UNIVERSITY OF ANTIOQUIA

MASTER THESIS

Colombian climate teleconnections from complex networks and information transference

---

Author:
Nicole RIVERA-PARRA

Advisors:
PhD. Boris RODRIGUEZ,
PhD. Isabel HOYOS

A thesis submitted in fulfillment of the requirements
for the degree of Master in Physics

in the

Fundamentos y Enseñanza de la Física y los Sistemas Dinámicos
Exact and Natural Sciences Faculty

Medellín, 2024

Declaration of Authorship

I, Nicole RIVERA-PARRA, declare that this thesis titled, "Colombian climate teleconnections from complex networks and information transference" and the work presented in it are my own. I confirm that:

- This work was done wholly or mainly while in candidature for a research degree at this University.
- Where any part of this thesis has previously been submitted for a degree or any other qualification at this University or any other institution, this has been clearly stated.
- Where I have consulted the published work of others, this is always clearly attributed.
- Where I have quoted from the work of others, the source is always given. With the exception of such quotations, this thesis is entirely my own work.
- I have acknowledged all main sources of help.
- Where the thesis is based on work done by myself jointly with others, I have made clear exactly what was done by others and what I have contributed myself.

Signed: 

Date: 27/5/2024

"There is no man, however wise, who has not at some period of his youth said things, or lived in a way the consciousness of which is so unpleasant to him in later life that he would gladly, if he could, expunge it from his memory. And yet he ought not entirely to regret it, because he cannot be certain that he has indeed become a wise man — so far as it is possible for any of us to be wise — unless he has passed through all the fatuous or unwholesome incarnations by which that ultimate stage must be preceded."

Marcel Proust

UNIVERSITY OF ANTIOQUIA

Abstract

Exact and Natural Sciences Faculty
Physics Institute

Master in Physics

Colombian climate teleconnections from complex networks and information transference

by Nicole RIVERA-PARRA

This study leverages complex network analysis and information transfer quantifiers to characterize the climate variability teleconnection spatial patterns across Colombia, based on causal relationships. Causality between six oceanic indices (Niño 3.4, Niño 3, SOI, TNA, NTA, and CAR) and three local climatic variables (total precipitation [TP], two-meter temperature [T2M], and vertical convergence of moisture flux [VIMFLUX]) is estimated through the Liang-Kleeman Maximum Likelihood Estimator (LKMLE), which works as a quantifier of information transfer between time series. Six climate complex networks are built based on the strength of these teleconnections, summarizing the overall causal influence of each index on the chosen variables.

The spatial distribution of teleconnections between El Niño-Southern Oscillation (ENSO) indices and TP captures the strong connectivity over eastern and southeastern Colombia, as in previous studies. Nevertheless, our findings suggest that ENSO does not have a statistically significant causal relationship with T2M. On the other hand, Atlantic Ocean-related indices are found to have significant causal teleconnections with northern, western, and northwestern Colombia, especially over the Orinoco and Amazon's basins for TP, in agreement with previous studies.

Our model results in simple, undirected, and unweighted graphs. The presence of several isolated vertices for all indices suggests sparse networks. This indicates that most of the territory exhibits limited teleconnection-based connectivity. Despite this sparseness, the properties of the graphs hint at a highly complex underlying network structure that may not readily fit into the currently established categories of canonical complex networks.

Keywords: Complex Network Analysis, Atmospheric Teleconnections, Climate Networks, Colombia

Acknowledgements

To the universe's beating heart,
Thank you for the whispers of wonder.
Every breath, a privilege to learn.

University of Antioquia, fertile ground,
Estudiante Instructor programm,
This journey is possible thanks to it.

Isabel and Boris, guiding lights,
Physics of knowledge, humanity of science,
Your lessons etched deeper than texts.

Roots of love, mom, dad, abuelo,
Your steady grip in sun and storm,
A thousand thanks for your steadfast hold.

Laughter's melody, Alejandra, Camila,
Stefania, Laura, Ximena, Julietta,
Friendship's strength through tears and cheers.

Caffeinated wisdom, Sebastian, Andres,
Coffee and knowledge, shared journeys fueled,
With every cup, a passion kept alive.

Love's surprise duet, Alejandro,
Your harmony made this manuscript sing,
And my heart, beautifully bloom with your care.

Curious readers, a hand outstretched,
Thank you for joining this journey,
One small piece of knowledge for Earth, an universe to explore.

Contents

List of Figures

List of Tables

List of Abbreviations

AMO	**A**tlantic **M**ultidecadal **O**scillation
AWC	**A**rea **W**eighted **C**onnectivity
C3S	**C**opernicus **C**limate **C**hange **S**ervice
CAR	**Car**ibbean SST Index
CLLJ	**C**aribbean **L**ow-**L**evel **J**et
Chocó LLJ	**C**hocó **L**ow-**L**evel **J**et
ECMWF	**E**uropean **C**entre for **M**edium-Range **W**eather **F**orecasts
EN	**E**l **N**iño event
ERE	**E**xtreme-**R**ainfall **E**vent
ES	**E**vent **S**ynchronization
ESN	**E**vent **S**ynchronization **N**etwork
GCM	**G**eneral **C**irculation **M**odels
IOD	**I**ndian **O**cean **D**ipole
IPCC	**I**ntergovernmental **P**anel on **C**limate **C**hange
ITCZ	**I**n**t**ertropical **C**onvergence **Z**one
LKMLE	**L**iang-**K**leeman **M**aximum **L**ikelihood **E**stimator
LN	**L**a **N**iña event
MJO	**M**adden-**J**ulian **O**scillation
NOAA	**N**ational **O**ceanic and **A**tmospheric **A**dministration
NTA	**N**orth **T**ropical **A**tlantic
NWS	**N**ational **W**eather **S**ystem
PDF	**P**robability **D**ensity **F**unction
PDO	**P**acific **D**ecadal **O**scillation
RCM	**R**egional **C**limate **M**odels
SAD	**S**outh **A**tlantic **D**ipole
SAM	**S**outhern **A**nnular **M**ode
SLP	**S**ea **L**evel **P**ressure
SST	**S**urface **S**ea **T**emperatures
SOI	**S**outhern **O**scillation **I**ndex
T2M	**T**wo-Meter **T**emperature
TAD	**T**ropical **A**tlantic **D**ipole
TNA	**T**ropical **N**orth **A**tlantic
TP	**T**otal **P**recipitation
VIMFLUX	**V**ertical Convergence of **M**oisture **Flux**

List of Symbols

N	Number of Nodes
L	Number of Links
k_v	Vertex Degree
C	Clustering Coefficient
$\langle C \rangle$	Mean Clustering Coefficient
$\mathcal{D}$	Density of the graph
$\mathcal{P}$	Average Path Length
A	Adjacency Matrix
$\langle k \rangle$	Average Graph Connectivity
$\mathcal{H}$	Entropy of the graph
u, v	Nodes
$d(u, v)$	Distance Between Nodes
$\epsilon(v)$	Node Eccentricity
$\overline{\epsilon(v)}$	Average Node Eccentricity
$\mathbf{R}$	Radius of the graph
$\mathbf{D}$	Diameter of the graph
l_i	Mean Geodesic Distance
$\mathcal{C}_i$	Closeness Centrality
$\mathcal{B}_i$	Betweenness Centrality
$T_{Y \to X}$	Information Transfer Flow from Y to X
$\tau_{Y \to X}$	Relative Information Transfer Flow from Y to X

Para mi Abuelo…

Introduction

In the wake of unprecedented global climate change, our planet faces an uncertain ecological equilibrium, as affirmed by the Intergovernmental Panel on Climate Change (IPCC, Lee et al., 2023). Nowhere is more concern than in regions harboring endangered biodiversity hotspots like Colombia (Pabón, 2003; Ruiz et al., 2008), a country characterized by the convergence of ecological diversity and climate susceptibility. This work studies the far-reaching atmospheric interactions and their structure in Colombia through the construction of complex networks of climate teleconnections. This research aims to provide valuable insights for safeguarding the biodiversity-rich nation and offering valuable scientific information to face our evolving climate.

Teleconnections describe the relationships between climate anomalies at large distances (often thousands of kilometers). These linkages were first introduced in the climate context by Ångström (1935) to explain how the variability of large-scale atmospheric circulation patterns influence precipitation and temperature in tropical and extratropical regions, on various timescales (Lorenz, 1963; Schneider and Dickinson, 1974; Easterling et al., 2000; Donner et al., 2009; Solman, 2013). Due to climate's high complexity, one critical challenge in our time lies in distinguishing between natural climate fluctuations and human-induced climate change (Hulme et al., 1999). Understanding teleconnections is crucial for accurately predicting both climate variability and climate change (Kucharski et al., 2010). Different approaches have been proposed to quantify teleconnections, including atmospheric circulation models (ACMs) and statistical methods based on correlations and mutual information (Trenberth et al., 1998).

The Earth's climate is a constantly changing system driven by external forcings, such as the Earth's astronomical position relative to the sun or anthropogenic activities. It dissipates energy through various processes, exhibits unpredictable behavior (chaotic), and exists in a non-equilibrium state, meaning it is not perfectly balanced thermodynamically (Ghil and Lucarini, 2020). Natural variability arises from a complex interplay of amplifying and dampening effects (positive and negative feedback), alongside mechanisms that limit change (saturation mechanisms) and variables interacting at different spatiotemporal scales (Ghil, 2002). Therefore, a complex systems framework is a valuable tool for studying global climate and its teleconnections.

Complex network theory is a recently introduced framework for modeling climate teleconnections (Tsonis et al., 2006). For instance, Tsonis et al. (2008) employed a climate network to analyze

teleconnection patterns by calculating the Empirical Orthogonal Functions (EOFs) of observed extratropical 500-hPa flow. However, their method exhibits limitations in the tropics, as these regions appear to have a similar number of connections to all other locations on Earth. Conversely, Liu et al. (2023) established a teleconnection between the Amazonian rainforest, the Tibetan Plateau, and the West Antarctic ice sheet —exceeding 16,000 kilometers apart— by constructing a climate network using near-surface air temperature fields.

All methods, including complex networks, have to reflect all we already know about the physical mechanisms of teleconections, including The Tropical Oceans-Global Atmosphere (TOGA) diagnostic research (Schneider, 2006). For example, it is well known that the Atmosphere influences the Oceans primarily through anomalies in surface winds (Wallace et al., 1989; Hayes et al., 1989). In contrast, the Ocean influences the Atmosphere through anomalies in Surface Sea Temperatures (SST) and upward fluxes of latent and sensible heat (Peixoto and Oort, 1992). In the tropics, fluctuations in moist deep convection occur on timescales of weeks to months (intraseasonal) and are linked to the large-scale atmospheric circulation (Stan et al., 2017). This connection arises partly from the redistributed mass by tropical convection. This redistribution is associated with both global and regional overturning patterns and cycles of atmospheric angular momentum, and eastward and poleward propagating Rossby wave trains observed in the mid-latitudes (Sardeshmukh and Hoskins, 1988; Pinault, 2022).

This highly complex atmospheric process is modeled using General Circulation Models (GCMs) and Regional Climate Models (RCMs). These models are widely employed to understand and predict the climate state. They utilize the equations of motion derived from fluid mechanics and incorporate various parametrizations for physical processes, including cloud microphysics, radiation-matter interaction, chemical components, and more (Ghil et al., 2008). However, GCMs and RCMs require significant computational resources, restricting their accessibility primarily to well-funded, high-tech institutions that can manage and absorb the associated budget. Additionally, these models may not directly capture the full complexity of climate characteristics, such as teleconnections (Shackley et al., 1998; Claussen et al., 2002; Morrison and Lawrence, 2023). Therefore, the application of a low-cost, complex-systems-based framework that accounts for these climate characteristics becomes necessary.

This research is focused on a portion of Northern South America centered in Colombia, a biodiversity hotspot and second main holder of freshwater supply in the continent (Pacific, 2018). Its diverse topography, ranging from soaring Andean peaks to lush Amazonian lowlands, presents a kaleidoscope of climate scenarios. The nation's unique geographic position accentuates its susceptibility to climate variations, making it a living laboratory for researchers striving to decode the causality relationships of global variability atmospheric phenomena with the territory's climate evolution. Several studies have characterized the teleconnections with global variability like El Niño-Southern Oscillation (ENSO) (Poveda et al., 2002; Poveda et al., 2011; Hoyos et al., 2013a; Hoyos et al., 2018; Bedoya-Soto et al., 2019; Hoyos et al., 2019; Cai et al., 2020; Canchala et al., 2020b; Canchala et al.,

2020a; Arias et al., 2021; Cerón et al., 2021; Reboita et al., 2021; Builes-Jaramillo et al., 2023). Hence, this research's model can be validated through comparisons with existing research findings.

Unraveling the complex dynamic of Colombia's climate is not merely an academic pursuit but a vital necessity for sustainable development and adaptive strategies in the face of a changing climate. Recent research points out that global variability influences key activities like coffee production (Sanderson et al., 2022) an emblematic product from our land –representing around 8% of total exports– and the main income for over 2.2 million people (FedeCafeteros, 2020). Food security and other agricultural activities rely highly on the nation's capability to predict and management of extreme climatic events (Botero and Barnes, 2022). Additionally, 68% of electric production in Colombia depends entirely on Hydroelectric dams (XM, 2019), whose vulnerability to droughts or heavy rainfall events is critical.

The methodology is based on a fundamental characteristic of complex networks, particularly pertinent to climate networks: the presence of subsets of nodes that exhibit strong interconnections among themselves while maintaining weak connections to the broader network. These cohesive node sets are commonly referred to as *communities* (Fortunato, 2010; Newman et al., 2011). Communities assume significance in the context of complexity reduction and system coarse-graining as nodes within the same community can be considered a coherent component of a larger macro-system, whereas interconnections between nodes from different communities can be abstracted into links connecting these communities. This approach enables the construction of a network of communities, offering a simplified representation that encapsulates pertinent *macroscopic* insights from the original network.

This research explores a new method for identifying teleconnections in climate data. The proposed method leverages information transference quantifiers, a class of tools capable of analyzing time series data from climate variables and establishing causal relationships between them. Following the identification of teleconnections, this research delves further to unravel the underlying structure of these linkages. This is achieved by constructing a complex network representation of the Colombian climate system, where the identified teleconnections map onto connections between nodes within the territory's network.

The thesis is divided into four main parts:

1. **Theoretical Foundations (Chapters 1 & 2)**

    - Networks in Climate introduces key elements of complex network theory, including the specific graph properties that will be calculated in this study. It also provides a brief introduction to climate networks.

- Information Theory in Dynamical and Complex Systems delves into the concepts of causality and entropy. It then introduces the Liang-Kleeman Information Transference Maximum Likelihood Estimator (LKMLE), the key mathematical foundation for this thesis.

Note: This is a concise introduction. Readers seeking a more in-depth exploration of these topics are encouraged to consult the references provided throughout the text.

2. **Data and Methods (Chapters 3 & 4)**

   - Colombia Climate Characteristics introduces relevant climate characteristics of the study area, along with the global variability modes that will be analyzed. Additionally, it details the data sources, resolution, and timescale used in the research.
   - Beyond data: Constructing a Climate Network based on Information Transference describes the construction of the climate complex network based on the LKMLE.

3. **Results (Chapters 5 & 6)**

   - Mapping Teleconnections: Information Transference towards Colombia details the characteristics of the identified teleconnections and discusses similarities and differences with findings from other studies.
   - Understanding the Link Between Climate Variability and Complex Networks explores the constructed climate networks based on their properties.

4. **Finally, Conclusions and Perspectives**

The proposed methodology holds broad relevance for the physics community and various applied fields, owing to its universal applicability to spatially extended dynamical systems. Timely within the context of the ongoing scenario on climate change, this approach offers a novel perspective for evaluating the regional connectivity and complexity of the climate system, drawing upon global, rather than exclusively regional, knowledge. This thesis's methodology has the potential to make significant contributions to our comprehension of local responses to extreme events and tipping points in the Earth system within a holistic global framework.

Chapter 1

Networks in Climate

In this section, fundamental concepts of complex networks applied to climate are introduced. A summary of common construction methods, research findings, and applications within the realm of climate studies is provided. Subsequently, the primary properties used to calibrate the model are elaborated, focusing on their mathematical underpinnings and physical significance. Finally, the general methodology for constructing climate networks is presented.

1.1 Elements of Complex Networks Theory

In the realm of network science, the term complex networks belongs to a category derived from empirical observations and characterized by intricate topological features (Watts and Strogatz, 1998; Barabási and Albert, 1999). These networks serve as a powerful framework for unraveling the fundamental structural attributes inherent to diverse real-world systems, encompassing domains as broad as the internet, ecological ecosystems, neural connectivity patterns, interpersonal relationships, and climatological phenomena, among others (Strogatz, 2001; Albert and Barabási, 2002; Caldarelli, 2007; Barabási, 2009; Newman et al., 2011). The exploration and comprehension of complex networks contribute significantly to our capacity to figure out the complexities of the real world, providing a deeper understanding of the underlying dynamics in complex systems (Pósfai and Barabási, 2016).

In network theory, a network is a collection of elements, called nodes (vertices), connected by links (edges), as it is shown in Figure 1.1. These connections can be visualized as a graph. Links can be directed (indicating a specific flow or asymmetry) or undirected (symmetry). Additionally, networks can be weighted (with links carrying values) or unweighted. Finally, networks can be simple (excluding connections between a node and itself) or include self-loops (Barabási, 2016a). One of the main advantages of analyzing complex systems through complex networks is the ability to identify patterns in their dynamics (regular or irregular) based on the network's topology (Battiston et al., 2020). This approach offers an elegant and versatile tool for examining the interconnections, dependencies, and emergent behaviors between elements within a system (Mitchell, 2006). Moreover, by analyzing the network's structure, researchers can establish statistical and physical properties governing these

relationships, ultimately revealing intricate interdependencies within the system (Majhi et al., 2022).

The complexity of a network is related to its topological properties. Complex networks are often characterized by a vast number of non-regular connections. A central challenge within this modeling domain is the building of a network that faithfully replicates the actual connectivity inherent to the system under investigation. This requires careful consideration of methodological approaches. Two well-known and extensively studied classes of complex networks are scale-free networks (Barabási and Bonabeau, 2003; Barabási, 2009) and small-world networks (Watts, 2004). While these classes are the focus of this study, understanding the properties of random networks also has important mathematical and modeling implications (Erdős, Rényi, et al., 1959; Erdős, Rényi, et al., 1960). For further insight into random networks, readers can refer to Barabási (2016b) although this specific class of networks is not extensively used in this manuscript.

Scale-free networks were introduced by Barabási and Albert (1999) and have been widely studied after that. This type of network exhibits a distinctive structural pattern where a small fraction of nodes possess an exceptionally high number of connections (hubs), while the majority of nodes have relatively few links. This connectivity distribution follows a power-law, indicating a probabilistic decay in the likelihood of nodes having a certain number of connections. This phenomenon is observed in diverse real-world systems such as the World Wide Web (Pastor-Satorras and Vespignani, 2001), social networks (Ebel et al., 2002), and biological networks like protein-protein interactions (Albert, 2005). Notably, scale-free networks display increased robustness to random node failures but heightened vulnerability to targeted attacks on the highly connected hubs (Li et al., 2015).

On the other hand, small-world networks were introduced by Watts and Strogatz (1998) while exploring models that exhibit a connection topology that is neither fully regular nor fully random. Therefore, these networks are characterized by high local clustering and short average path lengths, forming a balance between local connectivity and global efficiency. This property is exemplified in various real-world systems such as social networks, neuronal networks in the brain (Bassett and Bullmore, 2017; Liao et al., 2017), and power grids (Pagani and Aiello, 2014). In social networks, for instance, individuals tend to be closely connected to their immediate contacts while maintaining short paths to any other person in the network (Arney, 2010).

Complex network understanding based on graph theory developments and several applications to diverse scientific and computational problems has become a novel and accessible modeling framework that unveils high-dimensional systems structure and connectivity (Newman, 2003; Boccaletti et al., 2006). Its wide scope of applications easily adaptable construction methods and variety of measures Strogatz (2001) make this a good complex-systems-based framework to study climate (Steinhaeuser et al., 2011). Previous research has discussed notorious climate networking challenges and

advantages that incentivize more novel developments in this area (Tsonis et al., 2006), especially selecting the appropriate linkage criteria and threshold selection (Steinhaeuser et al., 2010).

1.1.1 Complex Networks Properties

The complex network framework approach unveils a range of network measures, spanning from local metrics like vertex degree centrality (k_v), which quantifies the normalized number of first neighbors of a vertex, to mesoscopic properties like the clustering coefficient (C), and finally, to global indicators such as the average path length ($\mathcal{P}$) (Dijkstra et al., 2019). Local degree centrality and associated metrics provide an instrument to identify super-nodes, denoting regions with notably high degree centrality (Tsonis et al., 2006). In the subsequent discussion, we enumerate and detail the specific network properties that we prioritize to evaluate our climate network, emphasizing their significance in enhancing our comprehension of Colombia's climate dynamics. All definitions are based on Pósfai and Barabási (2016) and Newman (2018).

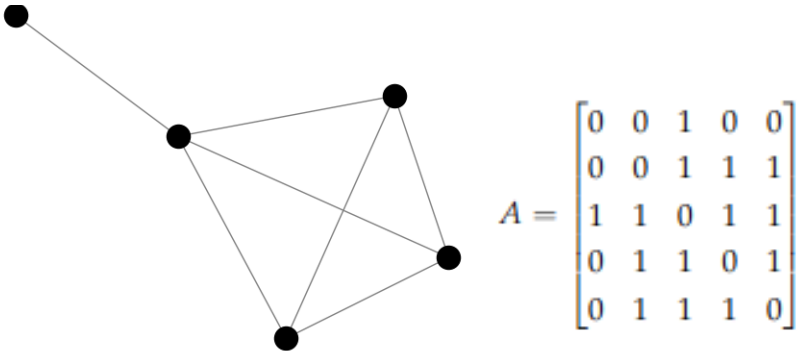


FIGURE 1.1: Simple undirected unweighted network representation for a graph with 5 nodes ($N = 5$) and 7 links ($L = 7$) (left) and its corresponding adjacency matrix (right).

Let N be the total number of nodes in the network and L the number of links, then the **graph density** for simple undirected graphs is defined as $\mathcal{D} = 2L/N(N-1)$. This metric serves as an indicator of the network's connectivity, with a maximum value of 1 indicating that every node is linked to every other node. In the realm of complex networks, densities fall below 0.5 (or even more), signifying that the network is not characterized by ubiquitous connections but is instead distinguished by the presence of shorter paths or *small-world* connections between nodes that promote efficient information transfer.

Consider that all the networks used in this work are simple unweighted graphs (see Figure 1.1), then their adjacency matrices $\boldsymbol{A}$ will have dimensions $N \times N$, with their elements $a_{ij} = 1$ if there exists a link between nodes i and j. Hence, its main diagonal $a_{ii} = 0$ for all $i = 1, \ldots, N$ and $\boldsymbol{A} = \boldsymbol{A}^T$. This matrix description facilitates the quantification of network properties (Marwan et al., 2009).

The **degree of the i-th node**, denoted by k_i, is the count of links associated with a given node, reflecting the number of links that either meet at or terminate at that node. A node exhibiting a degree of 0 is known as an isolated node. A node with a degree of 1 is referred to as a leaf node or end node. For an undirected graph, the total number of links (L) can be expressed as half the sum of all node's degrees (k_i):

$$L = \frac{1}{2}\sum_{i=1}^{N} k_i.$$

The average degree of the graph, often denoted by $\langle k \rangle$, given the degree of a node, can be calculated using Equation 1.1. This property holds significance in various domains such as information transmission, where a higher average degree fosters more efficient communication and robustness, as networks with a greater degree tend to be more resistant to random failures:

$$\langle k \rangle = \frac{2L}{N}. \tag{1.1}$$

For each node, the degree to which every linked node tends to group, or cluster can be measured using the local clustering coefficient C_i defined in Equation 1.2 and discussed in Wang et al. (2017), were is computed for a binary, undirected, simple graph (the same type of graph we construct in this research).

$$C_i = \frac{1}{k_i(k_i - 1)} \sum_{jk} A_{ij} A_{jk} A_{ki}. \tag{1.2}$$

Higher C_i indicates nodes that link with nodes that also link between them. For an undirected simple graph, the mean Clustering coefficient $\langle C \rangle$ can be found averaging over all nodes N:

$$\langle C \rangle = \frac{1}{N}\sum_{i=1}^{N} C_i. \tag{1.3}$$

According to Freitas et al. (2019), one suitable and clear measure of the network's degree distribution heterogeneity is the Network Entropy $\mathcal{H}$. They propose its calculation based on the normalization of the probability distribution of a random walk from node i towards node j:

$$\mathcal{H} = \frac{1}{N \log(N-1)} \sum_{i=1}^{N} \log k_i. \tag{1.4}$$

Thus, $\mathcal{H} \in [0, 1]$ where 0 indicates a sparse network and 1 a fully connected one.

It is also important to understand how close or far is one node from another. Hence, the **distance between two nodes u and v**, $d(u, v)$ is defined as the number of links in the shortest path that links them (this is also called a graph geodesic). When two nodes are disconnected, then its distance is set equal to infinity. The greatest distance from a node v to any other node within the graph is called the **node eccentricity** $\epsilon(v)$. This last concept provides insight into the relative centrality and spread of the

vertices within the network. The main focus is to discuss mainly the diameter and the radius of each graph, given by the maximum and the minimum eccentricity respectively.

The **diameter** represents the graph's overall size and maximal longest shortest path length, whereas the **radius** is the minimal shortest path length. Both metrics provide insights into the centrality and compactness of the network. A small diameter suggests effective overall connectivity, while a small radius indicates a more centrally concentrated and compact structure (see Figure 1.2).

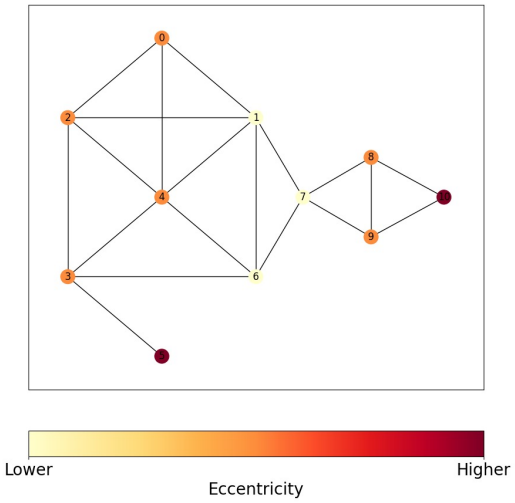


FIGURE 1.2: Illustration of the network's eccentricity measure intensity given in color scale. Circles represent nodes and lines show links.

To characterize each node in the network, a discussion of the degree, eigenvector, closeness, and betweenness centrality measures is proposed (see Figure 1.3). These will help identify spatial areas that play a key role in the information flux of the Colombian territory, the structure, and the influence of the calculated teleconnections patterns towards the study area. More details about these measures can be found in Newman (2010).

Eigenvector centrality, sometimes called *eigencentrality*, evaluates a node's importance in a network by considering both the quantity and quality of its connections. This measure is an extension of the basic degree centrality since it measures the connections for a node but recognizes that not all connections are equal and gives higher weight to nodes connected to other highly central nodes. In a network, nodes with high eigenvector centrality are considered influential not only for their direct connections but also for their connections to other influential nodes, making this measure valuable for identifying key players in complex networks and the influence dynamics. A practical illustration of this concept is discussed in Negre et al. (2018).

Closeness centrality C_i exhibits a close mathematical connection to the concept of node-to-node distance within a network. The distance between any two nodes is precisely defined as the length of the shortest path linking these nodes. Meanwhile, *farness* is the aggregate sum of distances from a given node to all other nodes in the network. Closeness centrality stands as the reciprocal of this farness measure. It effectively identifies nodes with the capability to efficiently access any other node within a limited number of steps, as well as nodes that may be positioned quite distantly from others within the graph. In essence, if the cumulative sum of distances from a node is substantial, the closeness centrality assumes a smaller value, and vice versa. Thus, a node characterized by high closeness

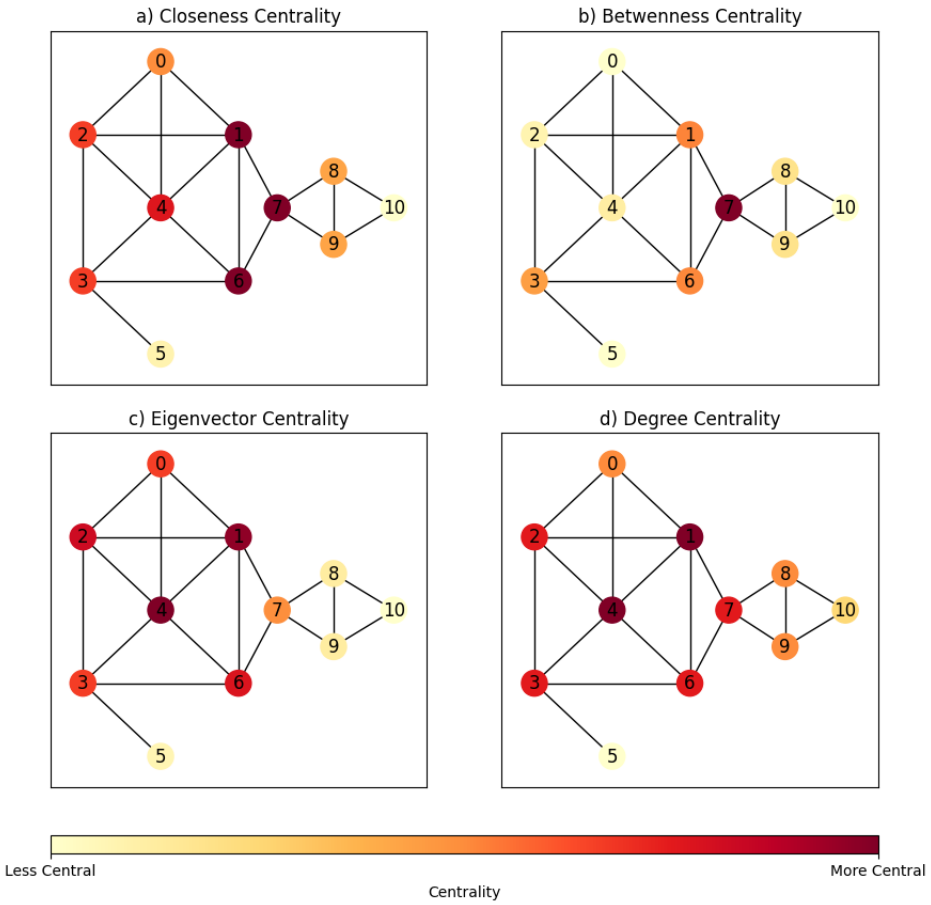


FIGURE 1.3: Illustration of the closeness, betweenness, eigenvector, and degree centrality measures used to understand a graph topology. Circles represent nodes and lines show links. Darker nodes represent higher centrality.

centrality implies its close relational proximity to a multitude of other vertices within the network. This measure is estimated by taking the *mean geodesic distance* from i to j, denoted by l_i:

$$l_i = \frac{1}{N} \sum_{j}^{N} d_{ij}. \tag{1.5}$$

Then, the closeness centrality is:

$$\mathcal{C}_i = \frac{1}{l_i}. \tag{1.6}$$

Finally, **Betweenness Centrality** quantifies the extent to which a target node, denoted as u, plays a pivotal role in the passage of shortest paths between all possible pairs of nodes within a graph. Mathematically, this score is calculated by considering the number of shortest paths, $\sigma_{v,w}(u)$, that traverse through the target node u (see Equation 1.7). The computation is then normalized by the total count of shortest paths, $\sigma_{v,w}$, existing between any pair of nodes in the graph. The betweenness

centrality of a particular node is higher when it is traversed by a larger number of shortest paths.

$$\mathcal{B}(u) = \sum_{u \neq v \neq w} \frac{\sigma_{v,w}(u)}{\sigma_{v,w}}. \tag{1.7}$$

The betweenness centrality measure serves as an indicator of the number of paths a node is situated upon, and further elucidates how many paths it actively participates in. In essence, it reflects the node's capacity to forge connections with various groups within the graph. This concept finds significant applicability in domains such as social media influence analysis (Golbeck, 2015).

1.2 Climate Networks

Climate networks are essentially complex networks. They are characterized by subsets of highly interconnected nodes that maintain relatively weak connections to the broader network. These subsets are commonly known as communities (Fortunato, 2010; Newman et al., 2011). In these networks, nodes represent geographical regions, and links represent relationships defined by a calculated metric between the climatic variables that describe each node. Previous work has shown that climate networks have small-world attributes on a global scale. This feature is attributed to the presence of long-range connections represented as edges linking geographically distant nodes. These long-range connections play a pivotal role in stabilizing and facilitating efficient information transfer within the climate system (Tsonis and Roebber, 2004; Tsonis et al., 2006; Tsonis and Swanson, 2008).

Climate networks can be categorized into three folds: i) ***Functional networks*** are created through correlation analysis, whether linear or nonlinear, employing techniques such as cross-correlations or mutual information. These networks may exhibit either undirected or directed links. The directionality of links is based on factors such as correlation time lags, measures of information transfer, or Granger causality. ii) ***Flow networks*** originate from the examination of transport processes within flows. iii) ***Event synchronization networks***, formed through the analysis of correlations and time delays associated with extreme events (Froyland et al., 2014; Baudena et al., 2022).

The fundamental procedure in establishment a climate network involves the computation of either the correlation function or mutual information between time series data associated with network nodes, followed by a thresholding process to determine the presence or absence of network links. This is a challenging task that can be achieved using available software like ClimNet (Deza and Ihshaish, 2015). However, construction methods and novel linkage algorithms need to be analyzed to improve efficiency and add spatio-temporal relationships to networks (Steinhaeuser et al., 2010).

Donges et al. (2009) conducted a comparative analysis of two construction methods for climate networks using HadCM3 Surface Air Temperature data, and their findings revealed only minor deviations in the resulting network characteristics when employing either Pearson correlation or mutual

information as the linking criterion (see Figure 1.4). Further research demonstrated that both data characteristics and linkage methods must be taken into account to increase the reliability of Climate Networks outputs (Hlinka et al., 2013).

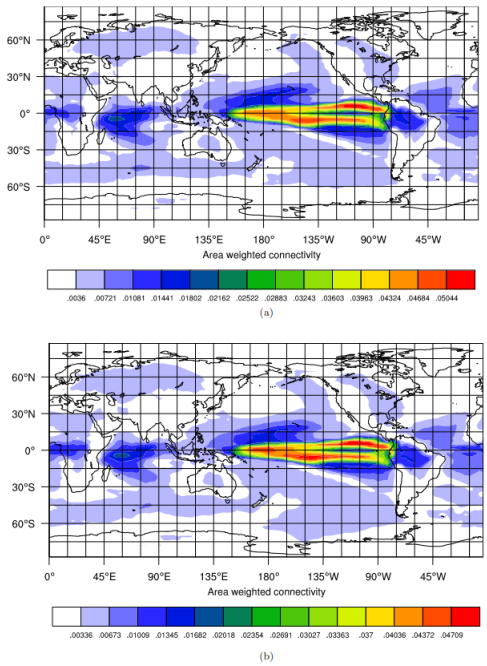


FIGURE 1.4: Example of climate network from Donges et al., 2009 represented using the Area weighted connectivity fields for global HadCM3 SAT networks using (a) Pearson correlation and (b) mutual information linkages criteria.

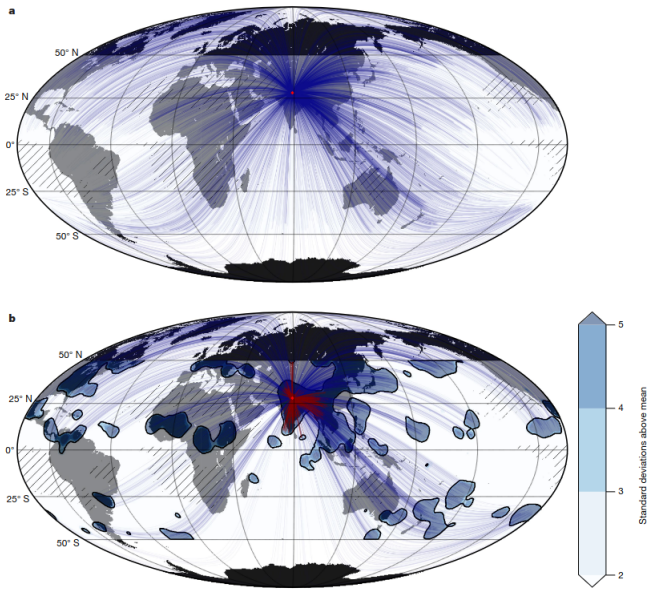


FIGURE 1.5: Example of Event Synchronization Network (ESN) from Boers et al., 2019 showing the south-central Asia teleconnection pattern for rainfall events above the 95th percentile.

The event synchronization (ES) method offers an alternative approach for constructing networks (ESNs) from climate observations. To illustrate this methodology, Boers et al. (2019) employed event synchronization to quantify the synchronicity of rainfall events in South-central Asia. They identified highly significant synchronizations, limited to a maximum time delay of ten days, resulting in the formation of a network encompassing 576.000 nodes, corresponding to the total spatial grid cells within the TRMM dataset. Network connections were established between pairs of nodes based on the significance level of the synchronization values, with a $p - value < 0.05$. Their analysis revealed significant structural patterns within the network, leading to the identification of teleconnections associated with Extreme-Rainfall Events (EREs) in the South-central Asian region (refer to Figure 1.5). This approach shows the utility of event synchronization in unveiling important climate-related phenomena and their interconnections.

Generally, the climate networks presented above adhere to the methodology proposed by Donges et al. (2015), where the construction of the climate network is divided into five stages: 1) selection of spatial grid, or node location and size, 2) selection of the climatological variables to study, 3) linkage criteria based on a statistical similarity measure and thresholding, 4) resulting network analysis based on graph theory measures and metrics, and 5) climatological interpretation of the inferred links and network topology (see Figure 1.6).

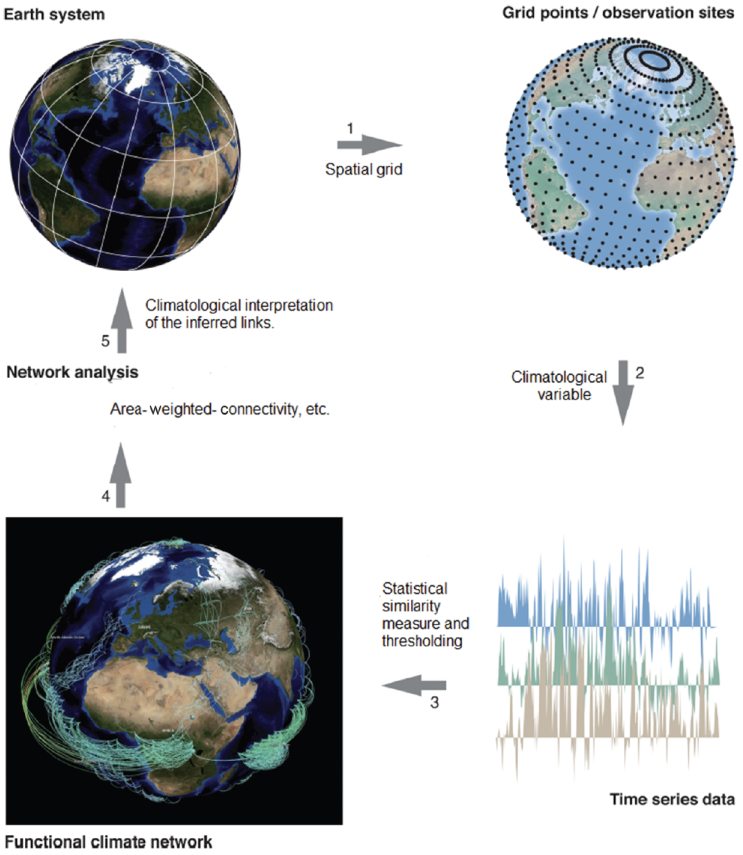


FIGURE 1.6: Methodology used to extract climate networks from climatic signals shown in Dijkstra et al. (2019). This thesis proposal changes only step 3 introducing the Liang-Kleeman causality measure to establish the network's links.

In consequence, the primary advancements and innovations in Climate network construction protocols focus on altering the linkage criteria. The traditional reliance on simple and linear statistical similarity measures has proven insufficient in grasping the intricacies of the climate system, often failing to capture the complex interactions and feedback loops inherent in climatic phenomena. To address this limitation, the incorporation of causality quantifiers provides an improvement in the representation of dynamical connections in climate networks. By doing so, the analysis moves beyond basic statistical correlations and immerses itself in the directional relationships among different

components of the climate system. This shift is particularly crucial for comprehending the dynamic nature of climate networks, enabling the identification of influential nodes and the discernment of causal links that actively shape the behavior of the system.

Chapter 2

Information Theory in Dynamical and Complex Systems

Entropy is a measurable physical quantity associated with the information content of a system from variables and connections that describe its dynamics (Volkenstein, 2009). The measure of information between two variables X and Y depends on their interactions and, in turn, indicates causality or the existence of coupling mechanisms. The transfer of information between the components of a dynamic system is an important concept for the analysis of nonlinear and multivariate time series. Many studies have focused on mutual information or entropy functionals to estimate the information transfer of nonlinear, stochastic, and highly complex systems such as neural connections in the brain (Heisz et al., 2012) or the weather (Bhaskar et al., 2017).

In this contribution, the main focus is on information theory-based estimators of causality, since they rely on the Probability Density Function (PDF) instead of the time series, so they can assess both linear and nonlinear statistical relationships between the ensemble prediction and observation. In contrast, conventional signal-to-noise ratio-based methods primarily gauge their linear correlation, leading to an underestimation of nonlinear statistical relationships (Kleeman, 2011).

This chapter provides an overview of the initial theoretical methodologies and mathematical advancements related to predictability, with a focus on the utilization of relative entropy. Relative entropy holds paramount significance in the context of the Liang-Kleeman formalism, which is detailed in Section 2.2, and which forms the foundational basis for constructing the climate network through an information theory framework within this thesis.

2.1 Relative Entropy and Causation

Entropy stands as a cornerstone in the realm of physics, intricately intertwined with the Second Law of Thermodynamics, which asserts that the total entropy of an isolated system evolves towards augmentation until reaches the thermodynamic equilibrium (Boltzmann, 1974; Lieb and Yngvason, 1999). The genesis of entropy as a theoretical construct harks back to the pioneering contributions

of Rudolf Clausius and Ludwig Boltzmann during the 19th century (Balibrea et al., 2016). Clausius (1866) conceived entropy as a property limiting the energy available to do mechanical work. In contrast, Boltzmann (1896) proffered a statistical interpretation of entropy, establishing a link with the microscopic configurations compatible with a macroscopic state. The Boltzmann entropy is defined as $S = k \log W$, where k denotes the Boltzmann constant, and W represents the count of microstates compatible with the macrostate of entropy S. Entropy finds multifaceted utility in fields such as information theory, where the pioneering work of Shannon (1938) provides a mathematical formulation of entropy, that quantifies information uncertainty (Shannon, 1948). Entropy is a foundational concept in the dynamics of physical systems and information processing, thereby yielding far-reaching implications across diverse scientific and engineering domains.

In the context of climate dynamics, the concept of entropy has been actively examined to evaluate the predictability of potential states that climate variables might traverse based on specific initial conditions. Pioneering research by Lorenz (1963) highlighted the high sensitivity of weather predictions to the initial conditions, leading to the realization that detailed weather forecasts become inherently impracticable beyond a certain temporal threshold –later research established to be approximately two weeks– (Krishnamurthy, 2019). This observation prompted in-depth investigations into the fundamental constraints governing the predictability of weather and climate conditions, even when employing idealized models. It is essential to underscore that the efficacy of weather forecasts is intricately linked to the performance of the underlying models, in addition to being influenced by factors such as the observational network, initialization procedures, and forecasting methodologies (Brotzge et al., 2023). Therefore, a comprehensive understanding of predictability requires an approach that acknowledges the inherent challenges posed by data inaccuracies and the inherently chaotic nature intrinsic to complex systems like the Earth's climate.

A novel approach, based on the amount of physical information transferred from one system to another under interaction, is useful to quantify the utility of a prediction (Liang, 2013a). In practical scenarios, prior knowledge typically exists concerning the historical or climatological behavior of the system, against which predictions are evaluated. Information theory offers a metric for this purpose, known as relative entropy, denoted as R (e.g., Cover and Thomas, 1991).

$$R = \sum_i p_i \ln \left(\frac{p_i}{q_i} \right), \tag{2.1}$$

where q_i is the climatological distribution and p_i is that for the prediction. This definition of R is also known as the Kullback-Leibler distance (Kullback and Leibler, 1951) since it measures the distance between the q and p distributions and only goes to zero when they are identical. Relative entropy quantifies the lack of information encoded when the assumption of climatology is upheld in the presence of an accessible prediction distribution. In essence, it gauges the informational gain achieved through prediction.

Kleeman (2002) introduced a novel measure of prediction utility for dynamical systems, rooted in the framework of a perfect model approach. This innovative measure served as the foundation for the subsequent Liang-Kleeman Maximum Likelihood Estimator (LKMLE) (Liang, 2013b), a central element characterized and employed as a linkage norm within the complex network framework explored in this thesis. Kleeman's approach drew upon the formal definition of relative entropy and incorporated considerations regarding the physical accuracy of prediction models. Additionally, it addressed the challenges inherent in determining the climatological (mean) state distribution in real-world scenarios.

2.2 Liang-Kleeman Information Transference

In recent years, Liang (2014) and Liang (2016) introduced a novel theoretical framework for establishing causal relationships between two time series employing an information transfer quantifier. This represents a conceptual advancement over conventional correlation analysis, including lag correlation, due to the asymmetric associations between correlation and causation (Pearl, 1997). In the field of climate studies, this approach offers the potential to define more robust dynamic teleconnections than those based on linear correlations. Specifically, it enables inquiries into whether a global climate phenomenon exerts influence on a local process. In this section, a summary of this method is provided, while readers seeking in-depth insights are encouraged to refer to Liang's seminal articles including (Liang, 2014; Liang, 2015; Liang, 2016).

Complex systems are fundamentally composed of a large number of subsystems that interact across different spatio-temporal scales. This allows access to the phase space to be statistically represented by a probability distribution, denoted by ρ. As highlighted by Schreiber (2000), the dynamics of these processes are captured by the transition probabilities. These probabilities correspond to the conditional probabilities within a Markov process. The transition probabilities themselves are determined by the dynamical interactions between the system's components. These interactions can be understood as an asymmetric transfer of information, meaning they are not necessarily equal in both directions between components. Information transfer serves as a measure of the dynamic influence one system exerts on another. In physics, this concept is referred to as causality (Liang, 2016).

Consider a bi-dimensional stochastic dynamical system composed by the processes $\boldsymbol{S} = \{X, Y\}$, its governing equation is given by:

$$\frac{d\boldsymbol{S}}{dt} = \boldsymbol{F}(\boldsymbol{S}, \boldsymbol{\theta}) + \boldsymbol{B}(\boldsymbol{S}, \boldsymbol{\theta}) \frac{d\boldsymbol{W}}{dt}, \tag{2.2}$$

where $\boldsymbol{F} = \{F_X, F_Y\}$ is the differentiable flux vector, $\boldsymbol{\theta}$ is the parameter vector, $\boldsymbol{B}$ is the diffusion coefficient matrix with elements b_{ij}, and $\boldsymbol{W}$ is the white noise vector. Let ρ be the joint probability

density of X and Y, and suppose that it and its derivatives have compact support. Following Liang (2008) and Liang (2014), the derivation of the information transfer from X to Y needs the marginal density of X written as:

$$\rho_X(t;X) = \int_{\mathbb{R}} \rho \ dY,$$

and the marginal (Shannon) entropy,

$$H_X = -\int_{\mathbb{R}} \rho_X \ln \rho_X \ dX,$$

where H_X varies as the system moves forward and the use of natural logarithm sets the information units in nats. The variation of the marginal entropy of X is due to two mechanisms, the one due to itself, named $dH_{X\not Y}$, and another due to Y, named $T_{Y\to X}$, the information transference from Y to X that we are interested in. Later in this section (subsection 2.2.1) we explore the influence of an additional mechanism called *the stochastic part* which accounts for other variables' influence and noise (Liang, 2015).

Considering the time evolution of the joint entropy H, a system following the Louville equation fulfills:

$$\frac{dH}{dt} = E(\nabla \cdot \boldsymbol{F}), \tag{2.3}$$

where E is the mathematical expectation with respect to ρ. Building on the detailed mathematical demonstration provided in Liang (2016), involving the evaluation of the Frobenius-Perron operator (Liang, 2013a) and the establishment of a kind of Fokker-Planck equation for the Y-excluded system (Liang, 2008), it can be established that for the 2-dimensional system, $dH_{X\not Y}/dt$ takes the form:

$$\frac{dH_{X\not Y}}{dt} = E\left[\frac{\partial F_X}{\partial X}\right], \tag{2.4}$$

and, the information transference for the 2-dimensional system is

$$T_{Y\to X} = -E\left[\frac{1}{\rho_x}\frac{\partial(F_x\rho_x)}{\partial_x}\right] + \frac{1}{2}E\left[\frac{1}{\rho_x}\frac{\partial^2(b_{xx}^2 + b_{xy}^2)\rho_x}{\partial x^2}\right]. \tag{2.5}$$

A value of $T_{Y\to X}$ equal to zero signifies the absence of causality from Y to X, while a nonzero value indicates that Y is causally connected to X, with information transfer flowing from Y to X. This information flow metric maintains an inherent asymmetry in the direction in which information flows.

The final form of Equation 2.5 depends on the governing dynamics of the specific system in question. Liang (2014) proposes the simplest linear system description $\boldsymbol{F} = \boldsymbol{f} + \boldsymbol{AS}$, where $\boldsymbol{f} = (f1, f2)^T$, $\boldsymbol{A} = (a_{ij})$, and $\boldsymbol{B}$ is a diagonal matrix with values b_1 and b_2. If the initial state of the variables (X, Y) follow a bivariate Gaussian distribution, then it is easy to find a re-writing of equation 2.5 in terms of the now known marginal density of X, its variance and mean. From there, the maximum likelihood

estimation (MLE) is used to estimate the parameters $\boldsymbol{f}$, $\boldsymbol{A}$ and $\boldsymbol{B}$ from the time series of X and Y. The corresponding mathematical derivation is left to read on pages 3 to 4 of Liang (2014).

Finally, the Liang-Kleeman Maximum Likelihood Estimator of Information Transference flow rate (LKMLE for simplicity) from Y to X can be written as:

$$T_{Y\to X} = \frac{C_{XX}C_{XY}C_{Y\dot{X}} - C_{XX}C_{X\dot{X}}}{C_{XX}^2 C_{YY} - C_{XX}C_{XY}^2}. \tag{2.6}$$

where C_{XY} is the sample covariance between the time series of the processes, and $C_{Y\dot{X}}$ is the sample covariance between Y and the first time derivative of X. This is an asymmetric quantifier that allows the distinction of information flow directionality and a nonzero information transfer from Y to X implies causality from Y to X. If the information transfer flow from Y to X is zero ($T_{Y\to X} = 0$), then X is not causally related to Y.

The sign of a nonzero information transference can be either positive or negative. If $T_{Y\to X} < 0$ then Y is said to indeterminate X, i.e., Y expands the possible accessible space states for X; and, if $T_{Y\to X} > 0$ then Y is said that make X more predictable, i.e., Y limits the possible accessible space states for X. In general, the asymmetry nature of this quantifier implies that $T_{Y\to X} \neq T_{X\to Y}$, thus each information flow direction must be analyzed separately. In this thesis, emphasis is placed on determining the direction of transference information flow from oceanic indices (Y) towards local climate variables (X) so that teleconnections can be established.

Within the framework of our calculations utilizing the LKMLE, a negative $T_{Y\to X}$ signifies a decrease in the change of entropy over time compared to the contribution from other sources. A decrease in relative entropy compared to other sources implies that the negative information flow from one variable to another, say from variable Y to variable X, pushes X away from its most probable state of maximum entropy. Consequently, the number of accessible microscopic configurations (or equivalently, accessible states) of variable X increases due to the influence of Y. An increase in the accessible states for variable X due to the influence of Y means that Y's action on X makes it more uncertain. In terms of predictability, this type of teleconnection can be seen as an indicator of sources of uncertainty, as variable Y serves as a source of uncertainty for variable X.

As noted by Kleeman (2011) regarding the predictability of dynamical systems, a source of uncertainty, in the case of the teleconnections found, can be understood as those places (and variables) where small variations significantly increase the possible states (values) that another place (variable) can take. Thus, classical determinism is lost. Increasing our understanding of sources of uncertainty can help better determine the probability distribution of the possible accessible states of those variables it affects. And this is what we mathematically seek with an improvement in predictability.

The LKMLE $T_{Y\to X}$ is a valuable instrument for unveiling complex spatial interconnections across the Earth (Vannitsem and Liang, 2022). A notable illustration of this application is found in the work of Hagan et al. (2019), who extended the Liang-Kleeman formalism to encompass a time-varying variant. This time-varying formalism retains analogous properties to its time-invariant counterpart and is applied in causal relationships within the context of soil moisture–air temperature coupling. Their application of this approach yielded successful results, underscoring the potential of the Liang-Kleeman formalism as a versatile analytical tool for unravelling complex causal structures in climate dynamics.

2.2.1 Normalized Information Transference

In subsequent developments, Liang (2015) proposes a normalization to the information transference flow, denoted by $Z_{Y\to X}$, to assess the importance of the influence of Y on X relative to other processes. By considering a similar fundamental deduction as the one employed in Liang (2016), they distinguish between three types of mechanisms that contribute to the evolution of the marginal entropy of X. Let H_X^{noise}/dt be the time evolution of stochastic effects, H_X^*/dt be the time rate of change of H_X due to itself in the absence of stochasticity, and $T_{Y\to X}$ the information transference flow from Y to X deduced previously in this section, then, by considering the evolution of X and Y in terms of this decomposition (see Figure 2.1), the normalization term goes by:

$$Z_{Y\to X} \equiv |T_{Y\to X}| + \left|\frac{dH_X^*}{dt}\right| + \left|\frac{dH_X^{noise}}{dt}\right| \tag{2.7}$$

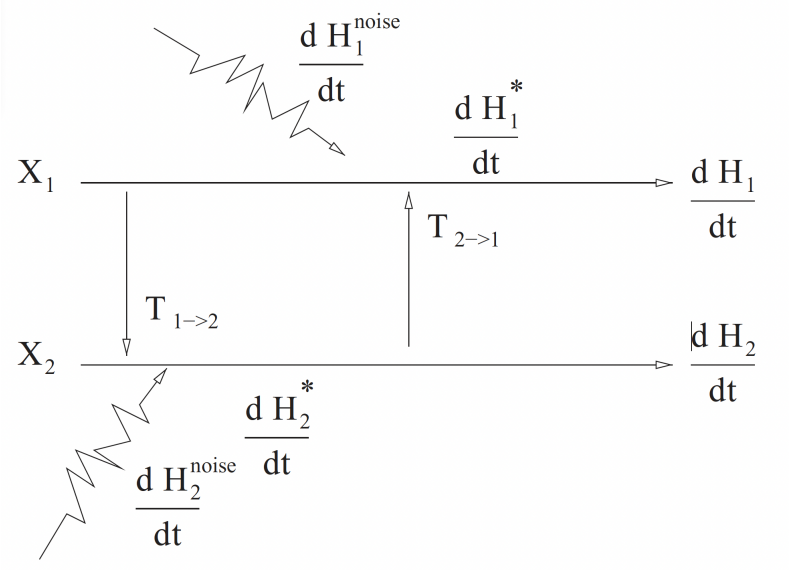


FIGURE 2.1: Schematic of the marginal entropy evolution and information flows in the system of X (X_1) and Y (X_2) from Liang (2015).

This $Z_{Y\to X}$ is no less than $T_{Y\to X}$ in magnitude and only sets to zero if X does not evolve in time –which is excluded within this analysis–. Therefore, the relative information flow is defined as:

$$\tau_{Y\to X} = \frac{T_{Y\to X}}{Z_{Y\to X}}. \tag{2.8}$$

If $\tau_{Y\to X} = 1$, the variation of H_X is 100% due to the information flow from Y; if $\tau_{Y\to X} = 0$, Y is not the cause. This relative information flow may only be used to compare effects for the same variable's time series, since comparison of the information flow between the two series, i.e. $\tau_{Y\to X}$ and $\tau_{X\to Y}$, does not belong to the same normalization considerations given the asymmetry of the proposed estimator.

In the context of this thesis, we will use both, the net information transference $T_{Y\to X}$ and the relative $\tau_{Y\to X}$, to map Colombia's teleconnections with global variability phenomena and to build our proposed information transference-based complex climate network. The latter was chosen as the similarity measure for linkage establishment in the network since, as mentioned before, $\tau_{Y\to X}$ assesses the importance of the influence of the studied global variability on the local climate variables in a $[0, 1]$ interval that makes easier the thresholding of minimum information transference.

Chapter 3

Colombia Climate Characteristics

This section describes the study area, with a focus on Colombia as the target country, encompassing crucial elements such as its topography, ecological characteristics, and socio-economic conditions. Also, a description of the global climate variability modes is presented, which exert a substantial influence on moisture flux patterns in this country of Northern South America. Finally, the six climate indices representing the global phenomena related to anomalies over both the Pacific and Atlantic Oceans are presented.

3.1 The Study Area and Data

Colombia's climate is featured by a convergence of unique characteristics that collectively are globally meaningful. This country hosts high biodiversity and high biological connectivity (Myers et al., 2000; Bass et al., 2010; Sánchez-Cuervo et al., 2012), and stands as the sixth country globally in freshwater resources -and the second in South America- (Pacific, 2018), playing a pivotal role in the global hydrological cycle. The Colombian terrain experiences a consistent influx of solar radiation that leads to convective atmospheric processes dominating short-term dynamics, resulting in a smaller variation in the annual temperature cycle compared to the diurnal cycle (Mesa S et al., 1997; IDEAM and Desarrollo Territorial, 2005; Hoyos et al., 2013b).

Colombian climate unfolds through a diverse topography, shaped prominently by the Andes mountain, which splits into three branches, contributing to a varied landscape, including snow peaks, plateaus, canyons, rainforests, and valleys, that impact atmospheric dynamics and regional climate patterns (Espinoza et al., 2020; Poveda et al., 2020; Arias et al., 2021), as well as diverse ecological ecosystems such as tropical moist rainforests, tropical grasslands, tropical dry forests, deserts, and mangroves (Sánchez-Cuervo et al., 2012; Aldana-Domínguez et al., 2017). Seasonality is chiefly defined by the annual rainfall regime, capturing the migration of the Intertropical Convergence Zone (ITCZ) and its interaction with regional orography (Poveda et al., 2011; Urrea et al., 2019; Pabón-Caicedo et al., 2020).

Colombia is surrounded by the basins of the Amazon and Orinoco rivers, the Pacific Ocean, and the Caribbean Sea. The complex interplay of these geographical factors and regional circulation results in diverse rainfall patterns, atmospheric transport across the Americas, and a heightened regional sensitivity to global climate phenomena across various temporal scales (Restrepo and Kjerfve, 2000; Pabón, 2003; Sakamoto et al., 2011; Poveda et al., 2011; Hoyos et al., 2013a; Arias et al., 2015; Cordoba-Machado et al., 2015; Córdoba-Machado et al., 2015; Hoyos et al., 2018; Martinez et al., 2019; Escobar et al., 2022). The study area is particularly sensitive to the variability modes of the Atlantic and Pacific oceans, as detailed in Hoyos et al. (2019).

Colombia has one the most climate change-endangered ecosystems, such as the Páramos, which harbor unique flora and fauna, further accentuating the nation's role as a living laboratory for studying climate-induced impacts on specialized habitats and freshwater supply (Díaz-Granados Ortiz et al., 2005; Rincón, 2015). Understanding Colombia's climate is important not only for local environmental stewardship but also for its broader implications in the context of global climate research, freshwater resource management, and biodiversity conservation strategies (Poveda et al., 2020).

The geographical scope of this thesis encompasses northern South America, a region delineated within the coordinates of 5°S to 15°N latitude and 90°W to 60°W longitude, centered on Colombia (Figure 3.1) and encompassing Nicaragua, Costa Rica, Panama, Venezuela, and Ecuador, along with a northeastern section of Peru and northwestern Brazil.

A homogeneous coverage of data across the study area ensures the robust construction of teleconnection patterns through information transference. The Liang-Kleeman Maximum Likelihood Estimator (LKMLE) requires long-term and well-behaved time series for both the climate variables and global indices. In the present work, data from the ERA5 reanalysis, the fifth-generation dataset from the European Centre for Medium-Range Weather Forecasts (ECMWF), is included as a regional climate database. This reanalysis offers hourly estimates for diverse atmospheric, land, and oceanic climate variables, covering the Earth on a 30 km grid with vertical resolution spanning 137 levels from the surface to 80 km height. The dataset spans from 1959 to the present (Hersbach et al., 2020a; Hersbach et al., 2020b). In this study, the regional climate state is represented in terms of a set of target variables: two-meter temperature (T2M), total precipitation (TP), and vertical convergence of moisture flux (VIMFLUX) for the time window 1950 to 2020.

3.2 Representing the Influence of Global Variability in Colombia

The intrinsic Ocean-Atmosphere coupled dynamics, which drive the entire Earth's circulation patterns, exhibit a pronounced reliance on various timescale-dependent variability modes distributed across the globe. For instance, some of the primary sources of variability arising from anomalies

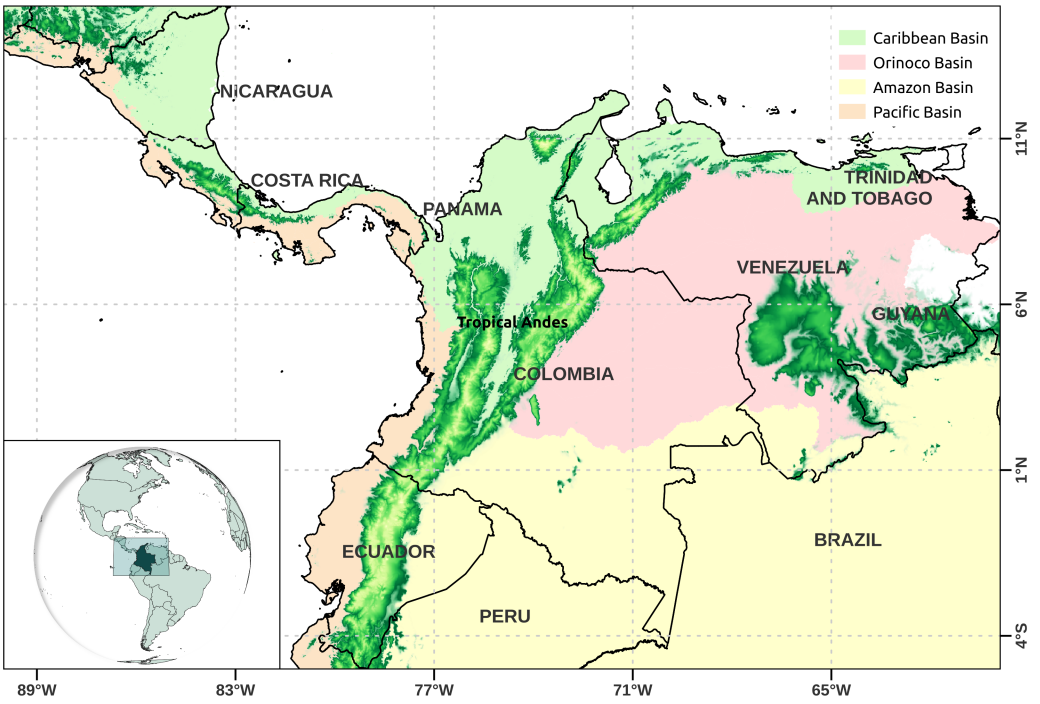


FIGURE 3.1: Study area with Orography and Basins of interest used to locate the discussion of the results.

in this ocean-atmosphere system that significantly influence Colombia's climate are those occurring within the influence regions of the El Niño-Southern Oscillation (ENSO), the Atlantic Multidecadal Oscillation (AMO), the Tropical Atlantic Dipole (TAD), the South Atlantic Dipole (SAD), and the annual oscillations of the Intertropical Convergence Zone (ITCZ), which traverses the northern part of South America (Schneider et al., 2014).

The predominant sources of precipitation seasonality and spatial variations within the study area are chiefly influenced by regional atmospheric transport and its interaction with local topography (Hoyos et al., 2018). Subsequent investigations have revealed the key moisture flux-related indices for Colombia, with TNA, NTA, and CAR emerging as the most causally linked variables on the primary mode of Moisture Flux Divergence (MFD) variability, while Niño 3, Niño 3.4, and SOI hold prominence on the second mode of MFD variability (Hoyos et al., 2019). This can be attributed to the heightened susceptibility of Colombia's coastline to changes in the Caribbean Atlantic Ocean and the heterogeneous ENSO response in the continental area. Consequently, a more intricate analysis is warranted to provide a comprehensive depiction of ENSO effects on the region.

For this study, a set of six climate indices was chosen based on Hoyos et al. (2019). They found that the following present the most substantial contributions to the rate of information flow between

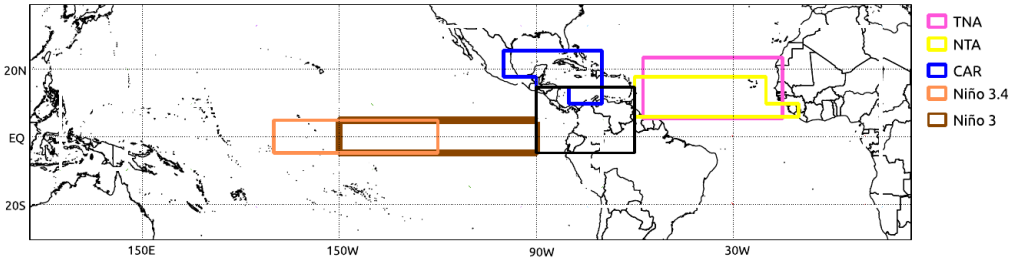


FIGURE 3.2: Extension of the area involved in climate index calculation for TNA, NTA, CAR, Niño 3.4, and Niño3. Highlighted in black is the study area.

oceanic indices, specifically linked to the variability modes of global climate dynamics, and the first three Principal Components (PCs) characterizing the regional moisture flux divergence field. The six indices, distributed by the National Oceanic and Atmospheric Administration (NOAA), include Niño 3, Niño 3.4, Southern Oscillation Index (SOI), Caribbean SST Index (CAR), Tropical North Atlantic (TNA), and North Tropical Atlantic (NTA).

Indices related to Atlantic (Pacific) variability: TNA, NTA, AMO, and CAR (the Niño 3, Niño 3.4, and SOI). Indices CAR, TNA, NTA, Niño 3.4, and Niño 3 are calculated using SST anomalies over the specific regions shown in Figure 3.2. SOI is based on the observed sea level pressure (SLP) differences between Tahiti and Darwin, Australia.

Chapter 4

Beyond data: Constructing a Climate Network based on Information Transference

This chapter presents the details of the methodology used throughout this thesis. This framework follows the general Climate Network construction steps suggested by Dijkstra et al. (2019), where the statistical measure used to construct the adjacency matrix was replaced with the relative information transference flow. The primary goal of this change is to quantify the regions in Colombia connected by the shared importance of causality from global phenomena towards the local climate variables.

4.1 Information transference network construction details

The procedure to construct our proposed climate network is illustrated in Figure 4.1, where the study area is selected. In this case, the area consists of a portion of Northern South America that includes the Colombian territory described in section 3.1. This region defines 9801 nodes corresponding to each one of the $0.25^\circ \times 0.25^\circ$ grid squares over the coordinates -90°W -60°W and 15°N 5°S, following the spatial resolution of the ERA5 reanalysis. The time series for the three climatic variables (TP, T2M and VIMFLUX) and the six global climate variability indices (CAR, TNA, NTA, Niño 3, Niño 3.4, SOI) have monthly time-scale resolution over the period 1950-2022.

The causal similarity measure from the relative information transference flow from the global climate indexes towards the gridded climate variables is used for establishing the links between nodes (or grid cells). A 0.5% threshold is suggested for $\tau_{2->1}$, such that $\tau_{2->1} \in (-0.005, 0.005)$ are set as zero in the building process. Additionally, two different nodes i, j are connected if, for a given index h and *any* variable k, the relative information transference for that node ($\tau^i_{h\to k}$) is within a 0.005 distance of another node's $\tau^j_{h\to k}$, so for each oceanic index h the elements of the adjacency matrix A of its

associated causal climate network follows the linkage criterion for $i \neq j$:

$$a_{ij}^{(h)} = \begin{cases} 1, & \text{if } \tau_{h\to k}^{i} \in [\tau_{h\to k}^{j} - 0.005, \tau_{h\to k}^{j} + 0.005] \ \forall k \\ 0, & \text{otherwise} \end{cases}. \tag{4.1}$$

As noted in 4.1, the diagonal elements are zero. This reflects our focus on inter-location relationships, not self-causal effects, which could be a topic for future research. The threshold of 0.005 was chosen based on a statistical analysis that indicated it represents the minimum detectable influence within the study area using the LKMLE quantifier.

The obtained adjacency matrices are symmetric and their corresponding graphs are simple, unweighted and undirected. Each graph captures a single index's overall influence on the three studied variables (TP, T2M, and VIMFLUX) within the study area. Consequently, we have six distinct networks, one for each index (CAR, TNA, NTA, Niño 3, Niño 3.4, and SOI).

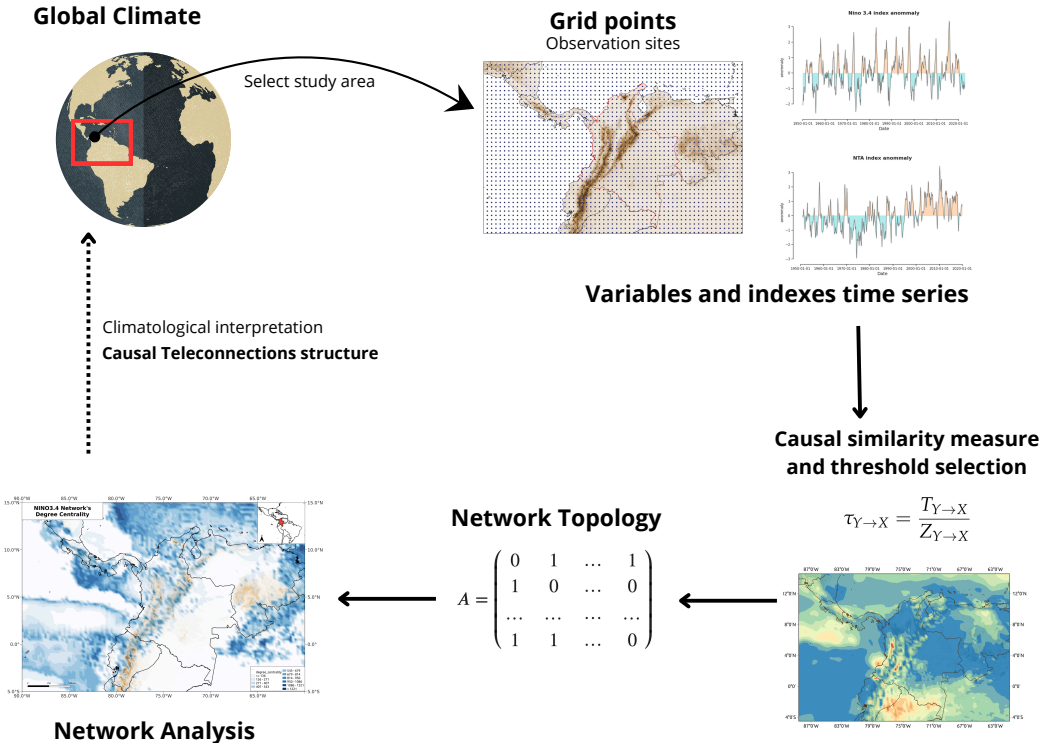


FIGURE 4.1: Scheme of the step-by-step process to build the proposed climate networks based on information transference. Modified from (Dijkstra et al., 2019).

Chapter 5

Mapping Teleconnections: Information Transference towards Colombia

This chapter presents the results of the computed Liang-Kleeman information transfer flows for all indices and variables. Starting with the general description of the spatial distribution of the net information transference flow and the meaning given by its sign, the objective is to understand the climate teleconnections of the Colombian territory. Then, the approximate distribution of these values is shown and the differences between each index for the same variable are compared. Finally, the results for the relative information transference values are presented to give insight into the importance of these long-range phenomena compared to noise and self-causality for each variable.

5.1 Colombian climate unveiled from a teleconnection perspective

Figure 5.1 presents a visual summary of the regional information transference for all indices and variables. These results reveal a high variability of the information transfer flow across the study area, represented by a large difference between the maximum and minimum values. Moreover, there is a considerable number of extreme values, shown by the outliers. This section focuses on describing the spatial distribution of the LKMLE outliers. Values near zero for Information Transfer do not definitively establish causal relationships. In the later section 5.2, by studying the normalized LKMLE, we can clearly establish.

A high amount of almost zero information transference can be seen in Figure 5.2. Here, high kurtosis around zero and positive skewness show the dominance of negative values, which implies an indetermination of the state space of the variable. This also shows that the majority of nodes do not have a clear causal relationship through LKMLE.

The information transference flow rate for VIMFLUX is mainly negative for all indices, showing that the overall causal relation between the studied indices and this variable is to indeterminate its state. Niño 3 and Niño 3.4 display the highest magnitudes of $T_{Y \to VIMFLUX}$ being -165.37 and -176.06

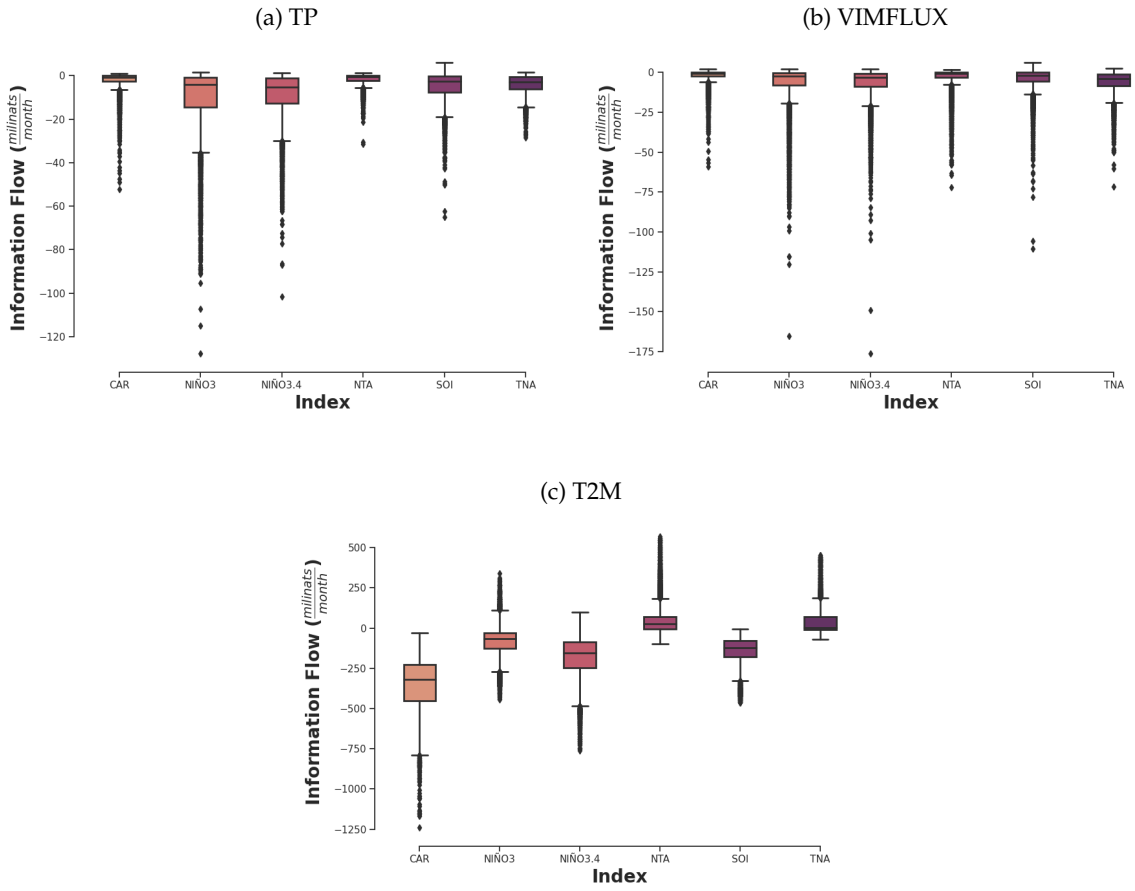


FIGURE 5.1: Boxplot illustrating the spread of (a) TP, (b) VIMFLUX and (c) T2M variables information transference flow from the CAR, Niño 3, Niño 3.4, NTA, SOI and TNA indices. The box represents the middle 50% of the data, with the line inside indicating the median. Whiskers show the overall range, and any dots outside indicate outliers.

respectively. The outliers below the lower limit for these indices indicate that the higher negative values for these two indices are located along an area over the western branch of the Colombian Andes, the Colombian Pacific Ocean coastline near Panama (see Figures 5.3e and 5.3d), and another one similar to the Chocó Low-Level Jet (Chocó LLJ) moisture path of moisture income (Yepes et al., 2019).

Niño 3 and Niño 3.4 LKMLE distributions for all indices exhibit the same characteristics, with long heavy tails of negative values and sharped peaks around zero for the $T_{Y \to VIMFLUX}$ (see Figure 5.2). Remarkable differences state in the minimum values, being CAR the lowest with -58.89 milinats per month (Figure 5.1b). Nevertheless, those small values cover small punctual areas almost not appreciated in the map (see Figure 5.3a).

It can also be noticed that Niño 3 and Niño 3.4 exhibit small and zero values of information transference over the Orinoco, Amazon, and Caribbean Basins, meaning no causal relationship over wide areas on the western side of the Andes and in Northern Colombia (see Figures 5.3d and 5.3e).

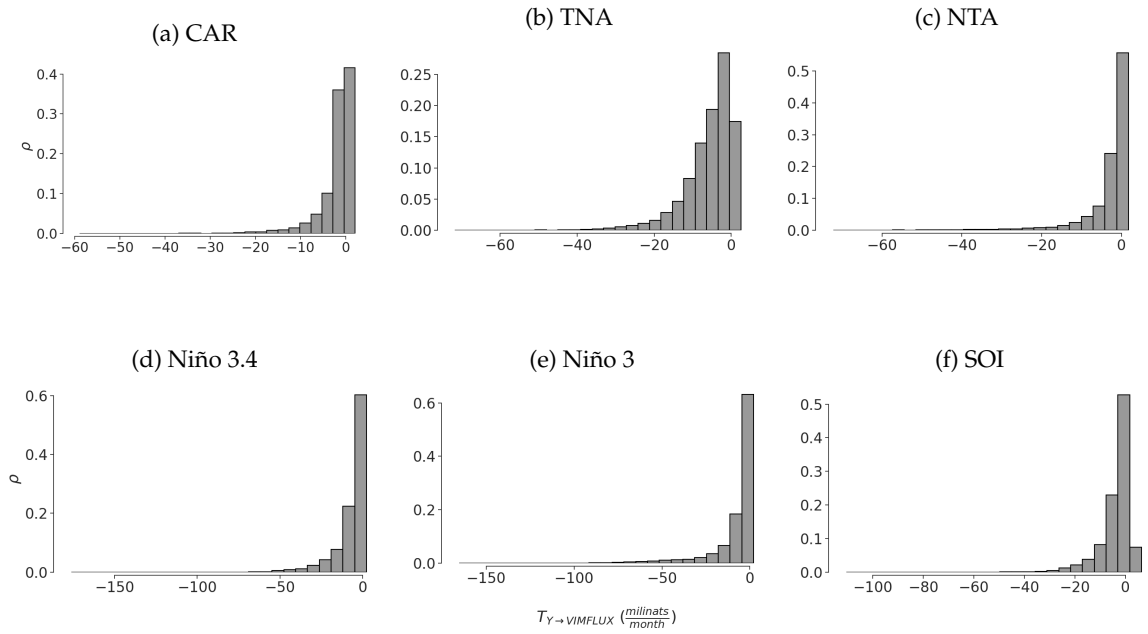


FIGURE 5.2: Information transference probability distribution for the vertical integral of moisture flux (VIMFLUX) from the (a) CAR, (b) TNA, (c) NTA, (d) Niño 3.4, (e) Niño 3, and (f) SOI indices in milinats per month. The majority of data is centred around zero.

TNA and NTA $T_{Y\to VIMFLUX}$ have similar minimums, -71.71 and -72.17 milinats per month respectively, and maximum values, 2.52 and 1.76 milinats per month respectively. For TNA, the negative outliers are located in northern Peru and some areas around the Maracaibo Lake in Venezuela and la Guajira in Colombia (see 5.3b). NTA outliers cover small areas similar to the ones of TNA over the Caribbean Basin (see 5.3c).

Outliers of the ENSO-related indices (Niño 3, Niño 3.4, and SOI) $T_{Y\to TP}$ are located over the Colombian Andes, the Cold Tongue of the Pacific and the Caribbean near Venezuela's northern coast (see Figures 5.4e, 5.4d, 5.4f. This suggests a causal teleconnection between high-altitude mountains located in the Andes and ENSO, linking the indetermination of precipitation over the portion of territory most densely populated with this global variability phenomenon. SOI also presents a low positive $T_{Y\to TP}$ towards TP over a small portion of the Amazon basin in southern Colombia (see Figure 5.4f).

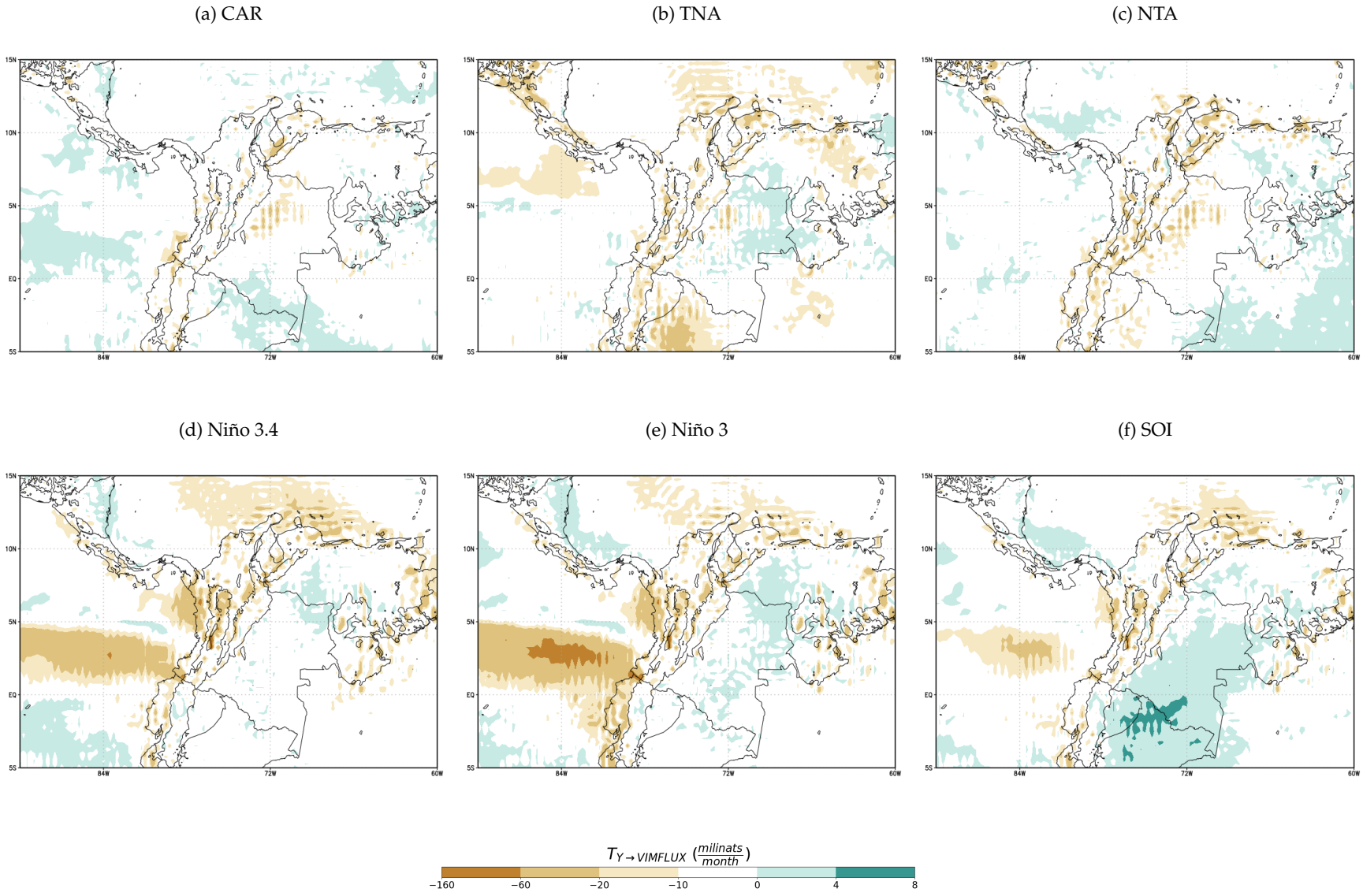


FIGURE 5.3: Information transference towards vertical convergence of moisture flux (VIMFLUX) from the (a) CAR, (b) TNA, (c) NTA, (d) Niño 3.4, (e) Niño 3 and (f) SOI indices in milinats per month. Darker zones indicate outlier values for $T_{Y \to VIMFLUX}$.

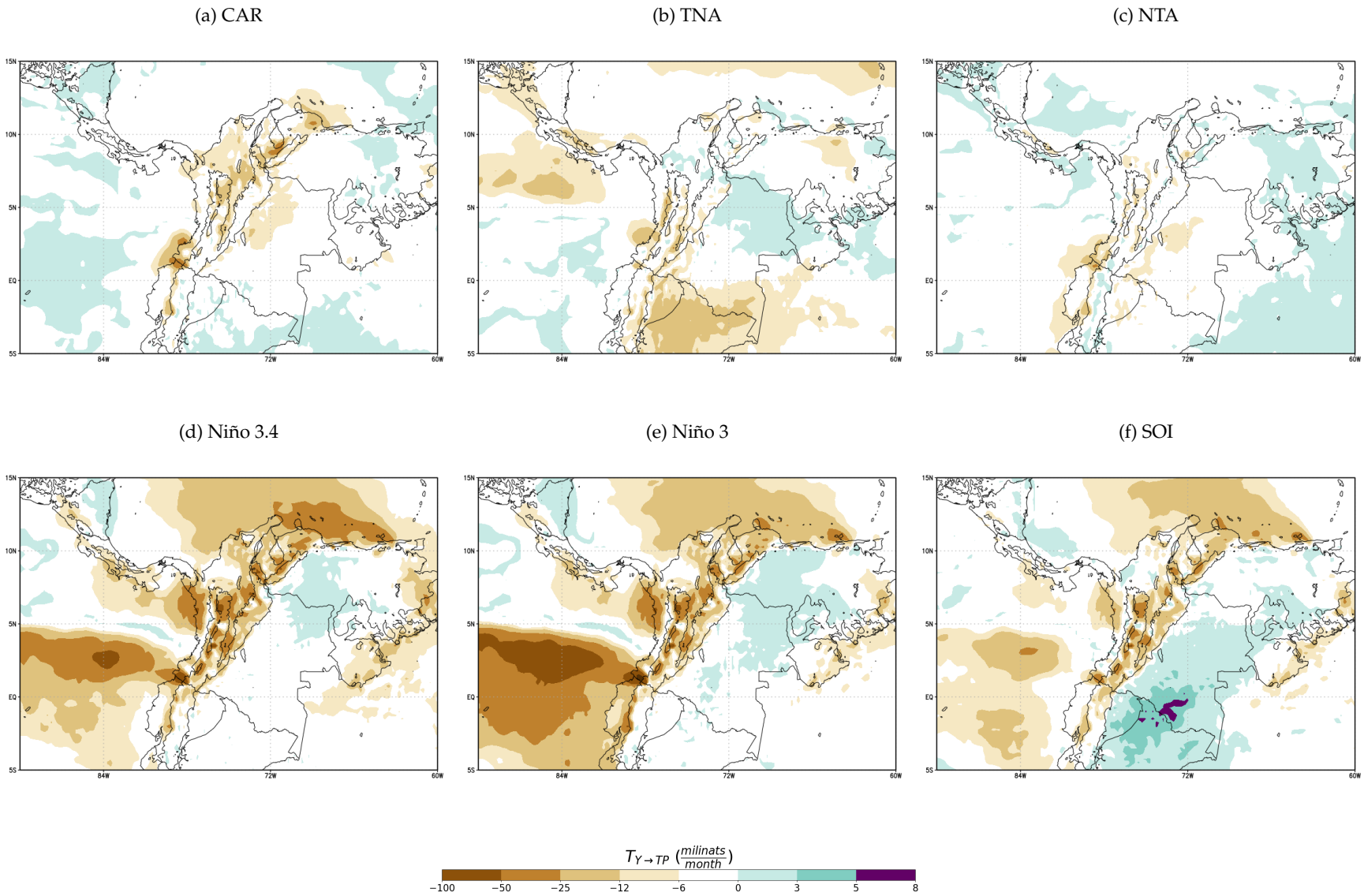
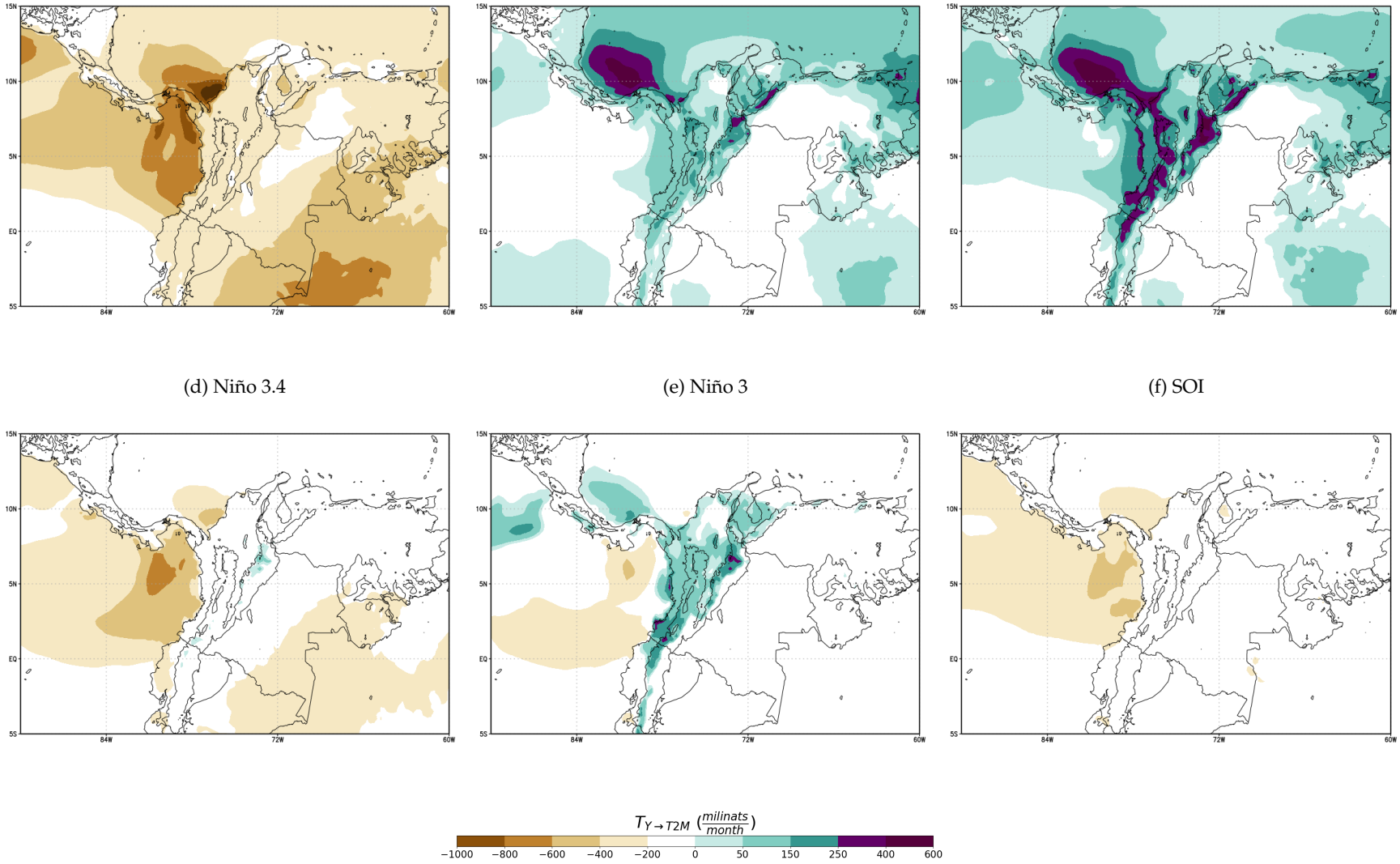


FIGURE 5.4: Information transference towards Total Precipitation (TP) from the (a) CAR, (b) TNA, (c) NTA, (d) Niño 3.4, (e) Niño 3 and (f) SOI indices in milinats per month. Darker zones indicate outlier values for $T_{Y \to TP}$.

(a) CAR (b) TNA (c) NTA



(d) Niño 3.4 (e) Niño 3 (f) SOI

FIGURE 5.5: Information transference towards two-meter Temperature (T2M) from the (a) CAR, (b) TNA, (c) NTA, (d) Niño 3.4, (e) Niño 3 and (f) SOI indices in milinats per month. Darker zones indicate outlier values for $T_{Y \to T2M}$.

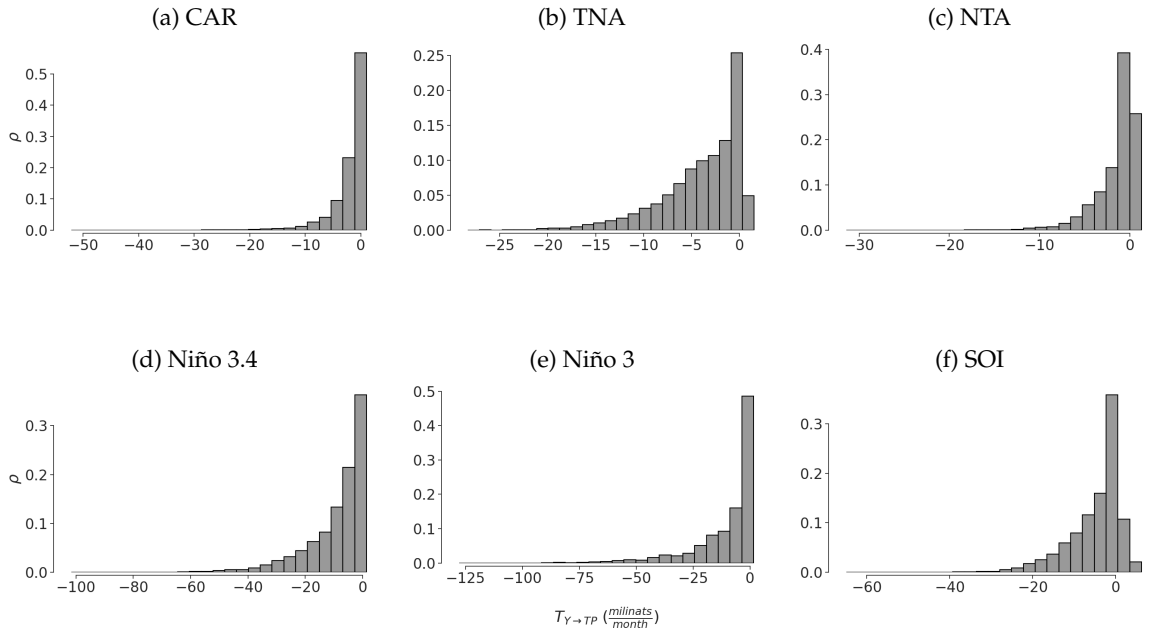


FIGURE 5.6: Information transference probability distribution for the Total Precipitation (TP) from the (a) CAR, (b) TNA, (c) NTA, (d) Niño 3.4, (e) Niño 3 and (f) SOI indices in milinats per month. The majority of data is centred around zero.

TP shows a spatial pattern in the LKMLE analysis that closely resembles the pattern for VIMFLUX, where the $T_{Y\to TP}$ sign decreases the predictability of this variable. In this case, TNA and NTA hold the lower minimums with -28.31 and -31.41 milinats per month respectively. Their distributions also have kurtosis around zero and positive skewness but, the TNA trend towards zero is softer than the other indices (see Figure 5.6b).

Different influence zones are found for the Atlantic's indices, with CAR being the one with higher negative values but a smaller influence area surrounding only the Colombia-Ecuador western frontier and the Maracaibo Lake in Venezuela (see Figure 5.4a). TNA negative valued outliers for $T_{Y\to TP}$ are located over the Amazon Basin, with values between -14 and -28 milinats per month, establishing a teleconnection between this portion of an important South America macro basin and the Atlantic. NTA's $T_{Y\to TP}$ values flatten rapidly after -5 milinats per month (see 5.6c), stating fewer outliers and this can also be seen by the wide white areas, indicating no causal teleconnection, with the studied area.

The causal relationship from all Pacific-related indices towards TP also covers a significant area of the Caribbean Sea coast, with values ranging from -6 to -30 milinats per month. This suggests a role of the Pacific's variability in the rain patterns of this area. Further research can be performed to understand, in terms of information transference, the teleconnection between these two important sources of moisture for the study area.

On the other hand, information transference flow towards T2M exhibits significantly different behavior. High positive values are found for TNA and NTA indices, being 571 and 454 milinats per month the maximum LKMLE values respectively (see 5.7). This suggests a causal teleconnection between these two global phenomena that determine the state space of T2M over the Colombian Andes and the Caribean coast of Costa Rica and Panamá (see Figures 5.5b, 5.5c). Focusing our attention on Colombian territory, NTA has the widest covered territory with the highest values of information transference flow, suggesting an important role of the Atlantic. This will be discussed later in section 5.2.

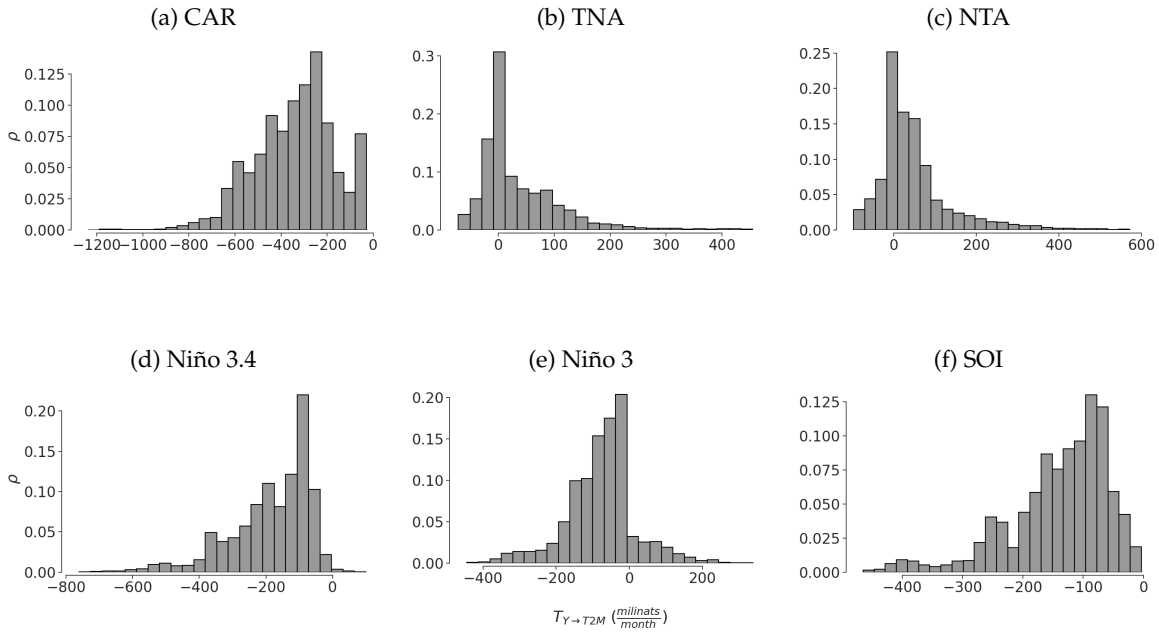


FIGURE 5.7: Information transference probability distribution for the two-meter Temperature (T2M) from the (a) CAR, (b) TNA, (c) NTA, (d) Niño 3.4, (e) Niño 3, and (f) SOI indices in milinats per month.

For CAR index, the highest overall negative values for $T_{Y\to T2M}$, reaching -1238 milinats per month, were found around the coastlines of Colombia near Panamá and the Amazon Basin (see Figure 5.5a). The former represents one of the most interesting teleconnections for this index since Morales et al., 2017 relates the important role of the Caribbean Low-Level Jet (CLLJ) in the transport of moisture towards Central America but the role of the Caribbean with temperature was not clear. Further research is needed to understand the meaning and usefulness of LKMLE applied to Low-Level Jets in this area.

Over the same influence area of CAR, Niño 3.4 also has negative values of $T_{Y\to T2M}$, suggesting an indetermination of temperature due to these two teleconnections (see Figure 5.5d). On the contrary,

Niño 3 $T_{Y\to T2M}$ takes positive high values over the Colombian Andes ranging from 112 to 338 milinats per month (see figure 5.5e). The inversion of the $T_{Y\to T2M}$ sign concerning TP in Niño 3 suggests that the anti-correlation studied between these two variables (Trenberth and Shea, 2005; Berg et al., 2015) is somewhat captured by the LKMLE of information transference, but it does not hold for all the indices and it is not quantitative proportional. Further studies might be carried out to establish this inversion of sign in terms of $T_{Y\to X}$ for TP and T2M.

Notice that, unlike the behavior of VIMFLUX and TP, the values of the LKMLE for T2M are not centred around zero for most of the indices. Also, the peaks are not as sharp. This indicates a higher number of teleconnections with the studied global phenomena with this variable, which might have a more strong causal relationship.

5.2 Role of Teleconnections in Shaping Colombia's Climate

The previous section focused on mapping information flow and identifying potential teleconnections (remote climate influences) on Colombia's climate. However, understanding the relative strength of each influence is crucial. This section dives deeper by analyzing the spatial distribution and normalized magnitude of information transfer for each climate variability mode towards the three key variables (TP, T2M, VIMFLUX). This approach provides a clearer picture of how much each teleconnection previously identified, contributes to the Colombian climate and highlights the specific locations (future nodes) considered in the analysis. Furthermore, by comparing these results with previously characterized teleconnections, we can validate this method.

Building upon the boxplots presented in the previous section (Figure 5.1), Figure 5.8 displays the normalized LKMLE ($\tau_{Y\to X}$) for all indices and variables. We observe that the box shapes and signs of the relative information transfers remain consistent with Figure 5.1. However, the percentage values of the normalized LKMLE are considerably smaller compared to the large net values found previously. This emphasizes the importance of information transference normalization. Figures 5.9, 5.10 and 5.11 present the spatial distribution of the calculated $\tau_{Y\to X}$. Values below 0.05% are considered negligible and set to zero ($-0.005 < \tau_{Y\to X} < 0.005$).

Figure 5.9 maps the relative influence on VIMFLUX across Colombia. The influence values range from -13% to 1%, with higher values indicating stronger influence. Compared to other variables (see Figure 5.8), VIMFLUX shows the greatest overall influence. This highlights an interesting point: even though the net influence values ($T_{Y\to X}$) found earlier were small, they have a relatively strong impact on the overall changes in Colombia's moisture patterns. For example, the Caribbean coast and the Brazilian Amazon (areas with lower net influence for TNA) show the strongest relative influence from the Tropical North Atlantic (TNA) index according to Figure 5.9b (compare with net influence

in Figure 5.3b). This suggests that while the overall impact of TNA might seem small, it plays a significant role in specific regions.

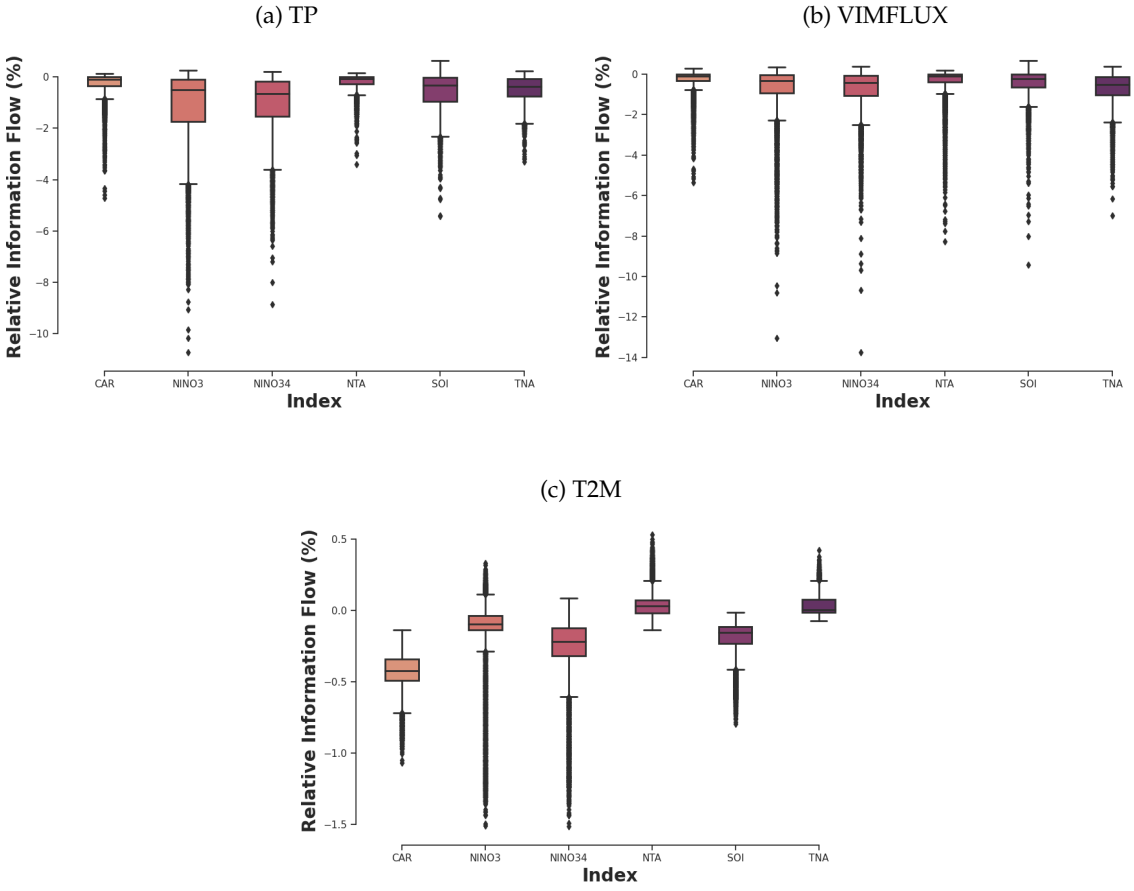


FIGURE 5.8: Boxplot illustrating the spread of (a) TP, (b) VIMFLUX and (c) T2M variables relative information transference flow from the CAR, Niño 3, Niño 3.4, NTA, SOI and TNA indices. The box represents the middle 50% of the data, with the line inside indicating the median. Whiskers show the overall range, and any dots outside indicate outliers.

Additionally, if the relative causal contributions to VIMFLUX for all indices over the Colombian Orinoco basin are added, the relative information transference flowing from global variability towards this area is less than 10%. This observation aligns with prior research findings, indicating that the primary source of moisture within the study area predominantly stems from the recycling process, involving evapotranspiration within the Andean mountain range, as well as in the Orinoco and Amazonian Basins (Hoyos et al., 2018; Escobar et al., 2022).

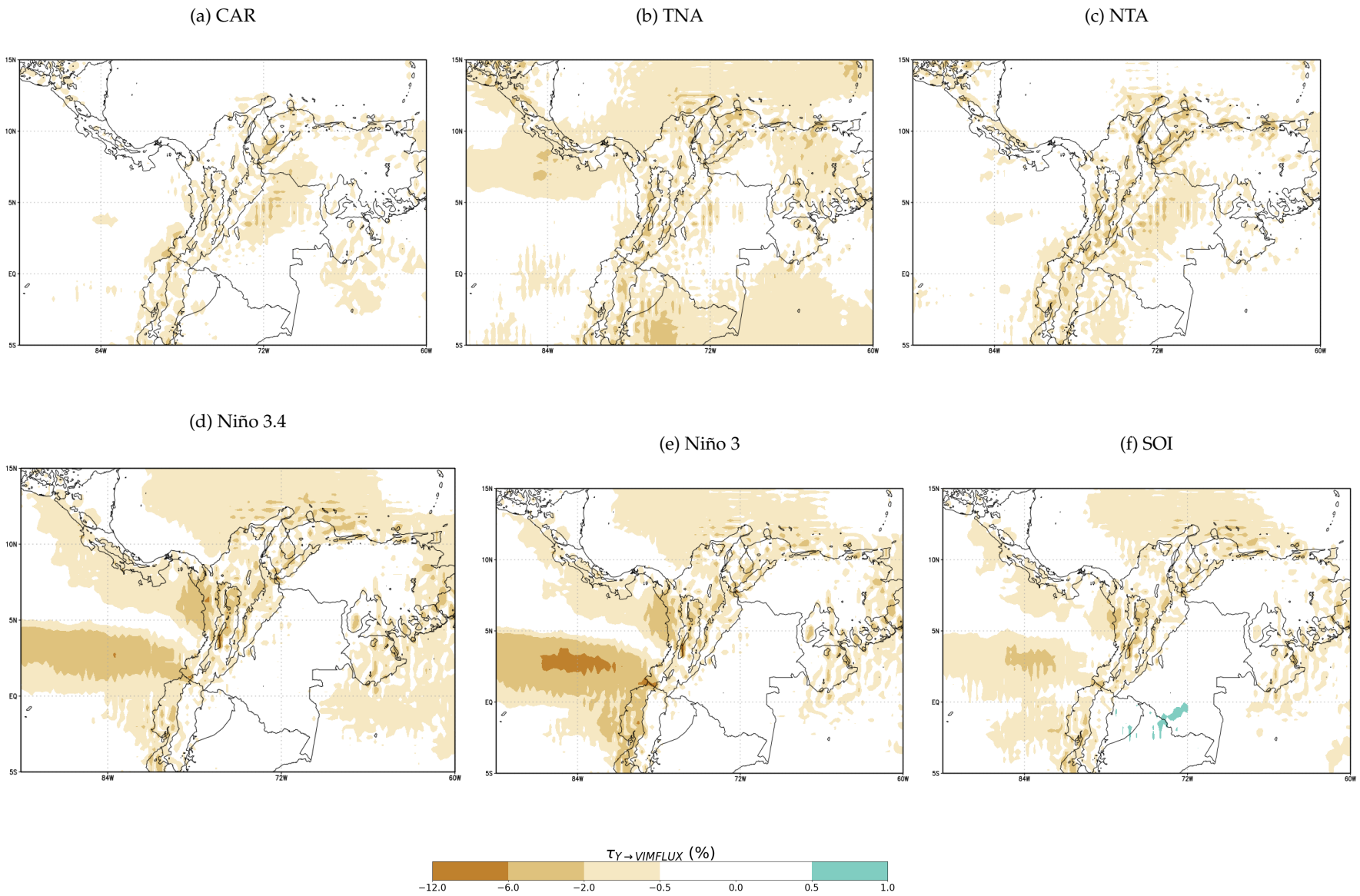


FIGURE 5.9: Relative information transference towards vertical integral of moisture flux (VIMFLUX) from the (a) CAR, (b) TNA, (c) NTA, (d) Niño 3.4, (e) Niño 3 and (f) SOI indices in %. Darker zones indicate outlier values for $\tau_{Y \to VIMFLUX}$.

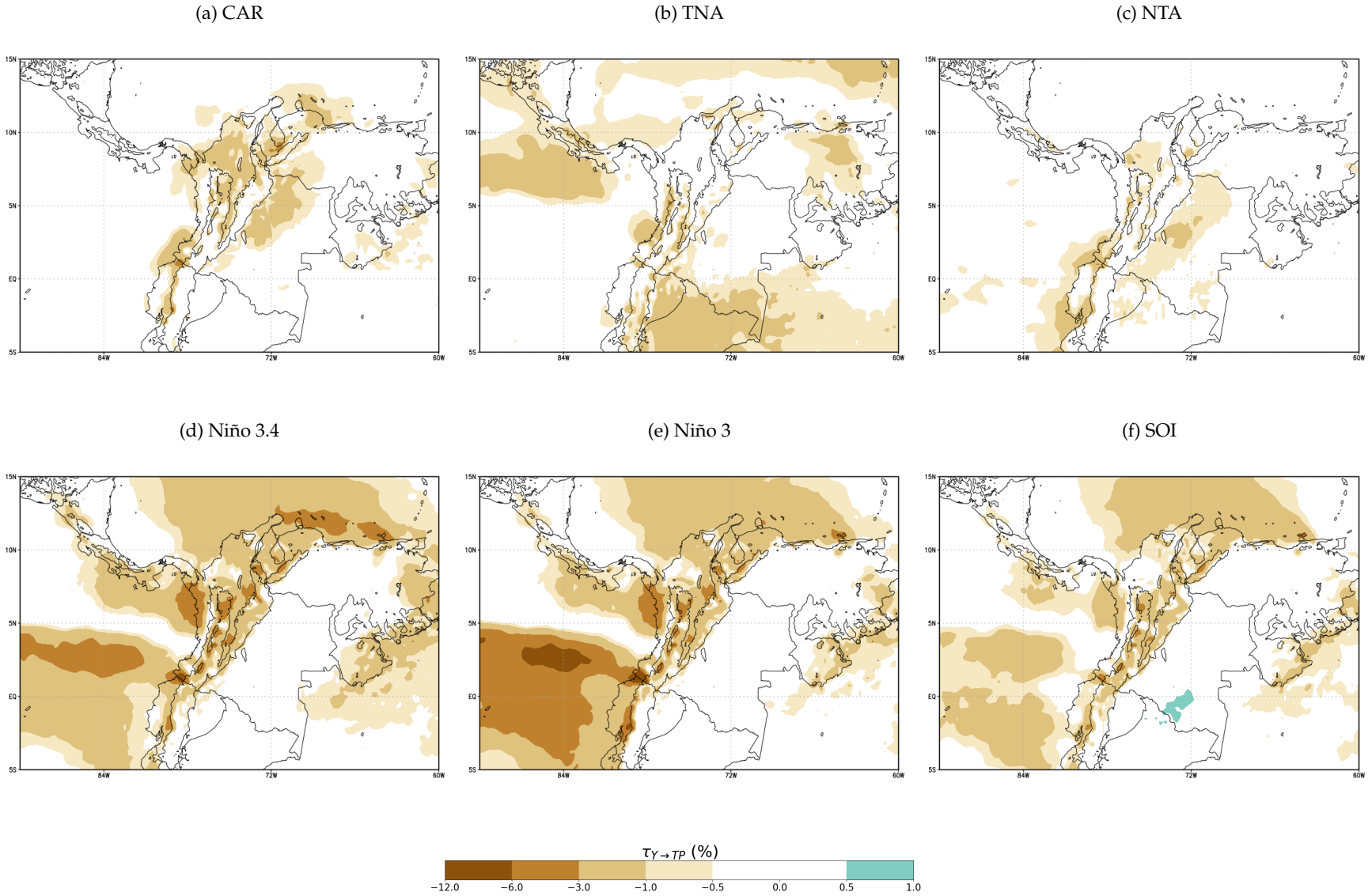


FIGURE 5.10: Relative information transference towards Total Precipitation (TP) from the (a) CAR, (b) TNA, (c) NTA, (d) Niño 3.4, (e) Niño 3 and (f) SOI indices in %. Darker zones indicate outlier values for $\tau_{Y \to TP}$.

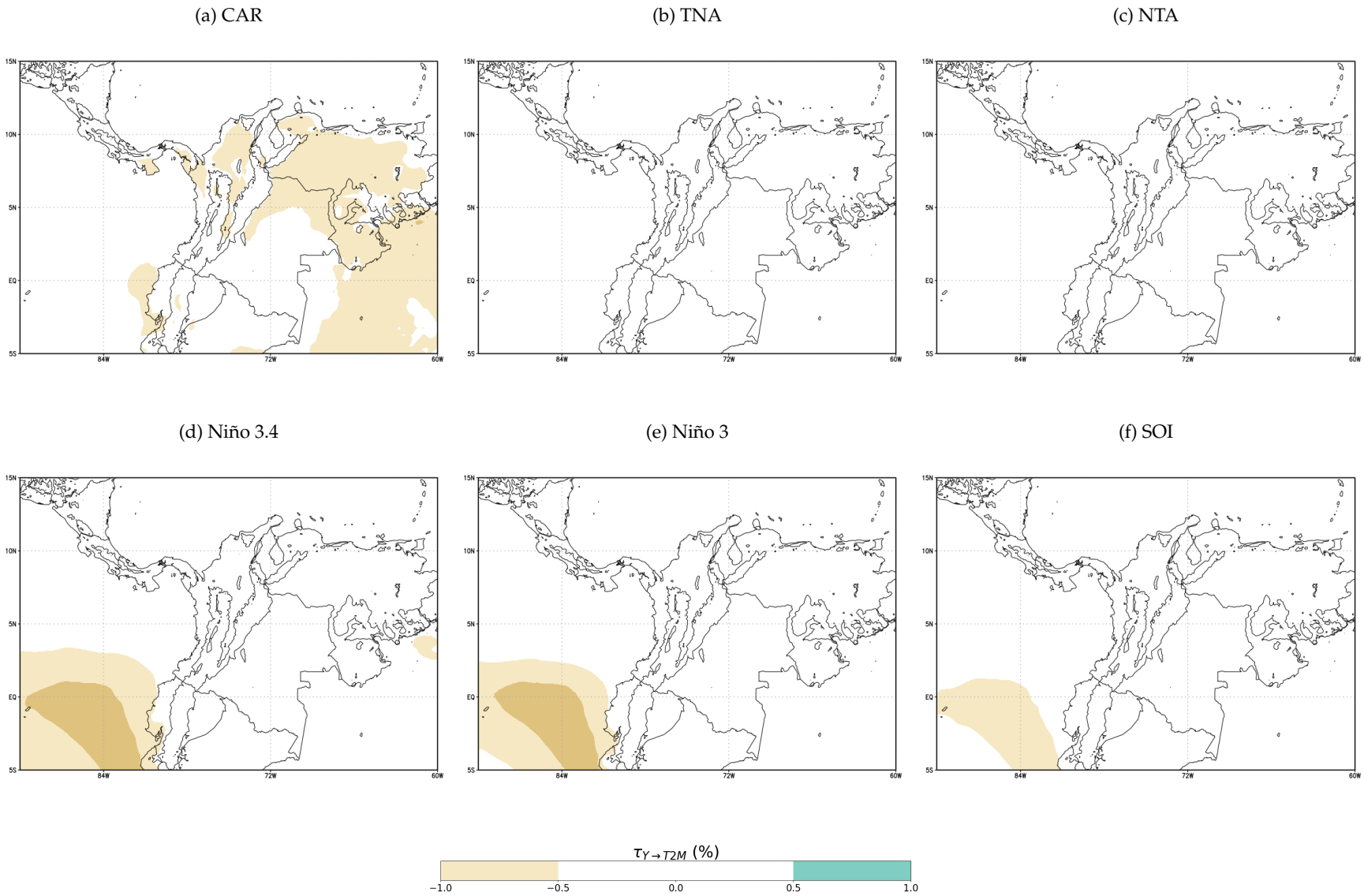


FIGURE 5.11: Relative information transference towards two-meter Temperature (T2M) from the (a) CAR, (b) TNA, (c) NTA, (d) Niño 3.4, (e) Niño 3 and (f) SOI indices in %. Darker zones indicate outlier values for $\tau_{Y \to T2M}$.

Figure 5.8a shows that the relative influence $\tau_{Y \to X}$ on Total Precipitation (TP) ranges from -10.8% to 0.7%. Nino 3 and Nino 3.4 indices, linked to the Pacific's SST anomalies, have the most significant negative LKMLE normalized values, particularly in the Cold Tongue, the Andes, the Caribbean and Pacific coastal areas (seen in Figure 5.10e). This suggests a strong and clear teleconnection (remote climate influence) impacting overall precipitation uncertainty. SOI, another Pacific Ocean index, shows similar results. However, its influence pattern over the Pacific doesn't directly match the Cold Tongue region. Interestingly, a small area in the Amazon basin also exhibits low positive LKMLE values (see Figure 5.10e).

In the literature, ENSO and TP is a good representation of tropics-tropics teleconnections since they are consistent with precipitation deficiency or abundance during El Niño (EN) or La Niña (LN) events over Northern South America. Our analysis confirms a strong causal influence of ENSO on precipitation over some parts of the Colombian territory, consistent with previous findings across broader regions (e.g., Kousky et al., 1984; Aceituno, 1988; Souza and Ambrizzi, 2002; Poveda et al., 2006; Grimm and Tedeschi, 2009). Importantly, the LKMLE approach allows for spatially differentiated causality analysis, enabling further targeted investigations in areas exhibiting the strongest ENSO-precipitation linkages (as identified by high normalized LKMLE values).

Canchala et al. (2020a) found a teleconnection based on Pearson's Correlation between the el Niño indices (ONI, MEI, Niño1 + 2, Niño3, Niño3.4, Niño4, and PDO) and precipitation in the department of Nariño in Southwestern Colombia, being the correlation stronger with Niño3.4 and Nino4. Similarly, we found high information transference from Niño 3 and Nino 3.4 towards this region (see Figure 5.10d and 5.10e darker areas over the Pacific coast). Our contribution states that the information transference is negative, meaning a causal relationship with ENSO variability that amplifies the TP uncertainty.

In a similar work, Canchala et al. (2020b) found a strong ENSO teleconnection towards streamflow (water carried by streams for rivers) in eastern Colombia, using Kendall's Tau Correlation and Cross-Correlation. This is due to the direct effect towards the Choco Jet (Chocó LLJ) during EN (weakening, i.e., reduction of rainfall) or LN (intensifying i.e., higher convection and rainfall) events (Poveda and Mesa, 1999). In agreement with them, we found this teleconnection with TP and VIMFLUX to be stronger for El Niño 3 and El Niño 3.4 indices (see figure 5.9d,5.9e and 5.10d, 5.10e), with negative information transference sign.

In agreement with these two studies, subsequent prediction analysis for this region must include the climate-oceanographic-indices, but focus on its influence on the space state expansion of rainfall. Further research into this region of Colombia and considering interannual and annual time scales is suggested to differentiate the causal relationship with EN or LN events.

The Caribbean (CAR) and North Tropical Atlantic (NTA) indices have the least widespread influence on TP (see Figures 5.10a and 5.10c). Their small negative influence is limited to the Caribbean Basin and the eastern Andes (values between -0.5 and -3). Interestingly, the Tropical North Atlantic (TNA) index shows similar influence but in a completely different region (compared to net LKMLE), around Northern Peru (see Figure 5.10b).

Surprisingly, except for the Caribbean (CAR) index, all other indices have minimal influence on two-meter temperature (T2M) within the study area (see Figure 5.11). This is unexpected because T2M had the highest overall influence (net $T_{Y \to X}$) in the previous analysis (compare with Figure 5.5). This suggests that for T2M, "noise" and the influence of the location itself on its own changes (self-contribution) are much stronger than the influence of these teleconnections.

Chapter 6

Understanding the Link Between Climate Variability and Complex Networks

In this chapter, the main findings of this thesis are presented. First, the network's complexity is analyzed using the graph properties of the climate network built based on the LKMLE. In the second section, the complex climate network structure is discussed in terms of the centrality measures: degree, eigenvector, closeness, and betweenness. Findings suggest that the highly clustered networks exhibit similar measures and structures for all ENSO-related indices, with the Colombian Andes playing a key role in the connectivity with this phenomenon. Meanwhile, Atlantic oceanic indices display different features and spatial distributions among them, with higher centralities in the Orinoco, Amazon, and Caribbean basins.

6.1 Graph Structure: Unveiling Complexity

The relative information transference flow allows the estimation of differential contributions of global climate processes to regional climate variability instead of the net value of the estimator. A 0.5% relative information transference is a suitable threshold for establishing a climate link among grid cells for all indices. A summary of the overall properties for each index is presented in Table 6.1, followed by the adjacency matrix structure (see Figure 6.1), and its corresponding node degree distribution (see Figure 6.2).

Six simple, undirected, and unweighted networks were obtained. There are around 50% and 70% isolated vertex within the study area for each index. NTA shows the maximum isolated vertices with 7209, followed by CAR and SOI; these three indices have the resulting smaller networks with 2592, 3846, and 4714 nodes, respectively. On the other hand, TNA, Niño 3.4 and Niño 3 graphs contain 5800, 6283, and 5444 nodes respectively. All graphs are neither sparse nor dense since all their densities state around 50% of the maximum possible connections. Niño 3 has a lower density with 35% of the maximum possible connections.

Table 6.1: Summary of Network's measures and properties highlighting changes when omitting isolated nodes for the mean degree centrality $\langle k_v \rangle$ and the diameter $\mathcal{D}_{isolated}$.

Network	L	$\mathcal{D}$	$\langle k \rangle$	$\langle C \rangle$	$\mathcal{H}$	$\overline{\epsilon(v)}$	**R**	**D**	**Isolated nodes**	$\mathcal{D}_{isolated}$	$\langle k_v \rangle$
CAR	3871346	0.5236	790.0	0.833	0.319	2.875	2	3	5955	0.0806	0.081
NTA	1374263	0.4092	280.4	0.781	0.195	3.223	2	4	7209	0.0286	0.029
TNA	9961331	0.5923	2032.7	0.795	0.517	2.458	2	4	4001	0.2074	0.207
Niño3.4	8706501	0.4412	1776.7	0.705	0.545	3.119	3	5	3518	0.1813	0.181
Niño3	5234314	0.3533	1068.1	0.690	0.448	3.028	2	4	4357	0.1089	0.109
SOI	5940966	0.5348	1212.3	0.769	0.402	3.805	3	6	5087	0.1237	0.124

Considering the 9801 grid points covering the study area, the obtained network densities show a sparse graph structure due to the high amount of isolated vertices. In this case, the lowest density corresponds to NTA with around 2% of all possible connections and the highest density is for TNA with 20%. This finding shows a strong differentiation between the resulting spatial distribution of the teleconnections pattern from SST anomalies from similar areas of the Atlantic, and it is a clue about the mechanisms of information transference from these areas.

The average eccentricity for all indices (Table 6.1) suggests that for the big graphs obtained (more than 2k nodes and 1m edges each) it takes a maximum of 4 links to reach any other node on the network. This is the first indicator of shortcuts and complex structure in our networks because there are emerging paths that minimize the distance between points of space in terms of the shared relative information transference from global variability phenomena. This property is supported by the graph radius and diameter, with a maximum eccentricity of the whole network between 3 and 6 links for all indices.

This heterogeneous and complex structure of the obtained graphs is confirmed for some indices by the network's entropy, quantifying the heterogeneity of the network's degree distribution. Figure 6.1 shows the complex structure pattern discussed in terms of all graph properties, as well as clustering areas and isolated patterns. For TNA, Niño 3, Niño 3.4, and SOI, the average normalized entropy is around 0.5, which reinforces that the network is neither sparse nor fully connected, and does not have dominant peaks or tails on the distributions, as seen in Figures 6.2 b, d, e and f. On the other hand, CAR and NTA have lower entropy, showing that flattened and peaked shapes in their degree distributions are expected (Figure 6.2 a and c).

The average Clustering coefficient $\langle C \rangle$ for all networks suggests a high clustering structure among all networks. CAR shows the highest value with 0.833 and Niño 3 the lowest with 0.690. Clustering values are very similar for NTA and TNA indices, with 0.78 and 0.79 respectively. The modular structure indicated by these high ($\langle C \rangle$) values is also visualized in the display of adjacency matrices (see

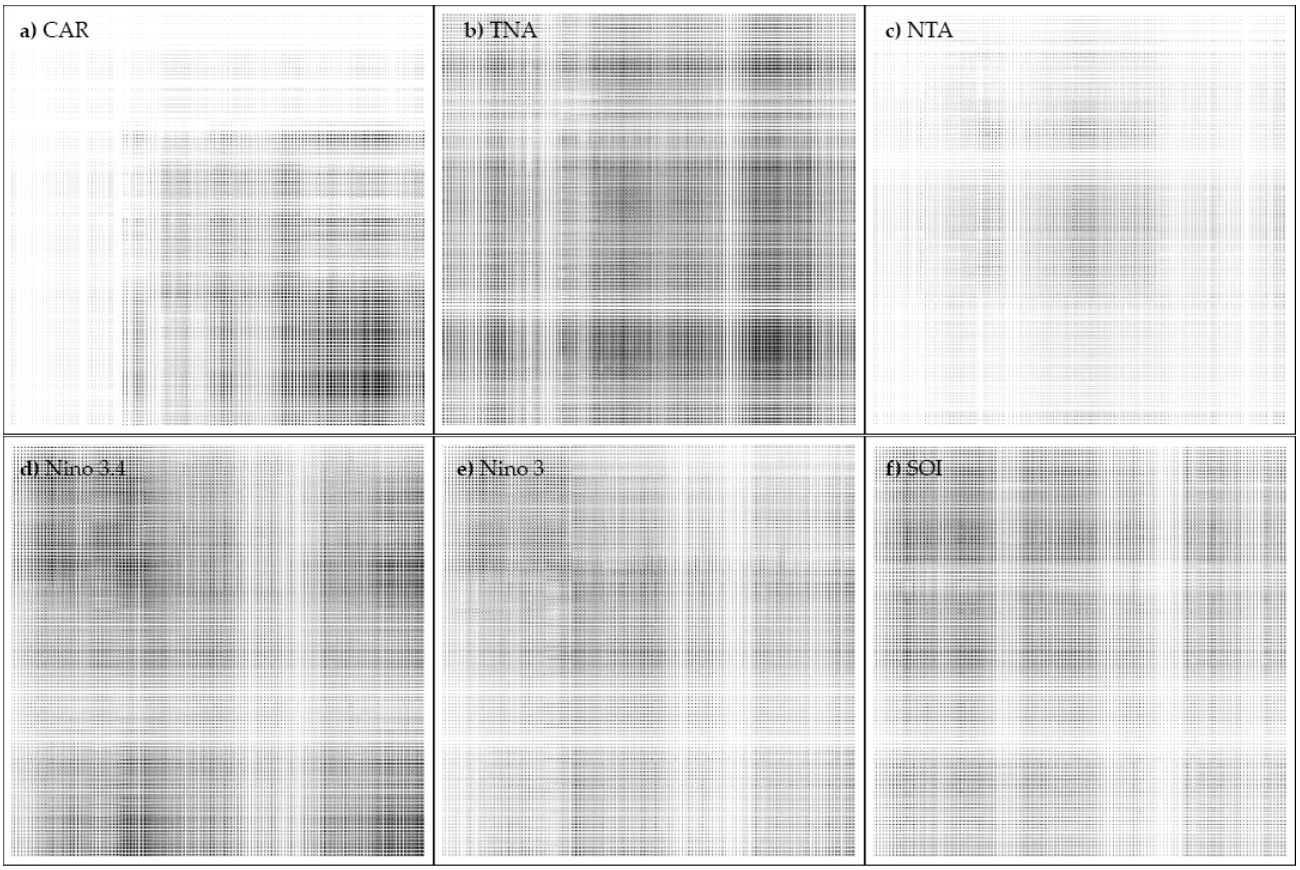


FIGURE 6.1: Adjacency matrix representation for (a) CAR, (b) TNA, (c) NTA, (d) Niño 3.4, (e) Niño 3 and (f) SOI indices networks. Dimension is 9801×9801 ($N \times N$) and black points represent links.

Figure 6.1). This type of structure also means that the network has the smallest possible average distance among different nodes, a characteristic previously discussed in terms of the graph's eccentricity.

The probability distribution of the networks, estimated by the histograms in figure 6.2, show that hyper-connected nodes are highly probable in TNA, Niño 3.4, Niño 3, and SOI. These nodes serve as an important connectivity bridge for the whole network and will be discussed in the next section. Nevertheless, neither distribution follows a canonical known shape for the degree distribution (see vertex-degree-dist), since neither Power Law nor Binomial shape distributions are seen in the ρ_k plots.

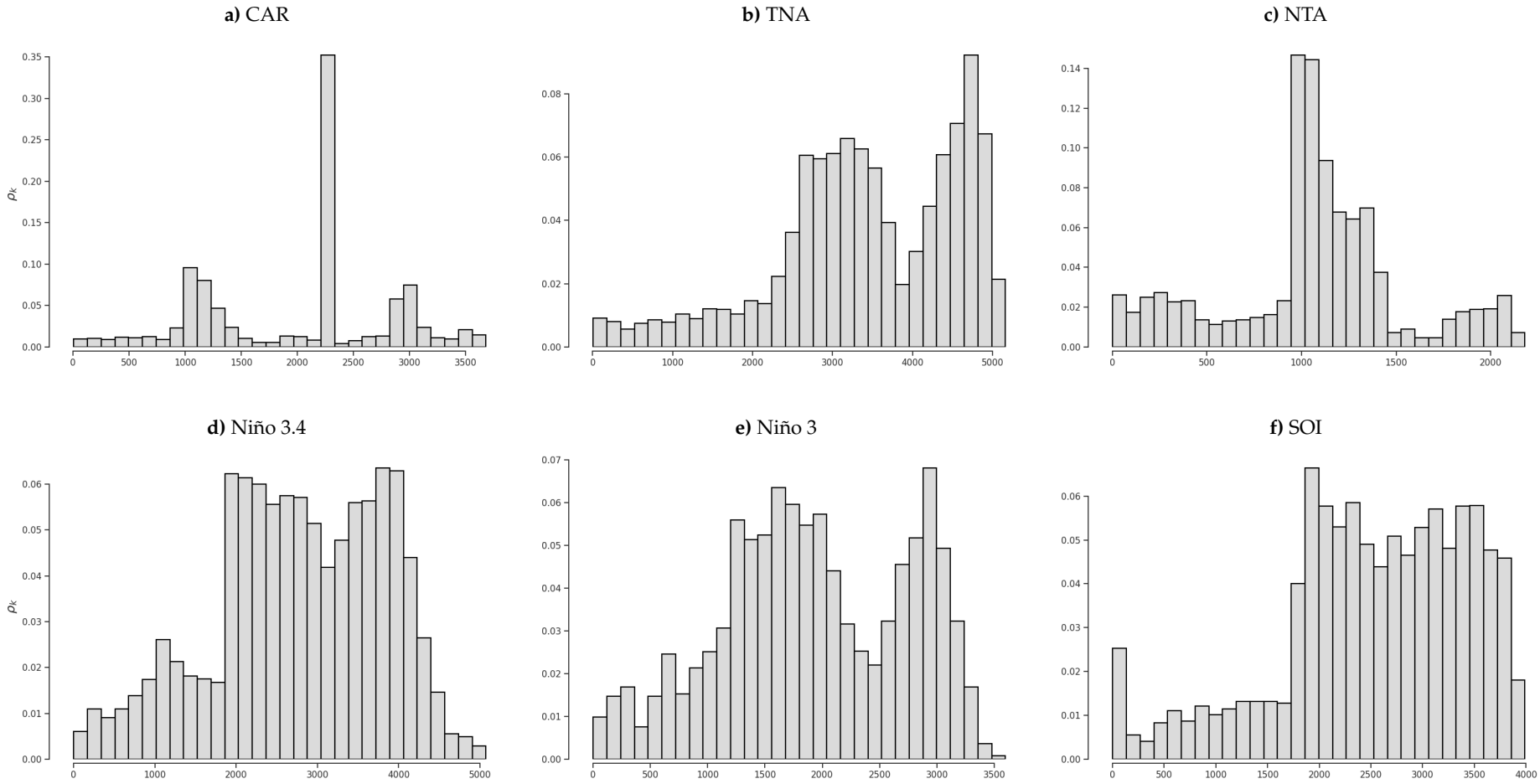


FIGURE 6.2: Estimation of the probability distribution function for the a) CAR, (b) TNA, (c) NTA, (d) Niño 3.4, (e) Niño 3 and (f) SOI indices networks vertex degree k. The distributions indicate certain regions with nodes with several connections, showing a high probability of finding nodes with a high number of links.

CAR and NTA have a marked maximum probability for the vertex degree, at 2100 and 1000, respectively. On the other hand, TNA, Niño 3, Niño 3.4, and SOI have several probability modes, which translates into several regions with a large number of nodes connected, while intermediate nodes connect such regions. Degrees centralities below 1000 nodes are less probable for all networks, indicating other regions in the network with fewer connected nodes. These regions will be then observed by the centrality measures, calculated in section 6.2.

Characteristics of complex networks have been described for each obtained network in terms of the key network properties, the visualization of the networks (Figure 6.1), and the vertex degree distribution (Figure 6.2). Further research must be done to establish the topological details of the presented measures in the context of a Climate Complex Network and check additional properties for real-world systems. In the scope of this work, clustering structure and shortcuts were found as important features of these networks, making them robust against stochastic perturbations.

6.2 Interpreting Colombian Variability from a Climate Complex Network Approach

In the previous section, the complexity of these networks was discussed. Here, the properties maps of the degree, eigenvector, closeness, and betweenness centrality measures are shown to place the key nodes of the networks in specific points of Colombian territory and its surroundings.

Figure 6.3 presents the degree centrality for the six constructed networks. Darker zones represent higher vertex degrees, i.e., the ones with more links in the network. ENSO-related indices exhibit a similar structure, with higher connectivity in the Caribbean region and along the cold tongue of the Pacific. There is also high connectivity over southern Venezuela for Niño 3.4 and Niño 3.

On the other hand, CAR, TNA, and NTA exhibit differentiated structures. The TNA vertex degree is maximum over the southern coast of Costa Rica and the western portion of the Amazon rainforest in Brazil. CAR maximums are distributed over the Colombian Andes, with peaks in the Orinoco Basin near the eastern branch and northwestern Brazil. NTA exhibits a smaller degree of centrality from all indices, being its maximal area of influence around the valleys near the Colombian Andes.

White areas in Figure 6.3 display the location of isolated nodes in each network. ENSO-related indices show the isolation of the Orinoco and Amazonian basins, which agrees with the significance of the teleconnections established in the previous chapter. This is also valid for CAR, TNA and NTA isolated nodes over the Pacific.

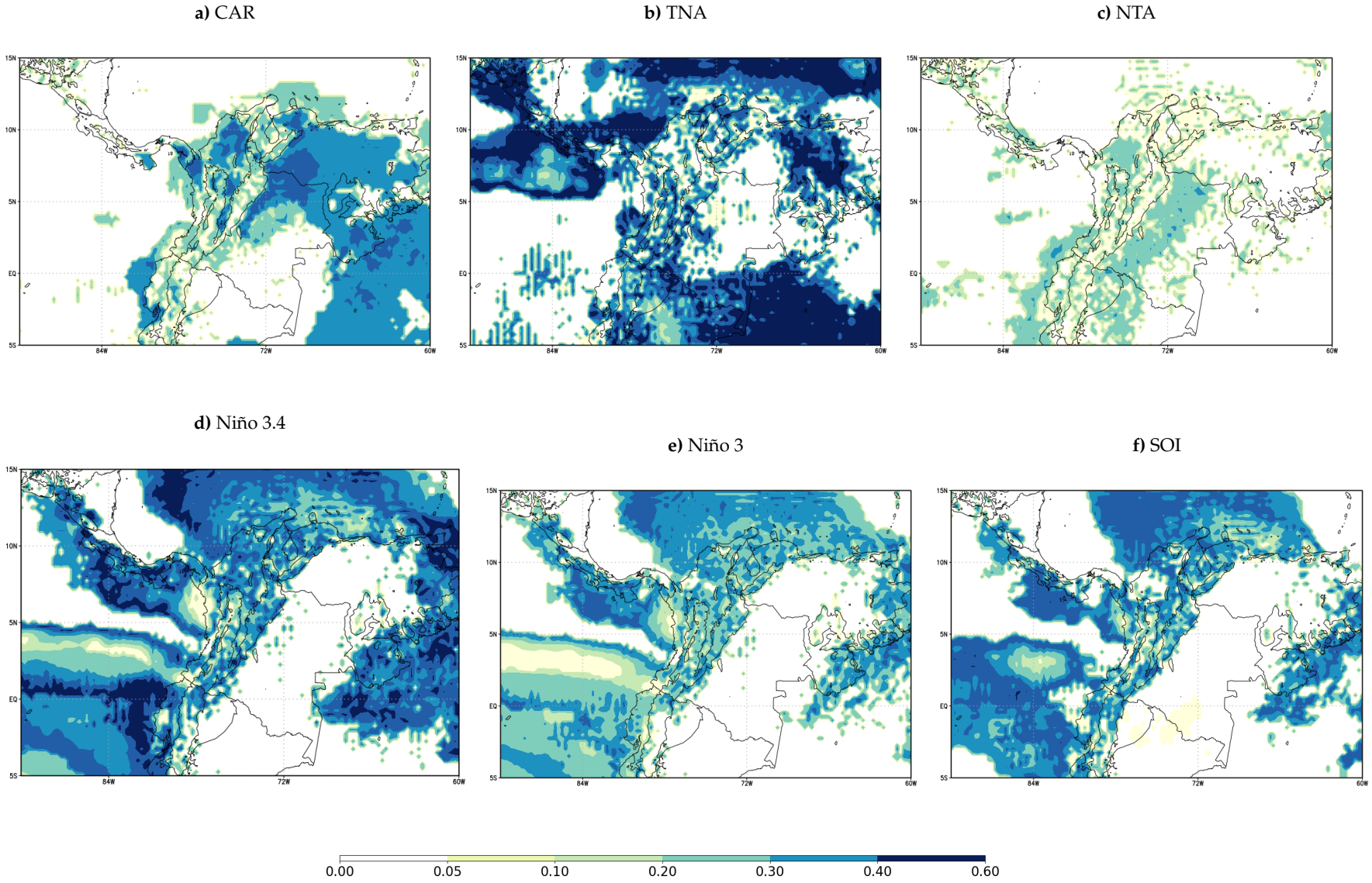


FIGURE 6.3: Degree centrality measure's spatial distribution over the study area for the a) CAR, (b) TNA, (c) NTA, (d) Niño 3.4, (e) Niño 3 and (f) SOI indices networks. Darker zones indicate nodes with higher vertex degree k.

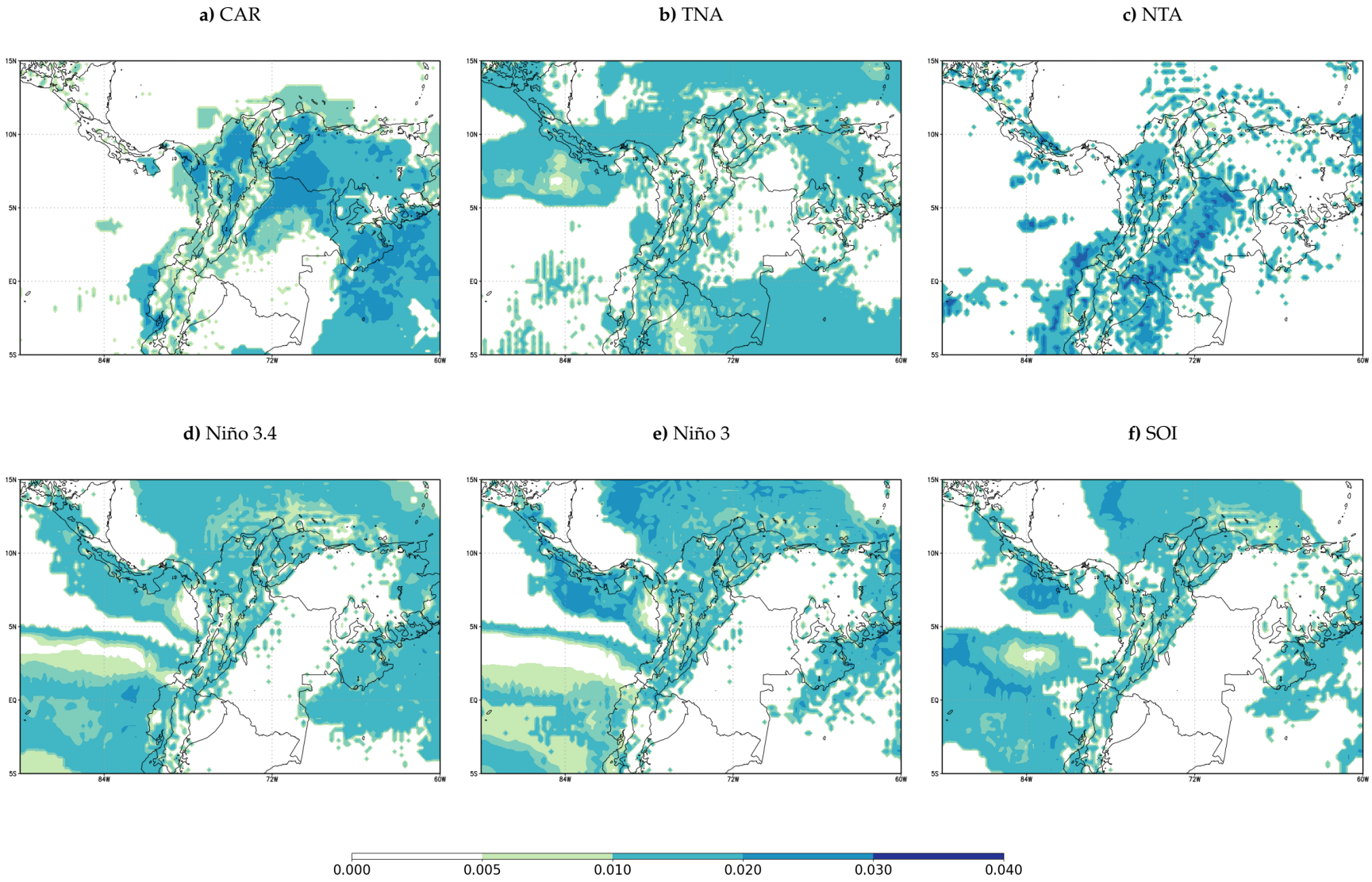


FIGURE 6.4: Eigenvector centrality measure's spatial distribution over the study area for the a) CAR, (b) TNA, (c) NTA, (d) Niño 3.4, (e) Niño 3 and (f) SOI indices networks. Darker zones indicate nodes with higher meaningful links.

TNA high probability for great node degrees is attributed to the wide areas around all the studied, with degree centrality over 0.4. This is attributed to similar $\tau_{2\to1}$ values in VIMFLUX and TP, weakly connecting these places due to the chosen 0.5% threshold. Further research is needed to establish the sensitivity of this network's connectivity to the threshold employed.

In addition to the degree, the importance of the linkages each node has is discussed using the eigenvector centrality displayed in Figure 6.4. It is observed that the spatial structures are the same, but the color gradient changes in terms of the importance of those connections. For instance, NTA low maximal degree connections are important within the network structure, so the valleys surrounding the Andes play a key role in the information transference for its associated variability phenomena. In contrast, TNA high maximal degree areas display similar centrality importance.

All ENSO-related indices show similar eigenvector centrality (Figure 6.4), suggesting that the importance and quantity of connections within these networks preserve the same topology for the study area. This finding is interesting since it is suggested that no matter the index you use to discuss the causal effect of ENSO towards Colombia, the causal structure in terms of a network's topology will be very similar over the Andean mountain range. Differences over the Pacific Cold Tongue for these three indices may unravel slightly important paths for the information transference flow, but further research is needed.

Closeness centrality indicates nodes with the capability to efficiently access any other node within a limited number of steps, as well as nodes that may be positioned quite distantly from others within the graph. Figure 6.5 exhibits which places are close to other parts of the study area by the connections established with the causal relationship of the shared teleconnection with the Oceanic index. To make this more clear, darker zones are understood as being closer to the rest of the network and lighter ones as farther away from other nodes. For instance, the Andes are close to the rest of the network in TNA, Niño 3.4, Niño 3, and SOI. For CAR the highly central nodes are distributed among the Orinoco Basin.

Additionally, the importance of nodes given their role as crucial passing spots of short paths between nodes is shown in Figure 6.6. Here, darker zones highlight important structural nodes that, if removed, will certainly affect the eccentricity of the network. Higher values are found over the Orinoco basin in the Colombian frontier with Venezuela for CAR, highlighting a structurally important role of this area for the teleconnection structure of the territory with the Caribbean (see Figure 6.6 a). Southern to this area, following the shape of the Andes NTA displays some of the most structural points for its network (see Figure 6.6 c).

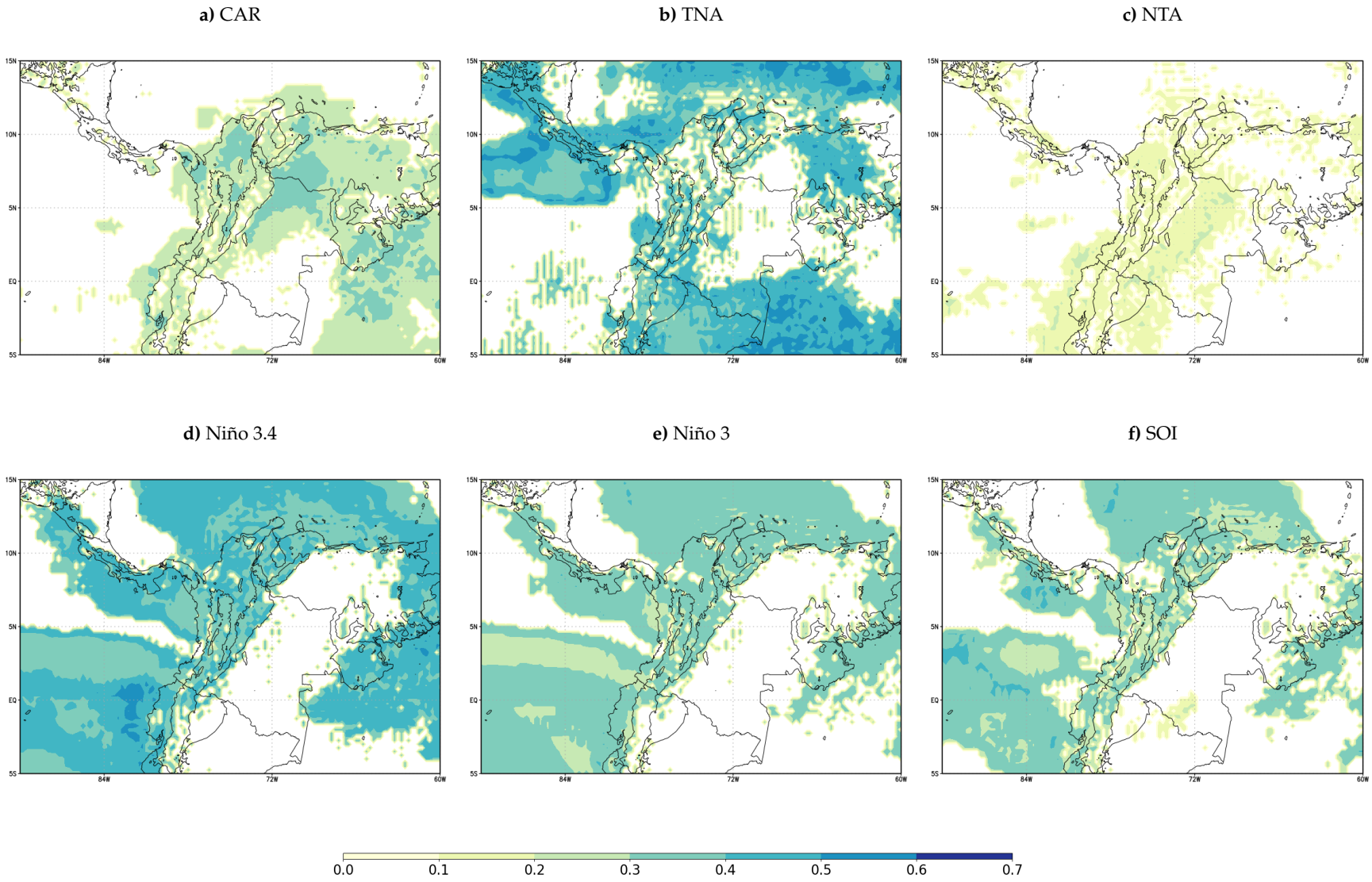


FIGURE 6.5: Closenness centrality measure's spatial distribution over the study area for the a) CAR, (b) TNA, (c) NTA, (d) Niño 3.4, (e) Niño 3 and (f) SOI indices networks. Darker zones indicate nodes closer to any other node in the network.

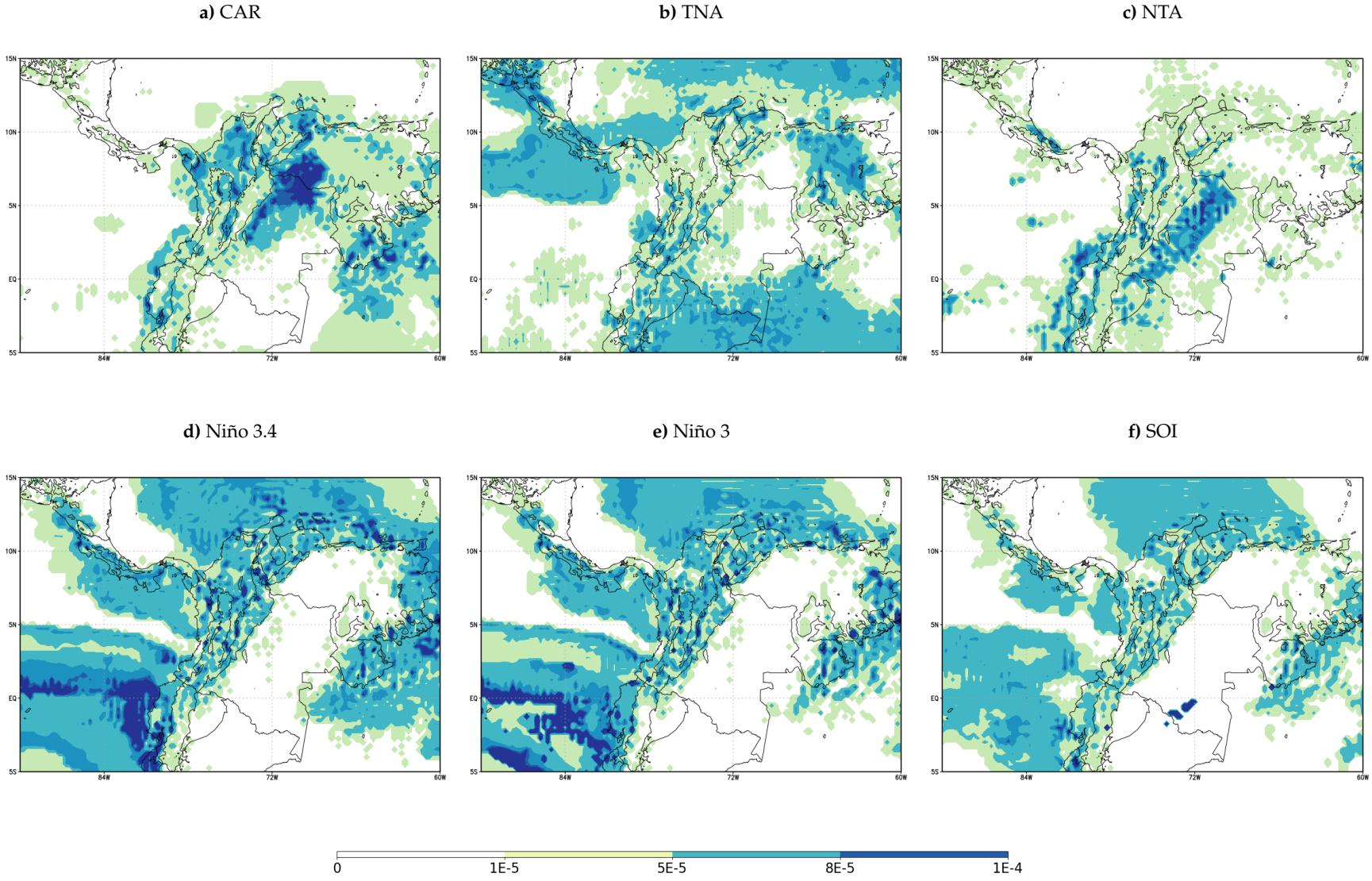


FIGURE 6.6: Betweenness centrality measure's spatial distribution over the study area for the a) CAR, (b) TNA, (c) NTA, (d) Niño 3.4, (e) Niño 3 and (f) SOI indices networks. Darker zones indicate structurally important nodes through which many short paths pass in the network.

TNA has some important -but not crucial- structural nodes over the north-western Amazon and the Pacific near Panama and Costa Rica (see Figure 6.6 b). A high betweenness centrality over this area suggests a dependency on the Amazon rainforest to preserve the graph structure for this index. On the other hand, for ENSO-related indices, the structure of betweenness is similar to previous centrality measures. The Colombian Andes show a key structural role in these networks (see Figure 6.6 d, e and f).

Considering all the above centrality measures, some of the most important places in the study area for the constructed networks are the Colombian portion of the Andean Mountain range, which is transversal for all indices except CAR and NTA. The Colombian part of the Orinoco Basin for the former and the Pacific Cold Tongue for all ENSO-related indices. The Amazon Basin is isolated from the former but significant for TNA, which opens up further discussion into the connection between this important South American Basin and the Atlantic.

Conclusions and Perspectives

This work studied causal teleconnections underlying Colombian climate variability related to the Atlantic and Pacific Ocean's anomalies in the scope of the network frame. Considering the Liang-Kleeman maximum likelihood quantifier for information transference revealed the spatial structure of causal relationships. ENSO-related indexes exhibit a robust causal teleconnection with precipitation, amplifying the accessible states. Conversely, AMO-related indexes showcase a causal relationship with temperature, contributing to its certainty.

Based on the Liang-Kleeman quantifier, the present analysis remarks a heterogeneous spatial distribution of information transference within the study area. Regions demonstrating maximum causality are concentrated along the Pacific coast and the Andean Mountain range portion of Colombia. In contrast, negligible causal relationships were identified along the Amazonian and Orinoco Basins for ENSO across all variables. Particularly, Sea Surface Temperature index anomalies in the Atlantic, exhibit significant causal relationships in the eastern and southern parts of Colombia's Andes. These findings underscore the complex nature of the teleconnections, offering valuable insights for understanding and predicting Colombian climate dynamics in the context of oceanic sources of variability.

The 0.5% chosen threshold is suitable to discern between negligible and significant causal relationships through Liang-Kleeman's relative information transference flow, showing meaningful teleconnections within the study area. Further research applications to evaluate this formalism should be carried out in other tropical and non-tropical areas to explore the causal relationship around the globe.

Undirected, unweighted, and simple graphs were derived from the proposed climate network based on information transference. The graph's measures and description suggest a complex network topology that is not canonical. Further research should explore deeper into other real-world network characteristics so our analysis contributes to the development of a more robust wide network theory to unravel climate teleconnections structure.

Finally, the climate network based on information transfer for Colombia contributes to the explanation of causal connectivity between the territory and the global climate system. This network unveils distinct spatial structures associated with well-known phenomena, such as the Cold Tongue of the Pacific and the Low-Level Jets in the Pacific and Orinoco Basin. The outcomes generated through this framework provide valuable insights for unraveling the most significant causally related spatial

locations of interest. Moreover, they shed light on how global teleconnections manifest in specific structures that influence climate development in a particular region, all achieved within a framework of lower computational cost.

The established framework presented in this thesis will serve as a valuable tool for replicating similar analyses in diverse geographical regions. To facilitate widespread use, all codes have been made open-source and are compatible with personal laptops. Given the unprecedented era of climate change, it is imperative for the scientific community to collaboratively enhance our comprehension of the climate system for a more informed and nuanced understanding of our decisions and impacts within our territory and the whole world.

Appendix A

Information transference calculation code

The computations undertaken in this thesis were executed utilizing CDO (Climate Data Operators), Python, and Mathematica, with each programming language serving distinct roles in the various stages of implementing the complete formalism of the Complex Network based on information transference. The corresponding source codes are accessible in the GitHub repository: https://github.com/nicolerivera1/colombian-teleconnections. While a concise overview of the repository's structure and functionality is provided in this appendix, a detailed elucidation is available within the repository and each code.

The repository's hierarchical organization aligns with the distinct computational stages involved in constructing the networks. The *inf-transference* folder encompasses Bash scripts dedicated to executing Liang-Kleeman's information transference rate calculations between two *cdf* files, leveraging the CDO library. Both files must exhibit temporal and spatial consistency to ensure the proper functioning of the algorithms. Additionally, for readers seeking to replicate the thesis results, instructions for acquiring ERA5 data from the Climate Service Storage and obtaining the NOAA indexes are provided in this section.

Within the *adjacency-matrices* folder, all Python codes pertinent to the analysis of information transference flow rate probability density distributions, as well as the corresponding fitting and visualization using Mathematica, are housed. For reproduction purposes, the primary algorithm, *build_adjacency_matrix.py*, encompasses calculations for causal similarity thresholding, linkage establishment based on the threshold, and final adjacency matrix construction for each index.

Lastly, the *network-analysis* folder houses all codes for calculating the network's properties. Here, we also store the original properties maps alongside the corresponding Mathematica scripts. Notably, all maps in this thesis were crafted using QGIS Software for map construction and layer management.

Bibliography

Aceituno, Patricio (1988). "On the functioning of the Southern Oscillation in the South American sector. Part I: Surface climate". In: *Monthly Weather Review* 116.3, pp. 505–524.

Albert, Réka (2005). "Scale-free networks in cell biology". In: *Journal of cell science* 118.21, pp. 4947–4957.

Albert, Réka and Albert-László Barabási (2002). "Statistical mechanics of complex networks". In: *Reviews of modern physics* 74.1, p. 47.

Aldana-Domínguez, Juanita, Carlos Montes, María Martínez, Nicolás Medina, Joachim Hahn, and Maritza Duque (2017). "Biodiversity and ecosystem services knowledge in the Colombian Caribbean: progress and challenges". In: *Tropical Conservation Science* 10, p. 1940082917714229.

Ångström, Anders (1935). "Teleconnections of climatic changes in present time". In: *Geografiska Annaler* 17.3-4, pp. 242–258.

Arias, Paola A, René Garreaud, Germán Poveda, Jhan Carlo Espinoza, Jorge Molina-Carpio, Mariano Masiokas, Maximiliano Viale, Lucia Scaff, and Peter J Van Oevelen (2021). "Hydroclimate of the Andes part II: Hydroclimate variability and sub-continental patterns". In: *Frontiers in Earth Science* 8, p. 666.

Arias, Paola A, J Alejandro Martínez, and Sara C Vieira (2015). "Moisture sources to the 2010–2012 anomalous wet season in northern South America". In: *Climate dynamics* 45, pp. 2861–2884.

Arney, Chris (2010). "Six degrees: The science of the connected age". In: *Mathematics and Computer Education* 44.3, p. 260.

Balibrea, Francisco et al. (2016). "On Clausius, Boltzmann and Shannon notions of entropy". In: *Journal of Modern Physics* 7.02, p. 219.

Barabási, Albert-László (2009). "Scale-free networks: a decade and beyond". In: *science* 325.5939, pp. 412–413.

Barabási, Albert-László and Réka Albert (1999). "Emergence of scaling in random networks". In: *science* 286.5439, pp. 509–512.

Barabási, Albert-László and Eric Bonabeau (2003). "Scale-free networks". In: *Scientific american* 288.5, pp. 60–69.

Barabási, Albert-László (2016a). "Graph Theory". In: *Network Science*. Cambridge University Press.

— (2016b). "Random Networks". In: *Network Science*. Cambridge University Press.

Bass, Margot S, Matt Finer, Clinton N Jenkins, Holger Kreft, Diego F Cisneros-Heredia, Shawn F McCracken, Nigel CA Pitman, Peter H English, Kelly Swing, Gorky Villa, et al. (2010). "Global conservation significance of Ecuador's Yasuní National Park". In: *PloS one* 5.1, e8767.

Bassett, Danielle S and Edward T Bullmore (2017). "Small-world brain networks revisited". In: *The Neuroscientist* 23.5, pp. 499–516.

Battiston, Federico, Giulia Cencetti, Iacopo Iacopini, Vito Latora, Maxime Lucas, Alice Patania, Jean-Gabriel Young, and Giovanni Petri (2020). "Networks beyond pairwise interactions: Structure and dynamics". In: *Physics Reports* 874, pp. 1–92.

Baudena, Alberto, Enrico Ser-Giacomi, Isabel Jalón-Rojas, François Galgani, and Maria Luiza Pedrotti (2022). "The streaming of plastic in the Mediterranean Sea". In: *Nature Communications* 13.1, p. 2981.

Bedoya-Soto, Juan Mauricio, Germán Poveda, Kevin E Trenberth, and Jorge Julián Vélez-Upegui (2019). "Interannual hydroclimatic variability and the 2009–2011 extreme ENSO phases in Colombia: from Andean glaciers to Caribbean lowlands". In: *Theoretical and Applied Climatology* 135, pp. 1531–1544.

Berg, Alexis, Benjamin R Lintner, Kirsten Findell, Sonia I Seneviratne, Bart van Den Hurk, Agnès Ducharne, Frédérique Chéruy, Stefan Hagemann, David M Lawrence, Sergey Malyshev, et al. (2015). "Interannual coupling between summertime surface temperature and precipitation over land: Processes and implications for climate change". In: *Journal of Climate* 28.3, pp. 1308–1328.

Bhaskar, Ankush, Durbha Sai Ramesh, Geeta Vichare, Triven Koganti, and S Gurubaran (2017). "Quantitative assessment of drivers of recent global temperature variability: an information theoretic approach". In: *Climate Dynamics* 49.11-12, pp. 3877–3886.

Boccaletti, Stefano, Vito Latora, Yamir Moreno, Martin Chavez, and D-U Hwang (2006). "Complex networks: Structure and dynamics". In: *Physics reports* 424.4-5, pp. 175–308.

Boers, Niklas, Bedartha Goswami, Aljoscha Rheinwalt, Bodo Bookhagen, Brian Hoskins, and Jürgen Kurths (2019). "Complex networks reveal global pattern of extreme-rainfall teleconnections". In: *Nature* 566.7744, pp. 373–377.

Boltzmann, Ludwig (1896). "Entgegnung auf die wärmetheoretischen Betrachtungen des Hrn. E. Zermelo". In: *Annalen der physik* 293.4, pp. 773–784.

— (1974). "The second law of thermodynamics". In: *Theoretical physics and philosophical problems: selected writings*. Springer, pp. 13–32.

Botero, Hernan and Andrew P Barnes (2022). "The effect of ENSO on common bean production in Colombia: a time series approach". In: *Food Security* 14.6, pp. 1417–1430.

Brotzge, Jerald A, Don Berchoff, DaNa L Carlis, Frederick H Carr, Rachel Hogan Carr, Jordan J Gerth, Brian D Gross, Thomas M Hamill, Sue Ellen Haupt, Neil Jacobs, et al. (2023). "Challenges and Opportunities in Numerical Weather Prediction". In: *Bulletin of the American Meteorological Society* 104.3, E698–E705.

Builes-Jaramillo, Alejandro, Juliana Valencia, and Hernán D Salas (2023). "The influence of the El Niño-Southern Oscillation phase transitions over the northern South America hydroclimate". In: *Atmospheric Research* 290, p. 106786.

Cai, Wenju, Michael J McPhaden, Alice M Grimm, Regina R Rodrigues, Andréa S Taschetto, René D Garreaud, Boris Dewitte, Germán Poveda, Yoo-Geun Ham, Agus Santoso, et al. (2020). "Climate

impacts of the El Niño–southern oscillation on South America". In: *Nature Reviews Earth & Environment* 1.4, pp. 215–231.

Caldarelli, Guido (2007). *Scale-free networks: complex webs in nature and technology*. Oxford Finance.

Canchala, Teresita, Wilfredo Alfonso-Morales, Wilmar Loaiza Cerón, Yesid Carvajal-Escobar, and Eduardo Caicedo-Bravo (2020a). "Teleconnections between monthly rainfall variability and large-scale climate indices in Southwestern Colombia". In: *Water* 12.7, p. 1863.

Canchala, Teresita, Wilmar Loaiza Cerón, Felix Frances, Yesid Carvajal-Escobar, Rita Valéria Andreoli, Mary Toshie Kayano, Wilfredo Alfonso-Morales, Eduardo Caicedo-Bravo, and Rodrigo Augusto Ferreira de Souza (2020b). "Streamflow variability in colombian pacific basins and their teleconnections with climate indices". In: *Water* 12.2, p. 526.

Cerón, Wilmar L, Mary T Kayano, Rita V Andreoli, Teresita Canchala, Yesid Carvajal-Escobar, and Wilfredo Alfonso-Morales (2021). "Rainfall variability in southwestern Colombia: changes in ENSO-related features". In: *Pure and Applied Geophysics* 178.3, pp. 1087–1103.

Clausius, Rudolph (1866). "I. On the determination of the energy and entropy of a body". In: *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science* 32.213, pp. 1–17.

Claussen, Martin, L Mysak, A Weaver, Michel Crucifix, Thierry Fichefet, M-F Loutre, Shlomo Weber, Joseph Alcamo, Vladimir Alexeev, André Berger, et al. (2002). "Earth system models of intermediate complexity: closing the gap in the spectrum of climate system models". In: *Climate dynamics* 18, pp. 579–586.

Cordoba-Machado, Samir, Reiner Palomino-Lemus, Sonia Raquel Gamiz-Fortis, Yolanda Castro-Díez, and Maria Jesus Esteban-Parra (2015). "Assessing the impact of El Niño Modoki on seasonal precipitation in Colombia". In: *Global and Planetary Change* 124, pp. 41–61.

Córdoba-Machado, Samir, Reiner Palomino-Lemus, Sonia Raquel Gámiz-Fortis, Yolanda Castro-Díez, and María Jesús Esteban-Parra (2015). "Influence of tropical Pacific SST on seasonal precipitation in Colombia: prediction using El Niño and El Niño Modoki". In: *Climate Dynamics* 44, pp. 1293–1310.

Cover, Thomas M and Joy A Thomas (1991). "Network information theory". In: *Elements of information theory*, pp. 374–458.

Deza, J Ignacio and Hisham Ihshaish (2015). "The construction of complex networks from linear and nonlinear measures–Climate Networks". In: *Procedia Computer Science* 51, pp. 404–412.

Díaz-Granados Ortiz, Mario A, Juan D Navarrete González, and Tatiana Suárez López (2005). "Páramos: hidrosistemas sensibles". In: *Revista de ingeniería* 22, pp. 64–75.

Dijkstra, Henk A, Emilio Hernández-García, Cristina Masoller, and Marcelo Barreiro (2019). *Networks in climate*. Cambridge University Press.

Donges, Jonathan F, Jobst Heitzig, Boyan Beronov, Marc Wiedermann, Jakob Runge, Qing Yi Feng, Liubov Tupikina, Veronika Stolbova, Reik V Donner, Norbert Marwan, et al. (2015). "Unified functional network and nonlinear time series analysis for complex systems science: The pyunicorn package". In: *Chaos: An Interdisciplinary Journal of Nonlinear Science* 25.11.

Donges, Jonathan F, Yong Zou, Norbert Marwan, and Jürgen Kurths (2009). "Complex networks in climate dynamics: Comparing linear and nonlinear network construction methods". In: *The European Physical Journal Special Topics* 174.1, pp. 157–179.

Donner, Reik, Susana Barbosa, Jürgen Kurths, and Norbert Marwan (2009). "Understanding the Earth as a Complex System–recent advances in data analysis and modelling in Earth sciences". In: *The European Physical Journal Special Topics* 174, pp. 1–9.

Easterling, David R, Gerald A Meehl, Camille Parmesan, Stanley A Changnon, Thomas R Karl, and Linda O Mearns (2000). "Climate extremes: observations, modeling, and impacts". In: *science* 289.5487, pp. 2068–2074.

Ebel, Holger, Lutz-Ingo Mielsch, and Stefan Bornholdt (2002). "Scale-free topology of e-mail networks". In: *Physical review E* 66.3, p. 035103.

Erdős, Paul, Alfréd Rényi, et al. (1959). "On random graphs I". In: *Publ. Math. Debrecen* 6, pp. 290–297.

— (1960). "On the evolution of random graphs". In: *Publ. math. inst. hung. acad. sci* 5.1, pp. 17–60.

Escobar, Maritza, Isabel Hoyos, Raquel Nieto, and Juan Camilo Villegas (2022). "The importance of continental evaporation for precipitation in Colombia: A baseline combining observations from stable isotopes and modelling moisture trajectories". In: *Hydrological Processes* 36.6, e14595.

Espinoza, Jhan Carlo, René Garreaud, Germán Poveda, Paola A Arias, Jorge Molina-Carpio, Mariano Masiokas, Maximiliano Viale, and Lucia Scaff (2020). "Hydroclimate of the Andes part I: Main climatic features". In: *Frontiers in Earth Science* 8, p. 64.

FedeCafeteros, Federación Nacional de Cafeteros (2020). *Coffee Grower Services*. URL: https://federac\iondecafeteros.org/wp/coffee-grower-services/?lang=en.

Fortunato, Santo (2010). "Community detection in graphs". In: *Physics reports* 486.3-5, pp. 75–174.

Freitas, Cristopher GS, Andre LL Aquino, Heitor S Ramos, Alejandro C Frery, and Osvaldo A Rosso (2019). "A detailed characterization of complex networks using Information Theory". In: *Scientific reports* 9.1, p. 16689.

Froyland, Gary, Robyn M Stuart, and Erik van Sebille (2014). "How well-connected is the surface of the global ocean?" In: *Chaos: An Interdisciplinary Journal of Nonlinear Science* 24.3.

Ghil, Michael (2002). "Natural climate variability". In: *Encyclopedia of global environmental change* 1, pp. 544–549.

Ghil, Michael, Mickaël D Chekroun, and Eric Simonnet (2008). "Climate dynamics and fluid mechanics: Natural variability and related uncertainties". In: *Physica D: Nonlinear Phenomena* 237.14-17, pp. 2111–2126.

Ghil, Michael and Valerio Lucarini (2020). "The physics of climate variability and climate change". In: *Reviews of Modern Physics* 92.3, p. 035002.

Golbeck, Jennifer (2015). *Introduction to social media investigation: A hands-on approach*. Syngress.

Grimm, Alice M and Renata G Tedeschi (2009). "ENSO and extreme rainfall events in South America". In: *Journal of Climate* 22.7, pp. 1589–1609.

Hagan, Daniel Fiifi Tawia, Guojie Wang, X San Liang, and Han AJ Dolman (2019). "A time-varying causality formalism based on the Liang–Kleeman information flow for analyzing directed interactions in nonstationary climate systems". In: *Journal of Climate* 32.21, pp. 7521–7537.

Hayes, SP, MJ McPhaden, and JM Wallace (1989). "The influence of sea-surface temperature on surface wind in the eastern equatorial Pacific: Weekly to monthly variability". In: *Journal of Climate* 2.12, pp. 1500–1506.

Heisz, Jennifer J, Judith M Shedden, and Anthony R McIntosh (2012). "Relating brain signal variability to knowledge representation". In: *Neuroimage* 63.3, pp. 1384–1392.

Hersbach, Hans, Bill Bell, Paul Berrisford, Per Dahlgren, András Horányi, J Munoz-Sebater, Julien Nicolas, Raluca Radu, Dinand Schepers, Adrian Simmons, et al. (2020a). "The ERA5 Global Reanalysis: achieving a detailed record of the climate and weather for the past 70 years". In: *European geophysical union general assembly*, pp. 3–8.

Hersbach, Hans, Bill Bell, Paul Berrisford, Shoji Hirahara, András Horányi, Joaquín Muñoz-Sabater, Julien Nicolas, Carole Peubey, Raluca Radu, Dinand Schepers, et al. (2020b). "The ERA5 global reanalysis". In: *Quarterly Journal of the Royal Meteorological Society* 146.730, pp. 1999–2049.

Hlinka, Jaroslav, David Hartman, Martin Vejmelka, Jakob Runge, Norbert Marwan, Jürgen Kurths, and Milan Paluš (2013). "Reliability of inference of directed climate networks using conditional mutual information". In: *Entropy* 15.6, pp. 2023–2045.

Hoyos, Isabel, Astrid Baquero-Bernal, and Stefan Hagemann (2013a). "How accurately are climatological characteristics and surface water and energy balances represented for the Colombian Caribbean Catchment Basin?" In: *Climate dynamics* 41, pp. 1269–1290.

Hoyos, Isabel, Astrid Baquero-Bernal, Daniela Jacob, and Boris A Rodríguez (2013b). "Variability of extreme events in the Colombian Pacific and Caribbean catchment basins". In: *Climate dynamics* 40, pp. 1985–2003.

Hoyos, Isabel, J Cañón-Barriga, T Arenas-Suárez, F Dominguez, and BA Rodríguez (2019). "Variability of regional atmospheric moisture over Northern South America: patterns and underlying phenomena". In: *Climate Dynamics* 52, pp. 893–911.

Hoyos, Isabel, F Dominguez, J Cañón-Barriga, JA Martínez, R Nieto, L Gimeno, and PA Dirmeyer (2018). "Moisture origin and transport processes in Colombia, northern South America". In: *Climate Dynamics* 50, pp. 971–990.

Hulme, Mike, Elaine M Barrow, Nigel W Arnell, Paula A Harrison, Timothy C Johns, and Thomas E Downing (1999). "Relative impacts of human-induced climate change and natural climate variability". In: *Nature* 397.6721, pp. 688–691.

IDEAM and Ministerio de Ambiente Vivienda y Desarrollo Territorial (2005). *Atlas climatológico de Colombia*. IDEAM (Instituto de Hidrología, Meteorología y Estudios Ambientales).

Kleeman, Richard (2002). "Measuring dynamical prediction utility using relative entropy". In: *Journal of the atmospheric sciences* 59.13, pp. 2057–2072.

— (2011). "Information theory and dynamical system predictability". In: *Entropy* 13.3, pp. 612–649.

Kousky, Vernon E, Mary T Kagano, and Iracema FA Cavalcanti (1984). "A review of the Southern Oscillation: oceanic-atmospheric circulation changes and related rainfall anomalies". In: *Tellus A* 36.5, pp. 490–504.

Krishnamurthy, V (2019). "Predictability of weather and climate". In: *Earth and Space Science* 6.7, pp. 1043–1056.

Kucharski, Fred, In-Sik Kang, David Straus, and Martin P King (2010). "Teleconnections in the atmosphere and oceans". In: *Bulletin of the American Meteorological Society* 91.3, pp. 381–383.

Kullback, Solomon and Richard A Leibler (1951). "On information and sufficiency". In: *The annals of mathematical statistics* 22.1, pp. 79–86.

Lee, Hoesung, Katherine Calvin, Dipak Dasgupta, Gerhard Krinner, Aditi Mukherji, Peter Thorne, Christopher Trisos, José Romero, Paulina Aldunce, Ko Barret, et al. (2023). "IPCC, 2023: Climate Change 2023: Synthesis Report, Summary for Policymakers. Contribution of Working Groups I, II and III to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change [Core Writing Team, H. Lee and J. Romero (eds.)]. IPCC, Geneva, Switzerland." In.

Li, Rui-qi, Shi-wen Sun, Yi-lin Ma, Li Wang, and Cheng-yi Xia (2015). "Effect of clustering on attack vulnerability of interdependent scale-free networks". In: *Chaos, Solitons & Fractals* 80, pp. 109–116.

Liang, X. San (2008). "Information flow within stochastic dynamical systems". In: *Physical Review E—Statistical, Nonlinear, and Soft Matter Physics* 78.3, p. 031113.

— (2013a). "Local predictability and information flow in complex dynamical systems". In: *Physica D: Nonlinear Phenomena* 248, pp. 1–15.

— (2013b). "The Liang-Kleeman information flow: Theory and applications". In: *Entropy* 15.1, pp. 327–360.

— (2014). "Unraveling the cause-effect relation between time series". In: *Physical Review E* 90.5, p. 052150.

— (2015). "Normalizing the causality between time series". In: *Phys. Rev. E* 92 (2), p. 022126. DOI: 10.1103/PhysRevE.92.022126.

— (2016). "Information flow and causality as rigorous notions ab initio". In: *Physical Review E* 94.5, p. 052201.

Liao, Xuhong, Athanasios V Vasilakos, and Yong He (2017). "Small-world human brain networks: perspectives and challenges". In: *Neuroscience & Biobehavioral Reviews* 77, pp. 286–300.

Lieb, Elliott H and Jakob Yngvason (1999). "The physics and mathematics of the second law of thermodynamics". In: *Physics Reports* 310.1, pp. 1–96.

Liu, Teng, Dean Chen, Lan Yang, Jun Meng, Zanchenling Wang, Josef Ludescher, Jingfang Fan, Saini Yang, Deliang Chen, Jürgen Kurths, et al. (2023). "Teleconnections among tipping elements in the Earth system". In: *Nature Climate Change* 13.1, pp. 67–74.

Lorenz, Edward N (1963). "Deterministic nonperiodic flow". In: *Journal of atmospheric sciences* 20.2, pp. 130–141.

Majhi, Soumen, Matjaž Perc, and Dibakar Ghosh (2022). "Dynamics on higher-order networks: A review". In: *Journal of the Royal Society Interface* 19.188, p. 20220043.

Martinez, J Alejandro, Paola A Arias, Chris Castro, Hsin-I Chang, and Carlos A Ochoa-Moya (2019). "Sea surface temperature-related response of precipitation in northern South America according to a WRF multi-decadal simulation". In: *International Journal of Climatology* 39.4, pp. 2136–2155.

Marwan, Norbert, Jonathan F Donges, Yong Zou, Reik V Donner, and Jürgen Kurths (2009). "Complex network approach for recurrence analysis of time series". In: *Physics Letters A* 373.46, pp. 4246–4254.

Mesa S, Oscar J, Germán Poveda J, and Luis F Carvajal S (1997). *Introducción al clima de Colombia*. Universidad Nacional de Colombia.

Mitchell, Melanie (2006). "Complex systems: Network thinking". In: *Artificial intelligence* 170.18, pp. 1194–1212.

Morales, José, Paola Arias, and John Martínez (2017). "Role of Caribbean low-level jet and Choco jet in the transport of moisture patterns towards Central America". In.

Morrison, Monica Ainhorn and Peter Lawrence (2023). "Understanding Model-Based Uncertainty in Climate Science". In: *Handbook of Philosophy of Climate Change*. Springer, pp. 1–21.

Myers, Norman, Russell A Mittermeier, Cristina G Mittermeier, Gustavo AB Da Fonseca, and Jennifer Kent (2000). "Biodiversity hotspots for conservation priorities". In: *Nature* 403.6772, pp. 853–858.

Negre, Christian FA, Uriel N Morzan, Heidi P Hendrickson, Rhitankar Pal, George P Lisi, J Patrick Loria, Ivan Rivalta, Junming Ho, and Victor S Batista (2018). "Eigenvector centrality for characterization of protein allosteric pathways". In: *Proceedings of the National Academy of Sciences* 115.52, E12201–E12208.

Newman, Mark (2018). *Networks*. Oxford university press.

Newman, Mark, Albert-László Barabási, and Duncan J Watts (2011). *The structure and dynamics of networks*. Princeton university press.

Newman, Mark E. J. (2010). "Measures and Metrics in Networks". In: *Networks: An Introduction*. Oxford University Press. Chap. 7.

Newman, Mark EJ (2003). "The structure and function of complex networks". In: *SIAM review* 45.2, pp. 167–256.

Pabón, José Daniel (2003). "El cambio climático global y su manifestación en Colombia". In: *Cuadernos de Geografía: Revista Colombiana de Geografía* 12, pp. 111–119.

Pabón-Caicedo, José Daniel, Paola A Arias, Andrea F Carril, Jhan Carlo Espinoza, Lluís Fita Borrel, Katerina Goubanova, Waldo Lavado-Casimiro, Mariano Masiokas, Silvina Solman, and Ricardo Villalba (2020). "Observed and projected hydroclimate changes in the Andes". In: *Frontiers in Earth Science* 8, p. 61.

Pacific, Institute (2018). *Data Table: Total renewable freshwater supply by country*.

Pagani, Giuliano Andrea and Marco Aiello (2014). "Power grid complex network evolutions for the smart grid". In: *Physica A: Statistical Mechanics and its Applications* 396, pp. 248–266.

Pastor-Satorras, Romualdo and Alessandro Vespignani (2001). "Epidemic spreading in scale-free networks". In: *Physical review letters* 86.14, p. 3200.

Pearl, Judea (1997). "The new challenge: From a century of statistics to the age of causation". In: *Computing Science and Statistics*, pp. 415–423.

Peixoto, José Pinto and Abraham H Oort (1992). "Physics of climate". In.

Pinault, Jean-Louis (2022). "A review of the role of the oceanic Rossby waves in climate variability". In: *Journal of Marine Science and Engineering* 10.4, p. 493.

Pósfai, Márton and Albert-László Barabási (2016). *Network Science*. Cambridge University Press Cambridge, UK:

Poveda, Germán, Diana M. Álvarez, and Óscar A. Rueda (2011). "Hydro-climatic variability over the Andes of Colombia associated with ENSO: a review of climatic processes and their impact on one of the Earth's most important biodiversity hotspots". In: *Climate Dynamics* 36.11, pp. 2233–2249.

Poveda, Germán, Jhan Carlo Espinoza, Manuel D Zuluaga, Silvina A Solman, René Garreaud, and Peter J Van Oevelen (2020). "High impact weather events in the Andes". In: *Frontiers in Earth Science* 8, p. 162.

Poveda, Germán and Oscar Mesa (1999). "La corriente de chorro superficial del Oeste ("del Chocó") y otras dos corrientes de chorro en Colombia: climatología y variabilidad durante las fases del ENSO". In: *Revista Académica Colombiana de Ciencia* 23.89, pp. 517–528.

Poveda, Germán, Oscar Mesa, Paula Agudelo, Juan Álvarez, Paola Arias, Hernán Moreno, Luis Salazar, Vladimir Toro, and Sara Vieira (2002). "Influencia del ENSO, oscilación Madden-Julian, ondas del este, huracanes y fases de la luna en el ciclo diurno de precipitación en los Andes Tropicales de Colombia". In: *Meteorología Colombiana* 5.0124-6984, pp. 3–12.

Poveda, Germán, Peter R Waylen, and Roger S Pulwarty (2006). "Annual and inter-annual variability of the present climate in northern South America and southern Mesoamerica". In: *Palaeogeography, Palaeoclimatology, Palaeoecology* 234.1, pp. 3–27.

Reboita, Michelle Simões, Tércio Ambrizzi, Natália Machado Crespo, Lívia Márcia Mosso Dutra, Glauber Willian de S Ferreira, Amanda Rehbein, Anita Drumond, Rosmeri Porfírio da Rocha, and Christie Andre de Souza (2021). "Impacts of teleconnection patterns on South America climate". In: *Annals of the New York Academy of Sciences* 1504.1, pp. 116–153.

Restrepo, JD and B Kjerfve (2000). "Magdalena river: interannual variability (1975–1995) and revised water discharge and sediment load estimates". In: *Journal of hydrology* 235.1-2, pp. 137–149.

Rincón, Laura Natalia Garavito (2015). "Los páramos en Colombia, un ecosistema en riesgo". In: *Ingeniare* 19, pp. 127–136.

Ruiz, Daniel, Hernán Alonso Moreno, María Elena Gutiérrez, and Paula Andrea Zapata (2008). "Changing climate and endangered high mountain ecosystems in Colombia". In: *Science of the total environment* 398.1-3, pp. 122–132.

Sakamoto, Meiry Sayuri, Tércio Ambrizzi, Germán Poveda, et al. (2011). "Moisture sources and life cycle of convective systems over western Colombia". In: *Advances in Meteorology* 2011.

Sánchez-Cuervo, Ana María, T Mitchell Aide, Matthew L Clark, and Andrés Etter (2012). "Land cover change in Colombia: surprising forest recovery trends between 2001 and 2010". In.

Sanderson, Michael, Katie Hodge, José Ricardo Cure, Daniel Rodríguez, Luigi Ponti, Andrew Paul Gutierrez, et al. (2022). "Impacts of the ENSO cycle on climate and coffee production in Colombia". In: *EGU General Assembly Conference Abstracts*, EGU22–13513.

Sardeshmukh, Prashant D and Brian J Hoskins (1988). "The generation of global rotational flow by steady idealized tropical divergence". In: *Journal of the Atmospheric Sciences* 45.7, pp. 1228–1251.

Schneider, Stephen H and Robert E Dickinson (1974). "Climate modeling". In: *Reviews of Geophysics* 12.3, pp. 447–493.

Schneider, Tapio (2006). "The general circulation of the atmosphere". In: *Annu. Rev. Earth Planet. Sci.* 34, pp. 655–688.

Schneider, Tapio, Tobias Bischoff, and Gerald H Haug (2014). "Migrations and dynamics of the intertropical convergence zone". In: *Nature* 513.7516, pp. 45–53.

Schreiber, Thomas (2000). "Measuring information transfer". In: *Physical review letters* 85.2, p. 461.

Shackley, Simon, Peter Young, Stuart Parkinson, and Brian Wynne (1998). "Uncertainty, complexity and concepts of good science in climate change modelling: are GCMs the best tools?" In: *Climatic change* 38, pp. 159–205.

Shannon, Claude E (1938). "A symbolic analysis of relay and switching circuits". In: *Electrical Engineering* 57.12, pp. 713–723.

Shannon, Claude Elwood (1948). "A mathematical theory of communication". In: *The Bell system technical journal* 27.3, pp. 379–423.

Solman, Silvina A (2013). "Regional climate modeling over South America: a review". In: *Advances in Meteorology* 2013, pp. 1–13.

Souza, EB DE and Tércio Ambrizzi (2002). "ENSO impacts on the South American rainfall during 1980s: Hadley and Walker circulation". In: *Atmósfera* 15.2, pp. 105–120.

Stan, Cristiana, David M Straus, Jorgen S Frederiksen, Hai Lin, Eric D Maloney, and Courtney Schumacher (2017). "Review of tropical-extratropical teleconnections on intraseasonal time scales". In: *Reviews of Geophysics* 55.4, pp. 902–937.

Steinhaeuser, Karsten, Nitesh V Chawla, and Auroop R Ganguly (2010). "Complex Networks in Climate Science: Progress, Opportunities and Challenges." In: *CIDU*, pp. 16–26.

— (2011). "Complex networks as a unified framework for descriptive analysis and predictive modeling in climate science". In: *Statistical Analysis and Data Mining: The ASA Data Science Journal* 4.5, pp. 497–511.

Strogatz, Steven H (2001). "Exploring complex networks". In: *nature* 410.6825, pp. 268–276.

Trenberth, Kevin E, Grant W Branstator, David Karoly, Arun Kumar, Ngar-Cheung Lau, and Chester Ropelewski (1998). "Progress during TOGA in understanding and modeling global teleconnections associated with tropical sea surface temperatures". In: *Journal of Geophysical Research: Oceans* 103.C7, pp. 14291–14324.

Trenberth, Kevin E and Dennis J Shea (2005). "Relationships between precipitation and surface temperature". In: *Geophysical Research Letters* 32.14.

Tsonis, Anastasios A and Paul J Roebber (2004). "The architecture of the climate network". In: *Physica A: Statistical Mechanics and its Applications* 333, pp. 497–504.

Tsonis, Anastasios A and Kyle L Swanson (2008). "Topology and predictability of El Nino and La Nina networks". In: *Physical Review Letters* 100.22, p. 228502.

Tsonis, Anastasios A, Kyle L Swanson, and Paul J Roebber (2006). "What do networks have to do with climate?" In: *Bulletin of the American Meteorological Society* 87.5, pp. 585–596.

Tsonis, Anastasios A, Kyle L Swanson, and Geli Wang (2008). "On the role of atmospheric teleconnections in climate". In: *Journal of Climate* 21.12, pp. 2990–3001.

Urrea, Viviana, Andrés Ochoa, and Oscar Mesa (2019). "Seasonality of rainfall in Colombia". In: *Water Resources Research* 55.5, pp. 4149–4162.

Vannitsem, Stéphane and X. San Liang (2022). "Dynamical dependencies at monthly and interannual time scales in the climate system: Study of the North Pacific and Atlantic regions". In: *Tellus A* 74, pp. 141–158.

Volkenstein, Mikhail V (2009). *Entropy and information*. Vol. 57. Springer Science & Business Media.

Wallace, John M, TP Mitchell, and C Deser (1989). "The influence of sea-surface temperature on surface wind in the eastern equatorial Pacific: Seasonal and interannual variability". In: *Journal of Climate* 2.12, pp. 1492–1499.

Wang, Yu, Eshwar Ghumare, Rik Vandenberghe, and Patrick Dupont (2017). "Comparison of different generalizations of clustering coefficient and local efficiency for weighted undirected graphs". In: *Neural computation* 29.2, pp. 313–331.

Watts, Duncan J (2004). *Small worlds: the dynamics of networks between order and randomness*. Vol. 36. Princeton university press.

Watts, Duncan J and Steven H Strogatz (1998). "Collective dynamics of 'small-world'networks". In: *nature* 393.6684, pp. 440–442.

XM, SAESP (2019). *Reporte integral de sostenibilidad, operación y mercado 2018*.

Yepes, Johanna, Germán Poveda, John F Mejía, Leonardo Moreno, and Carolina Rueda (2019). "Chocojex: A research experiment focused on the Chocó low-level jet over the far eastern Pacific and western Colombia". In: *Bulletin of the American Meteorological Society* 100.5, pp. 779–796.