Synchronization and interdependence between the cycles of Colombia's hydroclimatology and El Niño-Southern Oscillation

Hernán Darío Salas Parra

Universidad Nacional de Colombia
Facultad de Minas, Departamento de Geociencias y Medio Ambiente
Medellín, Colombia
2020

Synchronization and interdependence between the cycles of Colombia's hydroclimatology and El Niño-Southern Oscillation

Hernán Darío Salas Parra

Thesis submitted as partial requirement for the degree of:
Doctor en Ingeniería - Recursos Hidráulicos

Advisors:
Germán Poveda, Ph.D. and Oscar J. Mesa, Ph.D.

Research area:
Hydroclimatology
Universidad Nacional de Colombia
Facultad de Minas, Departamento de Geociencias y Medio Ambiente
Medellín, Colombia
2020

I dedicate this work to my family: My wife Marcela, my daughter Julieta, and my son Juan Miguel. Their love and patience did this possible.

Also, I dedicate this to Susann Reininger, whom my family and I deeply miss.

Aknowledgments

I express my deep gratitude to Professors Germán Poveda and Óscar J. Mesa for their patience, advice, motivation, and friendship. Without their deep questions and suggestions, this work would not have been possible.

Thanks to Professor Jürgen Kurths at Potsdam Institute for Climate Impact Research (PIK) in Germany for his kind hospitality and support during my stay there.

Thanks to Susann and Klaus Reininger, who did wonderful our stay in Caputh - Schwielowsee (Potsdam, Germany). My family and I have deep gratitude and appreciation.

Thanks to Niklas Boers and Norbert Marwan at PIK for their contribution to this research. I learned a lot in my talks with them.

Thanks to Professors: Francina Dominguez, Jorge M. Ramírez and Andrés Ochoa Jaramillo. For their participation in the qualification exam and Ph.D proposal.

Thanks to the Ph.D committee: Isabel Hoyos, Marcelo Barreiro, and A. Ochoa for constructive comments on this work.

Thanks to my friends in Emergente S.A.S - Energía Sostenible : Óscar A. Rueda, Joany Sánchez, Óscar Álvarez-Villa.

This thesis was developed with the financial support of the program *Becas para Doctorados Nacionales - Convocatoria 617–2* by COLCIENCIAS.

Synchronization and interdependence between the cycles of Colombia's hydroclimatology and El Niño-Southern Oscillation

Abstract

Hydroclimatology of Colombia is highly influenced by El Niño - Southern Oscillation (ENSO), which conditions the hydrological response over Colombia, increasing (decreasing) rainfall and streamflows during La Niña (El Niño) depending on the location in the country. This dissertation presents an approach based on synchronization techniques to study the interdependence between ENSO and hydrological variables in Colombia. To that end, we use synchronization techniques such as: Phase synchronization (PS) that is based on the physical properties of weakly coupled periodic oscillators, and Generalized Synchronization (GS) that is based on properties of recurrence of non-linear dynamical systems. Furthermore, we quantify interannual hydroclimatic anomalies (HyAns) using diverse methods to evaluate the sensitivity of linear and non-linear interdependence quantifiers. Our main findings reveal that: (1) we need of quantifying the uncertainty of HyAns in terms of magnitude, sign, timing, and phase of ENSO, because HyAns methods induce an important error source and bias on the interdependence analysis and modeling of climate time series; (2) We find that the positive (negative) HyAns experienced in the Pacific, the Caribbean and the Andean regions of Colombia, during La Niña (El Niño), are phase-locked with the ENSO. Moreover, we provide evidence that the ENSO signal is phase-locked with the annual cycle of rainfall in some regions of Colombia. Furthermore, other macro-climatic processes also show significant PS such as the Pacific Decadal Oscillation (PDO) and the North Atlantic Oscillation (NAO); (3) The Caribbean, the CHOCO and the Orinoco Low-Level Jets (LLJs), and the Cross-Equatorial Flow (CEF) constitute an interdependence mechanism and contribute to explain hydrological anomalies in Colombia during the phases of ENSO. During La Niña (El Niño), GS is strong (weak) for the Caribbean and the CHOCO LLJs whereas GS is moderate (strong) for the Orinoco LLJ. Moreover, moisture advection by the Caribbean and CHOCO LLJs exhibit synchrony with HyAns at 0 to 2 (2 to 4) month-lags over north-western Colombia and the Orinoco LLJ moisture advection synchonizes with HyAns at similar month-lags over the Amazon region of Colombia. This work provides new evidence on the non-linear interactions between hydro-climatic processes in Colombia and ENSO, and constitute an unexplored approach to the understanding of climatic anomalies in tropical South America.
Keywords: Hydrology, Climatology, El Niño - Southern Oscillation (ENSO), Rainfall, Streamflows, Anomalies, Phase Synchronization, Generalized Synchronization.

Sincronización e interdependencia entre los ciclos de la hidroclimatología de Colombia y El Niño-Oscilación del Sur

Resumen

La hidroclimatología de Colombia está fuertemente influenciada por El Niño - Oscilación del Sur (ENSO), el cuál condiciona la respuesta hidrológica en Colombia, incrementando (disminuyendo) la lluvia y los caudales durante La Niña (El Niño) según la región de Colombia. Esta disertación presenta un enfoque basado en técnicas de sincronización para estudiar la interdependencia entre ENSO y las variables hidrológicas en Colombia. Para tal fin, usamos técnicas de sincronización tales como: Sincronización de Fases (PS) que está basada en propiedades físicas de osciladores periodicos debilmente acoplados, y Sincronización Generalizada (GS) que está basada en propiedades de recurrencia de sistemas dinámicos no-lineales. Además, cuantificamos las anomalies hidroclimáticas a escala interanual (HyAns) usando diversos métodos para evaluar la sensibilidad de los cuantificadores de interdependencia lineales y no-lineales. Nuestros principales hallazgos revelan que: (1) Es necesario cuantificar la incertidumbre de las HyAns en términos de la magnitud, signo, momento de ocurrencia y fase de ENSO porque los métodos HyAns inducen una importante fuente de error y sesgo para el análisis de interdependencia y modelación de series climáticas; (2) Encontramos que las anomalías positivas (negativas) en las regiones Pacífico, Caribe, y Andes de Colombia, durante La Niña (El Niño), están enfasadas con el ENSO. Además, presentamos evidencia que la señal de ENSO está enfasada con el ciclo anual de la lluvia en algunas regiones de Colombia. Adicionalmente, otros procesos macroclimáticos también exhiben sincronización de fase significativa tales como la Oscilación Decadal de Pacífico (PDO) y la Oscilación del Atlántico Norte (NAO); (3) La advección de humedad por las corrientes de Chorro de bajo nivel del Caribe, CHOCO y Orinoco, así como el Flujo Cruzado Ecuatorial (CEF) constituyen mecanismos de interdependencia y contribuyen a explicar las anomalías hidrológicas en Colombia durante las fases de ENSO. Durante La Niña (El Niño), GS es fuerte (débil) para las corrientes del Caribe y CHOCO mientras GS is moderada (fuerte) para la corriente del Orinoco. Además, la advección de humedad por las corrientes del Caribe y CHOCO exhiben sincronización con las anomalías con 0 a 2 (2 a 4) meses de rezago sobre el noroccidente Colombiano y la advección de humedad por la corriente del Orinoco se sincroniza con las anomalias de lluvia con tiempos de rezago similares sobre la región amazónica de Colombia. Este trabajo proporciona nueva evidencia sobre las interaciones no-lineales entre los procesos hidroclimáticos en Colombia y ENSO, y además constituye un enfoque inexplorado para el entendimiento de las anomalias climáticas en el norte de Sur America

Palabras Clave: Hidrología, Climatología, El Niño - Oscilación del Sur, Lluvia, Caudales, Anomalías, Sincronización de Fases, Sincronización Generalizada.

Table of contents

Abbreviations

Abbreviations

Abbreviation	Definition
ENSO	El Niño - Southern Oscillation
HyAns	Hydroclimatic anomalies
TAC	Traditional Annual Cycle
FAC	F-Filtering of the Annual Cycle
MAC	Modulated Annual Cycle
SSA	Singular Spectrum Analysis
SSAC	Annual Cycle defined using SSA
PCA	Principal Component Analysis
EMD	Empirical Mode Decomposition
EEMD	Emsemble Empirical Mode Decomposition
PS	Phase Synchronization
SPS	Strength of Phase Synchronization
NAO	North Atlantic Oscillation
PDO	Pacific Decadal Oscillation
NTA	North Tropical Atlantic
NASH	North Atlantic Subtropical High
LLJ	Low Level Jet
STS	Synthetic Time Series
LN	La Niña; Cold state of ENSO
EN	El Niño; Warm state of ENSO
GS	Generalized Synchronization
AC	Annual Cycle
CDF	Cumulative Distribution Function
SIR	Solar Incoming Radiation
GPD	Generalized Phase Difference

1. Introduction

Northern South America is a highly relevant region to study hydro-climatic processes, their interrelations and their socio-economical, environmental and ecological impacts [*Poveda and Mesa*(1997), *Coelho et. al.*(2002), *Grimm et al.(2009)*, *Poveda et al.*(2011), Andreoli et al.(2017), *Bedoya-Soto et. al.*(2018)]. In particular, the continental territory of Colombia is located in the northwestern corner of South America and it is influenced by a suite of hydro-climatic processes from diurnal to inter-annual timescales [*Poveda et al.*(2004)]. The hydro-climatic variability of Colombia is influenced by climatic processes occurring in the tropical Pacific Ocean, the Atlantic Ocean and land-atmosphere-ocean feedbacks in conjunction with the orographic influence of the Andean mountains [*Poveda and Mesa*(1997), *Poveda et al.*(2005), *Bedoya-Soto and Poveda*(2017)]. The occurrence of both phases of ENSO at interannual timescales are associated with anomalies of diverse hydro-climatic variables and processes, depending on the location in the country [*Poveda and Mesa*(1997), *Poveda et al.*(2001), *Poveda et al.*(2011), *Bedoya-Soto et. al.*(2018)]. In general, according to the literature, negative anomalies in rainfall, river flows, soil moisture and vegetation activity are witnessed during the warm phase (El Niño), and the opposite during La Niña in most of the country [*Poveda and Mesa*(1997), *Poveda et al.*(2001), *Poveda et al.*(2011), *Bedoya-Soto et. al.*(2018)].

This dissertation is a novel approach in order to reveal new insights regarding the impacts of ENSO on northern South America. To that end, we use a set of non-linear methods because the complex dynamics of the relation between ENSO and Colombian rainfall makes hard the whole picture and also because traditional approaches are based on linear methods (composites and linear correlations). In this sense, previous works that have investigated the influence of ENSO on Colombian hydro-climatology can be classified in two large groups: (1) the physical mechanisms that explain the impact of Pacific Oceanic-Atmospheric dynamics over Colombia ([*Poveda and Mesa*(1997), *Poveda and Mesa*(1999), *Poveda et al.*(2001), *Poveda et al.*(2011), *Poveda et al.*(2014), *Muñoz et al.*(2017), *Jaramillo et al.*(2017), *Hoyos et al.*(2017)]; among others) and, (2) the statistical characterization and modeling of hydro-climatic variables, under ENSO influence, using linear and non-linear techniques ([*Carvajal et al.*(1998), *Poveda et al.*(2003), *Hurtado and Poveda*(2009), *Carmona and Poveda*(2014)], among others). It is remarkable emphasize that although not only ENSO has influence over hydro-climatic variables of Colombia, it is a key component to understand the Colombian water cycle dynamics. Then, although there is evidence of the interdependence between hydro-climatic variables of Colombia and the Pacific ocean-atmospheric dynamics at ENSO timescales [*Poveda et al.*(2001), *Poveda et al.*(2011), *Bedoya-Soto and Poveda*(2017), *Bedoya-Soto et. al.*(2018)], such interdependence has not been studied using an approach based on *the interaction between the oscillations* under the assumption that physical phenomena are generated by multiple oscillatory interacting systems or *synchronization* in the sense of [*Pikovsky et al.*(2001)]. There are different methods to investigate synchronization, based on recurrences [*Marwan et al.*(2007), *Donner et al.*(2010), *Arnhold et al.*(1999), *Le Van Quyen et al.*(1999), *Shiff et al.*(1996)], phase differences [*Rosenblum et al.*(1997), *Pikovsky et al.*(2001)], or the quasi-simultaneous appearance of events [*Stolbova et al.*(2014), *Malik et al.*(2012), *Rheinwalt et al.*(2016), *Agarwal et al.*(2017)]. In particular, the phase differences method has physical basis deduced from weakly coupled oscillators whereas the method based on recurrences is deduced on properties of the non-linear dynamical systems.

In Chapter 2, we study the definition of the so-called hydro-climatic anomalies (HyAns) because extracting the annual cycle of hydro-climatic data is a widely used procedure in Geosciences such as Meteorology, Hydrology, and Climatology, among others and, although diverse methodologies are used to filter out the seasonal cycle, little effort has been made to assess the uncertainties implied in the estimation of HyAns. To advance in dealing with this issue, it is necessary to compare diverse methods to estimate HyAns, which are based on diverse reference 'climatologies' in order to investigate the differences and biases among different HyAns methods (in terms of their sign, timing, and magnitude) as well as their repercussions for linear and non-linear analysis and modeling of climate time series.

In Chapter 3, we investigate Phase Synchronization (PS) between the El Niño - Southern Oscillation (ENSO) signal and the (inter-annual and annual) cycles of rainfall in Colombia. To that end, we carry out a detailed data analysis in order to define the annual cycle (AC), eliminating the seasonal residual frequencies in hydro-climatic anomalies (HyAns), assessing the statistical significance of PS metrics, and checking the sensitivity of the PS metrics in relation to diverse HyAns estimation methods. In this chapter, we characterize the seasonal regional patterns of rainfall anomalies in relation to the ENSO's states. We look for PS between HyAns in the five natural regions of Colombia and the ENSO signal as well as the PS between the AC of rainfall and the ENSO signal. Furthermore, we explore possible PS between HyAns and other macro-climatic processes such as the Pacific Decadal Oscillation (PDO) and the North Atlantic Oscillation (NAO), among others.

In Chapter 4, we use Recurrence Quantification Analysis (RQA) to study features of Generalized Synchronization (GS) between El Niño-Southern Oscillation (ENSO) and monthly HyAns (rainfall and streamflows) in Colombia. To that end, we check the sensitivity of the RQA concerning diverse HyAns estimation methods. In general, the GS and its sensitivity to HyAns methods are quantified by means of time-lagged joint recurrence analysis. Then, we link the GS results with the dynamics of some important physical mechanisms that modulate Colombia's hydroclimatology, including the Caribbean, the CHOCO and the Orinoco Low-Level Jets (LLJs), and the Cross-Equatorial Flow (CEF) over northwestern Amazonia (southern Colombia). In this chapter, we characterize the zones of Colombian territory un-

der the influence of mentioned moisture advection sources during the phases of ENSO and possible physical basis behind the synchronization patterns.
Finally, Chapter 5 contains general conclusions and future work.

1.1. Some hydroclimatic features of the region of study

The continental territory of Colombia can be divided into five major natural regions (see Fig 2-1) with different hydro-climatological, geomorphological and bio-ecological characteristics due to the influence of diverse geographical features and processes at different spatial and temporal scales.
Some remarkable aspects of annual and inter-annual rainfall and streamflow variability in those regions are influenced by ENSO as follows:
The **Pacific** region comprises 7% of the total Colombian territory. The Atrato and San Juan rivers drain most of the region to the Caribbean Sea and the Pacific Ocean. The region witness one of the `the rainiest regions on Earth`, which has been explained in terms of land-ocean-atmosphere interactions enhanced by the Choco low-level jet (Choco Jet) [*Poveda and Mesa*(2000), *Mapes et al.*(2003a), *Mapes et al.*(2003b), *Warner et al.*(2003), *Bedoya-Soto et. al.*(2018)]. [*Poveda and Mesa*(2000)] found that zonal wind at 925 hPa penetrates into western Colombia as a low-level westerly jet, located at 5^oN. Those winds are almost absent during February–March, attain its maximum core wind velocities during October–November (6-8 m/s) and decrease onwards, which explain combined with the activity of Mesoscale Convective Complexes the high annual precipitation in this region. The annual cycle of rainfall is bimodal with the rainiest periods in March–April–May (MAM) and September–October–November (SON) and, the minima rainy periods in December–January–February (DJF) and June–July–August (JJA). This characterization was based on IDEAM and TRMM datasets [*Hoyos et al.*(2017)]. However, the recent work by urrea shows that rainfall seasonality in the Pacific region depends also on the location in the region. The central Pacific region exhibits a non-seasonal annual cycle while, that in northern Pacific is unimodal, and in southern Pacific is bimodal. In particular, during El Niño, the strength of CHOCO jet winds diminish as a result of reduction of the temperature gradient between surface air temperatures over the Pacific coast off Colombia and the Niño 1+2 zone. Such weakening of CHOCO jet winds causes a reduction of moisture advection from Pacific to Colombia and contribute to explain negative anomalies in rainfall during El Niño [*Poveda and Mesa*(1997), *Poveda and Mesa*(2000), *Rueda and Poveda*(2006), *Bedoya-Soto et. al.*(2018)].
The **Caribbean** region constitutes 12% of the total Colombian territory. The western Caribbean region exhibits a bimodal annual cycle with the rainiest periods in MAM and SON, and the driest ones in JJA and DJF, whereas the eastern Caribbean region exhibits a mixed seasonality [*Urrea et al.*(2018)]. This region is strongly influenced by the Caribbean Low Level jet (CLLJ), whose easterly zonal wind at 925 hPa has a bimodal annual cycle, with maxima in July and February, in response to the sea level pressure changes in the

North Atlantic Subtropical High (NASH) [*Wang*(2007), *Poveda et al.*(2014)]. According to [*Hoyos et al.*(2017)], the main influence of the CLLJ in the region occurs at the end of the SON season, when the blocking effect of the CHOCO jet over the Tropical Pacific is weakened and the southward component of low-level transport is intensified when the ITCZ is in its southernmost position. Then, as a result, the low level transport that crosses Central America deflects towards western Colombia, enhancing the convergence moisture in this area in DJF [*Hoyos et al.*(2017)]. Notwithstanding, during La Niña, the easterly CLLJ winds are weak (strong) in JJA (DJF), whereas during El Niño, the easterly CLLJ winds are strong (weak) in JJA (DJF) [*Wang*(2007)]. Hence, the CLLJ influence over northern South America contributes to explain rainfall anomalies during ENSO over the Caribbean basin and the nearest continental areas [*Wang*(2007), *Muños et al.*(2008), *Arias et al.*(2015)].

The **Andean** region is delimited by the three branches of the Andes mountains, its area constitutes 19% of the total Colombian territory. The Magdalena and Cauca constitute the most important river basins draining river flows into the Caribbean Sea. Precipitation in this region exhibits bimodal annual cycles with the rainiest periods in March–April–May (MAM) and September–October–November (SON). In this region, the latitudinal migration of the ITCZ and its interactions with topography play an important role in the annual distribution of hydro-climatic variables, whose ACs spatially vary in agreement with the migration of the ITCZ and other physical mechanisms [*Poveda and Mesa*(1997), *Hoyos et al.*(2017), *Urrea et al.*(2018)]. At interannual time scales, precipitation and streamflows in this region exhibit strong correlations with ENSO, overall with severe droughts during El Niño and severe floods during La Niña [*Poveda and Mesa*(1997), *Poveda et al.*(2001), *Poveda et al.*(2011), *Poveda et al.*(2015), *Córdoba-Machado et al.*(2015), *Córdoba-Machado et al.*(2014), *Bedoya-Soto et. al.*(2018)]. The work by poveda11 on the relationship between the hydro-climatic anomalies over the Andes of Colombia and their relation to ENSO, concludes that the ENSO's effects are phase-locked to the seasonal cycle, being stronger during December–January–February (DJF) but weaker during March–April–May (MAM). Furthermore, such relation between ENSO and the seasonal cycle is evidenced in rainfall, river discharges, soil moisture and, the vegetation index (NDVI) as a surrogate of evapotranspiration [*Poveda et al.*(2011)].

The **Orinoco** region is a vast plain whose area constitutes 22% of the total Colombian territory. The main rivers of its fluvial system are the Orinoco, Guaviare and Meta and its hydro-climatology is strongly influenced by the Orinoco LLJ (also known as Corriente de los Andes Orientales (CAO) or Low Level Jet in the Eastern Plains) [*Montoya*(2001), *Torrealba and Amador*(2010), *Jiménez-Sánchez*(2019)], whose strong northeasterly flux crosses through the Orinoco and Amazon river basins. Such flux causes a localized weak moisture convergence over this area, which conjointly with the slower latitudinal migration of the ITCZ explains the unimodal annual cycle of precipitation with the rainiest and driest months in June and January, respectively [*Hoyos et al.*(2017)]. The most important sources of moisture of this region are the Atlantic Ocean, local recycling and terrestrial recycling from the Amazon and Orinoco Basins [*Poveda et al.*(2014), *Arias et al.*(2015)]. In this region, hydro-climatic variability can be influenced by the Cross-Equatorial winds, which have an annual cycle with negative values (northern winds) between October and March with peak in December and positive values (southern winds) between April and September, with peak in June. The V-index for the Cross-Equatorial winds explains a high percentage of the seasonal variance of precipitation (37%) in northern South America [*Wang and Fu*(2002)].

The **Amazon** region is located north of the Amazon Basin and south of the Colombian territory. Its area constitutes 40% of the total Colombian territory. The main rivers of its fluvial system are Amazonas, Caquetá, Putumayo, Guaviare, Apaporis and Vaupes and, the moisture transport from the Amazon basin to Colombia induces temporal and spatial water cycle variability [*Poveda and Mesa*(1997)]. The AC of precipitation is unimodal with maxima in May [*Hoyos et al.*(2017)], and its main sources of moisture, as in the case of Orinoco, are the Atlantic Ocean, local recycling and terrestrial recycling from the Amazon and Orinoco Basins [*Hoyos et al.*(2017),*Arias et al.*(2015)]. In this region as well as in the Orinoco region, the annual and inter-annual cycles of hydro-climatic variables are significantly influenced by the Cross-Equatorial winds [*Wang and Fu*(2002)]. Notwithstanding, the literature does not provide information regarding the response of the Cross-Equatorial winds during different ENSO phases.

2. On the definition of hydro-climatic anomalies

Hernán D. Salas, Germán Poveda, Óscar J. Mesa, Niklas Boers, and Jürgen Kurths

2.1. Abstract

The implications of the estimation of inter-annual hydro-climatic anomalies (HyAns) need careful consideration, but the importance of the process does not correspond with the usual practice. For instance, common estimation techniques do not consider error sources or induced biases in subsequent interdependency analysis with other time series or climatic indices. Consequently, there is room for ambiguity in the physical interpretation of results. To advance in dealing with this issue, we quantify rainfall anomalies over Colombia (in terms of their sign, timing, and magnitude) and their interdependence with the El Niño - Southern Oscillation (ENSO) as a paradigmatic case to illustrate the problems, to compare different methods to define HyAns, and to infer recommendations. In general, we compare five HyAns estimation methods that consider both constant and variable definitions of the annual cycle. Next, we use the Pearson correlation analysis as well as the composite analysis to investigate the influence of those HyAns methods on the analysis of time series. Our results reveal significant differences among the HyAns methods regarding the influence of ENSO on monthly precipitation in Colombia in terms of magnitude, timing, spatial patterns, spectral properties, correlations, and probability distribution functions. This assessment study allows us to conclude that the method to estimate inter-annual HyAns based on moving averages is most adequate because of its *(i)* clear theoretical foundation, *(ii)* non-constant annual cycle, and *(iii)* adequate filtering of high-frequencies, among other criteria. Our identification of the differences between the diverse HyAns methods reveals biases and error sources toward the analysis and modeling of climate time series, which generate ambiguities for the physical interpretation of results. Our study presents methods that can be used anywhere in the world and evidences the need to tackle this fundamental but often overlooked problem. In particular, we stress its importance in assessing ENSO's effects on hydro-climatic processes.

2.2. Introduction

The seasonal component is the strongest signal in the temporal dynamics of hydro-climatic data. In order to investigate other signals and processes at different timescales, it is customary to remove the seasonality to obtain the so-called hydro-climatic anomalies (HyAns). This de-seasonalization is a widely used procedure in Geosciences and especially Meteorology, Hydrology, and Climatology, among others [*Fraedrich and Müller (1992)*, *Chandler (1992)*, *Mason et al. (2001)*, *Brohan et al.(2006)*, *Hegerl et al.(2018)*] and, although there are diverse methodologies to filter out the seasonal cycle, comparably less effort has been made to assess the uncertainties implied in the estimation of HyAns. In this sense, recent studies have shown that the analysis of hydro-climatic time series requires careful pre-processing steps to filter out the seasonal cycle in order to look for robust results and interpretations, because minor differences in data pre-processing can produce contradictory or even spurious results [*Lange et al.(2018)*, *Salas et al.(2020)*].

To advance in dealing with this issue, we compare five widely used methods to estimate HyAns, which are based on diverse reference 'climatologies': *(i)* the Traditional Annual Cycle (TAC) or constant climatology, which use long-term monthly averages as the annual cycle, which in turn assumes a constant response of the climate system to the periodic forcing of the annual solar incident radiation (SIR), and does not consider the non-linear and non-stationary responses of the climate system to the external forcing of SIR [*Salas et al.(1980)*], (ii) Fourier filtering of the annual cycle (hereafter Fourier), *(iii)* filtering of the annual cycle by Moving Averages (hereafter FAC), whose mathematical properties allow to filter out the seasonal frequencies and their harmonics from the original climatic time series [Douglass (2011), Douglass (2015)], *(iv)* Singular-Spectrum Analysis to extract the annual cycle (hereafter SSAC), which enables us to decompose a time series into its elementary patterns using a Principal Component Analysis (PCA) in the vector space of delay coordinates. After this decomposition, the elementary temporal patterns are distinguished into noise, semi-annual, annual and inter-annual components [*Vautard et al.(1992)*, *Ghil (2002)*], *(iv)* Subtraction of the Modulated Annual Cycle (MAC), which is based on the Ensemble Empirical Mode Decomposition (EEMD) technique. This method decomposes the signal into seasonal frequencies and other oscillation modes [*Huang et al.*(1998), *Wu and Huang*(2009)]. Hence, HyAns defined using the TAC and Fourier methods imply a constant climatology, whereas to define HyAns using FAC, SSAC, and MAC a non-constant annual cycle is implied.

To carry out our comparative analysis among HyAns methods, we study interannual rainfall anomalies during two extreme phases of ENSO (El Niño and La Niña) over South America because, despite the progress in understanding climate interactions in this region, uncertainty remains regarding the robustness of the impacts [*Cai et al.(2020)*]. In particular, we focus our research on the continental territory of Colombia, located in the northwestern corner of South America (see Fig. **2-1**). A suite of hydro-climatic processes from diurnal to interannual timescales influence Colombia's climate [*Poveda et al.*(2004)]. Besides, it is a highly relevant region where the hydro-climatic anomalies associated with ENSO produce strong socio-economical, environmental and ecological impacts [*Poveda and Mesa*(1997), *Coelho et al.(2002)*, *Grimm et al.(2009)*, *Poveda et al.*(2011), Andreoli et al.(2017), *Bedoya-Soto et. al.*(2018), *Poveda et al.(2020)*]. The main controls of the hydro-climatic variability of Colombia are processes occurring in the tropical Pacific Ocean, the Atlantic Ocean, and land-atmosphere-ocean feedback in conjunction with the orographic influence of the Andean mountains [*Poveda and Mesa*(1997), *Poveda et al.*(2005), *Bedoya-Soto and Poveda*(2017)]. During both phases of ENSO, there are characteristic hydroclimatic anomalies, depending on location in the country [*Poveda and Mesa*(1997), *Poveda et al.*(2001), *Poveda et al.*(2011), *Bedoya-Soto et. al.*(2018)]. In general, negative anomalies in rainfall (decrease in rainfall), river flows, soil moisture, and vegetation activity occur during the warm phase (El Niño) and the opposite during La Niña. Hence, Colombia is an adequate and interesting region to study how to characterize HyAns and investigate the differences and biases among different methods to quantify HyAns and their implications for analysis and modeling of climate time series.

Depending on the estimation method, temporal and spatial properties of HyAns could considerably differ, leading to significantly different physical interpretations about the interannual variability. Furthermore, some HyAns methods are widely accepted in climate sciences research [*Salas et al.(1980)*, *Vautard et al.(1992)*], although they lack physical justification, especially given the non-linear character of the climate processes [*Wu et al.*(2008), *Fedorov et al.(2000)*, *Thoning and Tans (1989)*]. Hence, in order to quantify the differences between HyAns methods, we compare the probability distributions of the HyAns estimations using the k-sample tests, a powerful method based on likelihood ratio that improves the conventional Kolmogorov-Smirnov, Cramér-von Mises and Anderson-Darling tests [*Zhang and Wu (2007)*].

Moreover, we explore the influence of the studied HyAns methods through Pearson's correlations, a well-known and straightforward measure of linear association [*Pearson (1895)*]. Then, we quantify Pearson correlations between an index of ENSO in the tropical Pacific Ocean, the Multivariate ENSO Index (MEI) [*Wolter and Timlin (2011)*], and rainfall anomalies in Colombia using the five different HyAns methods. Furthermore, we investigate whether the HyAns methods can induce significant differences in the composite analysis, a frequently used tool to infer the relationships among different climate phenomena [*Terray et al.(2003)*, *Boschat et al.(2016)*, *Xie et al.(2017)*] and we validate our results using parametric and non-parametric tests [*Nicholls (2001)*, *Terray et al.(2003)*]. As a complement, we discuss the role of *the scale factor* and *the reference climatology* in the formal definition of climate anomalies.

The purpose of this work is threefold: *(i)* to quantify and compare rainfall anomalies throughout Colombia estimated by the five different HyAns methods, namely HyAns–TAC, HyAns–Fourier, HyAns–FAC, HyAns–SSAC, and HyAns–MAC. *(ii)* To investigate the influence of HyAns methods on the analysis of the mutual dependence between macro-climatic phenom-

ena and hydrological time series. (iii) To recommend a HyAns method for practical purposes based on essential criteria, such as firm theoretical foundation, variable climatology, and adequate filtering of inter-annual frequencies.
This work is structured as follows: Section 2.3 describes the region of study and the precipitation data sets. Section 4.4 presents definitions of hydro-climatic anomalies using the five methods described above, as well as comparison methods such as Pearson correlation, composite analysis, and k-sample tests. Results are discussed in Section 4.5 and the conclusions are drawn in Section 4.6.

2.3. Region of study

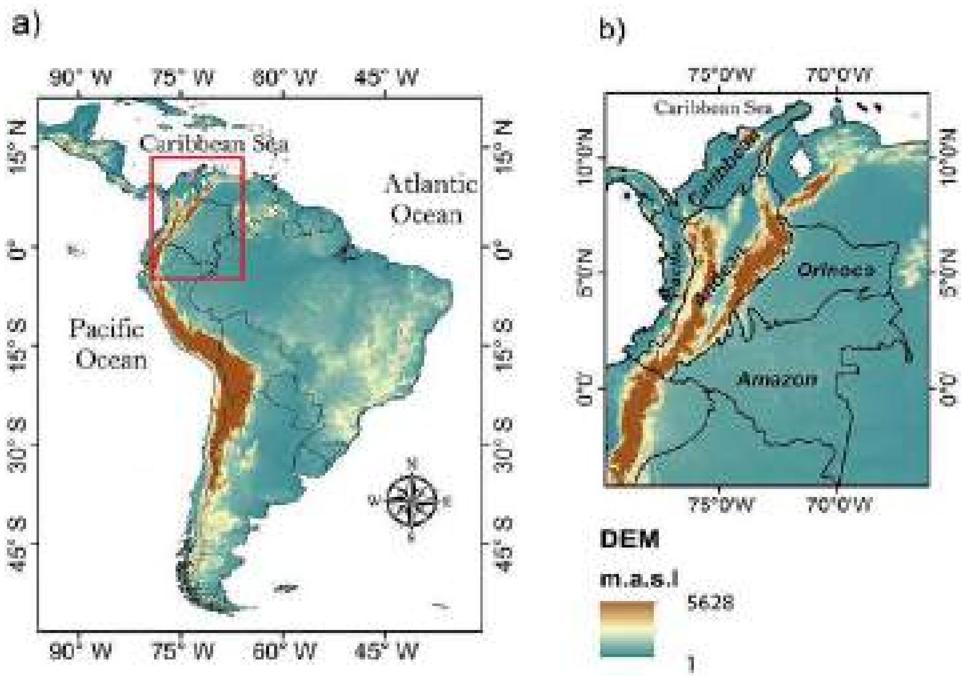
There are five regions in the continental territory of Colombia: Andes, Orinoco, Pacific, Amazon, and the Caribbean. Each exhibits different hydro-climatological, geomorphological, and bio-ecological characteristics due to the influence of diverse geographical characteristics and climate processes [*IGAC*(2002)]. Figure **2-1** shows the continental location of Colombia as well as the five different regions.

2.3.1. Data

Colombian precipitation

- **Dataset No. 1**: We use data from a precipitation reanalysis performed by Hurtado-Montoya and Mesa [*Hurtado and Mesa*(2014)], which produced monthly precipitation fields for Colombia, spanning the period from 1975 to 2006, at a spatial resolution of 5 minutes of arc, in the region 5^{o}S–15^{o}N and 80^{o}W–65^{o}W. Those authors used a set of 2270 rain gauges from diverse institutions including Instituto de Hidrología, Meteorología y Estudios Ambientales de Colombia (IDEAM), Empresas Públicas de Medellín (EPM), Centro Nacional de Investigaciones del Café (CENICAFE) and The Global Historical Climatology Network (GHCN). Besides, they also used satellite images from the Global Precipitation Climatology Project (GPCP)-V2, National Centers for Environmental Prediction (NCEP)/National Center for Atmospheric Research (NCAR), Tropical Rainfall Measuring Mission (TRMM), and the Geostationary Satellite System (GOES). Finally, the monthly precipitation fields resulted from an optimal combination of the available data and the implementation of the PRISM model [*Daly et al.(2002)*] for space interpolation, taking into account the topography of the Colombian territory. This dataset is freely available through personal communication with the authors [*Hurtado and Mesa*(2014)].

- **Dataset No. 2**: Climate Hazards Group InfraRed Precipitation with Station data (CHIRPS) is a high resolution (0.05^o), daily, pentad, and monthly quasi-global rainfall

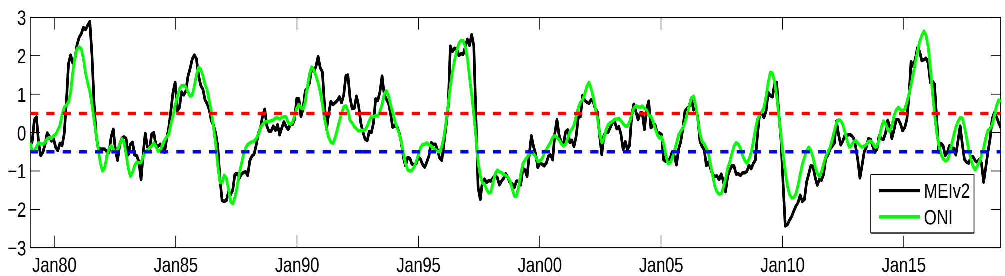


Figures 2-1.: Region of study. a) Continental context. b) National context including the five major natural regions [*IGAC*(2002)] (Andes, Orinoco, Pacific, Amazon and Caribbean). Figure was created using the void-filled elevation (30 sec DEM) raster obtained from the World Wildlife Fund HydroSHEDS project [*Lehner et al.(2008)*], and used under the following license. This product incorporates data from the HydroSHEDS database which is ©World Wildlife Fund, Inc. (2006-2013) and has been used herein under license. WWF has not evaluated the data as altered and incorporated within, and therefore gives no warranty regarding its accuracy, completeness, currency or suitability for any particular purpose. Portions of the HydroSHEDS database incorporate data which are the intellectual property rights of ©USGS (2006-2008), NASA (2000-2005), ESRI (1992-1998), CIAT (2004-2006), UNEP-WCMC (1993), WWF (2004), Commonwealth of Australia (2007), and Her Royal Majesty and the British Crown and are used under license. The HydroSHEDS database and more information are available at http://www.hydrosheds.org.

data set [*Funk et al.(2015)*]. It covers the 50^{o}S–50^{o}N latitudinal band (and all longitudes), ranging from 1981 to near-present. The purpose of CHIRPS was to support the United States Agency for International Development Famine Early Warning Systems Network (FEWS NET) [*Funk et al.(2015)*]. In this work, we use 0.05^o spatial resolution and the monthly time resolution from 1 January 1981 to 31 December 2018 over Colombia (5^{o}S–15^{o}N and 80^{o}W–65^{o}W).

ENSO signal

To characterize the ENSO signal, we use the monthly Multivariate ENSO Index (MEI) version 2.0 (MEI.v2) [*Wolter and Timlin (1998)*, *Wolter and Timlin (2011)*]. Five variables enter into the definition of the index: sea level pressure, sea surface temperature, surface zonal winds, surface meridional winds, and Outgoing Longwave Radiation (OLR). The MEI.v2 uses observations of OLR from NOAA Climate Data Record, whereas the original MEI used observations based on the International Comprehensive Ocean-Atmosphere Data Set (ICOADS) cloud cover fraction data. MEI is a more reliable estimator of the ENSO state compared to other indices such as Nino3.4 (based only on sea surface temperatures) or SOI (based on sea-level pressures) [Rasmusson and Carpener (1982), Trenberth and Stepaniak (2001)]. Values of the MEI index greater (lesser) than +0.5 (-0.5) define the occurrence of El Niño (La Niña) events, whereas the remaining periods are neutral. We classified the information of precipitation according to El Niño and La Niña events. See details on the classification of events per year at the webpage https://www.esrl.noaa.gov/psd/enso/mei.



Figures 2-2.: Multivariate ENSO Index (MEI) version 2.0 (MEIv2) and the Oceanic Niño Index (ONI). (dashed red line) values above the 0.5 for warm (El Niño) events. (dashed blue line) values below the -0.5 for cold (La Niña) events.

Besides, we also use the Oceanic Niño Index (ONI), above the +0.5^{o}C threshold for warm (El Niño) events and at or below the -0.5^{o}C anomaly for cold (La Niña) events, in order to verify the consistency of our results https://psl.noaa.gov/data/correlation/oni.data. Figure **2-2** shows the MEIv2 and ONI indices while Table **2-1** shows El Niño and La Niña events according to the mentioned criteria.

2.4. Methods

We use five methods to estimate hydro-climatic anomalies (HyAns), which depend on the way the annual cycle is defined. In those methods, the main idea is to subtract the seasonal components from data to filter out the effect of the external solar forcing, common to all hydro-climatic time series, and define the intra-annual and inter-annual variability of the processes [*Wu and Huang*(2009)]. In this work, we calculate HyAns as,

Tables 2-1.: El Niño and La Niña events according to the classification mentioned in Section 2.3.1.

El Niño	**La Niña**
1976-77	1975-76
1977-78	1983-84
1979-80	1984-85
1982-83	1988-89
1986-87	1995-96
1987-88	1998-99
1991-92	1999-00
1994-95	2000-01
1997-98	2005-06
2002-03	2007-08
2004-05	2008-09
2006-07	2010-11
2009-10	2011-12
2014-15	2016-17
2015-16	2017-18

$$Y^{(j)}_{v,\tau} = \frac{X_{v,\tau} - AC^{(j)}_{v,\tau}}{\sigma_T}, \tag{2-1}$$

where $X_{v,\tau}$ denotes the original data for year v at the time interval τ (in our case τ=1 month). $AC^{(j)}_{v,\tau}$ indicates the annual cycle (or seasonal component) estimated by the method j. In this study, σ_T, the standard deviation of the original time series, is the scale factor. Thus, in Eq. (2-1), the HyAns, $Y_{v,\tau}$, are defined using the same scale factor and varying the technique to estimate the seasonal component. After that, we classify HyAns according to the extreme phases of ENSO (La Niña and El Niño), as mentioned in section 2.3.1.

2.4.1. Hydro-climatic anomalies using TAC (HyAns–TAC)

The most popular technique to extract the annual (and semi-annual) components from hydro-climatic data is by monthly standardization. It is defined as,

$$Y_{v,\tau} = \frac{X_{v,\tau} - \hat{\mu}_\tau}{\hat{\sigma}_\tau}, \quad \tau = 1, 2, ..., 12 \text{ months}, \quad v = 1, 2, ..., n \text{ years}. \tag{2-2}$$

Here, $\hat{\mu}_\tau$ and $\hat{\sigma}_\tau$ denote the estimates of the monthly mean (or Traditional Annual Cycle - TAC) and the monthly standard deviation, respectively [*Salas et al.(1980)*]. Eq.2-2 assumes

cycle-stationary and constant climatology [*Wu et al.*(2008)]. Besides, it leaves aside the non-linear character of the climate system, and therefore the response to a periodic incoming solar radiation could not be perfectly periodic [*Wu et al.*(2008)]. Finally, it does not remove all of the seasonal frequencies [Douglass (2015)].
This approach attempts to separate the forced and the free responses of the climate system. Then, the TAC method looks for removing the part of the time series that is forced by cyclo-stationary astronomical driving. Given that such forcing is cyclo-stationary, one assumes that its variability is entirely predictable. Therefore, although this attempt is valuable in predicting the unforced component, the assumption of a linear response might bias the analysis, measures, and skills of prediction of models.

2.4.2. Hydro-climatic anomalies using Fourier transform (HyAns–Fourier)

This well-known method for analysis of time series considers that a time series is the superposition of oscillations of different frequencies. In this sense, a time series $X_{v,\tau}$ is given by

$$X_{v,\tau} = \alpha_o + \sum_{k=1}^{N/2} \alpha_k cos(2\pi f_k t) + \beta_k sin(2\pi f_k t) + \xi_t; \qquad t = 1, 2, 3, ..., N, \tag{2-3}$$

where t denotes the time, $f_k = \frac{k}{N}$ denotes the frequency, k is the harmonic of the fundamental frequency, and N is the number of observations [*Hashyap and Ramachandra (1976)*]. The parameter $\alpha_o = \bar{X}_{v,\tau}$ is the mean of the time series, and α_k and β_k are

$$\alpha_k = \frac{2}{N} \sum_{t=1}^{N} X_{v,\tau} cos(2\pi f_k t) \qquad k = 1, 2, ..., N/2, \tag{2-4}$$

$$\beta_k = \frac{2}{N} \sum_{t=1}^{N} X_{v,\tau} sin(2\pi f_k t) \qquad k = 1, 2, ..., N/2. \tag{2-5}$$

For $k > N/2$, one considers a variance spectrum by decomposing the variance of the process into a number of frequency bands. The spectral density I_k is the amount of variance per frequency band f_k defined as,

$$I_k = \frac{N}{2} \left[\alpha_k^2 + \beta_k^2\right], \tag{2-6}$$

and the angular frequency is,

$$\omega_k = \frac{2\pi k}{N} \qquad k = 1, 2, ..., M. \tag{2-7}$$

The plot of ω_k vs. I_k is the spectrum. Its smooth representation is called *Power Spectrum*, where prominent spikes indicate periodicity and important contribution to the variance of

the process at frequencies close to the peak. The total area under the spectrum is equal to the variance of the process.
Using the prominent spikes of the spectrum, one can detect the natural frequencies f_k of the series. Then one can test the significance of the periodicity following a procedure described by [*Hashyap and Ramachandra (1976)*]. In our case, we need to test the significance of periodicity at semiannual (6 months) and annual (12 months) timescales, separately. Hence, we built each new time series S_k with significant seasonal periodicity as

$$S_k = \mu + \alpha_k cos(\omega_k t) + \beta_k sin(\omega_k t), \tag{2-8}$$

where k is the harmonic of the fundamental frequency at which data exhibit significant periodicity (associated with frequencies of 6 and 12 months) and μ is the mean of time series. Finally, we define the HyAns as,

$$Y_{v,\tau} = \frac{X_{v,\tau} - S_k}{\sigma_T}, \tag{2-9}$$

where $Y_{v,\tau}$ are the anomalies calculated by subtracting the seasonal frequency bands from the original time series and σ_T denotes the scale factor.

2.4.3. Hydro-climatic anomalies using F-Filtering (HyAns–FAC)

A moving average is a method to filter out the seasonal component of hydro-climatic time series from the inter-annual components [Douglass (2011), Douglass (2015)]. As first step, we calculate $\hat{X}_{v,\tau} = X_{v,\tau} - \mu$, μ denoting the overall mean. After that, we apply a 12-month filter to $\hat{X}_{v,\tau}$, such that if we represent the time series by $\left\{\hat{X}_{v,\tau}\right\}_{\tau=0}^{m}$, $m = 11$ for monthly series, then a two-sided moving average of the time series is given by,

$$\mathbf{F}(\hat{X}_{v,\tau}) = \frac{1}{2k}\sum_{j=-k}^{k-1} \hat{X}_{v^*,(\tau+j)^*},\ \tau = k, k+1, \ldots, m-k. \tag{2-10}$$

where $\mathbf{F}(\hat{X}_{v,\tau})$ is a filtered time series based on averages of the original time series, $\hat{X}_{v,\tau}$, $m+1$ indicates the total number of months, and k is a positive integer. Note that the denominator, $2k$, in Eq. (2-10) denotes the size of the average window (in our case $k = 6$, for a 12-month average window). For the sub-indices, $(\tau+j)^* = (\tau+j)\text{mod}(m)$, and $v^* = v-1$ if $\tau+j \leq 0$, or $v^* = v+1$ if $\tau+j > m$, or $v^* = v$ in the other cases. The result is intended to be a season-free signal and the anomalies for this method are

$$Y_{v,\tau} = \frac{\mathbf{F}(\hat{X}_{v,\tau})}{\sigma_T}, \tag{2-11}$$

where $\mathbf{F}(\hat{X}_{v,\tau})$ is a low frequency signal without seasonal components. A high frequency signal that contains the seasonal and intra-seasonal frequencies and its harmonics is $\hat{X}_{v,\tau}$ −

$\mathbf{F}(\hat{X}_{v,\tau})$. In Eq. 2-11, σ_T is the standard deviation of $X_{v,\tau}$. According to Douglass [Douglass (2011), Douglass (2015)], this filter has two important mathematical properties: (1) it is similar to a low-pass filter that allows frequencies lower than 1.0 cycle/yr to pass and, (2) the Fourier transform of $\mathbf{F}(\hat{X}_{v,\tau})$ contains a factor $H12(f) = sin(12\pi f)/sin(\pi f)$, which has zeros at multiples of the frequency [*Smith (1997)*] $f = (1/12)$ month^{-1}. Thus, both the components of $\mathbf{F}(\hat{X}_{v,\tau})$ whose frequencies are exactly $f = (1/12)$ month^{-1} and all its harmonics are completely removed.
In this work, we use a 12-month filter that is also denoted as "6-1-5" where "1" is the reference [Douglass (2011)]. Hence, we lost six data points at the beginning of the time series and five at end of the time-series. That is, in Eq. (2-10), $k = 6$ and $\tau \geq 6$. It is necessary to point out that the filtered signal $\mathbf{F}(\hat{X}_{v,\tau})$ is a displaced signal in relation to $\hat{X}_{v,\tau}$ but this can be corrected using lagged-cross Pearson correlations between $\mathbf{F}(\hat{X}_{v,\tau})$ and $\hat{X}_{v,\tau}$.

2.4.4. Hydro-climatic anomalies using Singular Spectrum Analysis (HyAns-SSAC)

The basis of Singular-Spectrum Analysis (SSA) is the Principal Component Analysis (PCA) in the vector space of delay coordinates for a time series [*Vautard et al.(1992)*]. The classical PCA is used with multi-channel time series and gives the principal axes of a sequence of M-dimensional vectors x_i $(1 \leq i \leq N)$, by expanding it according to an orthonormal basis $(E^k, 1 \leq k \leq M)$.

$$X_{ij} = \sum_{k=1}^{M} a_i^k E_j^k, 1 \leq j \leq M, \tag{2-12}$$

where the coefficients a_i^k are the *principal components* and the vectors E^k are the *empirical orthogonal functions*. The vectors E^k are the eigenvectors of the cross-covariance matrix of the sequence (X_i). For a scalar time series denoted by x_i, $1 \leq i \leq M$, an equivalent expansion is

$$x_{i+j} = \sum_{k=1}^{M} a_i^k E_j^k, 1 \leq j \leq M. \tag{2-13}$$

The analogy is made by augmenting the single time series x_i into the multi-variate time series $X_i = (x_{i+1}, x_{i+2}, ..., x_{i+M})$. Therefore, there is no formal difference between Eqs. 2-12 and 2-13 [*Vautard et al.(1992)*, *Ghil (2002)*]. In the latter, M is the *window length*, chosen by the user, in contrast with the classical PCA, where M is the fixed dimension of data vectors. As we shall see, SSA allows extracting information embedded in the delay-coordinate phase space by decomposing the sequence of augmented vectors into simple patterns of behavior, which one can classify into trends, oscillatory patterns, and noise [*Ghil (2002)*]. A possible interpretation of SSA is a data-adaptive filter that decomposes the time series into components that are statistically independent at zero lag, with no use of prescribed waves (sines, cosines, or wavelets) as classical spectral analysis [*Vautard et al.(1992)*, *Ghil (2002)*]. In our case, we define $\hat{X}_{v,\tau} = X_{v,\tau} - \mu$, μ denoting the overall mean after SSA decomposition. After SSA processing, we define the standardized anomalies as,

$$Y_{v,\tau} = \frac{\hat{X}'_{v,\tau}}{\sigma_T}, \tag{2-14}$$

where $\hat{X}'_{v,\tau}$ is the reconstructed signal after SSA decomposition, with predominant frequencies at inter-annual timescales. σ_T is the scale factor, the standard deviation of $X_{v,\tau}$.

2.4.5. Hydro-climatic anomalies using MAC (HyAns–MAC)

Wu et al. [*Wu et al.*(2008)] propose a method to quantify time series anomalies based on Ensemble Empirical Mode Decomposition (EEMD) and the so-called Modulated Annual Cycle (MAC) [*Wu and Huang*(2009)]. This method consists of decomposing a signal in a set of Intrinsic Mode Functions (IFM), associated with different frequency bands, and consequently, to diverse physical oscillation modes. The IMF's containing the seasonal frequencies (i.e., annual and semiannual) constitute the MAC, subtracted from the original time series to obtain the anomalies. The rationale of this method comes from properties of non-linear dynamical systems, for which the response of a non-linear dynamical system, under an external periodical forcing, might not be perfectly periodic [*Wu et al.*(2008)]. Then, when the (non-linear) Earth climate system is excited by the seasonal solar incident radiation (SIR), most likely, such non-linearities prevent the response (the hydro-climatic variables) from being perfectly periodic [*Wu et al.*(2008)].

The Empirical Mode Decomposition (EMD) is an adaptive time-frequency data analysis method that does not require a priori functional basis because its basis functions are derived adaptively from data. This property of EMD allows its application to non-linear and non-stationary data. However, the adaptive advantage of EMD has as the price the difficulty of laying a firm theoretical foundation [*Huang and Shen (2005)*]. A summary of EMD is [*Huang et al.*(1998)]: (1) The extreme maximum and minimum values are identified in $x(t)$, (2) the maximal values are connected using a cubic spline to find an upper envelope, $U_{max}(t)$. A similar procedure is carried out to the minimal values resulting a lower envelope, $L_{min}(t)$, (3) the average value between $U_{max}(t)$ and $L_{min}(t)$ is calculated [i.e. $\bar{U}(t) = \frac{U_{max}(t) - L_{min}(t)}{2}$], (4) the time series $\bar{U}$ is subtracted from $x(t)$ [i.e. $d(t) = x(t) - \bar{U}(t)$], (5) the time series $d(t)$ is considered as the new $x(t)$ and the steps 1-4 are repeated until finding a signal with zero mean that constitutes the first IMF, (6) the IMF is subtracted from the original time series $x(t)$ and the procedure is repeated to obtain the next IMF, and (7) stop the iterative process when $d(t)$ has only a maximum or a minimum value, so it is not longer possible to extract more functions using $d(t)$.

However, after decomposing a signal into a set of IMFs by EMD, each IMF contains information at several frequency bands, making its physical interpretation difficult. To solve that problem, [*Wu and Huang*(2009)] proposed the EEMD method. A summary is (1) add a white noise series to the original data $x(t)$, (2) apply the EMD method to $x(t)$ with added noise to find the IMFs, (3) repeat steps (1) and (2) several times and, (4) calculate the (ensemble) mean for each IMF, until finding stable values of the IMF at an error level ϵ.
According to Wu and Huang [*Wu and Huang*(2009)], EEMD uses statistical properties of noise and the EMD method serves as a dyadic filter for various types of noise [*Flandrin et al.(2005)*, *Wu and Huang (2004)*], to enable the EMD method as a dyadic filter bank for any data. Then, the EEMD method largely eliminates the mode mixing problem and preserves the physical uniqueness of decomposition. In brief, the EEMD method works because the added white noise populates the whole time-frequency space uniformly with the constituting components of different scales. Then, when one adds the signal to this uniformly distributed white back-ground, its components are projected onto the reference scales established by the white noise. Hence, each trial may produce very noisy results, each EMD decomposition of the noise-added signal yields to IMFs, which contain information of the signal and the added white noise. Notwithstanding, while the noise in each trial is different in separate trials, it cancels out in the ensemble mean of enough trials. Therefore, as more and more trials go into the ensemble, the only persistent part is the signal [*Wu and Huang*(2009)].
Finally, denoting $\hat{X}'_{v,\tau}$ the summation of IMFs with predominant frequencies at inter-annual timescales, identified using Fourier spectral analysis (in our experiment), the standardized anomalies are

$$Y_{v,\tau} = \frac{\hat{X}'_{v,\tau}}{\sigma_T}, \tag{2-15}$$

where the $Y_{v,\tau}$'s are the standardized anomalies and σ_T is the standard deviation of $X_{v,\tau}$.

2.4.6. Pearson Correlations and Composite Analysis

Pearson linear correlations [*Pearson (1895)*] and composite analysis [*Terray et al.(2003)*, *Boschat et al.(2016)*] are two well-know methods to infer association between climatic variables. On the one hand, the Pearson cross-correlation coefficient for a time-lag τ quantifies the linear association between two time-series. Let X_t^i and Y_t^j two time-series in the locations i and j, then the τ-lag correlation coefficient $r_{X,Y}$ can be estimated as [*Salas et al.(1980)*],

$$r_{X,Y}(\tau) = \frac{\sum_{t=1}^{n-\tau}(X_t^i - \mu_t^i)(Y_{t+\tau}^i - \mu_{t+\tau}^i)}{\left[\sum_{t=1}^{n-\tau}(X_t^i - \mu_t^i)^2 \sum_{t=1}^{n-\tau}(Y_{t+\tau}^i - \mu_{t+\tau}^i)^2\right]^{1/2}}, \tag{2-16}$$

where $\tau = 0, 1, 2, ..., n$. For $i = j$, eqn. (3-16) is the temporal autocorrelation function (ACF).
On the other hand, the composite analysis is a tool to infer the relationships between different climate phenomena [*Terray et al.(2003)*, *Boschat et al.(2016)*, *Xie et al.(2017)*]. In general,

following Xie et al. [*Xie et al.(2017)*] to investigate the association between an index I and a field variable F, at first, we try to define some events, usually defined in terms of the index I. If I satisfies some criteria, the timings of those events constitute the key times. After that, we isolate the signal in the variable F corresponding to I by calculating the average of F in all key times. Finally, through statistical tests, we check whether there is a significant connection between those two variables or not [*Terray et al.(2003)*,*Boschat et al.(2016)*,*Xie et al.(2017)*]. In our experiment, we compare the average anomalies computed using the five different HA methods listed above, for El Niño and La Niña, to those estimated for neutral epochs (according to the classification described in section 2.3.1). Also, we test the null hypothesis of difference on average anomalies using the commonly-applied (parametric) Student's t-test [*Nicholls (2001)*] and also using a non-parametric test proposed by Terray et al. [*Terray et al.(2003)*].

2.4.7. k-sample tests based on likelihood ratio

In the vast set of methods to quantify differences between data sets, we select a powerful goodness-of-fit k-sample test, based on the likelihood ratio, in order to determine significant differences among probability distributions of rainfall anomalies obtained from the different HyAns methods. Let us denote as $Y^{(i)}_{v,\tau}$ and $Y^{(k)}_{v,\tau}$ the estimated anomalies through methods i and k, respectively. Then, it is possible to test the null hypothesis that $Y^{(i)}_{v,\tau}$ and $Y^{(k)}_{v,\tau}$ belong to the same probability distribution function using a two-sample test.

According to Zhang [*Zhang and Wu (2007)*], the new k-sample tests have two main advantages to more traditional tests (Kolmogorov-Smirnov, Anderson-Darling and, Cramér-von Mises). The first is that traditional tests are derived based on the chi-squared statistic (χ^2), whereas the new k-sample tests use the likelihood-ratio statistic. The second is that traditional tests are sensitive to location differences among distributions and are comparably less strong in detecting variations in shape, whereas k-sample tests are sensitive to both location and shape [*Zhang (2002)*,*Zhang and Wu (2007)*]. Finally, k-sample tests allow us to compare the probability distribution functions from the complete data sets instead of using metrics based on a unique statistic attributes such as mean absolute error (MAE), mean square error (MSE), root mean square error (RMSE), Minkowski distances (L_λ), among others. In this work, Z_K, Z_A and, Z_C denote the new k-sample Kolmogorov-Smirnov, Anderson-Darling and, Cramér-von Mises tests, respectively.

2.5. Results and Discussion

The following is a discussion about the spatial and temporal analysis of differences among the five methods to estimate hydro-climatic anomalies.

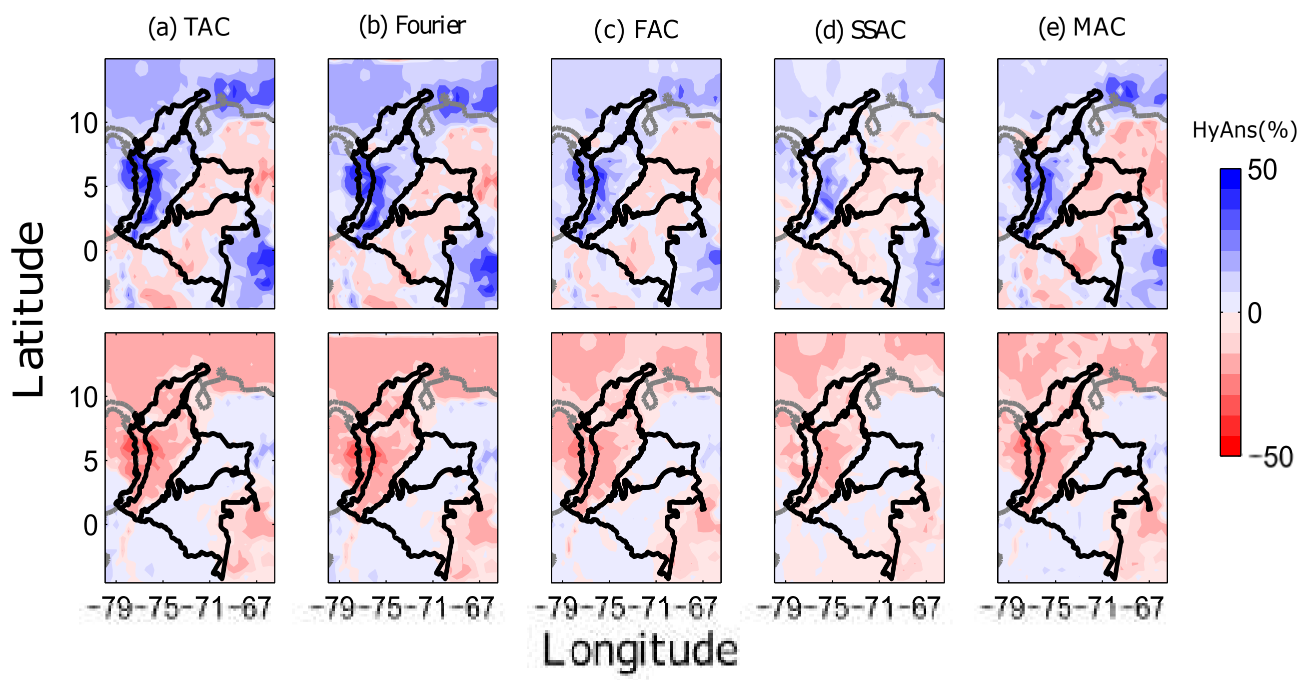
2.5.1. Spatial patterns of rainfall anomalies during ENSO at the national scale

Overall, the Andes, Caribbean, and Pacific regions exhibit negative (positive) rainfall anomalies during El Niño (La Niña). In contrast, the Amazon and Orinoco regions experience positive (negative) rainfall anomalies during the warm (cold) phase of ENSO [*Poveda et al.*(2011), *Bedoya-Soto et. al.*(2018), *Cai et al.(2020)*, *Salas et al.(2020)*]. Figure **2-3** suggests that the studied HyAns methods yield consistent results classifying anomalies in terms of their sign during both phases of ENSO. However, Fig. **2-3** also evidences that the methods are not consistent in the magnitude of rainfall anomalies although for all five HyAns methods we used the same scale factor. A couple of examples regarding the differences among the HyAns methods are (1) during La Niña, the HyAns–TAC, HyAns–Fourier, and HyAns–MAC methods exhibit high-positive anomalies in the Pacific region, whereas HyAns–FAC and HyAns–SSA show lower-positive anomalies in the same region (Fig. **2-3**, top row); (2) during El Niño, the HyAns–TAC, HyAns–Fourier, and HyAns–MAC methods exhibit high negative rainfall anomalies over the Pacific and Andes regions, whereas the other methods show lower negative rainfall anomalies (Fig. **2-4**, bottom panels).

Furthermore, we compare the spatial patterns of rainfall anomalies over the continental territory of Colombia using the five HyAns methods. To that end, we compare the cumulative probability distribution functions (CDFs) of the anomaly fields using the k-sample tests based on the likelihood ratio, at the 5% significance level. Our results indicate that: (1) During El Niño (Fig. **2-5** a), 70% (30%) of the study region experiences negative (positive) rainfall anomalies. All HyAns methods coincide with this feature. Moreover, the k-sample tests indicate that CDFs of HyAns–TAC and HyAns–Fourier are statistically equal, whereas the comparison among the CDFs of HyAns–FAC, HyAns–SSA and HyAns–MAC indicates that those produce statistically different results; (2) During La Niña (Fig. **2-5** b), 70% (30%) of the study region experiences positive (negative) rainfall anomalies, and all HyAns methods show this characteristic. Also, the k-sample tests indicate that CDFs of HyAns–TAC and HyAns–Fourier are statistically equal, whereas the statistical comparison among HyAns–FAC, HyAns–SSA and HyAns–MAC indicates that their CDFs are dissimilar.

A similar analysis was done using Dataset No. 2 (CHIRPS), Fig. **2-4** shows a good consistency with the results from Dataset No.1 over the regions that exhibit negative (positive) rainfall anomalies during El Niño (La Niña). Figure **2-5** (c-d) shows the CDFs for the CHIRPS dataset. k-sample tests indicate that CDFs of HyAns methods based on a constant annual cycle are similar (HyAns–TAC and HyAns–Fourier). In contrast, CDF's of HyAns methods based on a non-constant annual cycle differ from each other.

It is necessary to point out that the magnitude of rainfall anomalies could change significantly depending on the dataset used and the phase of ENSO (e.g., see Figure **2-5**b and Figure **2-5**d). Furthermore, the classification of the ENSO events using the MEI and ONI indices provide similar results in terms of the composite rainfall anomaly fields, which suggest that

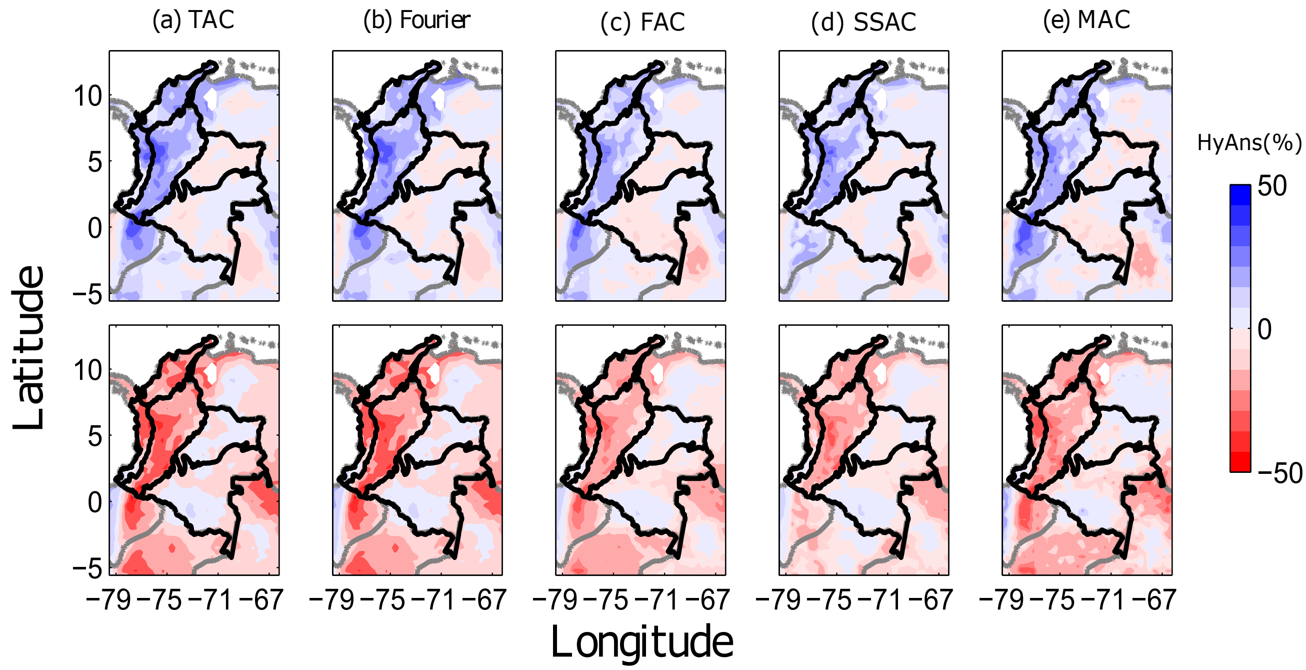


Figures 2-3.: Composite rainfall anomalies (HyAns) over Colombia (Dataset No. 1 – Hurtado-Montoya and Mesa (2014)), during ENSO, estimated by the five HyAns methods. (Top panels) HyAns during La Niña (LN). (Bottom panels) HyAns during El Niño (EN). (dark line) boundaries of the natural regions of Colombia. (gray line) northern South America boundaries of countries.

differences come from the HyAns methods of estimation (see Supplementary Information, Figures A-1 and A-2). It is remarkable that during El Niño (La Niña), k-sample tests over regions exhibiting only negative (positive) anomalies indicate that HyAns–FAC and HyAns–MAC have statistically similar CDFs. Our results reveal that composite HyAns fields calculated by different methods have significant statistical differences, although, at first glance, the differences seem to be minor.

2.5.2. Temporal patterns of rainfall anomalies during ENSO

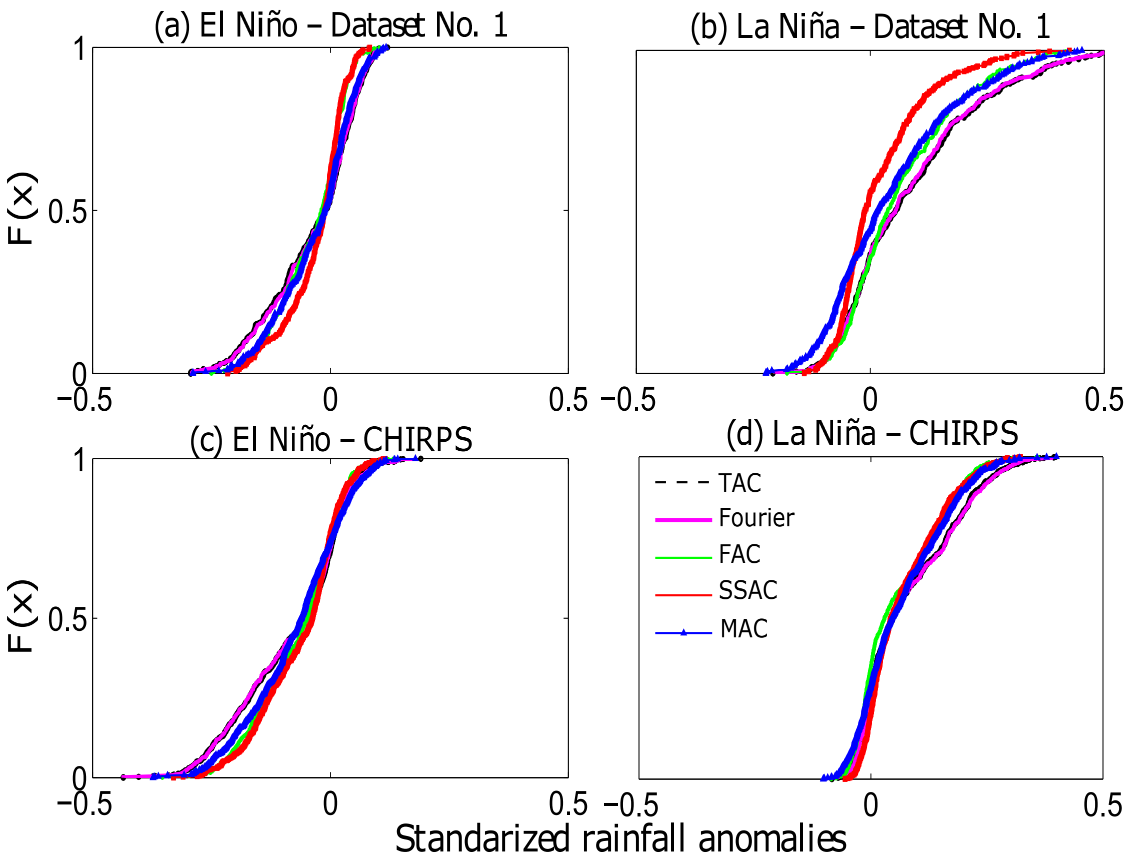
First, we analyze the Cumulative Distribution Functions (CDFs) of monthly rainfall anomalies averaged across the five regions of Colombia. These analyses evidence that: (1) During La Niña (Fig. **2-6** top row), rainfall anomalies are 70% (30%) of the time positive (negative) in the Pacific, Andean and Caribbean regions whereas in the other regions are 70% (30%) of the time negative (positive); (2) During El Niño (Fig. **2-6** bottom row), rainfall anomalies are 70% (30%) of the time negative (positive) in the Pacific, Andean and Caribbean, whereas in the other regions are 70% (30%) of the time negative (positive). Moreover, for



Figures 2-4.: Composite rainfall anomalies (HyAns) over Colombia (Dataset No. 2 – CHIRPS), during ENSO, estimated by the five HyAns methods. (Top panels) HyAns during La Niña (LN). (Bottom panels) HyAns during El Niño (EN). (dark line) boundaries of the natural regions of Colombia. (gray line) northern South America boundaries of countries.

both phases of ENSO, all k-sample tests at 0.05 significance level, suggest that CDFs of HyAns–TAC and HyAns–Fourier are similar. In contrast, the CDFs of the other HyAns methods are different to each other. Less rigorous k-sample tests, at the 0.10 significance level, indicate that HyAns–FAC and HyAns–MAC are similar in most cases. Those results show high sensibility in HyAns estimation depending on the method.

The autocorrelation functions (ACFs) in Fig **2-7** (top row) show that the HyAns–TAC and HyAns–Fourier rainfall anomalies exhibit a fast-decaying ACF. In contrast, the other HyAns methods show a slower decay. These mean that ACFs of HyAns methods based on a non-constant annual cycle (HyAns–FAC, HyAns–SSA and HyAns–MAC) are analogous. In contrast, the ACF of HyAns methods based on a constant climatology (HyAns–TAC and HyAns–Fourier) are different from those based on non-constant climatology. Regarding the power spectral density (PSD), Fig. **2-7** (bottom row) shows that HyAns–TAC and HyAns–Fourier are similar. Also, the PSDs exhibit significant differences among HyAns–FAC, HyAns–MAC and HyAns–SSA. Using the statistical test to detect significant periodic frequencies proposed by [*Ahdesmäki et al.(2005)*], one can demonstrate that HyAns–TAC and HyAns–Fourier leave significant residual frequencies at intra-annual and seasonal time

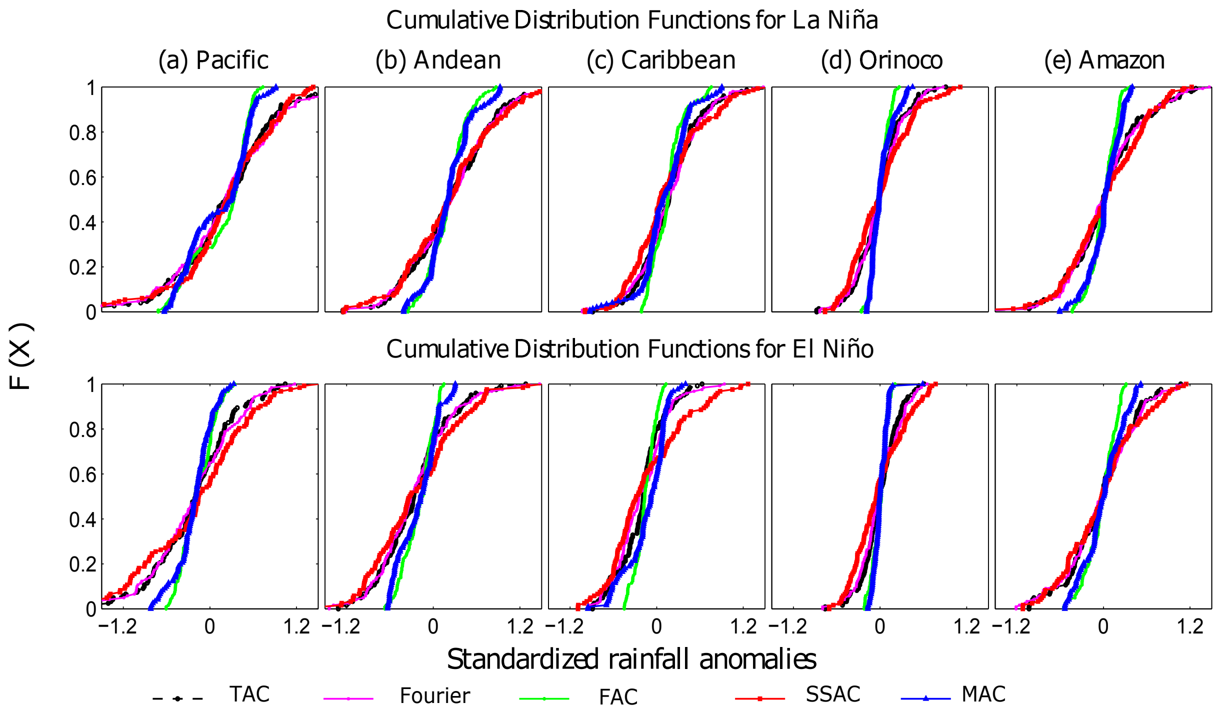


Figures 2-5.: Cumulative Distribution Functions (CDFs) of composite rainfall anomalies fields throughout the continental territory of Colombia, using diverse HyAns methods, during the extreme phases of ENSO (El Niño and La Niña). (a-b) CDFs using Data set No. 1. (c-d) CDFs using Data set No. 2 – CHIRPS.

scales, these residuals help explain the fast-decaying of ACF for these methods.
Moreover, the HyAns–SSA method mixes modes of variability with significant frequencies at intra-seasonal and seasonal time scales that are hard to link with physical processes (not shown). Meanwhile, the HyAns–FAC and HyAns–MAC methods allow removing the seasonal and intra-seasonal frequencies satisfactorily and contribute to attest to the reliability of the filtering to define inter-annual variability.

2.5.3. Repercussions of HyAns methods on Pearson correlation analysis between rainfall anomalies and the ENSO signal

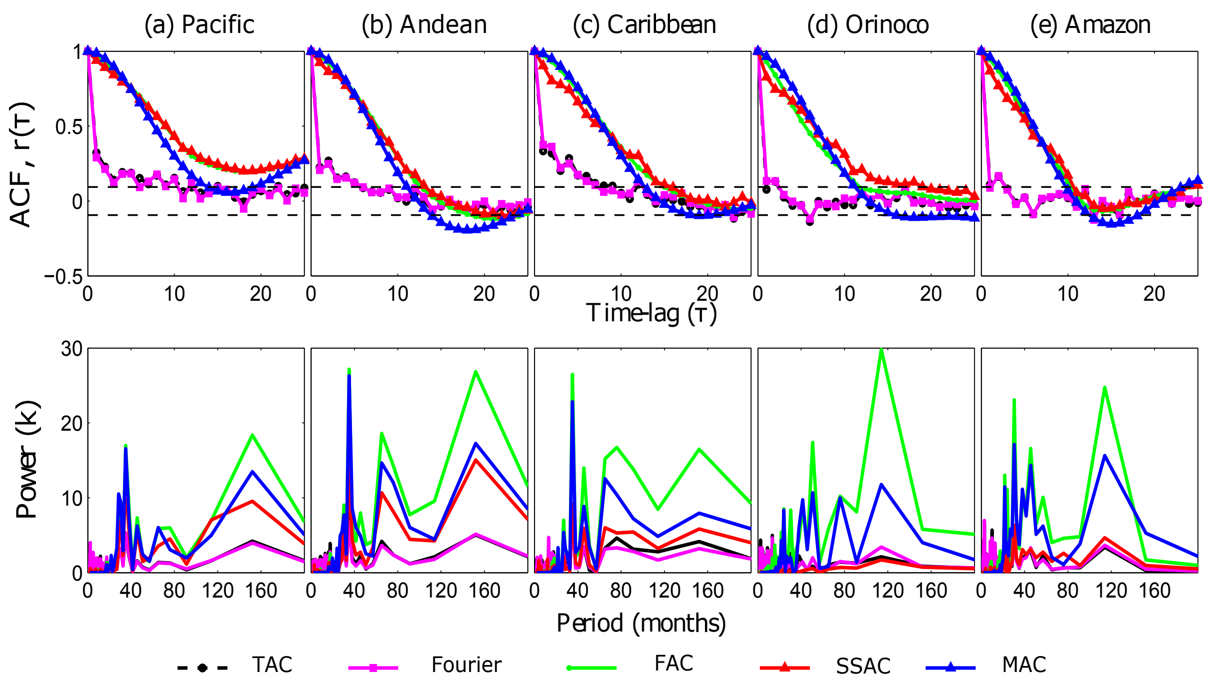
The choice of the HyAns method has a substantial impact on the interdependence analysis and modeling of climate time series. In order to illustrate this we study the association between ENSO and rainfall anomalies over Colombia using a linear correlation analysis, which



Figures 2-6.: Cumulative Distribution Functions (CDFs) of monthly rainfall anomalies (Dataset No. 2 –CHIRPS) averaged over the five regions of Colombia for La Niña (top row) and El Niño (bottom row). Each column denotes one of the Colombian regions.

is a well-known model to quantify the possible association between variables. Moreover, Pearson correlation has been used to investigate the relationship between ENSO indices and different hydro-climatic variables over Colombia (rainfall, streamflows, temperatures, vegetation indices, among others). Those investigations show significant linear correlations among the tropical Pacific Ocean dynamics and the land-atmosphere-ocean variability over northern South America [*Poveda and Mesa*(1997), *Poveda et al.*(2011), *Córdoba-Machado et al.*(2015), *Córdoba-Machado et al.*(2014), *Bedoya-Soto et. al.*(2018)].

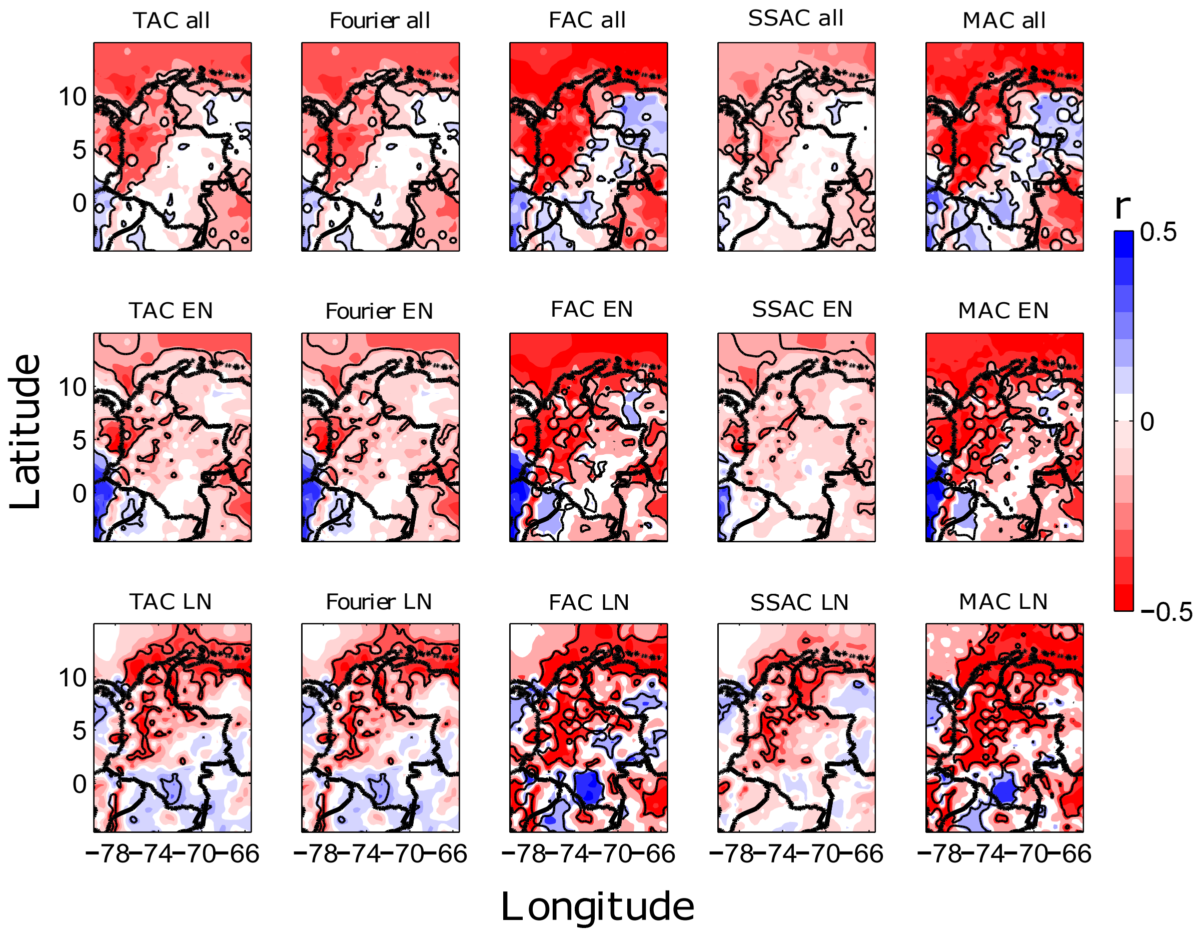
In our experiment, we quantify the Pearson correlation coefficient r, between the MEI index and rainfall anomalies over Colombia and its statistical significance. To that end, we use 1200 monthly time series of rainfall, each one corresponding to a pixel of the rainfall fields for Colombia, spanning the period from 1975 to 2006 [*Hurtado and Mesa*(2014)]. For each time series, we estimate the anomalies using the HyAns methods based on constant and non-constant annual cycles. Then, we calculate r between the series of rainfall anomalies and the MEI index. Also, we test the null hypothesis of no relationship between the observed phenomena. If the p-value is smaller than the significance level (default is 0.05), the corresponding correlation is significant. In this sense, we find that significant r typically correspond to values higher than 0.50 or lower than -0.50.



Figures 2-7.: Auto-correlation Functions, ACFs (top row) and Periodograms (bottom row) of monthly rainfall anomalies, spatially averaged over the five regions of Colombia using diverse HyAns methods.

In general, our results evidence that the Pearson correlation coefficients do not exhibit significant differences concerning the HyAns methods, when we use time series without classifying the data points according to the phases of ENSO. An explanation for those results is that 42% of the data points in the time series correspond to ENSO's neutral phase. In contrast, during the extreme phases of ENSO (El Niño and La Niña), maps of correlations exhibit significant differences with the HyAns methods (Fig.**2-8**). In such cases, Pearson correlations do not have a bias induced by neutral data points, where 41% (17%) of the data points correspond to El Niño (La Niña).

In particular, our findings allow us to conclude that: (1) the Andean, Pacific, and Caribbean regions of Colombia exhibit significant negative correlations, when we use the complete time series, for all HyAns methods as reported in several previous works [*Poveda and Mesa*(1997), *Poveda et al.*(2011)]. Furthermore, the Orinoco and Amazon regions show mainly non-significant correlations, although the HyAns–FAC and HyAns–MAC methods indicate some areas of positive significant r values (Fig.**2-8** top row). (2) During El Niño, the r-map using HyAns–TAC, HyAns–Fourier and HyAns–SSA suggest strong correlations across the Caribbean and western Colombia. In contrast, the r-map using HyAns–FAC and HyAns–MAC show a vast area of high significant r-values covering the continental territory



Figures 2-8.: Pearson Correlation, r, between the MEI index and rainfall anomalies over Colombia using diverse HyAns methods, each column denotes a HA method. (top row) r using complete time series without classification of data according to the ENSO phases. (middle row) r for the data points classified as El Niño (EN) following the MEI index. (bottom row) r for the data points classified as La Niña (LN) following the MEI index. (contour lines) Zones where the Pearson Correlation Coefficient, r, is significant at 5%.

of Colombia in the Pacific, Caribbean, and Andean regions, including the Caribbean sea zone in northern Colombia (Fig.**2-8** middle row). (3) During La Niña, the HyAns methods exhibit differences regarding the areas of r significance. r-maps using HyAns–TAC, HyAns–Fourier, and HyAns–SSA indicate smaller correlated areas over the country than r-maps using HyAns–FAC and HyAns–MAC. Besides, the r-map for most of the methods show high r-values covering the country to the north of 5^o N, whereas the r-map using HyAns–SSA suggests a weak connection towards the Caribbean Sea (Fig.**2-8** bottom row). (4) k-sample tests for datasets of significant r-values suggest equality of CDFs between methods based on constant annual cycle as well as equality of CDFs among methods based on a non-constant annual cycle. Furthermore, there is a non-equality of CDFs in other cases (CDFs not shown here).

2.5.4. Implications of HyAns methods on Composite Analysis considering the phases of ENSO

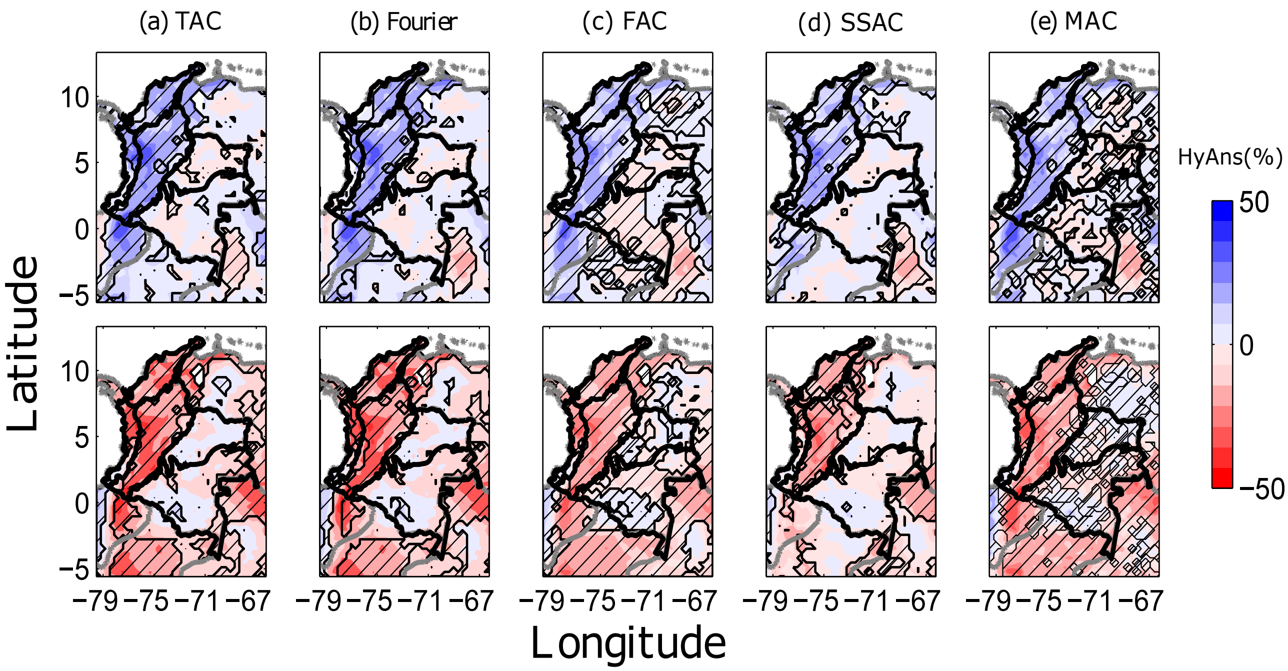
We carry out a composite analysis to verify how HyAns methods affect the areas of the statistical significance of the composite fields. To this end, we compare the average anomalies for La Niña and El Niño phases in relation to the Neutral ones. We compare composite using the t-test (parametric) and the Terray's test (non-parametric). In general, our results reveal that the HyAns computed by the FAC and MAC methods include areas of statistical significance that the HyAns by TAC, Fourier, and SSA are not able to capture (see Figure **2-9**). Results are similar for both statistical tests and data sets (see Supplementary Information, Figures A-3 to A-6). Hence, these results show that the structure of composites can change significantly and requires a careful analysis of sensitivity regarding the estimation of anomalies. Boschat et al. [*Boschat et al.(2016)*] suggested that *"the message contained in the structure of composites may be quite misleading unless interpreted carefully, and that physical hypotheses framed from these need rigorous tests with appropriate tools"*. In this sense, our analysis allows us to conclude that the definition of anomalies plays a fundamental role in the adequate inference through composite analysis.

2.5.5. The role of the scale factor and the reference climatology on traditional HyAns

At this point, it is worth mentioning that the scale factor affects the interpretation of HyAns. For example, the American Meteorological Society (AMS) defines **climate anomaly** as "*the difference between the average climate over a period of several decades or more, and the climate during a particular month or season*" [*AMS(2015)*]. Such a definition implies that the scale factor is unity. At the same time, some well-known climate indexes are standardized (e.g., Southern Oscillation Index –SOI) and take as scale factor the standard deviation of each month [*NOAA(2018)*], which are the so-called standardized anomalies. Even though both cases consider constant climatology, the latter provides more information about the magnitude of the anomalies because they are in terms of the dispersion.

Both definitions above use the original time series, subtracting a base value from it. As we mentioned in 2.4.1, the monthly mean of data is the most used value. However, as discussed by Kawale et al. [*Kawale et al.(2011)*], the true theoretical mean is unknown, and most of the time, a short reference interval defines the reference climatology, typically, 30 years. Sometimes a shorter reference period can produce spurious results [*Kawale et al.(2011)*].

Then, the official definition of **climate anomaly** is different from the standardized climate



Figures 2-9.: Composite analysis for rainfall anomalies (Dataset No. 2; CHIRPS) showing the zones where mean anomalies for La Niña (top row) and El Niño (bottom row) are statistically different to the mean anomalies for the neutral phases of ENSO. (hatched lines) zones of statistically different mean values using the Terray's test [*Terray et al.(2003)*] at the 0.05 significance level.

anomaly that provides information on the magnitude of anomalies in terms of the dispersion of climate processes. Furthermore, the reference period is a subjective criterion that does not consider the non-stationarity and non-linear dynamics of Earth climate processes. Also, the most used definition of climate anomalies does not consider the behavior of hydro-climatic variables in response to climate change.

Our results evidence that HyAns methods based on constant and non-constant annual cycles can be strongly affected by the scale factor selected. In particular, the HyAns–Fourier method exhibits high anomaly values if the scale factor considers the seasonal dispersion of the climate time series.

2.6. Concluding Remarks

We study spatial and temporal patterns of hydro-climatic anomalies at interannual timescales throughout Colombia during ENSO's extreme phases. To that end, we use diverse HyAns methods that use constant and non-constant definitions of the seasonal variability or annual cycle: Traditional Annual Cycle (TAC), Fourier filtering of the annual cycle (Fourier), filtering of the annual cycle by Moving Averages (FAC), annual cycle extracted by Singular Spectrum Analysis (SSA), and Modulated Annual Cycle (MAC). Our results unveil significant differences among time series of anomalies in terms of their: spatial patterns, magnitude, residual seasonal frequencies, the presence of noise, autocorrelations functions, spectral properties, and probability distribution functions.

Furthermore, we studied the influence of these HyAns methods to analyze interdependence among climate time series. We explored the relationship between ENSO and rainfall anomalies in Colombia using Pearson correlations. Our results evidence that the significant linear correlations might vary depending on the HyAns method and the phases of ENSO. For the composite analysis, our results allow us to conclude that the definition of anomalies plays a fundamental role in the proper inference of physical linkages through composites. Furthermore, the apparent minor differences in the HyAns estimation using diverse methods induce a significant error source and bias on the correlation analysis and modeling of climate time series. Moreover, these results evidence that data pre-processing for HyAns estimation requires verification of the adequate filtering of seasonal frequencies because some methods such as HyAns–TAC, HyAns–Fourier and HyAns–SSA do not remove satisfactorily such frequencies inducing differences in HyAns estimation and correlation metrics.

From these comparisons, for estimation of hydroclimatic anomalies at inter-annual timescale, we suggest using the filtering of the annual cycle using moving averages, here called HyAns–FAC because it (i) has a firm mathematical foundation, (ii) considers a varying climatology (annual cycle), (iii) adequately separates noise, intra-annual, seasonal and inter-annual components with (iv) not-significant residual seasonal frequencies in HyAns, (v) generates moderate extreme values of anomalies in comparison with the HyAns–SSA and HyAns–MAC methods, (vi) its autocorrelation function is comparable with HyAns–SSA and HyAns–MAC, and (vii) its computational cost is a one-tenth part of HyAns–SSA and hundreds of times less than HyAns–MAC.

Also, our results point out the need to quantify the uncertainty of HyAns in terms of magnitude, sign, timing, and phase of ENSO. It is also pertinent to revise the theoretical basis behind the diverse HyAns methods, not necessarily supported on physical grounds. Likewise, we discuss how the scale factor plays an essential role in the meaning of HyAns and how the period of the reference climatology affects HyAns estimations. e.g., climate reference varies depending on extreme phases of ENSO. On the other hand, HyAns can be affected by the datasets used, and hence we recommend to use different datasets in order to provide robustness to HyAns estimations.

Moreover, a large induced bias to correlation analysis of time series is a fundamental issue for future research, which needs consideration related to sophisticated non-linear methods reported in the literature. Likewise, it is essential to investigate the false "skill" in prediction models, whose input data contain a significant cyclo-stationary forced component. Our results evidence the need to tackle this fundamental overlooked problem regarding the effects of ENSO in hydroclimatic processes.

2.7. Acknowledgments

The work of H.D. Salas is supported by COLCIENCIAS – Grant for National Doctorates 617-2 and partially funded by the Humboldt University of Berlin IRTG 1740. The work of G. Poveda and O. J. Mesa is supported by Universidad Nacional de Colombia at Medellin. N.Boers acknowledges funding by the Alexander von Humboldt Foundation, the German Federal Ministry for Education and Research, and the German Science Foundation (DFG, Reference BO 4455/1-1). The work by J. Kurths is supported by PIK. We thank Profs. Brian Mapes (University of Miami) and David J. Raymond (New Mexico Tech) for their insightful comments and discussions. Thanks to: U.S. Geological Survey (USGS HydroSHEDS) for the Digital Elevation Models, NOAA for the ENSO index data and, A. F. Hurtado for reanalysis data of precipitation over Colombia.