campinas and Monte discharges and $\lambda = [\lambda_1 ... \lambda_n]^T$ is the set of weights for each measurement. For the study area, λ represents the relative drainage area of each discharge point. All areas were validated during November and December 2016.

The calibration process started by adjusting the input parameters to the most realistic interval, and then we selected the most sensitive ones using a Monte Carlo Analysis. Table 6.1 shows the most sensitive input parameters in the Capivari model. Using the Augmented Lagrangian Harmony Search Optimizer (ALSHO, (Geem et al., 2001)) from Perez et al. (2012), we calibrated the model for the three zones, from upstream to downstream, using the Nash-Sutcliffe efficiency (NSE, Nash and Sutcliffe (1970) as the objective function defined as:

$$NSE = 1 - \frac{\sum_{i=1}^{N} (Q_{obs,i} - Q_{sim,i})^2}{\sum_{i=1}^{N} (Q_{obs,i} - \overline{Q_{sim}})^2} \tag{6.17}$$

Where $Qobs$ and $Qrad$ are the model's observed and simulated streamflow, respectively, using the gauge-corrected weather radar as the input.

Model calibration and validation were additionally compared with the Root Mean Square Error ($RMSE$) (Eq. 6.8) and correlation coefficient (r) (Eq. 6.7). Table 6.1 describes the model performance over the three zones based on NSE, $RMSE$, R. Model results displayed satisfactory results for estimating hourly streamflow along the catchment with NSE ranging between 0.6/0.5 and 0.8/0.6 in calibration/validation and having a correlation coefficient higher than 0.76 and $RMSE$ lower than 20 mm on the evaluated areas.

Table 6.1: *Performance criteria of model calibration and validation at different stations based on NSE, RMSE and R*

Station	Campinas		Monte		Outlet	
	Calibration	Validation	Calibration	Validation	Calibration	Validation
NSE	0.78	0.54	0.61	0.58	0.73	0.52
RMSE	3.52	2.57	6.17	8.99	16.23	19.98
R	0.89	0.76	0.81	0.77	0.91	0.72

6.5. Results

6.5.1. Spatiotemporal errors of satellite-based CMORPH

Detecting systematic errors in satellite-based rainstorm events using ST-CORA

Figure 6.2 presents the object structure and rainstorm properties of the reference radar and the satellite-based CMORPH identified by ST-CORA for the short-lived rainstorm event that occurred on 1 January 2012 and the long-lived rainstorm event that occurred between 3 and 5 January 2011. Figure 6.2 a,b present rainstorm objects of the radar and satellite-based CMORPH for the short-lived event, while figures 6.2 d,e show the object structures for the long-lived event. Figures 6.2 c, f present the comparison of the scatter plots and the spatiotemporal properties of both objects for both rainstorm events.

According to the results, radar and CMORPH exhibit similar spatiotemporal structures but differ in the object's magnitude and location. CMORPH performance for the short-lived event showed an underestimation of high-intensity rain rates (Fig. 6.2c). However, the total rainstorm volume was higher compared to the rainstorm detected by the radar. The spatiotemporal properties are characterised by a mismatch in the orientation of the CMORPH event. In time, both objects presented the same duration (7h), while in space, CMORPH had the most considerable extension. The satellite performance for the long-lived event displayed a low scatter correlation, overestimating both the total rainstorm volume and the maximum volume intensity. Regarding the spatiotemporal characteristics, CMORPH and the weather radar shared the same

orientation but differed in the maximum extension and the duration of the rainstorm. CMORPH event was more extended with a longer duration (1 hour) than the weather radar rainstorm.

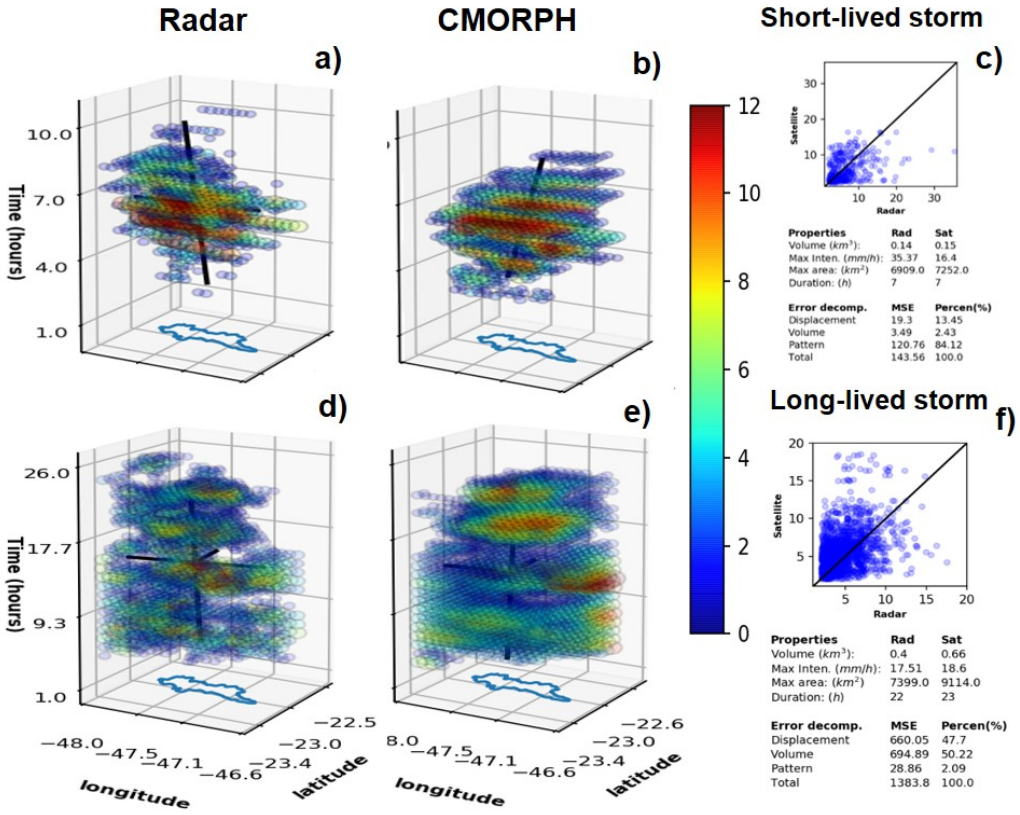


Figure 6.2: *Object structure and rainstorm properties of the reference Radar and the satellite-based CMORPH identified by ST-CORA. a), b), c) short-lived rainstorm and d), e), f) long-lived rainstorm events*

One of the main aims of an object-based approach is to decompose the total error into the sum of individual errors due to displacement, volume and pattern. The results obtained for the short-lived event indicate the pattern (aleatory error) of the primary error source (83%) followed by displacement (13%). In the long-lived rainstorm, volume and displacement contributed 47% and 50% of the total error, respectively.

Satellite error subtraction

After the rainstorm objects were identified, the satellite estimates due to location and magnitude were subtracted in the satellite data to create synthetic rainstorm scenarios. To illustrate the spatial distribution of the rainstorm scenarios over the Capivari catchment, Figure 6.3 presents the total rainstorm maps and the rainfall accumulation over the river streams based on radar CMORPH and synthetic rainstorm scenarios for the short-lived rainstorm. In contrast, figure 6.4 describes the corresponding maps for the long-lived rainstorm event.

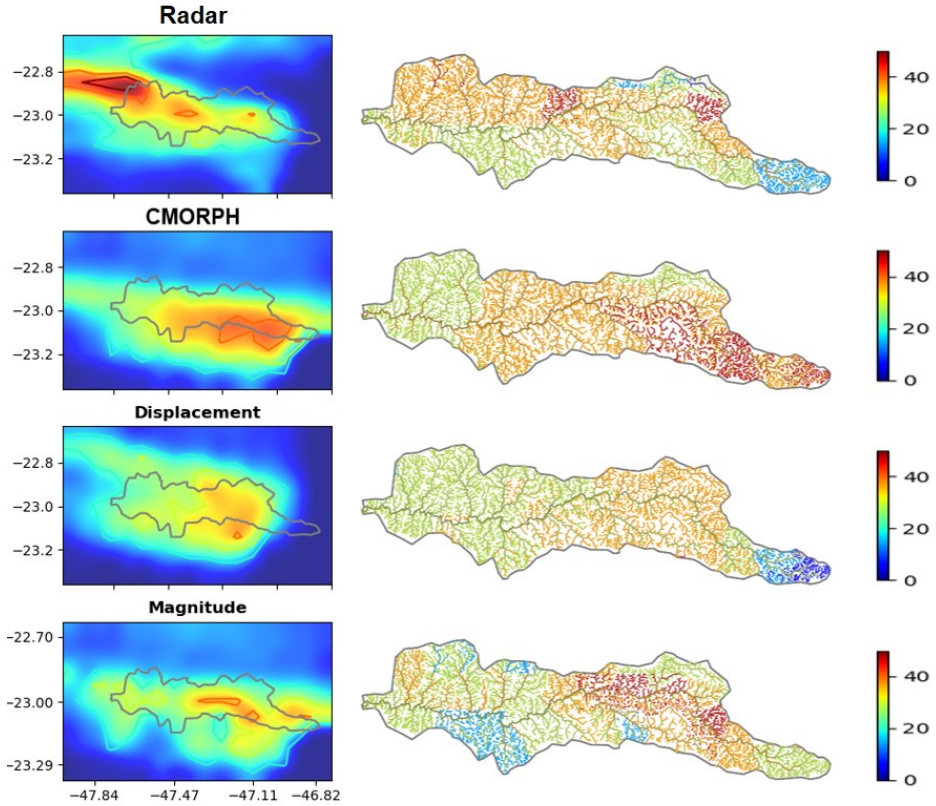


Figure 6.3: *Total event maps and rainfall accumulation by RADAR, CMORPH and synthetic rainstorm scenarios for short-lived rainstorm*

The results for the short-lived rainstorm scenario highlights the overestimation of CMORPH in the upstream area of the catchment. The location error scenario displaced the rainstorm event to the northwest of the catchment, smoothing the rainfall excess of the rainstorm event. Due to the magnitude scenario, the rainstorm scenario considerably reduced the overestimation in CMORPH, decreasing the total volume in the upstream area. The scenario for the long-lived rainstorm event (Figure 6.4) showed a significant rainfall overestimation in the northwestern part of the catchment, affecting the middle and lower areas. The location scenario slightly adjusted the rainstorm event in the northwest, reducing the rainfall excess of the upstream area. However, this scenario still overestimated rainfall in the middle part of the catchment. Lastly, the magnitude rainstorm error scenario effectively reduced extreme rainfall, overestimating rainfall over the catchment at small scales.

6.5.2. Hydrological impact

The hydrological response of systematic error due to location and magnitude was analysed over three zones along the catchment: Upper (Campinas), Middle (Monte), and Lower (Outlet). Figures 6.5 and 6.6 present the hydrological effect of systematic errors in CMORPH for the short- and long-lived scenarios using the hydrological simulation

based on the radar rainstorm for reference. Hydrographs have been normalised for the peak volume, Q_p, and the peak time, t_p, according to $(Q - Qmin)/Q_p$ and t/t_p for a better description of the hydrological impact. Table 6.2 summarises the evaluation of hydrological impact to the location and the magnitude scenarios in the streamflow along the catchment-based on r, $RMSE$, $MAPTE$, PPE and PVE.

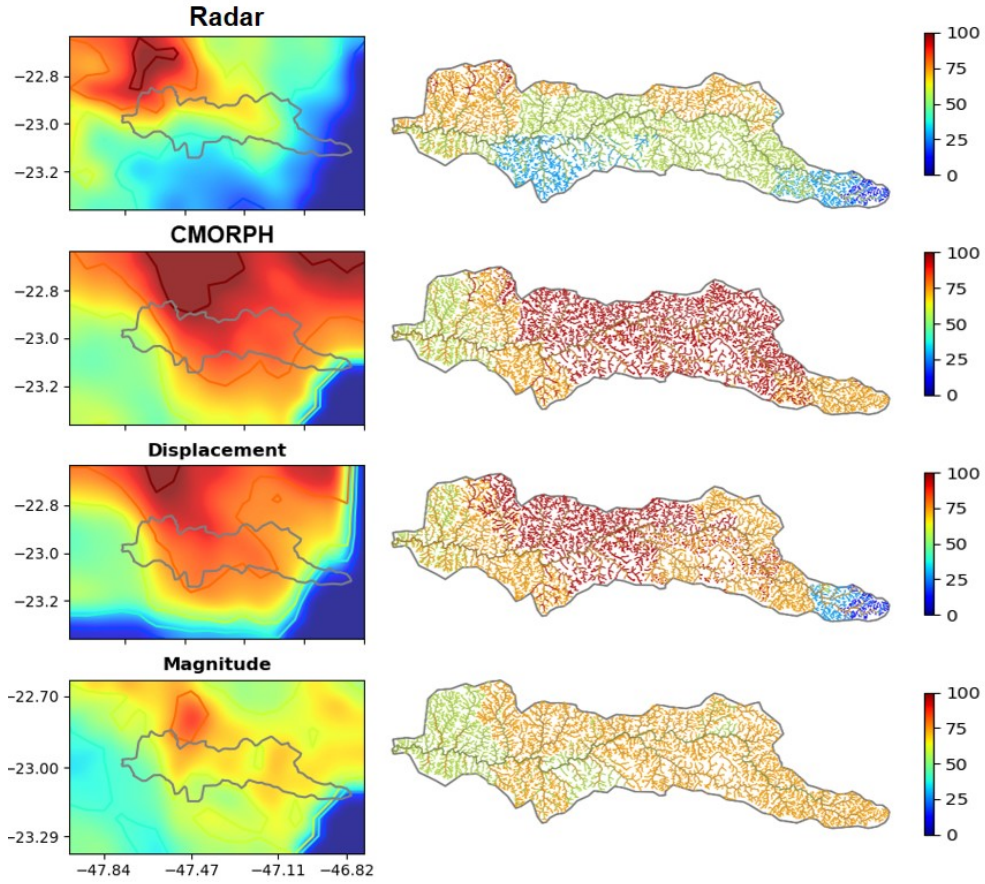


Figure 6.4: *Total event maps and rainfall accumulation by RADAR, CMORPH and synthetic rainstorm scenarios for long-lived rainstorm*

The results for the short-lived rainstorm event (Fig. 6.5) suggest that the overestimation of CMORPH in the rainfall rainstorm consequently increased the streamflow volume over the catchment. The upper and lower areas, corresponding to Campinas and Monte, showed an overestimation in the total volume of the rainstorm of 70% and 40%, respectively, due to the considerable overestimation of the peak flow. On the contrary, in the Outlet area, the level of overestimation was minimal, just 3% of the volume. Regarding the systematic errors, both scenarios decreased the streamflow volume and slightly increased the streamflow correlation over the catchment. The location scenario presented a significant decrease in RMSE in Campinas and Monte. However, the RMSE was high in the Outlet area due to underestimating 4% of the streamflow volume. Regarding the hydrological effect in the temporal response of the rainstorm, location and magnitude scenarios impacted the peak delay of 1 hour at Monte and 1

to 3 hours at Outlet.

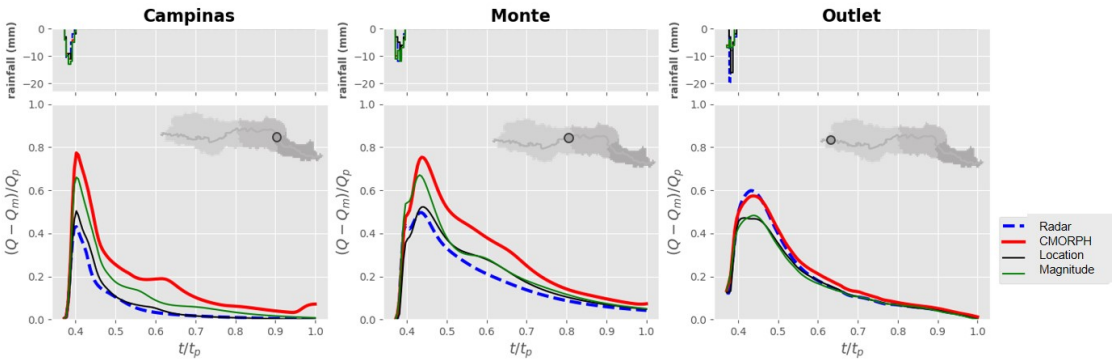


Figure 6.5: *hydrological response for short-live rainstorm. Upper plots represent the rainfall hietograms and lower plots represent the corresponding discharges for Campinas, Monte and outlet zones*

The hydrological response for the long-lived rainstorm event (Fig. 6.6) is marked by the significant effect of the rainfall overestimation by CMORPH over the catchment. CMORPH systematically over-predicted streamflow in Campinas, Monte and Outlet by 138%, 86% and 26%, respectively. The systematic error scenarios reduced the peak flow excess and consequently the RMSE in Campinas and Monte. However, the location did not affect the Outlet volume streamflow error. The location scenario had an essential impact in Campinas, displaying the smallest RMSE and the higher r. On the other hand, The magnitude scenario was more relevant in Monte, especially in Outlet, which had a small RMSE and higher r. For MAPTE, the hydrological response showed a systematic delay of the peak in CMORPH as much as 3 hours in Monte and 1 hour in Outlet. The location scenario did not seem to impact time delay, but the magnitude scenario exhibited a 1-hour additional delay in Monte and zero delays in Outlet.

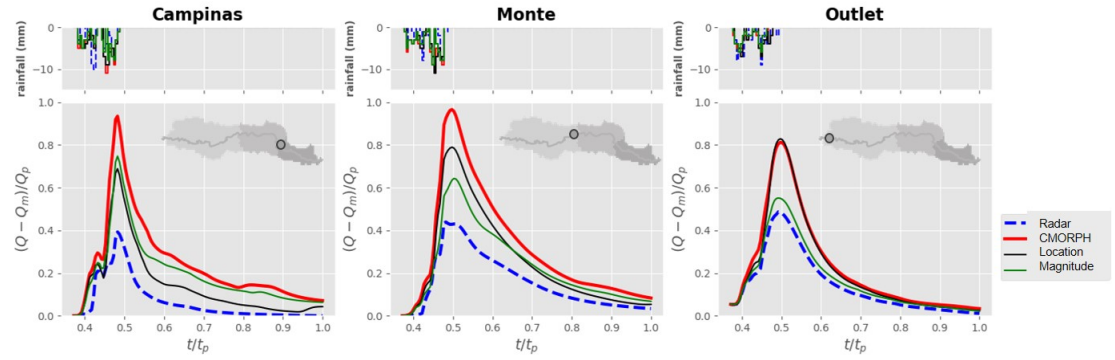


Figure 6.6: *hydrological response for long-lived rainstorm. Upper plots represent the rainfall hietograms and lower plots represent the corresponding discharges for Campinas, Monte and outlet zones*

Table 6.2: *Hydrological response evaluation for short and long-lived rainstorm at Campinas, Monte and Outlet zones corresponding to Original, Location and Magnitude scenarios*

rainstorm type	Statistical index	Campinas			Monte			Outlet		
		org	location	magnitude	org	location	magnitude	org	location	magnitude
Short-lived rainstorm	R	0.964	0.991	0.986	0.974	0.974	0.994	0.997	0.995	0.996
	RMSE (mm)	4.99	0.89	2.92	7.03	2.17	3.39	5.91	11.70	12.08
	MAPTE (h)	0	0	0	0	1	1	1	3	1
	PPE (%)	(64.40)	(13.80)	(43.20)	(44.88)	(4.63)	(30.38)	2.63	13.53	12.36
	PVE (%)	(68.59)	(6.26)	(34.78)	(41.84)	(9.23)	(18.80)	(2.67)	3.47	3.89
Short-lived rainstorm	R	0.938	0.961	0.938	0.992	0.993	0.991	0.996	0.990	0.999
	RMSE (mm)	12.60	6.16	8.79	25.69	16.84	11.31	73.71	71.71	18.90
	MAPTE (h)	0.00	0.00	0.00	3.00	3.00	4.00	1.00	1.00	0.00
	PPE (%)	(118.58)	(64.76)	(77.62)	(111.54)	(74.15)	(43.21)	(49.46)	(51.99)	(9.83)
	PVE (%)	(137.92)	(59.06)	(99.22)	(86.40)	(51.86)	(41.94)	(26.22)	(22.76)	(7.97)

6.5.3. Streamflow error comparison

Figure 6.7 describes the dispersion of the error variation corresponding to CMORPH and location and magnitude streamflow scenarios for short-lived (Fig 6.7a) and long-lived rainstorm (Figure 6.7b) events. Boxes represent the error variation dispersion between 25-75 percentiles, and dots represent values outside the 99.9% percentile (backlines). The positive percentile represents flow overestimation, while the negative percentile represents underestimation. Based on the results, overestimation prevails for both rainstorm events over the catchment. The location scenario displayed the lowest error variation (closest to zero) in the short-lived rainstorm, while the magnitude scenario showed the slightest variation in the long-lived rainstorm. These results demonstrate that the error due to location in CMORPH had a dominant influence on the hydrological response of short-lived rainstorm events. Meanwhile, error due to magnitude had a more considerable influence on the hydrological response of the long-lived rainstorm event.

6.6. Discussion

In the previous sections, we showed the evaluation of the impact of two systematic sources of error in CMORPH on the hydrological response in the Capivari catchment, Brazil. By using the ST-CORA method, we described the spatiotemporal characteristics of two meteorological rainstorm events. Differences between physical properties of observed and estimated objects allowed for the subtraction of two types of error: due to location and magnitude. Corresponding systematic scenarios were later propagated through a calibrated hydrological model to evaluate the individual hydrological impact of each source of error. In comparison to analytical approaches, this method analyses the error decomposition independently of the hydrological model configuration.

The results indicate that CMORPH overestimates the total volume of both rainstorm events, in agreement with multiple findings in South America (Ebert et al., 2007; Demaria et al., 2011; Laverde-Barajas et al., 2018). Aleatory errors exhibited an important error component in CMORPH, especially in the short-lived rainstorm event,

representing almost 85% of the total error. Regarding systematic errors, the location was the primary source of error for the short-lived rainstorm events, while magnitude had a more significant influence on the long-lived rainstorm event. Concerning the rainstorm structure, errors CMORPH for the short-lived event is mainly attributed to the description of the rainstorm orientation. In the case of the long-lived event, those are mainly focused on the spatiotemporal dimension of the rainstorm. Despite that, these results do not represent the general performance of CMORPH for different rainstorm types. These findings provide a framework to analyse the systematic error as a function of the type of rainstorm event.

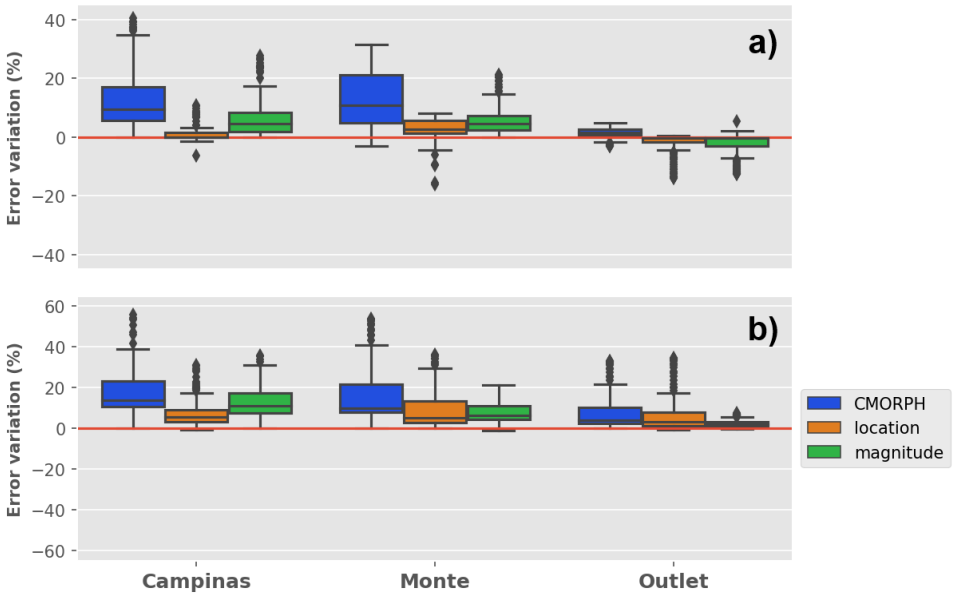


Figure 6.7: *Boxplots of the error distribution contribution for the original CMORPH event and location and magnitude stream-flow scenarios. a) short-lived rainstorm b) long-lived rainstorm*

Regarding the hydrological response, the systematic overestimation by the satellite consequently leads to a significant excess in the streamflow over the catchment. Errors in CMORPH due to location had a more substantial effect on the short-lived rainstorm, mainly delaying between 1 to 3 hours the time of the flow peak along catchment. In the analysis, the *RMSE* for location decreased by 70% and 80% RMSE in the upper and middle zones, respectively. Results in the long-lived event showed that location subtraction had a low influence on the hydrological response. For this event, the location subtraction slightly reduced the streamflow volume in Campinas and Monte, corresponding to elevated medium altitude areas and delaying the 1-hour peak flow of the streamflow in the Outlet area. On the other hand, Errors in CMORPH due to volume were more critical for the long-lived rainstorm. In all three zones, the

reduction in rainfall volume directly impacted the decrease in streamflow excess. This reduction was significant in Campinas and Monte, slightly affecting the peak time of the streamflow in both areas. For the short-live event, the volume scenario marginally reduced the streamflow in all three areas, and it had a negligible impact on the delay of the peak flow in Monte. These results suggest that there is a strong correlation between systematic errors inside satellite-based products and the streamflow properties, which are also found in similar studies using analytical approaches (e.g. Zoccatelli et al., 2011; Mei et al., 2017a)

We acknowledge certain limitations in the presented study. Firstly, the error in the hydrological model was not considered as an integral part of this research. Errors associated with modelling the rainfall-runoff process were reduced by using an intense optimisation process to calibrate the hydrological model over the catchment. Another limitation arises from the sensitivity rainfall intensity threshold and the Critical Mass Threshold for identifying and segmenting rainstorm objects. The characterisation parameters used in ST-CORA have potential implications for the rainstorm definition and segmentation. Laverde-Barajas et al. (2019) showed the sensitivity analysis of the Critical Mass threshold used for delineating the rainstorm object. Further studies could delineate the rainstorm event based on segmentation algorithms, such as multi-variable kernel segmentation or convolutions methods for considering the four dimensions of the rainfall object.

6.7. Conclusions

This chapter used the spatiotemporal object-based verification method, ST-CORA, to evaluate the hydrological impact of location and magnitude errors in CMORPH for rainstorm estimation. The results were obtained using a calibrated hydrological model of the Capivari river to reveal the primary sources of systematic errors for short-lived and long-lived rainstorm events. The most important conclusions of this study include the following:

1. Rainstorm events described as objects using ST-CORA provided a unique perspective to characterise the main spatiotemporal characteristics of the rainstorm in the catchment. Based on this information, differences between satellite and observed objects are calculated by separating the total error into three types: displacement, volume and pattern.

2. This study established location as the primary source of systematic error in the short-lived rainstorm, while volume was identified as the primary source in long-lived rainstorm events. These errors significantly impacted the shape phase and amplitude of the streamflow hydrograph. In the short-lived rainstorm, the subtraction of location error positively affected the error reduction in the upper

and middle zones. In the long-lived rainstorm, the subtraction of magnitude error in the rainstorm event lead to an error reduction over all three zones in the catchment.

3. This study presented a better understanding of the spatiotemporal characteristics of the systematic errors in satellites for rainstorm estimation and the impact on the hydrological response. However, errors relating to the hydrological model need to consider the potential "equifinality" of the optimal parameters used in the model.

4. Further research will incorporate the ST-CORA to improve the accuracy of satellite-based products in rainstorm detection. Location and magnitude error subtraction methods can be incorporated into bias correction approaches for reducing the systematic error of near-real-time satellite products in space and time. Additional satellite-based rainfall datasets can also be evaluated using the methodology, for example, the GPM-IMERG satellite-based product (Huffman et al., 2015) or the Climate Hazards Group Infrared Precipitation with Station datasets (CHIRPS) (Funk et al., 2015).

7

ST-CORAbico: A spatiotemporal object-based bias correction method for rainstorms detected by satellite

This chapter presents the Spatiotemporal Contiguous Object-based Rainfall Analysis for Bias Correction (ST-CORAbico). This method corrects errors in satellites due to displacement and volume in space and time. ST-CORAbico uses ST-CORA incorporated with a multivariate kernel density rainstorm segmentation to identify the main features of the rainstorm event. Displacement and volume are corrected by adjusting the spatiotemporal structure and the intensity distribution, respectively. ST-CORAbico was applied in the Lower Mekong basin in Thailand to correct short and long-lived rainstorm events detected by GPM-IMERG early version during the monsoon seasons from 2014 to 2017. The performance is compared against two widely used probabilistic methods. The results showed that ST-CORAbico considerably reduced RMSE and bias of GPM-IMERG due to displacement and magnitude outperforming both probabilistic approaches. This spatiotemporal bias correction method offers a new approach to enhance the accuracy of satellite-derived information for near real-time estimation of rainstorm events.

This chapter is partly based on the publication: Laverde-Barajas, M., Corzo, G. A., Poortinga, A., Chishtie, F., Meechaiya, C., Jayasinghe, S., Towashiraporn P., Markert A., Son L.H., Sothea K., Boonya-Aroonnet S., Chaowiwat W., R. U. Uijlenhoet, and Solomatine, D. P. (2020). ST-CORAbico: A spatiotemporal object-based bias correction method for rainstorm prediction detected by satellite. Remote Sensing, 12(21), 3538.

7.1. Introduction

Advances in near real-time rainstorm prediction using remote sensing have offered various opportunities for effective disaster management. Satellite-based measurements enable monitoring the dynamic of extreme rainfall events with high temporal and spatial resolution at a quasi-global scale. However, this information is subject to several sources of systematic and random errors that require correction (e.g Thiemig et al., 2012; Guo et al., 2015; Kimani et al., 2017). A wide range of bias correction methodologies has been developed to improve the performance of SPP, leveraging ground-based observations. Several examples include linear scaling, local intensity scaling, the power and distribution transformation methods, and Gamma Quantile mapping (e.g Tesfagiorgis et al., 2011; Vila et al., 2009; Habib et al., 2014). These methods all adjust SPP as a function of rainfall intensity values, ignoring important systematic errors, such as those that are caused by displacement and timing.

In the field of weather forecasting, displacement error in rainstorm prediction has been taken using spatial verification methods into account (Ebert and McBride, 2000; Davis et al., 2009b; Wernli et al., 2008). These "nontraditional" methods do not rely on point-to-point matches between the observed and estimated fields for avoiding double penalties (e.g. rainfall estimated but not observed and vice versa) that are commonly found in traditional approaches. Several studies have used spatial verification methods to analyse and correct the systematic error of SPP based on the characteristics of matched rainstorm objects, such as location, rotation, intensity, and shape (Skok et al., 2009; Li et al., 2015a, 2016). These methods have been useful for correcting the displacement errors when the grid resolution is high and the rainstorm event is small while preserving the higher spatial variability of SPP rainstorm (e.g. Demaria et al., 2011; Le Coz et al., 2019). However, these methodologies are constrained by the two-dimensional analysis of the rainstorm event.

The spatiotemporal analysis can provide a much deeper analysis on aspects of the entire life-cycle of the rainstorm event, including time, speed, evolution, among others. In the literature, error analysis using spatiotemporal approaches has been useful to evaluate the performance of several spatial rainfall products. For example, Clark et al. (2012, 2014) used the Object-Based Diagnostic Evaluation time-domain (MODE-TD; Bullock (2011)) to evaluate the convection-allowing forecast from the Weather Forecast Model over the United States. Recently, Laverde-Barajas et al. (2020b) used the Spatiotemporal Contiguous Rainfall Analysis (ST-CORA; Laverde-Barajas et al. (2019)) in the Southeast region of Brazil for analysing the error composition of the CMORPH SPP and evaluated the individual hydrological response of two systematic error sources: location and magnitude. This study demonstrated the importance of spatial and temporal rainstorm characteristics to analyse the main systematic error sources in SPP.

Spatiotemporal rainstorm analysis incorporated into bias correction methods is vital to reduce several sources of systematic error in SPP. In this study, we present a spatiotemporal object-based bias correction method to reduce several systematic errors in rainstorm events estimated by satellite. The method, called Spatiotemporal Contiguous Object-based Rainfall Analysis for Bias Correction (ST-CORAbico), uses the main rainstorm characteristics of satellite and observed events detected by the ST-CORA method to remove errors due to displacement in space and time and volume. This method is evaluated over the lower Mekong Basin in Thailand to correct several rainstorm event types in the Integrated Multi-satellitE Retrievals for GPM (GPM-IMERG) early version during the monsoon seasons from 2014 to 2017. The performance of ST-CORAbico is compared against two widely used probabilistic methods – Distribution Transformation and Gamma Quantile Mapping.

7.2. Study Area and Data Available

The study site selected for this research is the Lower Mekong basin area located in Thailand (Isan)(Figure 3.3). During the months from May to October, this area is highly impacted by heavy rainstorm events triggering flash floods and landslides (Delgado et al., 2012; Yang et al., 2019). The period evaluated corresponds to four years of monsoon periods (Jun to Oct) from 2014 to 2017. The Near real-time Integrated Multi-satellitE Retrievals for GPM (GPM-IMERG; Huffman et al. (2015)) early version. This version was selected because it is the lowest latency data product available (4 - 6 six hours), a crucial aspect of operational applications. In this study, we evaluated the early run of GPM-IMERG version V06B at a half-hourly temporal resolution (1 hourly aggregated). This GPM-IMERG version have recently incorporated several changes in maximum sensitivity rain rates of the algorithm from 110 mm/h to 200 mm/h impacting in the estimation of intense rainfall (Tan et al., 2019).

Ground measurements were obtained from a dense network of 138 rain gauge stations from the Thailand Integrated Water Resource Management System, operated by the Hydro-Informatics Institute (HII) in Thailand. Data were quality controlled by mirroring the density distribution functions of neighbouring stations to remove outlier values considered as noise. These observations were then further interpolated using the Ordinary Kriging interpolator from Software (2012) at 0.1 degrees for each hour to match the spatial and temporal resolution of the GPM-IMERG data. This method was selected due to the moderated topographic conditions of the study area and the density of the rain gauge measurements. Rain gauge data were interpolated using an exponential variogram with a sill of 1.14 (mm^2/h^2), a range of 4.4 km and a nugget of 0.66 (mm^2/h^2). It must be noted that interpolation methods for rainfall data are subject to uncertainty, e.g., Ly et al. (2012); Li and Heap (2014); however, in our case, a dense and optimal rain gauge distribution can reduce the level of uncertainty from

the interpolation method (Chen et al., 2008b).

7.3. Methodology

ST-CORAbico was developed to analyse the spatiotemporal characteristics of rainstorm events and bias correct the primary sources of systematic error in satellites. Figure 7.1 shows the methodology of ST-CORAbico. In this section, we describe the elements for rainstorm analysis and bias correction in ST-CORAbico.

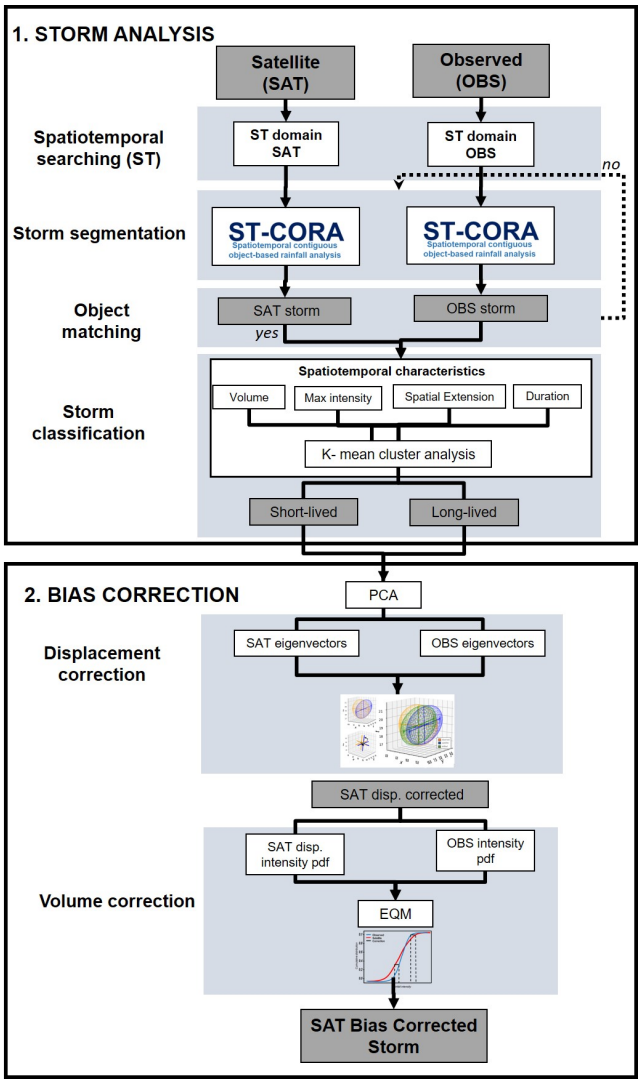


Figure 7.1: *Diagram of the Spatiotemporal Contiguous Object-based Rainfall Analysis for bias correction (ST-CORAbico) method. Grey boxes represent the input and output products while white boxes describe the methodological process for rainstorm analysis and bias correction components.*

7.3.1. Rainstorm Analysis

In the rainstorm analysis, ST-CORAbico uses ST-CORA to analyse the spatiotemporal characteristics of the rainstorm events observed and detected by satellites. This process

requires the definition of the spatial and temporal domain in order to reduce the computational time of ST-CORA. We applied a spatiotemporal searching algorithm to predetermine the region of analysis in ST-CORA. This algorithm uses the spatial searching algorithm concept that was proposed by Guttman (1984) to index areas with rainfall information in both datasets. The indexing is made in a two-dimensional space compressing the latitude and longitude dimensions using a maximum intensity value as a reference. Once the spatiotemporal domain is defined, we use ST-CORA in the observed and SPP dataset to identify rainstorms in the rainfall data. In this study, ST-CORA incorporates a multivariate kernel density function for rainstorm segmentation.

Rainstorm Segmentation using ST-CORA with Multivariate Kernel Density Segmentation

ST-CORA was applied to analyse the spatiotemporal characteristics of rainstorm events at the catchment scale (duration, spatial extent, magnitude, and centroid). This method enables the feature extraction of different rainstorm event types, classified based on hydrometeorological criteria. ST-CORA uses a multidimensional connected labelling component algorithm to associate connected voxels in space and time (a volume generalisation of pixels) into a disjoint object labelled with a unique classifier. Bethel et al. (2012) found that object segmentation while using image thresholding, such as the connected component labelling method, has limitations for edge detection in data with unknown topology. In the original ST-CORA, a size-filtering algorithm and morphological closing method are incorporated to remove small noisy objects and a false merging effect, respectively. However, this process is based on a binary object not taking into account the intensity value of voxels. To overcome this limitation, we have incorporated a Multivariate Kernel Density Estimation (KDE) approach to segment rainfall objects when considering their four dimensions. This method assumes a non-parametric probability density distribution technique for d-dimensional data. Notably, KDE has been widely used in many fields for image detection and object tracking, e.g.,(Chen et al., 2008a; Pereira et al., 2017; Wang et al., 2017; Zivkovic and Van Der Heijden, 2006; Berjón et al., 2018). Multivariate kernel density is estimated at point x from a random sample $X_1, X_2, ...X_n$ from a density function f.

$$\widehat{f_K}(x) = \frac{1}{n}\sum_{i=1}^{n} K_h(x - x_i) \tag{7.1}$$

where K corresponds to the kernel function and h is the bandwidth matrix. Choosing the bandwidth matrix can be restricted to a class of positive diagonal matrices (Hyndman et al., 2004). In the literature, there are several bandwidth methods selection methods for kernel density estimation (Hyndman et al., 2004; Zhang et al., 2006). For this approach, we use the standard reference rule-of-thumb proposed by Henderson and Parmeter (2012). This method estimates the bandwidth while assuming that the density distribution function follows a Gaussian distribution.

The process of edge detection using KDE is based on the Edge Detection by Density method that was developed by Pereira et al. (2017). This process evaluates the multivariate density distribution of the density of a four-dimensional (4D) rainfall object (Figure 7.2), and segments the object based on the density threshold, u. This threshold identifies the rainstorm edges that are lower than a probability percentage. This parameter is calculated by analysing the relationship between threshold delineation and the connected intensity value. We found that the 25th distribution percentile for u threshold showed promising results for rainstorm segmentation over the Lower Mekong Basin, especially for intense rainstorm events, characteristic of monsoon environments.

Spatiotemporal rainstorm event objects are defined based on two parameters: the rainfall intensity threshold IT for rain object identification and the Critical Mass Threshold (CMT), which is defined as the minimum volume of rainfall (km^3) necessary to be considered as rainstorms (Grams et al., 2006). The value of CMT is calculated locally based on the sensitivity between the spatial extent and the total object volume (Steiner et al., 1995; Demaria et al., 2011). In this analysis, we also incorporated the sensitivity of CMT to the maximum intensity of the rainstorm to evaluate the response of intense rainstorm events in the study area. Based on the sensitivity analysis of those parameters, we selected IT = 1mm/h to define rainfall objects (Ebert and Gallus, 2009) and CMT = 0.01 km^3 for rainstorm events with a maximum intensity greater than 10 mm/h. In the study area, these events correspond to rainfall objects bigger than 2000 km^2.

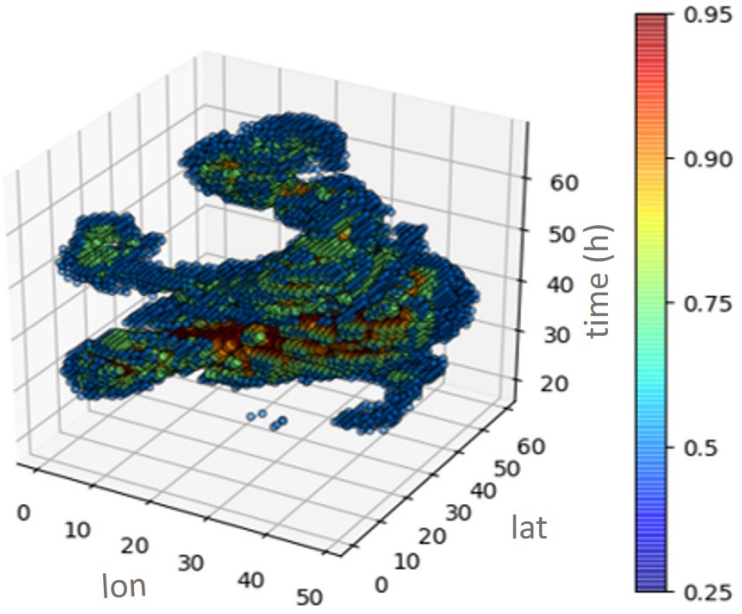


Figure 7.2: *Multivariable kernel density of a rainstorm object in space and time. Example for the rainstorm event 2014-07.*

Matching process

As a result of ST-CORA, multiple rainstorm events are identified in both observed and satellite data sets. However, it is necessary to determine the observed and estimated rainstorm matches. We used the Intersection-over-Union measure (IoU) to evaluate the level of similarity between predicted and observed data. IoU is defined as the ratio between the size of the intersection and the union of both objects represented by the following equation:

$$IoU = \frac{TP}{FP + TP + FN} \tag{7.2}$$

where TP represents true positives, and FP and FN are false positives and negatives, respectively. The selection of the intersection rate value determines the level of matching between objects. If the values are too low; multiple objects will be indexed with the same object. On the other hand, high values indicate that the object does not have any match. We found the IoU value to be 30 percent in the selected study area, which is a good balance for matching the observed and satellite rainstorm events.

rainstorm Classification

Once all of the rainstorm events are identified, ST-CORA classifies rainstorm events into two types: small convective systems with a short duration (short-lived) and long duration systems extended over large areas (long-lived) (Molini et al., 2009, 2011). We used an unsupervised K-means cluster analysis method to classify short- and long-lived rainstorm events based on the four main rainstorm characteristics (duration, spatial extent, maximum intensity, and total volume). This method divides n observations into k clusters in which each observation is a member of the cluster that minimises the objective function J, as follows:

$$J = \sum_{j=1}^{k} \sum_{i=1}^{n} \| X_{(x,y,z,c)}{}_i^j - C_j \|^2 \tag{7.3}$$

where X is the rainstorm with dimensions x, y, z, c corresponding to the rainstorm characteristics duration, spatial extent, maximum intensity and total volume, respectively. C is the centroid of the cluster k, and the absolute number represents the minimum Euclidean distance to C.

7.3.2. Bias Correction

Bias correction is the second component of the ST-CORAbico method. This component is based on the systematic error source extraction for SPP that was proposed by Laverde-Barajas et al. (2020b). Based on the error decomposition for rainstorm estimation defined by Ebert and McBride (2000), satellite error is composed of systematic

and aleatory errors due to displacement, volume, and pattern, as:

$$E_{total} = E_{displacement} + E_{volume} + E_{pattern} \tag{7.4}$$

where displacement and volume represent the systematic errors and pattern is the aleatory error calculated as follows:

$$E_{displacement} = E_{total} - E_{shifted} \tag{7.5}$$

$$E_{volume} = E_{total} - E_{magnitude} \tag{7.6}$$

$$E_{pattern} = E_{shifted} - E_{volume} \tag{7.7}$$

In Equation (7.5) and (7.6), location is the primary source of error due to displacement, while the magnitude is the corresponding source of error for volume. Using the error subtraction from Laverde-Barajas et al. (2020b), ST-CORAbico corrects displacement and volume error using the following process:

Displacement Correction:

Displacement correction corresponds to removing the shifting effect of the estimated rainstorm. In this step, the Principal Component Analysis method (PCA, Johnson and Hebert (1999)) is used in order to obtain the weighted centroid and orthogonal variables (eigenvectors, eigenspace) of the SPP rainstorm and reference data (Figure 7.3a). Once the geometric properties of the objects are obtained, the weighted centroid is matched, and the object is rotated accordingly to fit the eigenvectors of the reference rainstorm data.

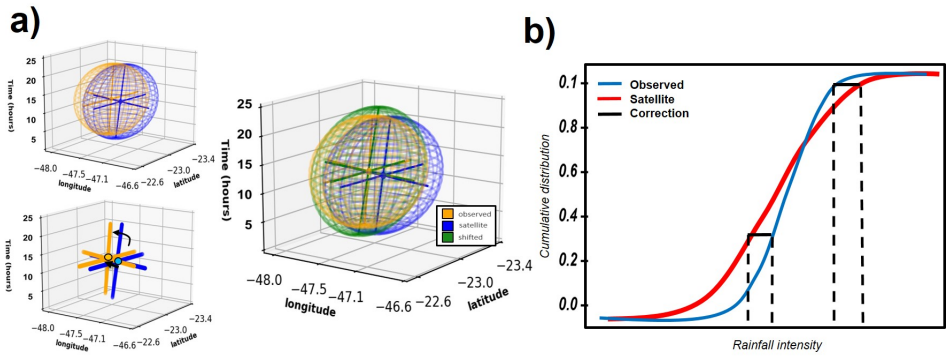


Figure 7.3: *ST-CORAbico systematic error subtraction. (**a**) In error correction, rainstorm centroid and eigenvectors derived from principal component of GPM-IMERG are fitted to the observed event; and, (**b**) magnitude subtraction using both satellite and observed empirical distribution functions, with respect to intensity.*

Volume Correction:

Volume correction corresponds to the subtraction of the magnitude source of error of the SPP. Using the statistical Empirical Quantile Method (EQM) Themeßl et al.

(2012), magnitude error is subtracted by adjusting all moments of the empirical cumulative distribution functions ($ecdfs$) of the SPP in terms of intensity, with respect to the reference data (Figure 7.3b). EQM builds the $ecdf$ for the observed ($ecdf_{obs}$) and the satellite ($ecdf_{sat}$) while using the intensity rainstorm distribution I_s, as:

$$EQM = ecdf_{obs}^{-1}(ecdf_{sat}(I_s)) \tag{7.8}$$

7.3.3. Evaluation of ST-CORAbico

The evaluation was done by comparing the bias-corrected results with two widely used probabilistic bias correction methods the Distribution Transformation (DT) method and the Gamma Quantile Mapping (GQM). The DT method was initially developed for the statistical downscaling of climate model data Brown et al. (2004). The method corrects the mean and difference in variation of the SPP by matching the satellite and the observed distribution based on Equation (7.9):

$$DT = (SAT(t) - \mu_{sat})DT_\tau + \tau_{sat} * DT_\mu \tag{7.9}$$

where μ and τ are the mean and standard deviation of the observed and satellite, respectively. DT_μ and DT_τ are the mean and standard deviation ratio between the observed and satellite data at time t.

The Gamma Quantile Mapping method uses the same methodology as the Empirical Quantile mapping method (eq. 7.8), based on the assumption that both observed OBS and satellite SAT intensity follows a gamma distribution (Piani et al., 2010). DT and GQM are implemented for each time step in order to correct the rainstorm event. The bias correction performance is evaluated based on three widely used error metrics: the Root Mean Square Error (Equation (7.10)) for evaluating the magnitude error, the bias level (Equation (7.11)) to evaluate the systematic bias, and the correlation coefficient (Equation (7.12)) in order to analyse the linear correlation between the observed and the bias-corrected rainstorm event.

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(OBS_i - SAT_i)^2} \tag{7.10}$$

$$Bias = \frac{\sum_{i=1}^{N}(SAT_i - OBS_i)}{\sum_{i=1}^{N}(OBS_i)} \tag{7.11}$$

$$r = \frac{\sum_{i=1}^{N}(SAT_i - \overline{SAT})(OBS_i - \overline{OBS})}{\sqrt{\sum_{i=1}^{N}(SAT_i - \overline{SAT})^2}\sqrt{\sum_{i=1}^{N}(OBS_i - \overline{OBS})^2}} \tag{7.12}$$

where *OBS* represents the rainfall values of the reference rain gauge data, and *SAT* are the satellite and the bias-corrected rainstorm obtained with each method.

7.4. Results

7.4.1. Rainstorm analysis

We identified 120 rainstorm events observed and estimated by GPM-IMERG at an hourly scale for 2014-2017 monsoon seasons. Figure 7.4 shows the scatter plot of the main rainstorm characteristics (total volume, duration, spatial extent, and maximum intensity) and classification between short- and long-lived rainstorm events using the k-means cluster analysis. For all events, 68 rainstorms (56%) were classified as short-lived rainstorms while 52 (44%) of rainstorms were classified as long-lived events. Short-lived events had a duration that ranged between three and 17 h, with a maximum spatial extent of 42 thousand km^2. Long-lived events ranged between 18 and 31 h and covered between 54 and 110 thousand km^2. In terms of total volume and maximum intensity, short-lived events have a total volume of up to 0.15 km^3 with low and intense rainstorms ranging from 3 to 82 mm/h. On the other hand, long-lived rainstorms comprise medium and high-intensity events with a total volume ranging from 0.27 to 0.65 km^3. Table 7.1 describes the observed rainstorm characteristics for short and long-lived event types.

Table 7.1: *Rainstorm characteristics for short- and long-lived event types.*

rainstorm Type	Statistics	Duration (h)	spatial Extent (km^2)	Maximum Intensity (mm/h)	Total Volume (km^3)
Short-lived rainstorm	mean	9	15,097	33.0	0.04
	min	3	1900	3.6	0.01
	max	17	42,300	82.0	0.15
long-lived rainstorm	mean	18	54,400	71.4	0.27
	min	10	24,300	31.6	0.07
	max	31	110,600	100.0	0.64

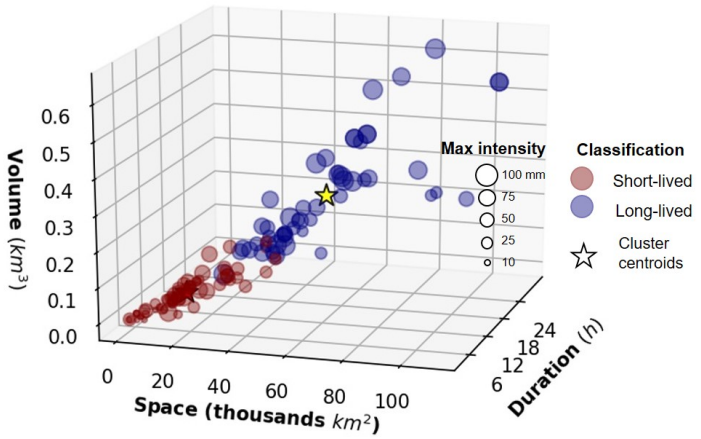


Figure 7.4: *Short- and long-lived cluster analysis classification for observed events during monsoon season 2014-2017.*

7.4.2. Bias Correction

We selected a short-lived and a long-lived rainstorm to describe the workflow for displacement and volume correction made by ST-CORAbico. Figures 7.5 and 7.6 present the bias correction steps for each rainstorm event type. Panel (a) shows the spatial distribution of the observed and satellite events and the bias-corrected satellite rainstorm events obtained from the correction of location and magnitude errors in ST-CORAbico. Panel (b) describes the displacement and volume corrections. Panel (c) presents the four-dimensional (4D) spatiotemporal evolution of the observed and satellite as well as the bias-corrected rainstorm (time in the z-axis). Panel (d) shows the bias and RMSE. Statistics as well the scatter and correlation between the observed rainstorm and original and bias-corrected satellite events.

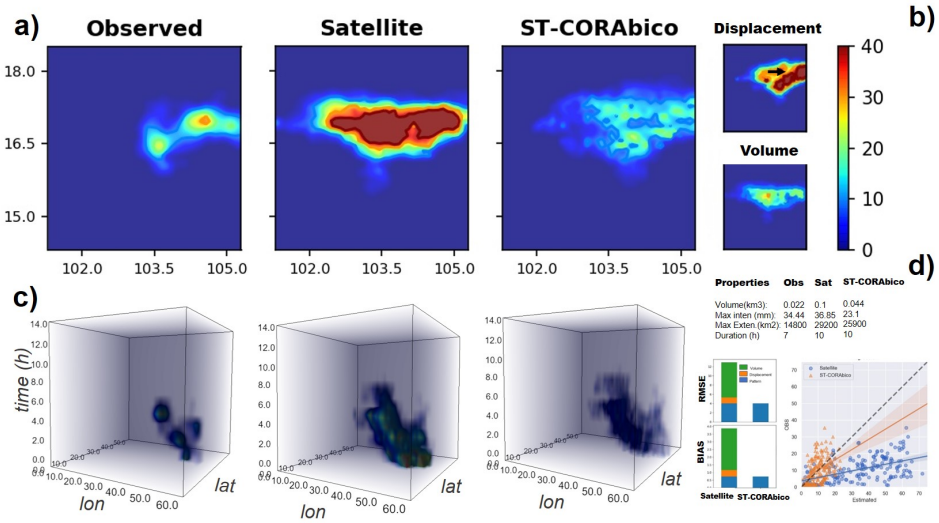


Figure 7.5: *Performance of ST-CORAbico for a short-lived rainstorm event (2014-08-27). (**a**) total events for observed, satellite and ST-CORAbico; (**b**) volume, displacement correction maps; (**c**) four-dimensional (4D) spatiotemporal evolution (lat, lon, time, intensity); and, (**d**) bias, RMSE statistics, and scatter and correlation between observed and estimated rainfall values.*

Both examples (Figures 7.5 and 7.6) show the importance of bias correction. In both the short and long-lived event scenarios, GPM-IMERG had a longer duration with a larger footprint. However, the long-lived event presented a better spatial agreement than the short-lived event. In terms of magnitude, GPM-IMERG considerably overestimated the total volume and rainfall intensity of the rainstorm. Overall, the performance of GPM-IMERG shows a positive bias and high RMSE, mostly being caused by an excess of rainfall. The correlation coefficients for short- and long-lived event scenarios were 0.7 and 0.5, respectively.

The corrections in displacement and volume made by ST-CORAbico displayed notable changes in the satellite rainstorm structure. In both scenarios, RMSE and bias mainly were reduced by correction due to volume, contributing 40 to 60% of the

RMSE reduction and around 70% of the total bias reduction for both events. Displacement correction had an essential impact on the reorientation of the satellite rainstorm. The individual correction contributed to 5% of the RMSE correction and 10% reduction of the total bias for the short-lived event. In the case of the long-lived scenario, displacement correction contributed almost 15% of the RMSE reduction and 20% of the total bias reduction. In terms of the correlation coefficient, displacement and volume correction made by ST-CORAbico showed a marginal impact on the spatial correlation for the short-lived events. For the long-lived scenario, this did not impact the spatial correlation.

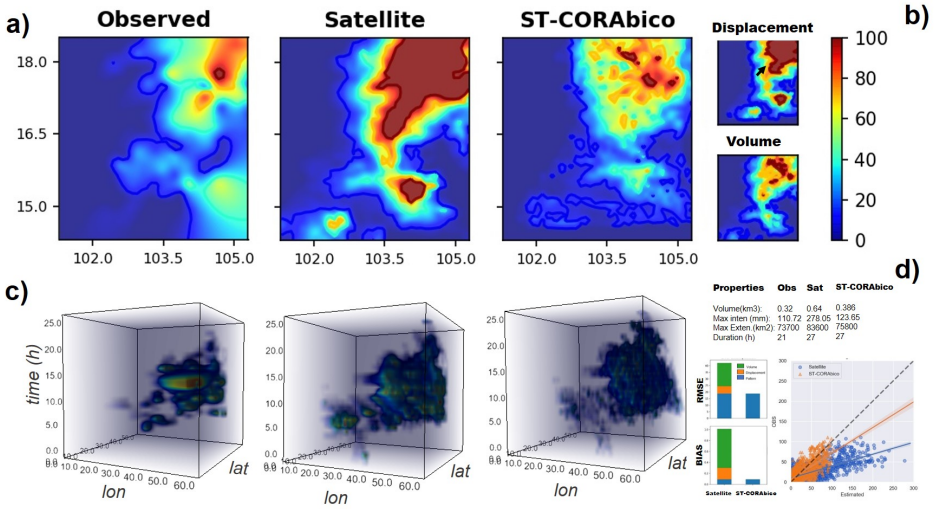


Figure 7.6: *Performance of ST-CORAbico for a long-lived rainstorm event (2014-07-21). (**a**) total events for observed, satellite and ST-CORAbico; (**b**) volume, displacement correction maps; (**c**) 4D spatiotemporal evolution (lat, long, time, intensity); and, (**d**) bias, RMSE statistics and scatter and correlation between observed and estimated rainfall values.*

Figure 7.7 presents the performance of ST-CORAbico for short- and long-lived rainstorm events. This figure describes the density distribution of RMSE (a-b), bias (c-d), and correlation coefficient (e-f) of the short- and long-lived rainstorms estimated by GPM-IMERG, ST-CORAbico, and the individual corrections due to displacement and volume. It was found that ST-CORAbico has a smaller error distribution in RMSE and bias for short- and long-lived rainstorm events compared with the original GPM-IMERG. The correction mainly causes this error reduction due to volume. Displacement correction was an important factor in reducing the bias, especially for long-lived rainstorm events. The results from the correlation coefficient showed that ST-CORAbico had a marginal effect on the spatial correlation of the rainstorm event. Overall, it was found that ST-CORAbico considerably reduced the systematic error of GPM-IMERG.

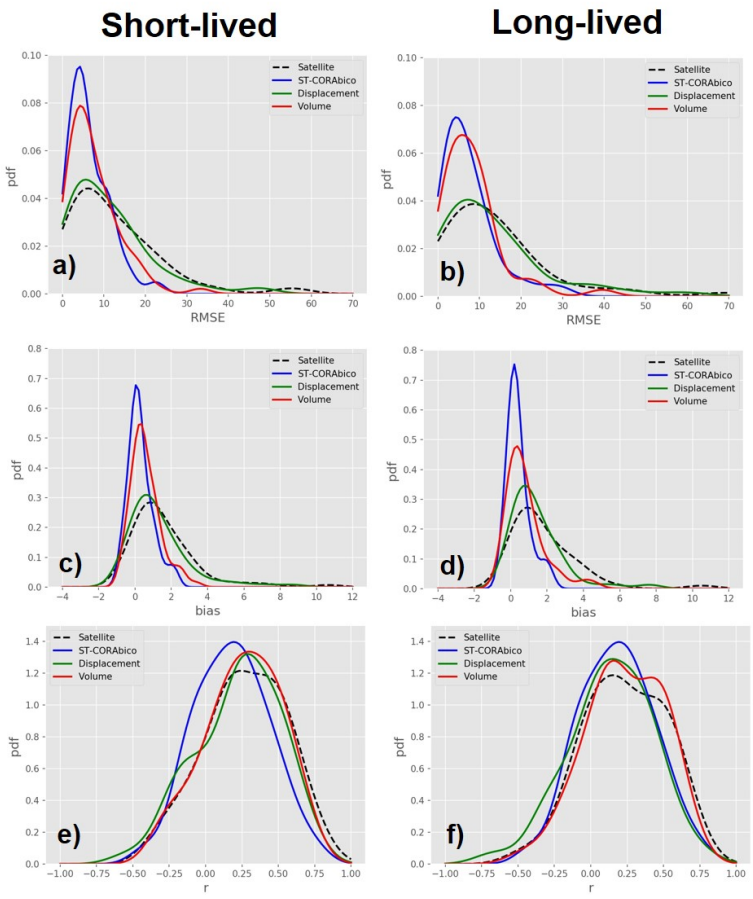


Figure 7.7: *Satellite and bias corrected error distribution for short and long-lived events during monsoon seasons 2014–2017: (**a**,**b**) RMSE; (**c**,**d**) bias; and, (**e**,**f**) correlation coefficient.*

7.4.3. Model Comparison

ST-CORAbico was compared with the Distribution Transformation method (DT) and the Gamma Quantile Mapping (GQM) method. Using the short- and long-lived rainstorm scenarios that are presented above, Figure 7.8 presents the spatial differences and linear correlation between the total observed rainstorms and the bias-corrected events obtained by ST-CORAbico, DT, and GQM. The results for both rainstorm event scenarios showed that ST-CORAbico had the lowest spatial difference among the evaluated methods. For the short-lived rainstorm scenario, ST-CORAbico displayed the highest correlation coefficient (r: 0.41) and the lowest RMSE and bias (RMSE: 4.05 mm; bias: 0.74) when compared with DT (r: 0.40; RMSE: 5.4 mm; bias: 1.17); and, GQM (r: 0.39 RMSE: 6.09mm and bias: 1.5). In the case of the long-lived rainstorm, ST-CORAbico and DT showed a notable error reduction compared to the GQM method that showed the biggest differences. For this rainstorm scenario, ST-CORAbico had the best performance (r: 0.71 RMSE: 18.02 mm; bias: 0.09), followed by DT (r: 0.68, RMSE: 23.0 mm; bias: 0.32), and finally GQM (r: 0.62, RMSE:

43.77 mm; bias: 0.97).

Figure 7.9 presents the comparison between ST-CORAbico, DT, and GQM for short- and long-lived rainstorm events. The boxplots show the distributions of the RMSE (Figures 7.9a,b), the bias (Figures 7.9c,d), and the correlation coefficient (Figures 7.9d,f) between the 25% and 75% percentiles for the original GPM-IMERG and the different bias correction methods. The dots represent the individual error for each rainstorm event. In comparison with the two probabilistic methods, we found that ST-CORAbico consistently had the lowest RMSE and the lowest bias for both short- and long-lived rainstorm events. ST-CORAbico and DT had a lower impact on the correlation coefficient, especially for short-lived events.

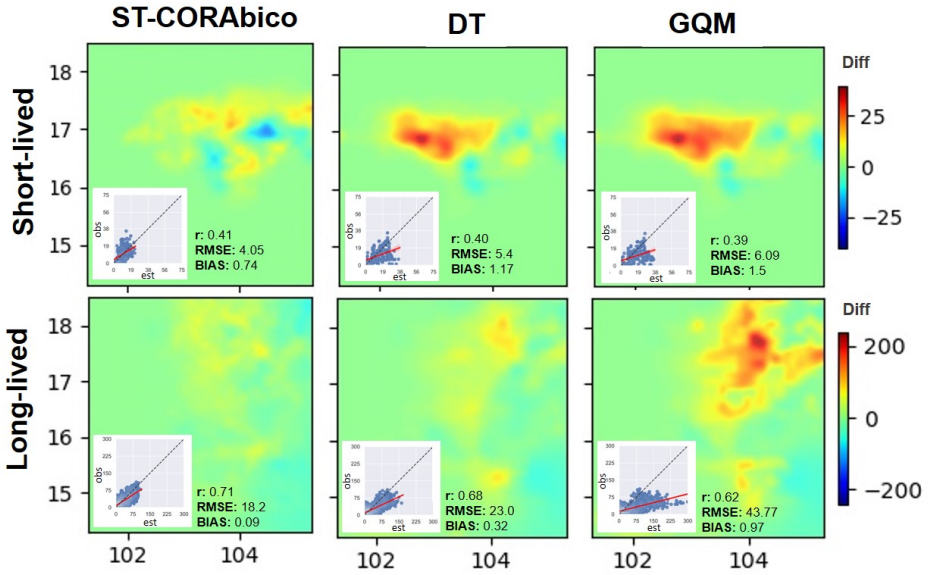


Figure 7.8: *Comparison between ST-CORAbico vs Distribution Transformation (DT) and Gamma Quantile Mapping (GQM).*

7.5. Discussion

ST-CORAbico is a spatiotemporal object-based bias correction method designed to reduce the displacement and volume systematic errors of rainstorm events detected by SPP. In comparison to spatial object-based bias correction methods, e.g., Demaria et al. (2011); Le Coz et al. (2019), the inclusion of the temporal component of the rainstorm event reduced additional error effects due to timing and orientation, improving the efficiency of the bias correction.

This research incorporated a multivariate kernel distribution algorithm into ST-CORA to segment the rainstorm event using the four dimensions of the rainstorm event. In comparison to binary segmentation in the previous version, ST-CORA with KDE segmentation was able to delineate intense rainstorm events by removing unreal rainstorm configurations due to false merging and false separation of rainstorms due

to the multidimensional connected labelling component algorithm. Based on the KDE threshold delineation and the connected intensity analysis, we found those rainstorm events segmented by the 25th percentile of the distribution showed a good result for segmenting intense rainstorms with a solid connection—however, further improvement is required.

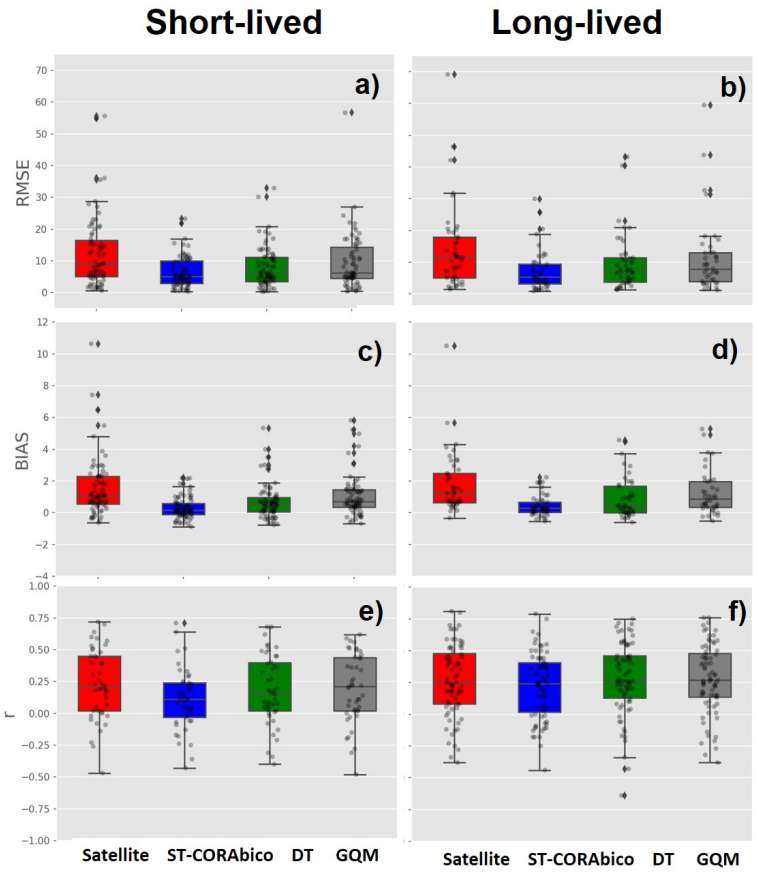


Figure 7.9: *Comparison between the satellite GPM-IMERG (red), ST-CORAbico (blue), Distribution Transformation (green) and Gamma Quantile Mapping (grey) error dispersion during monsoon seasons 2014–2017. (**a**,**b**) RMSE; (**c**,**d**) bias; and, (**e**,**f**) correlation coefficient.*

The implementation of ST-CORAbico described the individual error correction due to displacement and volume. Results in the Lower Mekong basin indicated that volume errors were the primary error correction, primarily resulting from the high overestimation of GPM-IMERG. These results agreed with multiple findings regarding hourly GPM-IMERG in monsoonal areas (Oliveira et al., 2016; Tang et al., 2016). Overall, volume and displacement errors effectively contributed to reducing bias and RMSE, demonstrating the importance of reducing both of these systematic errors in satellite correction.

We acknowledge certain limitations of the study. Firstly, the uncertainty arising from the spatial interpolation method used for rain gauge values was not fully addressed in this research. Volume and especially displacement corrections in ST-CORAbico can

be affected by the type of interpolation methods used to represent the spatiotemporal distribution of the observed rainstorm. A dense rain gauge network can reduce the level of uncertainty; however, it is essential to evaluate the impact of the type of interpolation method on the performance of ST-CORAbico, as mentioned above. Another limitation arises from the sensitivity of IoU percentage to match observed and estimated rainstorm events. Higher levels do not always correspond to similar events, which affects the bias correction. This process required an in-depth sensitivity analysis of IoU in order to reduce the automatic rainstorm matching. Additional analysis is required to identify why there is a strong correlation between observed and predicted rainstorms in a spatiotemporal environment. In this study, we validated the performance of ST-CORAbico by comparing its performance against two widely used probabilistic methods. However, error metrics were calculated using the observed values, as there is no independent validation dataset available. Further implementations should consider an independent dataset to validate the error correction of the ST-CORAbico method.

This study was conducted in collaboration with the SERVIR-Mekong project and the Mekong River Commissions (MRC). SERVIR-Mekong is harnessing space and geospatial technologies to help decision-makers and critical civil society groups integrate geospatial information into their decision-making, planning, and communication. The application of this methodology can be used for various scientific purposes, including flood risk and water management. More specifically, the methodology enhances the input rainfall data, which are a crucial component of flood and drought early warning systems, landslide monitoring, as well as other water-related decision support systems. Future work will include integrating machine learning technologies for near real-time bias correction of rainfall data when field data are scarce. In this regard, machine learning models will be trained and optimised using legacy field data and deployed in a near-real-time basis.

7.6. Conclusions

We proposed a new spatiotemporal bias correction method for rainstorm prediction detected by satellites. The method, called Spatiotemporal Contiguous Object-based Rainfall Analysis for bias correction (ST-CORAbico), analyses the main spatiotemporal characteristics of the observed and estimated rainstorm events to correct systematic error sources due to displacement and volume. This methodology has two main elements: rainstorm analysis for the segmentation and classification of rainstorm events; and bias correction for correcting errors due to displacement and volume. In the rainstorm analysis, we applied the ST-CORA method with a multivariate kernel segmentation in order to identify the spatiotemporal structure of the rainstorm event. This method was applied over the Lower Mekong basin in Thailand to correct the

GPM-IMERG Early version during the monsoon seasons from 2014 to 2017. The performance of ST-CORAbico was evaluated against the Distribution Transformation and the Gamma Quantile Mapping methods based on reducing RMSE, bias, and correlation coefficient. The results were divided by classifying the rainstorm events into short- and long-lived rainstorm events while using the k-means cluster analysis method.

We classified 68 rainstorms (56%) as short-lived rainstorms and 52 (44%) as long-lived events. The results of both rainstorm event types showed that ST-CORAbico reduced the RMSE and bias of GPM-IMERG. Volume correction was the primary error source due to the overestimation present in GPM-IMERG. Location error was most important in the reduction of the bias. ST-CORAbico displayed a marginal impact on the spatial structure of the satellite-derived rainfall, showing the original structure of the rainfall data.

The comparison of ST-CORAbico with the Distribution Transformation and the Gamma Quantile Mapping methods showed that ST-CORAbico had the lowest RMSE and the lowest bias in both short and long-lived events. In terms of the correlation coefficient, ST-CORAbico and DT had a lower impact on the correlation coefficient, especially for short-lived events.

ST-CORAbico improves the accuracy of satellite-derived near real-time information on rainstorm events. It can be used in various flood monitoring and water management applications.

8

Hybrid Machine learning/object-based approach to correct satellite data for operational purposes

Operational physically-based bias correction methods present several challenges when ground measurements are not yet available. This chapter presents an approach that combines the spatiotemporal object-based bias correction method with a machine learning classifier to reduce systematic errors of near real-time satellite-based precipitation products without corresponding rain gauge data. This hybrid approach corrects errors due to displacement and volume of satellite precipitation products using information from similar historical events identified by a K-nearest neighbour ML classifier. The proposed methodology is applied over the Lower Mekong Basin in Thailand to correct rainstorm events detected by GPM-IMERG early version during the monsoon period of 2018 based on historical monsoonal rainstorm information from 2014 to 2017. The performance for different rainstorm events types is compared against two probabilistic and ML-based bias correction methods. The results show that the operational ST-CORAbico improved the RMSE and bias without corresponding rain gauge data. Compared with both operational bias correction methods, this new approach provided a better description of most of the spatiotemporal features of the corrected rainstorms. The performance of this approach prompts more attention of researchers to using machine learning for operational error correction.

This chapter is partly based on the publication: Laverde-Barajas, M., Corzo, G. A., Poortinga, A., Chishtie, F., Meechaiya, C., Jayasinghe, S., ... Solomatine, D. P. .Hybrid Machine learning/object-based approach to correct satellite data for operational purposes. In preparation.

8.1. Introduction

Physically-based bias correction approaches have a significant advantage over traditional approaches for correcting errors in Satellite-based Precipitation Products (SPP) due to intensity and reducing errors due to location, rotation, and shape. However, the dependence on the corresponding ground data, which is not always available, make them unsuitable in operational applications. Probabilistic and machine learning approaches have overcome this limitation using historical ground-based information to correct SPPs. Distribution mapping methods such as the Gamma Quantile Mapping use the empirical cumulative distribution function (CDF) from historical ground-based intensity values as a reference to correct SPPs operationally. These methods are not only suitable for operational bias SPP bias correction but also the correction of forecast data from General Circulation Models and (e.g. Hashino et al., 2007; Piani et al., 2010; Zhao et al., 2017). On the other hand, machine learning methods such as K-Nearest Neighbour, Random Forest and Neural networks use historical data for their training from different sources (ground-based rainfall, soil moisture and topographic) to correct the rainfall data (e.g. Gagne et al., 2014; Yang and Luo, 2014; Kumar et al., 2019; Zarei et al., 2021; Tao et al., 2016; Le et al., 2020).

Both methods share a standard limitation to understand the physical dynamics of a rainstorm event. While probabilistic bias correction approaches are prone to mislead the physical trends of the rainfall data (Cannon et al., 2015b), ML methods do not consider the physical laws of the rainfall to train the model (Corzo Perez, 2009; Wang et al., 2020). As a result, combining physical and ML learning bias correction methods can solve both the lack of the corresponding ground data and the limited capacity to understand the details of the phenomenon's physics, respectively.

This study uses the recently developed spatiotemporal object-based bias correction method, ST-CORAbico, combined with an ML k-nearest neighbours classifier to bias correct SPP rainstorm events in operational applications without the corresponding rain gauge data. This hybrid approach corrects errors due to displacement and volume of the satellite using the information from similar historical rainstorm events identified by the kNN classifier. The kNN classifier identifies the closest historical event based on six spatiotemporal features of the rainstorm detected by the satellite: centroid (lat, long), maximum intensity and spatial extension, duration and total volume. Once the historical event is identified, we use ST-CORAbico to correct the satellite's systematic errors based on the observed historical data. This method is applied over the Lower Mekong Basin in Thailand to correct the integrated Multi-satellite Retrieval for GPM (GPM-IMERG; Huffman et al. (2015)) early version during the monsoon season of 2018 based on the historical rainstorms database from 2014 to 2017. The performance to correct different rainstorm event types is compared against two operational bias correction methods - the probabilistic Gamma Quantile mapping and Machine learning-

based K-nearest Neighbour Machine.

8.2. Study area and Available Data

The study area corresponds to the region of the Lower Mekong Basin (LMB) that covers Thailand. In chapter 3, the main physical, geographical and atmospheric characteristics of this area are described in detail. We evaluated the bias correction of rainstorm events detected by the GPM-IMERG early version during the monsoon event of 2018 based on a historical rainstorms database obtained by Laverde-Barajas et al. (2020a). This data set is comprised of an analysis of 120 observed and satellite-based events that impacted the study area during the monsoon seasons from 2014 to 2017 using the Spatiotemporal Object-based Contiguous Rainfall Analysis method (ST-CORA; Laverde-Barajas et al. (2019)). Five different spatiotemporal features were identified in each event (duration, spatial extent, magnitude, and centroid). In chapter 7, we have described this database and the observed data used for this analysis.

8.3. Methodology

The methodology is composed of a hybrid rainstorm analysis and spatiotemporal bias correction. Figure 8.1 illustrates the schematics of the proposed method. In the following subsection, we will describe the main elements for each methodological component.

8.3.1. Hybrid rainstorm analysis

This component combines the object-based rainstorm analysis method used in ST-CORAbico called ST-CORA with a machine learning-based k-nearest neighbours classifier (kNN). ST-CORA (Laverde-Barajas et al., 2019) describes the spatiotemporal features of the satellite-based rainstorm event while the kNN classifier uses this information to identify the historical rainstorm event with the closest features.

Object-based analysis using ST-CORA with KDE

We used ST-CORA with KDE segmentation to analyse the spatiotemporal features of rainstorm events detected by the GPM-IMERG early version. This method is an updated version of the original ST-CORA presented in chapter 5 to segment rainstorm events in 4-Dimensions (lat, long, time, intensity) (Laverde-Barajas et al., 2020a). The methodology combines the multidimensional connected-component labelling algorithm to identify rainfall objects in space and time with a multivariate Kernel Density Estimation algorithm to segment rainstorm events based on the probability distribution of the rainfall object. A detailed description of this method is presented in chapter 7.

The implementation of ST-CORA starts by defining the spatiotemporal temporal domain for searching rainfall objects. The definition of this domain is used to reduce the computational time of ST-CORA. However, in operational bases, the temporal

domain is associated with the temporal searching windows near real-time. In this study, we defined the temporal domains during the monsoon season of 2018 using the spatiotemporal searching algorithm implemented in ST-CORAbico. This algorithm compresses the latitude and longitude dimensions of the SPP data to index areas with rainfall information concerning time.

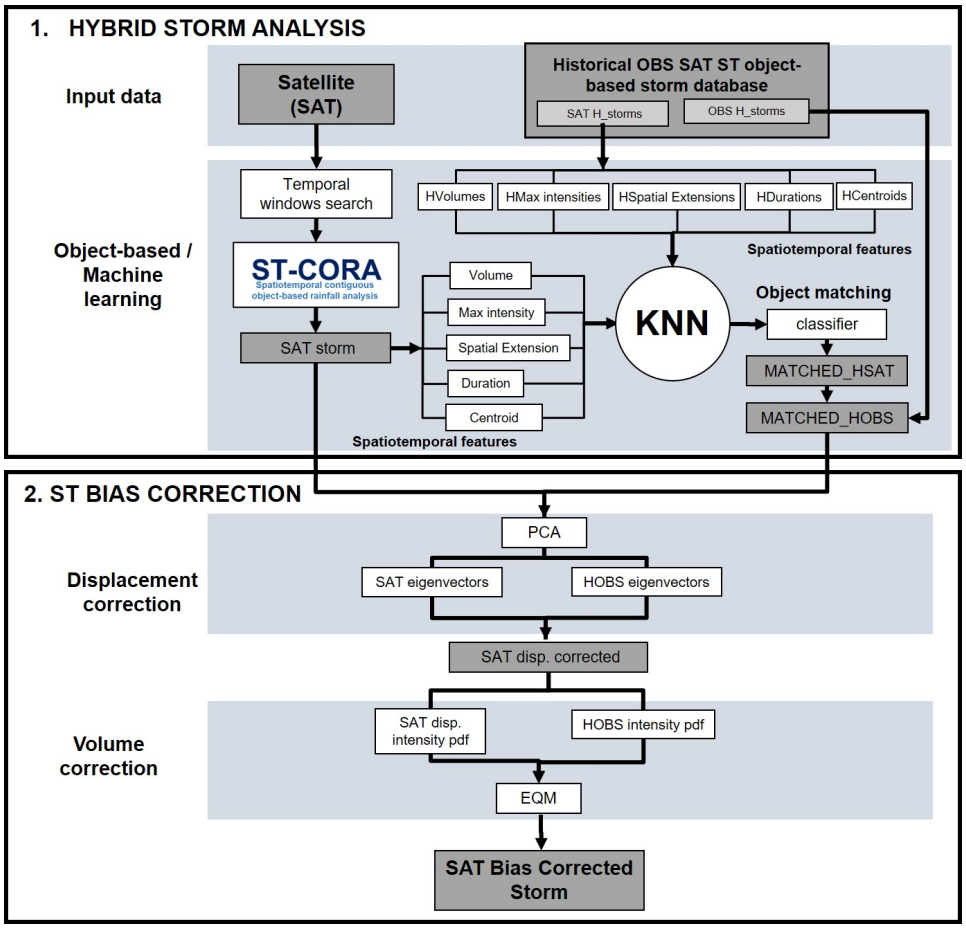


Figure 8.1: *Scheme of the operational ST-CORAbico*

ST-CORA with KDE is composed of two main processes - spatiotemporal rainfall object extraction and rainstorm identification. In the first process, spatiotemporal rainfall objects are defined based on the rainfall intensity threshold (IT) and percentage threshold for KDE (u). IT identifies the threshold of the voxel to be considered as "effective rainfall". In contrast, u is the threshold percentage of the KDE to segment the rainfall object in 4D. On the other hand, rainstorm events are identified based on the Critical Mass Threshold (CMT). This parameter corresponds to the minimum volume required by the rainfall object to be considered an extreme event. Laverde-Barajas et al. (2020a) incorporated maximum intensity of the rainstorm (MI) as a threshold to evaluate the response of intense rainstorm events. Based on the rainstorm analysis made by Laverde-Barajas et al. (2020a) in the Lower Mekong basin in Thailand during the monsoon seasons from 2014 to 2017. The parameters of ST-CORA with KDE used

in this research are IT= 1mm/h, u= 25th, CMT= 0.01 km^3 and MI= 10 mm/s.

Once all events are identified, the last step of ST-CORA corresponds to the classification of several rainstorm event types following hydro-meteorological and dynamic criteria (Molini et al., 2011). According to this principle, rainstorm events can be classified into short and long-lived events based on their main spatiotemporal characteristics. Laverde-Barajas et al. (2020a) implemented a K-means cluster analysis method to classify short and long-lived rainstorm events based on the four rainstorm features (duration, spatial extent, maximum intensity, and total volume). Based on this analysis over the Thailand area of the LMB, short-lived events are considered convective systems with duration ranging between three and 17 h, with a maximum spatial extent of 42 thousand km^2. In comparison, long-lived events corresponded to extensive extreme precipitation systems with a duration ranging between 18 and 31 h and covered between 54 and 110 thousand km^2. This study used those ranges to identify short and long-lived rainstorm events in the study area.

Machine learning-based analysis

In the absence of a corresponding observed data, the machine learning-based k-Nearest Neighbour classifier (kNN) analyses the spatiotemporal features of historical rainstorms detected by satellites to estimate a corresponding observed event for bias correction. The kNN classifier is one of the most widely used ML methods for classification due to its low complexity and fast implementation. One advantage of kNN is that it does not need any training data points for model generation, making it faster in the learning phase.

The kNN algorithm is designed to find the object with the k nearest neighbour pattern based on the distance between the features of query object QO and all training objects TO (figure 8.2). An object is classified based on the distance between points located in a multidimensional space. $k = 1$ is the number of nearest neighbours. If $k = 1$, the object is classified as the closest train object. Several distance metrics can be used to calculate the distance between points (e.g. Hamming, Euclidean, and Manhattan distance measures). We used the the Euclidean distance metrics in 'n'-Dimensional space, which is defined as:

$$D_{ij}^2 = \sum_{v=1}^{n} (X_{vi} - X_{vj})^2 \tag{8.1}$$

Where the 'n'-dimensions represent the number of object features, and X_{vi} and X_{vj} are also the query and the train object values, respectively. In a high dimensional representation of the kNN, one of the common challenges is ensuring the diversity of the feature selection (e.g. for avoiding the redundancy or overlap among features) (Li, 2009). The selection of the number of rainstorms features to train the kNN classifier is based on six spatiotemporal features to represent the position, geometry and magnitude of rainstorms detected by GPM-IMERG - centroid longitude and latitude for

the position, duration ad maximum extension for geometry and total volume and maximum intensity for magnitude. We standardised the values of the feature to remove the scale effect caused by the use of multiple features with different units using the following equation:

$$X_{standardised} = \frac{X - \mu}{\sigma} \tag{8.2}$$

where X is the rainstorm feature and μ, and σ are the mean and the standard deviation, respectively. The kNN classifies each SPP rainstorm assigning them a unique label identifier number. Once trained, kNN identifies the historical satellite-based event with the closest spatiotemporal features to the query rainstorm extracted by ST-CORA. The corresponding observed event is used to correct the systematic errors due to the displacement and volume of the SPP rainstorm event.

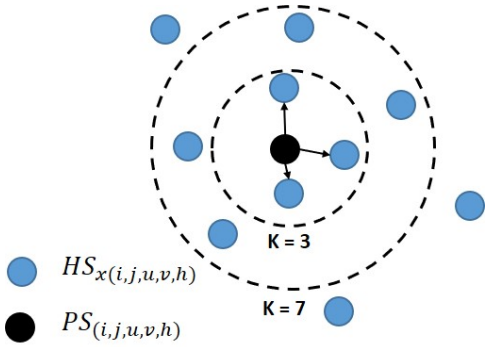


Figure 8.2: *Scheme of the machine learning-based k-Nearest Neighbour classifier (kNN) algorithm*

8.3.2. Spatiotemporal bias correction

Once both the satellite event and its corresponding historical observed rainstorm are identified by the object-based/machine learning method, we used the spatiotemporal bias correction applied in ST-CORAbico to correct the primary sources of systematic error in the SPP IMERG data. This bias correction approach uses rainstorm error decomposition defined by Ebert and McBride (2000) to correct two systematic sources - displacement and volume. ST-CORAbico independently corrects both systematic errors via individual subtraction of the errors due to location for displacement and magnitude for volume by the following process:

Displacement correction

This correction corrects the shifting of the rainstorm event detected by the SPP. The methodology analyses the Principal Components (weighted centroids and eigenvectors and eigenspace orthogonal variables) of both objects to fit the geometric characteristics of the SPP event concerning the observed data. The process first matches the weighted

centroid and then uses the eigenvectors to rotate the SPP in space and time. In contrast, in ST-CORAbico, where timing shifting can be addressed, this displacement correction only uses the weighted centroid and the eigenvectors in space.

Volume correction

Volume correction corresponds to the subtraction of the magnitude source due to the inherent bias of the satellite rainstorm. For ascertaining this systematic error, ST-CORAbico adjusts all moments of the satellite cumulative probabilistic function ($ecdfs$) in terms of intensity using the statistical Empirical Quantile Method (EQM) Themeßl et al. (2012) concerning the reference data. EQM models the $ecdf$ for observed ($ecdf_{obs}$) and the satellite ($ecdf_{sat}$) data using the intensity rainstorm distribution I_s as:

$$EGM = ecdf_{obs}^{-1}(ecdf_{sat}^{-1}(I_s) \tag{8.3}$$

8.3.3. Performance evaluation

The performance of the proposed methodology is evaluated against a single kNN machine learning and probabilistic Gamma Quantile Mapping approach operational bias correction approaches. The single kNN method follows the same methodology used in the machine learning-based rainstorm analysis previously described. This method uses the information from the historical rainstorm database to search for the observed event with the closest spatiotemporal features to the SPP rainstorm. Compared with the hybrid ST-kNNbico method, this classification approach uses the observed corresponding object as the bias-corrected rainstorm instead of using it for further correction. We trained the KNN classifier based on six standardised spatiotemporal features detected by GPM-IMERG: centroid (longitude, latitude), duration, total volume, maximum rainstorm extension and intensity. We used the Euclidean distance metric to identify the rainstorm with the nearest neighbour pattern (k=1).

The operational Gamma Quantile method adjusts all moments of the gamma distribution function of the query rainstorm event estimated by SPP based on the quantile-quantile relationship between historical observed and estimated cumulative probabilistic functions (cdf). There are several statistical transformation functions for modelling the quantile-quantile relationship (Gudmundsson et al., 2012). In this study, we applied the non-parametric transferring function solved in Equation 8.3. Based on the rainstorm event classification for GPM-IMERG obtained by Laverde-Barajas et al. (2020a), the performance of the operational bias correction methods is evaluated for short and long-lived events using three error metrics: the Root Mean Square Error for magnitude error (RMSE), the level of bias (BIAS)and the correlation coefficient to analyse the linear correction between the observed and the bias-corrected rainstorm events.

8.4. Results

Based on the observed and satellite rainstorm database obtained by Laverde-Barajas et al. (2020a) during the monsoon seasons from 2014 to 2017, the performance of the hybrid bias correction method ST-KNNbico is evaluated during the monsoon season of 2018. This section will describe the performance of this method for correcting systematic errors in GPM-IMERG due to displacement and volume and the comparison against two operational bias correction approaches for short and long-lived rainstorm event types.

8.4.1. Systematic bias correction in the absence of a matched observed data

During the monsoon season of 2018, 47 rainstorm events from GPM-IMERG were identified by ST-CORA with KDE. For each rainstorm event, the ML composed identified the event with the closest spatiotemporal features to correct errors due to displacement and volume. Figure 8.3 describes the bias correction made by the hybrid ST-kNNbico method for one rainstorm event detected by GPM-IMERG using a historical observed rainstorm event estimated by the machine learning k-NN classifier. Panels (a) and (b) show the total rainfall event and the 4D rainstorm evolution of the satellite IMERG and the bias-corrected rainstorm event, including the individual corrections due to displacement and volume. On the other hand, panels (c) and (d) present the real and the kNN observed event used to correct the SPP-based rainstorm. This example shows that ST-kNNbico improved several features of the SPP-based rainstorm event. In terms of performance, RMSE and BIAS were considerably reduced. However, it is noted that the bias correction led to a negative bias due to the rainstorm underestimation.

Figure 8.4 shows the comparison between the RMSE, BIAS and r for the events from GPM-IMERG and the metrics obtained by ST-kNNbico and the individual corrections due to displacement and volume. According to the results, ST-kNNbico effectively reduced the RMSE and bias for all evaluated events. Errors due to volume had a significant influence on the correction of the rainstorm events. In the case of displacement, the individual correction had a marginal effect on the spatial correlation of the rainstorm event.

8.4.2. Performance comparison

Figure 8.5 shows the differences between the spatiotemporal features from the observed events and the features detected by GPM-IMERG and the operational bias correction methods ST-kNNbico, the single Machine learning-based kNN classifier (kNN) and the operational Gamma Quantile Mapping (GQM). The results show that all methods could improve the rainstorm characterisation in the absence of corresponding observed

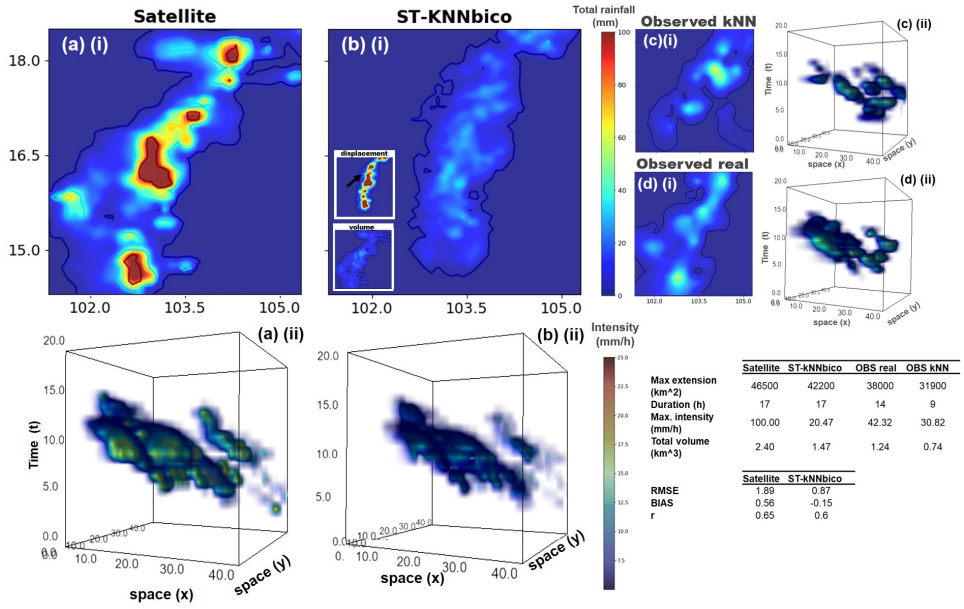


Figure 8.3: *Description of error correction and performance of ST-KNNbico for correcting a rainstorm event over the Mekong basin in Thailand during 19 Sept 2018. (a) satellite-based IMERG rainstorm; (b) bias corrected rainstorm with the individual correction due to displacement and volume; (c) corresponding historical event identified by the kNN method; (d) real observed event; subindexes (i) and (ii) are Total rainfall map and 4D rainstorm evolution of the rainstorm event respectively*

data. ST-kNNbico had the lowest difference in the total volume and the maximum spatial extension among the evaluated methods. It is found that all bias correction methods tended to underestimate the maximum intensity of several rainstorm events. kNN was the method with the lowest difference, followed by the GQM and the ST-kNNbico. The differences in duration showed that ST-kNNbico and GQM did not correct the errors due to timing, while kNN showed an underestimation in the duration of the rainstorm events.

The performance of each bias correction method for short- and long-lived rainstorm events is presented in Figure 8.6. Boxplots describe the mean error and distribution between the 25% and 75% percentiles for RMSE (Figures 8.6a,b), BIAS (Figures8.6c,d), and the correlation coefficient (Figures8.6a,b). Dots show the performance of each rainstorm event. Short- and long-lived events showed that the hybrid ST-kNNbico method had the lowest RMSE and bias among the evaluated methods. The single kNN displayed a lower RMSE and BIAS than GQM for the long-lived events. However, in short-lived events, GQM showed a better performance than kNN for RMSE and BIAS. In terms of the correlation, bias correction methods displayed a reduction in correlation coefficient, especially for short-lived events. In comparison, GQM has the highest correlation coefficient for both rainstorm events types followed by ST-kNNbico and finally, the single kNN method.

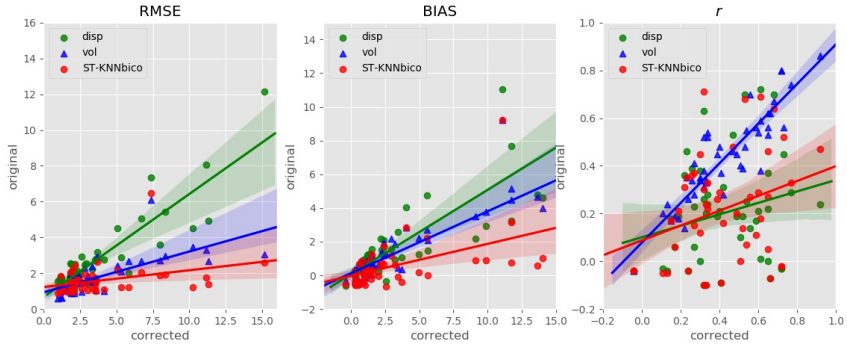


Figure 8.4: *Comparison between the RMSE, BIAS and r for the events from GPM-IMERG and the metrics obtained by ST-kNNbico and the individual corrections due to displacement and volume*

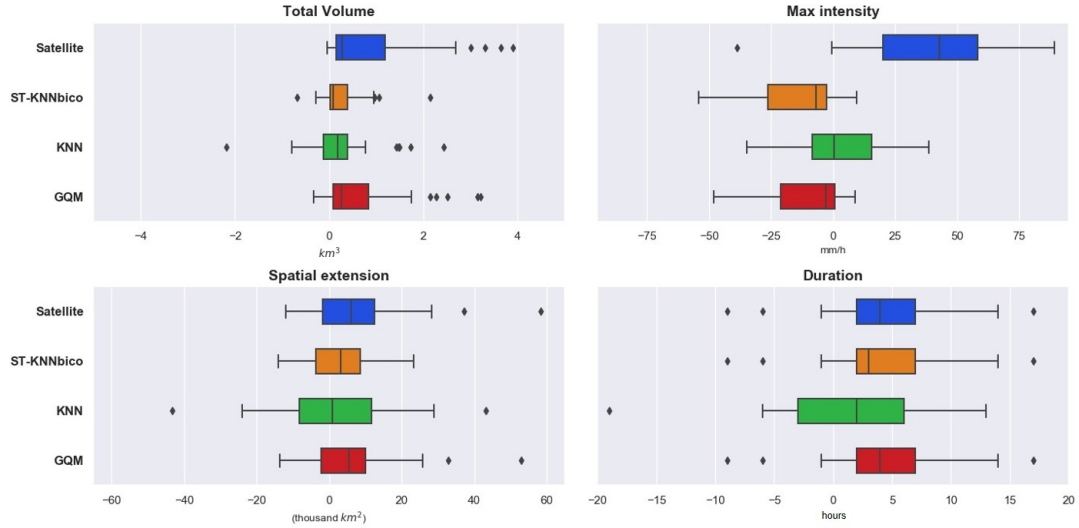


Figure 8.5: *Differences between the observed and satellite-based spatiotemporal features of rainstorm events obtained by GPM IMERG (blue), ST-kNNbico (orange), single kNN (green) and GQM (red) methods.*

8.5. Discussion

The development of this hybrid object-based Machine-learning bias correction method addressed the limitations of current physically-based methods when corresponding observed information is not available. Results from correcting multiple rainstorm events based on historical information indicated that ST-kNNbico can be helpful in operational applications or when practical information is not yet available.

The performance of ST-kNNbico over the Mekong basin in Thailand during the monsoon season of 2018 showed that displacement and volume corrections made by this method not only improved several spatiotemporal features of the rainstorm events detected by GPM-IMERG but also substantially reduced the RMSE and the level of BIAS of the SPP data. However, it is found that some mismatches in the location of

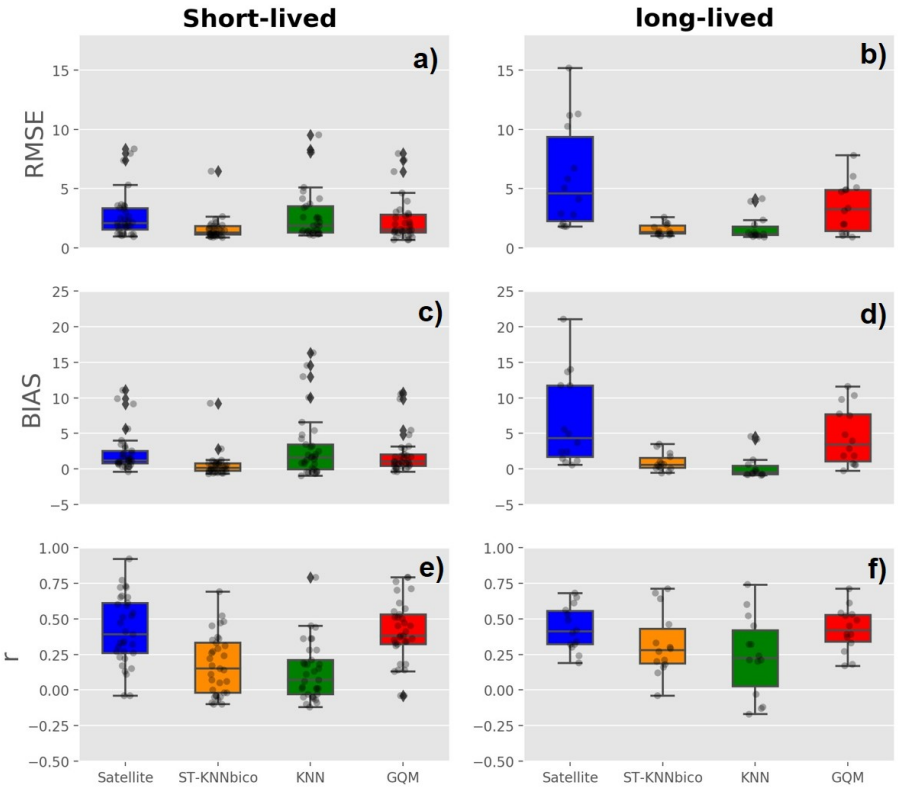


Figure 8.6: *Performance of GPM-IMERG(blue) ST-kNNbico (orange), single kNN (green) and GQM (red) for short- and long-lived rainstorm events. (a,b) RMSE; (c,d) bias; and, (e,f) correlation coefficient*

observed events led to reducing the correlation coefficient of the SPP event.

Overall, compared to the machine learning-based kNN classifier and the operational GQM, this hybrid approach showed a better performance for short and long-lived rainstorm events. In general, accounting for the total volume and the spatial extension of the rainstorm event were positively impacted the performance of the bias correction methods. However, all evaluated methods tended to underestimate the maximum intensity of the rainstorms. Regarding improvements in the duration, ST-kNNbico and the GQM consider the error due to timing, and the single KNN did not capture the total duration of the rainstorms.

We acknowledge further limitations in the proposed framework. Firstly, it is noted that the bias correction with k-NN is based on the assumption that past error structure determines its future configuration. In this sense, the capability of this method is limited to the number and diversity of the historical database of corresponding observed and SPP rainstorm events. Our research was based on 120 pairs of observed and SPP events analysed by Laverde-Barajas et al. (2020a) in the study area. However, it is recommended that a more extensive inventory is required to improve the performance of the ST-kNNbico method. Second, the user-defined formula to identify the param-

eters for rainstorm identification and segmentation in object-based approaches could be subjective, leading to imprecise rainstorm extractions. For this reason, a sensitivity analysis of the model parameters is fundamental for implementations in new areas.

Further analysis could enhance the machine learning component of ST-kNNbico by using more robust methods such as deep learning methods or incorporating a probabilistic analysis to identify the occurrence of matched corresponding observed rainstorms derived from multiple configurations of the rainstorm event (e.g. Johnson and Wang, 2012). Additional research lines could be pursued by applying this method into blending techniques for rainfall forecasting (e.g. rank-based quantile mapping approach in CHIRPS-GEFS, Shukla et al. (2020)) to improve the accuracy of global models for extreme weather forecasts.

8.6. Conclusions

We have proposed a hybrid object-based/Machine-learning bias correction method for rainstorm detection to correct systematic errors due to displacement and volume in the absence of a corresponding observed data. This method called ST-kNNbico incorporates a hybrid machine learning-based k-near Neighbour classifier into the recently developed Spatiotemporal Object-based bias correction method, ST-CORAbico, for reducing systematic errors in rainstorm events in SPPs using historical data. The performance for correcting 47 short-and long-lived rainstorm events across the Lower Mekong Basin in Thailand, detected by GPM-IMERG during the monsoon of 2018 based on the historical monsoonal rainstorm information from 2014 to 2017, was evaluated.

The results show that this method substantially reduced the RMSE and bias derived from systematic errors due to displacement and volume. Errors due to volume had the most considerable influence on the SPP correction. However, in terms of the correction due to displacement, several mismatches between observed events led to a marginal reduction in the correlation coefficient of the SPP, especially for short-lived rainstorm events. Compared to the two operational bias correction approaches, ST-kNNbico had superior performance in reducing RMSE and BIAS for short and long-lived rainstorm events, while GQM had the highest coefficient correlation among evaluated methods.

9

Conclusions, reflections and further developments

This concluding chapter summarises the main findings of the previous chapters based on the research questions presented in Chapter 1. Following this summary, the chapter discusses the methodological approaches used in this dissertation, outlining the technology developed in future research lines. The chapter ends by describing a further development of the spatiotemporal object-based methodology for rainfall analysis (STCORA with KDE) for the NASA/USAID SERVIR-Mekong program to monitor the severity of rainstorm events for the Lower Mekong Basin and generate alerts.

9.1. Conclusions

The development of this dissertation is framed according to five research questions. Below, we discuss how the research outputs address these questions.

The first question was: how important are the spatiotemporal dynamics of extreme rainfall events in satellite performance, with regards to error? In Chapter 4, we addressed this question by evaluating the performance of multiple near real-time SPPs to represent the spatial and temporal characteristics of extreme rainfall. This research demonstrated that spatiotemporal characteristics are a vital determining factor in the overall performance of the SPP data. It was found that short temporal and spatial extreme rainfall systems had more significant errors than extensive and lengthy duration events. Comparisons between VIS/IR and IRW/PMW satellites revealed the role of the type of retrievals in the accuracy to estimate extreme rainfall events. IRW/PMW-based SPP displayed a superior performance than VIS/IR-based SPP mainly due to the additional information of the hydrometeor provided by the PWM sensors.

In addition, we discussed the limitations of pixel-based approaches to analyse the spatial and temporal evolution of extreme precipitation events detected by SPP data

and to evaluate the performance in near real-time applications.

The second question was: how can spatiotemporal rainstorm dynamics be integrated into error estimation? This question was answered in Chapter 5. The Spatiotemporal Object-based Rainfall Analysis method, ST-CORA, addresses the limitations of pixel-based approaches to analyse the dynamics of extreme precipitation events in space and time. This physically-based approach provides a complete analysis of rainstorm events characterising many features such as volume, extension, duration, intensity, orientation, and speed. With ST-CORA, the performance of multiple gridded rainfall products can be evaluated by comparing individual features between observed and estimated events or directly measuring the level of error by using standard and categorical metrics. Applications of this method for error verification can be seen in Chapter 6 using the error decomposing proposed by Ebert (2005) to analyse the SPP error due to displacement, volume and displacement. In Chapter 7 and 8, an updated version of ST-CORA is presented incorporating a 4D analysis using the Multi-variable Kernel Density Estimation function algorithm (KDE) to improve the segmentation of rainstorm events at catchment scale.

The third question was: what is the effect of spatiotemporal errors on the hydrological response? In Chapter 6, we answered this question by evaluating the hydrological impact of systematic error sources (location and magnitude) from different rainstorm event types detected by the SPP CMORPH. Using the catchment of Capivari river as the study area, the hydrological responses for short and long-event rainstorm scenarios revealed that error sources due to location and magnitude from the SPP affect the shape, phase and amplitude of the streamflow in the catchment. In the short-lived rainstorm event, the error extraction due to location had a significant impact on reducing the streamflow error, while in the long-lived event, the error extraction due to magnitude led to an error reduction in the streamflow. These results demonstrated the deficiencies of satellites to estimate high convective rainstorm events and the vital role of the take into account the spatial and temporal error dynamics in SPP on the hydrological response.

The fourth question was: how can spatiotemporal error information be used to improve bias correction of satellite data? Chapter 7 addressed this error by developing the spatiotemporal object-based bias correction method for rainstorm analysis called ST-CORAbico. This method is designed to correct errors due to displacement and volume for rainstorm events detected by SPPs. This method evaluated in the Lower Mekong Basin in Thailand substantially reduced the RMSE and bias for several short- and long-lived rainstorm events, with a marginal impact on the correlation coefficient of the SPP data. Compared to two probabilistic bias correction methods, the distribution transformation method and the Gamma Quantile Mapping method, ST-CORAbico had the highest performance during the monsoon season from 2014 to 2017.

The fifth question was: how can artificial intelligence improve spatiotemporal bias

correction methods in operational applications? Chapter 8 answered a common challenge of object-based operational approaches when a parallel observed data set is not available. The hybrid object-based/Machine-learning bias correction method called ST-kNNbico was developed to correct systematic error due to displacement and volume for rainstorm events detected by SPPs using observed historical data. This approach applied to the Lower Mekong basin in Thailand showed that ST-kNNbico could be useful in operational applications or when practical information is unavailable. During the monsoon season of 2018, this method overcame other operational bias correction approaches by improving several spatiotemporal features of the rainstorm events detected by GPM-IMERG and substantially reducing the RMSE and the level of bias of the SPP data. However, it is essential to note that some mismatches in the location of historical data led to a reduction in the correlation coefficient of the SPP rainstorm event.

9.2. Reflections

The definition of rainstorm events

There is no universal definition of an extreme rainfall event. The term 'extreme rainfall system' or 'rainstorm event', like other environmental processes, has a statistical and physical definition based on the field of knowledge that surrounds it. In hydrology, for instance, this term is associated with a statistical concept defining "extreme rainfall events" as individual local weather variables exceeding critical levels on a continuous scale. In meteorology, on the other hand, the definition is more associated with the physics of the phenomenon. Under this concept, extreme events are defined according to extreme meteorological conditions derived from changes in atmospheric variables such as wind speed, moisture pressure, and temperature, among others (Stephenson et al., 2008).

Statistical and physical approaches have in common the selection of a threshold to identify a critical or extreme meteorological condition. However, both concepts differ in the method to calculate it. While in statistics, the definition of a extreme value is calculated by a punctual factor derived from statistical parameters such as percentiles (e.g AghaKouchak et al., 2011; Ringard et al., 2015; Boers et al., 2013) or the occurrence of rainfall (e.g. Gao and Liu, 2013; Miao et al., 2015; Tan and Santo, 2018). In physics, extreme conditions are calculated based on the rainfall system attributes(e.g. rate, magnitude, timing, spatial scale) since one number cannot fully describe the multidimensional nature of extreme events (Stephenson et al., 2008).

With satellite-based precipitation products for hydrological applications, the need to combine both approaches to define extreme rainfall events has become more relevant to understanding water-related disasters such as floods and landslides. Under hydrom-

eteorological criteria, physics can analyse the multidimensional structure of rainfall systems, while statistics can evaluate population attributes of extreme conditions.

Potential 2D, 3D and 4D object-based methods to analyse rainstorm events

Object-based methods have shown to be effective in analysing the physical dynamics of extreme precipitation systems based on their characteristics in space and time. Throughout this dissertation, rainstorm events have been analysed using 2D, 3D and 4D object-based methods. In Chapter 4, extreme rainfall events detected by different NRT SPPs were evaluated in space and time using a 2D object-based algorithm. First in space, analysing the spatial extension of extreme rainfall patterns and subsequently in time, evaluating the temporal aggregated composition of rainfall in 2D. This methodology provided valuable information regarding the importance of the spatiotemporal components in the performance of NRT SPPs. However, it also showed several limitations for analysing the configuration of rainfall events.

Based on the limitations of 2D object-based approaches, chapter 5 incorporated the temporal component of precipitation systems to develop the Spatiotemporal Contiguous Object-based Rainfall Analysis method (ST-CORA). This 3D object-based methodology uses a multidimensional connected-labelling component to analyse the features of rainstorm events detected by satellites. In ST-CORA, issues in the connected-labelling component algorithm due to noisy data and false merging objects are addressed by incorporating several morphological image processing algorithms. Nonetheless, existing gaps in the algorithm to segment rainstorm objects using binary information remained open.

Object-based methodologies in 2D and 3D rely on a connected labelling component algorithm to group similar pixels in space and time. However, it is found that managing precipitation systems with unknown topology in a binary domain limit the detection of edges for segmenting the rainfall object. In order to overcome this constraint, Chapter 7 incorporated a Multivariate Kernel Distribution Estimator algorithm (KDE) into the ST-CORA method. This algorithm analyses the 4-dimensional properties of the rainfall object to segment the rainstorm event. This approach showed that the inclusion of this probabilistic approach was able to delineate rainstorm events more effectively by removing unreal shapes and false merging. Further developments could incorporate several optimisation methods to calibrate the parameters used in ST-CORA with KDE in other study areas.

9.3. Practical implementation: the Rainstorm Tracker System

The Rainstorm Tracker System is an operational rainstorm analyser tool developed to monitor and alert the severity of rainstorm events over the Lower Mekong basin in near real-time and real-time. The rainstorm tracker relies on a peer-reviewed spatiotemporal rainstorm recognition method called the Spatiotemporal Object-based Rainfall Analysis method (ST-CORA). This method uses a 4D object-based recognition approach to track and evaluate the severity of rainstorms detected by satellites. This tool uses the information from different satellite-based rainfall products to identify the main features of rainstorm events in two operational modes: near-real-time (4-6 hours) and real-time (30-60 min). The level of severity of each rainstorm event is calculated based on the maximum intensity and total volume (Figure 9.1).

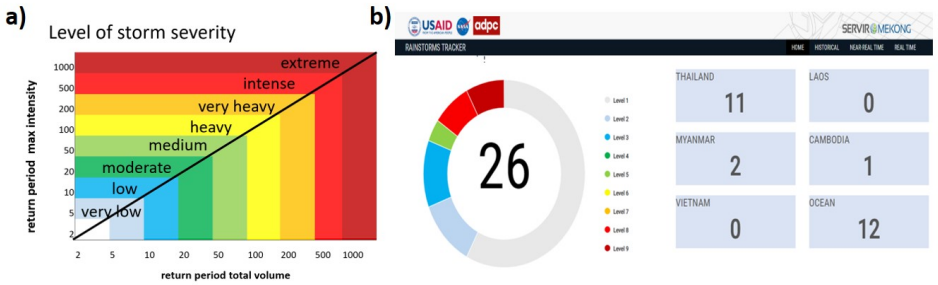


Figure 9.1: *Levels of rainstorm intensity based on the extremal analysis of maximum intensity and total volume; a) severity scale; b) display of rainstorm warnings per country in the rainstorm tracker system*

The Rainstorm Tracker is powered by the spatiotemporal rainstorm recognition method called the Spatiotemporal Object-based Rainfall Analysis method (ST-CORA). This method is developed by Laverde-Barajas et al. (2019) to analyse the main characteristics of the rainstorm events at the catchment scale. In Laverde-Barajas et al. (2020a), this method was successfully implemented over the Lower Mekong area in Thailand, incorporating a new multivariate probabilistic algorithm for rainstorm segmentation. Figure 9.2 presents the methodology steps of the rainstorm tracker using ST-CORA with KNN method, which can be summarised as follows:

1. Spatiotemporal Continuous Rainfall Analysis (ST-CRA): a multidimensional connected labelling component is implemented to identify convective rainfall systems in space and time.

2. Probabilistic rainstorm segmentation (KDE): rainstorm events are segmented using a multivariate edge detection based on Kernel Density Estimation (KDE).

3. Multi Threshold analysis: this process is applied to identify sub-rainstorm events over large convective systems.

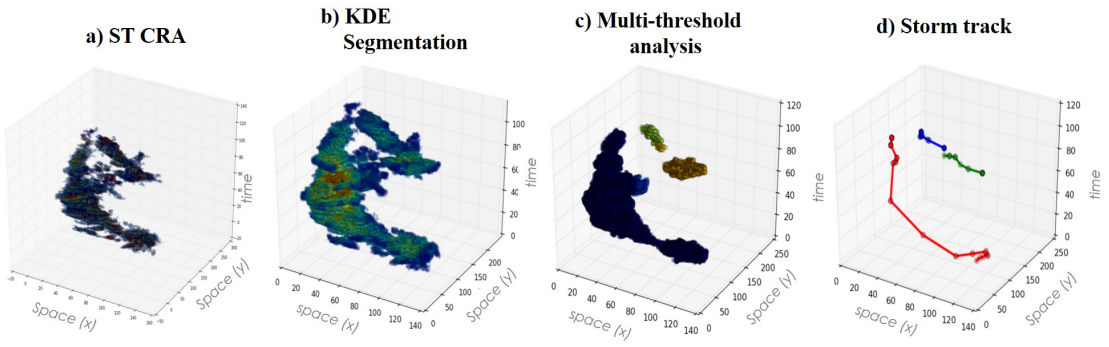


Figure 9.2: *Processes of ST-CORA with KDE: a) Spatiotemporal Continuous rainfall Analysis; b) Probabilistic rainstorm segmentation; c) Multi-threshold analysis; and c) rainstorm track detection for characterization of rainstorms/extreme events using 4D object detection.*

4. Rainstorm track detection: definition of the rainstorm track using the skeletal structure of the rainstorm event.

The geospatial information and the description of each rainstorm event are available through the web interface `http://rainstorms-servir.adpc.net`. In this portal, stakeholders can visualise and animate the spatial and temporal evolution of rainstorms in real and near-real-time. Figure 9.3 shows the interactive platform of the Rainstorm Tracker. The characteristics of the platform include: i) near-real-time and real-time monitoring; ii) filtering based on area, date, Storm severity, total volume, duration and max intensity; iii) Severity of the rainstorm; iv) Description of the main spatiotemporal features of the rainstorm; v) Download rainstorm event in format .png of the total storm event map and the raster file in format netCDF.

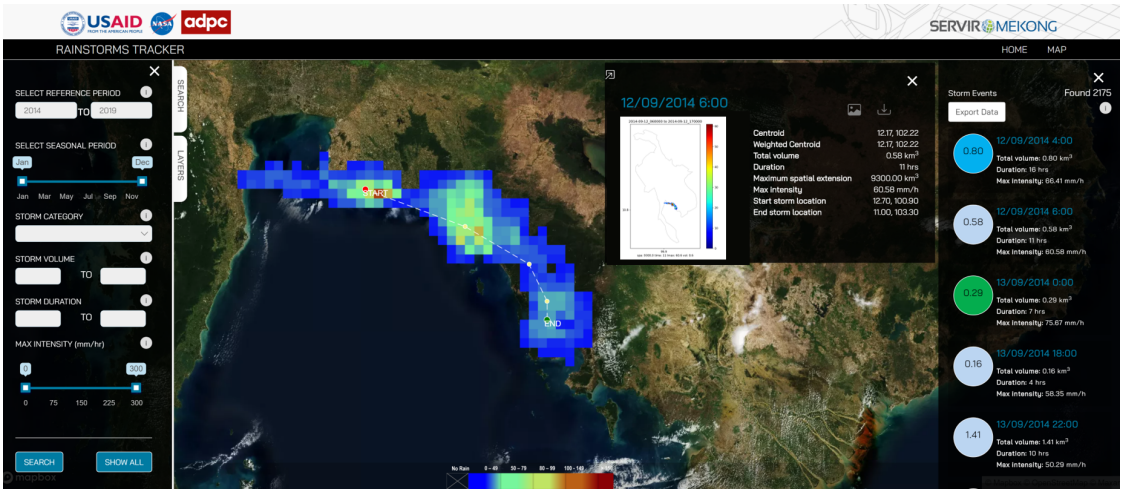


Figure 9.3: *Visualisation of the interactive platform of the Rainstorm tracker System*

The Historical information of the rainstorm event during monsoon season using the final version of GPM IMERG. Figures 9.4 presents the spatiotemporal features of two rainstorm events types over the Lower Mekong Basin in October 2020. Figure 9.4a

displays a mesoscale precipitation system over southeastern Asia, while Figure 9.4b shows a local rainfall rainstorm over Bangkok Thailand

In August of 2021, the Rainstorm Tracker System became fully operational.

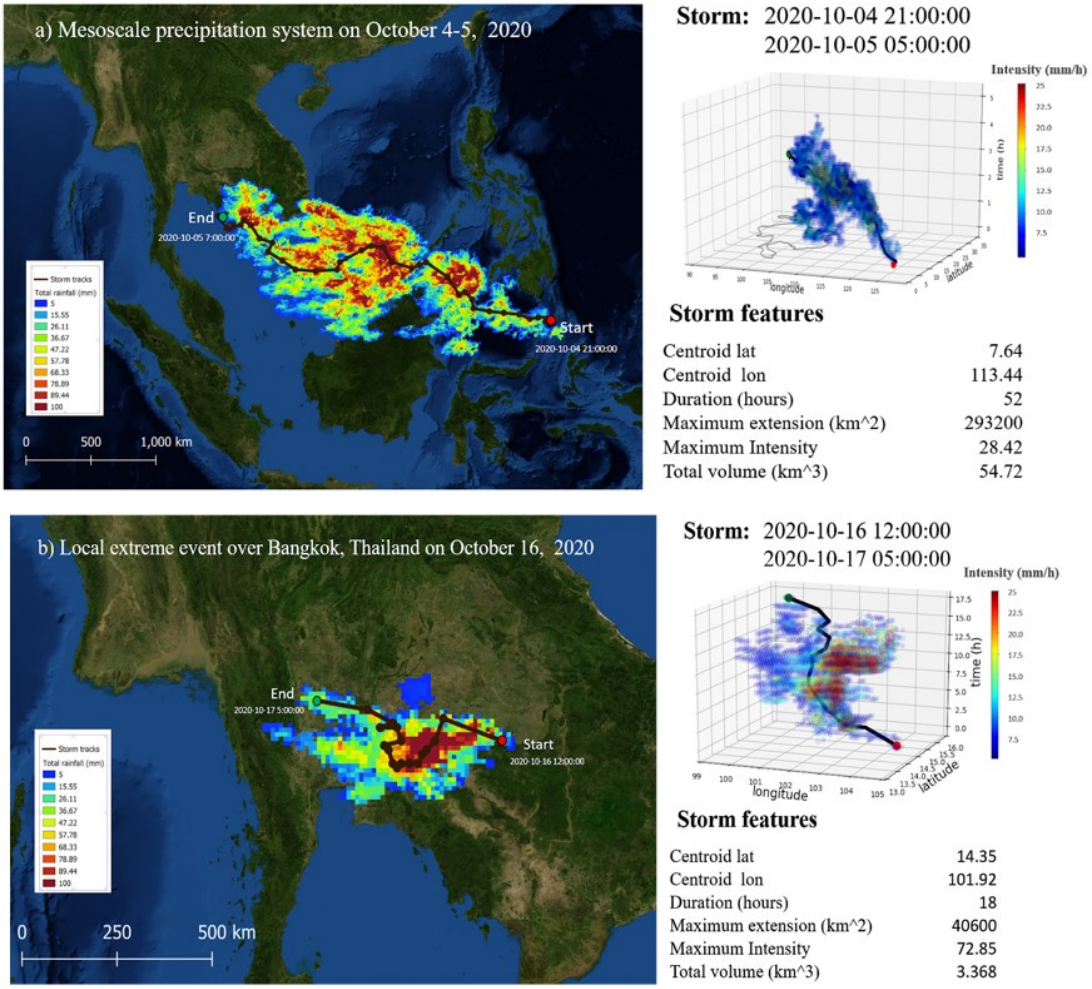


Figure 9.4: *Spatiotemporal features of two types of rainstorm events over the Lower Mekong Basin in October 2020: a) Mesoscale precipitation system; b) rainstorm event over Bangkok, Thailand.*

Acknowledgements

In this challenging but amazing journey of my PhD there countless people to whom I owe my deep gratitude for their unconditional support at all times. I especially want to thank:

I would like to acknowledge my supervisory team. Professor Dimitri Solomatine and Gerald Corzo from IHE Delft and Professor Remko Uijlenhoet from Wageningen University (now TUDelft). Thank all of you for your constant support and knowledge in every moment, and thank you for showing me the land when my mind was on the clouds. Gerald, I cannot express my gratitude to you with just simple words. Working with you has been an extraordinary and inspiring experience.

I would like to thanks Dr. Peeranan Towashiraporn from Asian Disaster Preparedness Centre (ADPC) in Bangkok Thailand. Dr Peeranan, many thanks for believe in me, and for trusting in the idea behind this research. Thanks the amazing team behind the SERVIR-Mekong program: Susantha, Hannah, Quyen, Kwan, Rishi, AJ, Nyein, Ankit, Ekapol, Book, Dimitris from ADPC; David, Farrukh and Ate from the Spatial Informatics, Arjen from Deltares and Amanda and Kel Markert from NASA coordination team. Thank you for the fruitfully discussions and support. Working with you all is a honour and a real pleasure. I also thanks the Regional Development Mission for Asia of USAID and NASA for the program support. We are keeping connecting space to village

I like to thank the Institute for Water Education IHE Delft and Delft University of Technology for being part of these both families. Especially thanks to Jolanda Boots, Anique Karsten and Floor Felix from IHE for their extraordinary support and for always show me your kind human touch in every step during my PhD.

This PhD research was made possible by the financial support from the Department of Science, Technology and Innovation of Colombia Colciencias. In addition, I would like to thanks Prof.dr.ir. Wim Bastiaanssen from the group of Water Accountant of IHE/TUDelft for the additional support obtained during the last part of my research.

Thanks to the local and regional organisations involved in this research. University of Campinas in Brazil especial thanks to Dr. José Gilberto Dalfré. The Hydroinformatics Institute in Thailand and the Mekong River Commission (MRC). I especially thank Dr. Winai Wangpimool, director of technical support division from the MRC Secretariat in Vientiane, Laos and Dr. Lam Hung Son, director of regional flood and drought management center (RFDMC) and his extraordinary team behind the RFDMC in Phnom Penh, Cambodia (Dr. Sothea, Anh, Kunthea, Sokong and Sameng). This project

was a big professional challenge for me and I glad to have you all for your constant assistance.

In general, I would like to thank the Netherlands and Thailand. Two countries that opened their doors to me during my doctoral stay and showed me the warmth of their people.

This PhD was a journey full of innumerable professional challenges and unforgettable life experiences. However, getting to this point would have been impossible without the help of my family and friends who have been with me in each of these moments. The incredible student community in Delft: The Latin and Italian community -Andresito, Juanse, Gonza, Omar, Alessio, Brego, Vero, Fer, Nata, Angie, Jessi, Alida, Juan Carlos, Neiler, Mario, Vitali, Thaine, Ana María, Irene, Pato, Andres Vargas, Paola, Adriana, Janice, Debora, Mauri, Irene, Yared (you are Colombian already), Alessandro, Luccia, Claudia, Santiago, Erika-, Claire and Stefan from IHE, and the lovely community of Floodies -Nikeh, Robert, Joanne, Okan, Tom, Solomon, Alex, Feroz, Marianne, Imra, Hadi, Josias, Fabio, Francesca, Nhilce, Hung- it is a pleasure having you in my life. I also thank my friends in Colombia and Mexico: Alvaro, Juan Pablo, Carolina, Cabula, Iskra, Miguelon, Ana, Irving. Especially thanks to Dr. Adrian Pedrozo-Acuña who was the person who showed me the value of research.

At last, and most important my family, my parents (Juan Manuel and Esperanza), my siblings (Alex, Andres y Keka), my aunts (Myriam, Rosita, Fanny and Carola); my cousins (Diana, Paola, Juank, Camilo, Julian and Laura) my Tata (you are always with me) and my lovely nephew Juanjo.

A mis gorditos: Gracias por darme las alas para poder volar. Ustédes son mi luz, mi vida y mi motivación mas grande. A mis hermanos: Ustedes han sido el mejor regalo de mis padres. Gracias por sus enseñanzas, por su aliento, por su comprension y por sobre todo su amor. A mi tias y primos: No puedo incontran un mejor aliado que ustedes. Gracias por su amor indondicional. A mi sobrino: Sueña, sueña muy alto y nunca pares de soñar. Siempre me tendras a tu lado como un aliado indondicional.

This final stage concludes with the beginning of a new stage in my life with an extraordinary person. Pin, my partner and now my wife, this degree belongs also to you. Thank you for teaching me never to give up and for showing me the light in moments of doubt. You were the kindest advisor and the harshest reviewer. *Te amo Osita*

About the author

Miguel Laverde-Barajas is a Colombian researcher specialised in hydroinformatics technologies with the focus on earth observations. During his professional carrier, Miguel has contributed in projects in water and disaster management in multiple countries, including the National Water Policy in Colombia, the Integral Water Plan of the State of Tabasco and Panuco in Mexico and the SERVIR-Mekong Program in Thailand, Laos, Vietnam, Myanmar and Cambodia. He holds BSc in Environmental Management from the District University of Bogota Colombia and MSc with honours in Coastal and River Engineering from the National Autonomous University of Mexico.

Since 2015, Miguel is a PhD research at the IHE Delft Institute for Water Education and Delft University of Technology in Netherlands founded by the Department of Science, Technology and Innovation of Colombia COLCIENCIAS. In 2019, he is working as a Geospatial Data Scientist at the Asian Disaster Preparedness Center in Bangkok Thailand under the NASA/USAID SERVIR-Mekong program. His interests include travelling, hiking and playing football.

List of Publications

1. **Laverde-Barajas, M. A.**, Corzo, G. A., Poortinga, A., Chishtie, F., Meechaiya, C., Jayasinghe, S. Towashiraporn P., Markert A., Son L.H., Sothea K., Boonya-Aroonnet S., Chaowiwat W., Uijlenhoet R. U., and Solomatine, D. P. (2020), ST-CORAbico: A spatiotemporal pattern recognition framework to bias correct satellite-based rainfall data for storm detection with or without observed data. *Journal of remote Sensing*, 12(21), 3538.

2. **Laverde-Barajas, M. A.**, Corzo Perez G. A., F. Chishtie, F., Poortinga, A., Uijlenhoet, R., and Solomatine, D. P. (2020) Decomposing the satellite-based rainfall errors in flood estimation: the hydrological responses using a spatiotemporal object-based verification method,*Journal of Hydrology* 12, 591, 125554.

3. **Laverde-Barajas, M. A.**, Corzo, G., Bhattacharya, B., Uijlenhoet, R., and Solomatine, D. P. (2019). Spatiotemporal analysis of extreme rainfall events using an object-based approach. In *Spatiotemporal Analysis of Extreme Hydrological Events* (pp. 95-112). Elsevier.

4. **Laverde-Barajas, M. A.**, Corzo Perez, G. A., Dalfré Filho, J. G., and Solomatine, D. P. (2018). Assessing the performance of near real-time rainfall products to represent spatiotemporal characteristics of extreme events: case study of a subtropical catchment in south-eastern Brazil.*International Journal of Remote Sensing* (21), 7568-7586.

5. Breña-Naranjo, J. A., **Laverde-Barajas, M. A.**, and Pedrozo-Acuña, A. (2017). Changes in pan evaporation in Mexico from 1961 to 2010. *International Journal of Climatology*, 37(1), 204-213.

6. Pedrozo-Acuña, A., Damania, R., **Laverde-Barajas, M. A.**, and Mira-Salama, D. (2015). Assessing the consequences of sea-level rise in the coastal zone of Quintana Roo, Mexico: the costs of inaction. *Journal of coastal conservation*, 19(2), 227-240.

Conference proceedings

1. **Laverde-Barajas, M. A.**, Meechaiya, C., Thannarot K., Khem S., Son L. H., Nguyen Q. A., Towashiraporn, P., Priestley H., Chishtie, F., Corzo, G., Uijlenhoet, R., Solomatine, D. P. (2021). The Rainstorm Tracker system: An operational monitoring and alert system for the Lower Mekong Region. AGU Fall Meeting 2021 (AGU2021), 13-17 December. New Orleans, United States.

2. **Laverde-Barajas, M. A.**, Meechaiya, C., Poortinga A., (Weigel) Markert, A., Nicolau, A., Jayasinghe, S., Basnayake, S. B., Chishtie, F., Towashiraporn, P., Corzo, G., Saah, D. S., Uijlenhoet, R., Solomatine, D. P. (2020). Climatology of monsoonal rainstorm events over the Lower Mekong Region: A regional analysis using a spatiotemporal Object-based rainfall method for GPM-IMERG. AGU Fall Meeting 2020 (AGU2020), 1-17 December. Online, United States.

3. **Laverde-Barajas, M. A.**, Corzo Perez, G. A., and Solomatine, D. P. (2019). The impact of satellite-based error sources for extreme rainfall events on the hydrological response. EGU General Assembly 2019 (EGU2019), 7-12 April; Vienna, Austria.

4. **Laverde-Barajas, M. A.**, Corzo Perez, G. A., and Solomatine, D. P. (2017). Uncertainty estimation of satellite rainfall products associated with spatiotemporal representation of extreme events. EGU General Assembly 2017 (EGU2017), 23-28 April; Vienna, Austria.

5. **Laverde-Barajas, M. A.**, Corzo Perez, G. A., and Solomatine, D. P. (2017). Pattern recognition techniques in estimation of rainfall extreme events spatiotemporal characteristic: case study of a subtropical catchment in south-eastern Brazil. EGU General Assembly 2017 (EGU2017), 23-28 April; Vienna, Austria.

References

Abdi, H. and Williams, L. J. (2010). Principal component analysis. *Wiley Interdisciplinary Reviews: Computational Statistics*, 2(4):433–459.

Abtew, W. (1996). Evapotranspiration measurements and modeling for three wetland systems in south florida. *JAWRA Journal of the American Water Resources Association*, 32(3):465–473.

Acharya, T. and Ray, A. K. (2005). *Image Processing: Principles and Applications*. John Wiley & Sons.

ADRC, A. D. R. C. (2021). *Natural Disaster Data Book 2019: An Analytical Overview*. ADR Center - adrc.asia.

AghaKouchak, A., Behrangi, A., Sorooshian, S., Hsu, K., and Amitai, E. (2011). Evaluation of satellite-retrieved extreme precipitation rates across the central united states. *Journal of Geophysical Research: Atmospheres*, 116:D02115.

AghaKouchak, A. and Mehran, A. (2013). Extended contingency table: Performance metrics for satellite observations and climate model simulations. *Water Resources Research*, 49(10):7144–7149.

AghaKouchak, A. and Nakhjiri, N. (2012). A near real-time satellite-based global drought climate data record. *Environmental Research Letters*, 7(4):044037.

Ahijevych, D., Gilleland, E., Brown, B. G., and Ebert, E. E. (2009). Application of spatial verification methods to idealized and NWP-gridded precipitation forecasts. *Weather and Forecasting*, 24(6):1485–1497.

Aksoy, B. and Ercanoglu, M. (2012). Landslide identification and classification by object-based image analysis and fuzzy logic: An example from the azdavay region (kastamonu, turkey). *Computers & Geosciences*, 38(1):87–98.

Allen, M. R. and Ingram, W. J. (2002). Constraints on future changes in climate and the hydrologic cycle. *Nature*, 419(6903):228–232.

Amatya, P., Kirschbaum, D., Stanley, T., and Tanyas, H. (2021). Landslide mapping using object-based image analysis and open source tools. *Engineering Geology*, 282:106000.

Amorati, R., Alberoni, P. P., and Fornasiero, A. (2012). Operational bias correction of hourly radar precipitation estimate using rain gauges. In *The 7th European Conference on radar in Meteorology and Hydrology (ERAD)*.

Aonashi, K., Awaka, J., Hirose, M., Kozu, T., Kubota, T., Liu, G., Shige, S., Kida, S., Seto, S., Takahashi, N., et al. (2009). Gsmap passive microwave precipitation retrieval algorithm: Algorithm description and validation. *Journal of the Meteorological Society of Japan. Ser. II*, 87:119–136.

Arnaud, P., Bouvier, C., Cisneros, L., and Dominguez, R. (2002). Influence of rainfall spatial variability on flood prediction. *Journal of Hydrology*, 260(1):216–230.

Artan, G., Gadain, H., Smith, J. L., Asante, K., Bandaragoda, C. J., and Verdin, J. P. (2007). Adequacy of satellite derived rainfall data for stream flow modeling. *Natural Hazards*, 43(2):167–185.

Ayat, H., Evans, J. P., Sherwood, S., and Behrangi, A. (2021). Are storm characteristics the same when viewed using merged surface radars or a merged satellite product? *Journal of Hydrometeorology*, 22(1):43–62.

Azarderakhsh, M., Rossow, W. B., Papa, F., Norouzi, H., and Khanbilvardi, R. (2011). Diagnosing water variations within the amazon basin using satellite data. *Journal of Geophysical Research: Atmospheres*, 116:D24107.

Baldwin, M. E. and Kain, J. S. (2006). Sensitivity of several performance measures to displacement error, bias, and event frequency. *Weather and forecasting*, 21(4):636–648.

Barrella, W. and Petrere Jr, M. (2003). Fish community alterations due to pollution and damming in tietê and paranapanema rivers (brazil). *River Research and Applications*, 19(1):59–76.

Beck, H. E., van Dijk, A. I. J. M., Levizzani, V., Schellekens, J., Miralles, D. G., Martens, B., and de Roo, A. (2016). MSWEP: 3-hourly 0.25 ◦ global gridded precipitation (1979–2015) by merging gauge, satellite, and reanalysis data. *Hydrol. Earth Syst. Sci. Discuss.*, 2016:1–38.

Beck, H. E., Vergopolan, N., Pan, M., Levizzani, V., Van Dijk, A. I., Weedon, G. P., Brocca, L., Pappenberger, F., Huffman, G. J., and Wood, E. F. (2017). Global-scale evaluation of 22 precipitation datasets using gauge observations and hydrological modeling. *Hydrology and Earth System Sciences*, 21(12):6201–6217.

Bennett, N. D., Croke, B. F., Guariso, G., Guillaume, J. H., Hamilton, S. H., Jakeman, A. J., Marsili-Libelli, S., Newham, L. T., Norton, J. P., Perrin, C., et al. (2013).

Characterising performance of environmental models. *Environmental Modelling & Software*, 40:1–20.

Berjón, D., Cuevas, C., Morán, F., and García, N. (2018). Real-time nonparametric background subtraction with tracking-based foreground update. *Pattern Recognition*, 74:156–170.

Bethel, E. W., Childs, H., and Hansen, C. (2012). *High performance visualization: Enabling extreme-scale scientific insight*. CRC Press.

Bhuiyan, M. A. E., Nikolopoulos, E. I., and Anagnostou, E. N. (01 Nov. 2019). Machine learning–based blending of satellite and reanalysis precipitation datasets: A multiregional tropical complex terrain evaluation. *Journal of Hydrometeorology*, 20(11):2147–2161.

Bitew, M. M. and Gebremichael, M. (2011). Evaluation of satellite rainfall products through hydrologic simulation in a fully distributed hydrologic model. *Water Resources Research*, 47(6).

Blaschke, T. (2010). Object based image analysis for remote sensing. *ISPRS journal of photogrammetry and remote sensing*, 65(1):2–16.

Blaschke, T., Burnett, C., and Pekkarinen, A. (2004). Image segmentation methods for object-based analysis and classification. In *Remote sensing image analysis: Including the spatial domain*, pages 211–236. Springer.

Boers, N., Bookhagen, B., Marwan, N., and Kurths, J. (2015). Spatiotemporal characteristics and synchronization of extreme rainfall in south america with focus on the andes mountain range. *Climate Dynamics*, 46(1):601–617.

Boers, N., Bookhagen, B., Marwan, N., Kurths, J., and Marengo, J. (2013). Complex networks identify spatial patterns of extreme rainfall events of the south american monsoon system. *Geophysical Research Letters*, 40(16):4386–4392.

Bouwer, L., Aerts, J., van de Coterlet, G., van de Giesen, N., Gieske, A., and Mannaerts, C. (2004). Evaluating downscaling methods for preparing global circulation model gcm data for hydrological impact modelling. In *Climate change in contrasting river basins: adaptation strategies for water, food and environment*, pages 25–47. CABI.

Breña-Naranjo, J. A., Pedrozo-Acuña, A., and Rico-Ramirez, M. A. (2015). World's greatest rainfall intensities observed by satellites. *Atmospheric Science Letters*, 16(3):420–424.

Brown, B. G., Bullock, Y. R., Davis, C. A., Gotway, J. H., Chapman, M. B., Takacs, A., Gillel, E., Manning, K., and Mahoney, J. L. (2004). New verification approaches for convective weather forecasts. In *Preprints, 11th Conference on Aviation, Range, and Aerospace, Hyannis*, pages 3–8.

Brunini, O. (2017). Rede meteorológica de superfície para o estado de são paulo.

Bui, Y. T., Orange, D., Visser, S., Hoanh, C. T., Laissus, M., Poortinga, A., Tran, D. T., and Stroosnijder, L. (2014). Lumped surface and sub-surface runoff for erosion modeling within a small hilly watershed in northern vietnam. *Hydrological processes*, 28(6):2961–2974.

Bullock, R. (2011). Development and implementation of mode time domain object-based verification. *Proc. 24th Conference Weather and Forecasting, (24–27 January 2011, Seattle, Washington)*.

Cai, W., McPhaden, M. J., Grimm, A. M., Rodrigues, R. R., Taschetto, A. S., Garreaud, R. D., Dewitte, B., Poveda, G., Ham, Y.-G., Santoso, A., et al. (2020). Climate impacts of the el niño–southern oscillation on south america. *Nature Reviews Earth & Environment*, 1(4):215–231.

Cannon, A. J. (01 Oct. 2016). Multivariate bias correction of climate model output: Matching marginal distributions and intervariable dependence structure. *Journal of Climate*, 29(19):7045–7064.

Cannon, A. J. (2018). Multivariate quantile mapping bias correction: An n-dimensional probability density function transform for climate model simulations of multiple variables. *Climate Dyn.*, 50:31–49.

Cannon, A. J., Sobie, S. R., and Murdock, T. Q. (2015a). Bias correction of gcm precipitation by quantile mapping: How well do methods preserve changes in quantiles and extremes? *Journal of Climate*, 28(17):6938–6959.

Cannon, A. J., Sobie, S. R., and Murdock, T. Q. (2015b). Bias correction of gcm precipitation by quantile mapping: How well do methods preserve changes in quantiles and extremes? *Journal of Climate*, 28(17):6938–6959.

Cao, C., Dragićević, S., and Li, S. (2019). Land-use change detection with convolutional neural network methods. *Environments*, 6(2):25.

Casati, B. (2010). New developments of the intensity-scale technique within the spatial verification methods intercomparison project. *Weather and Forecasting*, 25(1):113–143.

Casati, B., Ross, G., and Stephenson, D. (2004). A new intensity-scale approach for the verification of spatial precipitation forecasts. *Meteorological Applications*, 11(2):141–154.

Casati, B., Wilson, L., Stephenson, D., Nurmi, P., Ghelli, A., Pocernich, M., Damrath, U., Ebert, E., Brown, B., and Mason, S. (2008). Forecast verification: current status and future directions. *Meteorological Applications: A journal of forecasting, practical applications, training techniques and modelling*, 15(1):3–18.

Casse, C., Gosset, M., Peugeot, C., Pedinotti, V., Boone, A., Tanimoun, B., and Decharme, B. (2015). Potential of satellite rainfall products to predict niger river flood events in niamey. *Atmospheric Research*, 163:162–176.

Ceperuelo, M., Llasat, M., and Rigo, T. (2006). Rainfall events and hailstorms analysis program (rhap). *Advances in Geosciences*, 7.

Chang, W., Stein, M. L., Wang, J., Kotamarthi, V. R., and Moyer, E. J. (2016). Changes in spatiotemporal precipitation patterns in changing climate conditions. *Journal of Climate*, 29(23):8355–8376.

Chaudhary, S. and Dhanya, C. (2020). Decision tree-based reduction of bias in monthly imerg satellite precipitation dataset over india. *h2oj*, 3(1):236–255.

Chen, F., Yang, X., Ji, C., Li, Y., Deng, F., and Dong, M. (2019). Establishment and assessment of hourly high-resolution gridded air temperature data sets in zhejiang, china. *Meteorological Applications*, 26(3):396–408.

Chen, T.-B., Lu, H. H.-S., Lee, Y.-S., and Lan, H.-J. (2008a). Segmentation of cdna microarray images by kernel density estimation. *Journal of biomedical informatics*, 41(6):1021–1027.

Chen, Y.-C., Wei, C., and Yeh, H.-C. (2008b). Rainfall network design using kriging and entropy. *Hydrological Processes: An International Journal*, 22(3):340–346.

Chung, U. and Yun, J. I. (2004). Solar irradiance-corrected spatial interpolation of hourly temperature in complex terrain. *Agricultural and forest meteorology*, 126(1-2):129–139.

Clark, A. J., Bullock, R. G., Jensen, T. L., Xue, M., and Kong, F. (2014). Application of object-based time-domain diagnostics for tracking precipitation systems in convection-allowing models. *Weather and Forecasting*, 29(3):517–542.

Clark, A. J., Kain, J. S., Marsh, P. T., Correia Jr, J., Xue, M., and Kong, F. (2012). Forecasting tornado pathlengths using a three-dimensional object identification algorithm applied to convection-allowing forecasts. *Weather and Forecasting*, 27(5):1090–1113.

Corzo, G. and Solomatine, D. (2007). Baseflow separation techniques for modular artificial neural network modelling in flow forecasting. *Hydrological Sciences Journal*, 52(3):491–507.

Corzo Perez, G. (2009). *Hybrid models for hydrological forecasting: integration of data-driven and conceptual modelling techniques*. UNESCO-IHE, Institute for Water Education.

CRED, U. (2015). The human cost of weather related disasters. *The United Nations Office for Disaster Risk Reduction (UNISDR) https://doi.org/10.1017/CBO9781107415324*, 4.

Davis, C. A., Brown, B., and Bullock, R. (2009a). Spatial and temporal object-based evaluation of numerical precipitation forecasts. In *Preprints, 23rd Conf. on Weather Analysis and Forecasting/19th Conf. on Numerical Weather Prediction, Omaha, NE, Amer. Meteor. Soc. A*, volume 5.

Davis, C. A., Brown, B. G., Bullock, R., and Halley-Gotway, J. (2009b). The method for object-based diagnostic evaluation (MODE) applied to numerical forecasts from the 2005 NSSL/SPC spring program. *Weather and Forecasting*, 24(5):1252–1267.

de Moraes Gonçalves, N., da Cal Seixas, S. R., de Moraes Hoeffel, J. L., and de Lima, F. B. (2015). Floods and socioenvironmental vulnerability in the municipality of atibaia, sp–brazil. *Revista Meio Ambiente e Sustentabilidade*, 9(4):100–124.

DeGaetano, A. T. and Belcher, B. N. (2007). Spatial interpolation of daily maximum and minimum air temperature based on meteorological model analyses and independent observations. *Journal of Applied Meteorology and Climatology*, 46(11):1981–1992.

Dekens, L., Parey, S., Grandjacques, M., and Dacunha-Castelle, D. (2017). Multivariate distribution correction of climate model outputs: A generalization of quantile mapping approaches. *Environmetrics*, 28(6):e2454.

Delgado, J., Merz, B., and Apel, H. (2012). A climate-flood link for the lower mekong river. *Hydrology and Earth System Sciences*, 16(5):1533–1541.

Demaria, E., Rodriguez, D., Ebert, E., Salio, P., Su, F., and Valdes, J. B. (2011). Evaluation of mesoscale convective systems in south america using multiple satellite products and an object-based approach. *Journal of Geophysical Research: Atmospheres*, 116(D8).

Dembélé, M., Schaefli, B., van de Giesen, N., and Mariéthoz, G. (2020). Suitability of 17 rainfall and temperature gridded datasets for largescale hydrological modelling in

west africa. *Hydrol. Earth Syst. Sci. Discuss., https://doi. org/10.5194/hess-2020-68, in review*.

Dembélé, M. and Zwart, S. J. (2016). Evaluation and comparison of satellite-based rainfall products in burkina faso, west africa. *International Journal of Remote Sensing*, 37(17):3995–4014.

Derin, Y. and Yilmaz, K. K. (2014). Evaluation of multiple satellite-based precipitation products over complex topography. *Journal of Hydrometeorology*, 15(4):1498–1516.

Desclée, B., Bogaert, P., and Defourny, P. (2006). Forest change detection by statistical object-based method. *Remote sensing of environment*, 102(1-2):1–11.

Dinku, T., Ceccato, P., Grover-Kopec, E., Lemma, M., Connor, S. J., and Ropelewski, C. F. (2007). Validation of satellite rainfall products over east africa's complex topography. *International Journal of Remote Sensing*, 28(7):1503–1526.

Dinku, T., Ruiz, F., Connor, S. J., and Ceccato, P. (2010). Validation and intercomparison of satellite rainfall estimates over colombia. *Journal of Applied Meteorology and Climatology*, 49(5):1004–1014.

Dixon, M. and Wiener, G. (1993). TITAN: Thunderstorm identification, tracking, analysis, and nowcasting—a radar-based methodology. *J. Atmos. Oceanic Technol.*, 10(6):785–797.

Donat, M. G., Lowry, A. L., Alexander, L. V., Ogorman, P. A., and Maher, N. (2016). More extreme precipitation in the world[rsquor]s dry and wet regions. *Nature Clim. Change*, advance online publication.

Dottori, F., Salamon, P., Bianchi, A., Alfieri, L., Hirpa, F. A., and Feyen, L. (2016). Development and evaluation of a framework for global flood hazard mapping. *Advances in water resources*, 94:87–102.

DSMW-FAO (1960). Fao digital soil map of the world.

Dunkerley, D. (2008). Rain event properties in nature and in rainfall simulation experiments: a comparative review with recommendations for increasingly systematic study and reporting. *Hydrological Processes*, 22(22):4415–4435.

Dunkerley, D. L. (2010). How do the rain rates of sub-event intervals such as the maximum 5- and 15-min rates (i5 or i30) relate to the properties of the enclosing rainfall event? *Hydrological Processes*, 24(17):2425–2439.

Duro, D. C., Franklin, S. E., and Dubé, M. G. (2012). A comparison of pixel-based and object-based image analysis with selected machine learning algorithms for the

classification of agricultural landscapes using spot-5 hrg imagery. *Remote sensing of environment*, 118:259–272.

Ebert, B. (2005). CRA (entity-based) verification.

Ebert, E. and McBride, J. (2000). Verification of precipitation in weather systems: Determination of systematic errors. *Journal of Hydrology*, 239(1-4):179–202.

Ebert, E. E. and Gallus, W. A. (2009). Toward better understanding of the contiguous rain area (CRA) method for spatial forecast verification. *Weather and Forecasting*, 24(5):1401–1415.

Ebert, E. E., Janowiak, J. E., and Kidd, C. (2007). Comparison of near-real-time precipitation estimates from satellite observations and numerical models. *Bulletin of the American Meteorological Society*, 88(1):47–64.

Ehret, U. and Zehe, E. (2011). Series distance-an intuitive metric to quantify hydrograph similarity in terms of occurrence, amplitude and timing of hydrological events. *Hydrology & Earth System Sciences*, 15(3).

Falck, A. S., Maggioni, V., Tomasella, J., Vila, D. A., and Diniz, F. L. (2015). Propagation of satellite precipitation uncertainties through a distributed hydrologic model: A case study in the tocantins–araguaia basin in brazil. *Journal of Hydrology*, 527:943–957.

Fang, G., Yang, J., Chen, Y., and Zammit, C. (2015). Comparing bias correction methods in downscaling meteorological variables for a hydrologic impact study in an arid area in china. *Hydrology and Earth System Sciences*, 19(6):2547–2559.

Fenicia, F., Solomatine, D. P., Savenije, H. H. G., and Matgen, P. (2007). Soft combination of local models in a multi-objective framework. *Hydrol. Earth Syst. Sci.*, 11(6):1797–1809.

Ferraro, R. and Smith, T. (2013). Global precipitation monitoring. In *Satellite-based Applications on Climate Change*, pages 81–93. Springer.

Foufoula-Georgiou, E. and Vuruputur, V. (2001). *Patterns and organization in precipitation*. Cambridge University Press.

Funk, C., Peterson, P., Landsfeld, M., Pedreros, D., Verdin, J., Shukla, S., Husak, G., Rowland, J., Harrison, L., Hoell, A., et al. (2015). The climate hazards infrared precipitation with stations: a new environmental record for monitoring extremes. *Scientific Data*, 2:150066.

Gagne, D. J., McGovern, A., and Xue, M. (2014). Machine learning enhancement of storm-scale ensemble probabilistic quantitative precipitation forecasts. *Weather and Forecasting*, 29(4):1024–1043.

Gao, Y. and Liu, M. (2013). Evaluation of high-resolution satellite precipitation products using rain gauge observations over the tibetan plateau. *Hydrology and Earth system sciences*, 17(2):837–849.

Gash, J. H., Lloyd, C., and Lachaud, G. (1995). Estimating sparse forest rainfall interception with an analytical model. *Journal of Hydrology*, 170(1-4):79–86.

Gebere, S. B., Alamirew, T., Merkel, B. J., and Melesse, A. M. (2015). Performance of high resolution satellite rainfall products over data scarce parts of eastern ethiopia. *Remote Sensing*, 7(9):11639–11663.

Gebregiorgis, A. S. and Hossain, F. (2015). How well can we estimate error variance of satellite precipitation data around the world? *Atmospheric Research*, 154:39–59.

Geem, Z. W., Kim, J. H., and Loganathan, G. V. (2001). A new heuristic optimization algorithm: harmony search. *simulation*, 76(2):60–68.

Gilleland, E., Ahijevych, D., Brown, B. G., Casati, B., and Ebert, E. E. (2009). Intercomparison of spatial forecast verification methods. *Weather and Forecasting*, 24(5):1416–1430.

Grams, J. S., Gallus Jr, W. A., Koch, S. E., Wharton, L. S., Loughe, A., and Ebert, E. E. (2006). The use of a modified ebert–mcbride technique to evaluate mesoscale model qpf as a function of convective system morphology during ihop 2002. *Weather and forecasting*, 21(3):288–306.

Grayson, R. and Blöschl, G. (2001). *Spatial Patterns in Catchment Hydrology: Observations and Modelling*. CUP Archive. Google-Books-ID: ZPQ8AAAAIAAJ.

Gudmundsson, L., Bremnes, J. B., Haugen, J. E., and Engen-Skaugen, T. (2012). Downscaling rcm precipitation to the station scale using statistical transformations–a comparison of methods. *Hydrology and Earth System Sciences*, 16(9):3383–3390.

Gumindoga, W., Rientjes, T. H., Haile, A. T., Makurira, H., and Reggiani, P. (2019). Performance of bias-correction schemes for cmorph rainfall estimates in the zambezi river basin. *Hydrology and earth system sciences*, 23(7):2915–2938.

Guo, H., Chen, S., Bao, A., Hu, J., Gebregiorgis, A. S., Xue, X., and Zhang, X. (2015). Inter-comparison of high-resolution satellite precipitation products over central asia. *Remote Sensing*, 7(6):7181–7211.

Guttman, A. (1984). R-trees: A dynamic index structure for spatial searching. In *Proceedings of the 1984 ACM SIGMOD international conference on Management of data*, pages 47–57.

Habib, E., Haile, A. T., Sazib, N., Zhang, Y., and Rientjes, T. (2014). Effect of bias correction of satellite-rainfall estimates on runoff simulations at the source of the upper blue nile. *Remote Sensing*, 6(7):6688–6708.

Habib, E., Haile, A. T., Tian, Y., and Joyce, R. J. (01 Dec. 2012). Evaluation of the high-resolution cmorph satellite rainfall product using dense rain gauge observations and radar-based estimates. *Journal of Hydrometeorology*, 13(6):1784–1798.

Haile, A., Rientjes, T., Habib, E., Jetten, V., and Gebremichael, M. (2011). Rain event properties at the source of the blue nile river. *Hydrology and earth system sciences*, 15(3):1023–1034.

Han, L., Fu, S., Zhao, L., Zheng, Y., Wang, H., and Lin, Y. (2009). 3d convective storm identification, tracking, and forecasting—an enhanced TITAN algorithm. *Journal of Atmospheric and Oceanic Technology*, 26(4):719–732.

Hasan, M. M., Sharma, A., Johnson, F., Mariethoz, G., and Seed, A. (2014). Correcting bias in radar z–r relationships due to uncertainty in point rain gauge networks. *Journal of Hydrology*, 519:1668–1676.

Hashino, T., Bradley, A., and Schwartz, S. (2007). Evaluation of bias-correction methods for ensemble streamflow volume forecasts. *Hydrology and Earth System Sciences*, 11(2):939–950.

He, L., Ren, X., Gao, Q., Zhao, X., Yao, B., and Chao, Y. (2017). The connected-component labeling problem: A review of state-of-the-art algorithms. *Pattern Recognition*, 70:25–43.

Held, I. M. and Soden, B. J. (2006). Robust responses of the hydrological cycle to global warming. *Journal of climate*, 19(21):5686–5699.

Hempel, S., Frieler, K., Warszawski, L., Schewe, J., and Piontek, F. (2013). A trend-preserving bias correction–the isi-mip approach. *Earth System Dynamics*, 4(2):219–236.

Henderson, D. J. and Parmeter, C. F. (2012). Normal reference bandwidths for the general order, multivariate kernel density derivative estimator. *Statistics & Probability Letters*, 82(12):2198–2205.

Heumann, B. W. (2011). An object-based classification of mangroves using a hybrid decision tree—support vector machine approach. *Remote Sensing*, 3(11):2440–2460.

Hoffman, R. N., Liu, Z., Louis, J.-F., and Grassoti, C. (1995). Distortion representation of forecast errors. *Monthly Weather Review*, 123(9):2758–2770.

Hu, Z., Hu, Q., Zhang, C., Chen, X., and Li, Q. (2016). Evaluation of reanalysis, spatially interpolated and satellite remotely sensed precipitation data sets in central asia. *Journal of Geophysical Research: Atmospheres*, 121(10):5648–5663.

Huffman, G. J., Adler, R. F., Bolvin, D. T., and Nelkin, E. J. (2010). The TRMM multi-satellite precipitation analysis (TMPA). In *Satellite rainfall applications for surface hydrology*, pages 3–22. Springer.

Huffman, G. J., Bolvin, D. T., Nelkin, E. J., et al. (2015). Integrated multi-satellite retrievals for gpm (imerg) technical documentation. *NASA/GSFC Code*, 612(47):2019.

Huffman, G. J., Bolvin, D. T., Nelkin, E. J., Wolff, D. B., Adler, R. F., Gu, G., Hong, Y., Bowman, K. P., and Stocker, E. F. (2007). The TRMM multisatellite precipitation analysis (TMPA): Quasi-global, multiyear, combined-sensor precipitation estimates at fine scales. *Journal of Hydrometeorology*, 8(1):38–55.

Hyndman, R. L., Zhang, X., King, M. L., et al. (2004). Bandwidth selection for multivariate kernel density estimation using mcmc. In *Econometric Society 2004 Australasian Meetings*. Econometric Society.

IBGE (2017). *Monitoramento da cobertura e uso da terra : 2000 - 2010 - 2012 - 2014 : em grade territorial estatística*. Instituto Brasileiro de Geografia e Estatística.

Johnson, A. and Wang, X. (2012). Verification and calibration of neighborhood and object-based probabilistic precipitation forecasts from a multimodel convection-allowing ensemble. *Monthly weather review*, 140(9):3054–3077.

Johnson, A. E. and Hebert, M. (1999). Using spin images for efficient object recognition in cluttered 3d scenes. *IEEE Transactions on pattern analysis and machine intelligence*, 21(5):433–449.

Johnson, J. T., MacKeen, P. L., Witt, A., Mitchell, E. D. W., Stumpf, G. J., Eilts, M. D., and Thomas, K. W. (1998). The storm cell identification and tracking algorithm: An enhanced WSR-88d algorithm. *Wea. Forecasting*, 13(2):263–276.

Joy, C. (2012). The impact & management of floods & droughts in the lower mekong basin & the implications of possible climate change. *Vientiane, Laos*.

Joyce, R. J., Janowiak, J. E., Arkin, P. A., and Xie, P. (2004). Cmorph: A method that produces global precipitation estimates from passive microwave and infrared data at high spatial and temporal resolution. *Journal of hydrometeorology*, 5(3):487–503.

Kayastha, N. (2014). *Refining the committee approach and uncertainty prediction in hydrological modelling*. TU Delft, Delft University of Technology.

Keil, C. and Craig, G. C. (2007). A displacement-based error measure applied in a regional ensemble forecasting system. *Monthly Weather Review*, 135(9):3248–3259.

Keil, C. and Craig, G. C. (2009a). A displacement and amplitude score employing an optical flow technique. *Weather and Forecasting*, 24(5):1297–1308.

Keil, C. and Craig, G. C. (2009b). A displacement and amplitude score employing an optical flow technique. *Weather and Forecasting*, 24(5):1297–1308.

Kidd, C., Bauer, P., Turk, J., Huffman, G. J., Joyce, R., Hsu, K. L., and Braithwaite, D. (2012). Intercomparison of high-resolution precipitation products over northwest europe. *Journal of Hydrometeorology*, 13(1):67–83.

Kidd, C., Heinemann, T., Levizzani, V., et al. (2008). International precipitation working group (ipwg): Inter-comparison of regional precipitation products.

Kidd, C., Levizzani, V., Turk, J., and Ferraro, R. (2009). Satellite precipitation measurements for water resource monitoring 1. *JAWRA Journal of the American Water Resources Association*, 45(3):567–579.

Kidd, C., Shige, S., Vila, D., Tarnavsky, E., Yamamoto, M. K., Maggioni, V., and Maseko, B. (2020). The ipwg satellite precipitation validation effort. In *Satellite Precipitation Measurement*, pages 453–470. Springer.

Kim, M., Yhang, Y.-B., and Lim, C.-M. (01 Feb. 2019). Gaussian copula method for bias correction of daily precipitation generated by a dynamical model. *Journal of Applied Meteorology and Climatology*, 58(2):269–289.

Kimani, M. W., Hoedjes, J. C. B., and Su, Z. (2017). An assessment of satellite-derived rainfall products relative to ground observations over east africa. *Remote Sensing*, 9(5).

Kirstetter, P.-E., Hong, Y., Gourley, J., Chen, S., Flamig, Z., Zhang, J., Schwaller, M., Petersen, W., and Amitai, E. (2012). Toward a framework for systematic error modeling of spaceborne precipitation radar with noaa/nssl ground radar–based national mosaic qpe. *Journal of Hydrometeorology*, 13(4):1285–1300.

Kişi, Ö. (2007). Streamflow forecasting using different artificial neural network algorithms. *Journal of Hydrologic Engineering*, 12(5):532–539.

Kolluru, V., Kolluru, S., Wagle, N., and Acharya, T. D. (2020). Secondary precipitation estimate merging using machine learning: development and evaluation over krishna river basin, india. *Remote Sensing*, 12(18):3013.

Kumar, A., Ramsankaran, R., Brocca, L., and Munoz-Arriola, F. (2019). A machine learning approach for improving near-real-time satellite-based rainfall estimates by integrating soil moisture. *Remote Sensing*, 11(19):2221.

Laverde-Barajas, M., Corzo, G., Bhattacharya, B., Uijlenhoet, R., and Solomatine, D. P. (2019). Spatiotemporal analysis of extreme rainfall events using an object-based approach. In Corzo, G. and Varouchakis, E. A., editors, *Spatiotemporal Analysis of Extreme Hydrological Events*, pages 95 – 112. Elsevier.

Laverde-Barajas, M., Corzo, G. A., Poortinga, A., Chishtie, F., Meechaiya, C., Jayasinghe, S., Towashiraporn, P., Markert, A., Saah, D., Son, L. H., et al. (2020a). St-corabico: A spatiotemporal object-based bias correction method for storm prediction detected by satellite. *Remote Sensing*, 12(21):3538.

Laverde-Barajas, M., Corzo Perez, G., Dalfré Filho, J., and Solomatine, D. (2018). Assessing the performance of near real-time rainfall products to represent spatiotemporal characteristics of extreme events: case study of a subtropical catchment in south-eastern brazil. *International Journal of Remote Sensing*, 39(21):7568–7586.

Laverde-Barajas, M., Corzo Perez, G., and Solomatine, D. (2017). Pattern recognition techniques in estimation of rainfall extreme events spatiotemporal characteristic: case study of a subtropical catchment in south-eastern brazil. In *Geophysical Research Abstracts*, volume 19, page 1564.

Laverde-Barajas, M., Perez, G. C., Chishtie, F., Poortinga, A., Uijlenhoet, R., and Solomatine, D. (2020b). Decomposing satellite-based rainfall errors in flood estimation: Hydrological responses using a spatiotemporal object-based verification method. *Journal of Hydrology*, 591:125554.

Le, X.-H., Lee, G., Jung, K., An, H.-u., Lee, S., and Jung, Y. (2020). Application of convolutional neural network for spatiotemporal bias correction of daily satellite-based precipitation. *Remote Sensing*, 12(17):2731.

Le Coz, C., Heemink, A., Verlaan, M., ten Veldhuis, M.-c., and van de Giesen, N. (2019). Correcting Position Error in Precipitation Data Using Image Morphing. *Remote Sensing*, 11(21):2557.

Lenderink, G., Buishand, A., and van Deursen, W. (2007). Estimates of future discharges of the river rhine using two scenario methodologies: direct versus delta approach. *Hydrology and Earth System Sciences*, 11(3):1145–1159.

Li, J. and Heap, A. D. (2014). Spatial interpolation methods applied in the environmental sciences: A review. *Environmental Modelling & Software*, 53:173–189.

Li, J., Hsu, K., AghaKouchak, A., and Sorooshian, S. (2015a). An object-based approach for verification of precipitation estimation. *International Journal of Remote Sensing*, 36(2):513–529.

Li, J., Hsu, K.-L., AghaKouchak, A., and Sorooshian, S. (2016). Object-based assessment of satellite precipitation products. *Remote Sensing*, 8(7):547.

Li, L., Hong, Y., Wang, J., Adler, R. F., Policelli, F. S., Habib, S., Irwn, D., Korme, T., and Okello, L. (2009). Evaluation of the real-time trmm-based multi-satellite precipitation analysis for an operational flood prediction system in nzoia basin, lake victoria, africa. *Natural hazards*, 50(1):109–123.

Li, L., Li, Y., and Li, Z. (2020a). Object-based tracking of precipitation systems in western canada: the importance of temporal resolution of source data. *Climate Dynamics*, 55(9):2421–2437.

Li, S. (2009). Random knn modeling and variable selection for high dimensional data.

Li, X., Cheng, X., Chen, W., Chen, G., and Liu, S. (2015b). Identification of forested landslides using lidar data, object-based image analysis, and machine learning algorithms. *Remote sensing*, 7(8):9705–9726.

Li, X., Zhang, Q., and Xu, C.-Y. (2014). Assessing the performance of satellite-based precipitation products and its dependence on topography over poyang lake basin. *Theoretical and applied climatology*, 115(3):713–729.

Li, Z., Wright, D. B., Zhang, S. Q., Kirschbaum, D. B., and Hartke, S. H. (2020b). Object-based comparison of data-driven and physics-driven satellite estimates of extreme rainfall. *Journal of Hydrometeorology*, 21(12):2759 – 2776.

Liu, H., Sorooshian, S., and Gao, X. (2015). Assessment of the spatial and seasonal variation of the error–intensity relationship in satellite-based precipitation measurements using an adaptive parametric model. *Journal of Hydrometeorology*, 16(4):1700–1716.

Liu, W., Li, X., and Rahn, D. A. (2016). Storm event representation and analysis based on a directed spatiotemporal graph model. *International Journal of Geographical Information Science*, 30(5):948–969.

Ly, S., Sohier, C., Charles, C., and Degré, A. (2012). Effect of raingage density, position and interpolation on rainfall-discharge modelling. In *Geophysical Research Abstracts*, page 2592. European Geophysical Society.

Maggioni, V. and Massari, C. (2018). On the performance of satellite precipitation products in riverine flood modeling: A review. *Journal of Hydrology*, 558:214–224.

Maggioni, V., Meyers, P. C., and Robinson, M. D. (2016a). A review of merged high-resolution satellite precipitation product accuracy during the tropical rainfall measuring mission (TRMM) era. *Journal of Hydrometeorology*, 17(4):1101–1117.

Maggioni, V., Meyers, P. C., and Robinson, M. D. (2016b). A review of merged high-resolution satellite precipitation product accuracy during the tropical rainfall measuring mission (trmm) era. *Journal of Hydrometeorology*, 17(4):1101–1117.

Marcuzzo, F. (2020). Bacia hidrográfica do rio tietê: precipitação pluviométrica espacializada. *Geographia Meridionalis*, 5(3):243–266.

Marra, F., Morin, E., Peleg, N., Mei, Y., and Anagnostou, E. N. (2017). Intensity–duration–frequency curves from remote sensing rainfall estimates: comparing satellite and weather radar over the eastern mediterranean. *Hydrology and Earth System Sciences*, 21(5):2389–2404.

Marsigli, C., Montani, A., and Paccagnella, T. (2006). Verification of the cosmoleps new suite in terms of precipitation distribution. *COSMO Newsletter*, 6:134–141.

Masson-Delmotte, V., Péan, P. Z. A. P. S. L. C. C., Berger, S., Caud, N., Chen, Y., Goldfarb, L., Gomis, M. I., Huang, M., Leitzell, K., Lonnoy, E., Matthews, J. B., Maycock, T. K., Waterfield, T., Yelekçi, O., Yu, R., and Zhou, B. (2021). Climate change 2021: The physical science basis. *Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change*.

Mazzoleni, M., Brandimarte, L., and Amaranto, A. (2019). Evaluating precipitation datasets for large-scale distributed hydrological modelling. *Journal of Hydrology*, 578:124076.

Mehran, A. and AghaKouchak, A. (2014). Capabilities of satellite precipitation datasets to estimate heavy precipitation rates at different temporal accumulations. *Hydrological Processes*, 28(4):2262–2270.

Mei, Y., Anagnostou, E. N., Nikolopoulos, E. I., and Borga, M. (2014). Error analysis of satellite precipitation products in mountainous basins. *Journal of Hydrometeorology*, 15(5):1778–1793.

Mei, Y., Anagnostou, E. N., Shen, X., and Nikolopoulos, E. I. (2017a). Decomposing the satellite precipitation error propagation through the rainfall-runoff processes. *Advances in Water Resources*, 109:253–266.

Mei, Y., Nikolopoulos, E. I., Anagnostou, E. N., and Borga, M. (2016a). Evaluating satellite precipitation error propagation in runoff simulations of mountainous basins. *Journal of Hydrometeorology*, 17(5):1407–1423.

Mei, Y., Nikolopoulos, E. I., Anagnostou, E. N., Zoccatelli, D., and Borga, M. (2016b). Error analysis of satellite precipitation-driven modeling of flood events in complex alpine terrain. *Remote Sensing*, 8(4):293.

Mei, Y., Shen, X., and Anagnostou, E. N. (2017b). A synthesis of space-time variability in multicomponent flood response. *Hydrology & Earth System Sciences*, 21(5).

Miao, C., Ashouri, H., Hsu, K.-L., Sorooshian, S., and Duan, Q. (2015). Evaluation of the persiann-cdr daily rainfall estimates in capturing the behavior of extreme precipitation events over china. *Journal of Hydrometeorology*, 16(3):1387–1396.

Milano, M., Reynard, E., Muniz-Miranda, G., and Guerrin, J. (2018). Water supply basins of são paulo metropolitan region: hydro-climatic characteristics of the 2013–2015 water crisis. *Water*, 10(11):1517.

Mittermaier, M. and Bullock, R. (2013). Using mode to explore the spatial and temporal characteristics of cloud cover forecasts from high-resolution nwp models. *Meteorological Applications*, 20(2):187–196.

Mittermaier, M. P. (2006). Using an intensity-scale technique to assess the added benefit of high-resolution model precipitation forecasts. *Atmospheric Science Letters*, 7(2):36–42.

Moazami, S., Golian, S., Kavianpour, M. R., and Hong, Y. (2013). Comparison of persiann and v7 trmm multi-satellite precipitation analysis (tmpa) products with rain gauge data over iran. *International journal of remote sensing*, 34(22):8156–8171.

Moghim, S. and Bras, R. L. (2017). Bias correction of climate modeled temperature and precipitation using artificial neural networks. *Journal of Hydrometeorology*, 18(7):1867–1884.

Molini, L., Parodi, A., Rebora, N., and Craig, G. (2011). Classifying severe rainfall events over italy by hydrometeorological and dynamical criteria. *Quarterly Journal of the Royal Meteorological Society*, 137(654):148–154.

Molini, L., Parodi, A., and Siccardi, F. (2009). Dealing with uncertainty: an analysis of the severe weather events over italy in 2006. *Nat. Hazards Earth Syst. Sci*, 9:1775–1786.

Moon, S.-H., Kim, Y.-H., Lee, Y. H., and Moon, B.-R. (2019). Application of machine learning to an early warning system for very short-term heavy rainfall. *Journal of Hydrology*, 568:1042–1054.

NASA (2015). Meet the members of nasa's gpm constellation).

NASA, G. (2014). The global precipitation measurement mission (gpm).

Nash, J. E. and Sutcliffe, J. V. (1970). River flow forecasting through conceptual models part i—a discussion of principles. *Journal of hydrology*, 10(3):282–290.

Nguyen, P., Ombadi, M., Gorooh, V. A., Shearer, E. J., Sadeghi, M., Sorooshian, S., Hsu, K., Bolvin, D., and Ralph, M. F. (01 Dec. 2020). Persiann dynamic infrared–rain rate (pdir-now): A near-real-time, quasi-global satellite precipitation dataset. *Journal of Hydrometeorology*, 21(12):2893–2906.

Nikolopoulos, E. I., Anagnostou, E. N., and Borga, M. (2013). Using high-resolution satellite rainfall products to simulate a major flash flood event in northern italy. *Journal of Hydrometeorology*, 14(1):171–185.

NOAA, N. O. a. A. A. (2011). State of the climate: Global hazards for january 2011. Report, NOAA.

Oliveira, R., Maggioni, V., Vila, D., and Morales, C. (2016). Characteristics and diurnal cycle of gpm rainfall estimates over the central amazon region. *Remote Sensing*, 8(7):544.

Otsuka, S., Kotsuki, S., Ohhigashi, M., and Miyoshi, T. (2019). Gsmap riken nowcast: Global precipitation nowcasting with data assimilation. *Journal of the Meteorological Society of Japan. Ser. II*.

Paiva, R. C., Durand, M. T., and Hossain, F. (2015). Spatiotemporal interpolation of discharge across a river network by using synthetic swot satellite data. *Water Resources Research*, 51(1):430–449.

Pan, M., Li, H., and Wood, E. (2010). Assessing the skill of satellite-based precipitation estimates in hydrologic applications. *Water Resources Research*, 46(9).

Peleg, N., Marra, F., Fatichi, S., Paschalis, A., Molnar, P., and Burlando, P. (2018). Spatial variability of extreme rainfall at radar subpixel scale. *Journal of Hydrology*, 556:922–933.

Peña-Arancibia, J. L., van Dijk, A. I., Renzullo, L. J., and Mulligan, M. (2013). Evaluation of precipitation estimation accuracy in reanalyses, satellite products, and an ensemble method for regions in australia and south and east asia. *Journal of Hydrometeorology*, 14(4):1323–1333.

Pereira, O., Torres, E., Garcés, Y., and Rodríguez, R. (2017). Edge detection based on kernel density estimation. In *Proceedings of the International Conference on Image Processing, Computer Vision, and Pattern Recognition (IPCV)*, pages 123–128. The Steering Committee of The World Congress in Computer Science, Computer ….

Perez, R. E., Jansen, P. W., and Martins, J. R. (2012). pyopt: a python-based object-oriented framework for nonlinear constrained optimization. *Structural and Multidisciplinary Optimization*, 45(1):101–118.

Phongsapan, K., Chishtie, F., Poortinga, A., Bhandari, B., Meechaiya, C., Kunlamai, T., Aung, K. S., Saah, D., Anderson, E., Markert, K., et al. (2019). Operational flood risk index mapping for disaster risk reduction using earth observations and cloud computing technologies: a case study on myanmar. *Frontiers in Environmental Science*, 7:191.

Piani, C., Haerter, J., and Coppola, E. (2010). Statistical bias correction for daily precipitation in regional climate models over europe. *Theoretical and Applied Climatology*, 99(1-2):187–192.

Pinto, J. O., Grim, J. A., and Steiner, M. (2015). Assessment of the high-resolution rapid refresh model's ability to predict mesoscale convective systems using object-based evaluation. *Weather and Forecasting*, 30(4):892–913.

Pompermayer, R. d. S. et al. (2003). Aplicação da análise multicritério em gestão de recursos hídricos: simulação para as bacias dos rios piracicaba, capivari e jundiaí.

Poortinga, A., Bastiaanssen, W., Simons, G., Saah, D., Senay, G., Fenn, M., Bean, B., and Kadyszewski, J. (2017). A self-calibrating runoff and streamflow remote sensing model for ungauged basins using open-access earth observation data. *Remote Sensing*, 9(1):86.

Prat, O. P. and Barros, A. P. (2010). Assessing satellite-based precipitation estimates in the southern appalachian mountains using rain gauges and trmm pr. *Advances in Geosciences*, 25:143–153.

Prat, O. P. and Nelson, B. R. (2020). Satellite precipitation measurement and extreme rainfall. In *Satellite Precipitation Measurement*, pages 761–790. Springer.

Prein, A. F., Liu, C., Ikeda, K., Bullock, R., Rasmussen, R. M., Holland, G. J., and Clark, M. (2017a). Simulating north american mesoscale convective systems with a convection-permitting climate model. *Climate Dynamics*, pages 1–16.

Prein, A. F., Liu, C., Ikeda, K., Trier, S. B., Rasmussen, R. M., Holland, G. J., and Clark, M. P. (2017b). Increased rainfall volume from future convective storms in the us. *Nature Climate Change*, 7(12):880–884.

Qiao, L., Hong, Y., Chen, S., Zou, C. B., Gourley, J. J., and Yong, B. (2014). Performance assessment of the successive version 6 and version 7 tmpa products over the climate-transitional zone in the southern great plains, usa. *Journal of Hydrology*, 513:446–456.

Revilla-Romero, B., Hirpa, F. A., Pozo, J. T.-d., Salamon, P., Brakenridge, R., Pappenberger, F., and De Groeve, T. (2015). On the use of global flood forecasts and satellite-derived inundation maps for flood monitoring in data-sparse regions. *Remote Sensing*, 7(11):15702–15728.

Ringard, J., Becker, M., Seyler, F., and Linguet, L. (2015). Temporal and spatial assessment of four satellite rainfall estimates over french guiana and north brazil. *Remote Sensing*, 7(12):16441–16459.

Roberts, N. (2005). An investigation of the ability of a storm scale configuration of the met office nwp model to predict floodproducing rainfall. *UK Met Office Tech. Rep*, 455:80.

Sapiano, M. and Arkin, P. (2009). An intercomparison and validation of high-resolution satellite precipitation estimates with 3-hourly gauge data. *Journal of Hydrometeorology*, 10(1):149–166.

Saulnier, G.-M. and Le Lay, M. (2009). Sensitivity of flash-flood simulations on the volume, the intensity, and the localization of rainfall in the cévennes-vivarais region (france). *Water Resources Research*, 45(10).

Schellekens, J. (2018). wflow documentation.

Schneider, U., Fuchs, T., Meyer-Christoffer, A., and Rudolf, B. (2008). Global precipitation analysis products of the gpcc. *Global Precipitation Climatology Centre (GPCC), DWD, Internet Publikation*, 112.

Scofield, R. A. and Kuligowski, R. J. (2003). Status and outlook of operational satellite precipitation algorithms for extreme-precipitation events. *Weather and Forecasting*, 18(6):1037–1051.

SEDEC (2013). Secretaria nacional de proteção e defesa civil, sistema integrado de informações sobre desastres - s2id. accessed February,2019.

Sedgewick, R. (1998). *Algorithms in C, Parts 1-4: Fundamentals, Data Structures, Sorting, Searching*. Addison-Wesley Professional, 3 edition edition.

Sellars, S., Nguyen, P., Chu, W., Gao, X., Hsu, K.-l., and Sorooshian, S. (2013). Computational earth science: Big data transformed into insight. *Eos, Transactions American Geophysical Union*, 94(32):277–278.

Sellars, S. L., Gao, X., and Sorooshian, S. (2015). An object-oriented approach to investigate impacts of climate oscillations on precipitation: A western united states case study. *Journal of Hydrometeorology*, 16(2):830–842.

Serrat-Capdevila, A., Valdes, J. B., and Stakhiv, E. Z. (2014). Water management applications for satellite precipitation products: Synthesis and recommendations. *JAWRA Journal of the American Water Resources Association*, 50(2):509–525.

Shukla, S., Landsfeld, M., Anthony, M., Budde, M., Husak, G., Rowland, J., and Funk, C. (2020). Enhancing the application of earth observations for improved environmental decision-making using the early warning explorer (ewx). *Frontiers in Climate*, 2:34.

Sillmann, J., Thorarinsdottir, T., Keenlyside, N., Schaller, N., Alexander, L. V., Hegerl, G., Seneviratne, S. I., Vautard, R., Zhang, X., and Zwiers, F. W. (2017). Understanding, modeling and predicting weather and climate extremes: Challenges and opportunities. *Weather and climate extremes*, 18:65–74.

Sinclair, S. and Pegram, G. (2005). Combining radar and rain gauge rainfall estimates using conditional merging. *Atmospheric Science Letters*, 6(1):19–22.

Skok, G., Tribbia, J., Rakovec, J., and Brown, B. (2009). Object-based analysis of satellite-derived precipitation systems over the low-and midlatitude pacific ocean. *Monthly weather review*, 137(10):3196–3218.

Software, G. S. (2012). Surfer version 11: Reference manual.

Sondergaard, L., Luo, X., Jithitikulchai, T., Arin, T., Poggi, C., Lathapipat, D., et al. (2016). Getting back on track: reviving growth and securing prosperity for all. *Bangkok: World Bank Thailand*, page 158.

Sorooshian, S., Hsu, K.-L., Gao, X., Gupta, H. V., Imam, B., and Braithwaite, D. (2000). Evaluation of persiann system satellite-based estimates of tropical rainfall. *Bulletin of the American Meteorological Society*, 81(9):2035–2046.

Sprissler, T. (2011). Flood risk brazil: Prevention, adaptation and insurance.

Steiner, M., Houze Jr, R. A., and Yuter, S. E. (1995). Climatological characterization of three-dimensional storm structure from operational radar and rain gauge data. *Journal of Applied Meteorology*, 34(9):1978–2007.

Stensrud, D. J. and Wandishin, M. S. (2000). The correspondence ratio in forecast evaluation. *Weather and forecasting*, 15(5):593–602.

Stephenson, D. B., Diaz, H., and Murnane, R. (2008). Definition, diagnosis, and origin of extreme weather and climate events. *Climate extremes and society*, 340:11–23.

Stisen, S. and Sandholt, I. (2010). Evaluation of remote-sensing-based rainfall products through predictive capability in hydrological runoff modelling. *Hydrological Processes: An International Journal*, 24(7):879–891.

Su, F., Hong, Y., and Lettenmaier, D. P. (2008). Evaluation of trmm multisatellite precipitation analysis (tmpa) and its utility in hydrologic prediction in the la plata basin. *Journal of Hydrometeorology*, 9(4):622–640.

Sun, Q., Miao, C., Duan, Q., Ashouri, H., Sorooshian, S., and Hsu, K.-L. (2018). A review of global precipitation data sets: Data sources, estimation, and intercomparisons. *Reviews of Geophysics*, 56(1):79–107.

Syed, T. H., Lakshmi, V., Paleologos, E., Lohmann, D., Mitchell, K., and Famiglietti, J. S. (2004). Analysis of process controls in land surface hydrological cycle over the continental united states. *Journal of Geophysical Research: Atmospheres*, 109(D22).

Szeliski, R. (2010). *Computer vision: algorithms and applications*. Springer Science & Business Media.

Tan, J., Huffman, G. J., Bolvin, D. T., and Nelkin, E. J. (2019). Imerg v06: Changes to the morphing algorithm. *Journal of Atmospheric and Oceanic Technology*, 36(12):2471–2482.

Tan, M. L. and Santo, H. (2018). Comparison of gpm imerg, tmpa 3b42 and persiann-cdr satellite precipitation products over malaysia. *Atmospheric Research*, 202:63–76.

Tang, G., Ma, Y., Long, D., Zhong, L., and Hong, Y. (2016). Evaluation of gpm day-1 imerg and tmpa version-7 legacy products over mainland china at multiple spatiotemporal scales. *Journal of hydrology*, 533:152–167.

Tangune, B. F. and Escobedo, J. F. (2018). Reference evapotranspiration in so paulo state: Empirical methods and machine learning techniques. *International Journal of Water Resources and Environmental Engineering*, 10(4):33–44.

Tao, Y., Gao, X., Hsu, K., Sorooshian, S., and Ihler, A. (2016). A deep neural network modeling framework to reduce bias in satellite precipitation products. *Journal of Hydrometeorology*, 17(3):931–945.

Teng, J., Potter, N. J., Chiew, F. H. S., Zhang, L., Wang, B., Vaze, J., and Evans, J. P. (2015). How does bias correction of regional climate model precipitation affect modelled runoff? *Hydrology and Earth System Sciences*, 19(2):711–728.

Tesfagiorgis, K., Mahani, S., Krakauer, N., and Khanbilvardi, R. (2011). Bias correction of satellite rainfall estimates using a radar-gauge product–a case study in oklahoma (usa). *Hydrology and Earth System Sciences*, 15(8):2631–2647.

Teutschbein, C. and Seibert, J. (2012). Bias correction of regional climate model simulations for hydrological climate-change impact studies: Review and evaluation of different methods. *Journal of hydrology*, 456:12–29.

Themeßl, M. J., Gobiet, A., and Heinrich, G. (2012). Empirical-statistical downscaling and error correction of regional climate models and its impact on the climate change signal. *Climatic Change*, 112(2):449–468.

Thiemig, V., Rojas, R., Zambrano-Bigiarini, M., Levizzani, V., and De Roo, A. (2012). Validation of satellite-based precipitation products over sparsely gauged african river basins. *Journal of Hydrometeorology*, 13(6):1760–1783.

Tong, Y., Gao, X., Han, Z., Xu, Y., Xu, Y., and Giorgi, F. (2020). Bias correction of temperature and precipitation over china for rcm simulations using the qm and qdm methods. *Climate Dynamics*, pages 1–19.

Uijlenhoet, R. and Berne, A. (2008). Stochastic simulation experiment to assess radar rainfall retrieval uncertainties associated with attenuation and its correction. *Hydrology and Earth System Sciences*, 12(2):587–601.

Van Zyl, J. J. (2001). The shuttle radar topography mission (srtm): a breakthrough in remote sensing of topography. *Acta Astronautica*, 48(5-12):559–565.

Vergara, H., Hong, Y., Gourley, J. J., Anagnostou, E. N., Maggioni, V., Stampoulis, D., and Kirstetter, P.-E. (2014). Effects of resolution of satellite-based rainfall estimates on hydrologic modeling skill at different scales. *Journal of Hydrometeorology*, 15(2):593–613.

Vertessy, R. A. and Elsenbeer, H. (1999). Distributed modeling of storm flow generation in an amazonian rain forest catchment: Effects of model parameterization. *Water Resources Research*, 35(7):2173–2187.

Vicente, G. A., Scofield, R. A., and Menzel, W. P. (1998). The operational GOES infrared rainfall estimation technique. *Bulletin of the American Meteorological Society*, 79(9):1883–1898.

Viglione, A., Chirico, G. B., Komma, J., Woods, R., Borga, M., and Blöschl, G. (2010). Quantifying space-time dynamics of flood event types. *Journal of Hydrology*, 394(1-2):213–229.

Vila, D. A., De Goncalves, L. G. G., Toll, D. L., and Rozante, J. R. (2009). Statistical evaluation of combined daily gauge observations and rainfall satellite estimates over continental south america. *Journal of Hydrometeorology*, 10(2):533–543.

Vrac, M. and Friederichs, P. (01 Jan. 2015). Multivariate—intervariable, spatial, and temporal—bias correction. *Journal of Climate*, 28(1):218–237.

Wackernagel, H. (2013). *Multivariate Geostatistics: An Introduction with Applications*. Springer Science & Business Media. Google-Books-ID: vM3sCAAAQBAJ.

Wang, L., Lu, J., Li, X., Huan, Z., Liang, J., and Chen, S. (2017). Learning arbitrary-shape object detector from bounding-box annotation by searching region-graph. *Pattern Recognition Letters*, 87:171–176.

Wang, N., Zhang, D., Chang, H., and Li, H. (2020). Deep learning of subsurface flow via theory-guided neural network. *Journal of Hydrology*, 584:124700.

Wang, Z., Fan, C., and Xian, M. (2021). Application and evaluation of a deep learning architecture to urban tree canopy mapping. *Remote Sensing*, 13(9):1749.

Wernli, H., Paulat, M., Hagen, M., and Frei, C. (2008). Sal—a novel quality measure for the verification of quantitative precipitation forecasts. *Monthly Weather Review*, 136(11):4470–4487.

Wesselung, C. G., Karssenberg, D.-J., Burrough, P. A., and van Deursen, W. P. (1996). Integrating dynamic environmental models in gis: the development of a dynamic modelling language. *Transactions in GIS*, 1(1):40–48.

Wilks, D. S. (2011). *Statistical methods in the atmospheric sciences*, volume 100. Academic press. Google-Books-ID: IJuCVtQ0ySIC.

Yang, R., Zhang, W.-K., Gui, S., Tao, Y., and Cao, J. (2019). Rainy season precipitation variation in the mekong river basin and its relationship to the indian and east asian summer monsoons. *Climate dynamics*, 52(9-10):5691–5708.

Yang, Y. and Luo, Y. (2014). Using the back propagation neural network approach to bias correct tmpa data in the arid region of northwest china. *Journal of Hydrometeorology*, 15(1):459–473.

Yilmaz, K. K., Hogue, T. S., Hsu, K.-L., Sorooshian, S., Gupta, H. V., and Wagener, T. (2005). Intercomparison of rain gauge, radar, and satellite-based precipitation estimates with emphasis on hydrologic forecasting. *Journal of Hydrometeorology*, 6(4):497–517.

Yu, B., Liu, H., Wu, J., Hu, Y., and Zhang, L. (2010). Automated derivation of urban building density information using airborne LiDAR data and object-based method. *Landscape and Urban Planning*, 98(3):210–219.

Yuan, Q., Thorarinsdottir, T. L., Beldring, S., Wong, W. K., Huang, S., and Xu, C.-Y. (2019). New approach for bias correction and stochastic downscaling of future projections for daily mean temperatures to a high-resolution grid. *Journal of Applied Meteorology and Climatology*, 58(12):2617–2632.

Zarei, M., Najarchi, M., and Mastouri, R. (2021). Bias correction of global ensemble precipitation forecasts by random forest method. *Earth Science Informatics*, pages 1–13.

Zeng, J. and Zhang, Q. (2020). The trends in land surface heat fluxes over global monsoon domains and their responses to monsoon and precipitation. *Scientific reports*, 10(1):1–15.

Zhang, C., Sargent, I., Pan, X., Li, H., Gardiner, A., Hare, J., and Atkinson, P. M. (2018). An object-based convolutional neural network (ocnn) for urban land use classification. *Remote sensing of environment*, 216:57–70.

Zhang, X., King, M. L., and Hyndman, R. J. (2006). A bayesian approach to bandwidth selection for multivariate kernel density estimation. *Computational Statistics & Data Analysis*, 50(11):3009–3031.

Zhao, G., Pang, B., Xu, Z., Peng, D., and Xu, L. (2019). Assessment of urban flood susceptibility using semi-supervised machine learning model. *Science of The Total Environment*, 659:940–949.

Zhao, T., Bennett, J. C., Wang, Q., Schepen, A., Wood, A. W., Robertson, D. E., and Ramos, M.-H. (2017). How suitable is quantile mapping for postprocessing gcm precipitation forecasts? *Journal of Climate*, 30(9):3185–3196.

Zhou, J. and Lau, K. (1998). Does a monsoon climate exist over south america? *Journal of climate*, 11(5):1020–1040.

Zivkovic, Z. and Van Der Heijden, F. (2006). Efficient adaptive density estimation per image pixel for the task of background subtraction. *Pattern recognition letters*, 27(7):773–780.

Zoccatelli, D., Borga, M., Viglione, A., Chirico, G., and Blöschl, G. (2011). Spatial moments of catchment rainfall: rainfall spatial organisation, basin morphology, and flood response. *Hydrology and Earth System Sciences*, 15(12):3767.

Netherlands Research School for the
Socio-Economic and Natural Sciences of the Environment

D I P L O M A

for specialised PhD training

The Netherlands research school for the
Socio-Economic and Natural Sciences of the Environment
(SENSE) declares that

Miguel Angel Laverde Barajas

born on 13th September 1985 Bogota, Colombia

has successfully fulfilled all requirements of the
educational PhD programme of SENSE.

Delft, 11 April 2022

Chair of the SENSE board

Prof. dr. Martin Wassen

The SENSE Director

Prof. Philipp Pattberg

The SENSE Research School has been accredited by the Royal Netherlands Academy of Arts and Sciences (KNAW)

KONINKLIJKE NEDERLANDSE
AKADEMIE VAN WETENSCHAPPEN

The SENSE Research School declares that **Miguel Angel Laverde Barajas** has successfully fulfilled all requirements of the educational PhD programme of SENSE with a work load of 49.7 EC, including the following activities:

SENSE PhD Courses

- Environmental research in context (2015)
- A2 project: 'Organizing Symposium: Remote Sensing in Hydrological Sciences (2017)
- Masterclass Git, GitHub and Markdown in an R-environment (2018)

Selection of Other PhD and Advanced MSc Courses

- Data-driven modelling and real-time control of water systems, UNESCO-IHE Delft (2015)
- Where there is little data: How to estimate design variables in poorly gauged basins, UNESCO-IHE Delft (2016)
- Geostatistics for Water Resources and Environmental Science, UNESCO-IHE Delft (2016)
- Artificial Intelligence for Earth System Science Summer School, NCAR/UCAR (2020)
- High-performance computing helping to solve water-related problems, IHE Delft (2015)

Selection of External training at a foreign research institute

- Introduction to 'Synthetic Aperture Radar', NASA 'Applied Remote Sensing Training Program ARSET', (2017-2018)
- SAR for Water and Disaster Applications NASA Asian Disaster Preparedness Center, Bangkok Thailand (2019)
- SWOT Early Adopter Virtual Hackathon, NASA CNES (2020)

Selection of Management and Didactic Skills Training

- Supervising MSc student with thesis (2021)
- Teaching in BSc/MSc/PhD course 'Error analysis in geospatial sciences: Geostatistics for Water Resources and Environmental Science (2017-2021)

Selection of Oral Presentations

- The impact of satellite-based error sources for extreme rainfall events on the hydrological response. EGU General Assembly Conference, 7–12 April 2019, Vienna, Austria
- Advances in short and middle-term riverine flood forecasting over the Lower Mekong Basin: Implementation of new bias-corrected near-real-time and forecast rainfall information into the Mekong River Commission Flood Early Warning System. Water Security and Climate Change Conference. 01-04 March 2021, Hanoi, Vietnam

SENSE coordinator PhD education

Dr. ir. Peter Vermeulen

Satellite Precipitation Products (SPP) have been revolutionary in water resources management and flood-related disaster response. However, estimating extreme rainfall is subject to multiple systematic and aleatory errors that need to be corrected. This dissertation addresses errors in satellite data to estimate extreme rainfall events in space and time beyond the pixel. The Spatiotemporal Contiguous Object-based Rainfall Analysis method (ST-CORA) is developed to analyse errors in SPP for rainstorm estimations based on their main physical features in space and time (volume, intensity, duration, extension, orientation, speed, among others). Using ST-CORA, systematic errors due to volume and displacement in space and time are corrected in a novel bias-corrected method called ST-CORAbico. Case studies in two monsoonal areas in South America and Southeast Asia have been used to analyse the hydrological response of systematic errors in flood predictions and evaluate error reduction in non-operational and operational bias correction applications. Finally, the dissertation describes further implementations of ST-CORA in developing an operational system for rainstorm monitoring called Rainstorm tracker. This web-based platform is designed to monitor and alert decision-makers about the severity of rainstorm events over the Lower Mekong basin in near-real and real-time.

This book is printed on paper from sustainably managed forests and controlled sources

9 789073 445406